Biosemiotics
DOI 10.1007/s12304-011-9111-7

REVIEW

# Natural Genome Editing from a Biocommunicative Perspective

**Guenther Witzany**

Received: 28 October 2010 / Accepted: 28 January 2011


**Abstract** Natural genome editing from a biocommunicative perspective is the competent agent-driven generation and integration of meaningful nucleotide sequences into pre-existing genomic content arrangements, and the ability to (re-)combine and (re-) regulate them according to context-dependent (i.e. adaptational) purposes of the host organism. Natural genome editing integrates both natural editing of genetic code and epigenetic marking that determines genetic reading patterns. As agents that edit genetic code and epigenetically mark genomic structures, viral and subviral agents have been suggested because they may be evolutionarily older than cellular life. This hypothesis that viruses and viral-like agents edit genetic code is developed according to three well investigated examples that represent key evolutionary inventions in which non-lytic viral swarms act symbiotically in a persistent lifestyle within cellular host genomes: origin of eukaryotic nucleus, adaptive immunity, placental mammals. Additionally an abundance of various RNA elements cooperate in a variety of steps and substeps as regulatory and catalytic units with multiple competencies to act on the genetic code. Most of these RNA agents such as transposons, retroposons and small non-coding RNAs act consortially and are remnants of persistent viral infections that now act as co-opted adaptations in cellular key processes.

**Keywords** Natural languages/codes · Code concepts · Viral origins · Ribo-agents consortia

> The RNA is not the genome of a pathogen—the RNA is the pathogen and the entire infectious agent.
> *Roger W. Hendrix: Pittsburgh Bacteriophage Institute*

G. Witzany (✉)
Telos—Philosophische Praxis, Vogelsangstraße 18c, A-5111 Buermoos, Austria
e-mail: Witzany@sbg.at

Published online: 14 February 2011

## Introduction

For decades, there was no doubt that to speak about life meant to speak about living cells and cellular assemblies. This view changed little with the formulation of the RNA-world hypothesis with an abundance of competing and cooperating ribozymes. Recent findings in virology elucidated the fact that viruses code for typical features that are not part of cellular life, which seems to date viruses as older then cellular life. Additionally there are some indicators such as the origin of the eukaryotic nucleus, the origin of adaptive immunity and the role of endogenous retroviruses in placental mammals, that viruses are agents that edit the genome in host organisms. Additionally research on the components of transcripts from the DNA information storage medium demonstrated the presence of an RNA world with high diversity and an abundance of ribozymatic and regulatory functions in all key steps and even substeps of cellular replication such as expression, transcription, translation and repair.

In the following article, we will investigate theoretical thoughts on how to understand the genetic code, or as termed by Manfred Eigen, "the molecular syntax of the genetic language". That makes it necessary to change from an atomistic view on the DNA genetic storage medium to a more dynamic and less mechanistic view. In the end, we have to decide on the basis of the current knowledge whether one will follow the proponents of cell biology or those of virology, dependent on which data seem to have key role in evolution, development and genetic regulation. In both cases, natural genetic engineering and natural genome editing serve as appropriate designations for the highly dynamic processing of the genetic text (Witzany 2009a).

## Common Features of Natural Languages/Codes

It is now evident that the genetic code is not solely a genetic storage medium but offers highly active, dynamic networks of interactions during transcription. After transcription, a great variety of RNA interactions checks, changes and rearranges nucleotide sequences. The "code without commas" (Crick) has long been seen exclusively as a molecular assembly with underlying physical and chemical laws. Manfred Eigen explicitly relied on "syntax and semantics of the molecular language" (Eigen and Winkler 1975) and introduced a linguistic vocabulary into molecular biology in detail.

Patrick Forterre suggests defining life as both "ribosome encoding organisms and capsid-encoding organisms and their ancestors" (Forterre and Prangishvili 2009; Forterre 2010). Yet both encode using the same molecular alphabet, albeit encoding different products. They share a competence, i.e. to encode, 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 in the case of DNA serves as an information storage medium, and in the case of RNA serves also as information-based editing agents, as demonstrated below.

How should we think about evolution editing biological codes? Is the result that we identify in any genetic code sequence a consequence of gene duplication,

deletion, erosion, mutation, error or damage in the realm of physical and chemical laws according to chance and necessity? What would be needed?

If we look at all identified natural languages or codes, we find certain features which are empirically common to all kinds of natural languages and codes (Witzany 2000):

First of all, language/code doesn't speak (write) or code itself, i.e. there must be living agents which use the language/code to communicate for organisation and co-ordination interactions. Without living agents, there is no natural language or code.

Second, all languages/codes consist of a certain set of signs. These signs (icons, indices or symbols) are combined by sign-using agents to devise sign sequences which transport complex informational content with meaning-function for possible receivers.

Third, because sign-using agents are interwoven in continuous real-life scenarios, the context determines the meaning of sign sequences. This means that in natural languages, there is no context-free syntax or semantics.

This makes sense according to the energy costs of language use also. Language/code using living agents can generate identical sign sequences to transport different meanings in different situations. The humble dance in bees gives direction and energy costs in food gathering. Exactly same dance figures during the hive search where context transports the meaning of an appropriate hive (Witzany 2010). The sentence "Shooting of the hunters" in human communication may transmit the message that hunters shoot, conversely, that hunters are shot. Without context, contradictory semantics of sign sequences cannot be identified. Syntactic analyses alone are insufficient to identify semantic content.

Sign using agents which generate sign sequences to transport meaning within various contextual needs, i.e. living agents, communicate via language or code to coordinate and organise. As opposed to abiotic nature, language/code using living agents follow semiotic rules as described above. Combinatorial rules of sign assembles are referred to as syntactic rules, and the different contents of sign sequences are generated according semantic rules (meaning function). Different interactional contexts which are generated by sign-using agents (e.g. mating, development, attack, defence, food gathering, breeding, etc.) follow pragmatic rules (Witzany 1995) and are investigated in pragmatic analyses.

Any known natural language/code inherently represents the complementarity of these three levels of rules (syntax, pragmatics and semantics). As demonstrated by Charles Morris in the 1940s, if one level is really missing, there is no possibility to speak about a natural language or code (Morris 1938, 1946). As later demonstrated by Ludwig Wittgenstein, it is impossible for only one agent to follow a semiotic rule, i.e. rule following is inherently a consortial interaction (Wittgenstein 1975).

