

UC Irvine

UC Irvine Previously Published Works

Title

That is life: communicating RNA networks from viruses and cells in continuous interaction

Permalink

<https://escholarship.org/uc/item/4sn3f81d>

Journal

Annals of the New York Academy of Sciences, 1447(1)

ISSN

0077-8923

Authors

Villarreal, Luis P
Witzany, Guenther

Publication Date

2019-07-01

DOI

10.1111/nyas.14040

Peer reviewed

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Special Issue: *Genetic Novelty and Genomic Variation by RNA Networks and Viruses*

PERSPECTIVE

That is life: communicating RNA networks from viruses and cells in continuous interactionLuis P. Villarreal¹ and Guenther Witzany²¹Department of Molecular Biology and Biochemistry, University of California, Irvine, California. ²Telos–Philosophische Praxis, Buermoos, Austria

Address for correspondence: Guenther Witzany, Telos–Philosophische Praxis, Vogelsangstraße 18 c, 5111 – Buermoos, Austria. witzany@sbg.at

All the conserved detailed results of evolution stored in DNA must be read, transcribed, and translated via an RNA-mediated process. This is required for the development and growth of each individual cell. Thus, all known living organisms fundamentally depend on these RNA-mediated processes. In most cases, they are interconnected with other RNAs and their associated protein complexes and function in a strictly coordinated hierarchy of temporal and spatial steps (i.e., an RNA network). Clearly, all cellular life as we know it could not function without these key agents of DNA replication, namely rRNA, tRNA, and mRNA. Thus, any definition of life that lacks RNA functions and their networks misses an essential requirement for RNA agents that inherently regulate and coordinate (communicate to) cells, tissues, organs, and organisms. The precellular evolution of RNAs occurred at the core of the emergence of cellular life and the question remained of how both precellular and cellular levels are interconnected historically and functionally. RNA networks and RNA communication can interconnect these levels. With the reemergence of virology in evolution, it became clear that communicating viruses and subviral infectious genetic parasites are bridging these two levels by invading, integrating, coadapting, exapting, and recombining constituent parts in host genomes for cellular requirements in gene regulation and coordination aims. Therefore, a 21st century understanding of life is of an inherently social process based on communicating RNA networks, in which viruses and cells continuously interact.

Keywords: RNA networks; quasispecies; viruses; genomic parasites; cells; communication; disease

Introduction

Erwin Schrödinger held a lecture 75 years ago in Dublin entitled “What is life?” His answer as a physicist was “Life is physics and chemistry.” Later definitions of life in the 20th century assembled various descriptions based on a generally used materialistic paradigm determined by physico-chemical laws and theoretical systems underlying the main narratives, such as “life is a self-sustained chemical system capable of undergoing Darwinian evolution.”^{1,2} Another more recent definition is that organisms are open systems that maintain homeostasis, are composed of cells, have a life cycle, undergo metabolism, can grow, adapt to their

environment, respond to stimuli, reproduce, and evolve.⁴

With the rise of astrobiology and research on extraterrestrial life by NASA and the European Space Agency, the limitations of the definitions of life from the 20th century became obvious and their completeness more and more untenable.^{3,4} Is it still appropriate in the 21st century to look for only cell-based organisms or should we extend our perspective to integrate the essential RNA agents within the roots and stem of the tree of life that includes the neglected communicating features of RNA and viruses?^{5–7}

^aPortions of the article are taken from the authors’ previous papers.

RNA populations and their historical background

Any definition of life that lacks RNA functions and RNA networks misses these agents that inherently regulate and coordinate cells, tissues, organs, and organisms. Such definitions ignore the role of communication and networks. The precellular evolution of RNAs at the core of the evolution of cellular life is well documented, and the question remains how both levels are interconnected historically and functionally.^{8–10}

RNAs bind to other RNAs, DNA, and proteins.¹¹ In the early RNA world, proteins that bind to RNA served for RNA stability and therefore RNA function. RNAs such as the ribosomal, transfer, and messenger RNAs have long been known for their crucial roles in cellular replication processes through their helper functions in generating proteins out of DNA, the genetic storage medium. The evolutionary timeline of emergence ranges from the (co-opted and exapted) two halves of tRNAs, viroids, and RNA viruses, to the subunits of ribosomal RNAs and messenger RNAs.^{12–14} Essential features of such RNA structures are self-folding, catalysis, subunit assembly, association of large stem-loop groups, cooperative evolution, and surface proteinization.¹⁵ There is some evidence for direct codon (RNA loop)-to-amino acid interactions. This would indicate a pre-tRNA origin of a preliminary genetic code functioning as an emerging network of RNA–amino acid interactions.^{16–18}

Ribosomes are ribonucleoprotein complexes responsible for protein synthesis in all forms of life. They polymerize polypeptide chains templated by the nucleotide sequences in messenger RNA and mediated by transfer RNAs. Messenger RNA and transfer RNAs move rapidly through the ribosome while maintaining the translational reading frame. This process is accompanied by large- and small-scale conformational rearrangements in the ribosome, mainly in its rRNA.¹⁹ Interestingly, as essential agents of life, ribosomes are agents that may have self-assembled in an early RNA world, possibly out of prebiotic matter.^{20,21}

Noncoding RNAs can interact with and regulate each other through various molecular interactions based on complementary base-pairing of their nucleotides, generating a complex network including different species of RNAs (e.g., snRNAs, snoR-

NAs, mRNAs, miRNAs, lncRNAs, and circRNAs) and mobile genetic elements.^{22–25} Such a network of RNA–RNA (also competitive) interactions pervades and modulates the physiological functioning of canonical protein-coding pathways involved in proliferation, differentiation, and even metastasis in cancer. Interestingly, the noncoding RNAs in host genomes are more conserved than the protein coding sequences;²⁶ does this indicate that the regulatory functions are of greater importance than the proteins?

RNA interaction motifs: parasitizing, networking, immunity, and identity

A main characteristic of RNA identities is that the membership of RNA stem-loop groups can never be fully specified because it can always be (1) parasitized by yet unknown parasites or (2) internally reorganized and split into two from the original population. The most important consequence of this uncertainty is that it provides the inherent capacity for novelty, that is, the precondition for evolutionary innovation, such as new RNA groups and increased complexity.^{27–29}

If we look at some interacting motifs of RNA agents to form consortial biotic structures and share a functional identity, we must look at the formation of groups of RNA stem-loop structures.³⁰ It has been found that single stem-loops interact in a pure physicochemical mode without selective forces, independently of whether they are derived randomly or are artificially constructed.^{31,32} In contrast to this, if such single RNA stem-loops become part of group formation, they transcend purely physicochemical interaction patterns, and biological selection emerges. As a result, we can find biological identities capable of self/nonself-identification and preclusion, immune-like functions, and dynamically changing (adapting) membership/identity roles.³³ The primary emergence of biological selection occurred here.

Single alterations in a base-pairing RNA stem that leads to new loops or bulges may dynamically alter not only this single loop or bulge but may also change the whole group identity of which this section of single-stranded RNA is part.³⁴ This means that any RNA group transcribed out of the DNA storage medium may play a new, modular, unexpected role in cooperating or competing RNAs, genetic regulations, counter-regulations, and

diseases or infection events. Such noncoding RNAs most often derive from mobile genetic elements and are the major contributors to regulatory ncRNAs, which include endogenous retroviruses, their defective derivatives, and other persistent genetic parasites (not forgetting their degraded parts).^{35–40}

Another highly interesting interaction motif of RNAs is the RNA fragments that self-ligate into self-replicating ribozymes, which spontaneously form cooperative networks. It was found that the three-membered RNA networks showed highly cooperative growth dynamics. When such cooperative RNA networks compete directly against selfish autocatalytic cycles, the former grows faster. As a result, it was shown that RNA cooperation outcompetes selfishness.^{31,41–43}

