Genomics Insights

PERSPECTIVE

The DNA Habitat and its RNA Inhabitants: At the Dawn of RNA Sociology

Luis P. Villarreal[1] and Guenther Witzany[2]

[1]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, USA. [2]Telos–Philosophische Praxis, Buermoos, Austria. Corresponding author email: witzany@sbg.at

Abstract: Most molecular biological concepts derive from physical chemical assumptions about the genetic code that are basically more than 40 years old. Additionally, systems biology, another quantitative approach, investigates the sum of interrelations to obtain a more holistic picture of nucleotide sequence order. Recent empirical data on genetic code compositions and rearrangements by mobile genetic elements and noncoding RNAs, together with results of virus research and their role in evolution, does not really fit into these concepts and compel a reexamination. In this review, we try to find an alternate hypothesis. It seems plausible now that if we look at the abundance of regulatory RNAs and persistent viruses in host genomes, we will find more and more evidence that the key players that edit the genetic codes of host genomes are consortia of RNA agents and viruses that drive evolutionary novelty and regulation of cellular processes in all steps of development. This agent-based approach may lead to a qualitative RNA sociology that investigates and identifies relevant behavioral motifs of cooperative RNA consortia. In addition to molecular biological perspectives, this may lead to a better understanding of genetic code evolution and dynamics.

Keywords: agents, genetic content, RNA stem loops, quasispecies, social membership

Genomics Insights 2013:6 1–12

doi: 10.4137/GEI.S11490

This article is available from http://www.la-press.com.

Introduction: Conceptual Realms of the Past Decades and Current Empirical Data

For at least 4 decades, it was common use in molecular biology to investigate the physical chemical properties of nucleic acid sequence molecules. It was basically the concepts of physicians and chemists such as Max Delbrück or Manfred Eigen that shaped paradigmatic horizons and oriented the concepts of the following generations in molecular biology. The essential assumptions were that natural laws determine statistical fluctuations in the molecular sequence order and errors in cellular genetic replications lead to variants in DNA sequences that are the object of selection processes.[1] This successful model of molecular biology was coherent with more theoretical assumptions of modern evolutionary synthesis, that random mutations are the main reason for genetic variations.[2–4] In the 1980s, increasing amounts of empirical data did not fit into this concept:

- Barbara McClintock's thoughts and experiments found mobile genetic elements that actively influence the shape of genomes and their internal interactions.[5,6]
- The central dogma of molecular biology "DNA-RNA-Protein everything else," is obsolete nowadays; the reverse direction is also in common use.[7]
- The core concept of DNA biology, that DNA determines protein information and some remnants of former evolutionary stages remained as useless junk DNA in the genomes of cell-based organisms is also outdated. The genes coding for proteins in the human genome represent 1.5% of the total, whereas the number of noncoding DNA reaches 98.5%.[8,9] Noncoding RNAs are actively engaged in nearly all cell processes for both evolutionary and developmental needs.
- The detection of a third domain of life by Carl Woese was a very important step in the development of our picture on the animate world, as was the serial endosymbiotic theory of Lynn Margulis, ie, cooperation of merged members constituting eukaryotic cells with a true nucleus.[10,11]
- Viruses and their various roles in evolution and developmental processes as obligate inhabitants of cellular genomes change our view on the evolution and function of the nucleic acid language and its interconnection with the amino acid language.[12,13] In line with this thinking, some authors suggest the existence of a fourth domain of life, the virosphere.[14]
- If we look at the flexible and dynamic expansion of information processes through epigenetic programming and reprogramming, alternative splicing and RNA-editing,[15–18] we must confess today that investigations of the genomic contents of organisms with the molecular biology tools of the last 5 decades have yielded more questions than answers.
- Investigations on spatiotemporal coordination between cells, organs, and organisms in all domains have shown that communication processes are the core reason for this coordinative ability.[19]

Interestingly, there are some approaches that try to integrate this new data into research on RNA populations and viruses because it allows a more appropriate description and understanding of their interactions.[20–23]

In this review, we therefore focus on some aspects of the consortia networks of viruses and subviral agents: RNA stem-loop swarms and viral and subviral networks (SINEs, LINEs, Alus) that cooperate and coordinate (regulate) within cellular genomes as infection-derived modular tools such as noncoding RNAs. Some noncoding RNAs built complementary consortia such as rRNAs, tRNAs, spliceosomes, editosomes, toxin/antitoxin modules, restriction/modification modules, and insertion/deletion modules. Such networks determine regulation complexity in various ways and may coadapt to different functions than they originally evolved for.[24–27]

Key Players that Organize Genetic Content Composition

Today we can identify several consortia of key players that coordinate and organize the genetic content compositions of host organisms: endogenous viruses and defectives, transposons, retrotransposons, long terminal repeats, nonlong terminal repeats, long interspersed nuclear elements, short interspersed nuclear elements, group I introns, group II introns, phages, and plasmids are currently investigated examples that use genomic DNA as their preferred living habitat.[24,28,29] This means that DNA is not solely a genetic storage

medium that serves as an evolutionary protocol but is also a species-specific ecological niche.[30–32] A great variety of such mobile genetic elements have been identified during the last 40 years as obligate inhabitants of all genomes, either prokaryotic or eukaryotic.[33,34] They infect, insert, and delete; some cut and paste, others copy and paste, and both spread within the genome. They change host genetic identities either by insertion, recombination, or the epigenetic regulation or re-regulation of genetic content and coevolve with the host and interact in a modular manner.[33–35]