Bodies of living agents are built of cells which communicate via chemical signalling molecules and physical interactions to co-ordinate intracellular and intercellular behaviour. If communication is functioning, living bodies will prosper and be healthy; if communication is disturbed or damaged, living bodies will get sick, diseased or die. Intracellular as well as intercellular (cell-cell) communication follows syntactic, semantic and pragmatic rules (Witzany 2010). This feature is absent in abiotic nature. For example, no semiotic rules are present or even necessary as water freezes to ice.

If we summarise the essential characteristics of natural languages/codes we must notice consortia of living agents which are competent to use the language/code according to three levels of semiotic rules, i.e. the repertoire of signs are combined into sequences according to syntactic rules and situational contexts are generated according to adaptational purposes following pragmatic rules which determine content-specific rules, i.e. the meaning function of sign sequences.

What then does this mean if we speak about the genetic code, the DNA genetic storage medium, transcription, non-coding RNAs, RNA editing, epigenetic imprinting, alternative splicing, ribosomal assembly and translation into an amino acid language?

We certainly have to look at various concepts which deal with appropriate descriptions of the genetic code (intra-cellular communication) as well as appropriate description of cell-cell communication via molecular signalling.

## Concepts of the Genetic Code

After the exploration of the universal syntax and the structural code of DNA, the entanglement of linguistics and genetics was unavoidable. The genetic code integrates rules of informational encoding within DNA or RNA and its translation into proteins. The single nucleotides are assembles to three partite syntactic structures (codons) that correspond to certain amino acids. In this way genes are encoded and translated into proteins. Beneath the genetic code that encodes genes for proteins there are non-coding regions which bear regulatory elements, in most cases non-coding RNAs as well as start and stop signals. These regulatory sequences are integrated between or beside coding regions. Before a coherent protein coding sequence is generated all regulatory non-coding sequences are spliced out.

Noam Chomsky's approach linguistically constructed a formal meaning-free syntax which paved the way for bioinformatics and systems biology (Chomsky 1964, 1972). Manfred Eigen identified the understanding of the "molecular language" as a fundamental competence to understand the genetic code, but Eigen worked solely with a formal theory of language, i.e. "language" as a quanitifiable and measurable set of signs that depicts reality in a 1:1 fashion (Eigen and Winkler 1975). This was a purely mechanistic approach concerning the von Neumann automaton theory (Witzany 1995).

The systems biology approach investigates biological systems (cells, tissues, organs, organisms, and ecospheres) concerning the quantitative properties of their elementary building blocks and adapts them to the statistical methods of computer science (Baldi and Brunak 2001; Cristianini and Hahn 2006; Keedwell 2005). Such building blocks, e.g. a cellular network, can be modelled mathematically using the methods of chemical kinetics and control theory. Because of the large number of parameters, variables and constraints in cellular networks, numerical and computational techniques are used. This quantitative approach uses syntactic/semantic analyses but not pragmatically ones.

To use an appropriate tool for an explanation of the read and write properties in genetic sequences and to use terms of electronic computation cannot sufficiently serve as explanational pattern for cellular communication (and

organisation) because electronic computation depends on digitised processes according binary coding- and Turing machine- principles (Turing 1950). They rely on mathematical theories of language (Shannon and Weaver 1949), i.e. algorithm-based formalisable linguistic sequences which are not met in real-life languages but are artificial constructs in principle.

Bioinformatics interprets and investigates genetic structures in the light of categories of information theory (Popov et al. 1996; Ji 1997, 1999; Searls 2002; Chomsky 2004; Zhang 2006), and use statistic, computational, mathematical and therefore algorithm-based methods to identify sequence orders for measurements of sequence length and content homologies. Bioinformatics investigates language as a quantifiable set of signs and beliefs that it would be possible to extract semantic contents by analyses of the syntactic order.

This makes sense in genetic sequence comparison and comparative genomics, but has less ability for understanding the evolution of coded sequences because analyses of syntax of the genetic code do not tell us anything about the context in which the content bearer of genetic information is interwoven in real life. This context plays an important role in epigenetic imprinting and methylation patterns which are subject to alternative splicing that alters protein production.

Synthetic biology similar to systems biology aims to construct novel biotic functions by rewiring genetically-determined components, modules and networks (Serrano 2007; Mueller et al. 2009). Synthetic biology attempts to test hypotheses about necessary and sufficient components of a (natural) biological system by attempting to reconstruct some parts of such a system and therefore aims to decompose biological matter into smaller parts and components and divide them from each other into brick- or module-like parts that can be synthetically interchanged, rearranged and reassembled into new functional tools (Dymond et al. 2009).

The invention of new and complex genetic data sets and the coherent integration of new genes or gene blocks in pre-existent genetic content arrangements is not part of synthetic biology because innovative generation of non-random genetic content cannot be deduced out of mathematical concepts of language, i.e. formalisable procedures such as algorithms. Additionally, synthetic biology prefers formal systems with which they exclude real communicative acts and interactions, the a priori of natural language use.

Biosemiotics investigates semioses, i.e. sign-processes in biological processes including meaning and interpretation. As Semiotics biosemiotics focuses on codes in the world of living beings. Prominent biosemioticians designte that the use of semioses differentiates living from the world of non-living. The evolution of living nature therefore seems to start with the evolution of natural codes. The codes of life we find in a great variety starting from the genetic code, ribozymatic code, protein code, epigenetic code until the organismal codes in communication of bacteria, archaea, unicellular eukaryotes, fungi, animals and plants. Biosemiotics does not represent a unified theory that can be applied to empirical investigations (Favareau 2010). Parallel with this, biosemiotics is represented by diverse concepts (Sebeok and Umiker-Sebeok 1992; Hoffmeyer 1996; Barbieri 2001, 2007; Markos 2002; Brier 2008; Kull et al. 2009) with a natural science background such as mechanicism, physicalism, materialism, objectivism, information theory, systems

theory as well as other metaphysical constructions such as ontology or even a Peirce-derived pansemioticism (everything is a sign). Most biosemiotic investigations are focused on signs or the ontology of the relationship between signs or between signs and the signified something. The crucial role of pragmatics, i.e. the role of the real sign-user being part of the identity of a community of sign-users until now has not been part of the concepts of genetic code described above and most biosemiotic investigations (except e.g. Dario Martinelli, Karel Kleisner, Dominique Lestel).

## The Concept of Biocommunication

The biocommunicative approach integrates basic knowledge about the functioning of natural languages/codes by competent living agents (Witzany 2000, 2007, 2010). The biocommunicative approach investigates both communication processes within and among cells, tissues, organs and organisms as sign-mediated interactions, and nucleotide sequences as code, i.e. language-like text, which follows in parallel three kinds of rules: combinatorial (syntactic), context-sensitive (pragmatic) and content-specific (semantic) (Witzany 2010).

