NATURAL GENOME-EDITING COMPETENCES OF VIRUSES

Günther Witzany

telos – Philosophische Praxis, Vogelsangstr. 18c, 5111, Buermoos, Salzburg, Austria. Email: witzany@sbg.at

Received 4 October 2006; Accepted 22 December 2006

ABSTRACT

It is becoming increasingly evident that the driving forces of evolutionary novelty are not randomly derived chance mutations of the genetic text, but a precise genome editing by omnipresent viral agents. These competences integrate the whole toolbox of natural genetic engineering, replication, transcription, translation, genomic imprinting, genomic creativity, enzymatic inventions and all types of genetic repair patterns. Even the non-coding, repetitive DNA sequences which were interpreted as being ancient remnants of former evolutionary stages are now recognized as being of viral descent and crucial for higher-order regulatory and constitutional functions of protein structural vocabulary. In this article I argue that non-randomly derived natural genome editing can be envisioned as (a) combinatorial (syntactic), (b) context-specific (pragmatic) and (c) content-sensitive (semantic) competences of viral agents. These three-leveled biosemiotic competences could explain the emergence of complex new phenotypes in single evolutionary events. After short descriptions of the non-coding regulatory networks, major viral life strategies and pre-cellular viral life three of the major steps in evolution serve as examples: There is growing evidence that natural genome-editing competences of viruses are essential (1) for the evolution of the eukaryotic nucleus, (2) the adaptive immune system and (3) the placental mammals.

Key Words: non-coding DNA, repetitive elements, genome editing, persistent viral life-style

1. INTRODUCTION

Significant evolutionary modifications can occur by changing the repetitive elements that format the genome architecture rather than, as has been widely assumed so far, by altering protein-coding sequences. This implies that changes in the regulatory framework may lead to new complex phenotypes without even requiring a modification of the coding sequences.

For decades, non-coding regions of the genome have been ignored or declared as ''junk''-DNA. Over the past decade, however, scientists have realized that these regions incorporate decisive higher-order regulatory functions. Within the human genome sequences are only 3% coding for proteins and 97% non-coding ones. But these 3% of the protein-coding sequences match with those of the mouse by up to 99%. Highly developed genomes among mammals contain a nearly identical protein-coding vocabulary.

Acta Biotheoretica (2006) 54:235–253
DOI: 10.1007/s10441-007-9000-7

While bacterial genomes almost lack such repetitive elements, there is a ten-fold increase of such segments in the genome of the eukaryotic yeast cells; in the case of the fruit fly *Drosophila*, the repetitive elements already amount to 10%, while in humans they reach 97% (Villarreal, 2004: 304). These formatting, repetitive elements are a prerequisite for the decisive DNA-editing processes such as expression, replication, repair and recombination.

2. NON-CODING REGULATORY NETWORKS AND GENOME EDITING

Clearly, mobile sequences such as transposons and retroposons (Volff 2006) and non-coding repetitive elements such as LTRs, SINEs and LINEs (long terminal repeats, short interspersed elements, long interspersed elements) enable far-reaching DNA rearrangement and reorganization (Shapiro, 2002; Sternberg, 2002; Deepak *et al*., 2003; 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; Pollard *et al*., 2006). Depending on the organism's state of development, the varying chromatin markers are, thus capable – through different methylation patterns, histone modifications and alternative splicing – of creating a set of "multiple protein meanings" (Ast, 2005) from one and the same genetic data-set (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 even characterizes the rise of epigenetics, i.e., the view that phenotypic variations, which are heritable, need not be connected with 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 DNA, formerly termed "junk"-DNA?

Although we have 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). It subsequently became clear 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 micro-RNAs (Mattick, 2001, 2003, 2005; Mattick and Gagen, 2001). In 1992, Watson suspected these higher-order regulatory functions to be hidden in chromosomal structures (Watson *et al*., 1992). Interestingly, researchers analyzing the linguistic features of non-coding DNA sequences draw the right conclusions (Mantegna *et al*., 1994; Vendrami, 2004). They found coherence between coding DNA with similar structural features as artificial scientific languages and non-coding DNA with similar structural features as everyday language.

The genome, with both its coding and non-coding sections, serves as a genetic, epigenetic and computational cell storage medium that involves different kinds of storage techniques. These include short-time storage (dynamic rearrangements, remodifications, rapidly induced alterations for expression and repair), intermediate storage (epigenetic use) including multiple protein-meanings within the same chromosomal structures by different expression modes through different

methylation patterns, and long-time storage, i.e., stable inheritable datasets for many generations (Shapiro and Sternberg, 2005). This undoubtedly means that the DNA storage medium in its possible functions depends on different means, i.e., interpretations, of the organism inputs on this medium rather than on the inherent information. This enables changing inheritable priorities, whether computational, intermediate, or long-time storage data – or combinatorial patterns among them (Carroll, 2005; Jablonka and Lamb, 1989, 2006; Ryan, 2006).

The repetitive sequences are highly species-specific and more suitable for the determination of species than the coding sequences (Villarreal, 2005). Each taxon organizes 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 guarantees that these different actions are performed and maintained successfully: DNA sequences must be read at the right site and at the right time. The precise spatio-temporal co-ordinations are essential in order to sustain vital processes; nonetheless, such co-ordinations 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 codes are the keys in DNA editing and the basis of evolutionary processes. The contemporary term would be ''natural genetic engineering'' (Shapiro, 2004). Wittgenstein noted that ''to understand a sentence means to understand a language. To understand a language means to be master of a technique'' (Wittgenstein, 1972). Nonetheless, there is a clear difference between natural and artificial genetic engineering. The former provides ontologically genuine products that are evident in all living beings and life processes, whereas the latter attempts to achieve modifications and improvements by copying the natural genome-editing competences.

