Chapter 13

Evolution of Genetic Information without Error Replication

Guenther Witzany
Telos-Philosophische Praxis, Austria
witzany@sbg.at

Darwinian evolutionary theory has two key terms, variations and biological selection, which finally lead to survival of the fittest variant. With the rise of molecular genetics, variations were explained as results of error replications out of the genetic master templates. For more than half a century, it has been accepted that new genetic information is mostly derived from random error-based events. But the error replication narrative has problems explaining the sudden emergence of new species, new phenotypic traits, and genome innovations as a sudden single event. Meanwhile, it is recognized that errors cannot explain the evolution of genetic information, genetic novelty, and complexity. Now, empirical evidence establishes the crucial role of non-random genetic content editors, such as viruses, diversity generating retroelements, and other RNA networks, to produce new genetic information, complex regulatory control, inheritance vectors, genetic identity, immunity, new sequence space, evolution of complex organisms, and evolutionary transitions.

1. Introduction

In the decades where physics and chemistry dominated biology, especially molecular biology and genetics, it was a common assumption that evolutionary novelty is a result of cumulation of beneficial replication errors ("mutations"), i.e., changes and variations of a master sequence structure.

The new synthesis in biology was the mainstream evolution narrative, in that genetic novelties and genomic variations could only be viewed as errors of previous original sequences. Although it was known early that all of the empirically measured and documented mutations cause deformed variants, most of them leading to disease or even death, very few seemed to represent neutral or even better variants. It was assumed that, in the long run of evolutionary times, such beneficial variants became an object to biological selection processes leading to better inheritable variants.

The lasting, dominant, new synthesis was constructed at a time before new and abundant data from (a) RNA biology, (b) comparative genomics, (c) phylogenetic analyses, and the (d) comeback of virology were available. Additionally, core paradigms such as (1) one gene–one protein concept, (2) central dogma of molecular biology (DNA–RNA–Protein), and (3) non-coding DNA representing "junk" were falsified in the meanwhile (Shapiro, 2009; Witzany, 2017). Concerning a coherent update of evolutionary theory, these are good reasons to look at alternatives to the error replication narrative. What are these alternatives? Let's start with a look at crucial deficits of the former "mutation" concept and the theory of quasispecies.

2. The Error Replication Narrative: Quasispecies Mutant Spectra

In a series of articles in the early 1970s, Manfred Eigen and Peter Schuster constructed a formalizable (mathematically based) explanation of how basic features of life and their complexities can be explained solely by physical properties of matter: If certain chemical properties exist on a planet with certain physical conditions, life will start by two self-reproducing macromolecular cycles in a complementary way. First, there are information-carrying nucleic acids building one reproductive cycle. Second, there are functional amino acids building the other cycle. Together they build a catalytic hypercycle, which is the basis of self-reproductivity of life (Eigen, 1971; Eigen and Schuster, 1978).

Both parts can be reconstructed physically in a formalizable depiction. The system of nucleic acids (information) and proteins (function) represents a closed loop, because no function is available without information, and information represents meaning by function. A clear hierarchy of reaction cycles leads to all empirically known features of life. The resulting

evolutionary dynamics is an optimizing process, in that Darwinian selection values the fittest results of error replications (mutations). Mutations with selective advantage are instabilities in this system that can be explained sufficiently by irreversible thermodynamic processes.

A series of such mutations of master copies lead to quasispecies, i.e., mutant distributions of primitive replicating RNA entities. Such dynamic distributions of genomes that share genetic variation, competition, and selection generate fittest types (i.e., master copies) or, in the extreme case, "error tresholds", i.e., too high mutation rates/variations, in which information can't further reproduce in the case of too much mutational load.

Eigen followed the opinion of his time that a natural language as the genetic code is determined by a universal and context-free grammar as outlined in great detail by Noam Chomsky. Such a context-free grammar underlies natural laws strictly because it represents the universal logic of the material reality. The core functions of natural languages or codes are formalizable, predicable, and computable. The only scientific tool needed to depict the material reality is mathematics. Therefore, the molecular genetic code can be investigated and described sufficiently in a formalizable way (Schuster, 2011).

Today, we know that Eigen's concept of natural codes is not coherent with current knowledge on natural language or code use (Witzany, 1995, 2014b, 2015). First, no natural language speaks itself as no natural code codes itself (Witzany and Baluska, 2012; Witzany, 2016). The sequence structures of natural languages or codes are not a statistical, random assembly. Empirical data of all known observations indicate that all the natural languages or codes are the result of competent usage by groups of living agents. Second, natural languages or codes are not the inventions of single biotic agents, but they exclusively emerge in biotic interacting groups. This means, emergence of natural languages or codes is essentially a social event. Additionally, it is important not to forget the following: All known agents using natural code follow three kinds of semiotic rules (not laws), such as combinatorial (syntax), content-related (semantics), and context-dependent (pragmatics) rules. As proved by Charles Morris, if one level of rules is missing, nobody can reliably speak of a natural code or language.

In Eigen's concept, the syntax of the master copy determines the semantics (meaning). This is not coherent with the empirical fact that, in natural languages and codes, pragmatics (context) determine semantics. Pragmatic rules are completely absent in Eigen's concept. Several other

derivatives of mathematical theories of language at that time, such as systems theory, information theory, game theory, and later on, synthetic biology, and even biolinguistics, all share this deficit.

3. Non-coding RNAs are Derived from Invading Genetic Parasites

Current knowledge indicates that DNA is not solely a genetic storage medium but also the main ecological niche for invading RNA species. Could it replace the error replication narrative to more coherently explain genetic novelty and genomic variations?

Endogenous viruses and their defectives, transposons, retrotransposons, long terminal repeats, non-long terminal repeats, long interspersed nuclear elements, short interspersed nuclear elements, ALUs, group I introns, group II introns, phages, and plasmids are the currently investigated examples that use DNA genomes as their preferred live habitat (Witzany, 2015).

Some of these genome-invading agents have been termed as mobile genetic elements and an abundance of such agents has been identified during the last 40 years as obligate inhabitants of all genomes (Frost *et al.*, 2005; Villarreal and Witzany, 2018). It is important to note that all of them share behavioral motifs that cannot be found on abiotic planets: They rely on a natural code; they infect, insert, delete, cut and paste, or copy and paste, and spread within the genome. They change host genetic identities either by insertion/deletion and recombination, or the epigenetic markings of genetic content, co-evolve with the host, and interact in a typically module-like manner (Villarreal 2009a; Shapiro, 2004; Koonin *et al.*, 2015; Koonin and Krupovic, 2017).

Together with non-coding RNAs, these agents shape genome architecture and regulation. In this respect, they are agents of change not only in evolutionary time but also in real time (Mattick, 2009).

Many of the currently known non-coding RNAs are derived from such invading genetic parasites (Villarreal, 2005). Non-coding RNAs interact with DNA, RNA, and proteins and play important roles in nuclear organization, transcription, post-transcriptional modifications (RNA editing, splicing), repair, immunity, and epigenetic modifications (Mattick 2001, 2003; Chuong *et al.*, 2016). Non-coding RNAs transcribe in the sense and antisense directions and are expressed with relation to cell type, subcellular compartment, developmental stage, and environmental stimuli (Zinad *et al.*, 2017). This means their function is context dependent (Witzany, 2009).

