

# The Origin and Evolution of Viruses

Vadim I. Agol

**Abstract** The lecture covers three main topics: (i) Viruses: properties, place in the living world, and possible origin; (ii) Molecular basis of viral variability and evolution; and (iii) Evolution of viral pathogenicity and emerging viral infections.

**Keywords** Selfish genetic elements, replication, transcription, mutation, recombination, plasmids, pathogenicity, fibroma/myxoma virus, influenza virus, SARS-coronavirus, human immunodeficiency virus

## 1 Viruses: Properties, Place in the Living World, and Possible Origin

Viruses are selfish genetic elements existing in two forms, the passive extracellular form represented by a molecule(s) of nucleic acid (DNA or RNA) usually enclosed into a protein-containing shell, and the active intracellular replicating and gene-expressing form. Viruses are the most abundant biological entities on the Earth. According to some estimates, our planet harbors  $10^{31}$  virus particles (Breitbart and Rohwer, 2005). There are  $\sim 2 \times 10^8$  ton carbon in marine viruses only (=carbon content in  $7 \times 10^7$  blue whales). If stretched end to end, marine viruses would span  $\sim 10^7$  light years ( $\sim 100$  times the distance across our galaxy) (Suttle, 2005). The shape of virus particles is highly variable (e.g., rod-like, thread-like, with cubic symmetry, and many others). The viral genomes may be represented by either single-stranded or double-stranded DNA or RNA molecules, which may have either linear

---

Vadim I. Agol

M. P. Chumakov Institute of Poliomyelitis & Viral Encephalitides

Russian Academy of Medical Sciences and

A. N. Belozersky Institute of Physical-Chemical Biology

M. V. Lomonosov Moscow State University

Moscow Russia

or circular structures. The number of viral genes may vary from one (e.g., in hepatitis delta virus) to  $\sim 1,200$  (in mimivirus). In the latter case, the size of the viral genome is several-fold larger than that in the smallest cellular organisms.

A distinctive property of viruses is the absence of protein-synthesizing and energy-generating machineries. As a result, viruses are strictly dependent on the host cell. Another fundamental distinction between viruses and cells is the mode of their multiplication. Viral reproduction is based on disjunctive mechanisms, whereupon virus-specific proteins and genomic nucleic acids are accumulated as separate pools, from which multiple mature progeny viral particles (or their core nucleoproteins) are eventually assembled. On the other hand, cellular multiplication involves (usually binary) division. During a part of their life cycle, the genome of some viruses may be integrated into the host cell genome, existing and dividing as a cellular chromosomal segment.

Depending on the type of genome and mechanisms of its replication and expression, several major types of viral “strategies” can be discerned (Baltimore, 1971; Agol, 1974). Remarkably, all the theoretically possible replication/transcription systems based on the principle of complementarity appear to be exploited by viruses. The replication/transcription system used by genomes of cellular organisms comprises only a minor subset of the systems exploited by viruses.

Viruses have very close relatives, selfish genetic elements (DNA or RNA): viroids (small infectious non-coding circular RNA), plasmids (various non-chromosomal DNA or RNA elements), transposons (DNA elements moving between different positions of a genome; mobile genetic elements), retrotransposons (mobile DNA elements exploiting reverse transcription) and some others. Viruses alike, they are fully dependent on cellular translation and energy-generating machineries and use the same expression and replication mechanisms as viruses do. They may be integrated into cellular DNA or RNA or exist separately, may be amplified in a cell, move between different parts of the genetic material of a cells or migrate between cells. Therefore, these elements can be combined with viruses into a common domain of life. In distinction from viruses, however, they lack protein coats and stable extracellular forms.

There are two major problems related to the origin of viruses. What are relationships between viruses and cellular organisms? In other words, is there a place(s) for viruses on universal trees of life proposed for its three kingdoms, Archea, Eubacteria, and Eukarya? And secondly, are viruses monophyletic, i.e., is there a single evolutionary tree or branch of viruses? During the first century after discovery of viruses, several purely speculative but quite imaginative hypotheses had been proposed: according to them, viruses originated from escaped (“mad”) cellular genes, or from degenerated cells, or they are descendants of precellular genetic elements.

