

Two genetic codes: Repetitive syntax for active non-coding RNAs; non-repetitive syntax for the DNA archives

Guenther Witzany

Telos - Philosophische Praxis, Buermoos, Austria

ABSTRACT

Current knowledge of the RNA world indicates 2 different genetic codes being present throughout the living world. In contrast to non-coding RNAs that are built of repetitive nucleotide syntax, the sequences that serve as templates for proteins share—as main characteristics—a non-repetitive syntax. Whereas non-coding RNAs build groups that serve as regulatory tools in nearly all genetic processes, the coding sections represent the evolutionarily successful function of the genetic information storage medium. This indicates that the differences in their syntax structure are coherent with the differences of the functions they represent. Interestingly, these 2 genetic codes resemble the function of all natural languages, i.e., the repetitive non-coding sequences serve as appropriate tool for organization, coordination and regulation of group behavior, and the non-repetitive coding sequences are for conservation of instrumental constructions, plans, blueprints for complex protein-body architecture. This differentiation may help to better understand RNA group behavioral motifs.

ARTICLE HISTORY

Received 14 February 2017
Accepted 16 February 2017

KEYWORDS

DNA; repetitive sequences;
RNA; RNA stem loops; RNA
group identities

Introduction

With the detection of the genetic code, a completely new perspective arose in biology. This new paradigm changed the previous views on heredity and cell biology. Molecular biology and genetics started their success stories. The “code without commas” (Francis Crick) consisted of a nucleotide alphabet with 4 characters and their complementary base pairs storing biological information to be translated into proteins by living cells. Later on, Manfred Eigen insisted that this genetic code represents a real language and not just a metaphor.¹ At this time, it was thought that every natural language or code was determined by the rules of its syntax, i.e., how the single characters could be correctly combined into a line up to build extended sequences of information. The meaning (semantics) of a certain DNA sequence is defined by the protein product as the result of gene expression, transcription and translation into the amino acid code.

Additionally, with progress in genetic research, some hallmarks of molecular biology and genetics were defined, such as one gene codes for one protein, the central dogma (DNA-RNA-protein-everything else), that non-protein coding DNA is junk, and last but not least, the fact that error replications drive genetic variations.²

The concept of Noam Chomsky of a context-free universal grammar being present in the nucleotide sequences

inspired bioinformatics to study what Manfred Eigen proposed in the early 70s: life is physics and chemistry and information.

But in the 80s of the last century, empirical evidence proved the fact that natural languages and codes are determined by a third level of rules besides grammar and semantics, namely pragmatics. Pragmatics represents the level of rules that interconnects the real character users of a natural language or code with its real life situations; this means the context in which a natural language or code user is interwoven. This became clear with the empirical fact that no natural language speaks itself as no natural code codes itself. In any known case, there are living agents being competent to generate and use a natural language or code.³ Pragmatics, therefore, represents the context-dependence of meaning which was formerly presumed to depend on the syntax structure of a natural language or code sequence. Since few decades, epigenetics proved this fact that it was not the nucleotide syntax alone which defines the final protein product, but the epigenetic markings.

Nearly at the same time, the above mentioned hallmark assumptions in molecular biology and genetics were falsified through empirical evidence in the roles of non-coding RNAs and viruses in evolution and development of all domains of life. At the current stage, the

dominant role of RNA biology and the dramatic comeback of virology changes our picture of the genetic code.

In this article, we will have a look at some functions and group behavioral motifs of such RNAs with their inherent repetitive nucleotide syntax and their role in constituting and regulating genetic content of organisms.

No life would function with DNA alone

Without transcription from the genetic storage medium DNA into the current RNA world agents, no relevant genetic process can be initiated.^{4,5} RNAs format the expression of coding sequences and organize the coherent line-up of timely coordinated steps of replication.^{6,7} In addition, the transport of genetic information to the progeny cells is coordinated by these agents. Furthermore, they are crucial for the cooperation between groups of RNA-stem loops to constitute important nucleoprotein complexes such as ribosome, spliceosome, and editosome.⁸ Therefore, such RNA groups are essential for complex order of genome constructions. The construction, organization and regulation of cellular genomes, especially of high complexity, are special capabilities of such RNA groups.⁹ This does not mean that the coded content is crucial in higher eukaryotes but that the organizational regulation of content arrangements that determine protein translations seem more important. If we look at the number of genes of *C. elegans* and humans, we have approximately 20,000 protein coding genes, but a very different regulatory network.

With the focus on RNA-biology, it soon became obvious, that the syntax of non-coding RNAs and coded DNA differs fundamentally. It was recognized that the RNA sequences, in most cases, consisted of repetitive characters constituting the sequence structure. Additionally, it was soon mentioned that the repetitive RNA sequences are the main proponents of regulatory elements of gene expression in all domains of life.¹⁰ If we take a repetitive sequence such as, e.g., TATAAACGCC we know that these repeats are among the strongest binders of histones.¹¹ With the detection of mobile genetic elements, it also became obvious that these agents of genetic novelty are constituted by repetitive elements.¹²

RNA agents are generally constituted by ribonucleotide sequences with repetitive grammar. In contrast to protein coding sequences of the genetic text that do not represent repetitive sequence syntax, the non-coding RNAs all share this syntax feature. In this review, we will have a look at some functions and group behavioral motifs of such RNAs and their role in constituting and regulating genetic content of organisms.

Repetitive genetic syntax structures represent active agents

The language of the coding regions is the evolutionarily successful tool to conserve complex information within a storage medium whereas the non-coding repeat syntax is the essential tool to actively coordinate and organize interactions, such as regulations or even innovations.¹³

To download biologically relevant information into real life-world, it is necessary to transcribe and translate DNA-stored information into dynamic RNA groups and protein worlds again. This occurs by:

- Splitting up double stranded DNA into 2 single strands by helicase.
- Building of a copy DNA by polymerase and its counterpart; free RNAs bind to this sequence now.
- RNA polymerases do not duplicate the sequence exactly but, on the contrary, act highly “error prone,” i.e., extend the given sequence into a new alphabetical order which can serve as a new genetic identity contrasting its origin.
- Building of a mRNA after modifying the transcript by (i) RNA editing, and (ii) splicing out the (regulatory) non-coding RNAs, i.e., introns, to
- transport information content out of the nucleus to ribosomes (after forming a RNP complex of the 2 ribosomal subunits) in the cytoplasm followed
- by the tRNAs for translation into proteins with low error rate.
- Recycling and re-use of the remaining RNAs of certain RNA groups (e.g., spliced out introns, ribosomal subunits, tRNAs,) for further regulatory processes, such as important roles of microRNAs.