3. MAJOR VIRAL LIFE STRATEGIES

New research has shown that these non-coding repetitive sequences originated primarily from retroviral RNA (Villarreal, 2004, 2005). After persistent non-lytic viral infection, important protein coding regions become incurred within the replicating region of the host organism, whereas repetitive sequences are integrated into the non-coding sections (Villarreal *et al.*, 2000; Villarreal and DeFilippis, 2000). In this way, thousands of endogenous retroviral sequences have been integrated into the human genome, with 22 independent retroviral families currently having been identified (Villarreal, 2004, 2005; Ryan, 2006). The 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 recognized by their three genes *gag, pol* and *env* (Gao *et al.*, 2003; Ryan, 2004). The *gag* gene encodes structural proteins, *pol* encodes enzymes such as reverse transcriptase and integrase functions and *env* encodes envelope proteins.

At this stage we know roughly 3600 different kinds of viruses and approx. 30,000 viral subtypes (Villarreal, 2005: 6). The most numerous variety is found among single-stranded RNA viruses, followed by double-stranded DNA viruses, double-stranded RNA viruses and single-stranded DNA viruses (Villarreal, 2005: 5). This huge variety of viruses and viral components is even more astonishing when considering the fact that 1 ml of sea water contains about 1 million bacteria and that the number of viral components in the same amount of water is approx. 10 times higher (Villarreal, 2005: 55). However, their ecological niches are not just limited to purely aquatic or terrestrial environments, but also include species-specific tissues that differ significantly from each other (Villarreal, 2005: 56). As phylogenetic analysis shows, nearly all organisms of all kingdoms have been infected or highly colonized by viruses since the beginning of life.

Viruses can parasitize almost any replication system – even prebiotic ones – and probably emerged well before the appearance of cellular life forms. Viruses store information that not only pertains to (a) replication proteins but also to (b) morphology and (c) phenotypic diversity. 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: 11).

Since viruses with RNA genomes are the only living beings 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, 2006; 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. This proves the existence of negatively stranded viral RNA-replicases even before cellular life came into being (Villarreal, 2005: 11). DNA viruses, too, do not give any reference to a cellular origin. Phylogenetic analyses point to an older time scale, as DNA-repairing proteins of DNA viruses do not have any counterparts in cellular biota.

For a long time, viruses were not considered a part of the animated world because they were believed to be exclusively parasitic in nature. They were interpreted as causing an acute infection of the host organisms, using the host's cellular machinery to reproduce, and performing their lytic nature only in order to infect other cells. Although this narrative remains valid, it merely represents a special case of viruses that were unable to reach persistent status by a sessile lifestyle (Villarreal, 2004: 305). Most viruses, however, are stable, persistent living beings that do not colonize a host organism for simple selfish purposes. Rather, they display varying survival strategies which can differ according to the host they inhabit (Villarreal, 2005: 7):

- In bacteria we mainly find double-stranded DNA viruses. Similar viral types are also found in algae, but they are absent in fungi and plants.
- Fungal hosts house mainly double-stranded RNA viruses, whereas
- Plants contain predominantly single-stranded RNA viruses.
- Mammals, on the other hand, are colonized mainly by endogenous retroviruses (ERVs).

Viral persistence in host organisms is crucial because they reliably protect the host against similar parasites. Interestingly, competing viruses are in most cases of the same or related species, whereas unrelated viruses do not compete but interact symbiotically (Roossinck, 2005).

As for endogenous retrovirus, they confer the host a distinct and unique genetic identity, non-existent before the colonization. This persistence can either be endonucleic or cytoplasmatic. Persistent viruses are capable of continuous or episodic reproduction in the hosts (Villarreal, 2005: 5). The main direction is not to infect cells for reproduction with lytic consequences and often epidemic viral diseases, but a sessile life-style without harming the host. This includes different meanings of the term fitness. Thus, lytic life-styles are fit if the reproduction rate reaches epidemic dimensions, whereas the persistent life-style is fit if it does not harm the host and wards off competitive viral infections.

For far too long, viruses have only been considered as poisonous and highly infectious parasites. The fact that viruses are silent companions of virtually all organisms and that they play a decisive role in the evolution of the host has been largely ignored.

Gene functions of cellular biota, which undoubtedly are associated with the persistent life strategy of viruses, include (Villarreal, 2005: 21):

- Immunity (restriction and modification modules, toxic and antitoxic modules);
- Silencing functions/micro-RNAs, (methylation, suppression);
- Recognition functions (replicate expression, receptors, expression factors);
- Immune regulation (signal mediating, heredity, adaptation).

However, the host must be able to inhibit the replication of a persistent virus; i.e., it must prevent it from turning aggressive and exercising its lytic nature. Especially if the persistence does not reach a stable status, the move-countermove interaction between nucleic acid invader and a host genome may be an on-going process which leads to a special kind of immunity, a defensive RNA-silencing against competing viruses (Fire, 2006; Buchon and Vaury, 2006). This RNAi interactional pattern is documented to be a very old type of innate immune system.

4. EXAMPLES OF DIVERSE STRATEGIC PATTERNS OF VIRUSES

4.1. Virus Escape-Strategy

Mitochondria and chloroplast precursors take refuge in prokaryotic hosts where they are protected from cyanophagous colonization. Although the host is equipped with restriction and modification competence that can ward off viral colonization, endosymbiotic mitochondria and chloroplasts lack such competences. Such competences might have existed in the past, but they have been relocated to the hosting organism, thereby establishing a type of immune system against cyanophagous infestation, which is omnipresent in oceans (Villarreal, 2005: 127).

4.2. Wall-Off-Strategy

This is closely associated with calcification (encapsulation of the parasites) and provides some mechanical as well as chemical protection. It was initially based on

an RNAi defence system found even among simple marine species such as early protostomes and among their relatives such as crustaceans and insects. As in bivalves, this phenomenon is also found among insects: a chitinous or a mineralized barrier is formed around the parasitic intruder. This is a very archaic viral strategy that can still be observed today as a reaction to parasitic infections in mammals and in the form of a chronic inflammatory response in vertebrates (Villarreal, 2005: 204).