Additionally, RNA polymerases overlap in transcriptional contents, which means that each nucleotide can participate in varying transcriptional content arrangements according to varying contexts.

Non-coding RNAs regulate in a varying manner, coordinated or independent, autonomously or functionally interrelated, and can regulate individual genes as well as large genetic networks; they can precisely control the spatiotemporal deployment of genes that are executing neuronal processes with extreme cell specifity. Various classes of non-coding RNAs target each other for post-transcriptional regulation via alternative splicing, polyadenylation, 5′ capping, non-templated modifications, and RNA editing. In particular, RNA editing can transmit environmental information to the epigenome (Yablonovitch *et al.*, 2017). Additionally, it enables neuronal plasticity, which is essential for learning and memory (Qureshi and Mehler, 2012; Mercer and Mattick, 2013).

Non-coding RNAs can undergo nuclear-cytoplasmic, nuclear-mitochondrial, axodendritic trafficking via ribonucleprotein complexes that promote spatiotemporal distribution and function of various combinations of ncRNAs, mRNAs, and RNA-binding proteins (Spadoro and Bredy, 2012). Interestingly, the non-coding RNAs in host genomes are more conserved than the protein-coding sequences (Villarreal, 2004). Does this indicate that they are older in terms of evolution than coding sequences?

Non-coding RNAs are not the result of error replications, but in most cases derived from consortial interactions of RNA groups to stabilize or develop new identity traits embedded in a global network of the current RNA world, embedded also into a global virosphere of incredible density (Tycowski *et al.*, 2015). Infection and counter-defence are constantly changing according to the environmental context of these highly interactive groups that integrate to or preclude new, infectious agents. Let us never forget that 1 ml of sea water contains 1 million bacteria, but ten times more viruses and viral parts. If we take solely the number of phages present on this planet placed side by side, we would get 42 million light-years spanned by 10^{31} phage virions (Rohwer *et al.*, 2014).

4. From Physics to Biology: Emergence of Biological Selection

Single RNA stem loops react strictly according to physical and chemical laws, without any biotic characteristics. If such RNA stem loops built groups which can compete and cooperate, biological selection emerges

(Hayden and Lehman, 2006; Briones *et al.*, 2009; Vaidya *et al.*, 2012; Higgs and Lehman, 2015). How do agents emerge from chemicals to form biological identity and then form groups that learn membership (to belong or not to belong)? Single RNA stem-loop generation occurs by physical, chemical properties solely as demonstrated by natural and randomized RNA experiments. As mentioned above, if stem loops build complex consortia, they initiate social interactions not present in a pure chemical world, that is, biological selection emerges. This designates the crucial step from inanimate world to life.

In contrast to inanimate nature, RNA stem-loop groups actively generate behavioral motifs and patterns of interaction, that is, coordinate common behavior according to rules that lead to consortia of groups that differentiate self and non-self by integrating certain RNA stem-loop structures, which are beneficial for the stem-loop group, warding off others that do not fit into the consortia. We can find *de novo* initiation of behavior that cannot be deduced from former behavioral patterns, such as highly adaptive processes, retaining a contextual history, optimal energy costs (smart), solving problems beyond the capacity of its individual members, and fast-changing reactions against non-members (Villarreal and Witzany, 2013a, 2013b). These capabilities are all absent in inanimate nature.

5. RNA Group Interactions: Networking, Immune Reaction, Membership Roles

Interestingly, an RNA group's membership can never be fully specified, since it can always be further parasitized by as yet unknown parasites. This essential feature renders the ability to absolutely specify membership as basically undeniable. An RNA stem-loop group can never be fully secure from as yet unidentified genetic parasites. But the most important consequence from this insecurity is that it provides the inherent capacity for novelty, that is, the precondition for evolutionary innovation, such as new RNA groups and increased complexity (Villarreal, 2009a, 2011, 2015).

If we look at some interactional motifs of RNA agents to form consortial biotic structures, (i.e., RNA groups which together share a functional identity), we must look at the group building of RNA stem-loop structures (Witzany, 2014a). As mentioned above, it has been found that single stem loops interact in a pure physical, chemical mode without selective forces, independent of whether they are derived randomly or are artificially constructed under *in vitro* conditions (Vaidya *et al.*, 2012). In

contrast to this, if these single RNA stem loops build groups, they transcend a pure physical–chemical interaction pattern and biological selection emerges, causing biological identities capable of self/non-self-identification and preclusion, immune functions, and dynamically changing (adapting) membership roles.

A single alteration in a base-pairing RNA stem that leads to a new bulge may dynamically alter not only this single stem loop but may also change the whole group identity of which this stem loop is a part. This means that any RNA group transcribed out of the DNA storage medium, i.e., non-coding RNAs, mobile genetic elements, endogenous retroviruses and their defective and other persistent genetic parasites (not to forget their degraded parts), may play a new, modular, unexpected, and additionally non-predicable role in cooperating/competing RNAs, genetic regulations, counter-regulations, and diseases or infection events (Smit *et al.*, 2006; Manrubia and Briones, 2007; Geuking *et al.*, 2009; Stoddard, 2011; Villarreal and Witzany, 2015).

Simple self-ligating RNA stem loops can build much larger groups of RNA stem loops that serve to increase complexity (Salomon *et al.*, 2012). Significantly, RNA fragments that self-ligate into self-replicating ribozymes spontaneously form cooperative networks. For example, it was found that the three-membered networks showed highly cooperative growth dynamics. When such cooperative networks compete directly against selfish autocatalytic cycles, the former grow faster, indicating RNA populations evolve greater complexity through cooperation (Vaidya *et al.*, 2012; Higgs and Lehman, 2015). In this respect, cooperation outcompetes selfishness.

6. DNA Sequence Syntax with Several Different Meanings (Functions)

The genetic syntax, i.e., the nucleotide arrangement of genetic information is not unequivocal. This means that a given genetic sequence does not offer its final meaning. There are several processual steps in which the same sequence syntax may get several different meanings (functions). In natural languages and codes, this is a usual procedure and — additionally — it saves energy costs: A given sign sequence may represent various meanings/functions according to its contextual use. It is not necessary to generate a new sequence for every function. This indicates the primacy of pragmatics in determining the sequence structure, not its syntax (Witzany, 2014b; Cech, 2012; Doudna *et al.*, 1989). Several

context-dependent modifications of the meaning (function) of a given DNA sequence are currently known and investigated. Let's have a look at the various techniques to modify meaning (function) of given genetic sequences.

6.1. *Epigenetics: Context-dependent markings*

Although all cell types of any known organism contain the organism-specific genetic information, they are expressed according to their spatiotemporal position and their contextual (pragmatic) needs, such as developmental stages, stress, damage repair, or changing environments. This means that depending on the context, for a cell located within an organ of an organism, the expression leads to a tissue- and site-specific new cell at the right time (Slotkin and Martienssen, 2007; Turner, 2002).

Chromatin marking enables a kind of identity programming (Barlow, 2011). This means that a specific cell within an organism is able to obtain or even change its identity through epigenetics, if developmental, environmental, nutritional, or stress-related conditions make it necessary. RNAs are mobile and can serve as signals throughout the tissues, organs, and whole organisms. In this respect, the imprinting of new experiences that leads to variable meanings of genetic information depends on the ability of non-coding RNAs. With epigenetic marking, life has an appropriate technique for the emergence of memory and learning processes for faster adaptation (Mattick, 2010; Blaze and Roth, 2012; Marshall and Bredy, 2016; Kim and Kaang, 2017).