Together with noncoding RNAs they shape both genome architecture and regulation. In this respect, they are not only agents of change over evolutionary time but also in real time as domesticated agents.[36]

Noncoding RNAs interact with DNA, RNA, and proteins and play important roles in nuclear organization, transcription, posttranscriptional and epigenetic processes. Noncoding RNAs are transcribed in both the sense and antisense directions and may be expressed in a cell type, subcellular compartment, developmental stage, or environmental stimuli-specific, that is, context dependent manner. Specific RNA polymerases overlap in transcriptional contents, which means that each nucleotide can participate in varying transcriptional content arrangements according to varying contexts.[37]

Noncoding RNAs can be regulated in a varying manner, coordinated or independently, 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 specificity. Various classes of noncoding RNAs target each other for posttranscriptional regulation via alternative splicing, polyadenylation, 5′ capping, nontemplated modifications, and RNA editing. RNA editing especially can transmit environmental information to the epigenome and therefore enable neuronal plasticity with learning and memory. Additionally, noncoding RNAs can undergo nuclear-cytoplasmic, nuclear-mitochondrial, and axodendritic trafficking via ribonucleprotein complexes that promote the spatiotemporal distribution and function of various combinations of ncRNAs, mRNAs, and RNA binding proteins.[37]

Long ncRNAs are transcribed from intergenic regions in antisense, overlapping, intronic, and bidirectional orientations relative to protein coding genes. Long noncoding RNAs can integrate temporal, spatial, developmental, and stimulus-specific (contextual) cellular information. Small ncRNAs target RNAs in posttranscriptional regulation via RNAi. Such small ncRNAs include miRNAs, siRNAs, and piRNAs. Interestingly, we also find them in archaeal ribonucleoprotein complexes.[38]

Stem loops: basic tools of active RNAs

One of the most interesting discoveries in the last decades was certainly the fact that nucleic acid sequences are not only information bearing molecules that serve as a read-only storage medium, but they can be actively modified by several coplayers.[7] Additionally, we mentioned that DNA centrism, which investigates DNA as the dominant genetic information storage medium, might be wrong because the older RNAs are the dominant interactors as soon as transcription starts.[39] The ancient RNA world hypothesis is filled up nowadays with current RNA world facts and increasing knowledge about the abundance of different but compatible RNAs, most if not all of them in coacting processes via cooperation and suppression as well as amplification.[40]

In this world of life processes actively dominated by RNA, DNA is increasingly cast in the role of the "house" of genetic information storage, whereas the interacting RNAs seem to be the "social network" within this house.[30,32,41]

If we look at the "body" of these social networks, we can see a secondary structure that is shared by all these RNA nucleic acid sequences: it is the stem-loop structure or, in the case of more complex agents such as tRNA or ribosomal subunits, the ligated consortia of such stem loops. The RNA stem loops have several distinct parts/subunits: stems consisting of base-paired nucleic acids and loops/bulges/junctions consisting of unpaired regions limited by stems. Whereas the paired regions have a one-to-one ratio of pyramidines and purines, this is not the case in the loop structures of the unpaired regions. Interestingly, the different parts of the stem loops also have distinct base compositions.[42] This kind of variation in the base composition is similar in the two consortia of stem loops of the two ribosomal subunits throughout all domains of life, although evolutionarily divergent.

A rather astonishing conclusion of these facts is that randomly associated RNAs that have no evolutionary history show the same structure-dependent compositional bias as ribosomal RNAs. This means that the differences do not depend on selection processes but on the overall composition of the RNA consortium.[42,43]

Function, not sequence, determines composition

Additionally, we must consider a rather important fact, namely, that each RNA nucleotide of any natural RNA sequence is part of one of these secondary structure motifs: stem, loop, bulge, junction or end, or another unpaired region.

This means that the composition of RNA bases follows the stem-loop architecture as an appropriate tool, to store information and build more complex ensembles that together act catalytically as ribozymes. This results out of comparisons between natural RNA sequences and randomized artificial RNA sequences that indicate RNA sequence compatibility.[42,44,45] These compatible stem loops not only comprise the agents of the current RNA world but also include all forms of RNA viruses. Their modular "tool-like" composition opens a wide range of flexible compositional patterns that are combined via RNA self-ligation into a great variety of complementary interacting RNA structures such as subunits of ribosomes and tRNAs,[46,47] group I introns, group II introns, telomerase RNA, micro RNAs, and RNA viruses.

This means the stem-loop tool—with its base-pairing and nonpairing composition—is relevant to all agents of the current RNA world being active in cellular genomes after transcription.[48] They are not only remnants of the ancient RNA world but ongoing infection-derived modular tools for new cellular needs, that is, "domesticated" consortia of interacting RNA agents such as viruses, mobile genetic elements, and noncoding RNAs.[49–52] Whereas the single stem-loop emergence seems to not be subjected to selection processes, the consortia that are built with these tools are subjected to selection processes.