Modern hypotheses of viral origin are based on two major developments of the molecular biology: discovery of ribozymes (RNA-based enzymes) and formulation of the “RNA World” theory (RNA had been “invented” before proteins and DNA), on the one hand, and achievements of genomics (determination of the nucleotide sequences of a great number of cellular and viral genomes), on the other. These

hypotheses postulate a very significant contribution of viruses to the genetic information of cellular organisms.

According to the “Three viruses, three domains” hypothesis proposed by P. Forterre, the cellular Last Universal Common Ancestor (LUCA) had an RNA genome and harbored a variety of already created RNA-containing viruses. During the next step, DNA and DNA viruses had been invented. Three distinct DNA viruses, which had infected RNA genome-containing cells, gave rise to the three distinct domains of life, bacteria, archaea, and eukarya (Forterre, 2006). A detailed hypothetical evolutionary scenario was described by E. Koonin and colleagues (Koonin and Martin, 2005; Koonin et al., 2006). The authors suggest that life has originated in inorganic compartments serving as surrogate cells. RNA molecules with enzymatic activities (ribozymes) could attain there relatively high concentrations and could move between compartments and “infect” them. At this precellular stage, several fundamental inventions had been consecutively made. The invention of proteins led to the appearance of viruses with positive single-strand and with double-strand RNA genomes. The invention of DNA resulting from RNA-dependent DNA synthesis led to the appearance of retroviruses, retroid viruses and retroposons. Finally, DNA-dependent DNA synthesis had been invented, giving rise to the appearance of viruses with double-strand DNA genomes and DNA plasmids. Only after achieving such genetic variability and richness, the “pre-archeal” and “pre-bacterial” inorganic compartments had been replaced by cellular forms, which retained numerous ancient viruses. The engulfment of a bacterium by an archaeon was the starting point of eukarya, which acquired viruses of both of their “parents”. According to Koonin, the major classes of viruses do not have a common origin in the traditional sense but neither are they unrelated. On the other hand, a very significant proportion of the cellular genetic material has evolutionary relatedness to the genetic composition of the ancient world of RNA viruses. It is estimated that well over a half of mammalian genome is inherited from viruses. In other words, “the tree of life and its root are immersed in a viral ocean” (Bamford, 2003).

The above scenarios also suggest that modern viruses have inherited molecular mechanisms that have disappeared from modern DNA cells. That is why transcription and replication mechanisms in the viral world are more diverse than those in the cellular world. It cannot be excluded that many yet unknown molecular mechanisms exist in the current viral world. The exploration of the viral diversity is one of the major challenges of biology in this century.

## 2 Molecular Basis of Viral Variability and Evolution

After their “birth”, viral genomes had been and are still evolving by accumulation of point mutations and genome rearrangements (duplications, deletions, recombination). The frequency of mutations in DNA viruses may be several orders of magnitude lower than in RNA viruses, whose replicative enzyme, the RNA-dependent RNA polymerase, lacks proof-reading activity. The error frequency in RNA viruses

is also variable, and, in some of them, any newly synthesized genomic RNA molecule contains on an average a mutation. Such viruses exist on the edge of mutational catastrophe: even a few-fold increase in the error frequency may result in population extinction. Nevertheless, a certain level of replicative infidelity is highly advantageous for viruses since it confers a potential for changing and adaptation. For example, replicative errors are one of the major factors contributing to the development of drug-resistance and hence to the scarcity of efficacious antiviral drugs. Interestingly, a single point mutation in the poliovirus RNA-dependent RNA polymerase increases its fidelity several-fold. However, the more accurate mutant is less fit: not only it develops drug-resistance less readily, but is also less neurovirulent (Pfeiffer and Kirkegaard, 2003, 2005; Vignuzzi et al., 2006). A heterogeneous population, as a group, may have advantages over a more homogeneous population. The advantages may be due not only to a greater potential for adaptation but also to the possibility of mutual cooperation.

The genome instability challenges the very identity of viruses: how does virus X remain virus X? However, despite their intrinsic instability, viral genomes, even in the extreme case of RNA viruses, are remarkably robust. Several factors contribute to the genetic stability of viruses in nature. Adverse mutations are eliminated by negative selection, whereas fitness-increasing mutations are nearly not-existing in constant environment. Adaptive mutations may be positively selected upon environmental changes (including infection of a novel host species). The only major factor favoring accumulation of neutral mutations is bottlenecking (stochastic picking up a single viral particle, or a few of them, from a heterogeneous population), which is quite common during natural transmission of viruses. Consecutive bottlenecking events may result in a more or less marked decrease in viral fitness (the so called Muller's ratchet) because viral genomes picked up by chance from a heterogeneous population may well harbor detrimental mutations.