RNA group behavior generated the origin of a natural code

One of the crucial key features of single-stranded RNAs is their tendency to fold back on themselves and form a double strand based on the complementary base pairing of the involved nucleotides.^{44–46} This provides the basis for the historical evolution of the so-called stem-loops, a double-stranded RNA stem with a single-stranded RNA loop (or bulge). In contrast to the double-stranded RNA region, which is not the primary source of complementary binding to other RNAs, the single-stranded loops and bulges are prone to such binding. They are the interaction centers with other stem-loop structures.³¹ Interaction will be more likely if repetitive nucleotide sequences cluster than if it occurs in nonrepetitive sequence order. Therefore, repetitive sequence order was a beneficial biological selection for the growth of sequence space in the early RNA world.^{47,48} In contrast, a nonrepetitive nucleic acid code evolved, which is mainly used to code for protein structures only. This may be consistent with the role of proteins in an early RNA world: to stabilize RNA structures and maintain their function. Later on, proteins were exapted to build cells, tissues, organs, and organisms. The tRNAs especially interconnect(ed) the RNA world with the protein world.^{49–51} Interestingly, primary and secondary structure analyses indicate a common ancestry for tRNAs and parasitic RNAs such as, for example, ribosomal viroids.^{52,53}

The chemical binding of nucleotides able to better bind repetitive and distributed loop

sequences can lead to an RNA that interacts with larger RNA groups that have this repetition. This can then allow for selection of those RNAs that could participate in and build more extended RNA interactions and identities. An RNA group identity is characterized as a newly emerged social property that can protect the identity and reject the nonidentity of individual RNA agents, that is, self/nonself-differentiation. This looks like a kind of RNA-sensing or -monitoring property and can explain RNA quasispecies preclusion.^{34,54}

But this social–chemical binding can add a completely new and crucial communication feature to the population of chemical molecules: a semiotic interaction has started that is absent from all other known molecular structures. A semiotically determined code means that besides the physicochemical boundaries, a sign-based interaction (semiosis) now takes place with at least the predominant interaction motifs (such as RNA group identity).⁵⁵ Natural semiotic codes normally start out of social interactions, which requires agents and out of which emerge the complementary functions of combinatorial rules (syntax), contextual rules (pragmatics), and content-relevant rules (semantics). This start of a natural semiotic code gave rise to the genetic identities most relevant for the evolution of viruses and, later on, cells.

The better-selected RNA groups then inherently parasitize weaker groups or solitary stem-loops, constantly producing new sequence space by spontaneously generating binding-prone bulges and loops. The repetitive structures form much better interactions with (and parasitize) the complementary repetitive structures than nonrepetitive ones. The infectious lifestyle of later-derived viruses emerged from such parasites. However, this interactional modus is not what we might call life, but any life depends on these social abilities and features. Thus, an initial repetitive sequence grammar emerges that characterizes the code with which RNAs “talk” to each other, which means they communicate to generate group behavior.

The genetic identities of RNA stem-loop groups (RNA networks), such as those from group I introns, group II introns, viroids, RNA viruses, retrotransposons, LTRs, non-LTRs, and subviral networks, such as SINEs, LINEs, and Alu elements, have all invaded and mostly persist in host genomes.^{56,57} They provide complex RNA-mediated networks. In

addition, mixed consortia of RNA and DNA virus-derived parts (especially those encoding stem-loop RNAs) also integrated into host genomes. The highly dynamic RNA–protein networks, such as ribosome, editosome, and spliceosome, generate a large variety of results and core functions out of DNA content.^{30,58,59} These are all examples of complex RNA group functions; and we may conclude here that without socially interacting RNAs, there is no effective (code-based) communication and no evolution of viruses or cellular life.

Key features of RNA group evolution

Let us, for example, look at a single stem-loop RNA within an RNA consortium that undergoes replication (perhaps for several rounds). Each replication event (necessarily being low fidelity) produces its own particular version of diversified progeny. Let us say, we get a new bulge in the stem. This bulge then becomes available to provide a whole array of possible outcomes (including counteracting ones). It might:

- interact with the original template to either complement or inhibit it;
- provide an interaction point for other RNA progeny (including itself);
- provide a target site for cleavage or ligation;
- act in combination with other progeny to provide a more complex catalytic (ribozyme) function; or
- alter or provide a binding site to other participants, such as peptides (RNPs).

In other words, it now has a whole array of possible (and multiple) usages (positive and negative). It is important to note that the actual use will depend on the context (circumstances and history) of the population it is in. Then, add to this all the other diverse progeny from these few rounds of replication, all in their own peculiar RNA region, and all providing their own peculiar potential for use. Such a scenario very rapidly becomes too complex to follow the fate (fitness and usage) of any particular RNA.

But, if we now think in social terms, then the RNA population (quasispecies consortia) can be considered as a “culture” that retains a common natural code (with repetitive grammar) that provides a level of group coherence (quasispecies selection).⁶⁰ Each individual diverse RNA then becomes like a potentially new “word” for that “language” represented by

an agent within its population. The culture of a certain RNA population is then free to use it (possibly even with multiple “meanings”) however it can or to reject it. And if this culture changes by building new bulges or loops, formerly rejected RNA stem-loops may later on fit (be reused) into the assemblage.⁶¹ These RNA uses will also vary considerably with the history of prior RNAs, as well as any possible interactions with other quasispecies consortia. And these uses can vary (and be lost) with time as the culture adopts new meanings.

Whether or not viruses predate cellular life does not alter the fact that some of the simplest RNA viruses are built up entirely of RNA stem-loops.

Viruses and their relatives

Viruses have long been assumed to descend from cells as escaped parasites since they are not able to self-replicate, and therefore are not living entities.⁶² But as documented by several authors, this is not correct. Numerous genes can be found in viruses without any relation to cellular genes. This indicates clearly that viruses must be older than cellular life.^{6,63,64} At least with the discovery of mimiviruses (and other giant viruses), it might be plausible that giant viruses have cellular origins.^{65–70}

We must constantly remember that viruses and subviral infectious genetic parasites are the most abundant biological agents on the planet.^{71,72} They outnumber cellular life forms by 10 times, invade all cellular organisms, and serve as key agents in the generation of adaptive and innate immune systems, which are essential for the survival of cellular life forms since they are key for the capacity for self/nonself-differentiation.^{73–75} The invasion strategy of genomic parasites that results in persistence within host genomes provides novel evolutionary genetic identities not present prior to the invasion.^{65,76} This is not error-dependent evolution of novelty.

Importantly, fragmented parasitic genetic elements can also provide an abundance of distributed mobile genetic elements able to create new RNA networks and be directly relevant in gene regulation in all organisms.^{77,78} Best documented of these are the persistent lifestyles of retroviruses as endogenous retroviruses and their defective derivatives, such as LTRs, SINES, LINEs, and Alu elements.⁷⁹

They share all the variants of genetic sequence syntax from RNA to DNA, from single- to

double-stranded, and from repetitive to non-repetitive sequence order, and only living entities have these features. But the virus-related RNA-dependent RNA polymerases, specific reverse transcriptases, and specific RNases share the key ability to transfer RNA into DNA and to insert RNA interaction motifs into the DNA storage library of the host.^{80–82} Other transmissible ribozymes such as group II introns are related to spliceosomes and retrotransposons, not to mention editosomes and ribosomal subunits.⁸³ Interestingly, the nucleolus, which is the generator of ribosomal RNA subunits, represents an RNA skeleton structure that is built out of noncoding RNAs interconnected by Alu elements, which are also remnants and fragmented elements of former viral infections.^{84–86} Many, if not all, viruses also produce noncoding RNAs,⁸⁷ which can provide a source of host and viral regulation. One can now understand why the nucleolus has prominent roles in development and aging, given its noncoding RNA composition.⁸⁸

Viruses and subviral infectious genetic parasites remain conserved biological identities, dating to before the origins of the extant domains of cellular life.^{52,53,65} They are characterized by a great ability to modulate genomic content and by permanent evolutionary adaptability. They may recombine not only with each other but also with those that have single- or double-stranded RNA or DNA genomes.^{89–91}

They retain the ability to bridge RNA and DNA living domains. In bridging the RNA- and protein-based cellular worlds, they are also the main drivers of the adaptability of cells by introducing RNA interaction motifs into cells.^{87,92,93} Viruses are also masters of co-opting cellular noncoding RNAs.⁹⁴

Viruses and subviral infectious genetic parasites thus represent biotic identities that are competent to edit code.⁹⁵ As editors, they can cooperate, build communities, generate nucleotide sequences *de novo* and insert/delete them into/from host genetic content (without damaging host genetic content), remain as mobile genetic elements (or similar “defectives”), build counterbalancing “addiction modules” (T/A and R/M),^{96,97} and determine host genetic identities throughout all kingdoms of life, including the virosphere. No other entities have such an expansive capacity to edit and create code.