Natural genome editing from a biocommunicative perspective is competent agent-driven generation and integration of meaningful nucleotide sequences into pre-existing genomic content arrangements and the ability to (re-)combine and (re-) regulate them according to context-dependent (i.e. adaptational) purposes of the host organism (Witzany 2010).

If we speak about natural genome editing that is based on competent genetic code using and arranging living agents, we can both try to identify the cellular agents which are the responsible agents in intracellular and intercellular communication and investigate the agents which edit host genomes and determine identity, capabilities and competencies of cellular hosts (Witzany 2009a).

I will shortly review a few examples which we know of with these cellular properties/functions and capabilities clearly derived from persistent viral settlers of host organisms with the consequence that viruses are competent in natural genome editing.

## The Role of Viruses and Viral-like Agents

Although Salvator Luria expressed in the early 1950s that viruses could play major roles in the evolution of life, until now it has been mainstream thought that viruses are escaped genetic elements of cells that attained some autonomy from the cell that they left. Because they cannot replicate without cells, they must have originated later in evolution than the first cells. Only within the last decade has it become obvious that there are a variety of facts that do not fit in this picture, but fit better into the RNA world hypothesis (Domingo et al. 2008). According these data, RNA- and DNA-viruses have polyphyletic origins and represent a variety of features which are not present in cellular life (Villarreal and Witzany 2010).

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 yet exist (Forterre 2001, 2002, 2005, 2006; Villarreal 2005,

2009a; Koonin et al. 2006; Koonin 2009; Brüssow 2007; Jalasvuori 2010). Negative-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 via the last universal common ancestor (LUCA); however, there is no known similarity between RNA-viral replicases and those of any known cell type. DNA viruses also 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 life.

Because viruses combine a variety of ribozymatic competencies, they could be the agents of natural genome editing. Why should we consider these disease-causing parasites to be major drivers of evolution?

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 integration into the host genome or within the cell plasma, and are non-destructive during most life stages of the host (Villarreal 2007).

The persistent lifestyle allows the virus to transmit complex viral phenotypes to the host organism. Doing so enables the host to broaden its evolutionary potential that may well lead to the formation of new species.

The natural genome editing competencies of viruses are best characterized in bacteria in which the complete nucleotide word order is largely determined, combined, and recombined by viruses (Abedon 2011; Armon 2011). Hence, the main genomic novelties are found in the prokaryotic domain from where they originally evolved into higher life forms. Probably all basic enzymatic variations originated therein (Villarreal 2005, 2009a). Massive viral colonisation occurred from the very beginning of life. 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. 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).

Today, viruses are seen as the most abundant life forms in the oceans and the resource of most of the genetic diversity in the sea. It is estimated that $10^{30}$ viruses live in the ocean and $10^{23}$ infection events occur per second. They are the major source of mortality to all living agents in the sea, but are major settlers in genomes of sea organisms that serve as immune functions against infections by closely related viruses, and are therefore a major source of non-lytic viral settlement of host genomes (Suttle 2007).

## Viral Origin of the Eukaryotic Nucleus

The origin of the eukaryotic nucleus is a very old evolutionary event and serves as a rather impressive example for the important role of natural genome-editing

competencies 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 former free-living prokaryotes (see reviews Margulis 2004; Margulis and Sagan 2002). Mitochondria and other organelles clearly descended from these microorganisms (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 of mitochondria and chloroplasts. In fact, the nucleus has basic properties that are absent in any prokaryotic cells (Bell 2001, 2006).

All eukaryotic proteins involved in DNA replication differ from those found in prokaryotes. Hence, the nuclear properties of eukaryotes are completely different from those of prokaryotes (Villarreal 2005). These differences include:

- the use of linear chromosomes, with repetitive termination points and several origins for replication,
- transcription and translation separated by multiple membranes,
- the existence of complex nuclear pore structures that actively mediate RNA translocation,

All these properties represent complex phenotypes that require complex coordination of numerous protein functions. None of these functions are present in prokaryotes.

The eukaryotic nucleus contains three kinds of DNA-dependent RNA polymerases that differ significantly from the RNA polymerases of prokaryotes (Villarreal 2005). Even the three kinds of splicing group I-introns DNA-transposase, reverse transcriptase and micro-RNAs (see below) are largely non-existent in prokaryotes, although 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. The eukaryotic pore structure of the nuclear envelope likewise has no counterpart in the prokaryotic world (Villarreal 2009a).

The membrane-bound separation of transcription and translation is a characteristic of the poxviruses, more specifically of the Vaccinia and other DNA viruses (Takemura 2001). Moreover, these viruses have a very simple pore structure that has been actively 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, such as TTV1 and phycodnaviruses (Villarreal 2009a).

It has become 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 colonised a prokaryotic host (Bell 2001, 2006)

## Viral Origin of Adaptive Immunity in Bony Fish

The evolution of the adaptive immune system occurred long after the evolution of the eukaryotic nucleus during the transition from Urochordates (sea squirts/tunicates)

to bony fish. Here, one can still find many common tissue types; however, the bony structure required a huge increase in the genome. This was been predominantly achieved by retroposons, which originated from endogenous retroviruses. Urochordates, on the other hand, possess only an innate immune system.

Bony fish are the first in this line of descent that possess both an innate and an adaptive immune system (Villarreal 2005). It must be concluded that the entire complexity of an adaptive immune system was acquired at the very beginning of the vertebrate line of descent. At the same time, we can see that evolution simultaneously put forward jawbones, a vertebral spine and formation of the skull. Concomitantly, new viral families emerged that had not been found in any of the previous life forms; that is to say, four different kinds of negative-strand ssRNA virus families (rhabdoviruses, bunyaviruses, paramyxoviruses, orthomyxoviruses), and interestingly the first consortia of non-defective, autonomous and abundant retroviruses (Villarreal 2005). Hence, the adaptive immune system represents an interlinked network of proteins that tag inflammation and other acute processes (by cytokines and chemokines), and their receptors and signal transmission systems, that stimulate the humoral and cellular antigen-specific immune response pattern and must have been acquired in a single event because it is monophyletic (Villarreal 2009b).

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. This phenotype consists of a self-forming and a dynamically adapting genetic system that recognises "non-self" elements, and is thereby able to promptly attack and—interestingly—at the same time prevent fatal auto-aggression. In the context of a manifold self-identification system, this acquired gene-set strategy was developed in order to detect novel non-self agents. Once the system recognises 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 to detect such non-self agents. This kind of genetic recombination on a genetic level is found only from this point onwards and not in any evolutionary predecessors (Villarreal 2009b).