Emergence of RNA identities and the original roles of tRNA

To go deeper into a world of self-replicating RNAs prior to cellular organisms, we can start with forerunners of the tRNAs. Regarding RNAs and genetic language, we should think on Tomasello's investigations about the origin of the human language out of gestures inherent in social interactions.[53] Basically, a symbolic code must emerge from a non-symbolic or more operational instruction. Following this thinking regarding tRNA and the triplet genetic code, we can propose that the original role for a tRNA-like molecule was "operational" in that it was part of a ribozyme that also provided self-priming and self-identity for related replicators.[54–56] With the emergence of parasitic replicators, however, the network would be under selective pressure to further differentiate self identity. Thus, the covalent attachment of a 5′ amino acid (or peptide) to either the sense or antisense RNA would function as an additional primer-identity marker for the RNA, setting up selection for amino acid coding. Eventually, continued selective pressure by parasitic replicators that adapted to these enhanced covalent changes led to the emergence of additional symbolic information whose role was no longer operational in an enzymatic sense, but needed to specify network membership.[13]

The former dualism of information-bearing molecules that must be translated into executive molecules changes into an all-in-one perspective. tRNAs and other noncoding RNAs originally emerged as operational stem-loop consortia, but the selective pressure through replicative parasites is constantly high, and the attachment of amino acids as additional "primer-identity" markers leads to an amino acid code coherent to the nucleic acid language, possibly explaining the emergence of protein modules that stabilize RNA structures such as ribonucleoproteins.[57,58]

The emergence of enzymatic nonoperational sequence structures (symbolic) to specify network membership increases evolutionary potentials geometrically not arithmetically. It allows sets of relatively small agents to retain common behaviors and identities (such as replication). Thus, it now seems plausible that the basic coding of proteins and their variable regulations (conserved in DNA) are the collective results of an abundance of complementary RNA agents.

These RNA agents provide an unlimited potential for new productivity since there is no fundamental limit to be further parasitized by additional agents. From this perspective, the accumulation of junk

DNA would indeed scale with accumulating RNA society complexity. Thus, we must look at organisms as phenomena in which complex and dynamic "protein bodies" are regulated by RNA agents that build sequences that have new protein meanings and can act as consortial ribozymatic agents to create networks. The highly stable DNA cellular world is then an omnipresent, comfortable, chemically stable but enzymatically vacant habitat. In this scenario, self-replicating RNA agents or self-replicating viruses would remain as evolutionary curiosities because of the energy costs needed to dynamically maintain themselves.

It was proposed some time ago that tRNA evolution was from two different stem loops.[54] It now appears to us that two RNA replication primers (identifiers) merged at one time, and that amino acid or peptide marking of RNA was an early chemical system for providing replication chemical priming and identification, as were ribonucleoproteins. SINEs evolution is often considered from the perspective of a tRNA-like structure. But if we instead focus on the stem loops (not tRNA) as more fundamental and accept that the pre-tRNA stem-loop structural tag is really the main feature driving evolution of these structural RNAs,[57] then we can also apply this thinking to more recent retrotransposons such as SINEs and LINEs. The stem-loop primers of these agents have become a central component for agent identity. These structures (not sequences) can also be conditional (via pseudoknots, etc.). Thus, their link to hybrid dysgenesis or their presence as RNPs in the brain[58] can be better understood as agents of a network identity. Of course, all members would be expected to operate via a consortial quasispecies network identity that uses addiction modules to set group identity.[34]

An addiction module defines infection-derived viral encoded features that formerly competed for insertion and later on became counterbalancing (mutually dependent) features within the host organism. We can find them as antagonizing modules such as insertion/deletion, toxin/antitoxin, restriction/modification, SINE, LINE, and Alu sequences. In most cases, such addiction modules shape the genomic identity of the host. If one balanced part is weakened (eg, via stress, environmental influences, or infection), the other part can get out of balance and cause disease or lethal consequences.[34,35,59,60] Single addiction modules represent successfully domesticated (adapted) RNA consortia.

7S RNA ancestors of Alu and interactions with retroviruses

To go one step further with this, lets continue with 7S RNA as the likely ancestor of Alu elements.[61] All domains of life have their own peculiar and related 7S like RNAs, but only eukaryotes and some prokaryotes have the Alu domain (stem loops) at the 5′ end.[62] This RNA appears to have been present in an RNA quasispecies-like world with conserved domains of dsRNA and bound proteins. Clearly it established a DNA habitat along with the invention of DNA as the cellular genome. It assembles into an RNP and has two known activities: getting cleaved by various nucleases and affecting translation and transport of "marked" proteins across membranes to the "exterior" or endoplasmic reticulum in eukaryotes. This latter function is called signal transduction.

