

Revolutionary Struggle for Existence: Introduction to Four Intriguing Puzzles in Virus Research

Matti Jalasvuori

Abstract Cellular life is immersed into an ocean of viruses. Virosphere forms the shadow of this cell-based tree of life: completely dependent on the tree for existence, yet, the tree is equally unable to escape its ever evolving companion. How important role has the shadow played in the evolution of life? Is it a mere ethereal partner or a constitutive factor? In this chapter four puzzles in virus research are taken under the scope in order to probe some of the intriguing ways by which viruses can help us understand life on Earth. These puzzles consider the origin of genetic information in viruses, viruses as symbiotic partners, the structural diversity of viruses and the role of viruses in the origin of cellular life. More than providing answers, this introduction exemplifies how viruses can be approached from various angles and how each of the angles can open up new ways to appreciate their potential contributions to life.

1 Introduction

Life on Earth is composed of multitude of cellular organisms, some of them being as tiny as bacteria, others as complex as humans. Yet, this cellular way of living is overwhelmed in both number and genetic diversity by non-cellular entities, each of which is capable of enforcing cellular organisms to fulfill their selfish needs. A word *virus*, a Latin term for poison, commonly refers to this strategy for survival.

M. Jalasvuori (✉)

Division of Ecology, Evolution and Genetics, Research School of Biology,
Centre of Excellence in Biological Interactions,
Australian National University, Canberra, Australia

Department of Biological and Environmental Science, University of Jyväskylä,
P.O. Box 35,40014, Jyväskylä, Finland
e-mail: matti.jalasvuori@jyu.fi

And for a poison they are often treated. This is of no surprise, given that the apparent simplicity and inanimate nature of deadly viruses (van Regenmortel 2000; Moreira and López-García 2009) may lead us to intuitively neglect or completely ignore them in our approaches to understand the evolutionary spectacle that living things have to offer. Yet, while being relatively simple in comparison to cells, there is much that we do not know about viruses or their roles in evolutionary processes. Viruses have been here for a long time (Forterre and Prangishvili 2009a), and studies suggest that viruses appear to have played a part in events such as the origin of cellular life (Koonin et al. 2006) and the evolution of mammals (Gifford 2012). But what has their role been exactly? When does the inclusion of viruses into the frame of analysis lead to evolutionary insights? Or even breakthroughs?

Unfortunately in many instances we are still after on a mere hunch. For this reason, instead of providing you with a set of scientifically chewed and grounded answers, I introduce you to a four selected puzzles in virus research in an attempt to scope where the limits of some of our contemporary knowledge lies. The presented questions revolve around themes such as the origin of new genetic information, the origin of new types of symbiotic relationships, and even the origin of life as we know it. Naturally profound puzzles as these are horribly difficult ones to address in a complete and comprehensive manner. Yet, in the spirit of this book, these puzzles can help determining whether viruses could be considered truly as essential agents of life.

1.1 Viruses and Virions: What Is the Difference?

First, however, a relatively commonly adopted misconception on what a biological virus actually is must be resolved because it has been behind many of the misunderstandings on viruses. The heart of the issue lies in the notion that a virus often refers only to the protein-formed protective capsid, which encloses viral genomic information in the extracellular environment (see discussion in Jacob and Wollman 1961; Forterre and Prangishvili 2009b; Villarreal and Witzany 2010; Moreira and López-García 2009; Jalasvuori 2012). This infectious particle is known as a virion and they are generally regarded to be dead (in many depressingly unfruitful discussions). Virions are entities that intrude and assume the control of cellular organisms in order to produce more virions. But should this dead virion actually be considered equal to a virus? And what then would a virus be, if not a virion? The seemingly trivial difference between a virus and a virion needs to be tackled as it allows us to appreciate viruses as evolutionary players, or even as living organisms (Forterre and Prangishvili 2009b; Villarreal and Witzany 2010; Forterre 2011; Jalasvuori 2012). In any case and regardless of our opinions on their living status, viruses are part of the evolving biosphere and therefore a relevant factor in various evolutionary processes.

Virion is the extracellular step in the life cycle of a virus. Virion is the traditional picture that every book offers for depicting a virus. Virion is the transient stage by which the viral genetic information gets from one host organism to another. This virion, however, lacks the *life* of the virus since it is only the dormant and inactive form of viral genetic information (Brüssow 2009). For this reason viruses might appear as toxic substances that have the capability to occasionally cause the demise of cellular organisms but that are essentially just another environmental factor of only minor interest from evolutionary point of view.

However, arguably, the actual virus is more than its dead shell in the environment. Virus is part of a living organism when it is inside a host cell. And the phenotype of this organism is partly expressed by the virus (Forterre and Prangishvili 2009a; Forterre 2010; Jalasvuori 2012). Many viruses maintain the potential for producing inanimate virions during their endure within the cellular organism, but virus itself should be considered to be its full reproductive cycle including both external and internal parts (Villarreal and Witzany 2010; Jalasvuori 2012). Yet, strictly speaking, only the within-cell reproductive cycle is required for the survival of the viral genetic information (Krupovic and Bamford 2010; Jalasvuori 2012). And this requirement lets us approach viruses as a genuine form of life that can exploit foreign cell-vehicles for preserving and propagating their genetic information (Forterre 2010, 2011).

In other words, virus should not be mistaken only for their non-essential extracellular form given that viruses are equally dependent on cells with all other genetic replicators – being those chromosomes, plasmids or anything else. Virus just is not dependent on any particular cell due to their capability transfer themselves from one cell to another via virions. And due to this extracellular form of existence, viruses are not terminated even if their replication causes the demise of the current host organism. However, jumping from this notion to the conclusion that viruses are dead and thus irrelevant partners of evolutionary processes is unwarranted. Naturally, our definitions of viruses include the infectious extracellular part, but for thorough understanding of viral life it must be noted that any such definitions are in the end artificial. Virus is one of the ways by which genetic information have adapted to survive in this biosphere. From the viewpoint of cellular organisms, this way of struggle for existence is much more complex than the presence of chemical substances in the environment would be. Viruses, unlike poisons, are capable of evolving genetically and going extinct. Sometimes they can also form more or less permanent mutually benefiting relationships with their hosts.