The overall result of such processes are cells that produce new classes of molecules that are able to (a) bind and suppress non-self agents or (b) let amoeboid cytotoxic cells find and neutralise agent-containing cells. Most of these characteristic properties of an adaptive immune system were not available before the evolution of vertebrates.

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). However, this tunicate system is missing 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, i.e. an adaptive immune system becomes devastating for non-self elements and at the same time protects 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 (Villarreal 2009b). Hosts which incorporated it acquired a system with demolishing capabilities. This ability to

kill is comparable to a potent toxin that lyses any cell exposed to it. On the other hand, through this process of self-recognition, the host must be able to prevent autolysis, comparable to an antitoxin (Villarreal 2005). Luis P. Villarreal introduced this concept of "addiction module" for the counterbalanced feature of former competing viruses in persistent lifestyle in host organisms. Additionally he could demonstrate that addiction modules such as toxin/antitoxin or restriction/modification in bacteria exactly represent counterbalanced viral properties within bacterial host organisms that originated by multiple phage infections (Villarreal 2005, 2009a)

Similar to other addiction modules (Melderen and Saavedra De Bast 2009) the (a) lethal, toxic ability of the adaptive immune system is considered as stable and long-lasting, while (b) the ability to express antitoxic characteristics (self-recognition) is only a temporary feature and 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 varying stability of the toxin with regards to the protecting antitoxin. Many of these individual elements that are used for an adaptive immune system evolved individually. They definitely had not existed previously in cellular organisms but rather in different types of viruses (Villarreal 2009b).

## Viral Origin of Placental Mammals

The evolution of the placenta is a relatively late event. The precursor cells of the human placenta for example, trophoblasts, first appear four days after fertilisation as the outer layer of cells of the blastocyst. Trophoblast cells differentiate into cytotrophoblast cells and syncytiotrophobalst cells. The role of syncytin in trophoblast development is (a) the protection of the embryo from the maternal immune system and (b) the regulated invasion of maternal tissue. Syncytin is coded by an endogenous retrovirus (Mallet et al. 2004). In contrast to mammalian tumour cells which have lost the capability to react to stop signals which prevent the invasion of tissues, trophoblast cells stop at a certain level of invaded maternal tissue (Prudhomme et al. 2005). The close ties between the human genome and the colonisation with repetitive sequences (LTRs, SINEs, LINEs) become obvious when considering the Y chromosome (Villarreal 2004, 2009b). 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, 2009)

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). 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 (Prudhomme et al. 2005). The egg is protected by trophoblasts against an immunoreactive response by the mother. These characteristics are unknown to monotreme mammals and marsupials. The acquisition of such abilities must have been a remarkable evolutionary event (Villarreal 2004).

In turn, the trophectoderm is a very complex tissue that is, surprisingly, not of maternal origin, but a derivative of the fertilised egg. It even develops before the egg becomes implanted into the uteral lining. Experiments that suppressed the expression

of syncytin encoding 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).

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 as imprinted gene silencing (Villarreal 2005; Hudson et al. 2010). Some endogenous retroviruses are epigenetically regulated according to developmental stages (Gimenez et al. 2009). There are, however, clear references for evolutionary 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 the sebaceous glands of normal skin (Ryan 2004, 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 division. They safeguard not only the formation of the placenta but also of the most important tissues of the foetus.

## Present RNA-Agents: Co-opted Remnants of Persistent Viral Infections

### Mobile Genetic Agents Share Repeat-sequences

Some are active as their RNA intermediate (retrotransposons), while some are active as mobile DNA (DNA transposons). Mobile genetic elements are interacting genetic agents (Bapteste and Burian 2010) that replicate by either cut and paste (DNA transposons; Class II elements) or copy and paste (retroposons; Class I elements) processes. Mobile genetic elements are flanked by repeat sequences.

Mobile genetic elements share repeat sequences as essential parts of their identity. This is an important feature, because non-repeat sequences are the most relevant part of protein coding sequences of translational mRNAs, a coherent protein coding line-up of exons in which all intronic sequences are spliced out (Jurka et al. 2007). The intronic sequences are known as regulatory elements of great diversity. Additionally, it must be mentioned that transposable elements may contribute to major evolutionary processes in altering and expanding host genomes by changes to gene regulation (Baertsch et al. 2008). Domesticated transposons play important roles in activation of adaptive immunity (O'Donnell and Burns 2010) But transposable elements must also be suppressed sufficiently so as to not cause disease (Schumann 2007), and are therefore the subject of epigenetic regulation (Lisch 2008). Suppression is stabilised by epigenetic regulation such as RNA interference, DNA methylation and histone modifications (Slotkin and Martienssen 2007).

We now know that there are many overlapping genes and dispersed genetic elements which together constitute mobile genetic elements. Long terminal

repeats (LTRs) constitute compact nuclear structures like centromeres and the related telomeres (Blackburn 2000, 2006; Witzany 2008). Repetitive genetic elements delineate centromeres and form telomeres by non-LTR-retroposons (Shapiro and Sternberg 2005). A large part of genomic repetitive DNA is reverse transcribed, and plays a major role in the physical structured order of the genome as well as formatting functions for expression, replication, transmission, repair, restructuring, cell division and differentiation (Sternberg and Shapiro 2005; Sciamanna et al. 2009). If we look at the essential roles of repetitive genetic elements, such as

- transcription (promoters, enhancers, silencers, transcription attenuation, terminators and regulatory RNAs),
- post-transcriptional RNA processing (mRNA targeting, RNA editing),
- translation (enhancement of SINE mRNA translation),
- DNA replication (origins, centromeres, telomeres, meiotic pairing and recombination),
- localisation and movement, chromatin organisation (heterochromatin, nucleosome positioning elements, epigenetic memory, methylation, epigenetic imprinting and modification),
- error correction and repair (double-strand break repair by homologous recombination, methyl-directed mismatch repair) and
- DNA restructuring (antigenic variation, phase variation, genome plasticity, uptake and integration of laterally transferred DNA, chromatin diminution, VDJ recombination, immunoglobulin class switching)

it is apparent that the essential agents are all retroelements such as LINEs, SINEs, LTR-retroposons, non-LTR-retroposons and ALUs (Sternberg and Shapiro 2005). Whereas in mammals, genome formatting occurs mainly by SINEs, LINEs and ALUs, in plant genomes, LTR retroposons play major roles in this respect (Weiner 2006). Genome size in both is determined by repetitive DNA abundance. Interestingly, the distance between coding sequences and regulatory repetitive sequences is an important parameter (Zuckerkandl 2002). Additionally it should be noticed that epigenetic marking is a commonly shared feature of all viruses (Villarreal 2005). If viruses truly predate evolution of cellular life it seems most likely that this feature has been co-apted from viruses by cellular life forms via persistent infection events to expand genetic reading patterns out of given genetic datasets.