However, we suggest a different idea: 7S RNA serves as a molecular identity that initially provided ribosome-to-protein identity. It is a function (addiction or immune module) that will eliminate (export) proteins that are not marked as members of the proper ribosomal "cytoplasm." But it can also have domains such as Alu that are targets for endonuclease. With the further evolution of prokaryotes, it became adapted for protein export. This transport function has been retained in all life forms, although in eukaryotes, 7S RNA was also adapted to pol II transcription control.[63]

Viral quasispecies as precoordinated networks of integrating agents brought Alu

The integration of infecting agents into host genomes is not a randomly occurring process but needs the features and competences of consortia of RNA stem loops, that is, what Eigen termed quasispecies populations.[1] Most events in vertebrate genome colonization (ERVs, LINEs, Alus) must have been via sexual process involving colonization of the early embryo (which maintains ERV and LINE activity). Necessarily this needs a network approach to understanding the role of these agents in evolution (especially their regulatory complexity). In the human situation, ERVs, LTRs, LINEs, Alu elements, and most probably other

human specific viruses all need to be considered as basically interacting.[13]

Fundamentally, the new network was most likely initiated by HERV embryo colonization, but this necessarily involved LINE-1 and Alu elements working both to support and oppose the change. Alu elements alone are estimated to have inserted into 75% of all human genes.[64] This network is expected to have derived from addiction module action.[12,60]

The resulting changes were predominantly in regulatory regions (introns, 5′ noncoding, 3′ noncoding). These changes had to superimpose a diffuse new network onto already existing identity networks (via small RNA-based consortia of innate immunity). Only a viral quasispecies population-based process can provide such a diffuse but precoordinated network of integrating RNA agents. Point selection alone cannot achieve this. Since all retroviruses preferentially package 7S RNPs into virions (ie, how 7S RNA was discovered), they are the natural providers of new Alu subsequences. This packaged 7S RNA is efficiently reverse transcribed via alternative tRNA primers.[61] The product of the RT reaction is a subregion of the 7S RNA that is only slightly larger than the Alu subdomain.

However, sequence analysis of Alu integrations does not indicate that LTRs were directly involved. Instead, LINE-1 RT seems to have mediated Alu integration.[65,66] LINE-1 often interrupt ERVs and are highly and specifically induced and amplified in early embryos[67–69] and can downregulate HERV expression.[70] Early embryos are also active for Alu and LTR mediated expression.[71,72] Thus, if functional LINEs are responding to embryonic retroviruses, they can provide RT for Alu elements' subdomain integration coming from a retrovirus.

Normally, such complex interactions and genome colonizations result in a more complex antiviral or antigenetic parasite system of identity.[59] The resulting Alu integrations should have accordingly provided such a new lineage specific system of RNA-based identity. But this new identity system acts mainly via antisense miRNA network[73,74] targets that the Alu elements have provided to the 3′ ends of numerous transcripts.[73] Normally, cells have only small amounts of pol III transcribed "sense" Alu RNAs in their cytoplasm (ie, Alu Y). However, once virus infected, these levels go up dramatically and can also be virus regulated.[75] The expression of Alu transcripts (often edited, hence generating mismatches), will provide the cytoplasm with both sense and antisense Alu elements that should result in dsRNA that will promote inhibition of translation. Like interferon, this would be an antiviral response.

It is thus interesting that these transcribed Alus are mostly of the young class, found predominantly in the Y chromosome[76,33] and associated with L1 LINEs.[77] We also know Alu ingression into other chromosomes of the genome seems to be male mediated.[78] In addition, the Y chromosomes of humans and chimps differ markedly not only due to Alu elements[79] but also to the large sets of tandem ERV integration events (HERV-W and HERV-K), which is the most distinct genetic difference between all human and chimp chromosomes.[78,33] In contrast, LINE-1 is not well represented in the Y chromosome. Thus, the concerted actions of at least these three families of RNA agents underlies much of the regulatory changes associated with recent human evolution. In addition, these Alu elements also now provide very handy signals for genomic content editing (recombination/deletion/insertion/point change),[80,37] which is expressed in early human embryo development.[71]

Origin of Social Group Behavior: Role of Member Agents

If we want to better understand interactional motifs of RNA stem-loop swarms and viruses, we can add sociological group membership features that are absent in the inanimate world. Modelling collective behavior has long been left to mathematical sociology trying to predict social behaviors (culture, behavioral norms) using various differential equations. Evolutionists abandoned the idea of group selection decades ago.[81,82] These social models are derived from both the individual and the collective perspective. However, emergent behaviors such as learning are not well explained by any such models. This deficit is similar to systems biology, which fails because one cannot deduce the emergence of new qualities merely out of an increasing complexity of quantities.[83,84]

Group membership (identity) is an issue that is seldom if ever considered in any of the above models.[34] A main assumption of most models is that self similarity will determine many group behaviors and membership. It might be helpful if social science–like

models instead approach an empirical concept based on communication: a key competence of social-bound individuals is their common language use in communicative interactions. These are the primary tools for promoting active coordination and cooperation. Unfortunately, current social science approaches are limited to humans or at best cellular organisms, such as animals, plants, fungi, and bacteria.[85–88] Agents that have nucleic acid sequence structures solely have not yet been considered. Such agent-based approaches integrate communicative and linguistic competences of individual member activities and consortia activities in evolutionary and developmental processes, in contrast to ontological conceptions such as reductionism or holism.

RNA consortia: networks, systems, social membership?

In current biology, a network (eg, the nodes in the net) is not evaluated from the perspective of agents. Individual participants are simply "inert" nodes in an interacting net. Proteomes, transcriptomes, and interactomes are all more or less presented as such simple collections in a systems biology view of such ensembles. It is the nature of the interaction that is specified (positive, negative, etc.). Social membership is not considered. Nor is the emergence of complex behavior or phenotype crucial. Interactions within and between networks simply must emerge from individual fittest type origins as suggested in former quasispecies concepts.[1,89,90] There are no group behavioral issues or principles that must be adhered to. The collective has no group fitness even though it can show emergent features. All this can be modelled by "game theory" (eg, self similar kin selection).