Viruses and subviral infectious genetic parasites are the main drivers of speciation in evolution-

ary processes since the beginning of life. Most importantly, they determine host genetic identities throughout all the kingdoms of life through various techniques, such as generating and installing addiction modules persistently. All known species (even highly related ones) have distinct patterns of colonization by viruses and subviral infectious genetic parasites. And all species have their peculiar pattern of susceptibility to persistent viruses and subviral infectious genetic parasites, and also their corresponding acute versions. As an evolutionary rule, colonization created a distinct group identity that initiates the process of speciation.²⁷

The main behavioral motif that interconnects communication of RNA groups, viruses, and cells

The crucial question remains: How did evolution connect the RNA world of interaction motifs with cell-based life via infectious genetic parasites such as viruses and their relatives? As noted above, we can find an abundance of counterbalancing modules in cellular genomes, such as T/A, R/M, and various similar counterbalancing agents, that are inherently regulatory and are abundantly employed for immune functions against genetic parasites (although immune systems themselves are mostly constituted by such parasitic agents).^{98–100}

Cell-based organisms, be they prokaryotic or eukaryotic, represent rare islands in a sea (or population) of viruses, virus-like agents, and RNAs.¹⁰¹ This situation is a classic feature of life. This means that cells are a rare resource for competing genetic parasites, persistent genome settlers, or similar integration-driven behavioral motifs. The competition in most cases is a rather complex one between several virus populations that try to invade host genomes.¹⁰² Whereas one virus population may try to invade the cell, another population also tries to invade or protect the host from the opposing invading cloud. It is an arms race between the various virus populations and the immune system of the host, which constantly tries to react and oppose nonself-agents that themselves seek to adapt to the host immune response.

Within these close, dense, and in many cases, fast and unexpected interactions that follow agent colonization, the acquisition of identification (self/nonself) agents also promotes the acquisition and emergence of new addiction modules and

makes this an objective of the biological selection processes. At the end of this process, the host has changed its genetic identity through integration of several modules, including antagonizing agents, that enriches its own genetic identity but which is absent from cells of related species that were not subjected to the same invasion process.^{103,104} For example, the enrichment by restriction/modification or toxin/antitoxin modules which protects the new host does not protect related organisms that do not have such a TA module, which means that related species may now be killed by a toxin originating from an infectious agent.^{105,106} In the long run of evolution, this divides a species into two different genetic identities, the noninfected (and nonprotected) one and the infected (and protected) one.¹⁰⁷ This may lead to the start of two different lineages.

RNA-dependent RNA polymerases (RdRp) are very ancient enzymes and crucial players for all viruses with RNA genomes.¹⁰⁸ Together with reverse transcriptase and various RNases they promote and maintain RNA networks, and are essential agents in key cellular processes, including generating DNA sequences, a basic requirement for cellular life.¹⁰⁹ The reverse transcriptases are also closely related to those of the group II introns—but lacking the intronic RNA sequence structure—as well as being associated with a type of CRISPR/Cas system.¹¹⁰ The CRISPR/Cas system is an effective prokaryotic adaptive immune system descended from mobile genetic elements and most likely represents an exapted T/A module, as presented above.¹¹¹

Protein-based cells

Protein-based cells are metabolizing protein bodies that have membranes, genetic information that is inherited, and replication processes for reproduction. Their origin during evolution occurred after the emergence of RNA networks and virus-like structures.^{112,113} In contrast, the origin of cell-based life without RNA networks and viruses first seems to be impossible because all regulatory elements in cells depend on varying degrees of RNA functions.¹¹⁴ Cells represent identities in the living world of either prokaryotic populations—with their own mostly unicellular history and ecological niche construction—or eukaryotes, with their emergence, most often, of social cellular identities from formerly free-living prokaryotes. Both of these

cellular identities are protein-based, genetically conserved, and reproducible.¹¹⁵ This means the former determinants of RNA or virus-based life are now dominated by the identity of the protein-based cell. Yet, these protein-based cells retain their constant entanglement with the virosphere, which means that permanent infectious events and counter-defense actions involving persistent genetic parasites constantly calibrating immune functions against competing genetic parasites, both acting with noncoding RNAs to adapt or reject. Thus, protein-based cellular identities remain strongly influenced by viruses and subviral infectious genetic parasites.

Cells have long been assumed to be the basic entity of all life. According to Woese's categorization of life into three domains, the most primitive cells were archaea followed by bacteria and later on the eukarya emerged. Previously, it was the mainstream opinion that it took innumerable rounds of mutations and natural selection for eukaryotes to evolve. Such evolution results from small changes in structure and function that accumulate over very long periods of time. However, the serial endosymbiotic theory of Lynn Margulis presents a different evolutionary scenario.¹¹⁵ According to this narrative, it is not mutations in unicellular prokaryotes over very long time periods that led to the emergence of eukaryotes. Instead, it is via the formation of cooperative symbiotic networks from formerly free-living prokaryotes that resulted in unicellular eukaryotes. This view requires new networks to be formed via symbiosis.

The emergence of protein-based cellular life—which constitutes what has been termed “life” exclusively for many centuries—also provides a new phenotype, with completely new protein interaction motifs that were absent on abiotic planets as well as in the early RNA world. Because protein bodies also depend on RNA-mediated regulation, this vastly extends the interactions between constituents of cells and also diversified how cells experience life-world interactions. This had large consequences on niche constructions, the adaptation processes, and competing and cooperating populations. But all such interactions remain linked to an inherently stable genetic identity of the cellular organism. This cellular interacting life world is a new level that determines cellular life more or equally than the determination by genetic information, which now has to be coordinated in an appropriate way to survive.

Reproduction pathways include several steps and substeps of transcription, translation, repair, and immunity that are conserved in cellular evolution. But DNA alone does not specify cell fates. The crucial evolutionary benefit derived from RNA stem-loop networking and genetic parasites is the posttranscriptional (epigenetic) modifications that modulate genetic content into a dynamic and highly adaptive behavioral modification of the stored genetic information.^{116,117}

Context-dependent natural genome editing in cells

Although all cell types of any known organism contain the same 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.¹¹⁸ This means that depending on the context, such as a cell being located within an organ of an organism, the resulting expression leads to a new tissue- and site-specific pattern at the right time and in the right place to generate a specific cell type and chromatin state.^{119,120}

Chromatin marking enables a kind of identity programming.¹²¹ This means that a specific cell within an organism is able to obtain or even change its identity through epigenetics, if developmental, environmental, nutrition, or stress-related conditions make it necessary. Because RNAs are mobile, they can serve as signals throughout tissues, organs, and even the whole organism. In this respect, it is the imprinting of new experiences that leads to variable meanings of genetic information, depending on the action of noncoding RNAs. With epigenetic marking, life has an appropriate technique for the emergence of memory and learning processes for faster adaptation.^{122–125} Additionally, this epigenetic memory tool plays important roles in transgenerational inheritance, which also represents an important evolutionary function.^{126–128}

Both small RNAs and long noncoding RNAs are able to direct chromatin changes through histone modifications and DNA methylation. These noncoding 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 nonself-elements for transcriptional silencing or

elimination.^{129,130} 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.^{131,132}

We can look at the shared behavioral motif of how RNA can modify meaning out of a given DNA sequence syntax and also how RNA affects RNA editing.^{133,134} RNA editing is a co- or post-transcriptional process that alters the RNA sequence derived complementarily from the DNA from which it was transcribed. Before RNA editing, the editosome (i.e., small nuclear RNAs complexed with a variety of proteins) must be assembled in a strictly coordinated process. RNA editing changes gene sequences at the RNA level.¹³⁵ The edited mRNA specifies an amino acid sequence that is different from the protein that would be expected based on the encoding of the genomic DNA into the primary transcript.¹³⁶ RNA-editing alterations of such transcribed RNA sequences occur by modification, substitution, and insertion/deletion processes.¹³⁷ Editing sites have to be identified individually to differentiate the A, T, G, C to be edited from the A, T, G, C that should not be edited.¹³⁸ The discriminating information can be found in the nucleotide sequence surrounding a given site. This means that context is crucial for identification. Thus, each editing site carries its own identification context.¹³⁹

RNA editing predates splicing (also in evolution) and is temporally and functionally interconnected. The editosome and spliceosome are important interacting agents but whose assembly is dependent on editing.^{140–142} In ribosome and editosome assembly, and also in spliceosome assembly, construction of the needed ribonucleoprotein complex occurs via various steps that cut out introns and splice exons together. The spliceosomal ribonucleoproteins are mainly small nuclear RNAs complexed with at least 300 different proteins to form five spliceosomal subunits.