This means, for every download of information being relevant for real life organisms, the DNA content has to be translated into RNA sequence again to be available for active RNA group behavior, i.e., RNA mediated sequence modulation, and re-arrangements or modifications of any kind. Not to forget that the ribosome, the most abundant cellular RNA species, have evolved as the catalytic, organizational regulatory hub of protein biosynthesis in all cells. Two subunits together decode mRNA and additionally synthesize corresponding peptide chains.¹⁴

Another interesting aspect is that infectious/invasive non-coding RNAs insert preferentially in non-coding DNA areas, whereas coding DNA usually is not the target of invading RNAs. This indicates that the insertion competence relates to interaction competence. In this perspective the non-coding DNA is the preferred habitat to settle down by infectious RNAs, e.g., γ -chromosome in human genomes,¹⁵ whereas the coding regions are not preferentially targeted.^{16,17}

If an invading RNA species meets a host genome it will sense repetitive sequences much more easier for insertion than non-repetitive ones. If we look at a host repeat sequence, e.g., AAAACCCC and an invading stem loop has e.g., UUGG, there would be some binding affinity on 4 bases, which will be not the case in a non-repetitive sequence. This may indicate that the preferred change in evolutionary processes occurs in regulatory sections and not in the information storage coding for proteins.

Meaning of genetic information depends on group interactions

RNAs remain the most ancient biological agents that connect information with meaning, in that they interact, based on their nucleic acid syntax that binds to complementary acids of the genetic alphabet. In this way, they create information bearing molecules. However, the meaning of these molecules does not depend on syntax but on pragmatics (context), i.e., its real life function.¹⁸⁻

²⁰ The meaning is determined by the context of the interacting participants which means the function that the information represents for the interacting agents.^{9,21} In this way, coded information may carry meaning for e.g., toxic components that kill invaded hosts. But the same information may carry meaning for immune functions (Antitoxins) to protect the host.²² A variety of meanings may be coded in information in the genetic syntax, but the living agents that are involved in real-life contexts are crucial for meaning to emerge from the information they represent. They generate meaning out of genetic information storage medium by real-life world interactions with themselves or non-themselves (hosts) as documented in the recent knowledge about epigenetics.²³ Without living agents, no information out of DNA would find its way to some meaning.

In the ancient RNA world, with building RNA stem loop consortia, a biological behavior emerged that underlies a biological selection which is absent in a pure physico-chemical world.^{24,25} The start of biological selection was the result of group behavior of RNA stem loops.^{26,27} After abundance of selection processes and compartmentalizations, we may agree that the first biological cells as organisms emerged.²⁸ We should keep in mind, that all functions of these organisms that enabled them to replicate, metabolize, defeat against invading or parasitizing living agents are regulated by the obligate RNA-based competencies in transcription, translation, repair, and immunity. Hence, the basic features of cellular life are processed by a DNA based organism in which all functions still depend on an orchestrated line up of coupled RNA functions.²⁹

If we look at the current stage of life history, we have a biological system in that cellular life based on DNA storage medium, and RNA regulatory networks dominate the planet.³⁰ RNA groups still play important roles in catalyzing biochemical reactions, the translation of mRNAs into proteins, the fine-tuned regulation of gene expression, and the steps and sub-steps of immune defense in the various ways of recognition of “non-self”-integration of identification tools into the self-protecting system to recognize invading identities and the targeting of defense tools on these invading agents.³¹

The coupling of the information and the contextual regulation of this information was the coupling of introns and exons, where the introns are representing the RNA repeat sequences within protein coding sections (non-repeat sequences). Interestingly, nature created a way to line up complex protein bodies in building a final non-repeat sequence of protein coding sections of the genetic text that serves as template; and splicing out the intronic sequences being relevant for regulatory functions.³²⁻³⁵

Important aspects of natural genetic editors

Since the early RNA world, maybe before the evolution of DNA-based or even RNA-based protein bodies, the RNAs not only catalyze biochemical processes but also bind to proteins in various forms and modes so that proteins stabilized the RNA structures in a certain way for its specialized functions.^{36,37} Protein itself has very limited possibilities to transmit information (such as prions). A crucial key for protein synthesis is the ribosome with both its RNA based ribozymatic functions and its protein body to stabilize and interconnect the 2 complex and complicated subunits of RNA consortia. Without ribosome, no protein body would exist.³⁸

- It does not really interest whether the RNA world was predated by another primordial system, the facts outline RNA to be the only successful and essential natural code that provides the preconditions for RNA replication and DNA-based cellular life.³⁹
- Besides the ribosome that translates RNA exon line up into protein code after splicing out the regulatory introns, RNAs transport information as a messenger in mRNA and the “readymade” transcript out of DNA storage medium to an RNA template. Nevertheless, whereas 90% of the e.g., human genome are transcribed into RNAs only 1.5% serve as intron free template for protein production.^{40,41}
- Interestingly, RNAs can bind to small metabolites to use the binding energy to switch from one RNA structure to another. Additionally, very small RNAs act as ribozymes that may self-cleave. Larger

ribozymes can act in complex RNA splicing interactions such as group II introns.^{42,43}

- More often, RNAs work in concert with proteins, such as RNA splicing, RNA editing. We must not forget that a primary transcript involves approximately 200 proteins together with at least 5 small nuclear RNAs. Approximately 95% of say, the human genes, underlie splicing activities.⁴⁴
- Another important complementary interaction tool is the telomerase a certain form of reverse transcriptase being an essential function for creating DNA sequences in general, or how should we think of the emergence of DNA from a primordial RNA world?⁴⁵
- One interesting action on how RNAs act on each other is the antisense RNA action to inhibit RNA protein interactions by complementary base pairing to RNAs.^{46,47}
- Another important function is the regulation of the translatability of mRNAs by small double stranded RNAs. This is achieved by the recognition of complementary sequences in the mRNA, which serves as efficient tool also in RNA silencing.^{48,49}
- Additionally, the DNA storage medium condensed in chromatin may be manipulated in various ways by RNAs that attract proteins to modify the chromatin status—more globally or even specialized—to very local interactions affecting single genes or groups of genes. Interestingly, information that identifies invading genomes (such as in CRIPRS) is stored in DNA, but is then converted into small RNAs that recognize and interfere with these invaders. Therefore, the conversion of DNA into RNAs plays a crucial role for inheritance via RNA, and to remain, functions with the conversion process in the RNAs that process the action.³³

Evolutionary results of RNA productivity: DNA and protein bodies

The evolution of protein bodies, i.e., the translation of nucleic acid sequences into amino acid sequences, leads to the emergence of what we called “life,” i.e., the visible structures of organisms built out of cells, tissues, and organs.^{37,50} If we look at the results of this strain of biological evolution, we can assume that no organism can function vitally without RNA- consortia interactions that dominate all regulatory processes within the living cells and cell societies such as tissues and organs.⁵¹

To this dependency on the RNA world agents, the protein bodies also communicate by themselves but at a different level of signaling interactions, relatively independent of RNA world influences. Hence, this level of interactions

(although dependent on RNA world activities at the base of all biologically relevant functions), is somehow independent from RNA-determined functions, because bodies interact within their phenosphere more than on their genosphere interactions which are focused on the bodies to which they relate. The phenosphere interactions are primarily more relevant to behavioral response behaviors between bodies, than the genosphere activities within the body.¹⁸ The feedback of the “living DNA archives” to the current RNA networks constantly fine-tune their regulation and innovation capabilities.