Now this perhaps more allowing perspective to viral life sets a more appropriate stage to consider any virus related puzzles. Each of the presented questions approach viruses from different angles and hopefully provide an intriguing introduction to the diversity of ways by which viruses may help us understand the evolution of our biosphere. However, I wish to note that I consciously retained from drowning the reader in supporting evidence in order to keep the text fast pacing and relatively easy to digest.

2 Can Genes Emerge in Viruses?

Novel sequencing and sampling techniques have made it possible to determine the overall genetic information in any particular sample. Moreover, sequences of complete organisms have revealed the true genetic diversity of living entities. These studies have led to the revelation that many organisms harbor a variety of genes that are unknown to science (Mocali and Benedetti 2010). In other words, our biosphere is abundant with genetic information for which we cannot assign a role, function or evolutionary origin (Cortez et al. 2009). Interestingly, a fair portion of these novel genes are found from viral genomes (Yin and Fischer 2008; Prangishvili et al. 2006) or belong to genome integrating genetic elements (Cortez et al. 2009). How did these genes end up in viruses?

2.1 *Are Viruses Only Hitchhiking on Genetic Information?*

Viruses are completely dependent on cellular resources for reproduction. Viruses use cellular amino acids to make viral proteins and some acquire lipids from cellular membranes to assemble functional virions. All viruses embrace cellular nucleotides to produce copies of viral genetic information. Given the profoundly parasitic nature of viruses, it seems reasonable to assume that viruses are also completely dependent on cellular genes for evolution. Indeed, many viral genes appear to have been acquired from their hosts and thus viruses could be considered as genetic burglars, hitchhikers on the highway of genetic information. Viruses are something that themselves are not evolving but which are evolved by cells (Moreira and López-García 2009). The actual *de novo* origin of genetic information would happen within stable cellular beings such as bacteria.

However, many viral genes appear to have no cellular counterparts (Yin and Fischer 2008; Forterre and Prangishvili 2009b). Why is this? Do we need to sequence more bacterial genomes in order to find the common ancestor form a cellular chromosome? Yet, as the number of sequenced bacterial chromosomes has increased, the number of unknown genes in viruses has remained unchanged (Forterre and Prangishvili 2009b). Sometimes when some rare types of virus genes are finally discovered from host chromosomes, it turns out that the genes in the chromosomes actually belong to genome integrated viruses (Jalasvuori et al. 2009, 2010). Therefore the sequencing of bacterial chromosomes does not seem to provide an easy way out of the puzzle. Perhaps the genetic novelty of viruses is of genuine nature and there are no cellular homologies to be found. Or could it just be that the rapid evolutionary rates of genes in viruses is simply making the homology with cellular genes untraceable?

In principle, it is possible that majority of genes evolve in such a fast pace in viruses that the sequence can no longer be recognized to be of cellular origin (Forterre and Prangishvili 2009b). Indeed, general analyses of the divergences of amino acid sequences propose that even the most conserved proteins in our biosphere have not

discovered all potential ways to encode their function (Povolotskaya and Kondrashov 2010). Therefore there appears to be room in the sequence space into which the host-derived genes can evolve to in viral genomes.

However, comparison of nucleotide or amino acid sequences is not the only mean by which gene divergences can be studied. While the sequence on DNA or amino acid level may evolve rapidly, the three dimensional structure of the gene product, usually a protein, can remain relatively unchanged. Indeed, generally there is no selection to preserve any certain amino acid sequence but only the (whatever) function that is associated with the three dimensional conformation of the protein. Save for amino acids mediating chemical reactions, the same structural conformation can be acquired with a variety of different sequences.

Viruses seem to have genes that produce structurally and functionally conserved proteins, which have no apparent cellular ancestors (Bamford et al. 2005; Koonin et al. 2006; Keller et al. 2009). These genes have been within (relatively) independently evolving viral genomes perhaps for as long as billions of years and they can still be shown to share a common ancestry. Did these genes emerge in virus genomes in the first place? It seems possible, given that many of these conserved “hallmark” virus genes (Koonin et al. 2006) encode for viruses specific tasks such as capsid proteins or packaging enzymes that facilitate the transfer of viral genome into the capsid.

2.2 If Gene Emerges Within a Cell But Survives in Viral Genome, Is It a Viral Gene?

Naturally, the emergence of a gene in a virus does not indicate that the gene popped into existence within the protective capsid in an extracellular environment (Forterre and Prangishvili 2009b; Forterre 2010; Jalasvuori 2012). Rather, it would mean that a virus, while replicating in a cell, ended up having an altered genetic sequence. This altered sequence opened the road for the emergence and evolution of a new gene. In practice the gene would form through point mutations and other genetic changes similarly with any other emerging genes (Forterre and Prangishvili 2009b).

But if the new gene would emerge within a cell, is it not rather a cellular gene than a viral one (Moreira and López-García 2009)? Doesn't this indeed only enforce the view of cellular origin of viral genetic information? No, it does not, if we allow ourselves to consider viruses to be more than just their encapsulated extracellular forms (Forterre 2010). If the gene formed through mutations in a viral genome and the new gene was able to survive due to its benefits to the virus and not to the host, then it would seem only reasonable to consider the gene to be of viral origin (Jalasvuori 2012). Therefore, even if a cell serves the function of a vessel for the development of a new gene, the gene would remain in the global gene pool because of viruses. Eventually, when metagenomic studies, for example, are performed, these novel genes could be discovered from capsid enclosed genomes of viruses with no apparent counterparts in any cellular organisms.