Both the small RNAs and long non-coding RNAs are competent in terms of directing chromatin changes through histone modifications and DNA methylation. These non-coding RNAs are able to direct chromatin-modifying agents to specific targets. In small RNA-driven silencing pathways, the regulatory RNAs identify and mark potentially dangerous "non-self" elements for transcriptional silencing or elimination (Slotkin and Martienssen, 2007; Shapiro, 2014). In other networks, homology between the regulatory RNA and the target locus marks the region as "self" and protects it from silencing or elimination. Interestingly, epigenetic marking conceivably originally emerged to defend genomes against genetic invaders (Huda and Jordan, 2009; Huda *et al.*, 2010).

6.2. *RNA editing*

We can look at a shared behavioral motif of how to modify meaning out of a given DNA sequence syntax in RNA Editing (Göringer, 2008). RNA editing

is a co- or post-transcriptional process, which alters the RNA sequence derived complementarily to the DNA from which it was transcribed. Before RNA editing, the editosome, i.e., small nuclear RNAs interconnected with a variety of proteins, must be assembled in a strictly coordinated process.

RNA editing changes gene sequences at the RNA level (Homann, 2008). 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. RNA editing alterations of such transcribed RNA sequences occur by modification, substitution, and insertion/deletion processes (Smith, 2008). Editing sites have to be identified individually to differentiate an A,T,G,C to be edited, from an A,T,G,C which should not be edited. The discriminating information can be found in the nucleotide sequence surrounding a given site. This means that context is relevant for recognition. Thus, each editing site carries its own recognition context (Gott, 2003).

6.3. *Alternative splicing*

RNA editing predates splicing and is timely and functionally interconnected. Editosome and spliceosome are important interacting agents (Matlin and Moore, 2007; House and Lynch, 2008; Yang, 2015). As in ribosome and editosome assembly, in spliceosome construction also, a ribonucleoprotein complex is assembled in various steps that cut out introns and splice exons together. The spliceosomal ribonucleoproteins are mainly small nuclear RNAs that are interconnected with at least 300 different proteins to five spliceosomal subunits.

Interestingly, a variety of steps in which the subunits of the final spliceosome are produced are counterbalanced within competing parts that regulate the stepwise processing of subcomplexes. After this final splicing procedure of the mature spliceosome, the remaining RNA products are actively discharged from the spliceosome and the remaining ribonucleoprotein particles are recycled for further catalytic processes as multi-use modules. Dependent on these regulations, the end product may vary with respect to the context dependency of the regulation process, which is highly sensitive to various needs and circumstances. In consequence, spliceosomal regulation differentiates the inclusion (splicing enhancers) or exclusion (splicing silencers) of exons in the final mRNA (Matlin and Moore, 2007). Splicing regulation occurs due to competing *cis*-acting elements that precisely balance regulatory proteins (Pyle and Lambowitz, 2006).

6.4. *tRNA-derived fragments*

Transfer-RNA (tRNA) is a key player between the RNA world and the protein world. tRNA-derived RNA fragments are small RNAs that are not random degradation products, but are specifically cleaved from mature tRNA transcripts. tRFs are microRNA, like small RNAs, involved in the regulation of genome stability. A variety of gene structures and RNA secondary structures of pre-tRNAs and mature RNAs have been identified in all domains of life. Coding sequences for these tRNAs are either separated within introns, fragmented, or permuted at the genome level (Keam and Hutvagner, 2015; Fujishima and Kanai, 2014; Gebetsberger *et al.*, 2016; Kanai, 2015).

Although evolutionary origins of the tRNA gene disruptions are unknown, a variety of tRNA structures seem to be co-evolved with RNA splicing endonuclease. Interestingly, tRNA-derived small RNAs are able to identify and target transcripts of transposable elements. This may indicate their relation to viruses and similar invading genetic parasites (Martinez *et al.*, 2017; Wang *et al.*, 2013).

6.5. *Pseudo-knotting*

Pseudo-knotting is a simple way in which a single strand of RNA can fold back on itself. This plays fundamental roles in structurally organizing complex RNAs, such as the assembly of ribonucleoprotein complexes, and additionally, in translational regulation and recoding when present within messenger RNAs (Staple and Butcher, 2005; Peselis and Serganov, 2014).

Because pseudo-knotting is extremely context sensitive, methods to predict pseudo-knotting based on stochastic context-free grammars were not very successful until recently (Staple and Butcher, 2005). The roles pseudo-knotting plays include generating the catalytic core of various ribozymes, self-splicing introns, and telomerase (Peselis and Serganov, 2014; Tholstrup *et al.*, 2012). Additionally, pseudo-knots play crucial roles in altering gene expression by inducing ribosomal frameshifting in many viruses. Pseudo-knots also induce ribosomes to slip into alternative reading frames. Because this behavioral motif is similar to that of self-splicing Group II introns, this may indicate their descent to genome-invading agents (Toor *et al.*, 2008).

6.6. *Frameshifting*

A ribosomal frameshift is a natural technique to process alternative translation of an mRNA sequence by changing the open-reading frame. It is

usual in viruses, and enables the virus to encode multiple types of proteins from the same mRNA (Atkins *et al.*, 2016). Through frameshifting viruses are able to create many protein structures from small genomes. During frameshifting, the ribosome shifts forward or backward, in a way specific to the virus. Because three mRNA nucleotides always make a codon, which is translated into one amino acid, the frameshift makes every amino acid different, because it translates a different set of codons in the new reading frame (Giedroc and Cornish, 2009; Dinman, 2006). This means a different protein product is formed. Ribosomes typically translate mRNA without shifting the translational reading frame. Interestingly, a number of organisms have evolved mechanisms to cause site-specific or programmed frameshifting of the ribosome in either the +1 or −1 direction.

6.7. *Loop kissing*

An RNA group is built up of different members that represent a group identity (like gangs), as suggested by Villarreal (2015). One main behavioral motif here is the "kissing loop" motif. The single-stranded loop RNA is open for interactions with other (self or even non-self) single-stranded RNAs via base pairing, according to base-pairing combination rules (Brunel *et al.*, 2002; Cao *et al.*, 2014; Cao and Chen, 2011). This may have highly dynamic aspects such as the loop–loop interaction within an RNA stem-loop group (as represented by editosomes, spliceosomes, ribosomal subunits, and viruses) or between RNA stem loops with non-related RNA loops present within the cell, tissue, organ, or organism. Some of them may be remnants of formerly degraded RNAs. RNA recycling is an important resource for building and rebuilding RNA identity groups, i.e., formerly foreign (non-self) RNA group members may become relevant parts of a new (self) identity, according to the contextual need.

6.8. *Bypassing translation*

One of the most interesting explorations in the last decades was the multiple ways in which identical nucleotide sequences that serve as genetic storage mediums undergo a variety of dynamic changing modes during transcription as outlined above. The translational process from DNA via RNA and an abundance of regulatory processes of non-coding RNAs is object to a further process to dynamically modify the genetically stored meanings (functions).