In parallel, these phenosphere interactions feed back to the body regulating RNA- consortia, via impression patterns that may be relevant to epigenetic modifications in, say, stress. In stress situations, similar to stem cells poorly transcriptionally active and epigenetically determined, the silencing of mobile genetic elements is weakened, and the immune functions are opportunistic. We can look at such context-relevant situations in pre-implantation of placenta, trophoblast differentiation, as well as tumor growth.^{52,53} In all these situations, the genetic identity producers that are silenced or conserved as regulatory tools may become virulent again; active for change, de-regulation, and production of novel content arrangement. This may lead not only to genetic novelty, novel gene regulations, and changing dynamics but also to increased rates of disease caused by de-regulation of former counterbalanced networks of regulations, as described in Fig. 1.

If RNA network, i.e., RNA-consortia interactions become relevant, the “force” of life is still activated.⁵⁴ However, in contrast to former narratives, we now can identify this neither as a metaphysical force from a transcendent creator, nor as a result out of mechanistic error replication events by chance, but as an intrinsic capability of biological entities dependent on the RNA agents’ activities.⁵⁵ If something gets out of control, they are still activated. If all functions well without stress, the RNA agents are constantly serving as a mighty network of conserved functions that integrate the wide open horizon of all interactional motifs being existent in a biological species. The main motif is to promote and support life processes and not to destroy them.

Keyplayer in RNA to DNA transformation: Reverse transcriptase

In several attempts, biological features which cannot be found in inanimate nature have been reduced to the chemical level only. In this perspective, biological features are expanded organization systems at the macromolecular level, but in a strict natural law system. Authors were convinced that in this perspective,

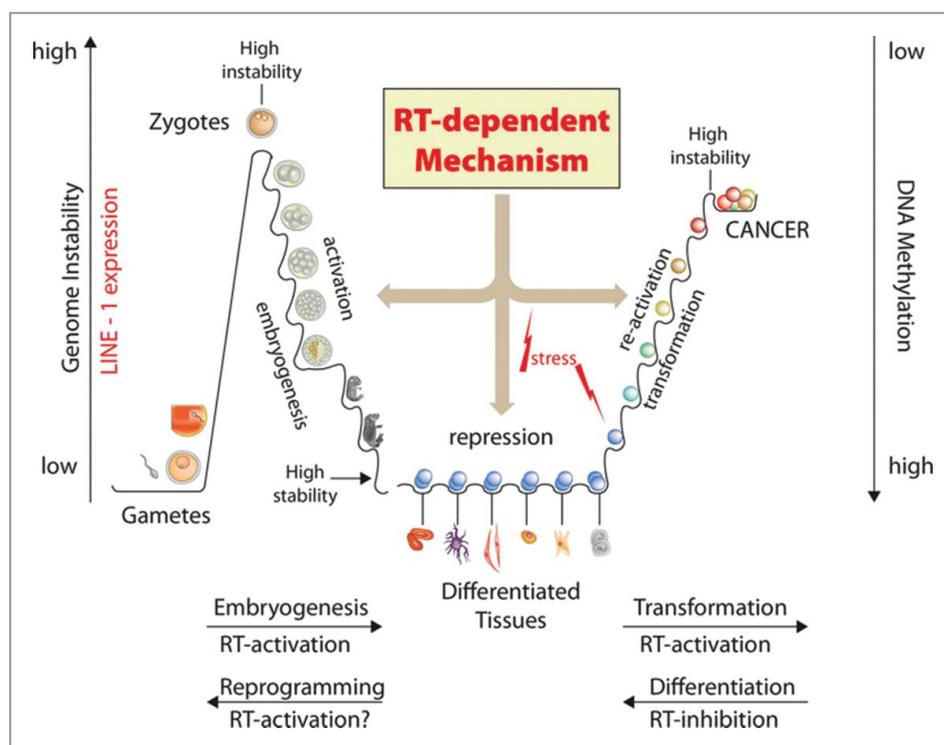


Figure 1. Reverse Transcriptase is the key player to transfer RNA sequences into DNA. This is a counter regulated process we can look at in embryogenesis and cell differentiation. In tumorigenesis, this gets out of control © New York Academy of Sciences. Reproduced by permission of New York Academy of Sciences. Permission to reuse must be obtained from the rightsholder.⁵²

evolution can be integrated into chemical kinetics, a merger of population processuality with molecular biology.^{56,57} Meanwhile, we know that the quantitative explanation of biological sign mediated interactions can identify side effects at its best, but cannot fully explain the main features of social interacting biological agents.⁵⁸

If we transfer this result to the quasispecies behavior at the RNA level, we can identify several RNA group behaviors that represent de novo generation activities that constitute the basis for evolutionary novelty.^{59,60} If we take reproduction only from the physico-chemical perspective, we have to realize that a clonal reproduction can occur only in an identical copy production of the former product, if there is some change in this “error replication” was the term for decades. Nevertheless, as demonstrated by the biocommunication theory—the generation of novelty by agents competent in editing biological codes—this is not an appropriate description.⁶¹ Similar to the generation of poems in a natural language by humans in this reductionistic light would be an error prone reproduction of former available sign sequences into new ones—a rather curious conclusion.

A key agent in several aspects of biology, especially in genetics, is the reverse transcriptase (RT). The reverse transcriptase derived from an RNA world, and correctly termed, is the primary tool to transport information from RNA to DNA before other polymerases.^{45,62,63} RTs can be

identified through the whole current RNA world activities in virions and RNA viruses such as retroviruses, and is the essential tool to place RNA information within DNA-based genomes.^{64,65} In this respect, it is an essential tool to install persistent viral agents in host genomes. The RTs care for virus replication cycles as well as suppressed viral-derived inhabitants that serve as regulatory elements of host genes. In special situations such as stress, the regulation of the host function may become weak and the original function may become virulent again (see Fig. 1). Especially if RTs get into this modus, they may change the genetic identity of host genomes and fix it into the DNA storage medium being part of an inheritable feature.^{66,67}

The important innovative perspective on these processes is, to keep in mind that these regulatory RNAs exist always in dynamic populations that are quite tissue specific. It is this social collective of RNA groups that interprets the (non-repetitive) ‘gene’ language of the coding sequences and provides the specific identity of each cell type.

Conserved repetitive sequences in DNA archives: Centromeres, telomeres

Interestingly, some repetitive sequences are conserved into DNA at important physiologic places of the genomes, such as at the center of the chromosomes or at the end (end protection). Out of centromeres and

telomeres, no transcription occurs, which indicates rather strict conservation and non-coding for proteins' functions.⁶⁸ We do not know why, but it is a fact that the repeat telomere syntax essentially protects chromosome ends from invasion of genetic parasites which degrade or damage chromosome ends.^{69,70} Maybe infection derived group II introns drove evolutionary novelty of linear chromosomes and co-adapted to genome end-protection.⁷¹ The repetitive syntax structure of telomeres seems to make them somehow immune to recombination and rearrangements by invading genetic parasites.

