Journal of Theoretical Biology 262 (2010) 698–710

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Viruses are essential agents within the roots and stem of the tree of life

Luis P. Villarreal [a], Guenther Witzany [b,*]

[a] Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697, USA
[b] Telos – Philosophische Praxis, Vogelsangstraße 18c, A-5111-Buermoos, Austria

ARTICLE INFO

Article history:
Received 29 May 2009
Received in revised form
28 September 2009
Accepted 8 October 2009
Available online 13 October 2009

Keywords:
Viral interactions
RNA world agents
Eukaryotic nucleus
Biocommunication

ABSTRACT

In contrast with former definitions of life limited to membrane-bound cellular life forms which feed, grow, metabolise and replicate (i) a role of viruses as genetic symbionts, (ii) along with peripheral phenomena such as cryptobiosis and (iii) the horizontal nature of genetic information acquisition and processing broaden our view of the tree of life. Some researchers insist on the traditional textbook conviction of what is part of the community of life. In a recent review [Moreira, D., Lopez-Garcia, P., 2009. Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol. 7, 306–311.] they assemble four main arguments which should exclude viruses from the tree of life because of their inability to self-sustain and self-replicate, their polyphyly, the cellular origin of their cell-like genes and the volatility of their genomes. In this article we will show that these features are not coherent with current knowledge about viruses but that viral agents play key roles within the roots and stem of the tree of life.

0. Introduction

Within the Darwin project of the European Space Agency (ESA) four or five spacecraft will search for planets similar to earth around other stars and analyse their atmospheres for the chemical signature of life. Its search for extraterrestrial life raise the question of an appropriate definition of life, because it could be expected to find certain indicators on other planets which differ from terrestrial life forms. Current debate on common characteristics of terrestrial living agents assemble traditional textbook convictions such as that life includes those cellular life forms which feed, grow, metabolise and reproduce, i.e. membrane-bound life forms which divide cytoplasmatic space from outer environments. All of these concepts share the opinion that life emerged by interacting non-living chemical components (Greener, 2008).

A useful definition of life must also coherently explain (i) the peripheral phenomena of life, such as cryptobiosis, where the metabolic activity is barely discernible, (ii) the contrasting concept to the selfish gene hypothesis, i.e. the well documented phenomena of symbiogenesis, (iii) the role of non-lytic but persistent virus life-strategies which serve as main regulatory elements in all cellular life forms such as mobile genetic elements and noncoding RNAs and last but not least (iv) the current knowledge of genetic information processing.

In a recent article, Moreira and Lopez Garcia defend the traditional organism based tree and outline their opinion that there are ten reasons to exclude viruses from this tree of life model. At the end of their article they reduce these to four relevant arguments. 'Taken together their inability to self-sustain and self-replicate, their polyphyly, the cellular origin of their cell-like genes and the volatility of their genomes through time make it impossible to incorporate viruses into the tree of life' (Moreira and Lopez Garcia, 2009, p. 311).

As we will see, there are several arguments which contradict this opinion. Viruses are not metabolising cells, true enough. But all the functions within living cells such as replication, transcription and repair as well as their fine-tuned regulatory order are now known to also be of viral origin (Tang et al., 1997; Villarreal, 2005). Therefore it must be considered that viruses have played and still play crucial evolutionary roles and are essential agents within the tree of life. Are the essential agents within the roots of a tree not part of a tree? What would remain of the tree of life if we would substract all viral properties? We will try to show that at least three of the four arguments of Moreira and Lopez-Garcia in particular indicate viruses to be essential parts of the tree of life.

In a first step we will look at prokaryotic genomes and their gene word order determined by their colonizing viral agents. The second step exemplifies the identification of viruses and viral lineages. The third point will argue that the cell first perspective has to outline as good arguments as the virus first perspective. In a fourth point we will see that there is a lot of virus–virus interaction. A fifth chapter will demonstrate that the predecessors of cells had to be polyphyletic. At the sixth point we will see that the ancient and the current RNA-world is a main force in life. A seventh point will show that more than prokaryotes eukaryotic cells also depend on evolutionary roles of viruses, and this offers

* Corresponding author. Tel./fax: +43 6274 6805.
E-mail address: witzany@sbg.at (G. Witzany).

0022-5193/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2009.10.014

solutions to: (i) the origin of the eukaryotic nucleus, (ii) the role of mobile genetic elements and (iii) the role of non-coding RNAs, i.e. key features with viral origins. Last but not least there is a eighth argument introducing semiotic principles that viruses are an essential part of the tree of life and provide coherent explanations of genetic information processing by the biocommunicative approach.

1. Prokaryotic genomes

The arguments of Moreira and Lopez-Garcia to exclude viruses of the tree of life are based on the comparison of viral features and cellular features. But as 'cellular' features they assume in most cases eukaryotic cells. Prior to them we have to look at the prokaryotic genomes and their gene word order if we shall have an appropriate basis of the roles of viruses in the evolution of prokaryotes as hypothesised predecessors of eukaryotes. Interestingly the serial endosymbiotic theory which identified key components of the eukaryotic cell such as mitochondria and chloroplasts as former free living bacteria is supported by the suggestion that the eukaryotic nucleus derived from a large doublestranded DNA virus, which represent the properties of eukaryotic nucleus not present in any known prokaryote (Villarreal, 2005). Additionally this is supported by reviews regarding the role of viral polymerases in the origin of mitochondria and chloroplasts (Filée and Forterre, 2005; Brussow, 2009).

The oceans (and the world) are intensely viral. All life must survive this viral laden habitat and survivors generally retain prophage (or provirus) or their defectives. This includes marine bacteria (Krupovic and Bamford, 2007), extremophiles in deep sea hydrothermal vents (Williamson et al., 2008a) as well as the organisms of the Antarctic (Angly et al., 2006; Dinsdale et al., 2008). If we imagine that 1ml of seawater contains one million bacteria and ten times more viral sequences it can be determined that 10^{31} bacteriophages infect 10^{24} bacteria per second (Tettelin et al., 2005). Since the beginning of life this has been an ongoing process. The enormous viral genetic diversity in the ocean seems to have established pathways for the integration of complete and complex genetic data sets into host genomes, e.g. acquisition of complex new phenotypes. A prophage can provide the acquisition of more than 100 new genes in a single genome editing event (Campbell, 2007; Canchaya et al., 2003a; Brussow et al., 2004; Villarreal, 2009; Ryan, 2009).

It is now 13 years since first bacterial genome was sequenced, and comparative genomics now provides us a very clear picture of prokaryotic evolution; both bacteria and archaea show dominant force of evolution is mediated by horizontal gene transfer (HTG) (Frost et al., 2005; Koonin and Wolf, 2008). Comparing metabolic pathways of 160 prokaryotic species, shows acquisition of gene sets not by point mutation but by rapid and massive acquisition of gene groups (Iwasaki and Takagi, 2009). Clearly, as sex is not common to most bacteria, this is mainly mediated by phage action as such changing gene clusters (aka phage islands) are adjacent to tRNA integration sites. This inherently symbiogenic situation was also apparent with the initial sequencing of *Bacillus subtilis* as the second complete bacterial genome (Sonenshein et al., 2002). Thus, the 'Tree of Life' concept has been severely undermined and cannot apply to such large scale lateral gene transfer processes (Bapteste et al., 2005; Lopez and Bapteste, 2009) or explain the role of viruses (Sinkvics, 2001; DeFilippis et al., 2001; Brussow, 2009). Yet a tree-like structure of genetic evolution is observed in all domains of life, including most viruses. Thus HGT is colonizing an existing tree from non-ancestral (viral) sources. However, 'Tree-thinking' which explains tree growth by ancestral variation

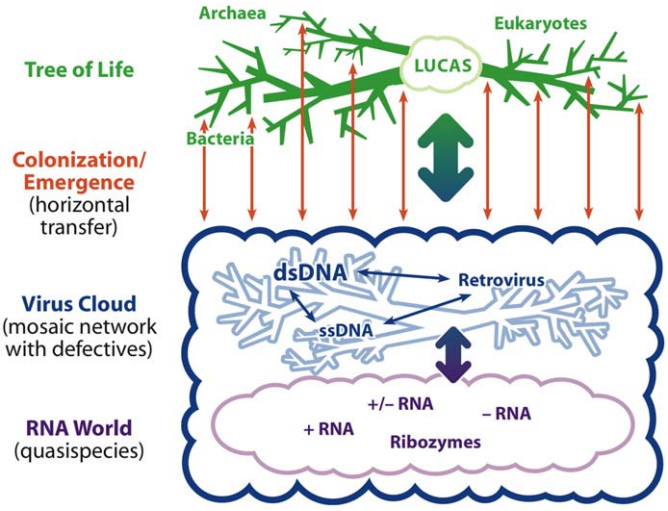

Fig. 1

and natural selection continues to be vigorously defended leading many to dismiss the prokaryotes as 'odd-balls' that evolve differently from other life. Evidence now compels us to revise our definition and vision of the Tree of Life to include viruses. Fig. 1 thus presents a schematic of how an inherently fuzzy virus community (shown as a cloud) provides the information and process of host colonisation. Reticulate evolution and symbiosis apply to all life and must now be incorporated into our conceptual framework (Beiko et al., 2008). We also clearly see recurrent endosymbiosis as a main creative force the evolution of eukaryotic life (Ryan, 2009).

2. We can trace viral genes

The opinions of Moreira and Lopez Garcia on viral lineages is that: (i) there is no viral phylogenetic tree, (ii) viral genes in general have been acquired from their host by HGT, (iii) viral lineages cannot have persistence in the presence of high HGT rates, because genetic contents are not stable but highly volatile and (iv) the cell to virus gene flux is quantitatively overwhelming in contrast to the opposite. But current knowledge on viral lineages contradicts this opinion clearly. The presence of viral specific tRNA genes, their distinct G/C composition and the Restriction–Modification word bias in viral genomes allows us to identify and differentiate viral genes from host using codon and word bias based methods and apply them to DNA viruses of both prokaryotes and eukaryotes (e.g., using non-BLAST methods). This was initially done when a set of T-even phage were first sequenced (Nolan et al., 2006), but has been followed by other viruses (Monier et al., 2007; Pride and Schoenfeld, 2008). Also, it is known that DNA viruses often have an unusually large number of small genes (Villarreal, 2005; Ogata and Claverie, 2007) which makes them good sources of novel but coordinated functions. Indeed, a general assessment of viral ORFs of unknown function indicates a clear trend to code for small genes (Yin and Fischer, 2008) and most ORFans in prokaryotic genomes apparently came from viruses (Cortez et al., 2009). Studies of viruses in the ocean, are consistent with other virus studies and show their evolution is mosaic involving reticulated gene exchange with many other viral lineages, but they tend to most conserve their capsid genes (Rohwer et al., 2000; Rohwer, 2003). We see lots of phage to phage DNA transfers and much evidence of intron related endonucleases involvement (Chibani-Chennoufi et al., 2004).

What has been missing in our thinking about host survival, however, are the effects of virus–virus interactions (i.e, lysogeny and lytic virus). In an intensely viral habitat, host survival depends fully on a dynamic system of virus–virus evolution that promotes the horizontal transfer of DNA. Such transfers of virus derived information are clearly seen in the genomes of all life (Villarreal, 2005, 2009; Ryan, 2009).