During the translational process, ribosomes may bypass and, therefore, ignore parts of the mRNA sequence (Samatova *et al.*, 2014;

Agirrezabala, 2017). The translation is blocked at a certain codon and a different stem-loop RNA ensures the further translation process at a certain stage of the mRNA. For example, in the mitochondrial genome of a yeast species, 81 translational bypassing elements were recently found. The mitochondrial bypassing resembles a sequence of the t4 phage gene 60. Interestingly, some codons remain unused in mitochondrial translation and can be used as variable module-like parts in new ways, such as serving as translational barriers for bypassing undesired sequences, or they may be reassigned to new amino acid identities by codon capture. Such bypassing elements resemble dynamic modes of mobile genetic elements.

6.9. *Competing endogenous RNAs*

Competing endogenous RNAs (ceRNAs) designate a variety of RNA consortia that compete for the wide range of micro RNAs, which regulate messenger RNAs (Salmena *et al.*, 2011). Such RNA consortia are pseudogenes, long non-coding RNAs, and circular RNAs. Long non-coding RNA-derived microRNAs are very abundant and regulate nearly half of the genomes. ceRNAs possess microRNA-binding sites (miRNA response elements) and ceRNAs may also cooperate to repress targets (Denzler *et al.*, 2014). Such competing endogenous RNAs indicate competitive behavior as a biological phenomenon derived from the early RNA world of an abundance of RNA stem-loops competing for limited resources initiating biological selection processes. Competing endogenous RNAs are a further level of regulation processes that interact in a concerted manner with other RNA-determined regulation processes. If their competition gets counterbalanced, diseases such as cancer may be the result (Yang *et al.*, 2016).

7. Genetic Novelty and Genomic Variations by Viruses

After looking at the various techniques to modify the meaning/function of a given genetic sequence, we must be aware of genome-invading agents, such as viruses and RNA networks, which represent a very large and dynamic source of genetic novelty.

They can cooperate, build communities, generate nucleotide sequences de novo, and insert/delete them into host genetic content (Stedman, 2013, 2015). Viruses and RNA networks often remain as mobile genetic elements or similar "defectives" and determine host genetic identities throughout all kingdoms, including the virosphere. But, inclusion of a transmissive viral

biology differs fundamentally from mainstream concepts of the last decades, in which it represents a vertical domain of life providing vast amounts of linked information not derived from direct ancestors (Villarreal, 2005).

Current knowledge about the virosphere and its role in evolution ("virolution") indicates interactions of RNA viruses, DNA viruses, viral swarms, and viral and RNA-based subviral networks that cooperate and coordinate (regulate) within cellular genomes either as replication-relevant coplayers or suppression-relevant silencers (Stedman, 2013, 2015; Ryan, 2009; Seligman and Raoult, 2016). Some represent infection-derived modular tools of non-coding RNAs which have built consortia of complementary agents that function together (Tycowski *et al.*, 2015).

Persistent viral lifestyles that we can identify as counterbalanced viral properties are found as addiction modules (see below), which transfer complete genetic datasets into host genomes and alter the DNA genetic identity of the host and, additionally, the formerly competing viral agents, without damage to the host genome content. As we saw earlier, RNA consortia may edit genetic code information without altering inherited DNA content. Transcribed from DNA cellular sequences, the RNA-activated inhabitants from former viral infection events act as modular tools for cellular needs in nearly all cellular processes (Witzany, 2009, 2016, 2017). They act not as lytic agents but as part of the host genetic identity because of domestication by counterbalanced competing genetic parasites (Villarreal, 2012a). Addiction modules not only change genetic identity but also enrich immune functions of host organisms to fight related parasites (Moelling and Broecker, 2015). We can easily identify these in the known complex adaptive immune systems (Villarreal, 2009b).

Endogenous retroviral competences in the persistent status are often characterized by features expressed only in the strict time window of a developmental process, such as axis formation, trophectoplast formation (as in the case of endogenous retroviruses in placental mammals), or the S phase of the cell cycle (Sciamanna *et al.*, 2009; Villarreal and Witzany, 2010; Villarreal, 2016b). In these highly specialized contexts, they are replicated through signaling, which blocks the suppression of the replication process. After the function is fulfilled, a signal once again initiates the suppressor function. If the signal is missing, as in the case of regulation damage, the whole process can get out of control.

Diversity generating retroelements — with their (1) higher-order regulatory functions, (2) ability for genetic creativity and (3) capacity for innovation of new regulatory patterns and combinations — are most probably

descended from retroviruses, which can easily be identified by their three essential parts: gag, pol, and env. Most endogenous retroviruses have been degraded into formerly connected domains, but they can still be recognized by retroposons or one of their three core genes (Boeke, 2003; Eickbush and Jamburuthugoda, 2008). This means that their formerly connected genomic content may be used by host organisms as single or networking modular tools for a variety of new regulatory functions (Belfort and Curcio, 2011; Witzany, 2012; Moelling *et al.*, 2017). Retroelements represent the most abundant genome invaders, especially in mammals.

8. How to Persist within an Infected Genome: Counterbalancing Addiction Modules

Addiction modules represent at least two competing genetic parasite consortia, which try to invade host genomes. The competition occurs not solely against each other, but primarily against immune functions of the host. In some cases, the competing genetic parasites and the third party, the host immune system, can together install persistent modus in which all remain as a counterbalancing module — more or less stable (Villarreal, 2009a, 2012b).

If these module-like agents are conserved into the genomic identity of the host, the former identity changes dramatically. New features are part of the host genomes which did not exist before. In such addiction module integration events, up until 100 new genes can be transferred into the host genome in a single event (Villarreal, 2005).

Addiction modules can be defined as features that consist, in general, of a stable, toxic component, which is counterbalanced by an unstable component inhibiting and suppressing the toxic component (Villarreal, 2012b). Both are necessary for transferring a feature to the host without harming the host. For example, in the case of a restriction/modification module, this means 52 restriction enzymes are counterbalanced by 52 modification enzymes. This indicates how complexly addiction modules are constructed and how difficult it can be to understand the evolution of such phenotypes (Kobayashi, 2001; Mruk and Kobayashi, 2014).

Several kinds of addiction modules are known. First of all, and most prominent, in bacterial life is the restriction/modification addiction module, which is a common feature with immune functions. One part consists of an antitoxic modification enzyme, which is an unstable beneficial (protective) agent. The counterpart consists of a toxic restriction enzyme component,

which is a stable but harmful (destructive) agent. Another kind of addiction module consists of two related features. There is an antitoxic antipore toxin, which represents the unstable protective agent and a toxic component with a toxic pore, which represents the stable but destructive agent (Villarreal, 2012b; Gerdes and Wagner, 2007). A third kind of addiction module consists of the antitoxic viral immunity component and the toxic component of viral-mediated lysis. This third kind is the most obvious viral-derived immune function, because it necessarily consists of a persistent genetic parasite and an external lytic phage (Villarreal, 2012, 2016a).

Most importantly, this new information representing genetic novelty by a persistent integration of counterbalanced coded genetic elements is not the result of error replication, but a result of module-like linked genetic contents. This fundamental difference to error replication narratives proposes new nucleic acid sequence constructions by integration of larger content arrangements into a coherent syntax without destroying the already existing sequence content (Witzany, 2014).