However, initiation of telomere DNA replication occurs frequently at telomere repeats.⁷² For instance, we find TTAGGG in vertebrates, humans, mice, *Xenopus*, filamentous fungi, *Neurospora crassa*, the slime molds *Physarum* and *Didymium*; TTGGGG in *Tetrahymena* and *Glaucoma*; TTGGG(T/G) in *Paramecium*; TTTTGGGG in *Oxytricha*, *Stylonychia* and *Euplotes*; TTAGGG(T/C) in the apicomplexan protozoan *Plasmodium*; TTTAGGG in *Arabidopsis thaliana*; TTTTAGGG in green algae *Chlamydomonas*; TTAGG in the insect *Bombyx mori*; TTAGGC in the roundworm *Ascaris lumbricoides*; TGTGGGTGTGGTG (from RNA template) in *Saccharomyces cerevisiae*; GGGGTCTGGGTGCTG in *Candida glabrata*; and GGTGTACGGATGTCTAACTTCTT in *Candida albicans*.⁷³

The telomerase, a subspecies of the reverse transcriptase family, cares for telomere replication at the ends of chromosomes and, therefore, not only is a key player of genome maintenance but is also part of an immunity function that provides telomere ends from infection by genetic parasites.⁷³ This is a new role of a reverse transcriptase and demonstrates that RNA agents serve for both, infection of host genomes and—as coapted exaptations—as immune function against (related) infectious agents. The evolutionary move from telomeres to centromeres in eukaryotes is an excellent example of re-invention of newly evolved features into another useful function.⁷⁴

Without infection there can be no evolution.⁷⁵ Infectious agents are the driving force to generate genetic novelty, generate immune functions against related parasites and, therefore, serve for protection of host organisms in an ongoing way.⁶³

Three of a perfect pair: Inanimate nature, RNA-consortia/viruses and protein bodies

At the current stage of understanding how evolution occurs, and the emergence of RNA groups, viruses and DNA-based cellular life representing what we call “Life,” we know that cooperativity between these 3 levels is a core behavioral motif.^{24,25,76,77} Single RNA stem loops,

being alone, only react in a physico-chemical reaction modus without any biological selection, whereas if they build groups, they actively evolve self/non-self-differentiation and compete in a biological selection modus.²¹

Viruses evolved in a constant competition of this RNA group selection, to become part of a group self and not to become a part in non-self, whether non-self by itself may represent another group. In the step from RNA consortia to protein-based cell bodies, the viruses played the essential driving role of infection/immunity/identity.¹⁵ The genetic code went from molecules to a semiotic code by group interactions which fits the general thesis that any natural language or code needs groups of competent agents that use (generate, innovate) such codes in social interactions,⁷⁸ whereas a natural code without competent users does nothing.

The evolution of DNA, and later on of cellular protein-based life, cared for the emergence of phenotypes that interact via a different level of signaling. Although based on RNA-coordinated body regulation, the phenotypic interactions of protein-based cellular life forms determine ecosphere habitats. The interactions that constitute a phenosphere of protein bodies depends on the basics of RNA biology as the RNA biology depends on the natural laws of physics and chemistry. However, RNA biology transcends physics and chemistry by (i) social group behavior, (ii) semiotic code biology and the (iii) biological selection, all of them being absent in inanimate nature.

Similarly, the phenosphere of protein bodies transcends RNA biology, because the phenosphere is mostly constituted by cellular signaling motifs of coordination, and organization, i.e., biocommunication.⁶¹ We must differentiate these levels of interaction because they all are relevant in a complementary way to define life in the 21st century: First, you need physics and chemistry. Second, you need cooperative agents that build consortia which underlie biological selection. Life starts here.⁷⁹ RNA consortia lead to DNA storage and cellular protein-based bodies. Thirdly, protein-based bodies interact and constitute the phenosphere of life, as we know it. Although all 3 levels derive and are constituted upon the former, they add a new level/sphere that is not present in the former one and cannot be deduced out of the former one. This means it is a real novelty.

These features are known to us and have been under investigation since several years; some, for decades. What was not clear is that all these RNA agents which all share repetitive nucleotide syntax derive from infectious/invasive agents such as viruses and virus-like agents.^{55,80} RNAs do not derive by chance mutations in error replication events. There is crucial difference to former pictures of how the biological world works. If we identify

RNA that work as key tools in cellular life, we must be aware that this is due to infection events, and additionally its counter effect of competing infectious agents that found a way to counter-regulate themselves via the immune function of the host. The genial concept to explain this was the “addiction module” of Luis Villarreal: It documents the variety of counterbalancing modules each constituted by a variety of RNA stem-loop groups with their volatile genetic identities that may change like “gangs” according actual needs (see Fig. 2) and that we currently know in its host persistence (toxin-antitoxin, restriction-modification, insertion/deletion etc.) and that will be detected in the future in all naturally counter-regulated biological processes.²²

As mentioned above, there is some crucial difference in the motifs and behavioral patterns of RNA consortia and protein-based cellular life. If we look at biocommunication

processes in cells, tissues, organs, and organisms, in which signaling is relevant to all coordination and organization affairs, we can identify similar behavioral patterns throughout all domains of life. Whether it may be bacteria, protozoa, fungi, animals, or plants: All are highly sensitive organisms that actively compete for environmental resources. They assess their surroundings, estimate how much energy they need for particular goals, and then realize the optimum variant. They take measures to control certain environmental resources. They perceive themselves and can distinguish between ‘self’ and ‘non-self’. They process and evaluate information and then modify their behavior accordingly.⁸¹⁻⁸⁶

If we look at RNA agents (which all share repetitive nucleotide syntax) that build consortia, we can identify behavioral motifs such as RNA binding, DNA binding, protein-binding, nucleotide assembly and fold back to