Even if the *de novo* origin of genes actually occurred in viruses, it would be only a starting point from which to approach other interesting questions. What do these novel genes do? There are countless unique genes in viruses, but are they also encoding countless unique functions. Or is it possible that they only have unique sequences while affecting very similar cellular processes? And what would that indicate?

Viruses of bacteria, also known as bacteriophages, can have genes for very different types of functions. Some phages encode transfer RNAs and other essential cellular functions (Miller et al. 2003). Others can carry genetic information for mediating photosynthesis (Mann et al. 2003) or producing lethal toxins (O'Brien et al. 1984). Much of the phage genes, however, affect genetic regulation, virion assembly and host-virus interactions. Yet, other viruses (like Mimivirus) have genes that were earlier considered to be only part of cellular chromosomes and thus blurred the line between what viruses can and what they can not do (Raoult et al. 2004).

Nevertheless, in principle, it seems possible that the product of a viral gene can influence any thinkable biological process. Some truly novel genetically encoded functions allowing, for example, exploitation of completely new types of resources or inhabit previously uninhabitable environments, may come into existence in the genome of a virus. Perhaps viral innovations can open new niches for cellular organisms to occupy: many of the novel genes in bacteria are taxonomically restricted and ecologically important (Wilson et al. 2005).

3 Can Viruses Become Symbionts?

Viruses are generally seen as parasites of cellular organisms. Viruses enter the host cell, utilize cellular resources for creating new viruses and then sacrifice (or damage) their temporary slaves in order to escape the scene of crime. How could this violent strategy ever turn into a mutually benefiting symbiosis?

In a mutualistic relationship the fitness of the two entities together is (often) higher than the fitness of either of the components alone. In other words, both of the symbionts would suffer from abandoning its partner. Therefore, if a virus was ever to be appreciated as a mutually benefiting partner, it should be counterproductive for the host cell to get rid of a virus that has integrated into genome of the host. This seems to be a problematic approach, given that the avoidance of parasites is considered to be one of the key drivers of evolution and responsible (at least partly) for the maintenance of such fundamental traits as sexual reproduction (Hamilton et al. 1990; King et al. 2011).

3.1 Endogenous Viruses: Fossils or Something More?

Nevertheless, viral genetic information is often found to be incorporated to cellular genomes (Holmes 2011). For example, human chromosomes contain more viral DNA than actual human genes. In fact, remnants of viruses are abundant in genomes

of many different organisms, ranging from animals to bacteria (Casjens 2003; Katzourakis and Gifford 2010; Jalasvuori et al. 2010). How did these viral elements get into all these organisms? What types of evolutionary processes may be responsible for these genomic fusions, and could they be of evolutionary importance?

Are the existing viral remnants in genomes mere evolutionarily insignificant left-overs of previous virus infections (Jern and Coffin 2008)? Were they so insignificant to the fitness of the hosting cell that there simply was no selection to get rid of the element? Many of the endogenous viruses are relatively conserved and have persisted over evolutionary times in various species, such as humans and our primate cousins, suggesting that the relatively error-free host polymerases that are used to replicate the endogenous viruses are able to preserve these sequences as viral fossils over evolutionary times (Duffy et al. 2008). However, many of the virus elements have also shown to accumulate inactivating mutations and thus they are evolving only as non-encoding pseudogenes in animal genomes (Katzourakis and Gifford 2010). Yet, other virus genes have remained functional, suggesting that there has been a purifying selection to maintain the correct sequence.

3.2 What Benefits Can Viral Elements Provide to the Host?

Could it be possible that some of these viral elements in cellular chromosomes resulted essentially from mutually benefiting although aggressive genetic fusions (Ryan 2009)? Can the symbioses of viruses with cells be evolutionarily favorable steps, not mere coincidences?

In order to be more precise, the question is not whether genetic fusions of the genomes of viruses and cells can improve the reproductive rate of cells *per se*. There are clear examples for this to be true. As a tragic example several viruses are known to cause the uncontrolled multiplication of human cells, which results in the formation of tumors. These virus-containing cells out-reproduce other human cells and thus they end up having much more descendants than the virus-free cells. Within this limited framework the virus-cell symbiotic can have the highest fitness. But by extending our perspective we notice that this short-term benefit rapidly backfires due to the demise of the hosting animal. The selfish behavior of some cells leads to a tragedy of commons, where the gain of few is decreasing the fitness of both host and the virus. Therefore, the real question is whether viruses and their hosts may form symbiotic relationship that can increase the fitness of the whole organism within a large-enough evolutionary frame. In other words, we can ask, for example, if the virus-host symbiont could invade a population of virus-free hosts because of the advantages that the virus provides to its hosts.

Some viruses that infect bacteria are known to form temporary mutually benefiting symbiotic relationships with bacterial cells (Roossinck 2011). These viruses enter the host cell and, instead of producing vast number of virions and destroying the cell, they take up residence within the host. During this latent infection temperate viruses replicate their genomes along with the cell but deter from making

virions. Only in the distress of their hosts they ignite the production of virions and they do it in order to escape the potentially doomed bacterium.

These temperate bacterial viruses may carry genes (e.g. for producing toxins) that can significantly improve the performance and thus the reproduction of their host bacteria. The combination of the bacterial virus and the bacterium can end up being the evolutionary winner in a competition against bacteria that did not have the latent viral infection. Therefore, among bacterial organisms such straightforward mutualistic relationships may emerge on regular basis (Roossinck 2011). Moreover, the short-term benefit provided by the phage does not backfire in the same sense as the spreading tumors do within animal hosts. But then, bacteria and humans are quite different in multiple respects. Are these symbioses limited only to single-celled beings or can such relationships emerge among more complex organisms that reproduce via specific germ cells? Indeed, despite of the all the movies, we do not know of any viruses that carry bacteriophage-like toxin genes, which would grant us some sort of superpowers. Therefore this bacterial approach may simply be ill-suited to understand symbiotic relationships in animals.