4.3. Addiction Module

Addiction modules include antagonistic components acquired in order to survive inside a host without damaging it (Villarreal, 2005: 302). Acquisition of addiction modules can be reconstructed as a massive viral colonization of the host. Addiction modules consist of a set of genes or functions that are harmless to the host and that include both toxin and antitoxin. In most cases the evolutionary aspect for the host organism is an acquired and hereditary immunity function, e.g., the restriction/modification phenotype. These modules probably paved the way for the evolution of new species, so persistent colonization led to permanent integration without being threatened by any other virus (Villarreal, 2005: 145); e.g., *Dictyostelium* (Villarreal, 2005: 178, 181), sharks, bryophytes (Villarreal, 2005: 238) and ferns (Villarreal, 2005: 228); this process takes place even today. However, the present diversity of these innovative phenotypes is in turn immensely colonized by viruses and, interestingly, by groups of novel viruses that were non-existent before.

4.4. Multiplicity Reactivation

Some of the oldest viruses, the phycodnaviruses and cyanophages, are able to repair UV-light induced damages of their genomes and can reproduce even if the host organisms are exposed to lethal doses of radiation. This ability to replicate is only possible due to active participation of virus-specific repair enzymes, which can re-establish the cell's competence to process macromolecules.

This capability of expressing subsets of genes is retained even if each of the viruses' genome experiences lethal radiation damage. Once functional complementation of expressed genes and other damaged viruses is established, the complementary sets are capable of initiating viral replication: the necessary combinations of defective viral genomes are utilized in order to assemble intact viruses. Therefore, viruses are the only living beings capable of meaningfully recombining text fragments of a damaged genome into a fully functional viral genome capable of self-replication (Villarreal, 2005:128).

4.5. Sexual Isolation

Once a species is massively infested by an acute and lytic pandemic virus, some of the surviving hosts manage to permanently include this viral genome into their own genome, thereby obtaining permanent immunity against such infections. As soon as they engage in intraspecific reproduction, they lethally infect those siblings that lack this additional immunity. This sexually isolates the immunized from the non-infected members. As an immediate result, the genetically altered

siblings experience an interruption of their line of ancestry by acquiring a complex immunological phenotype. This enables the immunized species to establish a novel species incapable of mating with the previously related members without killing them (Villarreal, 2004: 315, 2005: 361).

5. PRE-CELLULAR VIRAL LIFE

RNA genomes probably existed before the appearance of DNA genomes (Villarreal, 2005: 30). Hence, DNA can be considered to be a modification of RNA, yielding ribonocleotide reductase, followed by only two thymidylate synthases.

Initially, there were only two components, RNA molecules and RNA proteins, which constituted a RNA cell that contained only few, individual DNA components. In turn, they were colonized by parasitic RNA that facilitated the scission into two separate lineages – DNA viruses and RNA viruses – whereby a DNA virus was capable of infecting an RNA virus. This model would at least account for the existence of viral encoded DNA transaction proteins, to which no cellular counterpart exists. In addition, this model would explain the existence of two dissimilar DNA-replication systems.

On the one hand, the absence of repair enzymes within the RNA world gave it a more flexible and creative character. On the other hand, this inability is a drawback when a reliable storing capacity is required. Evidently, the presence of such addiction modules within the highly competitive viral RNA-DNA world must have an advantage. This is particularly true when a DNA virus infects an RNA virus, forcing the latter to establish a bi-layered cell-membrane and to encapsulate the genome by a porous nuclear envelope. Doing so still enables replication, repair and recombination (Villarreal, 2005).

Even though the DNA cell gained distinct selective advantages, the RNA parasites still had an astonishingly powerful genomic creativity (Ryan, 2006); 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 innovation of the RNA world complements that of the more conservative and stable DNA world.

Besides these processes, life also resulted from two alternative innovations: the creation of the archaeal domain and that of the eubacterial domain. Again, this would explain why archaeal lipids are so much different from eukaryotic/bacterial lipids. These lipids contradict eubacterial stereochemistry and differ in the backbone architecture of the hydrocarbon chain.

6. THE ORIGIN OF THE EUKARYOTIC NUCLEUS

The origin of the eukaryotic nucleus serves as a first example for the important role of natural genome-editing competences of viruses. Since the introduction of the Serial Endosymbiotic Theory (SET), it is generally accepted that the eukaryotic cell did not result from random mutations, but from the coordinated union of free-living prokaryotes (Margulis, 1996, 1999, 2004; Margulis *et al.*, 2000; Margulis and Sagan, 2002; Witzany, 2006b). Mitochondria and other

organelles clearly descended from these micro-organisms (Odintsova and Yurina, 2000, 2005) and the assumption was that the eukaryotic nucleus is also probably of archaeal or bacterial descent.

New evidence supports the idea that eukaryotic nuclei originated before the symbiogenetic integration with mitochondria and chloroplasts (Villarreal, 2005: 102). In fact, the nucleus has basic properties that are otherwise absent in eukaryotic cells (Bell, 2001, 2006).

Prokaryotes do have circular chromosomes with uniform, standardized 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: 306). These differences include:

- use of linear chromosomes, with repetitive termination points and several origins for replication,
- transcription and translation separated by multiple membranes,
- existence of complex nuclear pore structures that actively mediate RNA translocation,
- a tubulin-system that enables separation of duplicated chromosomes.

All these properties represent complex phenotypes that require complex co-ordination of numerous protein functions. None of these functions are present in prokaryotes even though they are considered as 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: 106). Even the three kinds of splicing group I-introns (DNA transposase, reverse transcriptase and micro-RNAs) are largely inexistent in prokaryotes, although they are present in viruses of prokaryotes. Bacteria never had introns in any of their coding genes. In addition, no single prokaryotic process is known to account for the tasks of membrane disintegration and restoration as observed in eukaryotes. Eukaryotic tubulin consisting of a highly conservative mitotic spindle is absent in all prokaryotic lines of descent, even though it is one of the most important components in eukaryotes. Prokaryotes lack a tubulin system. The eukaryotic pore-structure of the nuclear envelope likewise has no counterpart in the prokaryotic world (Villarreal, 2005: 119).