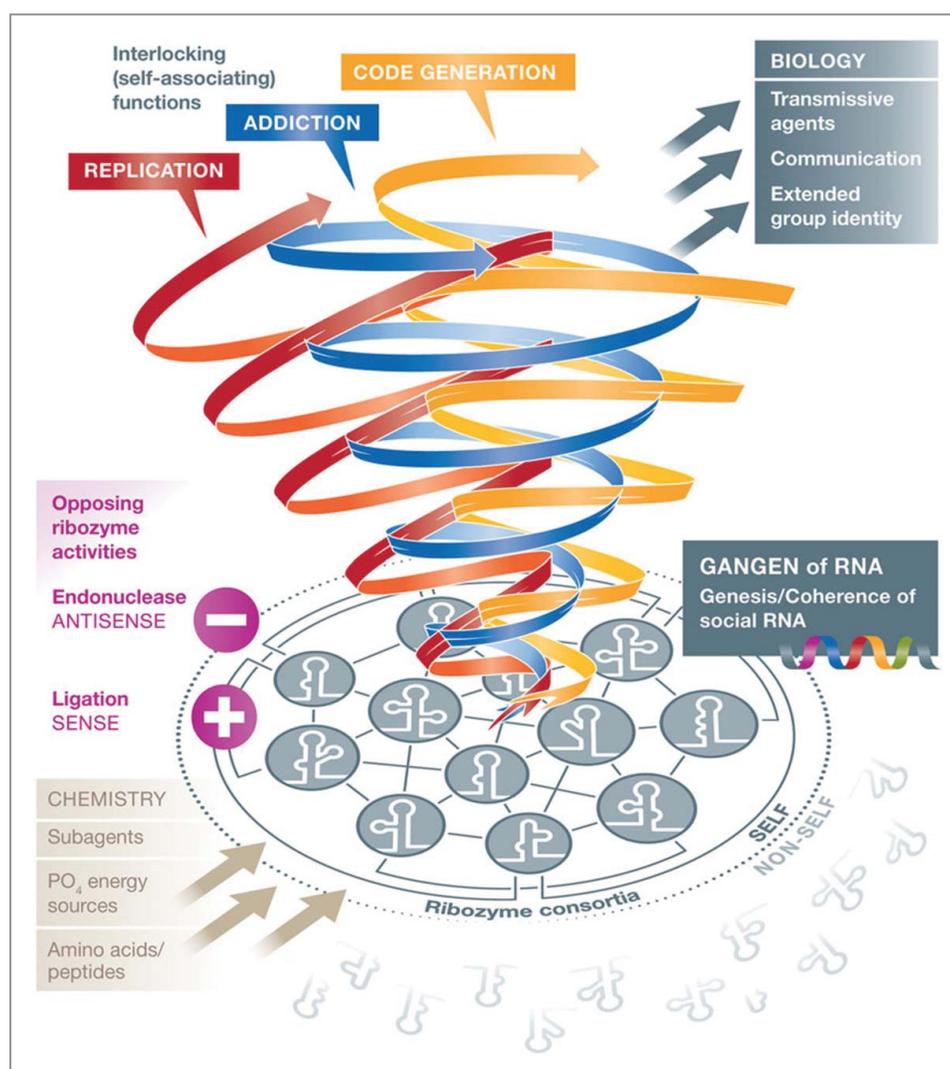


Figure 2. The RNA “gangen” hypothesis of Luis Villarreal coherently explains cooperativity, followed by group identity of an RNA collective that requires opposite functions for the genesis of life © New York Academy of Sciences. Reproduced by permission of New York Academy of Sciences. Permission to reuse must be obtained from the rightsholder.⁵⁵

build stem loops with single strand motifs and double strands, pseudoknots, and spontaneous loop-building out of stems being prone to other (single stranded) loop interactions.⁸⁷⁻⁹¹ However, some common behavioral motifs also are shared, like identification of “self” and “non-self,” invasive behavior that manipulates another, rejection of invasive behavior as the commonly shared immune functions, group building, etc.⁹²

Ribozymes, introns, mobile genetic elements, viruses and their defectives, viroids, maybe we can look at viruses and virus-like agents now as first full range semi-otic subjects, competent to follow interactional, behavioral rules in infection, colonization, immunity, recombination, context-dependent meaning fixation (methylation patterns) and finally sequence space innovation. Prior to protein-based cellular life viruses and virus-like genetic parasites seem to represent the evolutionary transition agents from the early (repetitive syntax constituted) RNA world to the living DNA archives.⁹²⁻⁹⁷

Nothing in RNA-biology makes sense except in the light of group-identity

To constitute RNA group identity, they must be able to assemble, to divide (nucleases), and to re-assemble (ligases) within new sequence orders in a rather fast and strictly ruled way. As mentioned before, they are able to build groups according to actual needs, just like “gangs” with their goals that may change within the next moment. Luis Villarreal outlined this within his “gangen” thesis:⁵⁵

From this perspective, some explanatory goals can be reached more easily than from previous ones, because the biological identity is in the main focus now:

- 1) Evolutionary novelty arises in and by groups of infectious agents that are competent to interact within a common set of rules of combination, rejection/defense, and cooperation. This includes the later integration of former rejected agents, the integration of recycled parts, and the subunits of former degraded nucleotide sequences of such agents. The re-use of RNAs in several lengths for several functions reminds us on their origin.
- 2) Evolutionary novelty also arises through exaptations of original novelties to other organisms as previously evolved; e.g., tRNAs, or the epigenetic marking of genomes led to memory and learning systems, or the pheromone receptors, as well as light receptors or the telomere repeats or centromeres.
- 3) The RNA world dominant interaction strategies of colonization via infection, defense/rejection and degradation and ligation competence need a new

matrix for more complexity: Protein-based bodies with memory storage medium DNA, i.e., RNA world escape islands. Here, competing and co-surgical agents with different strategies may reach the addiction module status, which means they are not eliminating each other but both survive in an instabil/stabil equilibrium status for both, the competing agents and the host (win-win situation). This makes the infectious agents become crucial drivers of the evolution of complexity.

- 4) Coherent sequence integration, without damage of protein coding sequence, may be transported by the LTRs of infectious agents. Later on, these tools are helpful in a real toolbox of similar and combinatorial modules for both infectious agents and host organisms.
- 5) Real revolutionary new organisms that represent success exemplars of new kingdoms such as in the case of sharks, *C. elegans*, glass sponges, sea urchins, ferns, etc. are nearly free of disease-causing viral parasites. This means the uniqueness of addiction modules as a result of persistent colonization within these new creations have reached a kind of hallmark immunity.
- 6) Such evolutionary constructions may reach highly complex tissue societies, as represented in organs of various organisms that vary/combine with fluid organs such as blood systems, together with other body secretion systems creating a real fascinating construction to embody complexity.
- 7) Change in environmental conditions (context) may change the invasive strategies of infectious genetic parasites and also vary the potential targets for invasion. e.g., exogenous retroviruses rise selective pressure on HERV env's which results in an increase of the neocortex tissue by the reprogramming of neuronal stem cells. This rather organic emergence pattern has a clear result on the phenotypes as we currently may recognize at ourselves!
- 8) The virus competence is rather ecological, i.e., tissue specific or better: cell-community-identity specific. If there are cell-communities as represented by tissues (e.g., organs) or fluids (e.g., blood) the virus and virus derived (defectives) inhabitants build a rather complex network of interactions for replication regulations, immune functions, repair competence, and also - if necessary - evolutionary innovations or its pre-steps to adapt.

In this perspective, nothing in RNA-biology makes sense, except in the light of group identity: How RNA groups (which all share repetitive nucleotide syntax) constitute genetic content of themselves being relevant for host genetic content. To belong or not to belong to a

group decides, whether an individual is part of an interactional group that share common capabilities that underlie biological selection, share interactional motifs, and attack and defense strategies. Any evolutionary relevant result can be seen within this light.