Additionally, the basic motif of any counterregulated modules, such as known toxin/antitoxin modules, reminds us of similar structures with similar features such as the nuclease/ligase regulation as basic competence for self/non-self-differentiation within RNA stem-loop groups (Witzany, 2016), the kinein/dynein retroviral movement motif (Witzany, 2009), and the viral origin of some ribonucleoproteins (Villarreal, 2012b). Also, the telomere/telomerase-dependent genome maintenance (Witzany, 2008) and the function of centrosomes and spindles (Alliegro, 2011) most seemingly represent counterregulated addiction modules, although their origin still remains in the dark. Such counterbalancing interactions of former competing genetic parasites, now serving as a beneficial tool within host organisms, can be found generally in insertion/deletion modules and in adaptive immune systems such as VDJ of complex eukaryotes or in the bacterial CRIPR/Cas immune system (Villarreal, 2016a).

9. Genetic Novelty by Domesticated Infectious Agents

Transposable elements shape the structure and function of infected genomes (non-LTRs, LINEs, SINEs, Alus) and generate genomic variations (Slotkin and Martienssen, 2007). This means that sequence syntax may change and such changes may cause phenotypic variations. Also post-integration recombinations generate structural variations. Additionally, host cells may regulate the invading RNA consortia by restriction motifs, epigenetic

methylation silencing, or by other ways to restrict retrotransposition (Richardson *et al.*, 2015). Depending on descent transposable elements may serve as regulatory elements for host cells, such as promotors, enhancers, or repressors.

Diversity-generating retroelements such as retrotransposons seem to derive from mobile self-splicing introns (a sister strand of retroviruses) together with retrons, retroplasmids, retrointrons, non-LTR and LTR retrotransposons, caulimoviruses, hepadnaviruses, and the DIRS and Bel family (Boeke 2003; Frost *et al.*, 2005). Their double role (as promotors and repressors) depends on the differences between retrotransposon families, even within the same family, whether they have co-evolved with host genomes, have been co-opted to serve in species-specific functions, and others that are not beneficial and may cause diseases (Robbez-Masson and Rowe, 2015). The evolutionarily original polymerase, the reverse transcriptase, with its dominant roles of retrotransposition and the transfer of RNA sequences into DNA sequences, plays important roles (Belfort and Curcio, 2011; Egan and Collins, 2012).

In this perspective, retrotransposons may change their function between repression and promotion of gene regulations in a highly species-specific manner without error replication, and act in a coordinated way to generate regulatory complexity for gene networks in developmental stages (Geuking *et al.*, 2009; Pathak and Nagy, 2009). If retrotransposons inactivate genes by inserting within them, or even more importantly, separating genes from a regulatory sequence, bringing new regulatory sequences to genes, or altering RNA splicing, or even RNA editing, they may generate genomic variations by incorporating themselves into the new sequences (Finnegan, 2012; Liscovitch *et al.*, 2017). According to the concerted temporal steps in the developmental processes, they may act in tissue-specific repression and activation complexity to coordinate the timely steps of tissue developments within organisms. As exapted and co-opted former infectious agents, they are essential tools of host genomes and they serve as drivers of genome plasticity and creation of novel genes (Villarreal and Witzany, 2015; Robbez-Masson and Rowe, 2015).

Additionally, all these interactional motifs play essential roles in warding off the invading genetic parasites. As they drive host evolution by genetic novelty and genomic variations, they may constantly serve as disease-causing enemies also. But some are co-opted and exapted, and they serve as essential parts in the dynamic adaptation of host immune systems (Goodier, 2016; Chuong *et al.*, 2016). Such sub-functionalization and

neo-functionalization are essential characteristics in this network modularity (Kuniak and Makalowska, 2017).

Interestingly, all these invading RNA elements described in the previous sections may recombine and may even recombine with DNA viruses. This serves as the "ultimate modularity" on the genetic level (Koonin *et al.*, 2015; Koonin and Krupovic, 2017).

10. Masters of the RNA Universe: Reverse Transcriptase/RNAse H

A key agent in several aspects of biology, especially in genetics, is the diversity-generating retroelement reverse transcriptase (RT). The reverse transcriptase, derived from the ancient RNA world, and correctly termed, is the primary tool to transport information from RNA to DNA before other polymerases. RTs can be identified through the whole current RNA world activities in virions (naked RNA stem-loop consortia) and RNA viruses such as retroviruses, and are the essential tool to place RNA information within DNA-based genomes (Nakamura and Cech, 1998; Sciamanna *et al.*, 2009). In this respect, it is crucial to install persistent viral agents in host genomes.

The RTs care for virus replication cycles as well as the suppressed viral-derived inhabitants that serve as regulatory elements of the 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. Especially if the RTs get into this modus, they may change the genetic identity of host genomes and fix it into the DNA storage medium as a part of an inheritable feature (Deng *et al.*, 2012; Spadafora, 2016). The telomerase, a subspecies of the reverse transcriptase family, cares for telomere replication at the ends of chromosomes and, therefore, is not only the key player of genome maintenance but is also a part of an immunity function that protects telomere ends from infection by genetic parasites (Nosek *et al.*, 2006; deLange, 2015; Podlevsky and Chen, 2016). This is a new role of a reverse transcriptase and it demonstrates that the RNA agents serve for both infection of the host genomes and — as coapted exaptations — immune functions against (related) infectious agents.

Reverse Transcriptase is the key player to transfer RNA sequences into DNA. This is a counterregulatory process that we can look at in embryogenesis and cell differentiation. In cancerogenesis, this gets out of control.

We can look at such context-relevant loss of gene control in preimplantation of placenta, trophectoblast differentiation, as well as tumor growth. In all these situations, the genetic identity producers that are silenced or conserved as regulatory tools may become virulent again, active for change, deregulation, 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 deregulation of former counterbalanced networks of regulations (Spadafora, 2015). In this respect, carcinogenesis is the result of communication breakdown. Interestingly, invading cancer cells behave according to the retroviral-mediated embryological program of placenta invasion (Villarreal, 2016b). In such a case, the regulatory network may get into an unregulated, wrong direction. This means that, e.g., cell growth still remains the goal, but the stop signal does not function and the growth program goes on in an epigenetically induced deregulated way (Smith *et al.*, 2017), or invasion of the placenta into the uterus is the coordinated behavioral goal, but the invasive signaling network does not stop and spreads into other organs (metastasis) (Spadafora, 2015, 2016).

Retroviral infections are essential for evolutionary novelty, in that they serve as the main editors for new genetic identities to host genomes. The key natural engineers in this respect are RT and RNase H (Moelling and Broecker, 2015; Moelling *et al.*, 2017). They can transport new cellular functions and contribute to the host immune defense. Their high production rate of new sequences, formerly termed "error prone", renders RT–RNases H the drivers of evolutionary innovation. Integrated retroviruses and the related transposable elements (TEs) represent up to 80% of eukaryotic genomes and are also present in prokaryotes. Endogenous retroviruses not only regulate host genes but also introduce novel genes, such as the syncytins, that mediate maternal–fetal immune tolerance, and may become infectious again, under certain circumstances.

The RT and the RNase H represent the most ancient and abundant protein folds (Moelling *et al.*, 2017). RNases H most probably evolved from ribozymes in the ancient RNA world, forming ribosomes, RNA replicases, and polymerases. RT and RNases H are found in bacterial group II introns, the ancestors of transposable elements. They regulate viral replication, antiviral defence in eukaryotes and prokaryotes, splicing, DNA repair, and serve as various mediators of small regulatory RNAs. The evolution of DNA-based life, as we know it, would not have been possible without the RTs and the RNase H competencies.