To what extent is the Muller's ratchet inevitable and irreversible? The robustness of viruses may be studied experimentally by damaging their genome (e.g., by introducing point mutations or deletions, randomization of portions of the nucleotide sequence, or even breaking the genome apart) followed by investigating viral viability and fitness as well as structure and stability of the mutated viral genome. A wealth of relevant information concerning numerous viruses has already been accumulated in literature. Some examples of such "interrogation under torture" carried out with picornaviruses in the author's laboratory are presented below.

The Theiler's murine encephalomyelitis virus (TMEV) is highly pathogenic for mice under certain conditions: a few wild type viral particles injected intracerebrally kill mice. For efficient viral reproduction in the brain, interaction of a translational control element (IRES) in the 5'-noncoding region (5NCR) of the viral RNA with a host protein, neural polypyrimidine tract-binding protein (nPTB), is essential (Pilipenko et al., 2001). A single point mutation in one of the nPTB-binding sites of the viral RNA led to a dramatic decrease in viral neurovirulence:  $\sim 10^4$  more mutant particles were now needed to kill a mouse. However, wild-type level of neurovirulence could readily be restored by either reversion or pseudoreversions during viral reproduction, e.g., by generation of a new nPTB-binding site in the viral 5NCR due

to a compensating point mutation (Pilipenko et al., 2001). An 8-nt deletion in the 5NCR of poliovirus RNA severely impaired viral fitness and nearly killed the virus. However, this defect could be readily fixed by natural pseudoreversions of at least three different kinds, including point mutations, insertions and extended deletions (Pilipenko et al., 1992; Gmyl et al., 1993).

The results of these and other experiments show that many mutations and their combinations are life-compatible and may not decrease viral fitness and, if they have adverse effects, these effects may often be repaired or compensated. Even broken molecules of genomic RNA may sometimes be repaired by recombination (Gmyl et al., 1999, 2003). Thus, viral RNA genomes are robust and exhibit remarkable vitality and phoenix-like phenotype. What does then force them to evolve and to form new species and genera?

A hypothesis is put forward according to which a significant decrease in fitness (caused for example by mutations or host changes) may result in genome instability, which in turn would produce a set of various low-fit variants. Under non-competitive conditions (e.g., in a small-sized population), such variants may serve as avid acceptors of novel elements acquired by recombination or by some other genetic modifications. Further fine genetic tuning may convert the new creature into a well fit new viral species. Thus, evolutionary “jumps” may not necessarily be due to a consecutive acquisition of improving mutations but rather they may result from injury of an important element followed by metastability of the viral genome, and acquisition of a novel element.

### **3 Evolution of Viral Pathogenicity and Emerging Viral Infections**

Viruses are not only major human, animal, and plant pathogens, sometime deadly, being the causative agents of small pox, HIV, influenza, foot-and-mouth disease and many other severe pathological conditions, but also they are a driver of global geochemical cycles through killing microplankton, bacteria etc. Although viruses have no special “desire” to hurt the host cell, they very often are doing so just by selfishness and negligence.

It is well known that specific viruses can infect only specific organisms and can damage in these organisms only specific organs and tissues. An important factor contributing to this specific pathogenicity is the viral host range, i.e., the ability or inability to infect a given type of cells. Main factors controlling the host range are availability of appropriate receptors on the cell surface, appropriate intracellular milieu, and status of the innate immunity. Virus-induced damage to cells may be caused by many factors and among them by competition for resources and for cellular infrastructure. The outcome of infection also depends on the availability of viral anti-defensive mechanisms. In some instances, pathology may result from the hyperactive host defense.