Conclusions

The repetitive non-coding RNA sequences resemble that of a natural everyday language, i.e., the essential tool to coordinate and organize common behavior. The non-repetitive coding DNA sequences, after being modified by the editosome and the spliceosome, serve as exon line-up for artificial constructions within an information storage medium for proteins that build the whole world of cellular-based life. Whereas the non-coding RNAs which all share repetitive nucleotide syntax build an abundance of groups that are active in nearly all biological processes, the DNA serves as a living archive. The step from pure physical-chemical molecules to “meaning” of genetic information depends on interacting RNA-groups with certain identities which underlie biological selection. We can find such RNA-based group-identities in viruses and virus-like infective agents that constantly invade DNA genomes as preferred habitat via a counterbalanced persistence strategy—the addiction module. For DNA based organisms an abundance of invading genetic parasites, infective clouds of RNA groups and their competitive counter-groups are the driving force of evolution and later on as co-adapted persistent agents serve as regulatory tools in nearly all cellular processes. Additionally, RNA-groups constantly produce new sequence space, which not only serve as adaptation tools for their cell-based host organisms but also plays crucial roles in evolutionary novelty.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

References

- [1] Eigen M, Winkler R. *The Laws of the Game: How the Principles of Nature Govern Chance*. Princeton: Princeton University Press; 1981.
- [2] Schuster, P. *Evolution in an RNA World*. In: Ord MG, Stocken LA (eds). *Foundations of Modern Biochemistry*. Stamford and London: Jai Press; 1998. 159-198 p.
- [3] Witzany G. *Language and Communication as Universal Requirements for Life*. In: Kolb V (ed). *Astrobiology: An Evolutionary Approach*. Boca Raton: CRC Press; 2014. 349-369 p.
- [4] Atkins JF, Gesteland RF, Cech TR, (eds). *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*. New York: Cold Spring Harbor Laboratory Press. 2010.
- [5] Cech TR, Steitz JA. The noncoding RNA revolution - trashing old rules to forge new ones. *Cell* 2014; 157:77-94; <http://dx.doi.org/10.1016/j.cell.2014.03.008>; PMID:24679528; <http://dx.doi.org/10.1016/j.cell.2014.03.008>
- [6] Shapiro JA, von Sternberg R. Why repetitive DNA is essential to genome function. *Biol Rev Camb Philos Soc* 2005; 80:227-250; PMID:15921050; <http://dx.doi.org/10.1017/S1464793104006657>
- [7] Jurka J, Kapitanov VV, Kohany O, Jurka MV. Repetitive sequences in complex genomes: Structure and evolution. *Annu Rev Genomics Hum Genet* 2007; 8:241-59; PMID:17506661; <http://dx.doi.org/10.1146/annurev.genom.8.080706.092416>
- [8] Witzany G. The agents of natural genome editing. *J Mol Cell Biol* 2011; 3:181-189; PMID:21459884; <http://dx.doi.org/10.1093/jmcb/mjr005>
- [9] von Sternberg R, Shapiro JA. How repeated retroelements format genome function. *Cytogenet Genome Res* 2005; 110:108-116; PMID:16093662; <http://dx.doi.org/10.1159/000084942>
- [10] Cech TR. The RNA worlds in context. *Cold Spring Harb Perspect Biol* 2012; 4(7):a006742; PMID:21441585; <http://dx.doi.org/10.1101/cshperspect.a006742>
- [11] Roychoudhury M, Sitlani A, Lapham J, Crothers DM. Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proc Natl Acad Sci* 2000; 97:13608-13613; <http://www.pnas.org/content/97/25/13608.long>; PMID:11095739; <http://dx.doi.org/10.1073/pnas.250476297>
- [12] Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005; 3:22-732; PMID:16138100; <http://dx.doi.org/10.1038/nrmicro1235>
- [13] Shapiro JA. Revisiting the central dogma in the 21st century. *Ann NY Acad Sci* 2009; 1178:6-28; PMID:19845625; <http://dx.doi.org/10.1111/j.1749-6632.2009.04990.x>
- [14] Doudna JA, Rath VL. Structure and function of the eukaryotic ribosome: the next frontier. *Cell* 2002; 109:153-156; PMID:12007402; [http://dx.doi.org/10.1016/S0092-8674\(02\)00725-0](http://dx.doi.org/10.1016/S0092-8674(02)00725-0)
- [15] Villarreal LP. *The origin of group identity. Viruses, addiction and cooperation*. New York: Springer; 2009.
- [16] Bushman FD. Targeting survival: Integration site selection by retroviruses and LTR-retrotransposons. *Cell* 2003; 115:135-138; PMID:14567911; [http://dx.doi.org/10.1016/S0092-8674\(03\)00760-8](http://dx.doi.org/10.1016/S0092-8674(03)00760-8)
- [17] Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2004; 2(8):E234; PMID:15314653; <http://dx.doi.org/10.1371/journal.pbio.0020234>
- [18] Witzany G. *Biocommunication and Natural Genome Editing*. Dordrecht: Springer; 2010.
- [19] Maizels N, Weiner AM. Phylogeny from function: Evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc Natl Acad Sci USA* 1994; 91:6729-6734; PMID:8041690; <http://dx.doi.org/10.1073/pnas.91.15.6729>
- [20] Doudna JA, Cormack BP, Szostak JW. RNA structure, not sequence, determines the 5' splice site specificity of a