3. Cell first vs. virus first perspectives

The authors of the mentioned tree-of-life article share the widely held opinion that viruses are products of cells and are evolved by cells. But phylogenetic analyses and comparative genomics contradict this opinion. Viruses are most abundant agents in the oceanic biosphere and metagenomic screens indicate that 3% of total protein in the oceans correspond with capsid genes of PSSM4-(T4-like) cyanophage. Viruses provide the largest reservoir of genes known in the biosphere (Comeau and Krisch, 2008; Filee et al., 2005) but were not, stolen' from host. Such capsids cannot be of host origin. But these viruses can also have host-like genes. The cyanophage may have a nearly universal presence of pbsA core photosynthetic genes (Bench et al., 2007; Sullivan et al., 2006), clearly resembling genes from host. This similarity of virus and host genes is uncritically cited by Moreira and Lopez Garcia as evidence that such genes are host derived based on the common belief that when viral and host genes cluster together, it is likely that the virus derived the gene from the host. Yet, this phage psb gene is within a highly conserved gene cluster associated with energy metabolism that also contains the mazG gene (Bryan et al., 2008). But this mazG gene does not cluster with related host genes thus it cannot be explained by acquisition from host. We can conclude from these results that co-clustering does not identify the direction of gene transfer. Thus, a viral origin of the psb-A genes can also be asserted. Indeed metagenomic measurements of photosynthetic genes in marine habitats indicate that the majority of all photosynthesis genes are viral derived (Sharon et al., 2007) and are undergoing selection independent of the host (Lindell et al., 2007). As these genes display virus-like codon bias, this further supports the idea that viral photosynthetic genes evolve independently of hosts (Zeidner et al., 2005; Weigele et al., 2007). This argument for viral origin of psb genes is fully developed in Villarreal (2009).

Cyanophage mediate cyanobacterial evolution. Indeed, comparative genomics of Prochlorococcus cyanobacterial biotypes (sub-species variants) indicates that they differ mainly by gene sets or 'phage islands' (Coleman et al., 2006). These surveys: 'lend strong support to the notion that viral-mediated gene acquisition is a common and ongoing mechanism for generating microbial diversity in the marine environment' (Williamson et al., 2008b). Accordingly, it is most commonly observed that recruitment of highly variable genome 'fragments' is the prevalent evolutionary process in the biosphere and such hypervariable genomic islands were too variable to assemble into coherent trees (Rusch et al., 2007).

Additionally it is well accepted by virologists that viruses often contain many complex genes (including core genes) that cannot be attributed to having been derived from host genes. Indeed, comparative viral genomics and the study of virus evolution often depends on precisely such non-host gene conservation (Villarreal, 2005; Domingo et al., 2008). Indeed, there are many indicators that the converse relationship is prevalent as demonstrated by phage genes that colonise host.

Comparative genomics of *Escherichia coli* dramatically demonstrate the importance of large scale genetic variation even within one bacterial 'species'. It is established that *E. coli* genome variation is from 4.6 to 5.5 Mbp and much of this variation appears to be of phage origin (Binnewies et al., 2006). The best bacterial example are the highly sequenced genomes of pathogenic *E. coli*, such as 0157(E2348/69) (Iguchi et al., 2009). Indeed, it has recently become clear that chromosomal integration hotspots are occurring adjacent to lueX tRNA indicating a heavy phage involvement in coli adaptation (Lescat et al., 2009; Kirsch et al., 2004). However, non-pathogenic comensual *E. coli* has similar phage mediated genomic plasticity (Oshima et al., 2008). And it is clear that highly conserved prophage elements are mediating the extraordinary adjacent genomic instability in both pathogenic and commensal isolates (Bielaszewska et al., 2007; Yang et al., 2009). From the comparative genomics of several pathogenic *E. coli* strains, it is also clear that independent infections with similar but distinct bacteriophages were deeply involved in the evolution of these *E. coli* (Ogura et al., 2007). Furthermore, it appears these Ler and Pch phage can orchestrate the coordination of the scattered transcription of various genes involved pathogenicity (Abe et al., 2008). Since a similar 'horizontal' tRNA(leuX) adjacent genetic pattern applies to the differences between related bacterial but distinct bacterial 'species' such as *E. coli* and *Salmonella enterica* (Bishop et al., 2005) or Yersinia (Rakin et al., 2001), the *E. coli* example appears represent a generalised and prominent mechanism of prokaryotic evolution. Indeed, the initial sequencing of a second bacterial species (*B. subtilis*) also identified tRNA adjacent gene sets as the main difference with *E. coli* (Sonenshein et al., 2002).

4. There is much evidence for virus–virus interactions

In prokaryotes, viruses appear to exist into two broad but interacting relationships with host; lytic and persistent (prophage) (Villarreal, 2005; Villarreal, 2007) with two distinct evolutionary dynamics (Gelfand and Koonin, 1997). T4 are strictly lytic phage, and are the iconic example of the T even phage. Although they frequently recombine with each other, T-even phage do not exchange very much DNA with *E. coli* (Nolan et al., 2006). These lytic phage are not 'moronic' (prophage derived small inserts into ORFs, next to promoter) like the persisting lambdoid phage (Hendrix et al., 2003). Their genomes are partitioned and mosaic, via extensive exchange with mostly other phage, but retain expression strategy and morphology (head and tail most conserved). Since phage replication is often recombination dependant, they have notoriously reticulated patterns of evolution. Yet even the most intensively studied phage of all (T4), retains many poorly characterized virus-specific genes (i.e. of the 300 genes, nearly 130 remain uncharacterized and unrelated to other genes). But some of these genes are known to affect virus–virus interactions.

Host fitness is not usually considered from viro-centric perspective. Yet we can assert with confidence that numerically viruses rule the world. Consider the well studies T4 SegB gene which is a homing endonuclease, is highly conserved in T4 (but not other phage) and is needed for preferred inheritance of T4 tRNA gene region (Brok-Volchanskaya et al., 2008). Homing endonuclease are found in both host and virus, but those in T4 are often used to preclude non-T4-like genomes. The T4 SegG appears to provide T2 exclusion (Liu et al., 2003) as does SegF. T4 has many such genes and similar use of endonucleases for genetic exclusion may apply to other T4 like phage, although the specific endonucleases are different (Kadyrov et al., 1997; Sandegren et al., 2005; see Brussow et al., 2004; Chibani-Chennoufi et al., 2004). Indeed group I introns are found in about half of all *Streptococcus*, *Lactococcus*, and *Lactobacillus* phage, often interrupting the lysine gene (Foley et al., 2000). This suggests

wide spread virus–virus based selection which should have big affects on host survival.

The prophage-lytic phage relationship is directly relevant to this issue and has been well examined experimentally with lactic acid bacteria (Desiere et al., 2001, 2002; Canchaya et al., 2003b). Here, it is clear that prophage are major mediators of lateral gene transfer as seen via tRNA adjacent integration (Brussow and Desiere, 2001; Canchaya et al., 2004; Cheetham and Katz, 1995; Brussow, 2007). Lysogenic conversion is also well established (Canchaya et al., 2003a). This type phage-host evolutionary process also appears to apply to the well studied bacterial pathogens (Brussow et al., 2004). And although extensive exchange of genes with host can be inferred, 80% of the viral orthologous groups have no host counterparts (Liu et al., 2006). Even cryptic prophage have been well established to exclude lytic T4 (Mehta et al., 2004; Toothman and Herskowitz, 1980a, b), thus, even viral 'junk' matters greatly to host survival.

5. The predecessors of the three domains of life had to be polyphyletic

There may be several non-cellular origins of viruses consistent with their polyphyletic nature. A precellular RNA world with self-replicating and self-cleaving ribozymes contains the essential host and parts for ancient RNA viruses. But beneath that we can find ancient but related DNA viruses infecting all three domains of life which completely differ from RNA viruses as well as retroviruses (a kind of RNA virus but with reverse reading direction).

If we look at the three domains of life we can identify three different DNA polymerases of the family B extension polymerases of the DNA replication complex which are clearly polyphyletic (Forterre, 2001, 2002, 2005, 2006a) similar to the DNA topoisomerases which also seem to be polyphyletic (Forterre and Gadelle, 2009). This means at the roots of the tree of life there must be at least three polyphyletic predecessors. Only this could explain the three remaining polymerases. Since RNA genomes most probably existed before the appearance of DNA genomes (Villarreal, 2005; Forterre, 2006b), DNA can be considered to be a modification of RNA, requiring ribonocleotide reductase, followed by only two thymidylate synthases.

Prior to DNA, there were only two informational components RNA molecules and proteins. Hypothesised proto-cells initially were colonised by only a few individual DNA components that still operated through RNA. In turn, these DNA agents facilitated the scission into two separate lineages—DNA viruses and RNA viruses—whereby a DNA virus was capable of infecting and persisting as stable DNA in an RNA virus host. Eventually, even more stable colonising genomes developed dsDNA which were recombinogenic and able to capture host genes and establish permanent persistence (symbiosis). This model would at least account for the existence of virally-encoded DNA transaction proteins for which no cellular counterpart exists. In addition, this model would explain the existence of two dissimilar DNA replication systems (Villarreal, 2005).

Even though the DNA cells gained distinct selective advantages, the RNA parasites still had an astonishingly powerful genomic creativity (Ryan, 2006, 2009); this imparted distinct survival advantages if environmental conditions changed considerably. Today we know that the DNA world, on its own, would not have brought forth such an incredible diversity by natural selection (Gabora, 2006), let alone established the necessary genetic precondition to create such a high degree of complexity. Overall, the genomic innovations of the RNA world complements that of the more conservative and stable DNA world. Thus the consortial (commonly shared) volatility and 'lack of structural continuity' (Moreira and Lopez-Garcia, 2009) of viral genomes should be an argument for the inclusion of viruses in the tree of life and provides the precondition for the evolution of the complexity of the tree of life as well as the emergence of a commonly used code for life. Even Darwin did not exclude a multiple origin of life: 'There is grandeur in this view of life, with its several powers, having been originally breathed by the creator into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved' (Charles Darwin, The Origin of Species).

6. The ancient and the current RNA world

In some sections Moreira and Lopez Garcia seem to suggest that the RNA world hypothesis remained a hypothesis and not well proven. However, the RNA-world research indicates a contrasting perspective. From the early RNA-world perspective the whole diversity of processes within and between cells (intracellular and intercellular cell communication) depends on various RNAs. Therefore the nature of modern RNA suggests a precellular RNA world which must have been dominated by quasispecies consortia-based evolution just like current RNA viruses (Domingo et al., 2008). A variety of RNAs can be identified at the roots of the tree of life such as

- genetic polymers inside membrane vesicles of a hypothesised protocell (Chen et al., 2006), riboswitches (Breaker, 2006),
- a variety of catalytic strategies of self-cleaving ribozymes (Ke and Doudna, 2006),
- the structure and function of group I introns (Hougland et al., 2006),
- the roles of RNA in the synthesis of proteins (Moore and Steitz, 2006).

Important features like the role of ribosomes in the translation from RNA into protein (Noller, 2006) and great diversity of action potential of RNAs in the modern DNA world like

- the role of ribonucleoproteins (RNPs) (Cech et al., 2006),
- the list of functions necessary for each single cellular process such as small nuclear RNPs (Tycowski et al., 2006),
- small nucleolar RNPs (Matera et al., 2007),
- the assemblies which build spliceosomes with viral origins,
- the insertion/deletion competence for site-specific modifications of RNA molecules (Simpson, 2006),
- the unique feature of all retroelements, i.e. reverse transcriptase and other telomerases (Blackburn, 2006).