Interestingly, the variety of steps in which the subunits of the final spliceosome are produced are counterbalanced by (formerly) competing genetic parasites, all of them persistently integrated within the host genome. 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 multiuse modules. Depending on these regulatory events, the end product may vary concerning the context dependency of the resulting regulation process, which is highly sensitive to various needs and circumstances. Consequently, spliceosomal regulation differentiates the inclusion (splicing enhancers) or exclusion (splicing silencers) of exons in the final mRNA.¹⁴⁰ Splicing regulation occurs by competing *cis*-acting elements that precisely balance regulatory proteins.¹⁴³

Intron-exon genetic sequence construction in cells

One of the most interesting aspects of RNA editing is the complexity of genetic sequence construction, with genes that code for proteins and noncoding sections containing regulatory sequences, that is, the division of genetic information into introns and exons.¹⁴⁴ Previously, introns that do not code for proteins had been viewed as meaningless remnants of former evolutionary stages remaining in the host genome.¹⁴⁵ Later on, it became clear that cellular genomes represent a rather limited resource and most likely do not represent senseless sequences. With the resurgence of virology in host evolution, especially the focus on genome-invading genetic parasites with a persistent status (and their relatives), introns more and more seemed to represent former genetic parasites that are co-opted or exapted for cellular genetic functions.¹⁴⁶

To generate a coherent messenger RNA sequence for use in a cellular replication process, the lining up of the translationally relevant genetic sequences coding for a protein requires that the introns be removed and the remaining exons be ligated into a protein-coding sequence. By our perspective, this means that the identity of a protein-coding gene can be found only if the introns, which represent remnants of previous infection events by genetic parasites with repetitive sequence syntax, are removed. Otherwise, the gene coding for a protein cannot be produced coherently. This reminds us of the early self/nonself-differentiation competence of repetitive RNA, the forerunner of every evolutionarily derived immune function and the core behavioral motif needed to generate a genetic identity.¹⁴⁷ The repetitive sequences of introns must be removed to get the nonrepetitive sequences to line up properly in order to code for a protein. A very complex and

strict regulation must govern this division of functions in every replication process throughout the living world.¹⁴⁸

The repetitive sequences of the introns also represent the preferred target sites of genetic parasites to invade the host genome, whereas the nonrepetitive sequences that code for proteins normally are not damaged or deformed by genome invading agents.^{54,149–151} Intriguingly, intronic regions that do not code for proteins in many cases serve as a rich source of RNAs that are used for defense against transposable elements, indicating additional roles in self/nonself-discrimination.^{152–154}

Communication is the key

The core paradigmatic assumptions of the 20th century biology, including (1) the central dogma of molecular biology (DNA to RNA to proteins), (2) noncoding repetitive DNA is junk, and (3) the “one gene, one protein” hypothesis, have been falsified and no longer play important roles in the 21st century.^{155,156} A similar situation also applies to the core concepts seeking to explain the genetic code, RNAs, viruses, and cellular life. These used mathematically determined concepts such as cybernetic systems theory, information theory, biophysics, and derivative concepts in an attempt to explain living processes mechanistically. Such approaches have been uniformly lacking in essential progress to understand the complexity of interactional (communicating) patterns.^{29,157} For example, the self-reproducing machine has been announced for more than half a century but not a single self-replicating machine has been constructed, observed, or demonstrated as yet.

Although cell–cell communication and numerous signaling processes within cells have been well known for over half a century, the explanations for this communication were subsumed under their corresponding physicochemical properties as well as under information- and formal systems-theoretical explanatory attempts. The word “system” can confuse many readers. As currently accepted in science, we must keep in mind that if we define cells or even life as a system, we do not speak about empirical observations but of a cybernetic, systems-theoretical perspective that exchanges “cells” for “systems” in a systems-theoretical (mathematical) construct. It is important to remember this and not to confuse such a theoretical explanatory model with existing life

or even confuse the theoretical term with observed behavior.

This is not a minor disagreement in nuanced word usage. Key vocabulary that was used for the description of the essential activities of cellular life (at least in the last six decades), including terms such as “genetic code,” “genetic information,” “cell–cell communication,” “nucleotide sequences,” “protein-coding sequences,” and “self/nonself-recognition,” all are using vocabulary from linguistics and communication science, and not vocabulary from chemistry, physics, or mathematics. But biology, as well as physics, chemistry, and mathematics, has yet to integrate the results of the discourse in the philosophy of science about the topics of communication and meaning and therefore is lacking a clear expertise on understanding language and communication.¹⁵⁸

Communication primarily is a kind of social interaction

Communication designates social interaction. Socially interacting living agents need tools so that interaction may lead to the coordination and organization of common behaviors to reach goals. In contrast to physicochemical interactions on an abiotic planet, communicative interactions on biotic planets are mediated by signs. In cell-based organisms, such signs must be uttered by bodily expressed movements, phonetics, audiovisuality, tactility (e.g., vibrational), or semiochemical sensing.¹⁵⁸

Communication as rule-governed sign-mediated interactions is different from interactions in a purely physicochemical world without any biotic agents. Communicating living agents share a limited repertoire of signs that are used according to a limited number of rules that must be followed to generate correct sign sequences to designate context-dependent content. Most interesting is the fact that such rules—although rather conservative—may be changed in extreme cases or if adaptation is necessary. Rule-following by living agents is rather flexible, in contrast to the natural laws that living agents strictly abide by. This means that communication is the essential tool to generate new signs, sign sequences, new rules for sign use, and generation of new content according to unexpected contextual circumstances. Communicating living agents are able, in principle, to generate new communica-

tive patterns for better or innovative adaptation to a new and unforeseeable situation.¹⁵⁹

Communication in all domains of life

If communication is the key in the 21st century to understanding life, it must be possible to identify communicative actions throughout all domains of life.¹⁶⁰ Until the middle of the last century, language and communication were thought to be the special tools of only humans. Meanwhile, we know many examples of nonhuman languages and communication processes.^{161–166} Therefore, the description of communication processes must be valid in principle in all organisms, from the simplest akaryote up to humans. The main characteristics of communication, namely its (1) social character, (2) dependence on signs accordingly, and (3) the three kinds of rules (combinatorial, context-specific, and content-coherent), are not compatible with the decades-long narrative suggested by information theory and systems theory, that is, the sender–receiver narrative (coding-decoding), which was wrong in several respects.¹⁶⁷ All empirical data clearly show that communication is not only an information-transfer process, but also an interaction mediated by signs.

No natural code codes itself

As we know today, communicative properties and communicative interactive agents cannot be sufficiently described by physicochemical analyses and mathematical theories of language and communication (such as systems theory and information theories) because:

- the hidden, context-dependent deep grammar that finally determines the meaning of the sequence of superficial (visible) sign sequences cannot be identified, because it may vary according to its concrete contextual use, and
- the dependence of natural communication processes on social interacting agents is primarily a sociological expertise and not one of physics, chemistry, or mathematics.

It is empirically evident that natural codes and natural languages do not code themselves or speak themselves. In all scientific observations, it is evident that there must be agents that use and edit natural codes, such as the living agents that use and generate natural sign-based languages.¹⁵⁹ Code or language characters do not build sequences by

statistical ensemble mechanics, and the genetic code is not the result of a sequence of selected errors caused by mechanistic and thermodynamic conditions.⁶⁰ The same is true for RNA group constructions and interactions as well as viruses and their interaction motifs with cell-based life forms. The tools to construct appropriate concepts that can fully integrate present empirical data cannot be found in the 20th century.¹⁶⁸

If we look for a coherent explanation and understanding of life, we must add the communicative aspects of sign-mediated interactions in RNA networks, viruses, and cells, although the interactional patterns of these three levels are quite different historically and functionally.¹⁶⁹

- In RNA communication, RNA stem-loops are functioning as both catalytic drivers of reactions and as signs in sequences by themselves, representing nucleotide sequences mainly in a repetitive sequence order that is stabilized by its binding to protein structures (RNPs).
- Cellular life is organized and coordinated exclusively by sign-mediated interactions in the uptake, interpretation, and release of chemical substances. This means that cellular life constantly interacts during its typical life cycle by continuous communication processes using semiochemicals, that is, molecules that are produced, released, and taken up through signaling processes.
- Viruses and subviral infectious genetic parasites can do both of the above: (1) they generate, take up, and release semiochemicals for communication and they mimic host communication between themselves and the host organism, and (2) they function as catalytic units and sign sequences (viroids). They are the ideal intermediate between RNA networks and cells that have their own goals.