11. Conclusions

In the decades where biology has been exclusively a subdiscipline of physics and chemistry, the only explanatory model to explain changes and novelties of replicated nucleotide sequences has been the "error replication" (mutations) narrative. This was the consequence of Erwin Schroedinger's answer "Life is physics and chemistry" to his question "What is Life". Within this mechanistic paradigm, natural genome editing, i.e., competent natural genetic content arrangement by cooperative consortia of RNA networks and viruses, is not part of theories and concepts in biology. The main problem of the error replication (mutations) narrative is that it cannot explain sufficiently these new and rich empirical data of the variety of RNA consortia and viruses as natural genome editors. No doubt they can insert and delete, can integrate themselves in complex genetic content without destroying relevant protein-coding content, arrange, rearrange, cut and paste, copy and paste, and serve as multiple reuseable modular tools in the genetic content architecture. In the old error replication narrative, all these consortia remain as randomly derived results of statistical events by genetic drift. While the old paradigm cannot integrate these new empirical data, the natural genome editing paradigm is essentially based on all these data. Current knowledge identified an abundance of processes with a variety of small interacting agents that actively edit genetic code sequences by different behavioral motifs; these can all be combined in a module-like manner. Natural genome editing is a technique that actively initiates "ultimate modularity" to generate genetic novelty and genomic variations, which can be fulfilled by cooperative interacting networks of viruses and subviral RNA consortia.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Agirrezabala, X., Samatova, E., Klimova, M., Zamora, M., Gil-Carton, D., Rodnina, M. V. and Valle, M. (2017). Ribosome rearrangements at the onset of translational bypassing, *Sci. Adv.* 3(6), e1700147.

Alliegro, M. C. (2011). The centrosome and spindle as a ribonucleoprotein complex, *Chromosome Res.* 19, 367–376.

Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. and Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use, *Nucleic Acids Res.* 44, 7007–7078.

Barlow, D. P. (2011). Genomic imprinting: A mammalian epigenetic discovery model, *Annu. Rev. Genet.* 45, 379–403.

Belfort, M., Curcio, M. J. and Lue, M. J. (2011). Telomerase and retrotransposons: Reverse transcriptases that shaped genomes, *Proc. Natl. Acad. Sci. USA* 108, 20304–20310.

Blaze, J. and Roth, T. L. (2012). Epigenetic mechanisms in learning and memory, *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 105–115.

Boeke, J. D. (2003). The unusual phylogenetic distribution of retrotransposons: A hypothesis, *Genome Res.* 13, 1975–1983.

Briones, C., Stich, M. and Manrubia, S. (2009). The dawn of the RNA World: Toward functional complexity through ligation of random RNA oligomers, *RNA* 15, 743–749.

Brunel, C., Marquet, R., Romby, P. and Ehresmann, C. (2002). RNA loop–loop interactions as dynamic functional motifs, *Biochimie* 84, 925–944.

Cao, S. and Chen, S. J. (2011). Structure and stability of RNA/RNA kissing complex: With application to HIV dimerization initiation signal, *RNA* 17, 2130–2143.

Cao, S., Xu, X. and Chen, S. J. (2014). Predicting structure and stability for RNA complexes with intermolecular loop-loop base-pairing, *RNA* 20, 835–845.

Carbonell, A., Flores, R. and Gago, S. (2012). Hammerhead ribozymes against virus and viroid RNAs, in *From Nucleic Acids Sequences to Molecular Medicine*, Erdmann V. A. and Barciszewski, J. (eds.), Springer, pp. 411–427.

Cech, T. R. (2012). The RNA worlds in context, *Cold Spring Harb. Perspect. Biol.* 4(7), a006742.

Cech, T. R. and Steitz, J. A. (2014). The noncoding RNA revolution — Trashing old rules to forge new ones, *Cell* 157, 77–94.

Chuong, E. B., Elde, N. C. and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses, *Science* 351, 1083–1087.

de Lange, T. (2015). A loopy view of telomere evolution, *Front Genet.* 6, 321.

Deng, Z., Wang, Z. and Lieberman, P. M. (2012). Telomeres and viruses: Common themes of genome maintenance, *Front. Oncol.* 2, 201.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D. and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance, *Mol. Cell* 54, 766–776.

Dinman, J. D. (2006). Programmed ribosomal frameshifting goes beyond viruses, *Microbe Wash. DC.* 1, 521–527.

Doudna, J. A., Cormack, B. P. and Szostak, J. W. (1989). RNA structure, not sequence, determines the $5'$ splice-site specificity of a group I intron, *Proc. Natl. Acad. Sci. U.S.A.* 86, 7402–7406.

Egan, E. D. and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins, *RNA* 18, 1747–1759.

Eickbush, T. H. and Jamburuthugoda, V. K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases, *Virus Res.* 134, 221–234.

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules, *Naturwissenschaften* 58, 465–523.

Eigen, M. and Schuster, P. (1978). The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle, *Naturwissenschaften* 65, 7–41.

Finnegan, D. J. (2012). Retrotransposons, *Curr. Biol.* 22, 432–437.

Frost, L. S., Leplae, R., Summers, A. O. and Toussaint, A. (2005). Mobile genetic elements: The agents of open source evolution, *Nat. Rev. Microbiol.* 3, 722–732.

Fujishima, K. and Kanai, A. (2014). tRNA gene diversity in the three domains of life, *Front. Genet.* 5, 142.

Gebetsberger, J., Wyss, L., Mleczko, A. M., Reuther, J. and Polacek, N. (2016). A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress, *RNA Biol.* 28, 1–10.

Gerdes, K. and Wagner, E. G. (2007). RNA antitoxins, *Curr. Opin. Microbiol.* 10, 117–124.

Geuking, M. B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., Zinkernagel, R. M. and Hangartner, L. (2009). Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration, *Science* 323, 393–396

Giedroc, D. P. and Cornish, P. V. (2009). Frameshifting RNA pseudoknots: Structure and mechanism, *Virus Res.* 139, 193–208.

Goodier, J. L. (2016). Restricting retrotransposons: A review, *Mob. DNA* 7, 16.

Göringer, H. U. (ed.) (2008). *RNA Editing*, Springer.

Gott, J. M. (2003). Expanding genome capacity via RNA editing, *C. R. Biol.* 326, 901–908.

Hayden, E. J. and Lehman, N. (2006). Self-assembly of a group I intron from inactive oligonucleotide fragments, *Chem. Biol.* 13, 909–918.

Higgs, P. G. and Lehman, N. (2015). The RNA world: Molecular cooperation at the origins of life, *Nat. Rev. Genet.* 16, 7–17.

Homann, M. (2008). Editing reactions from the perspective of RNA structure, in *RNA Editing*, Göringer, H. U. (ed.), Springer, pp. 1–32.

House, A. E. and Lynch, K. W. (2008). Regulation of alternative splicing: More than just the ABCs, *J. Biol. Chem.* 283, 1217–1221.

Huda, A. and Jordan, I. K. (2009). Epigenetic regulation of mammalian genomes by transposable elements, *Ann. N. Y. Acad. Sci.* 1178, 276–284.

Huda, A., Mariño-Ramírez, L. and Jordan, I. K. (2010). Epigenetic histone modifications of human transposable elements: Genome defense versus exaptation, *Mob. DNA* 1, 2.