- group I intron. *Proc Natl Acad Sci USA* 1989; 86:7402-7406; PMID:2678103; <http://dx.doi.org/10.1073/pnas.86.19.7402>
- [21] Witzany G. Crucial steps to life: From chemical reactions to code using agents. *Biosystems* 2016; 140:49-57; PMID:26723230; <http://dx.doi.org/10.1016/j.biosystems.2015.12.007>
- [22] Villarreal LP. The addiction module as a social force. In: Witzany G (ed). *Viruses: Essential Agents of Life*. Dordrecht: Springer; 2012. 107-45 p.
- [23] Mattick J. Deconstructing the dogma. a new view of the evolution and genetic programming of complex organisms. *Ann N Y Acad Sci* 2009; 1178:29-46; PMID:19845626; <http://dx.doi.org/10.1111/j.1749-6632.2009.04991.x>
- [24] Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. Spontaneous network formation among cooperative RNA replicators. *Nature* 2012; 491:72-77; PMID:23075853; <http://dx.doi.org/10.1038/nature11549>
- [25] Vaidya N, Walker SI, Lehman N. Recycling of informational units leads to selection of replicators in a prebiotic soup. *Chem Biol* 2013; 20:241-52; PMID:23438753; <http://dx.doi.org/10.1016/j.chembiol.2013.01.007>
- [26] Hayden EJ, Lehman N. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem Biol* 2006; 13:909-918; PMID:16931340; <http://dx.doi.org/10.1016/j.chembiol.2006.06.014>
- [27] Higgs PG, Lehman N. The RNA world: molecular cooperation at the origins of life. *Nat Rev Genet* 2015; 16:7-17; PMID:25385129; <http://dx.doi.org/10.1038/nrg3841>
- [28] Koonin EV. Viruses and mobile elements as drivers of evolutionary transitions. *Philos Trans R Soc Lond B Biol Sci* 2016; 371(1701) pii: 20150442; PMID:27431520; <http://dx.doi.org/10.1098/rstb.2015.0442>
- [29] Koonin EV. On the origin of cells and viruses: primordial virus world scenario. In: Witzany G (ed). *Natural Genetic Engineering and Natural Genome Editing*. Hoboken: John Wiley & Sons; 2010. 47-64 p.
- [30] Lehman N. The RNA world: 4,000,000,050 years old. *Life* 2015; 5:1583-86; []; PMID:26791312; <http://dx.doi.org/10.3390/life5041583>
- [31] Villarreal LP. Viruses and host evolution: virus-mediated self identity. In: Lopez-Larrea C (ed). *Self and Non-Self*. Austin: Landes Bioscience and Springer Science Business Media; 2011. 185-217 p.
- [32] Witzany G., Noncoding RNAs: persistent viral agents as modular tools for cellular needs. *Ann NY Acad Sci* 2009; 1178:244-67; PMID:19845641; <http://dx.doi.org/10.1111/j.1749-6632.2009.04989.x>
- [33] Wang X, Song X, Glass CK. The long arm of long non-coding RNAs: Roles as sensors regulating gene transcriptional programs. *Cold Spring Harb Perspect Biol* 2011; 3: a003756; PMID:20573714; <http://cshperspectives.cshlp.org/content/3/1/a003756.long>
- [34] Yang L. Splicing noncoding RNAs from the inside out. *Wiley Interdiscip Rev RNA* 2015; 6:651-660; PMID:26424453; <http://dx.doi.org/10.1002/wrna.1307>
- [35] Wilusz J. Repetitive elements regulate circular RNA biogenesis. *Mob Genet Elements* 2015; 5:1-7; PMID:26442181; <http://dx.doi.org/10.1080/2159256X.2015.1045682>
- [36] Moore PB, Steitz TA. The roles of RNA in the synthesis of protein. *Cold Spring Harb Perspect Biol* 2011; 3(11): a003780; PMID:21068149; <http://dx.doi.org/10.1101/cshperspect.a003780>
- [37] Noller H. Evolution of protein synthesis from an RNA world. *Cold Spring Harb Perspect Biol* 2012; 4(4): a003681; PMID:20610545;
- [38] Root-Bernstein M, Root-Bernstein R. The ribosome as a missing link in the evolution of life. *J Theor Biol* 2015; 367:130-158; PMID:25500179; <http://dx.doi.org/10.1016/j.jtbi.2014.11.025>
- [39] Sengupta S, Higgs PG. Pathways of genetic code evolution in ancient and modern organisms. *J Mol Evol* 2015; 80:229-243; PMID:26054480; <http://dx.doi.org/10.1007/s00239-015-9686-8>
- [40] Rederstorff M, Bernhart SH, Tanzer A, Zywicki M, Perfler K, Lukasser M, Hofacker IL, Hüttenhofer A. RNomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleoprotein particles. *Nucleic Acids Res* 2010; 38(10):e113; PMID:20150415; <http://dx.doi.org/10.1093/nar/gkq057>
- [41] Rissland OS. The organization and regulation of mRNA-protein complexes. *Wiley Interdiscip Rev RNA* 2016; June 21; PMID:27324829;
- [42] Lambowitz AM, Belfort M. Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol Spectr* 2015; 3: MDNA3-0050-2014; PMID:26104554; <http://dx.doi.org/10.1128/microbiolspec.MDNA3-0050-2014>
- [43] Lambowitz AM, Zimmerly S. Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* 2011; 3(8):a003616; PMID:20463000; <http://dx.doi.org/10.1101/cshperspect.a003616>
- [44] Giono LE, Moreno NN, Botto AEC, Dujardin G, Muñoz MJ, Kornblihtt AR. The RNA Response to DNA Damage. *J Mol Biol* 2016; 428:2636-2651; PMID:26979557; <http://dx.doi.org/10.1016/j.jmb.2016.03.004>
- [45] Blackburn EH, Collins K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol* 2011; 3(5):pii: a003558; PMID:20660025; <http://dx.doi.org/10.1101/cshperspect.a003558>
- [46] Gottesman S, Storz G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* 2011; 3(12):pii: a003798; PMID:20980440; <http://dx.doi.org/10.1101/cshperspect.a003798>
- [47] Updegrove TB, Shabalina SA, Storz G. How do base-pairing small RNAs evolve? *FEMS Microbiol Rev* 2015; 39:379-391; PMID:25934120; <http://dx.doi.org/10.1093/femsre/fuv014>
- [48] Joshua-Tor L, Hannon GJ. Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb Perspect Bio.* 2011; 3(10):a003772; PMID:20810548;
- [49] Ipsaro JJ, Joshua-Tor L. From guide to target: molecular insights into eukaryotic RNA-interference machinery. *Nat Struct Mol Biol* 2015; 22:20-28; PMID:25565029; <http://dx.doi.org/10.1038/nsmb.2931>
- [50] Noller HF. The driving force for molecular evolution of translation. *RNA* 2004; 10:1833-1837; PMID:15547132; <http://dx.doi.org/10.1261/rna.7142404>
- [51] Forterre P. To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life.