However, because reductionist approaches do not well explain emergent consortia or group behaviors, systems biology tried a more holistic approach to explain properties that emerge out of complex systems. Like systems theory, which investigates the capacity of formal systems, systems biology defines a system as a quantity of elements and a quantity of relations between these elements. Both assume that the relations between the elements of a system and its possibilities of behavior can be represented formally (mathematically) without respect to any kind of realization (circumstances, history). This means that the dynamic relations, as well as the quantities of elements that constitute these relations, are subject to formalizable (computational) procedures such as algorithms.

Systems biology tries to explain the emergence of new properties by the sum of interactions of its parts. But natural languages/codes and the historically grown real-life worlds of the code-using agents cannot be formalized or represented by computational procedures because no algorithm is available for coherent de novo generation of common behavior or commonly shared meanings of sign sequences. In systems biology, the single member and its crucial membership behavior based on social competence and active engagement is lost. Changes in group identity depend on individual membership (adaptational) change, but in holistic approaches, individual member behavior is determined by the interrelations of the whole.[83,84] Because in real-life worlds, the groups also depend on the active (learning) engagement of competent individuals, systems biology cannot identify these competencies. To understand the active engagement of single members with social competence within groups, we have to develop a qualitative approach that is not available with the above quantitative procedures.

RNA social membership is essential but basically indefinable

However, we can consider these same "systems" issues from the perspective of "agents."[91–93] A system then becomes a set of agents. The nature of "agent" interactions (behavior) will determine many "system" features. How then do we specify "set" membership? It has been observed that an RNA consortium can become "bound" to itself and preclusive of past members, as seen in quasispecies of RNA virus.[90]

Although set membership can be outlined to a large degree by various features (such as replicator activity/identity), it can never be completely specified, since it can always be further parasitized by as yet to be encountered members or parasites.[34] This feature renders the ability to absolutely specify membership (absolute immunity) as basically indefinable. Thus, a network can never be fully secure from as yet undefined parasite agents. But a crucial inference from this "insecurity" is that it provides the inherent capacity for novelty, that is, the precondition for greater complexity.

The core issue is thus to specify how "agents" emerge from chemicals (ribozymes) to form identity (such as for replicators) and then form groups that learn membership. For all extant life, these agents must have initially been RNA stem loops. Single RNA stem-loop generation occurs by physical chemical properties solely as demonstrated by natural and randomized RNA experiments.[42,44,45] If stem-loop consortia build complex consortia they initiate social interactions not present in a pure chemical world. The emergence of identity (self vs. nonself) is a crucial step from single RNA stem loops to RNA consortia; maybe the crucial step from inanimate world to life? This line of thought leads to the following relevant questions:

- What does it mean to be an RNA agent of a group?
- What are the relevant behavioral rules for an agent?
- When does a peptide also become a member of an RNA stem-loop society?
- What is needed to be a member of an RNA quasispecies collective?
- Is this membership always a dynamic state?
- Did the invention of DNA "lockdown" this dynamic and promote the individual fittest type?
- How does this dynamic system emerge from the "rules" of behavior?
- How do new "rules" emerge?

What are the behavioral motifs of active RNA social membership

An individual member must be coherent with the group. In groups that constitute by self/nonself differentiation, the exclusion processes are executed rather strictly by immune functions, for example, toxin/antitoxin modules, CRISPRs, restriction/modification.[34,59,94] The collective set of positive and negative interactions (identity/addiction) all must allow membership. A member may also be able to use what other members provide (cooperation between nonself similar agents). Members possess some behaviors that allow and maintain membership and the collective support of survival.

But membership is also dynamic, since the collective changes. This means a fittest type individual becomes a nonmember if it fails to keep up with membership changes (ie, does not participate in shared common experience by learning). The concept of virus involves the concept of transmission. Some of the fundamental questions for a "virus" to become an agent are as follows:

- Does membership "spread" from one consortia to another?
- Is it contagious?
- Can this spreading be modelled or is it too dependent on the specifics of the new "host"?
- Is this learning and transmission solely a chemical feature?

In quasispecies the ensembles of RNA stem loops (ribozymatic hairpin structures) represent not solely content that rather specify self from nonself, but in parallel—and this truly seems the start of living processes—they act as agents, obedient to chemical/physical laws and, in contrast to interactions in the inanimate world, to syntactic, pragmatic, semantic rules. It is exactly this latter feature that is absent in inanimate nature.[19] And exactly this feature is an inherent characteristic for the communicative and linguistic competencies of living agents in biocommunication processes throughout all kingdoms.[85–88]

Now we have molecular structures coherent with physical laws that store genetic information and actively generate behavioral motifs and patterns of interaction, that is, coordinate common behavior according to rules that lead to consortia of self and nonself groups. This resembles some kinds of social group behavior with shared features:

- bottom-up processes tolerant of and needing diversity
- spontaneous formation of group behavior
- de novo initiation of behavior that cannot be deduced from former behavioral patterns
- highly adaptive processes
- lacking central or fittest type control
- retaining a contextual history
- smart (optimal energy costs)
- solves problems beyond the capacity of its individual members
- fast changing reactions against nonmembers

Together these features are clearly and exclusively at the bottom of all living nature. If we were to eliminate these complementary competences

out of the life processes, would there remain a living organism, or would it now simply be a chemical state? It seems not; thus, social RNA agents are essential.