Disease: dysfunctional communication

Disease has long been assumed to be the result of mostly dangerous events for living organisms caused by abiotic or biotic influences that disturb or trigger a misleading mechanistic effect within the organism. In most cases, this is a rather complex event, because cascades of several distinct pathways are involved constantly and in collaboration with immune functions that try to restore health.

From the biocommunicative viewpoint, disease is the result of dysfunctional communication at one or more communication levels (i.e., RNA group communication, persistent virus communication, and cellular communication). As we have seen, these levels are rather complex and distinct from each other but are intertwined in all living organisms. RNA–RNA dysfunctions may be relevant to RNA–virus- and RNA–protein-based cell dysfunctional communication.

As we have seen, RNAs are key in nearly all functions of cell-based life. They have been integrated, combined, recombined, exapted, and co-opted by genetic parasites, such as viruses and their relatives (e.g., mobile genetic elements), which reached a persistent identity within host organisms. Persistence may be reached from fully functional viruses that have exapted old functional motifs for new traits—such as in the syncytin genes or the neuronal arc genes—to viral defectives, such as the great variety of transposable elements and other noncoding RNAs, all of which can be identified as repetitive sequence structures in either RNA or DNA.^{170–174}

Without a doubt, transposable elements represent key agents to change genome identities for various adaptive purposes in a rather flexible and module-based, and even social, manner. But this inherent capability to change genome integrity and sequence order with far-reaching consequences for the adaptability of the host organisms must be restricted to ensure an acceptable level of genome integrity (e.g., genome immunity). Mobile genetic elements otherwise may drive genome instability into deregulation of well-conserved cellular traits and may cause disease in various ways, which may lead to a paradoxical arms race between evolutionary genome plasticity and stability.¹⁷⁵

The crucial behavioral motif for such host integration events is the addiction module. As described above, we know them for counterbalancing (opposing) functions such as restriction/modification, toxin/antitoxin, insertion/deletion, amplification/silencing, endonuclease/ligase (anti-sense/sense), and various others, such as death/antideath programs.²⁹ The counterbalancing properties evolved by socially selective forces and represent astonishingly complex interaction motifs. For example, a bacterial genome with 51 restriction motifs is counterbalanced by 51 modification motifs.

A great problem in understanding life in all its complexity is its communication—the intertwined regulation of the detailed steps and substeps in transcription, translation, immunity, repair, replication, etc. Based on the three levels of communication described above, disease can be understood as dysfunctional communication in one or more levels that leads to a deregulation of the involved counterbalancing agents. This may lead to dysfunction on one level, which directs the nondysfunctional communication of another level into a wrong direction (i.e., death or uncontrolled growth), as occurs in cancer. For example, the same pathways used by trophoblasts in embryo uterus implantation can be used for metastasis in cancer cells in which those molecular pathways have lost their stop signal.^{176–178} We must therefore keep in mind the multiple roles of deregulated RNAs in cancer to understand the consequences of dysregulated RNA communication.^{179–181} From this perspective, errors in replication events may initiate such dysfunctional communication because it damages or deforms functional communication pathways.

Conclusions

Previously, the evolutionary beginning of life was defined by the emergence of self-replicating and metabolizing individual cells. Communication was simply a byproduct of physical interactions, and repetitive or parasitic DNA was simply junk. Now, we argue that life emerged from RNA networks capable of communication and self-identification via repetitive motifs. These networks underlie all extant life. They are constituted by three levels of interactions (see below). All cells are genetically regulated by RNA networks that were initially transferred into cellular host genomes as repetitive motifs via genome invading agents, such as viruses and subviral infectious genetic parasites (e.g., transposons). The persistence of genomic parasites results when the parasites introduce evolutionarily novel genetic identities absent before invasion (often via addiction modules). Formerly competing genetic parasites come together (via symbiosis) with host immune function to generate new regulatory tools that themselves are counterregulated. The three interaction levels are among (1) RNA groups, (2) viruses (both internal/external), and (3) cell-based organisms. These networks not only constitute identity and regulate phenotype, but are also inherently open

for subsequent parasite invasion that generates new, unpredictable, and thus noncomputable interaction profiles, such as *de novo* genetic sequences, new cooperation pathways, exaptation, and new traits from former parasite module-like parts that evolved for different purposes. Parasite-derived functions become degraded and available for reuse as regulatory modules and new behavioral motifs in all cellular life forms. If these conserved regulatory modules and their counterregulatory components become out of balance, dysfunctional communication takes place and disease may be the consequence.

Competing interests

The authors declare no competing interests.

References

- Luisi, P.L. 1998. About various definitions of life. *Orig. Life Evol. Biosph.* **28**: 613–622.
- Cleland, C.E. & C.F. Chyba. 2002. Defining 'life'. *Orig. Life Evol. Biosph.* **32**: 387–393.
- Des Marais, D.J., J.A. Nuth 3rd, L.J. Allamandola, *et al.* 2008. The NASA astrobiology roadmap. *Astrobiology* **8**: 715–730.
- Horneck, G., N. Walter, F. Westall, *et al.* 2016. AstRoMap European astrobiology roadmap. *Astrobiology* **16**: 201–243.
- Greener, M. 2008. It's life, but just as we know it. *EMBO Rep.* **9**: 1067–1069.
- Villarreal, L.P. & G. Witzany. 2010. Viruses are essential agents within the roots and stem of the tree of life. *J. Theor. Biol.* **262**: 698–710.
- Kolb, V. Ed. 2019. *Handbook of Astrobiology*. Boca Raton, FL: CRC Press.
- Mueller, S. 2015. Engineering of ribozymes with useful activities in the ancient RNA world. *Ann. N.Y. Acad. Sci.* **1341**: 54–60.
- Joyce, G.F. & J.W. Szostak. 2018. Protocells and RNA self-replication. *Cold Spring Harb. Perspect. Biol.* **10**. pii: a034801.
- Cech, T. 2018. A lifelong passion for all things ribonucleic. *Cell* **175**: 14–17.
- Mercer, T.R. & J.S. Mattick. 2013. Structure and function of long non-coding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **20**: 300–307.
- Maizels, N. & A.M. Weiner. 1994. Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl. Acad. Sci. USA* **91**: 6729–6734.
- Flores, R., S. Gago-Zachert, P. Serra, *et al.* 2014. Viroids: survivors from the RNA world? *Annu. Rev. Microbiol.* **68**: 395–414.
- Keam, S.P. & G. Hutvagner. 2015. tRNA-derived fragments (tRFs): emerging new roles for an ancient RNA in the regulation of gene expression. *Life (Basel)* **5**: 1638–1651.
- Draper, W.E., E.J. Hayden & N. Lehman. 2008. Mechanisms of covalent self-assembly of the *Azoarcus* ribozyme from