However, there is another way by which temperate viruses of bacteria boost the survival of their hosts. Whenever a bacterial virus resides within a bacterium, it renders the cell immune to infections by similar viruses. And this quality of viruses, the incapability of a single virus type to multiply infect an already-infected cell (i.e. the resistance of superinfection), appears to be very common among all viruses and therefore also applicable to other organisms (Berngruber et al. 2010). Prevention of superinfection allows viruses to establish latent infections that are especially important under conditions where chances for horizontal transfer of the virus are limited.

Among bacterial populations that are subjected to temperate viruses, the most rapid mean by which resistant host cells emerge are due to the latent infections by temperate viruses themselves. The presence of the virus therefore selects the bacterial population to become prevalent with integrated viruses. When there are both susceptible hosts and infective virions in the same environment, the resistant hosts have an apparent advantage (Roossinck 2011). Moreover, the genome integrated viruses sometimes produce virions and thus maintain the selection for the presence of the latent virus. The fact that viruses themselves contain genetic means to make host cells immune to the virus may prove to be the evolutionary superpower that can facilitate the formation of a symbiotic relationship also between a virus and its animal host.

However, even if viral infections can make the host animal resistant to further infections by similar types of viruses, it is not a heritable symbiosis. We are immune to chickenpox after an infection, but our children still need to get infected themselves in order to become resistant (or, alternatively, be vaccinated against the virus). Is it possible that the resistance would become inheritable so that the progeny of an infected individual would not need to face the severe effects of an infection?

Complex multi-cellular animals develop from a fertilized cell. This single cell divides and the divided cells specialize to different functions eventually producing a complete organism. The genetic information in all animal cells remains essentially

the same throughout the life of the organism even if the phenotypes of cells can vary tremendously. Therefore, if the virus was integrated already in the original germ cell, it would become inherited to every cell of the multi-cellular organism, including those that eventually become the germ cells of the next generation. In such a case the virus could both protect the organism from the external versions of the virus and be transmitted vertically to the next generation.

3.3 *Taming the Enemy into an Ally*

During a roaming virus epidemic, this integration of a virus to germ line cells could provide an advantage to an individual (Jern and Coffin 2008). Indeed, in many cases endogenous viruses appear to protect their hosts against exogenous viruses (Maori et al. 2007; Katzourakis and Gifford 2010). However, such endogenous viruses themselves seem to be able to reinfect the germ line cells (Belshaw et al. 2004). Nevertheless, the endogenous virus may be able to make the host organism to be able to ignore the ill-effects that the epidemic causes to other individuals. Naturally inheritable resistance against chickenpox is not a significant advantage but resistance against a more severe virus could be.

So, in principle and under certain conditions, germ line infection could prove to be a favorable trait within a population (Maori et al. 2007). The new *virus alleles* may even be able to invade the whole population, if the maintenance of the virus remains to improve the fitness of the virus-containing individuals over their virus-free counterparts (Katzourakis and Gifford 2010). Indeed, as with bacteriophages, endogenous viruses of animals can remain partly active even after endogenization (Coffin et al. 1997; Tarlinton et al. 2006) and thus the virus itself can maintain the pressure to retain the virus allele within the population.

In such a case, is it possible to consider that the virus has established a mutually benefiting relationship with its animal host. Maybe, given that it would be disadvantageous for the organism to get rid of the virus since it would make the organism susceptible to infections. Of course, this symbiotic partnership would exist mainly on the level of genetic information (Ryan 2009), but it would still emerge through a fusion of two distinct genetically reproducing entities. In the end, very little is still known about the endogenization process. Even if viruses could be considered to form symbiotic relationships via whatever mechanisms, several interesting questions remain. How does this new integrated virus affect the subsequent evolution of their hosts? Endogenous virus changes the genetic composition of the chromosomes and can, for example, regulate the expression of host genes (Jern and Coffin 2008). Some of the viruses are active elements and cannot be dismissed as irrelevant components of organisms. Indeed, some virus derived genes in mammals and other animals appear to have remained active for over tens of millions of years (Katzourakis et al. 2005; Katzourakis and Gifford 2010). But even then, it is difficult to say for certain how significant role did these viruses play in the evolution of their hosts. However, we are free to do little speculation.

Endogenous viruses can integrate repeatedly into various places within and among host chromosomes (Katzourakis et al. 2007). The number of elements and the site of integration can have significant effects on the phenotype of the host cell. The establishment of the viral genome into the host chromosome appears to be followed by in-genome evolution (Tarlinton et al. 2006; Katzourakis et al. 2007). Does this evolution select for the viruses to be integrated in positions where they induce the lowest possible cost on the host or, perhaps, even induce changes that increase the host fitness?

Sexual reproduction effectively filters genetic information to produce beneficial combinations. Could sexually reproducing individuals become favored over asexually reproducing phenotypes as the sexual recombination of genetic material allows the integrated virus to more rapidly settle within fixed beneficial locations in chromosomes? Or perhaps allow the hosts to tame the uncontrollably proliferating endogenous viruses (Katzourakis et al. 2005)? Could the subsequent evolution after virus endogenization induce notable changes in the phenotype of the organism as the genome stabilizes to cope with the presence of the new element?

Some or even most of the endogenous viruses may be just insignificant remnants of previous infections and as such they would not much affect the evolution of their host species. But other symbiotic viruses probably made a real difference. As an example of such, a virus derived gene, labeled as syncytin, appears to be crucially important for the morphogenesis of placenta (Mi et al. 2000). Did pregnancy as humans and other placental mammals experience it emerge as a result of viral endogenization?

4 Why Are There Only Few Types of Bacteriophages?

Viruses are known to evolve rapidly and viral genomes often contain unique genes for which no homologues can be determined. But are virions, the extracellular forms of viruses, composed of similarly diverse structures? Is there a novel structural design waiting whenever we pick up any of the 10^{31} or so virions (Suttle 2007) from the environment?