Viral proteins on the other hand bind both tubulin as well as actin, thereby triggering polymerization and mobility functions. Viral genes are directly involved in transposition tasks. All viruses mark their genome, their RNA and their proteins with virus-specific methylases, for example via enzymatic reactions known as base methylation (Villarreal, 2005: 122).

Viruses were long thought to be parasites capable of "stealing" host-specific properties. In the meantime, viruses are known to have created new genes as a result of their evolutionary line of descent. In the baculoviruses, for example, GenBank database investigations reveal that 80% of their genes are unique to this group and found nowhere else (Villarreal, 2005: 274). Gene losses have been documented in baculoviruses, but the 12 losses documented therein are opposed by a staggering acquisition of 255 new genes.

Between 1950 and 1980, scientists recognized that the T4 phage-polymerase proteins are much more similar to the eukaryotic DNA polymerase protein than to any prokaryotic polymerase. We now know that the eukaryotic DNA polymerase and the T4 DNA polymerase do have common origins. Indeed, the T4 represents a huge family of viruses that is capable of infecting both Bacteria and Archaea (Villarreal, 2004: 307). Hence, it is not surprising that T4-DNA polymerases are found in all three domains of life: Archaea, Bacteria and Eukarya.

Algae were among the first higher 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 must have included the characteristic of a protonucleus. The phycodnavirus is a case in point: this calls for examining the entire GenBank database for sequences that may be similar to the DNA polymerase of this particular virus (Villarreal, 2004: 308). Such sequences must include:

- replication polymerases of all higher eukaryotes as well as of all larger eukaryotic DNA-viruses,
- primer polymerases of eukaryotes,
- repair polymerases of both Archaea and Bacteria.

The DNA polymerase of the CSV1 virus is present at the origin of all eukaryotic replication DNA polymerases and is no doubt a precursor of all polymerases involved in replication of the eukaryotic genome. So far no other viral or prokaryotic DNA polymerase that shares these features is known (Villarreal, 2004: 310).

The membrane-bound separation of transcription and translation is a characteristics of the poxviruses, more specifically of the Vaccina and other DNA viruses. Moreover, these viruses have a very simple pore-structure that has actively been incorporated from the membrane-bound RNA into the host cytoplasm. A similar situation can be documented with the small chromatin proteins and the linear chromosomes along with their repetitive telomere tails that are so characteristic among various cytoplasmid DNA-viruses, TTV1 and phycodnaviruses. Even the highly complex function of tubulin, as an important coordinating element during chromosomal separation of duplicated strands, is present in DNA-viruses with exactly the same set of functions (Villarreal, 2004: 311).

It became increasingly obvious that all properties of the eukaryotic nucleus are derived from a large, stable and persistent DNA virus with linear chromosomes. The current interpretation is that the precursor of the eukaryotic nucleus was indeed a huge membrane-covered DNA virus that persistently colonized a prokaryotic host (Villarreal, 2004: 311). 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 most likely non-lytic because it coordinated both its own replication and its transcription genes, it had a double-layered membrane 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 corresponds to the sexual reproductive cycle of the host organism.

Interestingly, all these properties can be found in prokaryotic viruses such as cyanophage, archaeal phage, mycobacterial phage and eubacterial phage (Villarreal, 2005: 112).

7. THE ORIGIN OF THE ADAPTIVE IMMUNE SYSTEM

The second example for the important role of viral genome-editing competences is the origin of the adaptive immune system. In the case of persistent viruses, the host must actively be able to take control over viral replication. This requires the host to have some kind of defence/immune system (Villarreal, 2005: 209). On the other hand, the persistent virus must also be able to wall off similar viruses trying to take advantage of the host. Lower organisms achieve this with so-called addiction modules – sets of genes that impair the host when absent but are of advantage when present in the host (Villarreal, 2004: 302). Usually, the negative aspects of such gene sets involve toxic substances, while the positive involve the 'antidote'. At any rate, acute and lytic action might be encountered among closely related species, provided that they are persistent to a specific host. Only in such cases is the hosting organism protected. Such virus-based immune modules are complex genetic systems that include interrelated functional sets that can positively or negatively affect the host; certain environmental conditions or colonization by other viruses can alter the current persistent state by becoming acute, lytic and thereby damaging or even killing those that are not yet infected.

Prokaryotes of the archaeal and bacterial domain are the most adaptive individuals on earth. This raises an obvious question: how do these organisms protect themselves against viral agents? The answer involves utilizing restriction/modification enzymes: Restriction enzymes degrade unmodified DNA with modification enzymes acting only during DNA replication. This simple and innate immune system likewise originated from viruses. Prokaryotes colonized by such viruses possess specific immune response capabilities. The host organism acquires new complex immunologic properties once the viral colonization stabilizes (Villarreal, 2004: 303).

This, however, does not involve horizontal gene transfer (HTG), but rather large-scale viral inventions of new genes. Some find their way through host lineages and become permanently established within the host. Others have to assess the molecular texture of the bacterial genome as well as its evolutionary potential (Villarreal, 2004: 305).

A major change in evolution occurred during the transition from urochordates (sea squirts/tunicates) to bony fish. Many common tissue types are still evident, although the bony structure required a huge increase in the genome. This was predominantly achieved by retroposons, which originated from endogenous retroviruses. Urochordates, on the other hand, possess only an innate immune system.

7.1. The Acquisition of a Complex New Phenotype

Bony fish are the first in this line of descent that possess both an innate and an adaptive immune system (Villarreal, 2005: 203). The conclusion is that the entire complexity of an adaptive immune system was acquired at once, probably at the

very beginning of the vertebrate line of descent. At the same time, evolution simultaneously yielded jawbones, the vertebral spine, and the skull. Concomitantly, new viral families emerged that had never inhabited any of the previous life forms: four different kinds of RNA-virus families, non-defective, autonomous and abundant retroviruses (Villarreal, 2005: 204). The adaptive immune system therefore represents an interlinked network of proteins that tag inflammation and other acute processes (cytokine and chemokine), and their receptors and signal transmission systems, that stimulate the humeral and cellular antigen-specific immune response pattern. This must have been acquired in a single event because it is monophyletic (Villarreal, 2005: 205).