At the Dawn of a Qualitative RNA Sociology

If we take now a step away from mainstream molecular biology (including the physical chemical properties of quasispecies and mutant spectra) and systems biology (nonreductionistic physicalism) and move toward a more consortial thinking of cooperative ensembles of RNA stem loops, viruses, and subviral remnants of former infection events, we should move toward a kind of RNA sociology. Qualitative RNA sociology emphasizes understanding of "social phenomena" through analysis of active RNA sequences according to contextual needs, that is, the primacy of pragmatic interactional motifs, meaning functions prior to sequence syntax. In this respect the exaptations and domestications of former (infection derived) inventions to more appropriate needs fit ideally into these investigations, that being the highly dynamic modular perspective on these interacting agents. A qualitative RNA sociology additionally investigates the roles of viruses as direct descendants and mediators of RNA societies.

As mentioned in the introduction, the main reason for the difficulties in stepping from the old molecular biology concepts to current knowledge about the dynamics of genome biology is that the genetic code as a natural language was explained by inherent physical and chemical properties of matter. The meaning (semantics) of the information stored in the genetic code was assumed to be the result of selection of a varying molecular syntax.[95]

The crucial deficit in this assumption was ignorance of the agent perspective: no natural code codes itself, as no natural language speaks itself. In natural languages or codes, there must be living agents that generate signs and combine them into more complex sign sequences.[96] The living agents are involved in the context (pragmatics) of natural ecospheres and share social life worlds to coordinate and organize survival within a history. The concrete life world context in which living agents are interwoven determines the meaning (semantics) of the signs. In addition to these 3 levels of rules (syntax, pragmatics, semantics), in natural languages/codes a further most important precondition is necessary: it is not possible for only one agent to follow a rule only once. Rule-following basically is a kind of social interaction within consortia or between consortia.[96] Subsequently applied to RNA biology, an RNA sociology may lead to a better understanding of roles of RNA consortia in evolution and development.

Conclusion

The initial quasispecies and hypercycle concept of Manfred Eigen sought to connect physics (quantum mechanics) to Fisher's equation for natural selection through stochastic alterations ("error copies") of previously selected sequences ("master copies") that compete against their master copies. This highlights only a small fraction of the sources of genetic variation, via randomly derived changes due to error copying, such as mutations. This approach was not more helpful in identifying relevant agents that drive the evolution of new sequences and sequence space because it neglected the major source of genetic variation that derives from viral and viral-related infections that can alter the genetic order and even provide the dynamic regulations of genes. This is a completely different source of information than errors of copying. By trying to link these equations, Eigen also bypassed the more fundamental (and ongoing) requirement for consortial action of RNA agents in the emergence of life. In this respect, Eigen tried to deduce the emergence of life as a systems feature of self-replication of matter, that is, matter alone is able to emerge information.

In contrast to this, we have tried to exemplify the multiple roles of consortia of ribozymatic agents emerging from compatible stem-loop structures that work together to build RNA viruses and create infection-derived noncoding RNA networks without which life cannot function. The basic behavioral competences of these agents are absent in inanimate nature and resemble core sets of behavioral rules that are fundamentally shared by RNA agents but yet may change or adapt to varying adaptational purposes. This explains their module like character.

In this respect, we propose the alternative concept of qualitative RNA sociology that underlies the communicative competences of consortia of RNA stem-loop structures in constituting life.

Author Contributions

Wrote the first draft of the manuscript: LPV and GW. Contributed to the writing of the manuscript: LPV and GW. Agree with manuscript results and conclusions: LPV and GW. Jointly developed the structure and arguments for the paper: LPV and GW. Made critical revisions and approved final version: LPV and GW. All authors reviewed and approved of the final manuscript.