The proteins on the virion dictate whether or not viruses are able to attach to a suitable host cell and therefore there should be constant selection driving the evolution of these proteins (as well as their host counterparts) Weitz et al. 2005. This is indeed what has been observed: the genes responsible for encoding virion proteins that mediate host-cell attachment are the ones that evolve most rapidly (Saren et al. 2005; Paterson et al. 2010). Even closely related viruses may have completely different genes for producing the host-recognizing spikes on the virion (Jaakkola et al. 2012).

But virion is more than a mean to mediate host recognition. The capsid serves as the protective shell for genetic information in the extracellular environment and therefore viruses must also encode proteins (or other means) to produce this shell. Are the genes and the architectural principles for forming capsids equally diverse with host recognition genes?

While virions are extremely abundant and the genetic information they enclose can be very diverse, the capsids of a significant portion of virions in this biosphere may be arranged into just few conserved and homologous lineages (Krupovic and Bamford 2011). Given the astronomical number of virions on earth, this appears to be worth of a closer look.

4.1 Astronomical Number of Bacteriophages in a Handful of Lineages

Bacteria are the most abundant type of a cellular organism on earth and their viruses are equally common. Bacteriophages almost exclusively form virions with a spherical head on which a tail is attached to. The head beholds the genetic information of the virus whereas the tail serves as a tool for attaching onto new host cells and, sometimes, as an injection needle during the infection process. This homologous group of viruses is known as *Caudovirales* (Ackermann 1998). Other types of bacterial viruses also exist, but they are not many (Ackermann 2001): there are icosahedral viruses with inner – and outer membranes, amorphous viruses and helical viruses (Oksanen et al. 2010). Altogether, we have discovered only less than ten truly different types of virion-architectures from all currently known bacteriophages.

What is this architectural conservation trying to tell us? Why are there not a 100 different types of bacterial viruses, or 100 billion types? Even if there were 100 billion unique types of viruses, each of them would still have over billion billion virions. And such a large number of individuals could indeed retain a stable population over evolutionary times. This, however, is not the case. You can calculate the virion architectures of bacteriophages with your fingers. Viruses are generally considered to be of polyphyletic origin, indicating that there are multiple viral ancestor and not a single common one. Still, the apparently limited number of architectural types suggests that new virus types are not emerging on regular basis, since, if they were, we would be likely to find new viruses all the time. This leads to a question: when did these existing structural types emerge and why did they cease emerging?

We know that mankind may be facing a completely new and highly lethal epidemic any given day. HIV, SARS, Ebola and other doomsday candidates emerged out of the blue just to bring destruction to the world. Is it only bacterial viruses that are no longer emerging whereas higher organisms, like humans, can still have completely novel viruses? But are human viruses actually unique?

4.2 Deep Evolutionary Connections Between Viruses

In 1999 when the major structural proteins of bacterial virus PRD1 and human Adenovirus were compared on structural level, it was noticed, surprisingly, that they were highly similar (Benson et al. 1999). Despite of the sequence dissimilarity, both

viruses used a unique but respectively common type of interlinked protein-barrels (so-called double beta-barrels) for composing their protective capsids. The obvious question emerged: are these two viruses that infect very distantly related hosts (bacteria and humans) actually related to each other? Or is this just another case of convergent evolution where two entities independently evolved towards the same direction (Moreira and López-García 2009)?

Closer analysis of both of these viruses and their other relatives revealed more things in common (Krupovic and Bamford 2008). Vast majority of them had an inner lipid membrane beneath the protein capsid, a generally rare trait among viruses. Moreover, these viruses encode related ATPases (with certain specific motifs) which have been shown to facilitate the transfer of the viral genome into empty capsids. Later on similar viruses were found to infect thermophilic crenarchaea (Khayat et al. 2005) and reside in the genomes of thermophilic euryarchaea (Krupovic and Bamford 2008). In terms of genetic exchange, the Archaeal phylum of Crenarchaeota consists of deep-branching organisms that appear to have been evolving relatively isolated from all other life forms since the emergence of cellular life (Gribaldo and Brochier-Armanet 2006). Together these characteristics suggested that convergence appears to be an improbable cause to explain all the common features and thus it is reasonable to assume the existence of a common ancestor in some distant past. But this leads us to the same question as before: how distant are we actually talking about? 100 million years? A billion? Four billion?

Several analyses suggest that Bacteria and Eukaryote (a domain that includes us humans along with baking yeast) had their last common ancestor about four billion years ago. The same branching time applies to the divergence of Bacteria from Archaea. In other words, these double beta-barrel viruses infected all the domains of life and many deep branches within those domains. But are these viral lineages as old as their cellular hosts? Or is it possible that these viruses emerged later on just to spread to infect all domains of life? We know that viruses are very host specific and usually the viral tree of life corresponds quite well with the evolutionary tree of their hosts (McGeoch et al. 2005). However, there are exceptions and therefore this line of reasoning does not provide a way out of the problem.

Interestingly, several other domain-spanning lineages have been discovered. Herpes viruses have the same peculiar way to produce their capsids as do the extremely abundant tailed viruses that infect bacteria and archaea. Certain RNA-viruses such as bacterial cystoviruses and eukaryal reoviruses appear to be of common origin due to unique genome and capsid organization. There are also other lineages.

It seems that many viruses can have representatives infecting all basic cell types, but these representatives themselves have no recent common ancestors. Moreover, viruses appear to harbor genes that does seem to have been derived from none of the three domains of cellular life but which are very conserved and prevalent among viruses (Koonin et al. 2006). One possible way to explain all these features is to assume that the ancestor of these viruses may have emerged already before the separation of Bacteria, Archaea and Eukaryote into their independent domains.