In contrast to the evolutionary relevance of point mutations that is a relatively slow process, 45% of the human genome is composed of retroposed elements in which whole gene sequences may be integrated within one retrointegration event. Interestingly the formerly non-protein coding sequences have found various ways for exaptational processes and have acquired important protein coding functions (Brosius 1999; Baertsch et al. 2008). We can find also other transposable element-derived exaptations, i.e. a co-opted adaptation (Krull et al. 2005), as documented in the siRNA to microRNA evolutionary transition from a defence mechanism against transposable elements into a regulatory function of host gene expression (Doench et al. 2003; Piriyapongsa et al. 2007; Piryapongsa and King Jordan 2008). There are some indicators that the so-called pseudogenes, the molecular components of broken or damaged genes, can also be exapted in

changing circumstances (Balakirev and Ayala 2003). Transposable elements are clearly genetic parasites, but their action patterns are not selfish but instead co-evolve with the host and therefore act in an extremely species-specific manner (Wessler 2006).

## Module-like Riboagents: Editosome, Spliceosome, Ribosome

Important riboagents act as module-like consortia. Three RNA assemblages are currently known to play vital roles in editing genetic text as a read-and-write medium. This indicates sequence-specific identification competence for insertion and deletion activities, which alter semantic content (the function which leads to altered regulation or altered protein production) of primary transcripts out of the DNA storage medium. Editosomes, spliceosomes and ribosomes are constituted of a variety of components that counterregulate themselves during assembly. Slight changes in the composition structure delete their functionality, so it must be assumed that the assembly is a complementarity, i.e. that all parts play an important function. These subcomponents are ribomodules which constitute highly active riboagents with a variety of competencies to act on genetic transcripts.

RNA editing by editosomes is a co- or post-transcriptional process which alters the RNA sequence derived complementarily to the DNA from which it was transcribed (Gott 2003; Gott and Rhee 2008). RNA editing occurs in single celled and all kingdoms of multicellular eukaryotes. RNA editing changes gene sequences at the RNA level (Bass 2002). The edited mRNA specifies an amino acid sequence that is different from the protein that could be expected and is encoded by the genomic DNA of the primary transcript (Takenaka et al. 2008). RNA editing alterations of such transcribed RNA sequences occur by modification, substitution and insertion/deletion processes (Smith 2008; Homann 2008). According the suggestion of Grosjean, any sequence alteration that changes the genetic meaning of a transcript is termed editing, whereas structural changes solely are called modifications (Grosjean and Bjork 2004). The editosome is an assembly of ribonucleoproteins that have shown high sequence similarity to DNA repair enzymes (Panigrahi et al. 2003). Before the transcript is processed to the final mRNA for translation into protein, alternative splicing (Matlin and Moore 2007) occurs. As in ribosome and editosome assembly and also in spliceosome construction, a ribonucleoprotein complex is assembled in various steps that cuts out introns and splices exons together (Pyle and Lambowitz 2006). RNA editing predates splicing and is heavily interconnected, so that the editosome and spliceosome are important players in natural genome editing. The spliceosomal ribonucleoproteins are mainly small nuclear RNAs that are interconnected with at least 300 different proteins which are involved in mammalian pre-RNA splicing.

Ribosomes are composed of two thirds RNA and one third protein (Noller 2006). Ribosomes are assembled into a functional complex. As it is understood today, ribosomal proteins are vital for structural stabilisation. Around the catalytic site of the ribosome there are only RNAs and no ribosomal proteins. This means that the ribosome was originally a ribozyme, and proteins are not involved in the catalytic activity (Moore and Steitz 2006; Hamann and Westhof 2007).

## Fine-tuning Regulatory Agents: Non-coding RNAs

Non-coding RNAs that function in gene regulation coordinate and organise various actions, such as chromatin modification and epigenetic memory, transcriptional regulation, control of alternative splicing, RNA modification and RNA editing, control of mRNA turnover, control of translation and signal transduction (Matera et al. 2007; Amaral et al. 2008; Witzany 2009b).

Most of these non-coding RNAs are alternatively spliced and divided into smaller RNAs that are integral parts of ribonucleoprotein (RNP) complexes (Cech et al. 2006). They regulate nearly all aspects of gene regulation (Demongeot et al. 2009). Small RNA species include micro RNAs, small interfering RNAs, small nuclear RNAs and small nucleolar and transfer RNAs (Yazgan and Krebs 2007). Although recent research has tried to evaluate the enormous regulatory networks of small RNAs, the role of thousands of longer transcripts is not yet clear. We know that they play important roles in histone modification and methylation, that is, epigenetic control of developmental processes such as the mammalian HOX clusters (Amaral et al. 2008), and also transcriptional interference, promoter inactivation and effects on enzymatic pathways.

Small non-coding RNAs also share a special competence for epigenetic regulation of gene expression and are derived from repetitive genomic sequences (Ambros and Chen 2007; Farazi et al. 2008; Taft et al. 2009). The capacity for epigenetic regulation of gene expression includes the "recognition" (identification) of specific sequences in other nucleic acids and is common to RNAs (Filipowicz 2000; Chu and Rana 2007), especially small nuclear RNAs and tRNAs that (a) identify splice junctions in both pre-mRNAs and codons, and (b) process both the subunits of the spliceosome and the ribosome (Mattick and Gagen 2001). Micro RNAs are single-stranded RNAs 19–25 nucleotides in length and are generated from endogenous hairpin transcripts of 70 nucleotide precursor miRNAs (Kim 2005). The transcription of this pre-miRNA is processed by RNA polymerases pol II and pol III, whereas pol II produces messenger RNAs, small nucleolar and small nuclear RNAs of the spliceosome (Chen and Rajewsky 2007; Dieci et al. 2009). Pol III produces shorter non-coding RNAs, such as tRNAs, some rRNAs, and a nuclear RNA that is part of the spliceosome (Bartel 2004). They control not only developmental timing, hematopoiesis, organogenesis, apoptosis and cell proliferation, but also fat metabolism in flies, neuronal patterning in nematodes, and control of leaf and flower development in plants (Bartel 2004).