Funding

Author(s) disclose no funding sources.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Eigen M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*. 1971;10:465–523.
2. Mayr E. *The Growth of Biological Thought: Diversity, Evolution and Inheritance.* Cambridge, MA: Harvard University Press; 1992.
3. Beurton P, Falk R, Rheinberger HJ, editors. *The Concept of the Gene in Development and Evolution*. Cambridge, UK: Cambridge University Press; 2000.
4. Kutschera U. A comparative analysis of the Darwin-Wallace papers and the development of the concept of natural selection. *Theor Biosci*. 2003;122: 343–59.
5. Frost LS, Leplae R, Summers AO, et al. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol*. 2005;3:722–32.
6. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*. 2007;8:272–85.
7. Shapiro JA. *Evolution: A View From the 21st Century*. Upper Saddle River, NJ: FT Press Science; 2011.
8. Volff JN. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays*. 2006;28:913–22.
9. Mattick JS. Deconstructing the dogma. A new view of the evolution and genetic programming of complex organisms. *Ann NY Acad Sci*. 2009;1178: 29–46.
10. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci*. 1990;87:4576–9.
11. Sagan L. On the origin of mitosing cells. *J Theor Biol*. 1970;14:255–74.
12. Villarreal LP. *Viruses and the Evolution of Life*. Washington, DC: ASM Press; 2005.
13. Villarreal LP. Viruses and host evolution: virus-mediated self identity. *Adv Exp Med Biol*. 2012;738:185–217.
14. Witzany G, editor. *Viruses: Essential Agents of Life*. Dordrecht, Netherlands: Springer; 2012.
15. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33:245–54.
16. Zuckerkandl E, Cavalli G. Combinatorial epigenetics, 'junk DNA', and the evolution of complex organisms. *Gene*. 2007;390:232–42.
17. Matlin A, Moore MJ. Spliceosome assembly and composition. *Adv Exp Me Biol*. 2007;623:14–35.
18. Gott JM. Expanding genome capacity via RNA editing. *CR Biol*. 2003; 326:901–8.
19. Witzany G. *Biocommunication and Natural Genome Editing.* Dordrecht, Netherlands: Springer; 2010.
20. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009;136:215–33.
21. Breaker RR. Riboswitches and the RNA world. In: Gesteland RF, Cech Tr, Atkins JF, editors. *The RNA World*, 3rd ed. New York, NY: Cold Spring Harbor Laboratory Press; 2006:89–107.
22. Brosius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene*. 1999;238:115–34.
23. Forterre P, Prangishvili D. The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann NY Acad Sci*. 2009;1178:65–77.
24. Osborn AM, Boltner D. When phage, plasmids, and transposons collide: genomic islands, and conjugative—and mobilizable—transposons as a mosaic continuum. *Plasmid*. 2002;48:2002–12.
25. Cech TR, Moras D, Nagai K, et al. The RNP world. In: Gesteland RF, Cech Tr, Atkins JF, editors. *The RNA World*, 3rd ed. New York, NY: Cold Spring Harbor Laboratory Press; 2006:309–26.
26. Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet*. 2006;15: 17–29.
27. Jurka J, Kapitonov VV, Kohany O, et al. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genom Human Genet*. 2007;8:241–59.
28. Shapiro JA. Repetitive DNA, genome system architecture and genome reorganization. *Res Microbiol*. 2002;153:447–53.
29. Lambowitz AM, Zimmerly S. Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb Perspect Biol*. 2011;3(8):a003616.
30. Brookfield JFY. The ecology of the genome- Mobile DNA elements and their hosts. *Nat Rev Genet*. 2005;6:128–236.
31. Vetsigian K, Goldenfeld N, Woese C. Collective evolution and the genetic code. *Proc Natl Acad Sci*. 2006;103:10696–701.
32. Le Rouzic A, Dupas S, Capy P. Genome ecosystem and transposable elements species. *Gene*. 2007;390:214–20.
33. Baluška F. Cell-cell channels, viruses and evolution: via infection, parasitism and symbiosis toward higher levels of biological complexity. *Ann NY Acad Sci*. 2009;1178:106–19.
34. Villarreal LP. *Origin of Group Identity: Viruses, Addiction and Cooperation*. New York, NY: Springer; 2009.
35. Villarreal LP. The source of self. Genetic parasites and the origin of adaptive immunity. *Ann NY Acad Sci*. 2009;1178:194–232.
36. Mattick JS. Video Q&A: Non-coding RNAs and eukaryotic evolution—a personal view. *BMC Biol*. 2010;8:67.
37. Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci*. 2012; 13:528–41.