- four fragment oligonucleotides. *Nucleic Acids Res.* **36**: 520–531.
16. Yarus, M., J.J. Widmann & R. Knight. 2009. RNA-amino acid binding: a stereochemical era for the genetic code. *J. Mol. Evol.* **69**: 406–429.
 17. Yarus, M. 2017. The genetic code and RNA-amino acid affinities. *Life (Basel)* **7**: 13.
 18. Zagrovic, B., L. Bartonek & A.A. Polyansky. 2018. RNA–protein interactions in an unstructured context. *FEBS Lett.* **592**: 2901–2916.
 19. Petrov, A.S., B. Gulen B, A.M. Norris, *et al.* 2015. History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. USA* **112**: 15396–15401.
 20. Gwiazda, S., K. Salomon, B. Appel, *et al.* 2012. RNA self-ligation: from oligonucleotides to full length ribozymes. *Biochimie* **94**: 1457–1463.
 21. Agmon, I.C. 2016. Could a proto-ribosome emerge spontaneously in the prebiotic world? *Molecules* **21**. pii: E1701.
 22. Smith, C.M. & J.A. Steitz. 1997. Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell* **89**: 669–672.
 23. Cech, T.R. & J.A. Steitz. 2014. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **157**: 77–94.
 24. Witzany, G. 2011. The agents of natural genome editing. *J. Mol. Cell Biol.* **3**: 181–189.
 25. Kapusta, A., Z. Kronenberg, V.J. Lynch, *et al.* 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**: e1003470.
 26. Villarreal, L.P. 2004. Are viruses alive? *Sci. Am.* **291**: 100–105.
 27. Villarreal, L.P. 2009. *Origin of Group Identity: Viruses, Addiction and Cooperation*. New York, NY: Springer.
 28. Villarreal, L.P. 2011. Viruses and host evolution: virus-mediated self-identity. In *Self and Non-Self*. C. Lopez-Larrea, Ed.: 185–217. Austin: Landes Bioscience and Springer Science-Business Media.
 29. 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.
 30. Witzany, G. 2014. RNA sociology: group behavioral motifs of RNA consortia. *Life (Basel)* **4**: 800–818.
 31. Vaidya, N., M.L. Manapat, I.A. Chen, *et al.* 2012. Spontaneous network formation among cooperative RNA replicators. *Nature* **491**: 72–77.
 32. Yeates, J.A.M., P. Nghe & N. Lehman. 2017. Topological and thermodynamic factors that influence the evolution of small networks of catalytic RNA species. *RNA* **23**: 1088–1096.
 33. Koonin, E.V. & M. Krupovic. 2018. The depths of virus exaptation. *Curr. Opin. Virol.* **31**: 1–8.
 34. Villarreal, L.P. & G. Witzany. 2015. When competing viruses unify: evolution, conservation, and plasticity of genetic identities. *J. Mol. Evol.* **80**: 305–318.
 35. Smit, S., Yarus, M. & R. Knight. 2006. Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. *RNA* **12**: 1–14.
 36. Manrubia, S.C. & C. Briones. 2007. Modular evolution and increase of functional complexity in replicating RNA molecules. *RNA* **13**: 97–107.
 37. Geuking, M.B., J. Weber, M. Dewannieux, *et al.* 2009. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science* **323**: 393–396.
 38. Rowe, H.M. & D. Trono. 2011. Dynamic control of endogenous retroviruses during development. *Virology* **411**: 273–287.
 39. Villanueva-Cañas, L., J. Ruiz-Orera, M.I. Agea, *et al.* 2017. New genes and functional innovation in mammals. *Genome Biol. Evol.* **9**: 1886–1900.
 40. Rezelj, V.V., L.I. Levi & M. Vignuzzi. 2018. The defective component of viral populations. *Curr. Opin. Virol.* **33**: 74–80.
 41. Higgs, P.G. & N. Lehman. 2015. The RNA world: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**: 7–17.
 42. Gebetsberger, J., L. Wyss, A.M. Mleczko, *et al.* 2017. A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biol.* **14**: 1364–1373.
 43. Ariza-Mateos, A & J. Gómez. 2017. Viral tRNA mimicry from a biocommunicative perspective. *Front. Microbiol.* **8**: 2395.
 44. Briones, C, M. Stich & S.C. Manrubia. 2009. The dawn of the RNA world: toward functional complexity through ligation of random RNA oligomers. *RNA* **15**: 743–749.
 45. Lincoln, T.A. & G.F. Joyce. 2009. Self-sustained replication of an RNA enzyme. *Science* **323**: 1229–1232.
 46. Joyce, G.F. 2009. Evolution in an RNA world. *Cold Spring Harb. Symp. Quant. Biol.* **74**: 17–23.
 47. Jurka, J., V.V. Kapitonov, O. Kohany, *et al.* 2007. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **8**: 241–259.
 48. Robertson, M.P. & G.F. Joyce. 2012. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**. pii: a003608.
 49. Weiner, A.M. & N. Maizels. 1987. tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl. Acad. Sci. USA* **84**: 7383–7387.
 50. Bernhardt, H.S. & W.P. Tate. 2010. The transition from noncoded to coded protein synthesis: did coding mRNAs arise from stability-enhancing binding partners to tRNA? *Biol. Direct.* **5**: 16.
 51. Kanai, A. 2015. Disrupted tRNA genes and tRNA fragments: a perspective on tRNA gene evolution. *Life (Basel)* **5**: 321–331.
 52. Seligmann, H. & D. Raoult. 2016. 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.* **31**: 1–8.
 53. Seligmann, H. & D. Raoult. 2018. Stem-loop RNA hairpins in giant viruses: invading rRNA-like repeats and a template free RNA. *Front. Microbiol.* **9**: 101.
 54. Witzany, G. 2017. Two genetic codes: repetitive syntax for active non-coding RNAs; non-repetitive syntax for the DNA archives. *Commun. Integr. Biol.* **10**: e1297352.
 55. Witzany, G. 2016. Crucial steps to life: from chemical reactions to code using agents. *BioSystems* **140**: 49–57.
 56. Lambowitz, A.M. & S. Zimmerly. 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **3**: a003616.