Transposable elements in cellular genomes are most likely remnants of viral infection events (Villarreal and DeFilippis 2000; Villarreal 2004, 2005, 2009a). Additionally, the repeat sequences of mobile genetic elements such as LINEs, SINEs, LTR-retroposons, non-LTR-retroposons and ALUs are clearly related to retroviruses as are reverse transcriptases. Also, repeat sequences found in telomeres and centromeres are most likely of viral origin (Witzany 2008). There are strong indicators that the variety of introns and non-coding RNAs (Toor et al. 2001), because of their repetitive sequences, are of retroviral infection events and currently act as modular tools for cellular regulatory needs (Witzany 2009b).

## Conclusions

Persistent viral lifestyles are equilibrium interactions, counterbalanced viral properties (e.g. toxin-antitoxin modules), that transfer complete genetic datasets into a host genome and alter the genetic identity of the host and the formerly competing viral agents. Viruses can integrate their genomes into host genomes without damage to the host genome content-order that programs host metabolism. Viruses can integrate host genes within their genome also.

Transcribed out of DNA cellular sequences, the RNA activated inhabitants from former viral infection events act as modular tools for cellular needs in nearly all cellular processes. That they act not as lytic agents but as part of the host genetic identity is the result of inhabitation by counterbalanced competing genetic parasites. Addiction modules not only change genetic identity but enrich immune functions of host organisms against related parasites through Toxin/Antitoxin Modules.

Essential modules of RNA viruses are ribozymes and ribozymatic structures which autocatalyse and build *ribomodules*. Conserved structures like tRNAs and more the complex editosomes, spliceosomes and ribosomes consist of ensembles of such *consortia of riboagents* in which proteins stabilize its structure. As endogenised modules, they regulate genetic expression of host organisms.

Insertion, deletion, rearrangement, recombination, replication and repair of nucleotide sequences in all detailed steps are coordinated processes in a timely manner. An abundance of small non-coding RNAs act as competent agents in most cases combined within a network of other RNAs that bind proteins (Ribonucleo-proteins) which together build a fine-tuned network in the division of labour.

Some RNAs such as ribozymes are able to self-catalyse, eg. group II self splicing Introns. Reverse transcriptase is the only known enzyme that synthesises DNA out of RNA templates and is the essential feature of reverse transcribing viruses and for retroelements such as retrons, retrotransposons and retroplasmids.

For several decades neo-darwinistic concepts have dominated evolutionary theory discourse, i.e. the mutation-selection narrative. But this old narrative “mutations” (error, damage) and its selection cannot explain expanding genetic innovation, recombination, precise insertion, deletion, symbiogenetic fusion, hybridisation, gene transfer and integration of up until 100 new genes in a single event. If we follow Manfred Eigen and look at gene sequences as a language-like text with molecular syntax similar to the concepts of code described above we have to search for agents that are competent to edit the genetic code. As Charles Morris demonstrated convincingly in the late 40ies of the last century a coherent analysis of languge-like texts has to include all three levels of semiotic rules, i.e. syntaxc, pragmatics and semantics. No known natural code codes itself, nor is a natural code the result of a series of statistical errors. Therefore I tried to identify such agents in looking on evolutionary old genetic structures such as ribozymes and viruses that perform important functions in present living organisms as demonstrated above. The most prominent example in this field seem to be the non-lytic but persistent lifestyle of viruses in cellular host organisms.

**Acknowledgements** The author wants to thank Luis P. Villarreal, Director of Virus Research, Irvine, California for interesting comments and anonymous reviewer for helpful suggestions.

## References


Abedon, S. (2011). Communication among phages, bacteria, and soil environments. In G. Witzany (Ed.), *Biocommunication in soil microorganisms* (pp. 37–65). Berlin: Springer.

Amaral, P. P., Dinger, M. E., Mercer, T. R., & Mattick, J. S. (2008). The eukaryotic genome as an RNA machine. *Science, 319*, 1787–1789.

Ambros, V., & Chen, X. (2007). The regulation of genes and genomes by small RNAs. *Development, 134*, 1635–1641.

Armon, R. (2011). Soil bacteria and Bacteriophages. In G. Witzany (Ed.), *Biocommunication in soil microorganisms* (pp. 67–112). Berlin: Springer.

Balakirev, E. S., & Ayala, F. J. (2003). Pseudogenes: are they "Junk" or Functional DNA? *Annual Review of Genetics, 37*, 123–151.

Baldi, P., & Brunak, S. (2001). *Bioinformatics: The machine learning approach* (2nd ed.). Cambridge: MIT Press.

Bapteste, E., & Burian, R. M. (2010). On the need for integrative phylogenomics and some steps toward its creation. *Biology and Philosophy, 25*, 711–736.

Baertsch, R., Diekhans, M., Kent, W. J., Haussler, D., & Brosius, J. (2008). Retrocopy contributions to the evolution of the human genome. *BMC Genomics, 9*, 466. doi:10.1186/1471-2164-9-466.

Barbieri, M. (2001). *The organic codes. The birth of semantic biology*. Ancona: PeQuod.

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

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism and function. *Cell, 116*, 281–297.

Bass, B. L. (2002). RNA editing by adenosine deaminases that act on RNA. *Annual Reviews in Biochemistry, 71*, 817–846.

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

Bell, 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 eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. *Journal of Theoretical Biology, 243*, 54–63.

Blackburn, E. H. (2000). The end of the (DNA) line. *Nature Structural Biology, 7*, 847–850.

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

Brier, S. (2008). *Cybersemiotics: Why information is not enough*. Toronto: Toronto University Press.

Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene, 238*, 115–134.

Brüssow, H. (2007). *The quest for food. A natural history of eating*. New York: Springer Science and Business Media.

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

Chen, K., & Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nature Review Genetics, 8*, 93–103.

Chomsky, N. (1964). *Current Issues in linguistic theory*. London: The Hague, Mouton.

Chomsky, N. (1972). *Studies on semantics in generative grammar*. The Hague: Mouton.

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

Cristianini, N., & Hahn, M. (2006). *Introduction to computational genomics*. New York: Cambridge University Press.

Chu, C. Y., & Rana, T. M. (2007). Small RNAs: regulators and guardians of the genome. *Journal of Cell Physiology, 213*, 412–419.

Demongeot, J., Glade, N., Moreira, A., & Vial, L. (2009). RNA relics and origin of life. *International Journal of Molecular Science, 10*, 3420–3441.

Dieci, G., Preti, M., & Montanini, B. (2009). Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics, 94*, 83–88.

Doench, J. G., Petersen, C. P., & Sharp, P. A. (2003). siRNAs can function as miRNAs. *Genes & Development, 17*, 438–442.