Kanai, A. (2015). Disrupted tRNA genes and tRNA fragments: A perspective on tRNA gene evolution, *Life (Basel)* 5, 321–331.

Keam, S. P. and Hutvagner, G. (2015). tRNA-derived fragments (tRFs): Emerging new roles for an ancient RNA in the regulation of gene expression, *Life (Basel)* 5, 1638–1651.

Kim, S. and Kaang, B. K. (2017). Epigenetic regulation and chromatin remodeling in learning and memory, *Exp. Mol. Med.* 49, e281.

Kobayashi, L. (2001). Behavior of restriction-modification system as selfish mobile elements and their impact on genome evolution, *Nucleic Acid Res.* 29, 3742–3756.

Koonin, E. V. and Krupovic, M. (2017). Polintons, virophages and transpovirons: A tangled web linking viruses, transposons and immunity, *Curr. Opin. Virol.* 25, 7–15.

Koonin, E. V., Dolja, V. V. and Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity, *Virology* 479–480, 2–25.

Kubiak, M. R. and Makałowska, I. (2017). Protein-coding genes' retrocopies and their functions, *Viruses* 9(4), pii: E80.

Liscovitch-Brauer, N., Alon, S., Porath, H. T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E. Y., Rosenthal, J. J. and Eisenberg, E. (2017). Trade-off between transcriptome plasticity and genome evolution in cephalopods, *Cell* 169, 191–202.

Manrubia, S. C. and Briones, C. (2007). Modular evolution and increase of functional complexity in replicating RNA molecules, *RNA* 13, 97–107.

Marshall, P. and Bredy, T. W. (2016). Cognitive neuroepigenetics: The next evolution in our understanding of the molecular mechanisms underlying learning and memory? *NPJ Sci. Learn.* 1, 16014.

Martinez, G., Choudury, S. G. and Slotkin, R. K. (2017). tRNA-derived small RNAs target transposable element transcripts, *Nucleic Acids Res.* 45, 5142–5152.

Matlin, A. J. and Moore, M. J. (2007). Spliceosome assembly and composition, *Adv. Exp. Med. Biol.* 623, 14–35.

Mattick, J. S. (2001). Non-coding RNAs: The architects of eukaryotic complexity, *EMBO Rep.* 2, 986–991.

Mattick, J. S. (2003). Challenging the dogma: The hidden layer of noncoding RNAs in complex organisms, *BioEssays* 25, 930–939.

Mattick, J. S. (2009). Deconstructing the dogma: A new view of the evolution and genetic programming of complex organisms, *Ann. Acad, N. Y. Sci.* 1178, 29–46.

Mattick, J. S. (2010). RNA as the substrate for epigenome–environment interactions: RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning, and cognition, *BioEssays* 32, 548–552.

Mercer, T. R. and Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation, *Nat. Struct. Mol. Biol.* 20, 300–307.

Moelling, K. and Broecker, F. (2015). The reverse transcriptase-RNase H: From viruses to antiviral defense, in *DNA Habitats and Their RNA Inhabitants*, Witzany, G. (ed.), Wiley & Sons, pp. 126–135.

Moelling, K., Broecker, F., Russo, G. and Sunagawa, S. (2017). RNase H as gene modifier, driver of evolution and antiviral defense, *Front. Microbiol.* doi: 10.3389/fmicb.2017.01745.

Mruk, I. and Kobayashi, I. (2014). To be or not to be: Regulation of restriction-modification systems and other toxin-antitoxin systems, *Nucleic Acids Res.* 42, 70–86.

Nakamura, T. M. and Cech, T. R. (1998). Reversing time: Origin of telomerase, *Cell* 92, 587–590.

Nosek, J., Kosa, P. and Tomaska, L. (2006). On the origin of telomeres: A glimpse at the pre-telomerase world, *BioEssays* 28, 182–190.

Pathak, K. B. and Nagy, P. D. (2009). Defective interfering RNAs: Foes of viruses and friends of virologists, *Viruses* 1, 895–919.

Peselis, A. and Serganov, A. (2014). Structure and function of pseudoknots involved in gene expression control, *Wiley Interdiscip. Rev. RNA* 5, 803–822.

Podlevsky, J. D. and Chen, J. J. L. (2016). Evolutionary perspectives of telomerase RNA structure and function, *RNA Biol.* 13, 720–732.

Pyle, A. M. and Lambowitz, A. M. (2006). Group II introns: Ribozymes that splice RNA and invade DNA, in *The RNA World*, Gesteland, R. F., Cech, T. R. and Atkin, J. F. (eds.), 3rd edn., Cold Spring Harbor Laboratory Press, pp. 469–506.

Qureshi, I. A. and Mehler, M. F. (2012). Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease, *Nat. Rev. Neurosci.* 13, 528–541.

Richardson, S. R., Doucet, A. J., Kopera, H. C., Moldovan, J. B., Garcia-Perez, J. L. and Moran, J. V. (2015). The influence of LINE-1 and SINE retrotransposons on mammalian genomes, *Microbiol. Spectr.* 3(2), MDNA3-0061-201499.

Robbez-Masson, L. and Rowe, H. M. (2015). Retrotransposons shape species-specific embryonic stem cell gene expression, *Retrovirol.* 12, 45.

Rohwer, F., Youle, M., Maughan, H. and Hisakawa, N. (2014). *Life in Our Phage World*, Wholon.

Ryan, F. (2009). *Virolution*, Collins.

Salmena, L., Poliseno, L., Tay, Y., Kats, L. and Pandolfi, P. P. (2011). A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell* 146, 353–358.

Salomon, K., Appel, B. and Müller, S. (2012). RNA self-ligation: From oligonucleotides to full length ribozymes, *Biochimie* 94, 1457–1463.

Samatova, E., Konevega, A. L., Wills, N. M., Atkins, J. F. and Rodnina, M. V. (2014). High-efficiency translational bypassing of non-coding nucleotides specified by mRNA structure and nascent peptide, *Nat. Commun.* 5, 4459.

Schuster, P. (2011). Mathematical modeling of evolution. Solved and open problems, *Theory Biosci.* 130, 71–89.

Sciamanna, I., Vitullo, P., Curatolo, A. and Spadafora, C. (2009). Retrotransposons, reverse transcriptase and the genesis of new genetic information, *Gene* 448, 180–186.

Seligman, H. and Raoult, D. (2016). Univying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses, *Curr. Opin. Microbiol.* 31, 1–8.

Shapiro, J. A. (2004). A 21st century view of evolution: Genome system architecture, repetitive DNA, and natural genetic engineering, *Gene* 345, 91–100.

Shapiro, J. A. (2009). Revisiting the central dogma in the 21st century, *Ann. N.Y. Acad. Sci.* 1178, 6–28.

Shapiro, J. A. (2014). Epigenetic control of mobile DNA as an interface between experience and genome change, *Frontiers Genet.* 5, 1–16.

Slotkin, R. K. and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome, *Nat. Rev. Genet.* 8, 272–285.

Smit, S., Yarus, M. and Knight, R. (2006). Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories, *RNA* 12, 1–14.

Smith, H. C. (2008). Editing informational content of expressed DNA sequences and their transcripts, in *RNA Editing*, Göringer, H. U. (ed.), Springer, pp. 249–265.