57. McNeil, B.A., C. Semper & S. Zimmerly. 2016. Group II introns: versatile ribozymes and retroelements. *Wiley Interdiscip. Rev. RNA* **7**: 341–355.
58. Lykke-Andersen, J., M.D. Shu & J.A. Steitz. 2001. Communication of the position of exon–exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293**: 1836–1839.
59. Pawlicki, J.M. & J.A. Steitz. 2010. Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function. *Trends Cell Biol.* **20**: 52–61.
60. Villarreal, L.P. & G. Witzany. 2013. Rethinking quasispecies theory: from fittest type to cooperative consortia. *World J. Biol. Chem.* **4**: 79–90.
61. Vaidya, N., S.I. Walker & N. Lehman. 2013. Recycling of informational units leads to selection of replicators in a prebiotic soup. *Chem. Biol.* **20**: 241–252.
62. Moreira, D. & P. Lopez-Garcia. 2009. Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* **7**: 306–311.
63. Krupovič, M. & D.H. Bamford. 2010. Order to the viral universe. *J. Virol.* **84**: 12476–12479.
64. Forterre, P. & D. Prangishvili. 2013. The major role of viruses in cellular evolution: facts and hypotheses. *Curr. Opin. Virol.* **3**: 558–565.
65. Villarreal, L.P. 2005. *Viruses and the Evolution of Life*. Washington, DC: ASM Press.
66. Colson, P. & D. Raoult. 2012. Megavirales composing a fourth domain of life: Mimiviridae and Marseilleviridae. In *Viruses: Essential Agents of Life*. G. Witzany, Ed.: 217–242. Dordrecht: Springer.
67. Backstrom, D., N. Yutin, S.L. Jorgensen, *et al.* 2018. Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. <https://doi.org/10.1101/469403>.
68. Nasir, A., K.M. Kim & G. Caetano-Anollés. 2017. Phylogenetic tracings of proteome size support gradual accretion of protein structural domains and the early origin of viruses from primordial cells. *Front. Microbiol.* **8**: 1178.
69. Abrahão, J., L. Silva, L.S. Silva, *et al.* 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* **9**: 749.
70. Colson, P., A. Levasseur, B. La Scola, *et al.* 2018. Ancestrality and mosaicism of giant viruses supporting the definition of the fourth TRUC of microbes. *Front. Microbiol.* **9**: 2668.
71. Rohwer, F., M. Youle, H. Maughan, *et al.* 2014. *Life in Our Phage World*. San Diego, CA: Wholon.
72. Cobián Güemes, A.G., M. Youle, V.A. Cantú, *et al.* 2016. Viruses as winners in the game of life. *Annu. Rev. Virol.* **3**: 197–214.
73. Villarreal, L.P. 2009. The source of self: genetic parasites and the origin of adaptive immunity. *Ann. N.Y. Acad. Sci.* **1178**: 194–232.
74. Villarreal, L.P. 2011. Viral ancestors of antiviral systems. *Viruses* **3**: 1933–1958.
75. Koonin, E.V. & M. Krupovic. 2015. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.* **16**: 184–192.
76. Magiorkinis, G., R.J. Gifford, A. Katzourakis, *et al.* 2012. Env-less endogenous retroviruses are genomic super-spreaders. *Proc. Natl. Acad. Sci. USA* **109**: 7385–7390.
77. Feschotte, C. & C. Gilbert. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13**: 283–296.
78. Koonin, E.V. & M. Krupovic. 2017. Polintons, virophages and transpovirons: a tangled web linking viruses, transposons and immunity. *Curr. Opin. Virol.* **25**: 7–15.
79. Gifford, R.J., J. Blomberg, J.M. Coffin, *et al.* 2018. Nomenclature for endogenous retrovirus (ERV) loci. *Retrovirology* **15**: 59.
80. Witzany, G. 2008. The viral origins of telomeres, telomerases and their important role in eukaryogenesis and genome maintenance. *Biosemiotics* **1**: 191–206.
81. Moelling, K. & F. Broecker. 2015. The reverse transcriptase-RNase H: from viruses to antiviral defense. *Ann. N.Y. Acad. Sci.* **1341**: 126–135.
82. Moelling, K., F. Broecker, G. Russo, *et al.* 2017. RNase H as gene modifier, driver of evolution and antiviral defense. *Front. Microbiol.* **8**: 1745.
83. Dong, X., S. Ranganathan, G. Qu, *et al.* 2018. Structural accommodations accompanying splicing of a group II intron RNP. *Nucleic Acids Res.* **46**: 8542–8556.
84. Weber, S.C. & C.P. Brangwynne. 2012. Getting RNA and protein in phase. *Cell* **149**: 1188–1191.
85. Carmo-Fonseca, M. 2015. Assembly of the nucleolus: in need of revision. *EMBO J.* **34**: 2731–2732.
86. Caudron-Herger, M., T. Pankert, J. Seiler, *et al.* 2015. Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J.* **34**: 2758–2774.
87. Steitz, J., S. Borah, D. Cazalla, *et al.* 2011. Noncoding RNPs of viral origin. *Cold Spring Harb. Perspect. Biol.* **3**: pii: a005165.
88. Tikku, V. & A. Antebi. 2018. Nucleolar function in lifespan regulation. *Trends Cell Biol.* **28**: 662–672.
89. Diemer, G.S. & K.M. Stedman. 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Direct.* **7**: 13.
90. Stedman, K.M. 2013. Deep recombination: RNA and ssDNA virus genes in DNA virus and host genomes. *J. Mol. Evol.* **76**: 359–364.
91. Berliner, A.J., T. Mochizuki & K.M. Stedman. 2018. Astro-virology. Viruses at large in the universe. *Astrobiology* **18**: 207–223.
92. Guo, Y.E. & J.A. Steitz. 2014. Virus meets host microRNA: the destroyer, the booster, the hijacker. *Mol. Cell Biol.* **34**: 3780–3787.
93. Tycowski, K.T., E.G. Yang, N. Lee, *et al.* 2015. Viral noncoding RNAs: more surprises. *Genes Dev.* **29**: 567–584.
94. Heaton, N.S. & B.R. Cullen. 2017. Viruses hijack a long non-coding RNA. *Nature* **552**: 184–185.
95. Witzany, G. 2006. Natural genome-editing competences of viruses. *Acta Biotheor.* **54**: 235–253.
96. Iranzo, J., M. Krupovic & E. Koonin. 2017. A network perspective on the virus world. *Commun. Integr. Biol.* **10**: e1296614.
97. Harms, A., D.E. Brodersen, N. Mitarai, *et al.* 2018. Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol. Cell* **70**: 768–784.

98. Kobayashi, I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* **29**: 3742–3756.
99. Gerdes, K. & E.G. Wagner. 2007. RNA antitoxins. *Curr. Opin. Microbiol.* **10**: 117–124.
100. Mruk, I. & I. Kobayashi. 2014. To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res.* **42**: 70–86.
101. Ryan, F. 2009. *Virovolution*. London: Collins.
102. Mitchell, R.S., B.F. Beitzel, A.R.W. Schroder, *et al.* 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2**: e234.
103. Lehnherr, H., E. Maguin, S. Jafri, *et al.* 1993. Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. *J. Mol. Biol.* **233**: 414–428.
104. Villarreal, L.P. 2012. The addiction module as a social force. In *Viruses: Essential Agents of Life*. G. Witzany, Ed.: 107–145. Dordrecht: Springer.
105. Aizenman, E., H. Engelberg-Kulka & G. Glaser. 1996. An *Escherichia coli* chromosomal “addiction module” regulated by guanosine 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc. Natl. Acad. Sci. USA* **93**: 6059–6063.
106. Hazan, R., B. Sat, M. Reches, *et al.* 2001. Postsegregational killing mediated by the P1 phage “addiction module” phd-doc requires the *Escherichia coli* programmed cell death system mazEF. *J. Bacteriol.* **183**: 2046–2050.
107. Yahara, K., R. Horie, I. Kobayashi, *et al.* 2007. Evolution of DNA double-strand break repair by gene conversion: coevolution between a phage and a restriction-modification system. *Genetics* **176**: 513–526.
108. de Farias, S.T., A.P. Dos Santos Junior, T.G. Rêgo, *et al.* 2017. Origin and evolution of RNA-dependent RNA polymerase. *Front. Genet.* **8**: 125.
109. Moelling, K., F. Broecker & J.E. Kerrigan. 2014. RNase H: specificity, mechanisms of action, and antiviral target. *Methods Mol. Biol.* **1087**: 71–84.
110. Toro, N., F. Martínez-Abarca, A. González-Delgado, *et al.* 2018. On the origin and evolutionary relationships of the reverse transcriptases associated with type III CRISPR-Cas systems. *Front. Microbiol.* **9**: 1317.
111. Koonin, E.V. & K.S. Makarova. 2017. Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome Biol. Evol.* **9**: 2812–2825.
112. Koonin, E.V., T.G. Senkevich & V.V. Dolja. 2006. The ancient virus world and evolution of cells. *Biol. Direct.* **1**: 29.
113. Koonin, E.V., V.V. Dolja & M. Krupovic. 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* **479–480**: 2–25.
114. Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**: 397–405.
115. Witzany, G. 2006. Serial endosymbiotic theory (set): the biosemiotic update. *Acta Biotheor.* **54**: 103–117.
116. Villarreal, L.P. & G. Witzany. 2013. The DNA habitat and its RNA inhabitants. At the dawn of RNA sociology. *Genomics Insights* **6**: 1–12.
117. Boivin, V., G. Deschamps-Francoeur & M.S. Scott. 2018. Protein coding genes as hosts for noncoding RNA expression. *Semin. Cell Dev. Biol.* **75**: 3–12.
118. Shapiro, J.A. 2014. Epigenetic control of mobile DNA as an interface between experience and genome change. *Front. Genet.* **5**: 87.
119. Slotkin, R.K. & R. Martienssen. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**: 272–285.
120. Turner, B.M. 2002. Cellular memory and the histone code. *Cell* **111**: 285–291.
121. Barlow, D.P. 2011. Genomic imprinting: a mammalian epigenetic discovery model. *Annu. Rev. Genet.* **45**: 379–403.
122. Mattick, J. 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.
123. Blaze, J. & T.L. Roth. 2012. Epigenetic mechanisms in learning and memory. *Wiley Interdiscip. Rev. Cogn. Sci.* **4**: 105–115.
124. Marshall, P. & T.W. Bredy. 2016. Cognitive neuroepigenetics: the next evolution in our understanding of the molecular mechanisms underlying learning and memory? *NPJ Sci. Learn.* **1**: 16014.
125. Kim, S. & B.K. Kaang. 2017. Epigenetic regulation and chromatin remodeling in learning and memory. *Exp. Mol. Med.* **49**: e281.
126. Rassoulzadegan, M. & F. Cuzin. 2015. Epigenetic heredity: RNA-mediated modes of phenotypic variation. *Ann. N.Y. Acad. Sci.* **1341**: 172–175.
127. Klosin, A., E. Casas, C. Hidalgo-Carcedo, *et al.* 2017. Transgenerational transmission of environmental information in *C. elegans*. *Science* **356**: 320–323.
128. Gagliano, M., F. Baluska & G. Witzany, Eds. 2018. *Memory and Learning of Plants*. Dordrecht: Springer.
129. Lee, K.J., A.B. Conley, V.V. Lunyak, *et al.* 2012. Do human transposable element small RNAs serve primarily as genome defenders or genome regulators? *Mob. Genet. Elements* **2**: 1–7.
130. Mallet, P.L., M. Larochelle & F. Bachand. 2017. Multiple transcriptional and post-transcriptional pathways collaborate to control sense and antisense RNAs of Tf2 retroelements in fission yeast. *Genetics* **205**: 621–632.
131. Huda, A. & I.K. Jordan. 2009. Epigenetic regulation of mammalian genomes by transposable elements. *Ann. N.Y. Acad. Sci.* **1178**: 276–284.
132. Huda, A., L. Mariño-Ramírez & I.K. Jordan. 2010. Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. *Mob. DNA* **1**: 2.
133. Göringer, H.U., Ed. 2008. *RNA Editing*. Berlin, Heidelberg: Springer.
134. Gommans, W.M., S.P. Mullen & S. Maas. 2009. RNA editing: a driving force for adaptive evolution? *Bioessays* **31**: 1137–1145.
135. Homann, M. 2008. Editing reactions from the perspective of RNA structure. In *RNA Editing*. H.U. Göringer, Ed.: 1–32. Berlin, Heidelberg: Springer.