Last but not least we can find crucial parts of the present RNA world with clear viral origin such as

- group II introns with splicing competences (Pyle and Lambowitz, 2006),
- the important roles of SINEs and LINEs (Weiner, 2006),
- the whole range of non-coding RNAs (Witzany, 2009a),
- also the RNA ligase as found in T4 is found in all three domains of life (Ho and Shuman, 2002).

All of these agents can be identified as descendants of an early RNA (quasispecies dominated) world which evolved prior to cellular life and are predecessors of cellular life-functions present

since the last universal common ancestor of all cellular life (LUCA) in all three domains of life. On the basis of these RNAs we are able to reconstruct the emergence of the tree of life from its roots to its top.

Viruses can parasitise almost any replication system (i.e., other viruses including themselves)—even prebiotic ones—and probably emerged well before the appearance of cellular life forms. RNA viruses store crucial and dynamic information (in RNA-'clouds') that not only pertains to: (a) replication proteins but also to (b) morphology and (c) phenotypic diversity and retains a history of past selections. Based on this and the results of phylogenetic analyses and comparative genomics, it is possible to establish viral lines of ancestral origin. These lines of origin can also be non-linear because different parts of viruses contain different evolutionary histories (Villarreal, 2005; Domingo et al., 2008).

Since viruses with RNA genomes are the only living agents that use RNA as a storage medium, they are considered to be witnesses of an earlier RNA world, of a time when DNA did not exist yet (Forterre, 2001, 2002, 2005, 2006b; Koonin et al., 2006). Negatively stranded RNA viruses have genome structures and replication patterns that are dissimilar to all known cell types. As demonstrated by phylogenetic analyses, cellular replicases are related to each other; however, there is no similarity between RNA-viral replicases and those of any known cell types (Zanotto et al., 1996). This suggests the existence of negatively stranded viral RNA-replicases even before cellular life came into being (Villarreal, 2005). DNA viruses, too, do not give any reference to a cellular origin. Phylogenetic analyses point to an older timescale, as DNA-repairing proteins of DNA viruses do not have any counterparts in cellular biota.

7. The Virosphere is not restricted to prokaryotes

Eukaryotes also show that viruses mediated the most dynamic part of their genomes as well. But here, retrovirus related agents are much more active then the DNA prophage or DNA episomes of prokaryotes (see Fig. 1). Also, eukaryotic retroviruses more often seem to manipulate (reprogram) gene regulation and function, not invent new genes. There are also lots of eukaryotic large DNA viruses in the sea as reported by metagenomic screens (via DNA pol, Mimivirus and algal viruses) (Monier et al., 2007, 2008). And we can see major consequences to host by these viruses. For example, *Ostrecoccus tauri*, the smallest marine photosynthetic eukaryote, is host for large DNA virus. This virus lacks restriction/modification enzymes seen in other viruses of microalgae (Derelle et al., 2008). Yet, like cyanobacteria, it too is a highly gene dense chromosome that seems to evolve mostly via the action of horizontal transfer (Palenik et al., 2007). But this eukaryotic host also has two chromosomes that are structurally distinct, have biased G/C content and codon usage and one of which contains the majority of transposable elements (Derelle et al., 2006). Thus even in representatives of early eukaryotes, we can see a large expansion in the kind and quantity of (former parasitic) virus derived DNA. Thus they too have been molded by symbiotic viral mediated events (horizontal, reticulated and symbiotic acquisitions). But what really differs is the nature of virus involved in eukaryotes and how they control their host. Eukaryotes are not gene dense, do not support DNA prophage, but they have acquired regulatory genetic complexity via the action of mostly retroviral (and viral defective) colonisation. We can clearly demonstrate this assertion by comparative genomics. For example, a comparison of human and chimpanzee DNA establishes the prevalent and recent role of endogenous retroviruses (Ryan, 2009). Viruses contributed clearly our recent ancestors. So why would we choose to deny this incontrovertible fact and propose to exclude virus from the tree of life?

7.1. A viral-driven and symbiotic origin of eukaryotes

Serial Endosymbiotic Theory (SET) suggested that the eukaryotic cell did not result from random mutations, but from the coordinated union of former free-living prokaryotes (Margulis, 1996, 1999, 2004; Margulis et al., 2000; Margulis and Sagan, 2002; Witzany, 2006a). So far it is clear that mitochondria and other organelles descended from these micro-organisms (Odintsova and Yurina, 2000, 2005) and it has also been asserted that the eukaryotic nucleus is of archaeal or bacterial descent.

In the meantime, however, there are good reasons to support the idea that eukaryotic nuclei originated before the symbiogenetic integration with mitochondria and chloroplasts (see Villarreal, 2005, 2009). In fact, the nucleus has basic properties that are otherwise absent in prokaryotic cells (Bell, 2001, 2006).

Prokaryotes do have circular chromosomes with uniform standardised origins of replication. Their chromosomes are only loosely attached to chromatin proteins and have different control regions that coordinate and terminate DNA replication. All eukaryotic proteins involved in DNA replication differ from those found in prokaryotes. Hence, nuclear properties of eukaryotes are completely different from those of prokaryotes (Villarreal, 2004). These differences include, for example, use of linear chromosomes, with elaborately controlled and multiple origins of replication, repetitive termination points, transcription and translation which are separated via nuclear membrane and the existence of complex nuclear pore structures that actively mediate RNA translocation.

All these properties represent complex phenotypes, which require complex co-ordination of numerous protein functions. None of these functions can be found in prokaryotes even though they are considered to be the predecessors of the eukaryotic nucleus.

The eukaryotic nucleus contains three kinds of DNA-dependent RNA polymerases that differ significantly from RNA polymerases of prokaryotes (Villarreal, 2005). Even the three kinds of splicing group I-introns (DNA transposase, reverse transcriptase and micro-RNAs) are largely nonexistent in prokaryotes, but they are present in viruses of prokaryotes. In addition, no single prokaryotic process is known to account for the tasks of membrane disintegration and restoration as observed in eukaryotes, but this too can be seen in DNA viruses (Villarreal, 2005).

Viral genes are directly involved in tasks of transposition. Viruses generally mark their genome, their RNA and their proteins with various virus-specific enzymes, such as methylases; e.g. via enzymatic reactions known as base methylation (Villarreal, 2005).

Moreira and Lopez-Garcia suggest that viruses are gene-'robbers' of cellular life (Moreira and Lopez Garcia, 2009, p. 309). However, as asserted above, it is a well-known fact that viruses create new genes as a result of their evolutionary line of descent, via tremendous rates of recombination (often replication dependent) and a high tolerance for errors. Just by looking at the baculoviruses, for example, with reference to GenBank database investigations, we can find that 80% of their genes are unique to this group and found nowhere else (Herniou and Jehle, 2007; Herniou et al., 2001; Villarreal, 2005). Gene losses have been documented in baculoviruses but the 12 losses documented therein are countered by a staggering acquisition of 255 new genes. Similar observations of gene novelty apply to other families of large DNA viruses found in both prokaryotes (Hendrix, 2009) and eukaryotes (Domingo et al., 2008; Villarreal, 2009).

Between 1950 and 1980 scientists realised that the T4 phage-polymerase proteins are much more similar to the eukaryotic DNA polymerase protein than to any prokaryotic polymerase. Today we know that the eukaryotic DNA polymerase and the T4 DNA polymerase do have common origins. Indeed, the T4-like viruses (tailed icosahedral phage, Caudavirales) represent a huge family of viruses that is capable of infecting both bacteria and archaea (Villarreal, 2004). Hence, it is not surprising that T4-DNA polymerases (and capsids) are found in all three domains of life: archaea, bacteria and eukarya—although it must be noted that the phage HK97 capsid protein shared by Caudavirales does not indicate ancestral relationships.

Algae were among the first more complex eukaryotic organisms that had to deal with viruses. Thus, viruses that infected microalgae must have had a large adaptive potential that accompanied the evolutionary pathway and which must have included the characteristic of a protonucleus. Here in particular we can think of the phycodnavirus and therefore one has to examine the entire GenBank database for sequences that may be similar to the DNA polymerase of this particular virus (Villarreal, 2004). Such sequences must include replication polymerases of all higher eukaryotes as well as of all larger eukaryotic DNA viruses, primer polymerases of eukaryotes, and repair polymerases of both archaea and bacteria.

The DNA polymerase of the CSV1 virus can be found at the phylogenetic origin of any of the eukaryotic replication DNA polymerases and can be considered as a precursor of all polymerases which are involved in replication of the eukaryotic genome. So far no other viral or prokaryotic DNA polymerase that shares these features is known (Villarreal, 1999, 2004; DeFilippis and Villarreal, 2001).

The membrane-bound separation of transcription and translation is a characteristic of the pox viruses; more concretely, of the vaccinia and other large DNA viruses (Mimivirus) (Villarreal, 2005). Moreover, these viruses have a very simple pore structure that has actively been incorporated from the membrane-bound RNA into the cytoplasm of the host. A similar situation can be documented with the small chromatin proteins and the linear chromosomes along with their repetitive telomer tails that are so characteristic of various cytoplasmic DNA viruses, TTV1 and phycodna viruses. Even the highly complex function of tubulin as an important coordinating element during chromosomal separation of duplicated strands can be found in DNA viruses with exactly the same set of functions (Villarreal, 2004).

It became increasingly obvious that all properties of the eukaryotic nucleus are compatible with having been derived from a large, stable and persistent DNA virus with linear chromosomes. The precursor of the eukaryotic nucleus indeed appears to have been a huge membrane-covered DNA virus (similar to Mimivirus or poxvirus) that persistently colonised a prokaryotic host (Villarreal, 2004, 2009; Bell, 2001, 2006). Therefore, the hosting cell must have lost its cell wall with the virus incorporating the prokaryotic genes into its pre-nuclear genome: particularly in cases of encoding for metabolism and translation. This virus was probably non-lytic, as it coordinated both its own replication and its transcription genes, and it had a double-layered membrane (reminiscent of that seen in herpesviruses) and a tubulin system in order to wrap chromosomes. Its persistence and its reactivation would imply that: (a) the process of cell division (nuclear envelope dispersion and reformation), (b) mitotic duplication (doubling of the chromosomes and allocating them to the progeny cells) and (c) the viral DNA correspond to the sexual reproductive cycle of the host organism. Such an infectious origin for the nucleus would also be compatible for the prevalence of infectious nuclei, seen in many species of parasitic red algae that represent a basal eukaryote (Goff et al., 1997). Read algae are clearly the oldest eukaryote that can be found in the fossil record. Comparative genomics is consistent with this fossil evidence (Cole and Sheath, 1990). Interestingly all these properties can be found in various prokaryotic viruses such as cyanophage, archaeal phage, mycobacterial phage and eubacterial phage (Villarreal, 2005, 2009).

Not only can the eukaryotic nucleus, however, to be considered of viral origin, but as referenced above, RNA polymerase, DNA polymerase and DNA helicase which transcribe and replicate DNA in modern mitochondria may also be of viral origin (Filée and Forterre, 2005).