While urochordates do not possess any – except for one – of these features, bony fish possess them all. Sharks are among the earliest vertebrates that featured an adaptive immune system (Villarreal, 2005: 206). Interestingly, hardly any viruses are found in this group (except for a herpes type strain). They have a very rudimentary, adaptive immune system. Sharks are among the first organisms that include both oviparous as well as viviparous species (Villarreal, 2005: 207).

The reverse side of an adaptive immune system has a drawback: the non-related haploid chromosomal set is no longer rejected. From an immuno-related perspective it is not recognized as being "alien" – very much in analogy to the more recent development of placental mammals. Nonetheless, a large quantity of cytokines are transcribed within the placenta, i.e., the immune system is highly active, as is the complementary function of the addiction module. Because the mechanisms of self – and non-self-identification are extremely vigilant, they protect the forming progeny from external parasitic attacks.

The acquisition of an adaptive immune system is both a punctual and a variable evolutionary event of the animal kingdom. It enabled the expression of highly complex phenotypes. These phenotypes consist of a self-forming and a dynamically adapting genetic system that recognizes "non-self" elements and can thereby promptly attack and simultaneously prevent fatal auto-aggression. In the context of a manifold self-identification system, this acquisitional gene-set strategy was developed in order to detect novel non-self agents. Once the system recognizes the presence of "non-self" agents, it responds by developing a new molecular process that involves the generation of genetic diversity and clonal growth of specific cells capable of detecting such non-self agents. This kind of genetic recombination on a genetic level is found only from this point onwards, not in any predecessors (Villarreal, 2005: 209).

The overall result of such processes is cells that produce new classes of molecules that can (a) bind and suppress non-self agents or (b) enable amoeboid cytotoxic cells to find and neutralize agent-containing cells. Most of these characteristic properties of an adaptive immune system were lacking prior to vertebrate evolution.

Today, we know that tunicates, the precursors of urochordates, possess a polymorphous MHC-like system which is associated with cell-induced, non-adaptive, amoeboid hemolymphic lethal actions (Villarreal, 2005: 209). This tunicate system, however, lacks a molecular similarity to the vertebrate MHC system. In order to develop an adaptive immune system so common to vertebrates, the tunicate-like system had to acquire an adaptive component that allowed the non-self elements to be detected as alien. Accordingly, an adaptive

immune system destroys non-self elements while protecting those that are part of the system.

These related properties are characteristic for addiction modules. Indeed, the adaptive immune system is an elaborated addiction module. Hosts, which incorporated it acquired a system with destructive capabilities. This ability to kill is comparable to a potent toxin that lyses any cell exposed to it. Through this self-recognition process the host must be equally able to prevent autolysis, comparable to an antitoxin (Villarreal, 2005: 210).

Similar to other addiction modules, the (a) lethal, toxic ability of the adaptive immune system is stable and long-lasting, while (b) the ability to express antitoxic characteristics (self-recognition) is only a temporary feature acquired during the development stage of the immune cell. Thus, the adaptive immune system reveals two aspects of an addiction module (toxic and antitoxic) and a varying stability of the toxin with regard to the protecting antitoxin (Villarreal, 2005: 210).

Many of these individual elements of an adaptive immune system evolved individually early on in evolution. They definitely existed previously in different viral types rather than in cellular organisms.

What is the most important basic gene-function of an adaptive immune system that favors diversity (necessary for recognizing non-self elements), and where did these genes evolve from? The proteins resulting from recombinase activating genes (RAGs) are responsible for DNA adaptations and new combinations (Villarreal, 2005: 211). They are the driving agents establishing genetic diversity, thereby producing a manifold display of varying surface receptors. From this perspective, RAGs are the crucial starter proteins for the selection of adaptive immunity. Phylogenetic analyses of these RAGs suggest that precursors existed neither in eukaryotic nor prokaryotic genomes. Instead, RAGs are probably closely related to integrase genes of retroviral origin.

7.2. Where Can the Ancestral Origin of an Adaptive Immune System be Found?

So far, no specific viral family that can be considered as a precursor of an adaptive immune system has been isolated. No single virus family possesses all the necessary and required properties. The basic ability to acquire RAG functions and their controlling capability is also accompanied by the property of endogenous retroviral colonization of the host genome (Villarreal, 2005: 212). The above properties point to the acquisition of adaptive immunity rather than a complex punctual genetic event. This acquisition event however, was not the product of an individual genetic parasite but rather the joining of the stable colonization through entire sets of supplementary and defective viral agents (Villarreal, 2005: 212). These agents must have covered the immediate precursors of cartilaginous fish with their extremely complex addiction module. This enabled both a stable colonization and successfully excluded competing parasites. The host-predecessor was most certainly equipped with a recognition system quite similar to MHC – a system of cytotoxic cells that was colonized and regulated by a complementary set of newly colonizing parasitic agents.

This eventually resulted in the creation and evolution of the adaptive immune system (Villarreal, 2005: 213).

8. THE EVOLUTIONARY INNOVATION OF PLACENTAL MAMMALS THROUGH ENDOGENOUS RETROVIRUSES

The third example for the important role of genome-editing competences of viruses is the origin of placental mammals. The close ties between the human genome and the colonization with repetitive sequences of retroviral origin (LTRs, SINEs, LINEs) become obvious when considering the Y-chromosome (Villarreal, 2004: 314). It contains only 20–30 coding sequences (Villarreal, 2005: 338), approximately 225 genes, whereby the majority are non-coding sequences with higher-order regulatory functions. The remainder of endogenous retroviruses is mainly found on the Y-chromosome. Interestingly, the genome-editing competence of ALU elements is evident in that they can change their own gene expression by modifying their own genomic meythylation status (Batzer and Deininger, 2002; Ryan, 2004).