Killing the host organism does not confer to the infecting virus any advantage. Rather, co-evolution of viruses and hosts should likely result in a kind of equilibrium. This principle may be illustrated by co-evolution of a rabbit-pathogenic virus and wild rabbits in Australia in the mid of the last century (Fenner and Ratcliffe, 1965). To control the enormous population of imported European rabbits, which became major pests for agricultural industry, it was decided to use the fibroma/myxoma virus. This virus induces benign tumors in American rabbits but causes severe generalized lethal lesions in European ones. As expected, virus introduction resulted in mosquito-borne (summer) epizootics, with >99% of the infected rabbits dying in less than 2 weeks. Naturally occurring less virulent virus variants had more chances for overwintering, the circumstance resulting in selection of attenuated viruses. After a decade, the mortality of European rabbits in Australia caused by the evolved virus decreased roughly two-fold. Simultaneously, selection of resistant rabbits took place: mortality of such rabbits infected with the original virus decreased roughly fourfold.

On the other hand, newly emerging viral infections, caused often by viruses, which have been transferred to human populations from an animal reservoir, may exhibit a very high pathogenicity for humans. A remarkable example of this is the 1918 Spanish flu, which killed about 30–50 million people, or ~2% of those infected. In fact, influenza virus is in most cases a nonpathogenic or slightly pathogenic avian enteric virus. To infect a human, an avian flu virus should change its receptor specificity, which depends on the interaction of viral hemagglutinin (HA) with a cellular membrane glycoprotein receptor. Generally, it is sufficient to change only two amino acid residues in the avian HA to allow it to efficiently recognize the human receptor. Such a change in the host range may be achieved by either mutations in the avian HA or acquisition by an avian virus of the HA gene from human influenza virus as a result of genetic exchange (reassortment) between these viruses during mixed infections. Adaptation of flu viruses to humans may also require mutations in other viral genes. The severity of the infection depends not only on the efficiency of viral reproduction in the new host but also on the balance and interplay between host defense and viral counter-defense mechanisms.

Influenza virus populations are constantly changing. While undergoing immune pressure, they accumulate mutations, primarily in HA and neuraminidase (NA) (“antigenic drift”). Qualitatively new variants are emerging from time to time, primarily through reassortment (“antigenic shift”). The Spanish (1918) flu was just one of such “shifted” variants, as judged by the nucleotide sequence of its genomic RNA determined from archival formalin-treated paraffin-embedded as well as permafrost samples (Tumpey et al., 2005). It was established that the deadly virus was indeed of avian origin and that the “jump” into humans was associated with alterations in the HA and some other genes. The virus was reconstructed on the basis of known nucleotide sequence of its genome, and its pathogenic properties were confirmed and further investigated. The causative agents of the two next influenza pandemics, the “Asian” (1957) and “Hong-Kong” (1968) viruses, also resulted from reassortment between avian and human viruses. Currently newly emerging, highly pathogenic avian virus (H5N1) presents a potential threat of a new pandemic. Although cases

of human-to-human transmission of this virus are so far extremely rare and only occur during close family contacts, the possibility of acquisition of ominous mutations cannot be ruled out.

The severe acute respiratory syndrome (SARS) is caused by another newly emerged virus. It belongs to coronaviruses. Representatives of this viral family are etiological agents of relatively mild respiratory (e.g., common cold), enteric, and some other diseases in a wide variety of animals and humans. Again, the very dangerous virus “came” to humans from an animal (in this case, bat) and the inter-species transmission involved a change in receptor recognition (Lau et al., 2005; Li et al., 2005).

The human immunodeficiency viruses (HIV) originated from simian immunodeficiency viruses (SIV). SIVs are viruses that infect some 40 different nonhuman primate species in sub-Saharan Africa. Primates naturally infected with SIV do not appear to develop immunodeficiency. The immediate precursor of HIV-1 was a virus, SIVcpz, which infects chimpanzees of the subspecies *Pan troglodytes troglodytes* in West Central Africa. Interestingly, SIVcpz strains have been transmitted from chimps to humans on at least three independent occasions, with the current HIV-1 group M pandemic strain (>60 million people infected, >20 million deaths) resulting from just one of these transmissions. The trans-species transmission appeared to require multiple adaptation events, such as changed transmission mode (from bite/wounds to sexual), adaptation to receptors and overcoming some post-entry barriers (Gao et al., 1999; Heeney et al., 2006). In human populations and individuals,

HIV is rapidly evolving by mutations and recombination, making antiviral therapy a very difficult job.