7.2. Ongoing roles of retroviruses and mobile genetic elements in eukaryotes

Although DNA prophage no longer modify the eukaryotic genome as seen in prokaryotes, endogenous retroviruses (proviruses) do. The human genome has many more elements derived from retroviruses (LTR containing) then it has genes (ORFs) (Hughes and Coffin, 2001; Kim et al., 2004). And these endogenous retroviruses (ERVs) distinguish human DNA from all other primates (Andersson et al., 2002; Mayer and Meese, 2005; Polavarapu et al., 2006). Such ERVs have colonized Eukaryotic genomes from exogenous viral (non-ancestral) sources. All domains of eukaryotic life have their own peculiar pattern of colonisation by such viruses (Villarreal, 2005, 2009). And it has been experimentally verified that various crucial and complex functions of eukaryotes (vertebrates) are directly mediated by these ERVs (see Ryan, 2009). This includes the emergence of the adaptive immune system (Klein, 2004; Villarreal, 2009) as well as vivipary in mammals (Mi et al., 2000; Dupressoir et al., 2005; Dunlap et al., 2006). The most numerous mobile elements in the eukaryotic genome integrate into a host genome via an RNA intermediate, reverse transcriptase.

Copying from RNA into DNA generally involves reverse transcriptases. Mobile elements are important for genotype processing, with far-reaching consequences for phenotype expression during its various developmental stages (Jurka et al., 2007). Recent research has demonstrated that overlapping epigenetic marking in eukaryotic cells is an important evolutionary feature to silence the expression of mobility of these mobile elements (Slotkin and Martienssen, 2007). Mobile elements can silence single genes as well as larger chromosomal regions and therefore, play an important role in the evolution of diversity. They share their competence to recombine, rearrange and insert into genomic content with other retroelements (Coffin et al., 1997). They influence neighbouring genes through alternative splicing and are active agents as enhancers and promoters or act by polyadenylation patterns (Slotkin and Martienssen, 2007). Indeed, related elements appear to account for the origin of a substantial part of the regulatory sequences in the human genome (Jordan et al., 2003).

ERV related retroposons have direct repeats at its ends (LTR), others transposons do not (non-LTRs). Interestingly, the number of retroposons increases with every transposition (transposition duplication) so that they can expand host genomes: LINE-1 is 20% of the human genome (Maita et al., 2004). Like some ERVs, full LINEs contain a code for the transposase protein, which have been proposed to be evolutionary related to retroviral and ERV integrase (Capy et al., 1997). This enzyme identifies the terminal inverted repeats which flank mobile elements, excises them and integrates itself in place of them. The gap at the donor site is repaired in a cut-and-paste transposition or filled up with a copy of the transposon by a gap repair technique (Slotkin and Martienssen, 2007). Although many think LINES are independent of ERVs, deep genome database comparison to early ERVs (Chromoviruses, DIRs) suggests that most transposons and

retroposons originally descended from recombinant viruses which persistently integrated into host genomes and became defective (for references see Villarreal, 2004, 2005, 2009; Domingo et al., 2008; Weiss, 2006).

7.3. Non-lytic but persistent viruses as non-coding regulatory agents for cellular needs

Like persisting prophage of prokaryotes, the persistence of retrovirus derived elements in eukaryotes have major consequences. Principally, they provide a diffuse but coordinated system to edit and control the genome, but also affect virus susceptibility. Some thousands of endogenous retroviral sequences have been integrated into the human genome, and until now there are 22 independent retroviral families identified (Bannert and Kurth, 2004; Bromham, 2002; Buzdin et al., 2002; Hughes and Coffin, 2001; Khodosevich et al., 2002; Sverdlov, 2000; Villarreal, 2004, 2005; Ryan, 2006; Griffith and Voisset, 2008). A quantity of remaining former viral gene embedding repetitive elements embracing an enormous genetic diversity originally accompanied the protein coding sequences as control- and/or identification segments. Most endogenous retroviruses have been degraded into formerly connected domains, but they can still be recognised by their three genes gag, pol and env (Gao et al., 2003; Ryan, 2004).

It became obvious that mobile sequences such as transposons and retroposons (Volff, 2006) and non-coding repetitive elements such as LTRs, SINEs, LINEs (long terminal repeats, short interspersed elements, long interspersed elements) make far-reaching eukaryotic DNA rearrangement and reorganisation possible (Shapiro, 2002; Sternberg, 2002; Shapiro and Sternberg, 2005). Together, they play a decisive role in the evolution of new genomic structures (Shabalina and Spiridonov, 2004; Shapiro, 2004; Sternberg and Shapiro, 2005). Being dependent on the state of development, the varying chromatin markers are thus capable, through different methylation patterns, histone modifications and alternative splicing, of coming up with a set of multiple protein meanings, from one and the same genetic dataset (Turner, 2000, 2002; Jenuwein and Allis, 2001; Brett et al., 2002; Xu et al., 2002; Jaenisch and Bird, 2003; True et al., 2004; Wang et al., 2004). This marks even the rise of epigenetics and its research object that phenotypic variations, which are even heritable, must not depend on genetic alterations (Jablonka and Lamb, 1989, 2002, 2006; Van De Vijver et al., 2002; Van Speybroeck et al., 2002). The question arises as to how and why the evolution of higher genetic complexity is connected to non-coding virus derived DNA, that was formerly called ‘junk’ DNA.

Although it has been known for several decades that the unbelievable diversity of enzyme proteins is a practical tool for DNA editing processes, it was unclear by which rules or higher-order regulations they are governed (Witzany, 1995, 2000, 2005). Later on it became obvious that higher-order regulations such as co-suppression, suppression of transposition, position effect variegation, several start- and stop-signals, RNA interference, imprinting, chromosomal methylation, transvection and transcriptional and post-transcriptional gene silencing are processed by non-protein-coding RNAs, especially micro-RNAs (Mattick, 2001, 2003, 2005; Mattick and Gagen, 2001). New research indicates that these repetitive non-coding sequences originated primarily from retroviral RNA (Villarreal, 2004, 2005).

The ERV related and repetitive sequences are highly species-specific and more suitable for the determination of species than the corresponding coding sequences (Villarreal, 2005, 2009). Each taxon organises and formats its genome architecture differently, i.e. regulation of expression, transcription, replication, and translation are species-specific. Within each taxon, these processes along with the associated gene architecture must run in a highly coordinated manner so that they do not disturb each other. Only a co-ordination of the individual steps ensures that these different actions are performed and maintained successfully; i.e. DNA sequences must be read at the right site and at the right time. The precise spatio-temporal coordinations are essential in order to sustain vital processes; nonetheless, and in principle, these coordinations can also fail, either by sequence-damage (i.e. mutations) or by organism-induced translational, transductional, repairing or other rearrangement disturbances.

Experience has shown that: (1) excision, (2) insertion and (3) combination of the genomic texts are the keys in DNA editing and the basis of evolutionary processes. In contemporary terms, we could call it ‘natural genetic engineering’ (Shapiro, 2004).

8. The tree of life from a biocommunicative perspective

Finally, we should look at an additional argument for integration of viral consortia into the tree of life which is not part of cell biological and molecular biological but of bioinformatic perspectives. Although bioinformatics serves as important tool in detection and comparing sequences of genetic texts we must look at complementary perspectives on genetic information processing such as biolinguistics, biosemiotics or the biocommunicative approach. But is not it a commonly shared and widely accepted assumption that molecules of nucleic acid sequences are the result of randomly derived mixtures? How should we think about evolution editing biological codes? What would be needed? Agents like viral or virus-like consortia can coordinate their behavioural strategies via an inherent capacity to identify sequence specific content arrangements. Can this allow them to integrate a defense of an attack, and invent (i) de novo (elongate) code or (ii) edit code through rearranged new sequence combination (recombination), (iii) and/or provide alternative regulations? Can viruses thus provide new coordinated code sequences?

Nevertheless it seems to be reasonable to reflect on this feature because recent knowledge indicates the fundamental role of information processing in both viral life strategies and key regulatory novelty in cell biology.

If we look at life from an encoding and editing perspective in a recent article Patrick Forterre suggests defining life as both ‘ribosome encoding organisms and capsid-encoding organisms and their ancestors’ (Forterre and Prangishvili, 2009). Yet both encode (using the same molecular alphabet) albeit encoding different products. They share a competence, i.e. to code, in contrast with entities which do not share the competence to encode. But what does ‘encode’ mean? It has something to do with a code, i.e. the genetic code, which as DNA serves as an information storage medium, as RNAs serves as information-based editing agents. But ‘information’ is something which seems to be very useful the first time, but not the second time. There are approximately 60 different definitions of ‘information’, many of them incompatible with each other.

Several scientific approaches focus on the linguistic-like structure and function of the genetic code such as biolinguistics (Popov et al., 1996; Ji, 1997, 1999; Searls, 2002; Chomsky, 2004; Zhang, 2006), biohermeneutics (Chebanov, 1994; Markos, 2002), biosemiotics (Florkin, 1974; Sebeok and Umiker-Sebeok, 1992; Hoffmeyer, 1996; Barbieri, 2001, 2007), protein-linguistics (Gimona, 2006) and biocommunicative approach (Witzany, 1995, 2000, 2009b). Their common focus: a code is a language-like structure. This at first glance seems to be difficult because investigations on ‘language’ and its concrete use in ‘communication’ are not core

competences of natural sciences. Any trials of a coherent and sufficient definition of 'language' and 'communication' by information theoretical, cybernetic, systems theoretical, mathematical, statistical, and mechanistic methods failed (Witzany, 2000), because formal or even comparative analysis of the combinatorial patterns of sign sequences, i.e. the syntactical level, cannot identify the pragmatic *interactional* contexts in which living agents are interwoven that determine meaning/functions of sign sequences.

For example biolinguistics interprets and investigates genetic structures in the light of linguistic categories (Popov et al., 1996; Ji, 1997, 1999; Searls, 2002; Chomsky, 2004; Zhang, 2006). Similarly to bioinformatics they use statistical methods and algorithms to identify sequence orders for measurements of sequence-length and content homologies. Biolinguistics follows bioinformatics and its model of language as a quantifiable set of signs and is still convinced that it would be possible to extract semantic contents by analysis of the 'universal syntax'. In a limited sense this is possible, e.g. in genetic sequence comparison, i.e. comparative genomics. But an unambiguous determination of genetic semantics through analysis of the molecular syntax of genetic code is not possible in principle, because analysis of the molecular syntax does not tell us anything about the context in which the genetic content bearer of the genetic information is interwoven in real life. This context plays an important role in epigenetic imprinting and therefore in the construction of different methylation patterns which then are the determinants for alternative splicing pathways of the same genetic datasets. This crucial role of pragmatic contexts is not part of the methods of biolinguistics and bioinformatics.

One result of these deficiencies is that invention of new and even complex genetic data sets or, as they may be called, gene blocks and the coherent integration of new genes or gene blocks in pre-existent genetic content arrangements by competent agents is not part of bioinformatics or biolinguistics, because innovative generation of new genetic content which is not randomly assembled cannot be deduced out of a mathematic model of language, i.e. formalisable procedures such as algorithms.