The most active period of endogenous retroviral transcription occurs during the formation of placental tissue, during growth periods, and when trophoblasts join together (Villarreal, 2004: 314). Trophoblasts encapsulate the egg, help the egg nest properly, trigger processes that ensure nutrition, and prevent reactive responses by the mother's own immune system. The egg is protected by trophoblasts against an immuno-reactive response of the mother. These characteristics are unknown to monotreme mammals and marsupials. The acquisition of such abilities must have been a remarkable event (Villarreal, 2004: 314).

In turn, the trophectoderm is a very complex tissue that is, surprisingly, not of maternal origin, but a derivative of the fertilized egg – it even develops before the egg becomes implanted into the uteral lining. Experiments that suppressed expression of endogenous retroviruses inhibited implantation into the uteral lining. This implies that implantation of the embryo requires transcription of retroviral syncytin-coding genes. In humans, the HERV W *env* gene codes for syncytin (Dupressoir *et al.*, 2005), a molecule used by the host to join trophoblast cells with the tissue that eventually nourishes the embryo (Villarreal, 2004: 315).

Although these processes have already been known for over 30 years, the purpose of this reaction was unclear: it did not make sense that the evolutionary innovation of placental mammals was tied to the acquisition of a complex set of endogenous retroviruses. Since the trophectoderm is protected by the maternal immune system, it enables further growth into the placenta, thereby modifying blood flow and nutrient supply between mother and embryo. The trophectoderm is associated with extremely high expression rates of ERV genes that result in RNAs as well as in other gene products and retroviral corpuscles. ERVs are highly host-specific and are closely associated with LINEs and SINEs of placental species. Their expression is not suppressed in the trophectoderm.

Once the sex of the totipotent embryo is determined, the high ERV expression rates are stopped and DNA methylation starts functioning (Villarreal, 2005: 325). Interestingly, ERV competences protect the embryo from the maternal immune system until the embryo's sex is determined. Only then does DNA methylation mediate between growth and development.

There are, however, clear references for evolutive and physiologically relevant qualities. For example, the expression of HERV-3 is boosted because it involves many fetal tissues in humans such as the adrenal cortex, kidney tubules, tongue, heart, liver and central nervous system as well as sebaceous glands of normal skin (Ryan, 2006). Thus, important tissues are formed during the fetal stage and are mediated via the presence of human endogenous retroviruses that were expressed during early mitotic divisions. They safeguard not only the formation of the placenta but also of the most important tissues of the fetus. If these retroviral components were removed from our genomes, we would be already extinct.

9. SUMMARY

Phylogenetic analyses as well as GenBank database comparisons show that most natural genome-editing competences are not of cellular origin but represent original skills of viruses.

Far from being only stochastic results of randomly derived chance combinations as proposed earlier (Eigen *et al.*, 1981; Maynard-Smith, 1983; Wächtershäuser, 1992; Jortner, 2006; Deamer *et al.*, 2006; Szathmary, 2006), genome-editing competences are built on precise combinatorial rules (molecular syntax), highly sensitive to interactional contexts (pragmatic rules) that determine differences in the semantic content, e.g., multiple protein meanings generated according to higher-order regulatory functions (Witzany, 2000, 2006a). Natural genome editing of viruses depends primarily on their biosemiotic competences, i.e., their capability to generate and constitute genetic content as a language-like text according to syntactic, pragmatic and semantic rules in coherence with the laws of physics and chemistry. Through these competences the genetic code is used as a kind of innovation-sharing protocol (Vetsigian *et al.*, 2006).

Viruses have two completely different life strategies, which are clearly reflected in their genomes. Accordingly, acute viruses exhibit lytic action and can induce disease and even death, whereas the life strategy of persistent viruses implies compatible interactions with the host. The latter are either integrated into the hosting genome or the cell plasma, and act non-destructively during most life stages of the host. The persistent life-style allows the virus to transmit complex viral phenotypes to the hosting organism. Doing so enables the host to broaden evolutive potentials that can lead to the formation of new species.

The natural genome-editing competences of viruses are most complex in bacteria, 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

(Villarreal, 2006, personal communication). Massive viral colonization occurred from the very beginning of life, starting with the evolution of Bacteria and Archaea, later on of Protoctista and multicellular Eukaryota. The formation of all kingdoms, their families, genera, and species relies on the effects of viral colonization and results in diversified lineages and ultimately in the evolution of new species.

Increasing complexity and diversity involved genetic innovations, new combinatorial patterns of genetic content, non-coding regulatory networks and modifications of the genomic architecture. Interestingly, increasing complexity correlates with the expansion of non-coding DNA (Taft and Mattick, 2004). Combinatorial and rearranging processes in evolutionary and developmental processes occur non-randomly. They need successful and coherent adherence to the rules of molecular syntax (Eigen and Winkler 1975; Witzany, 1995). Genetic content within the genomic matrix depends on the situational contexts in which living organisms are involved in vivo (e.g., growth, mating, virulence, stress, etc.). This content can therefore produce multiple protein-meanings, i.e., different semantic contents of the same genetic data-set. This is the prerequisite for epigenetically induced evolutionary and developmental processes. These rule-governed processes may fail, with fatal consequences for the organism (Witzany, 2006a). If evolutionary processes are intertwined between different species, then 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.

From this perspective, evolutionary history emerges entirely differently than previously thought. This history is not an aggregation of chance mutations of the genetic text and its associated selection. Rather, it is a permanent and competent processing of genetic sequences to acquire previously unknown abilities and to ward off competing parasites via genomic innovation. This explains why scientists had difficulties understanding evolutive patterns before comparative genomics was established: the thesis of natural genome-editing competences of viruses would merely have been another curious hypothesis. Contemporary sequence analyses and intercomparisons underline the crucial evolutionary functions of viruses and their genome-editing competences for all life forms.