Domingo, E., Parrish, C. R., & Holland, J. J. (2008). *Origin and evolution of viruses* (2nd ed.). San Diego: Academic.

Dupressoir, A., Marceau, G., Vernochet, C., Benit, L., Kanellopoulos, C., Sapin, V., et al. (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 United States of America, 102*, 725–730.

Dymond, J. S., Scheifele, L. Z., Richardson, S., Lee, P., Chandrasegaran, S., Bader, J. S., et al. (2009). Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary Build-a-Genome course. *Genetics, 181*, 13–21.

Eigen, M., & Winkler, R. (1975). *Das Spiel. Naturgesetze steuern den Zufall*. München: Pieper.

Farazi, T. A., Juranek, S. A., & Tuschl, T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development, 135*, 1201–1214.

Favareau. D, (2010) (Ed). Essential readings in biosemiotics. Dortrecht: Springer.

Filipowicz, W. (2000). Imprinted expression of small nucleolar RNAs in brain: time for RNomics. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 14035–14037.

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.

Forterre, P. (2010) Manipulation of cellular syntheses and the nature of viruses: The virocell concept. Comptes Rendus Chimie (in press).

Forterre, P., & Prangishvili, D. (2009). The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Annals of the New York Academy of Sciences, 1178*, 65–77.

Gimenez, J., Montgiraud, C., Oriol, G., Pichon, J. P., Ruel, K., Tsatsaris, V., et al. (2009). Comparative Methylation of ERVWE1/Syncytin-1 and other human endogenous retrovirus LTRs in placenta tissues. *DNA Research, 16*, 195–211.

Gott, J. M. (2003). Expanding genome capacity via RNA editing. *CR Biology, 326*, 901–908.

Gott, J. M., & Rhee, A. C. (2008). Insertion/deletion editing in physarum polycephalum. In H. U. Göringer (Ed.), *RNA Editing* (pp. 85–104). Berlin: Springer Verlag.

Grosjean, H., & Bjork, G. R. (2004). Enzymatic conversion of cytidine to lysidine in anticodon of bacterial isoleucyl-tRNA–an alternative way of RNA editing. *Trends in Biochemical Sciences, 29*, 165–168.

Hamann, C., & Westhof, E. (2007). Searching genomes for ribozymes and riboswitches. *Genome Biology, 8*, 210. doi:10.1186/gb-2007-8-4-210.

Hoffmeyer, J. (1996). *Signs of meaning in the Universe*. Bloomington: Indiana University Press.

Homann, M. (2008). Editing reactions from the perspective of RNA structure. In H. U. Göringer (Ed.), *RNA editing* (pp. 1–32). Berlin: Springer Verlag.

Hudson, Q. J., Kulinski, T. M., Huetter, S. P., & Barlow, D. P. (2010). Genomic Imprinting mechanisms in embryonic and extraembryonic mouse tissues. *Heredity, 1*, 45–56.

Jalasvuori, M. (2010). *Viruses are ancient parasites that have influenced the evolution of contemporary and archaic forms of life*. Jyväskylä: University Printing House.

Ji, S. (1997). Isomorphism between cell and human languages: molecular biological, bioinformatic and linguistic implications. *Biosystems, 44*, 17–39.

Ji, S. (1999). The linguistics of DNA: words, sentences, grammar, phonetics and semantics. *Annals of the New York Academy of Sciences of the USA, 870*, 411–417.

Jurka, J., Kapitonov, V. V., Kohany, O., & Jurka, M. V. (2007). Repetitive sequences in complex genomes: structure and evolution. *Annual Review of Genomics and Human Genetics, 8*, 241–259.

Keedwell, E. (2005). *Intelligent bioinformatics: The application of artificial intelligence techniques to bioinformatics problems*. Chichester: Wiley.

Kim, V. N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews. Molecular Cell Biology, 6*, 376–385.

Koonin, E. V. (2009). On the origin of cells and viruses: primordial virus world scenario. *Annals of the New York Academy of Sciences of the USA, 1178*, 47–64.

Koonin, E. V., Senkevich, T. G., & Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biology Direct, 1*, 29.

Krull, M., Brosius, J., & Schmitz, J. (2005). Alu-SINE exonization: en route to protein-coding function. *Molecular Biology and Evolution, 22*, 1702–1711.
Kull, K., Deacon, T., Emmeche, C., Hoffmeyer, J., & Stjernfelt, F. (2009). Thesis an biosemiotics: prolegomena to a theoretical biology. *Biological Theory, 4*, 167–173.
Lisch, D. (2008). Epigenetic regulation of transposable elements in plants. *Annual Review of Plant Biology, 60*, 43–66.
Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., et al. (2004). The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 1731–1736.
Margulis, L. (2004). Serial endosymbiotic theory (SET) and composite individuality. *Transition from bacterial to eukaryotic genomes. Microbiol Today, 31*, 173–174.
Margulis, L., & Sagan, D. (2002). *Acquiring genomes. A theory of the origin of species*. New York: Basic Books.
Markos, A. (2002). *Readers of the book of life*. Oxford: Oxford University Press.
Matera, A. G., Terns, R. M., & Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews. Molecular Cell Biology, 8*, 209–220.
Mattick, J., & Gagen, M. J. (2001). The evolution of controlled multitasked gene net-works: the role of introns and other noncoding RNAs in the development of complex organisms. *Molecular Biology and Evolution, 18*, 1611–1630.
Melderen, L. V., & Saavedra De Bast, M. (2009). Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genetics, 5*, e1000437. doi:10.1371/journal.pgen.1000437.
Matlin, A. J., & Moore, M. J. (2007). Spliceosome assembly and composition. *Advances in Experimental Medicine and Biology, 623*, 14–35.
Moore, P. B., & Steitz, T. A. (2006). The roles of RNA in the synthesis of protein. In R. F. Gesteland, T. R. Cech, & J. F. Atkins (Eds.), *The RNAWorld, third ed* (pp. 257–285). New York: Cold Spring Harbor Laboratory Press.
Morris, C. W. (1938). *Foundations of the theory of signs*. Chicago: University Press.
Morris, C. W. (1946). *Signs, language, and behavior*. New York: Braziller.
Mueller, S., Coleman, J. R., & Wimmer, E. (2009). Putting synthesis into biology—a viral view of genetic engineering through de novo gene and genome synthesis. *Chemical Biology, 16*, 337–347.
Noller, H. F. (2006). Evolution of ribosomes and translation. In R. F. Gesteland, T. R. Cech, & J. F. Atkins (Eds.), *The RNAWorld, third ed* (pp. 287–307). New York: Cold Spring Harbor Laboratory Press.
Odintsova, M. S., & Yurina, N. P. (2000). RNA editing in plant chlorplasts and mitochondria. *Fisiologia Rastenij, 37*, 307–320.
Odintsova, M. S., & Yurina, N. P. (2005). Genomics and evolution of cellular organelles. *Russian Journal of Genetics, 41*, 957–967.
O'Donnell, K.A., & Burns, K.H. (2010). Mobilizing diversity: transposable element insertions in genetic variation and disease. Mobile DNA 1:21. http://www.mobilednajournal.com/content/1/1/21
Panigrahi, A. K., Schnaufer, A., Ernst, N. L., Wang, B., Carmean, N., Salavati, R., et al. (2003). Identification of novel components of Trypanosoma brucei editosomes. *RNA, 9*, 484–492.
Piriyapongsa, J., Marino-Ramirez, L., & King Jordan, I. (2007). Origin and evolution of human micro RNAs from transposable elements. *Genetics, 176*, 1323–1337.
Piryapongsa, J., & King Jordan, I. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA, 14*, 814–821.
Popov, O., Degal, D. M., & Trifonov, E. N. (1996). Linguistic complexity of protein sequences as compared to texts of human languages. *Biosystems, 38*, 65–74.
Prudhomme, S., Bonnaud, B., & Mallet, F. (2005). Endogenous Retroviruses and animal reproduction. *Cytogenetics and Genome Research, 110*, 353–364.
Pyle, A. M., & Lambowitz, A. M. (2006). Group II introns: Ribozymes that splice RNA and invade DNA. In R. F. Gesteland, T. R. Cech, & J. F. Atkins (Eds.), *The RNAWorld, third ed* (pp. 468–506). New York: Cold Spring Harbor Laboratory Press.
Roossinck, M. (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.
Ryan, F. P. (2009). *Virolution*. London: Collins.