As in any language there are certain characteristics of the genetic code which are fundamental and cannot be reduced to each other. This means they have complementary functions: a language/code needs an alphabet, i.e. signs which can be combined in certain ways with the result of different modes of combination. The rules (not laws) of correct combination we term syntax. Manfred Eigen spoke of the 'molecular syntax of nucleotide sequences' (Eigen and Winkler, 1975). Eigen also introduced the term quasispecies (in the oversimplified meaning of a 'master template'), but we now are able to understand this to be a consortial and dynamic system of virus information processing, adaptation and evolution with no master code (Domingo et al., 2008), but with coordinated capacity (i.e. competence) to edit. In terms of philosophy of science Eigen followed the most prominent 'linguistic turn' i.e. syntax analyses should offer understanding of semantic contents of code-sequences exclusively, and dismissed primacy of pragmatics, i.e. real-life situations in which sequence generating consortia are interwoven (Witzany, 1995).

A code or a language, however, does not code itself nor does a language speak (Witzany, 2000, 2006b). A code or a language (nucleic acid language) is generated and used by some living agents in real life-worlds to coordinate and organise via communication processes (Witzany, 2007), i.e. they are interwoven in constantly changing environmental processes. To balance life *interactionally* in real-life habitats is a kind of behaviour, in contrast with non-behavioural cause-effect reactions, underlying natural laws strictly. By using codes or languages, living agents can modulate the information that determines their behaviour. To generate different modes of behaviour in a coordinated way living agents follow not syntactic rules but pragmatic rules. According to different kinds of pragmatic situational-contexts living agents are able to use a limited number of signs (alphabet) to generate different messages, i.e. varying sign-sequence arrangements. This means that an identical sign sequence can be used to transport different messages according to the different needs of the living agent as a result of the different situational contexts in which it is interwoven. Therefore—this is really important—pragmatics determines the meaning/function of code-sequences or language-sequences (we term 'sentences').

Within the three semiotic (sign-theoretical) levels of rules (syntax, pragmatics, semantics) which are essential characteristics of any real code or language it must be noted that no single living agent is capable of generating or using a code or a language. Ludwig Wittgenstein once noted that is not possible for only one person to follow a rule only once (Wittgenstein, 1972). The capability to follow rules depends on a historically grown 'culture' (customs) of social interactors, i.e. group behaviour. The emergence of codes or languages needs, as a fundamental precondition, a consortium of living agents in principle. Therefore it seems likely that living agents being competent to encode coherent content sequences in the nucleic acid language emerged in parallel and in great number. In this, it is thus most interesting that RNA virus evolution and adaptation is fundamentally consortial (Domingo et al., 2008). And because they most likely evolved prior to DNA based cellular life forms also the assumption of a single last universal common ancestor (LUCA) as suggested by the authors of the current review is less plausible. If nucleic acid coding agents invented membrane bound cellular life which implies a commonly shared translational code it must be assumed that they emerged at once or in parallel in great numbers so that we must speak in the plural of LUCAs (Witzany, 2008) as being quasispecies based and reticulated.

Should we term these coding agents as viruses or subviral RNA species? Viruses themselves do not differentiate this issue very much. The multiplicity reactivation and/or complementation of defective individuals described in several articles (Bailey et al., 2004; Villarreal, 2005; Zeidner et al., 2005), is common and occurs at a critical level of viral parts such that these viral parts are able to recreate a complete virus after being damaged by e.g. UV radiation. This indicates that competent code editing could also occur on a subviral ribozymatic level by an otherwise defective consortia.

Not even one known natural code or language is the result of a random-like mixture of signs/parts of an alphabet. Any structure which functions as code or language has to be generated by a consortium of living agents which share a common competence to use signs according to basic semiotic (syntactic, i.e. combinatorial, pragmatic, i.e. context-sensitive, semantic, i.e. content-specific) rules. These rules in most cases are very conservative but in contrast with natural laws they are changeable in principle, which is an advantage for adaptational purposes as well as for the generation of new sequences *de novo* which never existed before. Viruses inherently provide this essential capacity.

This view about 'What is life?' avoids suggesting that the term 'organism' is crucial to the tree of life, because it may be paradox to describe an ancient hammerhead ribozyme with self-cleaving competence as an organism. 'Organism' from this point of view seems to be appropriate for cellular life forms. Although viruses and subviral agents are both driving forces in the evolution of organisms and represent key functions of all cellular organisms they cannot be considered to be organisms even if sometimes able to re-animate organisms. From the biocommunicative perspective current knowledge about viruses and their varying life strategies indicates their driving force and crucial role in the evolution of all

cellular life forms and key processes as well as key regulations of cells. Therefore the 'origin of life' at the roots of the tree of life from this perspective started with the origin of agents with encoding competences, a feature which is absent in non-animated nature. There are no (semiotic) rule-following agents if water freezes to ice (Witzany, 2009b).

From the biocommunicative aspect life starts with the competent interaction between swarms of ribozymes, i.e. based on identity/difference of encoded (molecular) syntactic order of nucleotides. This does not include the evolution of cellular metabolism at this stage even if some compartmentalisation is needed such as a different molecular syntactic structures. This means that we must decide whether ribozymatic swarms on extraterrestrial planets we term presence of life or presence of proto-life. This leads into a paradigmatic discussion with cellular textbook conviction on the one side and bioinformatic innovation at the other. At the current stage the position in our article is rather conservative but does not exclude further developments.

9. Conclusion

Phylogenetic analyses as well as GenBank database and genome comparison show that most adaptations are of external (horizontal) and mostly viral origin. Thus natural genome-editing competences are not of cellular origin but represent original skills of viruses. Viruses have two completely different life strategies, which are clearly reflected in their genomes. In comparison, acute viruses that exhibit lytic action induce disease and even death, whereas the life strategy of persistent viruses implies compatible interactions with the host, either by being integrated into the hosting genome or within the cell plasma, and act non-destructively during most life stages of the host. These two viral life strategies often oppose one another in a population dependent way.

The persistent lifestyle allows the virus to transmit complex viral phenotypes to the hosting organism. This enables the host to broaden evolutive potentials that may well lead to the formation of new species. This, along with providing resistance to lytic virus, is an advantage for host survival also for the persisting agent.

The natural genome-editing competences of viruses are most complex in prokaryotes, in which the complete nucleotide word order is largely determined, combined, and recombined by viruses. Hence, the main genomic novelties are found in the prokaryotic domain from where they originally evolved into the higher life forms. Probably all basic enzymatic variations originated therein. Massive viral colonisation occurred from the very beginning of life starting with the evolution of bacteria and archaea, and later that of protoctista and multicellular eukaryota. The formation of all kingdoms, their families, genera, and species relies on the effects of viral colonisation and results in diversified lineages and ultimately in the evolution of new species.

Increasing complexity and diversity are caused by genetic innovations, new combinatorial patterns of genetic content, non-coding regulatory networks and modifications of the genomic architecture. Interestingly, increasing eukaryotic complexity correlates with expansion of non-coding DNA (Taft and Mattick, 2004). Combinatorial and rearranging processes in evolutionary and developmental processes occur non-randomly. They need to successfully and coherently abide by the rules of molecular syntax (Eigen and Winkler, 1975; Witzany, 1995). Genetic content arrangements within the genomic matrix depends on situational contexts in which living organisms are involved in vivo (e.g. context of growth, mating, virulence, stress, etc.) and is therefore able to produce multiple protein meanings, i.e. different semantic contents of the same genetic dataset. Undoubtedly this promotes the development of regulatory complexity. This is the prerequisite for epigenetically-induced evolutionary and developmental processes. These rule-following processes may even fail, with fatal consequences for the organism (Witzany, 2006b). If evolutionary processes are intertwined between different species complexity is even more evident (Villarreal, 2005; Zhang, 2006). This indicates an important role of symbiogenetic processes in enhancing genetic, genomic and phenotypic complexity and diversity (Villarreal, 2007; Ryan, 2009).

Obviously, evolutionary history emerges from a totally different circumstance than previously thought. It must not be understood as an aggregation of chance mutations of the genetic text and its associated selection, but more as permanent and competent processing of genetic sequences for the purpose of acquisition of previously unknown abilities as an advantageous productivity to ward off competing parasites from host organisms via genomic innovation.

The very genetic volatility of viruses, used as argument to exclude them from the tree of life must now be considered as an essential *precondition* for life. Inventions of new genetic sequences and integration in host organisms change their genetic identity and are an advantage against competing genetic parasites which are not incorporated in host genomic contents. But inventions of new genetic sequences must occur through these competitors also, because otherwise they cannot colonise ever-changing host genomes. Genetic inventions by competent genetic content editors therefore must be considered as the driving force of evolution in both viruses and cellular host genomes. The changes of DNA-stored data by random mutations solely would never result in the great variety of new genetic content arrangements. Only the creative 'error-prone' consortial RNA gene pool (quasispecies) with its fast-changing genetic content arrangements and its selection (being stored in the DNA 'evolutionary protocols'. See Vetsigian et al., 2006) can serve as an appropriate explanatory concept. Thus we can start to understand why our own genome is composed of such a large and varied population of 'defective' virus which now act in their novel roles as 'effective' regulatory tools in host genomes.

A solution to the problem of cryptobiosis as mentioned in the introduction can be suggested now also. It is considered to be a contrasting phenomenon to actual life processes, because e.g. larvae of the African fly Polypedilum vanderplanki survive desiccation for up to 17 years and temperatures ranging from $-270\,^{\circ}C$ (liquid helium) to $106\,^{\circ}C$ (Watanabe et al., 2002) or some bacterial endospores are able to withstand almost any environment. Isolated endospores of thermophiles from cold lake sediments could be revived from samples some 100,000 years old (Nicholson et al., 2000). Also virions (in oceans) clearly represent a conserved possibility of living agents ready for actualisation under the appropriate environmental conditions such as virions of the ATV (Acidianus-Tailed Virus). These virions can go for months without any change of their morphology at normal temperatures. When they are exposed to higher temperatures ($70\,^{\circ}C$) they dramatically change their structure by forming two tails at the end of their central body (Häring et al., 2005) which may be a precondition for biologically required changes in morphology and structure (Prangishvili et al., 2006). If these environmental conditions are lacking cryptobiotic agents remain inactivated, a complete and long lasting absence of replication or any metabolising activity will occur. Would we assume them as not being part of the tree of life? Under appropriate environmental conditions they will start as typical for living agents. Their DNA data set is activated according the inherited regulatory pathways as no inactivation ever has occurred. At this point we should remember that the gene word order of the DNA datasets as well as the regulatory

ratio is the result of persistent viral colonisation. Should we assume them not being part of the tree of life?

Like any tree, the tree of life is not constituted by cells of the stem, branches and leaves alone. As we know today, the stem of actual trees is in continuous communication with the rhizosphere to coordinate growth and development. Without a balanced rhizosphere ecology trees will have serious problems of survival. This rhizosphere ecology is a particular example of non-selfish interactional (symbiotic) patterns and co-evolution (Witzany, 2006c). It crucially depends on communication processes which function in parallel between three different types of plant root cells, mychorizal fungi and rhizobia bacteria, all of them being capable of self/non-self differentiation to coordinate and organise group behaviour. Surely, this fundamental capability to coordinate group behaviour by communication processes in bacteria, fungi and plants must inherently be of viral origin. Persistent viral agents which colonised all of these organisms which are characteristic of rhizosphere ecology generated a great variety of host features with which these hosts are able to identify and communicate with group members. If the tree of life metaphor is to be useful in future, we have to remember that viruses have been and still are essential agents within the roots and stem of the tree of life.