APPENDIX

Box 1: Examples of viral genome editing by inventing new genes as shown by Villarreal

√	RNA, DNA
√	Replicase, polymerase, integrase
√	DNA repair
√	Restriction / modification
√	Methylation
√	Bilayer nuclear envelope
√	Eukaryotic nucleus
√	Division of transcription and translation
√	Nuclear pores
√	Tubulin-based chromosome duplication
√	Chitin, calcification
√	Linear chromosomes
√	Innate immune system (MHC-Komplex, RNAi)
√	Adaptive immune system
√	Cartilage, bones
√	Skin, dermal glands for poison, mucus and milk
√	Larvae, egg, placenta, flowering plants
√	Viviparous mammals

REFERENCES

Ast, G. (2005). The alternative genome. Scientific American 292: 58–65.

Batzer, M.A. and D.L. Deininger (2002). ALU repeats and human genomic diversity. Nature Reviews Genetics 3: 370–380.

Bell, P.J.L. (2001). Viral eukaryogenesis: Was the ancestor of the nucleus a complex DNA virus? Journal of Molecular Evolution 53: 251–256.

Bell, P.J.L. (2006). Sex and the eukaryotioc cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. Journal of Theoretical Biology, Doi: 10.1016/j.jtbi.2006.05.015.

Brett, D., H. Pospisil, J. Valcarcel, J. Reich and P. Bork (2002). Alternative splicing and genome complexity. Nature Genetics 30: 29–30.

Buchon, N. and C. Vaury (2006). RNAi: A defensive RNA-silencing against viruses and transposable elements. Heredity 96: 195–202.

Carroll, S.B. (2005). Evolution at two levels: On genes and form. PLoS Biology 3(7): e245 .

Deamer, D., S. Singaram, S. Rajamani, V. Kompanichenko and S. Guggenheim (2006). Self-assembly processes in the prebiotic environment. Philosophical Transactions of the Royal Society 361: 1809–1818.

Deepak, G., P.P. Majumder, C.B. Rao, S.K. Brahmachari and M. Mukerji (2003). Nonrandom distribution of alu elements in genes of various functional categories: Insight from analysis of human chromosomes 21 and 22. Molecular Biology and Evolution 20: 1420–1424.

Dupressoir, A., G. Marceau, C. Vernochet, L. Benit, C. Kanellopoulos, V. Sapin and T. Heidmann (2005). Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proceedings of the National Academy of Sciences of the USA 102: 725–730.

Eigen, M. and R. Winkler (1975). Das Spiel – Naturgesetze steuern den Zufall. München: Piper.

Eigen, M., P. Schuster, W. Gardiner and R. Winkler-Oswatitsch (1981). The origin of genetic information. Scientific American 244: 78–94.

Fire, A. (2005). Nucleic acid structure and intracellular immunity: Some recent ideas from the world of RNAi. Quarterly Reviews of Biophysics 38: 303–309.
Forterre, P. (2001). Genomics and early cellular evolution. The origin of the DNA world. Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie 324: 1067–1076.
Forterre, P. (2002). The origin of DNA genomes and DNA replication proteins. Current Opinion in Microbiology 5: 525–532.
Forterre, P. (2005). The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. Biochimie 87: 793–803.
Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. Virus Research 117: 5–16.
Gabora, L. (2006). Self-other organization: Why early life did not evolve through natural selection. Journal of Theoretical Biology 241: 443–450.
Gao, X., E.R. Havecker, P.V. Baranov, J.F. Atkins and D.F. Voytas (2003). Translational recoding signals between *gag* and *pol* in diverse LTR retrotransposons. RNA 9: 1422–1430.
Jablonka, E. and M.J. Lamb (1989). The inheritance of acquired epigenetic variations. Journal of Theoretical Biology 139: 69–83.
Jablonka, E. and M.J. Lamb (2002). The changing concept of epigenetics. In: Speybroeck, L.v., Vijver, G.V.d. and Waele, D.D. (Eds.). From Epigenesis to Epigenetics. The Genome in Context. Annals of the New York Academy of Sciences 981: 82–96.
Jablonka, E. and M.J. Lamb (2006). The evolution of information in the major transitions. Journal of Theoretical Biology 239: 236–246.
Jaenisch, R. and A. Bird (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. Nature Genetics 33: 245–254.
Jenuwein, T. and C.D. Allis (2001). Translating the histone code. Science 293: 1074–1080.
Jortner, J. (2006). Conditions for the emergence of life on the early earth: Summary and reflections. Philosophical Transactions of the Royal Society 361: 1877–1891.
Koonin, E.V., T.G. Senkevich and V.V. Dolja (2006). The ancient virus world and evolution of cells. Biology Direct 1: 29 .
Mantegna, R.N., S.V. Buldyrev, A.L. Goldberger, S. Havlin, S.C.K. Peng, M. Simons and H.E. Stanley (1994). Linguistic features of noncoding DNA sequences. Physical Review Letters 73: 3169–3172.
Margulis, L. (1996). Archaeal-eubacterial mergers in the origin of Eukarya: Phylogenetic classification of life. Proceedings of the National Academy of Sciences of the USA 93: 1071–1076.
Margulis, L. (1999). Die andere Evolution. Heidelberg: Spektrum Akademischer Verlag.
Margulis, L. (2004). Serial endosymbiotic theory (SET) and composite individuality. Transition from bacterial to eukaryotic genomes. Microbiology Today 31: 173–174.
Margulis, L., M.F. Dolan and R. Guerrero (2000). The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. Proceedings of the National Academy of Sciences of the USA 97: 6954–6959.
Margulis, L. and D. Sagan (2002). Acquiring genomes. A theory of the origin of species. New York: Basic Books.
Mattick, J.S. and M.J. Gagen (2001). The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms. Molecular Biology and Evolution 18: 1611–1630.
Mattick, J.S. (2001). Non-coding RNAs: The architects of eukaryotic complexity. EMBO Reports 2: 986–991.
Mattick, J.S. (2003). Challenging the dogma: The hidden layer of noncoding RNAs in complex organisms. BioEssays 25: 930–939.
Mattick, J.S. (2005). Das verkannte Genom-Programm. Spektrum der Wissenschaft 3: 62–69.
Maynard Smith, J. (1983). Models of evolution. Proceedings of the Royal Society of Biological Sciences 219: 315–325.
Odintsova, M.S. and N.P. Yurina (2000). RNA editing in plant chlorplasts and mitochondria. Fisiologia Rastenij 37: 307–320.
Odintsova, M.S. and N.P. Yurina (2005). Genomics and evolution of cellular organelles. Russian Journal of Genetics 41: 957–967.
Pollard, K.S., S.R. Salama, N. Lambert, M.A. Lambot, S. Coppens, J.S. Pedersen, S. Katzman, B. King, C. Onodera, A. Siepel, A.D. Kern, C. Dehay, H. Igel, M. Ares Jr, P. Vanderhaegen and D. Haussler (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature, DOI: 10.1038/nature05113.
Roossinck, M.J. (2005). Symbiosis versus competition in plant virus evolution. Nature Reviews Microbiology 3: 917–924.