Major lessons derived from studies of emerging viral infections are as follows. New human pathogenic viruses are usually originating from relatively low pathogenic animal viruses by changing host range (crossing interspecies barrier). They may become highly pathogenic in humans, in particular, because they, as unknown invaders, are met with a hyper-reactive defense reaction leading to extensive host damage, whereas the viruses carelessly employ in full their anti-defensive tools. Further rapid evolutionary rates may contribute to difficulty in combating these infections. On the other hand, long-term evolution of such viruses may lead to a decrease in their pathogenicity and human evolution may probably result in an increased resistance to such viruses.

We may blame viruses for severe diseases and even should actively combat them but we should never forget that we owe to viruses very much for the existence of the live Nature and for our own existence.

## References

- Agol, V. I., 1974, Towards the system of viruses, *Biosystems*, **6**:113–132.  
Baltimore, D., 1971, Expression of animal virus genomes, *Bacteriol. Rev.*, **35**:235–241.  
Bamford, D. H., 2003, Do viruses form lineages across different domains of life? *Res. Microbiol.*, **154**:231–236.

- Breitbart, M., and Rohwer, F., 2005, Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.*, **13**:278–284.
- Fenner, F., and Ratcliffe, F. N., 1965, Myxomatosis, Cambridge University Press, London and New York.
- Forterre, P., 2006, Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: A hypothesis for the origin of cellular domain. *Proc. Natl. Acad. Sci. U. S. A.*, **103**:3669–3674.
- Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C.M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H., 1999, Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature*, **397**:436–441.
- Gmyl, A. P., Pilipenko, E. V., Maslova, S. V., Belov, G. A., and Agol, V. I., 1993, Functional and genetic plasticities of the poliovirus genome: Quasi-infectious RNAs modified in the 5'-untranslated region yield a variety of pseudorevertants. *J. Virol.*, **67**: 6309–6316.
- Gmyl, A. P., Belousov, E. V., Maslova, S. V., Khitrina, E. V., and Agol V. I., 1999, Nonreplicative RNA recombination in poliovirus. *J. Virol.*, **73**:8958–8965.
- Gmyl, A. P., Korshenko, S. A., Belousov, E. V., Khitrina, E. V., and Agol, V. I., 2003, Nonreplicative homologous RNA recombination: Promiscuous joining of RNA pieces? *RNA*, **9**:1221–1231.
- Heeney, J. L., Dalgleish, A. G., and Weiss, R. A., 2006, Origins of HIV and the evolution of resistance to AIDS. *Science*, **313**:462–466.
- Koonin, E. V., and Martin, W., 2005, On the origin of genomes and cells within inorganic compartments. *Trends Genet.*, **21**:647–654.
- Koonin, E. V., Senkevich, T. G., and Dolja, V. V., 2006, The ancient virus world and evolution of cells, *Biology Direct.*, **1**:29.
- Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S. Y., Chan, K. H., and Yuen, K. Y., 2005, Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.*, **102**:14040–14045.
- Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Cramer, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S., and Wang L.-F., 2005, Bats are natural reservoirs of SARS-like coronaviruses. *Science*, **310**:676–679.
- Pfeiffer, J. K., and Kirkegaard, K., 2003, A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity, *Proc. Natl. Acad. Sci. U. S. A.*, **100**:7289–7294.
- Pfeiffer, J. K., and Kirkegaard, K., 2005, Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog.*, **1**:e11.
- Pilipenko, E. V., Gmyl, A. P., Maslova, S. V., Svitkin, Y. V., Sinyakov, A. N., and Agol, V. I., 1992, Prokaryotic-like cis element in the cap-independent internal initiation of translation on picornavirus RNA. *Cell*, **68**:119–131.
- Pilipenko, E. V., Viktorova, E. G., Guest, S. T., Agol, V. I., and Roos, R. P., 2001, Cell-specific proteins regulate viral RNA translation and virus-induced disease. *EMBO J.*, **20**:6899–6908.
- Suttle, C. A., 2005, Viruses in the sea. *Nature*, **437**:356–361.
- Tumpey, T. M., Basler, C. F., Aguilar, P. V., Zeng, H., Solorzano, A., Swayne, D. E., Cox, N. J., Katz, J. M., Taubenberger, J. K., Palese, P., and Garcia-Sastre, A., 2005, Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science*, **310**:77–80.
- Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., and Andino, R., 2006, Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, **439**:344–348.