- Stud Hist Philos Biol Biomed Sci 2016; 59:100-8 pii: S1369-8486(16)30008-5; PMID:26996409;
- [52] Spadafora C. A LINE-1-encoded reverse transcriptase-dependent regulatory mechanism is active in embryogenesis and tumorigenesis. *Ann NY Acad Sci* 2015; 1341:164-71; PMID:25586649; <http://dx.doi.org/10.1111/nyas.12637>
- [53] Villarreal LP. Viruses and the placenta: the essential virus first view. *APMIS* 2016; 124:20-30; PMID:26818259; <http://dx.doi.org/10.1111/apm.12485>
- [54] Gomez J, Ariza Mateos A, Cacho I. Virus is a signal for the host cell. *Biosemiotics* 2015; 8:483-49; <http://link.springer.com/article/10.1007/s12304-015-9245-0>; PMID:26640606; <http://dx.doi.org/10.1007/s12304-015-9245-0>
- [55] Villarreal LP. Force for ancient and recent life: viral and stem-loop RNA consortia promote life. *Ann NY Acad Sci* 2015; 1341:25-34; PMID:25376951; <http://dx.doi.org/10.1111/nyas.12565>
- [56] Eigen M, Winkler R. *The Laws of the Game: How the Principles of Nature Govern Chance*. Princeton: Princeton University Press; 1981.
- [57] Domingo E, Schuster P. (eds). *Quasispecies: From Theory to Experimental Systems*. Switzerland: Springer; 2016
- [58] Witzany G, Baluska F. Life's code script does not code itself. The machine metaphor for living organisms is outdated. *EMBO Rep* 2012; 13:1054-1056; PMID:23146891;
- [59] Villarreal LP, Witzany G. Rethinking quasispecies theory: from fittest type to cooperative consortia. *World J Biol Chem* 2013; 4:79-90; PMID:24340131
- [60] Villarreal LP, Witzany G. When competing viruses unify: evolution, conservation, and plasticity of genetic identities. *J Mol Evol* 2015; 80:305-18; PMID:26014757; <http://dx.doi.org/10.1007/s00239-015-9683-y>
- [61] Witzany G. Pragmatic turn in biology: from biological molecules to genetic content operators. *World J Biol Chem* 2014; 5:279-85; PMID:25225596; <http://dx.doi.org/10.4331/wjbc.v5.i3.279>
- [62] Moelling K, Broecker F. The reverse transcriptase-RNase H: from viruses to antiviral defense. *Ann NY Acad Sci* 2015; 1341:126-135; PMID:25703292; <http://dx.doi.org/10.1111/nyas.12668>
- [63] Villarreal LP. *Viruses and the evolution of Life*. Washington: ASM Press; 2005.
- [64] Qu G, Kaushal PS, Wang J, Shigematsu H, Piazza CL, Agrawal RK, Belfort M, Wang HW. Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol* 2016; 23:549-557; PMID:27136327; <http://dx.doi.org/10.1038/nsmb.3220>
- [65] Podlevsky JD, Chen JJ. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol* 2016; 30:1-13; PMID:27359343;
- [66] Sciamanna I, Vitullo P, Curatolo A, Spadafora C. Retrotransposons, reverse transcriptase and the genesis of new genetic information. *Gene* 2009; 448:180-186; PMID:19631262; <http://dx.doi.org/10.1016/j.gene.2009.07.011>
- [67] Spadafora C. Soma to germline inheritance of extrachromosomal genetic information via a LINE-1 reverse transcriptase-based mechanism. *Bioessays* 2016; 38:726-733; PMID:27315018; <http://dx.doi.org/10.1002/bies.201500197>
- [68] Deng Z, Wang Z, Lieberman PM. Telomeres and viruses: common themes of genome maintenance. *Front Oncol* 2012; 2:201; PMID:23293769; <http://dx.doi.org/10.3389/fonc.2012.00201>
- [69] Nakamura TM, Cech TR. Reversing time: origin of telomerase. *Cell* 1998; 92:587-90; PMID:9506510; [http://dx.doi.org/10.1016/S0092-8674\(00\)81123-X](http://dx.doi.org/10.1016/S0092-8674(00)81123-X)
- [70] Nosek J, Kosa P, Tomaska L. On the origin of telomeres: a glimpse at the pre-telomerase world. *Bioessays* 2006; 28:182-90; PMID:16435294; <http://dx.doi.org/10.1002/bies.20355>
- [71] de Lange T. A loopy view of telomere evolution. *Front Genet* 2015; 6:321; PMID:26539211; <http://dx.doi.org/10.3389/fgene.2015.00321>
- [72] Heyer WD. Regulation of recombination and genomic maintenance. *Cold Spring Harb Perspect Biol* 2015; 7(8): a016501; PMID:26238353; <http://dx.doi.org/10.1101/cshperspect.a016501>
- [73] Witzany G. The viral origins of telomeres and telomerases and their important role in eukaryogenesis and genome maintenance. *Biosemiotics* 2008; 1:191-206. <http://link.springer.com/article/10.1007/s12304-008-9018-0>
- [74] Villasante A, Abad JP, Méndez-Lago M. Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc Natl Acad Sci U S A* 2007; 104:10542-10547; PMID:17557836; <http://dx.doi.org/10.1073/pnas.0703808104>
- [75] Baluska F. Cell-cell channels, viruses, and evolution. Via infection, parasitism, and symbiosis toward higher levels of biological complexity. *Ann NY Acad Sci* 2009; 1178:106-119; PMID:19845631; <http://dx.doi.org/10.1111/j.1749-6632.2009.04995.x>
- [76] Müller S. Engineering of ribozymes with useful activities in the ancient RNA world. *Ann NY Acad Sci* 2015; 1341:54-60; PMID:25694225; <http://dx.doi.org/10.1111/nyas.12695>
- [77] Briones C, Stich M, Manrubia SC. The dawn of the RNA World: toward functional complexity through ligation of random RNA oligomers. *RNA* 2009; 15:743-749; PMID:19318464; <http://dx.doi.org/10.1261/rna.1488609>
- [78] Witzany G. From molecular entities to competent agents: Viral infection-derived consortia act as natural genetic engineers. In: Witzany G (ed). *Viruses: Essential Agents of Life*; Dordrecht: Springer; 2014. 407-419 p.
- [79] Witzany G. The biocommunication method: On the road to an integrative biology. *Commun Integr Bio* 2016; 9(2):e1164374; <http://dx.doi.org/10.1080/19420889.2016.1164374>
- [80] Seligmann H, Raoult D. Unifying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses. *Curr Opin Microbiol* 2016; 31:1-8; PMID:26716728; <http://dx.doi.org/10.1016/j.mib.2015.11.004>
- [81] Witzany G. Life is physics and chemistry and communication. *Ann NY Acad Sci* 2015; 1341:1-9; PMID:25557438; <http://dx.doi.org/10.1111/nyas.12570>
- [82] Witzany G (ed). *Biocommunication in Soil Microorganisms*. Heidelberg: Springer; 2011.
- [83] Witzany G, Baluska F (eds). *Biocommunication of Plants*. Heidelberg: Springer; 2012.
- [84] Witzany G (ed). *Biocommunication of Fungi*. Dordrecht: Springer; 2012

- [85] Witzany G (ed). *Biocommunication of Animals*. Springer; 2014.
- [86] Witzany G, Nowacki M (eds). *Biocommunication of Ciliates*. Dordrecht: Springer; 2016.
- [87] Smit S, Yarus M, Knight R. Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. *RNA* 2006; 12:1-14; PMID:16373489; <http://dx.doi.org/10.1261/rna.2183806>
- [88] Mercer, T.R; Mattick, J.S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Bio*. 2013; 20:300-307; <http://dx.doi.org/10.1038/nsmb.2480>
- [89] Gwiazda S, Salomon K, Appel B, Müller S. RNA self-ligation: from oligonucleotides to full length ribozymes. *Biochimie* 2012; 94:1457-1463; PMID:22465106; <http://dx.doi.org/10.1016/j.biochi.2012.03.015>
- [90] Müller S, Appel B, Balke D, Hieronymus R, Nübel C. Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? *F1000Res* 2016; 5. pii: F1000 Faculty Rev-1511; PMID:27408700; <http://dx.doi.org/10.12688/f1000research.8601.1>
- [91] Pressman A, Blanco C, Chen IA. The RNA World as a Model System to Study the Origin of Life. *Curr Biol* 2015; 25:R953-963; PMID:26439358; <http://dx.doi.org/10.1016/j.cub.2015.06.016>
- [92] Villarreal LP. Persistent virus and addiction modules: an engine of symbiosis. *Curr Opin Microbio* 2016; 31:70-79; <http://dx.doi.org/10.1016/j.mib.2016.03.005>
- [93] Forterre P, Prangishvili D. The major role of viruses in cellular evolution: facts and hypotheses. *Curr Opin Virol* 2013; 3:558-565; PMID:23870799; <http://dx.doi.org/10.1016/j.coviro.2013.06.013>
- [94] Villarreal LP, Witzany G. Viruses are essential agents within the stem and roots of the tree of life. *J Theor Biol* 2010; 262:698-710; PMID:19833132; <http://dx.doi.org/10.1016/j.jtbi.2009.10.014>
- [95] Krupovic M, Koonin EV. Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems. *Curr Opin Microbiol* 2016; 31:25-33; PMID:26836982; <http://dx.doi.org/10.1016/j.mib.2016.01.006>
- [96] Stedman K. Mechanisms for RNA capture by ssDNA viruses: grand theft RNA. *J Mol Evol* 2013; 76:359-364; PMID:23784142; <http://dx.doi.org/10.1007/s00239-013-9569-9>
- [97] Stedman KM. Deep Recombination: RNA and ssDNA Virus Genes in DNA Virus and Host Genomes. *Annu Rev Virol* 2015; 2:203-217; PMID:26958913; <http://dx.doi.org/10.1146/annurev-virology-100114-055127>