References

Abe, H., Miyahara, A., Oshima, T., Tashiro, K., Ogura, Y., Kuhara, S., Ogasawara, N., Hayashi, T., Tobe, T., 2008. Global regulation by horizontally transferred regulators establishes the pathogenicity of *Escherichia coli*. DNA Res. 15 (1), R25–R38.

Andersson, A.C., Venables, P.J., Tonjes, R.R., Scherer, J., Eriksson, L., Larsson, E., 2002. Developmental expression of HERV-R (ERV3) and HERV-K in human tissue. Virology 297, R220–R225.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J.M., Mueller, J.E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C.A., Rohwer, F., 2006. The marine viromes of four oceanic regions. PLoS Biol. 4 (11), e368.

Bailey, S., Clokie, M.R., Millard, A., Mann, N.H., 2004. Cyanophage infection and photoinhibition in marine cyanobacteria. Res. Microbiol. 155, R720–R725.

Bannert, N., Kurth, R., 2004. Retroelements and the human genome: new perspectives on an old relation. Proc. Natl. Acad. Sci. USA 101 (Suppl. 2), R14572–R14579.

Bapteste, E., Susko, E., Leigh, J., MacLeod, D., Charlebois, R.L., Doolittle, W.F., 2005. Do orthologous gene phylogenies really support tree-thinking?. BMC Evol. Biol. 5, R33.

Barbieri, M., 2001. The organic codes. The birth of semantic biology. PeQuod. Ancona.

Barbieri, M. (Ed.), 2007. Introduction to Biosemiotics. Springer, Dordrecht.

Beiko, R.G., Doolittle, W.F., Charlebois, R.L., 2008. The impact of reticulate evolution on genome phylogeny. Syst. Biol. 57, R844–R856.

Bell, P.J.L., 2001. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA Virus?. J. Mol. Evol. 53, R251–R256.

Bell, P.J.L., 2006. Sex and the eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. J. Theor. Biol. 243, 54–63.

Bench, S.R., Hanson, T.E., Williamson, K.E., Ghosh, D., Radosovich, M., Wang, K., Wommack, K.E., 2007. Metagenomic characterization of Chesapeake Bay virioplankton. Appl. Environ. Microbiol. 73 (23), R7629–R7641.

Bielaszewska, M., Dobrindt, U., Gartner, J., Gallitz, I., Hacker, J., Karch, H., Muller, D., Schubert, S., Alexander Schmidt, M., Sorsa, L.J., Zdziarski, J., 2007. Aspects of genome plasticity in pathogenic *Escherichia coli*. Int. J. Med. Microbiol. 297 (7–8), R625–R639.

Binnewies, T.T., Motro, Y., Hallin, P.F., Lund, O., Dunn, D., La, T., Hampson, D.J., Bellgard, M., Wassenaar, T.M., Ussery, D.W., 2006. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. Funct. Integr. Genomics 6 (3), R165–R185.

Bishop, A.L., Baker, S., Jenks, S., Fookes, M., Gaora, P.O., Pickard, D., Anjum, M., Farrar, J., Hien, T.T., Ivens, A., Dougan, G., 2005. Analysis of the hypervariable region of the *Salmonella enterica* genome associated with tRNA(leuX). J. Bacteriol. 187 (7), R2469–R2482.

Blackburn, E.H., 2006. Telomerase RNA. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 419–436.

Breaker, R.R., 2006. Riboswitches and the RNA World. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 89–107.

Brett, D., Pospisil, H., Valcarcel, J., Reich, J., Bork, P., 2002. Alternative splicing and genome complexity. Nat. Gen. 30, R29–R30.

Brok-Volchanskaya, V.S., Kadyrov, F.A., Sivogrivov, D.E., Kolosov, P.M., Sokolov, A.S., Shlyapnikov, M.G., Kryukov, V.M., Granovsky, I.E., 2008. Phage T4 SegB protein is a homing endonuclease required for the preferred inheritance of T4 tRNA gene region occurring in co-infection with a related phage. Nucleic Acids Res. 36 (6), R2094–R2105.

Bromham, L., 2002. The human zoo: endogenous retroviruses in the human genome. Trends Ecol. Evol. 17, R91–R97.

Brussow, H., 2007. Bacteria between protists and phages: from antipredation strategies to the evolution of pathogenicity. Mol. Microbiol. 65, R583–R589.

Brussow, H., 2009. The not so universal tree of life or the place of viruses in the living world. Trans. R. Soc. B 364, R2263–R2274.

Brussow, H., Desiere, F., 2001. Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol. Microbiol. 39, R213–R222.

Brussow, H., Canchaya, C., Hardt, W.D., 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68, R560–R602.

Bryan, M.J., Burroughs, N.J., Spence, E.M., Clokie, M.R., Mann, N.H., Bryan, S.J., 2008. Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. PLoS ONE 3 (4), e2048.

Buzdin, A., Khodosevich, K., Mamedov, I., Vinogradova, T., Lebedev, Y., Hunsmann, G., Sverdlov, E., 2002. A technique for genome-wide identification of differences in the interspersed repeats integrations between closely related genomes and its application to detection of human-specific integrations of HERV-K LTRs. Genomics 79, R413–R422.

Campbell, A., 2007. Phage integration and chromosome structure. A personal history. Annu. Rev. Genet. 41, R1–R11.

Canchaya, C., Fournous, G., Brussow, H., 2004. The impact of prophages on bacterial chromosomes. Mol. Microbiol. 53 (1), R9–R18.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., Brussow, H., 2003a. Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. 6 (4), R417–R424.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., Brussow, H., 2003b. Prophage genomics. Microbiol. Mol. Biol. Rev. 67 (2), R238–R276 table of contents.

Capy, P., Langin, T., Higuet, D., Maurer, P., Bazin, C., 1997. Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor?. Genetica 100, R63–R72.

Cech, T.R., Moras, D., Nagai, K., Williamson, J.R., 2006. The RNP World. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 309–326.

Chebanov, S., 1994. Man as participant to natural creation. Enlogue and ideas of hermeneutics in biology. Riv. Biol. 87, R39–R55.

Cheetham, B.F., Katz, M.E., 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. 18 (2), R201–R208.

Chen, I.A., Hanczyc, M.M., Sazani, P.L., Szostak, J.W., 2006. Protocells: genetic polymers inside membrane vesicles. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 58–88.

Chibani-Chennoufi, S., Canchaya, C., Bruttin, A., Brussow, H., 2004. Comparative genomics of the T4-Like *Escherichia coli* phage JS98: implications for the evolution of T4 phages. J. Bacteriol. 186 (24), R8276–R8286.

Cole, K.M., Sheath, R.G., 1990. Biology of the Red Algae. Cambridge University Press, New York.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F., Chisholm, S.W., 2006. Genomic islands and the ecology and evolution of Prochlorococcus. Science 311, R1768–R1770.

Comeau, A.M., Krisch, H.M., 2008. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. Mol. Biol. Evol. 25 (7), R1321–R1332.

Chomsky, N., 2004. Biolinguistics and the human capacity. Delivered at MTA, Budapest, May 17, 2004.

Coffin, J.M., Hughes, A.H., Varmus, H.E., 1997. Retroviruses. Cold Spring Harbor Laboratory 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 (6), R65.

DeFilippis, V., Villarreal, L.P., Salzberg, S.L., Eisen, J.A., 2001. Lateral gene transfer or viral colonization?. Science 293, 1048a.

DeFilippis, V.R., Villarreal, L.P., 2001. Virus evolution. In: Knipe, D.M., Howley, P.M., Griffin, D.E. (Eds.), Fundamental Virology fourth ed Lippincott Williams & Wilkins, Philadelphia, pp. 353–370.

Derelle, E., Ferraz, C., Escande, M.L., Eychenie, S., Cooke, R., Piganeau, G., Desdevises, Y., Bellec, L., Moreau, H., Grimsley, N., 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. PLoS ONE 3 (5), e2250.

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynie, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piegu, B., Ball, S.G., Ral, J.P., Bouget, F.Y., Piganeau, G., De Baets, B., Picard, A., Delseny, M., Demaille, J., Van de Peer, Y., Moreau, H., 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc. Natl. Acad. Sci. USA 103 (31), R11647–R11652.

Desiere, F., Lucchini, S., Canchaya, C., Ventura, M., Brussow, H., 2002. Comparative genomics of phages and prophages in lactic acid bacteria. Antonie Van Leeuwenhoek 82 (1–4), R73–R91.

Desiere, F., McShan, W.M., van Sinderen, D., Ferretti, J.J., Brussow, H., 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic Streptococci: evolutionary implications for prophage-host interactions. Virology 288 (2), R325–R341.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M.A., Nelson, K.E., Nilsson, C., Olson, R., Paul, J., Brito, B.R., Ruan, Y., Swan, B.K., Stevens, R., Valentine, D.L., Thurber, R.V., Wegley, L., White, B.A., Rohwer, F., 2008. Functional metagenomic profiling of nine biomes. Nature 452 (7187), R629–R632.

Domingo, E., Parrish, C.R., Holland, J.J., 2008. Origin and Evolution of Viruses, second ed Academic Press, San Diego.

Dunlap, K.A., Palmarini, M., Varela, M., Burghardt, R.C., Hayashi, K., Farmer, J.L., Spencer, T.E., 2006. Endogenous retroviruses regulate periimplantation placental growth and differentiation. Proc. Natl. Acad. Sci. USA 103, R14390–R14395.

Dupressoir, A., Marceau, G., Vernochet, C., Benit, L., Kanellopoulos, C., Sapin, V., Heidmann, T., 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc. Natl. Acad. Sci. USA 102, R725–R730.

Eigen, M., Winkler, R., 1975. Das Spiel-Naturgesetze steuern den Zufall. Piper, München.

Filée, J., Forterre, P., 2005. Viral proteins functioning in organelles: a cryptic origin. Trends Microbiol. 13, R510–R513.

Filee, J., Tetart, F., Suttle, C.A., Krisch, H.M., 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc. Natl. Acad. Sci. USA 102 (35), R12471–R12476.

Florkin, M., 1974. Concepts of molecular biosemiotics and of molecular evolution. In: Florkin, M., Stotz, E.H. (Eds.), Comparative Biochemistry. Elsevier, Amsterdam.

Foley, S., Bruttin, A., Brussow, H., 2000. Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. J. Virol. 74 (2), R611–R618.

Forterre, P., 2001. Genomics and early cellular evolution. The origin of the DNA world. CR. Acad. Sci. III 324, R1067–R1076.

Forterre, P., 2002. The origin of DNA genomes and DNA replication proteins. Curr. Opin. Microbiol. 5, R525–R532.

Forterre, P., 2005. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie 87, R793–R803.

Forterre, P., 2006a. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. 117, R5–R16.

Forterre, P., 2006b. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc. Natl. Acad. Sci. USA 103, R3669–R3674.

Forterre, P., Gadelle, D., 2009. Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. Nucleic Acids Res. 37, R679–R692.

Forterre, P., Prangishvili, D., 2009. The origin of viruses. Res. Microbiol. 160 (7), R466–R472.

Frost, L.S., Laplae, R., Summers, A.O., Toussaint, A., 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. 3, R722–R732.