Recently it was discovered that the double beta-barrel viruses appear to have evolved from a novel viral lineage, so-called single beta-barrel viruses, which

themselves form an independent domain spanning lineage (Krupovic and Bamford 2008; Jalasvuori et al. 2009; Ilona Rissanen personal communication). It is possible that these two viral lineages diverged already before the emergence of contemporary cellular domains. This on the other hand means that by studying viral lineages it might be possible to reach back to some past evolutionary events that occurred before the last universal common ancestor of cells. That period in the evolution of life is generally shrouded in unknown, given that the last common ancestor of cells have been considered as the ultimate boundary beyond which we cannot go by comparing differences between existing living organisms. But if we are not solely dependent on cells in our analyses, then this boundary may be breachable. Study of viral lineages and their origins can give us unique clues about the very first steps of life on Earth.

4.3 *Structural Diversity of Hot Archaeal Viruses*

Interestingly, while bacteriophages are either head-tail viruses or one of the few other types, the virions infecting hyperthermophilic crenarchaeal hosts are structurally very diverse (Prangishvili and Garrett 2004; Pina et al. 2011). There are lemon-shaped viruses, tulip-shaped viruses, bottle-shaped viruses, there are sticks with hooks and pleomorphic-viruses along with all sorts of globular, icosahedral and filamentous morphologies. Why is there such a variation especially among archeal viruses? Bacteria and archaea are so similar to each other that it was only recently that we were even able to distinguish them from one another.

Hyperthermophilic crenarchaea are very deeply branching organisms in the tree of life and their viruses are equally unique (Ortmann et al. 2006). They also inhabit extremely hot environments. Are these clues relevant for understanding the diversity of viral phenotypes? Indeed, when the viruses of less thermophilic archaeal organisms have been studied, they were found to be less diverse morphologically. Could it be possible that there was wider diversity of viral phenotypes during the early steps of the evolution of life? And has this diversity been somehow better prevailing among hyperthermophilic crenarchaeal organisms whereas it was lost among other prokaryotes (Jalasvuori and Bamford 2009)? The viruses of most deep-branching hyperthermophile bacterial families (like *Thermotoga* or *Aquifex*) have not been studied. It would be interesting to see if their viruses resemble only the usual head-tail viruses or whether they are more like the ones infecting crenarchaea – or something totally different.

It is likely that all contemporary life forms on earth have evolved from thermophilic ancestors (Di Giulio 2003). There are at least two potential explanations for this, both of which can be correct. First, life may have emerged within a hot habitat such as hydrothermal vents on the ocean floor. Second, life may have faced multiple near-extinction level catastrophes in which all the surviving organisms were thermophiles. Indeed, earth is known to have been under heavy bombardment of massive comets and asteroids during the Hadean period (ending about 3.8 billion years ago).

This bombardment must have elevated the temperature levels significantly, sweeping all non-thermophilic organisms.

If we assume that life has (repeatedly) evolved to adapt to survive in cooler conditions, it is then possible that only a portion of the original hot viruses have been able to follow their hosts. The original virosphere with all of its structural diversity may still be partially surviving among the most deeply branching and hot living entities. This suggests that the study of these viruses may give us a glimpse on the biosphere as it was very early in the history of life.

5 How Did Viruses Emerge?

As was noted in the previous section, majority or possibly even all of the virions in our biosphere may be arranged into few handfuls of structural lineages. These lineages span across different domains of life and possibly had their origins prior the emergence of the first true reproducing cell. Unfortunately, there is a serious problem in this line of reasoning.

How is it possible that viruses, which are completely dependent on cells to be able to reproduce, emerged before there were reproducing cells in our biosphere? In the introduction it was noted that the extracellular stage of a virus, the virion, is completely inactive unless it encounters a suitable host cell. The only way by which viruses can be considered as living entities is when the inclusion of their within-cell life cycle is taken into account. Therefore the idea of the pre-cellular origin of viruses appears to directly contradict with the very nature of viruses and thus it should falsify any reasoning that supports this virus-first scenario. Or should it?

5.1 *Viruses Before Cells?*

Cell theory states that biological life is composed of cells that reproduce by binary (or multiple) fission. And since the origin of cell theory in the mid nineteenth century, evolutionary biology as a discipline has focused mainly on what happens within and between cells, multi-cellular organisms or populations of organisms. Follow the evolutionary history of any given cell in our current biosphere and your voyage would ultimately end up in the early Earth where the first reproducing cell formed.

However, if any biologist is asked how this first independently reproducing cell came into existence, he or she would be likely to provide only clues to the potential answer. This is because our ideas of the origin of cells are currently only more or less vague hypotheses of potential scenarios. Therefore, as long as we do not know how the first cell (or cells) emerged, the modern life style of viruses cannot be used as a solid argument against the pre-cellular origin of viruses.

Even the most simple bacterium is far too complex for it to have popped out spontaneously within the life-time of our universe. However, evolution can yield increasingly complex systems in accessible timescales and therefore the first true

cell must have been a product of evolution already. Indeed, it might be possible that the contemporary types of cells and viruses are products of the same pre-cellular evolutionary process and thus understanding the origin of viruses as a part of this process may be critical for our understanding of the origin of cells themselves (Koonin et al. 2006; Jalasvuori and Bamford 2008). But if there were no reproducing cells, how did the system evolve?

The attempts to derive the actual nature of the last common ancestor of cells have lead to a strong indication that the ancestor was not any particular cell, but instead a last common community from which the modern domains of life eventually emerged (Doolittle 2000; Theobald 2010). This community appears to have been evolving mainly horizontally by swapping genetic information between proto-cells rather than in “Darwinian” manner by passing genes vertically to proto-cell offspring (Woese 1998, 2000, 2002; Koonin and Martin 2005). This suggest that the proto-cells themselves were not coherent genetic entities but instead more or less random collections of independent genetic replicators. The system probably evolved collectively, which might have maintained the common genetic code (Vetsigian et al. 2006). Physically the proto-cells could have been, for example, fixed inorganic formations that served as containers for enriching products of biochemical cycles and other essential resources (Koonin and Martin 2005).