Ryan, F.P. (2004). Human endogenous retroviruses in health and disease: a symbiotic perspective. Journal of the Royal Society of Medicine 97: 560–565.
Ryan, F.P. (2006). Genomic creativity and natural selection. A modern synthesis. Biological Journal of the Linnean Society 88: 655–672.
Shabalina, S.A. and N.A. Spiridonov (2004). The mammalian transcriptome and the function of non-coding DNA sequences. Genome Biology 5: 105e.
Shapiro, J.A. (2002). Genome organization and reorganization in evolution: Formatting for computation and function. In: Speybroeck, L.v., Vijver G.V.d. and Waele, DD. (Eds.). From Epigenesis to Epigenetics. The Genome in Context. Annals of the New York Academy of Sciences 981: 111–134.
Shapiro, J.A. (2004). A 21st century view of evolution: Genome system architecture, repetitive DNA, and natural genetic engineering. Gene 345: 91–100.
Shapiro, J.A. and R. Sternberg (2005). Why repetitive DNA is essential to genome function. Biological Reviews 80: 1–24.
Sternberg, R. (2002). On the roles of repetitive DNA elements in the context of a unified genomic-epigenetic system. Annals of the New York Academy of Sciences 981: 154–188.
Sternberg, R. and J.A. Shapiro (2005). How repeated retroelements format genome function. Cytogenetic and Genome Research 110: 108–116.
Szathmary, E. (2006). The origin of replicators and reproducers. Philosophical Transactions of the Royal Society 361: 1761–1776.
Taft, R.J. and J.S. Mattick (2004). Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. Genome Biology 5: 1.
True, H., I. Berlin and S.L. Lindquist (2004). Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. Nature 431: 184–187.
Turner, B.M. (2000). Histone acetylation and an epigenetic code. BioEssays 22: 836–845.
Turner, B.M. (2002). Cellular memory and the histone code. Cell 111: 285–291.
Van De Vijver G., L. Van Speybroeck and D. De Waele (2002). Epigentics: A challenge for genetics, evolution and development. In: Speybroeck, L.v., Vijver G.V.d. and D. DeWaele (Eds.). From Epigenesis to Epigenetics. The Genome in Context. Annals of the New York Academy of Sciences 981: 1–6.
Van Speybroeck, L., G. Van de Vijver and D. DeWaele (2002). Preface. In: Speybroeck, L.v., Vijver G.V.d. and D. De Waele (Eds.). From Epigenesis to Epigenetics. The Genome in Context. Annals of the New York Academy of Sciences 981: vii.
Vendrami, D. (2004). Noncoding DNA and the teem theory of inheritance, emotions and innate behaviour. Medical Hypotheses 64: 512–519.
Vetsigian, K., C. Woese, N. Goldenfeld (2006). Collective evolution and the genetic code. Proceedings of the National Academy of Sciences of the USA 103: 10696–10701.
Villarreal, L.P. (2004). Can viruses make us humans? Proceedings of the American Philosophical Society 148: 296–323.
Villarreal, L.P. (2005). Viruses and the Evolution of Life. Washington: American Society for Microbiology Press.
Villarreal, L.P., V.R. DeFilippis and K.A. Gottlieb (2000). Acute and persistent viral life strategies and their relationship to emerging diseases. Virology 272: 1–6.
Villarreal, L.P. and V.R. DeFilippis (2000). A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. Journal of Virology 74: 7079–7084.
Volff, J.N. (2006). Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28: 913–922.
Wächtershäuser, G. (1992). Groundworks for an evolutionary biochemistry: the iron–sulphur world. Progress in Biophysics and Molecular Biology 58: 85–201.
Wang, Y., W. Fischle, W. Cheung, S. Jacobs, S. Khorasanizadeh and C.D. Allis (2004). Beyond the double helix: Writing and reading the histone code. In: Bock, G. and Goode, J. (Eds.). Reversible Protein Acetylation. Novartis Foundation, Vol. 259, pp. 3–17.
Watson, J.D., J. Witkowski, M. Gilman and M. Zoller (1992). Recombinant DNA. Scientific American Books.
Wittgenstein, L. (1972). Philosophical Investigations. Oxford: Basil & Blackwell.
Witzany, G. (1995). From the 'logic of the molecular syntax' to molecular pragmatism. Explanatory deficits in Manfred Eigen's concept of language and communication. Evolution and Cognition 1: 148–168.
Witzany, G. (2000). Life: The communicative structure. A new philosophy of biology. Norderstedt: Libri Books on Demand.

Witzany, G. (2005). Natural history of life: History of communication logics and dynamics. SEED Journal 5: 27–55.
Witzany, G. (2006a). The *Logos* of the *Bios* 1. Contributions to the foundation of a three-leveled biosemiotics. Helsinki: Umweb.
Witzany, G. (2006b). Serial endosymbiotic theory (SET): The biosemiotic update. Acta Biotheoretica 54: 103–117.
Xu, Q., B. Modrek and C. Lee (2002). Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acids Research 30: 3754–3766.
Zhang, H.Y. (2006). The evolution of genomes and language. EMBO Reports 7: 248–249.