Smith, Z. D., Shi, S., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., Gnirke, A., Michor, F. and Meissner, A. (2017). Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer, *Nature* 549, 543–547.

Spadafora, C. (2015). A LINE-1–encoded reverse transcriptase–dependent regulatory mechanism is active in embryogenesis and tumorigenesis, *Ann. N.Y. Acad. Sci.* 1341, 164–171.

Spadafora, C. (2016). Soma to germline inheritance of extrachromosomal genetic information via a LINE-1 reverse transcriptase-based mechanism, *BioEssays* 38, 726–733.

Spadaro, P. A. and Bredy, T. W. (2012). Emerging role of non-coding RNA in neural plasticity, cognitive function, and neuropsychiatric disorders, *Front. Genet.* 3, 132.

Staple, D. W. and Butcher, S. E. (2005). Pseudoknots: RNA structures with diverse functions, *PLoS Biol.* 3(6), e213.

Stedman, K. M. (2013). Mechanisms for RNA capture by ssDNA viruses: Grand theft RNA, *J. Mol. Evol.* 76, 59–364.

Stedman, K. M. (2015). Deep recombination: RNA and ssDNA virus genes in DNA virus and host genomes, *Annu. Rev. Virol.* 2, 203–217.

Stoddard, B. L. (2011). Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification, *Structure* 19, 7–15.

Tholstrup, J., Oddershede, L. B. and Sørensen, M. A. (2012). mRNA pseudoknot structures can act as ribosomal roadblocks, *Nucleic Acids Res.* 40, 303–313.

Toor, N., Keating, K. S., Taylor, S. D. and Pyle, A. M. (2008). Crystal structure of a self-spliced group II intron, *Science* 320, 77–82.

Turner, B. M. (2002). Cellular memory and the histone code, *Cell* 111, 285–291.

Tycowski, K. T., Guo, Y. E., Lee, N., Moss, W. N., Vallery, T. K., Xie, M. and Steitz, J. A. (2015). Viral noncoding RNAs: More surprises, *Genes Dev.* 29, 567–584.

Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J. and Lehman, N. (2012). Spontaneous network formation among cooperative RNA replicators, *Nature* 491, 72–77.

Villarreal, L. P. (2004). Can viruses make us human? *Proc. Am. Phil. Soc.* 148, 296–323.

Villarreal, L. P. (2005). *Viruses and the Evolution of Life*, ASM Press.

Villarreal, L. P. (2009). *Origin of Group Identity: Viruses, Addiction and Cooperation*, Springer.

Villarreal, L. P. (2009). The source of self: Genetic parasites and the origin of adaptive immunity, *Ann. N.Y. Acad. Sci.* 1178, 194–232.

Villarreal, L. P. (2011). Viral ancestors of antiviral systems, *Viruses* 3, 1933–1958.

Villarreal, L. P. (2012a). Viruses and host evolution: Virus-mediated self identity, in *Self and Non-Self*, Lopez-Larrea C. (ed.), Landes Bioscience and Springer Science Business Media, pp. 185–217.

Villarreal, L. P. (2012b). The Addiction Module as a Social Force, in *Viruses: Essential Agents of Life*, Witzany, G. (ed.), Springer, pp. 107–145.

Villarreal, L. P. (2015). Force for ancient and recent life: Viral and stem-loop RNA consortia promote life, *Ann. N.Y. Acad. Sci.* 1341, 25–34.

Villarreal, L. P. (2016a). Persistent virus and addiction modules: An engine of symbiosis, *Curr. Opin. Microbiol.* 31, 70–79.

Villarreal, L. P. (2016b). Viruses and the placenta: The essential virus first view, *APMIS* 124, 20–30.

Villarreal, L. P. and Witzany, G. (2010). Viruses are essential agents within the roots and stem of the tree of life, *J. Theor. Biol.* 262, 698–710.

Villarreal, L. P. and Witzany, G. (2013a). The DNA habitat and its RNA inhabitants: At the dawn of RNA sociology, *Genom. Ins.* 6, 1–12.

Villarreal, L. P. and Witzany, G. (2013b). Rethinking quasispecies theory: From fittest type to cooperative consortia, *World J. Biol. Chem.* 4, 79–90.

Villarreal, L. P. and Witzany, G. (2015). When competing viruses unify: Evolution, conservation, and plasticity of genetic identities, *J. Mol. Evol.* 80, 305–318.

Villarreal, L. P. and Witzany, G. (2018). Editorial: Genome invading RNA networks, *Frontiers Microbiol.* 9, 581, doi: 10.3389/fmicb.2018.00581.

Wang, Q., Lee, I., Ren, J., Ajay, S. S., Lee, Y. S. and Bao, X. (2013). Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection, *Mol. Ther.* 21, 368–379.

Witzany, G. (1995). From the "logic of the molecular syntax" to molecular pragmatism. Explanatory deficits in Manfred Eigen's concept of language and communication, *Evol. Cog.* 1, 148–168.

Witzany, G. (2008). The viral origins of telomeres, telomerases and their important role in eukaryogenesis and genome maintenance, *Biosemiotics* 1, 191–206.

Witzany, G. (2009). Noncoding RNAs: Persistent viral agents as modular tools for cellular needs, *Ann. N.Y. Acad. Sci.* 1178, 244–267.

Witzany, G. (2012). From molecular entities to competent agents: Viral infection-derived consortia act as natural genetic engineers, in Witzany, G. (ed.) *Viruses: Essential Agents of Life*, Springer, pp. 407–419.

Witzany, G. (2014a). RNA sociology: Group behavioral motifs of RNA consortia, *Life(Basel)* 4, 800–818.

Witzany, G. (2014b). Language and communication as universal requirements for life, in *Astrobiology: An Evolutionary Approach*, Kolb, V. (ed.), CRC Press, pp. 349–369.

Witzany, G. (2015). Life is physics and chemistry and communication, *Ann. N.Y. Acad. Sci.* 1341, 1–9.

Witzany, G., (ed.) (2015). *DNA Habitats and Their RNA Inhabitants*, Wiley & Sons.

Witzany, G. (2016). Crucial steps to life: From chemical reactions to code using agents, *BioSystems* 140, 49–57.

Witzany, G. (2017). Two genetic codes: Repetitive syntax for active non-coding RNAs; non-repetitive syntax for the DNA archives, *Comm. Integr. Biol.* 10(2), e1297352.

Witzany, G. and Baluska, F. (2012). Life's code script does not code itself. The machine metaphor for living organisms is outdated, *EMBO Rep.* 13, 1054–1056.

Yablonovitch, A. L., Deng, P., Jacobson, D. and Li, J. B. (2017). The evolution and adaptation of A-to-I RNA editing, *PLoS Genet.* 13(11), e1007064.

Yang, L. (2015). Splicing noncoding RNAs from the inside out, *Wiley Interdiscip. Rev. RNA* 6, 651–660.

Yang, C., Wu, D., Gao, L., Liu, X., Jin, Y., Wang, D., Wang, T. and Li, X. (2016). Competing endogenous RNA networks in human cancer: Hypothesis, validation, and perspectives, *Oncotarget.* 7, 13479–13490.

Zinad, H. S., Natasya, I. and Werner, A. (2017). Natural antisense transcripts at the interface between host genome and mobile genetic elements, *Front Microbiol.* 8, 2292.