5.2 *What Good Is a Virus to Primordial Life?*

Regardless of the exact nature of the early evolutionary community, horizontal movement appears to have been a genuine feature of this system. How does a virus fit into this picture? Is it plausible that the viral strategy of survival may emerge within a primordial system even before any independently reproducing cells? Interestingly, all of the previous three questions and their possible answers may be relevant to answer this last question.

If viruses or virus-like replicators are able to come up with new genes, as was discussed in the first question, then viruses could have been one of the elements in the primordial community that produced new innovations. These innovations could have helped the virus-like replicators to, for example, harness resources or synthesize useful biomolecules that, in turn, improved the reproductive rate of the virus themselves. Therefore, it is possible that some of the emerging genes were selected due to their benefits on the survival of virus-like entities for very similar reasons as the novel genes in viral genomes may be doing even today.

Viruses also provide a possible explanation for the horizontal evolution of early life. This is because virions are essentially genetically encoded structures that mediate cell-to-cell transfer of genetic information. The different structural lineages of viruses, as discussed in the third question, may have emerged within this early community when selection favored any trait that allowed genetic information to get from one proto-cell to another. If the primordial system consisted of fixed set of proto-cells, then fitness of the replicator correlated to some extent with its capability

to distribute itself to all potential proto-cells of the community. Isolated virus-free proto-cells may have been prone to collapse under replication parasites (Bresch et al. 1980; Szathmáry and Demeter 1987). Maybe the system survived such parasite epidemics by distributing the contents of healthy cells where virus-production did not succumb to aggressive replication of parasites.

As the primordial system advanced, some of the first viruses may have established more permanent residence in some of the proto-cells in a similar manner as was speculated in the second question. Could these viruses have prevented the over-exploitation of cellular resources by selfish parasites by providing genetic means to prevent other viruses to super infect these proto-cells? Did these mutualistic relationships between proto-cells and viruses clear the way for some of the proto-cells to become more independent from the rest of the genetic community? And did these increasingly independent cells eventually serve as ancestors of modern cellular lineages? Or are we completely lost here and in reality it was something completely different that produced our contemporary cells?

There are plenty of intriguing questions for virus research to tackle. Yet, even if fundamental scientific puzzles like the ones introduced here are still buried into the ocean of uncertainties, the same puzzles can help realize the potential that virus research can have in helping to find the answers. In any case, only the study of viruses can tell us whether or not they are truly essential agents of life.

Acknowledgements This work was supported by The Academy of Finland (grant 251013 and The Centre of Excellence program in Biological Interactions).

References

- Ackermann HW (1998) Tailed bacteriophages: the order caudovirales. *Adv Virus Res* 51:135–201
- Ackermann HW (2001) Frequency of morphological phage descriptions in the year 2000 brief review. *Arch Virol* 146:843–857
- Bamford DH, Grimes JM, Stuart DI (2005) What does structure tell us about virus evolution? *Curr Opin Struct Biol* 15:655–663
- Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M (2004) Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci USA* 101:4894–4899
- Benson SD, Bamford JK, Bamford DH, Burnett RM (1999) Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* 98:825–833
- Berngruber TW, Weissing FJ, Gandon S (2010) Inhibition of superinfection and the evolution of viral latency. *J Virol* 84:10200–10208
- Bresch C, Niesert U, Harnasch D (1980) Hypercycles, parasites and packages. *J Theor Biol* 85:399–405
- Brüssow H (2009) The not so universal tree of life or the place of viruses in the living world. *Philos Trans R Soc Lond B Biol Sci* 364:2263–2274
- Casjens S (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 49:277–300
- Coffin JM, Hughes SH, Varmus HE (eds) (1997) *Retroviruses*. Cold Spring Harbor Lab Press, New York

- Cortez D, Forterre P, Gribaldo S (2009) A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol* 10:R65
- Di Giulio M (2003) The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. *J Theor Biol* 221:425–436
- Doolittle WF (2000) The nature of the universal ancestor and the evolution of the proteome. *Curr Opin Struct Biol* 10:355–358
- Duffy S, Shackleton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9:267–276
- Forterre P (2010) Giant viruses: conflicts in revisiting the virus concept. *Intervirology* 53:362–378
- Forterre P (2011) Manipulation of cellular syntheses and the nature of viruses: the virocell concept. *C R Chim* 14:392–399
- Forterre P, Prangishvili D (2009a) The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann NY Acad Sci* 1178:65–77
- Forterre P, Prangishvili D (2009b) The origin of viruses. *Res Microbiol* 160:466–472
- Gifford RJ (2012) Viral evolution in deep time: lentiviruses and mammals. *Trends Genet* 28:89–100
- Gribaldo S, Brochier-Armanet C (2006) The origin and evolution of archaea: a state of the art. *Philos Trans R Soc Lond B Biol Sci* 361:1007–1022
- Hamilton WD, Axelrod R, Tanese R (1990) Sexual reproduction as an adaptation to resist parasites. *Proc Natl Acad Sci USA* 87:3566–3573
- Holmes EC (2011) The evolution of endogenous viral elements. *Cell Host Microbe* 10:368–377
- Jaakkola ST, Penttinen RK, Vilén ST, Jalasvuori M, Rönnholm G, Bamford JK, Bamford DH, Oksanen HM (2012) Closely related archaeal haloarcula hispanica icosahedral viruses HHIV-2 and SH1 have nonhomologous genes encoding host recognition functions. *J Virol* 86:4734–4742
- Jacob F, Wollman EL (1961) Viruses and genes. *Sci Am* 204:93–107
- Jalasvuori M (2012) Vehicles, replicators and intercellular movement of genetic information: evolutionary dissection of a bacterial cell. *Int J Evol Biol* 2012:874153
- Jalasvuori M, Bamford JK (2008) Structural Co-evolution of viruses and cells in the primordial world. *Orig Life Evol Biosph* 38:165–181
- Jalasvuori M, Bamford JKH (2009) Did the ancient crenarchaeal viruses from the dawn of life survive exceptionally well the eons of meteorite bombardment? *Astrobiology* 9:131–137
- Jalasvuori M, Jaatinen ST, Laurinavicius S, Ahola-Iivarinen E, Kalkkinen N, Bamford DH, Bamford JK (2009) The closest relatives of icosahedral viruses of thermophilic bacteria are among viruses and plasmids of the halophilic archaea. *J Virol* 83:9388–9397
- Jalasvuori M, Pawlowski A, Bamford JK (2010) A unique group of virus-related, genome-integrating elements found solely in the bacterial family thermaceae and the archaeal family halobacteriaceae. *J Bacteriol* 192:3231–3234
- Jern P, Coffin JM (2008) Effects of retroviruses on host genome function. *Annu Rev Genet* 42:709–732
- Katzourakis A, Gifford RJ (2010) Endogenous viral elements in animal genomes. *PLoS Genet* 6:e1001191
- Katzourakis A, Rambaut A, Pybus OG (2005) The evolutionary dynamics of endogenous retroviruses. *Trends Microbiol* 13:463–468
- Katzourakis A, Pereira V, Tristem M (2007) Effects of recombination rate on human endogenous retrovirus fixation and persistence. *J Virol* 81:10712–10717
- Keller J, Leulliot N, Soler N, Collinet B, Vincentelli R, Forterre P, Van Tilbeurgh H (2009) A protein encoded by a new family of mobile elements from euryarchaea exhibits three domains with novel folds. *Protein Sci* 18:850–855
- Khayat R, Tang L, Larson ET, Lawrence CM, Young M, Johnson JE (2005) Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc Natl Acad Sci USA* 102:18944–18949