Gabora, L., 2006. Self-other organization: why early life did not evolve through natural selection. J. Theor. Biol. 241, R443–R450.

Gao, X., Havecker, E.R., Baranov, P.V., Atkins, J.F., Voytas, D.F., 2003. Translational recoding signals between gag and pol in diverse LTR retrotransposons. RNA 9, R1422–R1430.

Gelfand, M.S., Koonin, E.V., 1997. Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. Nucleic Acids Res. 25, R2430–R2439.

Gimona, M., 2006. Protein linguistics—a grammar for modular protein assembly?. Nat. Rev. Mol. Cell Biol. 7, R68–R73.

Goff, L.J., Ashen, J., Moon, D., 1997. The evolution of parasites from their hosts: a case study in the parasitic red algae. Evolution 51, R1068–R1078.

Greener, M., 2008. It's life but just as we know it. EMBO Rep. 9, R1067–R1069.

Griffith, D., Voisset, C., 2008. Human endogenous retroviruses: from ancestral pathogens to bona fide genes. Microbiol. Tod. 2, R32–R33.

Häring, M., Vestergaard, G., Rachel, R., Chen, L., Garrett, R.A., Prangishvili, D., 2005. Virology: independent virus development outside the host. Nature 436, R1101–R1102.

Hendrix, R.W., 2009. Jumbo bacteriophages. Curr. Top. Microbiol. Immunol. 328, R229–R240.

Hendrix, R.W., Hatfull, G.F., Smith, M.C., 2003. Bacteriophages with tails: chasing their origins and evolution. Res. Microbiol. 154, R253–R257.

Herniou, E.A., Jehle, J.A., 2007. Baculovirus phylogeny and evolution. Curr. Drug Targ. 8, R1043–R1050.

Herniou, E.A., Luque, T., Chen, X., Vlak, J.M., Winstanley, D., Cory, J.S., O'Reilly, D.R., 2001. Use of whole genome sequence data to infer baculovirus phylogeny. J. Virol. 75, R8117–R8126.

Ho, C.K., Shuman, S., 2002. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. Proc. Natl. Acad. Sci. USA 99, R12709–R12714.

Hoffmeyer, J., 1996. Signs of Meaning in the Universe. Indiana University Press, Bloomington.

Hougland, J.L., Piccirilli, J.A., Forconi, M., Lee, J., Herschlag, D., 2006. How the group I introns works: a case study of RNA structure and function. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 133–205.

Hughes, J.F., Coffin, J.M., 2001. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat. Genet. 29, R487–R489.

Iguchi, A., Thomson, N.R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I.R., Harris, D., Asadulghani, M., Kurokawa, K., Dean, P., Kenny, B., Quail, M.A., Thurston, S., Dougan, G., Hayashi, T., Parkhill, J., Frankel, G., 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J. Bacteriol. 191 (1), R347–R354.

Iwasaki, W., Takagi, T., 2009. Rapid pathway evolution facilitated by horizontal gene transfers across prokaryotic lineages. PLoS. Genet. 5 (3), e1000402.

Jablonka, E., Lamb, M.J., 1989. The inheritance of acquired epigenetic variations. J. Theor. Biol. 139, R69–R83.

Jablonka, E., Lamb, M.J., 2002. The changing concept of Epigenetics. In: Van Speybroeck, L., Van deVijver, G., Waele, D.D. (Eds.), From Epigenesis to Epigenetics. The Genome in Context. Ann. NYAS. 981, R82–R96.

Jablonka, E., Lamb, M.J., 2006. The evolution of information in the major transitions. J. Theor. Biol. 239, R236–R246.

Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. 33 (Suppl.), R245–R254.

Jenuwein, T., Allis, C.D., 2001. Translating the histone code. Science 293, R1074–R1080.

Ji, S., 1997. Isomorphism between cell and human languages: molecular biological, bioinformatic and linguistic implications. Biosystems 44, R17–R39.

Ji, S., 1999. The linguistics of DNA: words, sentences, grammar, phonetics and semantics. Ann. NYAS. 870, R411–R417.

Jordan, I.K., Rogozin, I.B., Glazko, G.V., Koonin, E.V., 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet. 19, R68–R72.

Jurka, J., Kapitonov, V.V., Kohany, O., Jurka, M.V., 2007. Repetitive sequences in complex genomes: structure and evolution. Annu. Rev. Genomics Hum. Genet. 8, R241–R259.

Kadyrov, F.A., Shlyapnikov, M.G., Kryukov, V.M., 1997. A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages. FEBS. Lett. 415 (1), R75–R80.

Ke, A., Doudna, J.A., 2006. Catalytic strategies of self-cleaving ribozymes: relics of an RNA World?. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 109–131.

Khodosevich, K., Lebedev, Y., Sverdlov, E., 2002. Endogenous retroviruses and human evolution. Comp. Func. Gen. 3, R494–R498.

Kim, T.H., Jeon, Y.J., Yi, J.M., Kim, D.S., Huh, J.W., Hur, C.G., Kim, H.S., 2004. The distribution and expression of HERV families in the human genome. Mol. Cells 18, R87–R93.

Kirsch, P., Jores, J., Wieler, L.H., 2004. Plasticity of bacterial genomes: pathogenicity islands and the locus of enterocyte effacement (LEE). Berliner und Muenchener tierarztliche Wochenschrift 117, R116–R129.

Klein, J., 2004. Did viruses play a part in the origin of the adaptive immune system?. Folia Biol. (Praha) 50 (3–4), R87–R92.

Koonin, E.V., Senkevich, T.G., Dolja, V.V., 2006. The ancient virus world and evolution of cells. Biol. Dir. 1, 29.

Koonin, E.V., Wolf, Y.I., 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36 (21), R6688–R6719.

Krupovic, M., Bamford, D.H., 2007. Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. BMC Genomics 8, R236.

Lescat, M., Calteau, A., Hoede, C., Barbe, V., Touchon, M., Rocha, E., Tenaillon, O., Medigue, C., Johnson, J.R., Denamur, E., 2009. A module located at a chromosomal integration hotspot is responsible for the multiple-resistance of a reference strain from *Escherichia coli* clonal group A. Antimicrob. Agents Chemother. 53 (6), R2283–R2288.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., Hess, W.R., Church, G.M., Chisholm, S.W., 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449, R83–R86.

Liu, J., Glazko, G., Mushegian, A., 2006. Protein repertoire of double-stranded DNA bacteriophages. Virus Res. 117, R68–R80.

Liu, Q., Belle, A., Shub, D.A., Belfort, M., Edgell, D.R., 2003. SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. J. Mol. Biol. 334, R13–R23.

Lopez, P., Bapteste, E., 2009. Molecular phylogeny: reconstructing the forest. C. R. Biol. 332 (2–3), R171–R182.

Maita, N., Anzai, T., Aoyagi, H., Mizuno, H., Fujiwara, H., 2004. Crystal structure of the endonuclease domain encoded by the telomere-specific long interspersed nuclear element, TRAS1. J. Biol. Chem. 279, R41067–R41076.

Margulis, L., 1996. Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. Proc. Natl. Acad. Sci. USA 93, R1071–R1076.

Margulis, L., 1999. Die andere Evolution. Spektrum Akademischer Verlag, Heidelberg.

Margulis, L., 2004. Serial endosymbiotic theory (SET) and composite individuality. Transition from bacterial to eukaryotic genomes. Microbiol. Tod. 31, R173–R174.

Margulis, L., Dolan, M.F., Guerrero, R., 2000. The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. Proc. Natl. Acad. Sci. USA 97, R6954–R6959.

Margulis, L., Sagan, S., 2002. Acquiring Genomes. A Theory of the Origin of Species. Basic Books, New York.

Markos, A., 2002. Readers of the Book of Life. Oxford University Press, Oxford, New York.

Matera, A.G., Terns, R.M., Terns, M.P., 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. 8, R209–R220.

Mattick, J.S., Gagen, M.J., 2001. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol. Biol. Evol. 18, R1611–R1630.

Mattick, J.S., 2001. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. 2, R986–R991.

Mattick, J.S., 2003. Challenging the dogma: the hidden layer of noncoding RNAs in complex organisms. BioEssays 25, R930–R939.

Mattick, J.S., 2005. Das verkannte Genom-Programm. Spektrum der Wissenschaft 3, R62–R69.

Mayer, J., Meese, E., 2005. Human endogenous retroviruses in the primate lineage and their influence on host genomes. Cytogenet. Genome Res. 110 R448–R456

Mehta, P., Casjens, S., Krishnaswamy, S., 2004. Analysis of the lambdoid prophage element e14 in the *E. coli* K-12 genome. BMC Microbiol. 4, R4.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., Keith Jr., J.C., McCoy, J.M., 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403, R785–R789.

Monier, A., Claverie, J.M., Ogata, H., 2007. Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. BMC Genomics 8, R456.

Monier, A., Claverie, J.M., Ogata, H., 2008. Taxonomic distribution of large DNA viruses in the sea. Genome Biol. 9 (7), R106.

Moore, P.B., Steitz, T.A., 2006. The roles of RNA in the synthesis of protein. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 257–285.

Moreira, D., Lopez-Garcia, P., 2009. Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol. 7, R306–R311.

Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J., Setlow, P., 2000. Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. Microbiol. Mol. Biol. Rev. 64, R548–R572.

Nolan, J.M., Petrov, V., Bertrand, C., Krisch, H.M., Karam, J.D., 2006. Genetic diversity among five T4-like bacteriophages. Virol. J. 3, R30.

Noller, H.F., 2006. Evolution of ribosomes and translation. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 287–307.

Odintsova, M.S., Yurina, N.P., 2000. RNA editing in plant chloroplasts and mitochondria. Fisiol. Rast. 37, R307–R320.

Odintsova, M.S., Yurina, N.P., 2005. Genomics and evolution of cellular organelles. Russ. J. Genet. 41, R957–R967.

Ogata, H., Claverie, J.M., 2007. Unique genes in giant viruses: regular substitution pattern and anomalously short size. Genome Res. 17 (9), R1353–R1361.

Ogura, Y., Ooka, T., Asadulghani, Terajima, J., Nougayrede, J.P., Kurokawa, K., Tashiro, K., Tobe, T., Nakayama, K., Kuhara, S., Oswald, E., Watanabe, H., Hayashi, T., 2007. Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. Genome Biol. 8 (7), R138.

Oshima, K., Toh, H., Ogura, Y., Sasamoto, H., Morita, H., Park, S.H., Ooka, T., Iyoda, S., Taylor, T.D., Hayashi, T., Itoh, K., Hattori, M., 2008. Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. DNA Res. 15 (6), R375–R386.

Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K., Otillar, R., Merchant, S.S., Podell, S., Gaasterland, T., Napoli, C., Gendler, K., Manuell, A., Tai, V., Vallon, O., Piganeau, G., Jancek, S., Heijde, M., Jabbari, K., Bowler, C., Lohr, M., Robbens, S., Werner, G., Dubchak, I., Pazour, G.J., Ren, Q., Paulsen, I., Delwiche, C., Schmutz, J., Rokhsar, D., Van de Peer, Y., Moreau, H., Grigoriev, I.V., 2007. The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. USA 104 (18), R7705–R7710.

Polavarapu, N., Bowen, N.J., McDonald, J.F., 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. Genome Biol. 7, R51.