Sciamanna, I., Vitulloa, P., Curatoloa, A., & Spadafora, C. (2009). Retrotransposons, reverse transcriptase and the genesis of new genetic information. *Gene, 448*, 180–186.
Searls, D. B. (2002). The language of genes. *Nature, 420*, 211–217.
Sebeok, T., & Umiker-Sebeok, J. (1992). *Biosemiotics: The semiotic web 1991*. Berlin: Mouton de Gruyter.
Serrano, L. (2007). Synthetic biology: promises and challenges. *Molecular Systems Biology, 3*, 158–163.
Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
Schumann, G. G. (2007). APOPEC3 proteins: major players in intracellular defence against LINE-1 mediated retrotransposition. *Biochemistry Society Transactions, 35*, 637–642.
Shapiro, J. A., & Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biological Reviews, 80*, 1–24.
Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Review Genetics, 8*, 272–285.
Smith, H. C. (2008). Editing informational content of expressed DNA sequences and their transcripts. In H. U. Göringer (Ed.), *RNA editing* (pp. 249–265). Berlin: Springer Verlag.
Sternberg, R., & Shapiro, J. A. (2005). How repeated retroelements format genome function. *Cytogenet Genome Research, 110*, 108–116.
Suttle, C. A. (2007). Marine viruses – major players in the global ecosystem. *Nature Reviews. Microbiology, 5*, 801812.
Taft, R. J., Glazov, E. A., Lassmann, T., Hayashizaki, Y., Carninci, P., & Mattick, J. S. (2009). Small RNAs derived from snoRNAs. *RNA, 15*, 1233–1240.
Takemura, M. (2001). Poxviruses and the origin of the eukaryotic nucleus. *Journal of Molecular Evolution, 52*, 419–425.
Takenaka, M., Van Der Merwe, J. A., Verbitskiy, D., Neuwirt, J., Zehrmann, A., & Brennicke, A. (2008). RNA editing in plant mitochondria. In H. U. Göringer (Ed.), *RNA editing* (pp. 105–122). Berlin: Springer Verlag.
Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.
Toor, N., Hausner, G., & Zimmerly, S. (2001). Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases RNA 7, 1142–1152
Villarreal, L. P. (2004). Can viruses make us humans? *Proceedings. 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. (2007). Virus-host symbiosis mediated by persistence. *Symbiosis, 44*, 1–9.
Villarreal, L. P. (2009a). *Origin of group identity: Viruses, addiction and cooperation*. New York: Springer Science and Business Media.
Villarreal, L. P. (2009b). The source of self. Genetic parasites and the origin of adaptive immunity. *Annals of the New York Academy of Sciences, 1178*, 194–232.
Villarreal, L. P., & DeFilippis, V. R. (2000). A hypothesis for DNA viruses as the origin of eu-karyotic replication proteins. *Journal of Virology, 74*, 7079–7084.
Villarreal, L. P., & Witzany, G. (2010). Viruses are essential agents within the roots and stem of the tree of life. *Journal of Theoretical Biology, 262*, 698–710.
Weiner, A. M. (2006). SINEs and LINEs: Troublemakers, saboteurs, benefactors, ancestors. In R. F. Gesteland, T. R. Cech, & J. F. Atkins (Eds.), *The RNAWorld, third ed* (pp. 507–534). New York: Cold Spring Harbor Laboratory Press.
Wessler, S. (2006). Eukaryotic transposable elements: teaching old genomes new tricks. In L. Caporale (Ed.), *The implicit genome* (pp. 139–162). New York: Oxford University Press.
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*. Norderstedt: Libri Books on Demand.
Witzany, G. (2007). *The logos of the bios 2. Bio-communication*. Helsinki: Umweb.
Witzany, G. (2008). The viral origins of telomeres, telomerases and their important role in eukaryogenesis and genome maintenance. *Biosemiotics, 2*, 191–206.
Witzany, G. (2009a) (ed). Natural genetic engineering and natural genome editing. Annals of the New York Academy of Sciences, Volume 1178.

Witzany, G. (2009b). Non-coding RNAs: persistent viral agents as modular tools for cellular needs. *Annals of the New York Academy of Sciences, 1178*, 244–267.
Witzany, G. (2010). *Biocommunication and natural genome editing*. Dordrecht: Springer.
Wittgenstein, L. (1975). *Philosophische Untersuchungen*. Frankfurt: Suhrkamp.
Yazgan, O., & Krebs, J. E. (2007). Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes. *Biochemistry and Cell Biology, 85*, 484–496.
Zhang, H. Y. (2006). The evolution of genomes and language. *EMBO Reports, 7*, 748–749.
Zuckerkandl, E. (2002). Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. *Genetica, 115*, 105–129.