- King KC, Jokela J, Lively CM (2011) Parasites, sex, and clonal diversity in natural snail populations. *Evolution* 65:1474–1481
- Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. *Trends Genet* 21:647–654
- Koonin EV, Senkevich TG, Dolja VV (2006) The ancient virus world and evolution of cells. *Biol Direct* 1:29
- Krupovic M, Bamford DH (2008) Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat Rev Microbiol* 6:941–948
- Krupovic M, Bamford DH (2010) Order to the viral universe. *J Virol* 84:12476–12479
- Krupovic M, Bamford DH (2011) Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. *Curr Opin Virol* 1:118–24
- Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003) Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* 424:741
- Maori E, Tanne E, Sela I (2007) Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology* 362:342–349
- McGeoch DJ, Gatherer D, Dolan A (2005) On phylogenetic relationships among major lineages of the gammaherpesvirinae. *J Gen Virol* 86:307–316
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC Jr, McCoy JM (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–789
- Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rürger W (2003) Bacteriophage T4 genome. *Microbiol Mol Biol Rev* 67:86–156
- Mocali S, Benedetti A (2010) Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Res Microbiol* 161:497–505
- Moreira D, López-García P (2009) Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol* 7:306–311
- O'Brien AD, Newland JW, Miller SF, Holmes RK, Smith HW, Formal SB (1984) Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* 226:694–696
- Oksanen HM, Poranen MM, Bamford DH (2010) Bacteriophages: lipid containing. In: eLS. Wiley, Chichester
- Ortmann AC, Wiedenheft B, Douglas T, Young M (2006) Hot crenarchaeal viruses reveal deep evolutionary connections. *Nat Rev Microbiol* 4:520–528
- Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B, Fenton A, Hall N, Brockhurst MA (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275–278
- Pina M, Bize A, Forterre P, Prangishvili D (2011) The archeoviruses. *FEMS Microbiol Rev* 35:1035–1054
- Povolotskaya IS, Kondrashov FA (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* 465:922–926
- Prangishvili D, Garrett RA (2004) Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochem Soc Trans* 32:204–208
- Prangishvili D, Garrett RA, Koonin EV (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res* 117:52–67
- Raoult D et al (2004) The 1.2-megabase genome sequence of mimivirus. *Science* 306:1344–1350
- Roossinck MJ (2011) The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol* 9:99–108
- Ryan F (2009) *Viroolution*. Collins, FPR-Books, Sheffield
- Saren AM, Ravantti JJ, Benson SD, Burnett RM, Paulin L, Bamford DH, Bamford JK (2005) A snapshot of viral evolution from genome analysis of the tectiviridae family. *J Mol Biol* 350:427–440
- Suttle CA (2007) Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812

- Szathmáry E, Demeter L (1987) Group selection of early replicators and the origin of life. *J Theor Biol* 128:463–486
- Tarlinton RE, Meers J, Young PR (2006) Retroviral invasion of the koala genome. *Nature* 442:79–81
- Theobald DL (2010) A formal test of the theory of universal common ancestry. *Nature* 465:219–222
- van Regenmortel MHV (2000) In: van Regenmortel MHV et al (eds) 7th report of the International Committee on Taxonomy of Viruses. Academic, San Diego, pp 3–16
- Vetsigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic code. *Proc Natl Acad Sci USA* 103:10696–10701
- Villarreal LP, Witzany G (2010) Viruses are essential agents within the roots and stem of the tree of life. *J Theor Biol* 262:698–710
- Weitz JS, Hartman H, Levin SA (2005) Coevolutionary arms races between bacteria and bacteriophage. *Proc Natl Acad Sci USA* 102:9535–9540. doi:[10.1073/pnas.0504062102](https://doi.org/10.1073/pnas.0504062102)
- Wilson GA, Bertrand N, Patel Y, Hughes JB, Feil EJ, Field D (2005) Orphans as taxonomically restricted and ecologically important genes. *Microbiology* 151(Pt 8):2499–2501
- Woese C (1998) The universal ancestor. *Proc Natl Acad Sci USA* 95(12):6854–6859
- Woese CR (2000) Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* 97:8392–8396
- Woese CR (2002) On the evolution of cells. *Proc Natl Acad Sci USA* 99:8742–8747
- Yin Y, Fischer D (2008) Identification and investigation of ORFans in the viral world. *BMC Genomics* 9:24