Popov, O., Degal, D.M., Trifonov, E.N., 1996. Linguistic complexity of protein sequences as compared to texts of human languages. Biosystems 38, R65–R74.

Prangishvili, D., Vestergaard, G., Haring, M., Aramayo, R., Basta, T., Rachel, R., Garrett, R.A., 2006. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. J. Mol. Biol. 359, R1203–R1216.

Pride, D.T., Schoenfeld, T., 2008. Genome signature analysis of thermal virus metagenomes reveals Archaea and thermophilic signatures. BMC Genomics 9, R420.

Pyle, A.M., Lambowitz, A.M., 2006. Group II introns: ribozymes that splice RNA and invade DNA. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 469–506.

Rakin, A., Noelting, C., Schropp, P., Heesemann, J., 2001. Integrative module of the high-pathogenicity island of Yersinia. Mol. Microbiol. 39 (2) R407–R415.

Rohwer, F., 2003. Global phage diversity. Cell 113, R141.

Rohwer, F., Segall, A., Steward, G., Seguritan, V., Breitbart, M., Wolven, F., Azam, F., 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. Limnol. Oceanograph. 45 R408–R418

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, J.E., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.H., Falcon, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M., Venter, J.C., 2007. The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol. 5 (3), e77.

Ryan, F.P., 2004. Human endogenous retroviruses in health and disease: a symbiotic perspective. J. R. Soc. Med. 97, R560–R565.

Ryan, F.P., 2006. Genomic creativity and natural selection. A modern synthesis. Biol. J. Linn. Soc. 88, R655–R672.

Ryan, F.P., 2009. Virolution. Harper Collins Publishers, London.

Sandegren, L., Nord, D., Sjoberg, B.M., 2005. SegH and Hef: two novel homing endonucleases whose genes replace the mobC and mobE genes in several T4-related phages. Nucleic Acids Res. 33 (19), R6203–R6213.

Searls, D.B., 2002. The language of genes. Nature 420, R211–R217.

Sebeok, T., Umiker-Sebeok, J., 1992. Biosemiotics: The Semiotic Web 1991. Mouton de Gruyter, Berlin.

Shabalina, S.A., Spiridonov, N.A., 2004. The mammalian transcriptome and the function of non-coding DNA sequences. Genome Biol. 5, 105e.

Sharon, I., Tzahor, S., Williamson, S., Shmoish, M., Man-Aharonovich, D., Rusch, D.B., Yooseph, S., Zeidner, G., Golden, S.S., Mackey, S.R., Adir, N., Weingart, U., Horn, D., Venter, J.C., Mandel-Gutfreund, Y., Beja, O., 2007. Viral photosynthetic reaction center genes and transcripts in the marine environment. ISME J. 1 (6), R492–R501.

Shapiro, J.A., 2002. Genome organization and reorganization in evolution: formatting for computation and function. In: Speybroeck, L., VandeVijver, G., Waele, D.D. (Eds.), From Epigenesis to Epigenetics. The Genome in Context. Ann. NYAS. 981, R111–R134.

Shapiro, J.A., 2004. A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. Gene 345, R91–R100.

Shapiro, J.A., Sternberg, R., 2005. Why repetitive DNA is essential to genome function. Biol. Rev. 80, R1–R24.

Simpson, L., 2006. Uridine insertion/deletion RNA editing as a paradigm for site-specific modifications of RNA molecules. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 401–417.

Sinkvics, J.G., 2001. The place of viruses in the "Tree of Life". Acta Microbiol. Immunol. Hung. 48 (1), R115–R127.

Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8, R272–R285.

Sonenshein, A.L., Hoch, J.A., Losick, R., 2002. *Bacillus Subtilis* and its Closest Relatives: from Genes to Cells. ASM Press, Washington.

Sternberg, R., 2002. On the roles of repetitive DNA elements in the context of a unified genomic-epigenetic system. In: Speybroeck, L., VandeVijver, G., Waele, D.D. (Eds.), From Epigenesis to Epigenetics. The Genome in Context. Ann. NYAS. 981, R154–R188.

Sternberg, R., Shapiro, J.A., 2005. How repeated retroelements format genome function. Cytogen. Genome Res. 110, R108–R116.

Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., Chisholm, S.W., 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol. 4, e234.

Sverdlov, E.D., 2000. Retroviruses and primate evolution. Bioessays 22, R161–R171.

Taft, R.J., Mattick, J.S., 2004. Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. Genome Biol. 5, R1.

Tang, R.S., Barton, D.J., Flanegan, J.B., Kirkegaard, K., 1997. Poliovirus RNA recombination in cell-free extracts. RNA 3, R624–R633.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., DeBoy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Ros, I.M., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J.B., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., Fraser, C.M., 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. 102 R13950–R13955

Toothman, P., Herskowitz, I., 1980a. Rex-dependent exclusion of lambdoid phages. I. Prophage requirements for exclusion. Virology 102, R133–R146.

Toothman, P., Herskowitz, I., 1980b. Rex-dependent exclusion of lambdoid phages. II. Determinants of sensitivity to exclusion. Virology 102, R147–R160.

True, H., Berlin, I., Lindquist, S.L., 2004. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. Nature 431, R184–R187.

Turner, B.M., 2000. Histone acetylation and an epigenetic code. BioEssays 22, R836–R845.

Turner, B.M., 2002. Cellular memory and the histone code. Cell 111, R285–R291.

Tycowski, K.T., Kolev, N.G., Conrad, N.K., Fok, V., Steitz, J.A., 2006. The ever-growing world of small nuclear ribonucleoproteins. In: Gestelandm, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 327–368.

Van De Vijver, G., Van Speybroeck, L., De Waele, D., 2002. Epigentics: a challenge for genetics, evolution and development. In: Speybroeck, L., Van De Vijver, G., DeWaele, D. (Eds.), From Epigenesis to Epigenetics. The Genome in Context. Ann. NYAS. 981, R1–R6.

Van Speybroeck, L., Van de Vijver, G., DeWaele, D., 2002. Preface. In: Speybroeck, L., Van DeVijver, G., De Waele, D. (Eds.), From Epigenesis to Epigenetics. The Genome in Context. Ann. NYAS. 981, vii.
Vetsigian, K., Woese, C., Goldenfeld, N., 2006. Collective evolution and the genetic code. Ann. NYAS. 103, R10696–R10701.
Villarreal, L.P., 1999. DNA virus contribution to host evolution. In: Domingo, E., Webster, R.G., Holland, J.J. (Eds.), Origin and Evolution of Viruses. Academic Press, San Diego, pp. 391–420.
Villarreal, L.P., 2004. Can viruses make us humans?. Proc. Am. Phil. Soc. 148, R296–R323.
Villarreal, L.P., 2005. Viruses and the Evolution of Life. American Society for Microbiology Press, Washington.
Villarreal, L.P., 2007. Virus-host symbiosis mediated by persistence. Symbiosis 44, R1–R9.
Villarreal, L.P., 2009. Origin of Group Identity: Viruses, Addiction and Cooperation. Springer, New York.
Volff, J.N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28, R913–R922.
Wang, Y., Fischle, W., Cheung, W., Jacobs, S., Khorasanizadeh, S., Allis, C.D., 2004. Beyond the double Helix: writing and reading the Histone code. In: Bock, G., Goode, J. (Eds.), Reversible Protein Acetylation. Novartis Foundation 2004, vol. 259, pp. 3–17.
Watanabe, M., Kikawada, T., Minagawa, N., Yukuhiro, F., Okuda, T., 2002. Mechanism allowing an insect to survive complete dehydration and extreme temperatures. J. Exp. Biol. 205, R2799–R2802.
Weigele, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., King, J., Hatfull, G.F., Lawrence, J.G., Hendrix, R.W., 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine Prochlorococcus and Synechococcus. Environ. Microbiol. 9, R1675–R1695.
Weiner, A.M., 2006. SINEs and LINEs: troublemakers, saboteurs, benefactors, ancestors. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), The RNA World third ed Cold Spring Harbor Laboratory Press, New York, pp. 507–534.
Weiss, R.A., 2006. The discovery of endogenous retroviruses. Retrovirology 3, R67.
Williamson, S.J., Cary, S.C., Williamson, K.E., Helton, R.R., Bench, S.R., Winget, D., Wommack, K.E., 2008a. Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. ISME J. 2 (11), R1112–R1121.
Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., Andrews-Pfannkoch, C., Fadrosh, D., Miller, C.S., Sutton, G., Frazier, M., Venter, J.C., 2008b. The Sorcerer II global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. PLoS ONE 3, e1456.
Wittgenstein, L., 1972. Philosophical Investigations. Basil & Blackwell, Oxford.
Witzany, G., 1995. From the, logic of the molecular syntax' to molecular pragmatism, explanatory deficits in Manfred Eigen's concept of language and communication. Evol. Cog. 1, R148–R168.
Witzany, G., 2000. Life: The Communicative Structure. A New Philosophy of Biology. Libri Books on Demand, Norderstedt.
Witzany, G., 2005. Natural history of life: history of communication logics and dynamics. S.E.E.D. J. 5, R27–R55.
Witzany, G., 2006a. Serial endosymbiotic theory (SET): the biosemiotic update. Acta Biotheor. 54 (2), R103–R117.
Witzany, G., 2006b. Natural genome-editing competences of viruses. Acta Biotheor. 54 (4), R235–R253.
Witzany, G., 2006c. Plant communication from biosemiotic perspective. Plant Signal Behav. 1 (4), R169–R178.
Witzany, G., 2007. The Logos of the Bios 2. Bio-Communication. Umweb, Helsinki.
Witzany, G., 2008. Bio-communication of bacteria and their evolutionary roots in natural genome editing competences of viruses. Open Evol. J. 2, R44–R54.
Witzany, G., 2009a. Non-coding RNAs—persistent viral agents as modular tools for cellular needs. In: Witzany, G. (Ed.), Natural Genetic Engineering and Natural Genome Editing. Ann. NYAS. 1178, R244–R267.
Witzany, G., 2009b. Biocommunication and Natural Genome Editing. Springer, Dordrecht.
Xu, Q., Modrek, B., Lee, C., 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acid Res. 30, R3754–R3766.
Yang, Z., Kim, J., Zhang, C., Zhang, M., Nietfeldt, J., Southward, C.M., Surette, M.G., Kachman, S.D., Benson, A.K., 2009. Genomic instability in regions adjacent to highly conserved pch prophage in *E. coli* O157:H7 generates diversity in expression patterns of the LEE pathogenicity island. J. Bacteriol. 191 (11), R3553–R3568.
Yin, Y., Fischer, D., 2008. Identification and investigation of ORFans in the viral world. BMC Genomics 9, R24.
Zanotto, P.M., Gibbs, M.J., Gould, E.A., Holmes, E.C., 1996. A reevaluation of the higher taxonomy of viruses based on RNA polymerases. J. Virol. 70 R6083–R6096
Zeidner, G., Bielawski, J.P., Shmoish, M., Scanlan, D.J., Sabehi, G., Beja, O., 2005. Potential photosynthesis gene recombination between Prochlorococcus and Synechococcus via viral intermediates. Environ. Microbiol. 7, R1505–R1513.
Zhang, H.Y., 2006. The evolution of genomes and language. EMBO Rep. 7, R248–R249.