136. Knisbacher, B.A. & E.Y. Levanon. 2015. DNA and RNA editing of retrotransposons accelerate mammalian genome evolution. *Ann. N.Y. Acad. Sci.* **1341**: 115–125.
137. Smith, H.C. 2008. Editing informational content of expressed DNA sequences and their transcripts. In *RNA Editing*. H.U. Göringer, Ed.: 249–265. Berlin, Heidelberg: Springer.
138. Yablonovitch, A.L., P. Deng, D. Jacobson, *et al.* 2017. The evolution and adaptation of A-to-I RNA editing. *PLoS Genet.* **13**: e1007064.
139. Gott, J.M. 2003. Expanding genome capacity via RNA editing. *C. R. Biol.* **326**: 901–908.
140. Matlin, A.J. & M.J. Moore. 2007. Spliceosome assembly and composition. *Adv. Exp. Med. Biol.* **623**: 14–35.
141. House, A.E. & K.W. Lynch. 2008. Regulation of alternative splicing: more than just the ABCs. *J. Biol. Chem.* **283**: 1217–1221.
142. Yang, L. 2015. Splicing noncoding RNAs from the inside out. *Wiley Interdiscip. Rev. RNA* **6**: 651–660.
143. Pyle, A.M. & A.M. Lambowitz. 2006. Group II introns: ribozymes that splice RNA and invade DNA. In *The RNA World*. 3rd ed. R.F. Gesteland, T.R. Cech & J.F. Atkins, Eds.: 469–506. New York, NY: Cold Spring Harbor Laboratory Press.
144. Mattick, J.S. 1994. Introns: evolution and function. *Curr. Opin. Genet. Dev.* **4**: 823–831.
145. Steitz, J.A. 1988. “Snurps”. *Sci. Am.* **258**: 56–60.
146. Lambowitz, A.M. & M. Belfort. 1993. Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62**: 587–622.
147. Marraffini, L.A. & E.J. Sontheimer. 2010. Self versus non-self discrimination during CRIPR RNA-directed immunity. *Nature* **463**: 568–571.
148. Tycowski, K.T., M.D. Shu & J.A. Steitz. 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**: 464–466.
149. Lambowitz, A.M. & S. Zimmerly. 2004. Mobile group II introns. *Annu. Rev. Genet.* **38**: 1–35.
150. Lewinski, M.K., M. Yamashita, M. Emerman, *et al.* 2006. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* **2**: e60.
151. Simon, D.M., N.A. Clarke, B.A. McNeil, *et al.* 2008. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA* **14**: 1704–1713.
152. Cowley, M. & R.J. Oakey. 2013. Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet.* **9**: e1003234.
153. Boivin, V., G. Deschamps-Francoeur, S. Couture, *et al.* 2018. Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. *RNA* **24**: 950–965.
154. Lewis, S.H., K.A. Quarles, Y. Yang, *et al.* 2018. Panarthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat. Ecol. Evol.* **2**: 174–181.
155. Shapiro, J.A. 2009. Revisiting the central dogma in the 21st century. *Ann. N.Y. Acad. Sci.* **1178**: 6–28.
156. Mattick, J.S. 2009. Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Ann. N.Y. Acad. Sci.* **1178**: 29–46.
157. Witzany, G. 2015. Life is physics and chemistry and communication. *Ann. N.Y. Acad. Sci.* **1341**: 1–9.
158. Witzany, G. 2010. *Biocommunication and Natural Genome Editing*. Dordrecht: Springer.
159. Witzany, G. 2014. Language and communication as universal requirements for life. In *Astrobiology: An Evolutionary Approach*. V. Kolb, Ed.: 349–369. CRC Press.
160. Witzany, G. 2016. Key levels of biocommunication. In *Biocommunication: Sign-Mediated Interactions between Cells and Organisms*. R. Gordon & J. Seckbach, Eds.: 37–61. Singapore: World Scientific.
161. Witzany, G., Ed. 2011. *Biocommunication in Soil Microorganisms*. Heidelberg: Springer.
162. Witzany, G., Ed. 2012. *Biocommunication of Fungi*. Dordrecht: Springer.
163. Witzany, G. & F. Baluska, Eds. 2012. *Biocommunication of Plants*. Berlin, Heidelberg: Springer.
164. Witzany, G., Ed. 2014. *Biocommunication of Animals*. Dordrecht: Springer.
165. Witzany, G. & M. Nowacki, Eds. 2016. *Biocommunication of Ciliates*. Dordrecht: Springer.
166. Witzany, G., Ed. *Biocommunication of Archaea*. Dordrecht: Springer.
167. 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.
168. Shapiro, J.A. 2017. Living organisms author their read-write genomes in evolution. *Biology (Basel)* **6**: 42.
169. Villarreal, L.P. & G. Witzany. 2018. Editorial: genome invading RNA networks. *Front. Microbiol.* **9**: 581.
170. Pérot, P., P.A. Bolze & F. Mallet. 2012. From viruses to genes: syncytins. In *Viruses: Essential Agents of Life*. G. Witzany, Ed.: 325–361. Dordrecht: Springer.
171. Meyer, T.J., J.L. Rosenkrantz, L. Carbone, *et al.* 2017. Endogenous retroviruses: with us and against us. *Front. Chem.* **5**: 23.
172. Erez, Z., I. Steinberger-Levy, M. Shamir, *et al.* 2017. Communication between viruses guides lysis-lysogeny decisions. *Nature* **541**: 488–493.
173. Häfner, S.J., T.G. Talvard & A.H. Lund. 2017. Long noncoding RNAs in normal and pathological pluripotency. *Semin. Cell Dev. Biol.* **65**: 1–10.
174. Pastuzyn, E.D., C.E. Day, R.B. Kearns, *et al.* 2018. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell* **173**: 275.
175. Ernst, C., D.T. Odom & C. Kutter. 2017. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nat. Commun.* **8**: 1411.
176. Kämmerer, U., A. Germeyer, S. Stengel, *et al.* 2011. Human endogenous retrovirus K (HERV-K) is expressed in villous and extravillous cytotrophoblast cells of the human placenta. *J. Reprod. Immunol.* **91**: 1–8.

177. Grandi, N. & E. Tramontano. 2018. HERV envelope proteins: physiological role and pathogenic potential in cancer and autoimmunity. *Front. Microbiol.* **9**: 462.
178. Balestrieri, E., A. Argaw-Denboba, A. Gambacurta, *et al.* 2018. Human endogenous retrovirus K in the crosstalk between cancer cells microenvironment and plasticity: a new perspective for combination therapy. *Front. Microbiol.* **9**: 1448.
179. Gingold, H., D. Tehler, N.R. Christoffersen, *et al.* 2014. A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**: 1281–1292.
180. Barriocanal, M. & P. Fortes. 2017. Long non-coding RNAs in hepatitis C virus-infected cells. *Front. Microbiol.* **8**: 1833.
181. Costanzo, V., A. Bardelli, S. Siena, *et al.* 2018. Exploring the links between cancer and placenta development. *Open Biol.* **8**. pii: 180081.