38. Dennis PP, Omer A. Small non-coding RNAs in Archaea. *Curr Opin Microbiol*. 2005;8:685–94.
39. Wolf YI, Koonin E. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Dir*. 2007;2:14.
40. Atkins JF, Gesteland RF, Cech TR, editors. RNA Worlds. From Life's Origin to Diversity in Gene Regulation. *Cold Spring Harbor Laboratory Press*, New York; 2011.
41. Vennera S, Feschotte C, Biemonta C. Transposable elements dynamics: toward a community ecology of the genome. *Trends Genet*. 2009;25: 317–23.
42. Smit S, Yarus M, Knight R. Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. *RNA*. 2006;12:1–14.
43. Przybilski R, Hammann C. The tolerance to exchanges of the Watson Crick base pair in the hammerhead ribozyme core is determined by surrounding elements. *RNA*. 2007;13:1625–30.
44. Gevertz J, Gan HH, Schlick T. In vitro RNA random pools are not structurally diverse: A computational analysis. *RNA*. 2005;11:853–63.
45. Schultes EA, Spasic A, Mohanty U, Bartel DP. Compact and ordered collapse of randomly generated RNA sequences. *Nat Struct Mol Biol*. 2005;112:1130–6.
46. Gwiazda S, Salomon K, Appel B, Müller S. RNA self ligation: From oligonucleotides to full length ribozymes. *Biochimie*. 2012;94:1457–63.
47. Müller S, Appel B, Krellenberg T, Petkovic S. The Many Faces of the Hairpin Ribozyme: Structural and Functional Variants of a Small Catalytic RNA. *IUBMB Life*. 2012;64:36–47.
48. Witzany G. Non-coding RNAs: persistent viral agents as modular tools for cellular needs. *Ann NY Acad Sci*. 2009;1178:244–67.
49. Daubin V, Ochman H. Start-up entities in the origin of new genes. *Curr Op Genet Dev*. 2004;14:616–9.
50. Britten RJ. Coding sequences of functioning human genes derived entirely from mobile element sequences. *Proc Natl Acad Sci*. 2004;101:16825–30.
51. Moore PB, Steitz TA. The Roles of RNA in the Synthesis of Protein. *Cold Spring Harb Perspect Biol*. 2011;3(11):a003780.
52. Wang X, Song X, Glass CK, Rosenfeld MG. The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. *Cold Spring Harb Perspect Biol*. 2010;3:a003756.
53. Tomasello M. *Origins of Human Communication*. Cambridge, MA: MIT Press; 2008.
54. Dick TP, Schamel WWA. Molecular Evolution of Transfer RNA from two Precursor Hairpins: Implications for the Origin of Protein Synthesis. *J Mol Evol*. 1995;41:1–9.
55. Weiner AM. SINEs and LINEs: troublemakers, saboteurs, benefactors, ancestors. In: Gesteland RF, Cech Tr, Atkins JF, editors. *The RNA World*, 3rd ed. New York, NY: Cold Spring Harbor Laboratory Press; 2006:507–34.
56. Rodin AS, Szathmary E, Rodin SN. On origin of genetic code and tRNA before Translation. *Biol Dir*. 2011;6:14.
57. Sun FJ, Fleurdepine S, Bousquet-Antonelli C, Caetano-Anolles G, Deragon JM. Common evolutionary trends for SINE RNA structures. *Trends Genet*. 2006;23:26–33.
58. Harish A, Caetano-Anolles G. Ribosomal History Reveals Origins of Modern Protein Synthesis. *PLoS One*. 2012;7(2012):e32776.
59. Villarreal LP. Viral ancestors of antiviral systems. *Viruses*. 2011;3: 1933–58.
60. Villarreal LP. The addiction module as a social force. In: Witzany G, editor. *Viruses–Essential Agents of Life*. Dordrecht, Netherlands: Springer; 2012:107–45.
61. Chen PJ, Cywinski A, Taylor JM. Reverse Transcription of 7S L RNA by an Avian Retrovirus. *J Virol*. 1985;54:278–84.
62. Li WY, Reddy R, Henning D, Epstein P, Busch H. Nucleotide Sequence of 7 S RNA. The *J Biol Chem*. 1981;257:5136–42.
63. Marz M, Donath A, Verstraete N, Nguyen VT, Stadler PF, Bensaude O. Evolution of 7SK RNA and Its Protein Partners in Metazoa. *Mol Biol Evol*. 2009;26:2821–30.
64. Kim TM, Hong SJ, Rhyu MG. Periodic explosive expansion of human retroelements associated with the evolution of the hominoid primate. *J Korean Med Sci*. 2004;19:177–85.
65. Dewannieux M, Esnault C, Heidmann T. LINE-mediated Retrotransposition of Marked Alu Sequences. *Nat Genet*. 2003;35:41–8.
66. Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and alu repeats by particular l1 subfamilies in ancestral primates. *Genome Biol*. 2003;4:R74.
67. van den Hurk JA, Meij IC, Seleme MC, et al. L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet*. 2007;16: 1587–92.
68. Vitullo P, Sciamanna I, Baiocchi M, Sinibaldi-Vallebona P, Spadafora C. LINE-1 retrotransposon copies are amplified during murine early embryo development. *Mol Reprod Dev*. 2012;79:118–27.
69. Wissing S, Muñoz-Lopez M, Macia A, et al. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum Mol Gen*. 2012;21:208–18.
70. Oricchio E, Sciamanna I, Beraldi R, Tolstonog GV, Schumann GG, Spadafora C. Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene*. 2007;26: 4226–33.
71. Osenberg S, Paz Yaacov N, Safran M, et al. Alu sequences in undifferentiated human embryonic stem cells display high levels of A-to-I RNA editing. *PLoS One*. 2010;5:e11173.
72. Macfarlan TS, Gifford WD, Driscoll S, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*. 2012;487: 57–63.
73. Daskalova E, Baev V, Rusinov V, Minkov I. 3′UTR-located Alu elements: donors of potential miRNA target sites and mediators of network miRNA-based regulatory interactions. *Evol Bioinf Online*. 2006;2:103–20.
74. Smalheiser NR, Torvik VI. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet*. 2006;22:532–6.
75. Panning B, Smiley JR. Activation of expression of multiple subfamilies of human Alu elements by adenovirus type 5 and herpes simplex virus type 1. *J Mol Biol*. 1995;248:513–24.
76. Hormozdiari F, Alkan C, Ventura M, et al. Alu repeat discovery and characterization within human genomes. *Genome Res*. 2012;21:840–9.
77. Macia A, Muñoz-Lopez M, Cortes JL, et al. Epigenetic control of retrotransposon expression in human embryonic stem cells. *Mol Cell Biol*. 2011;31:300–16.
78. Jurka J, Krnjajic M, Kapitonov VV, Stenger JE, Kokhanyy O. Active Alu elements are passed primarily through paternal germlines. *Theor Popul Biol*. 2002;61:519–30.
79. Ferrucci L, Romano E, de Stefano GF. The AluI-induced bands in great apes and man: implication for heterochromatin characterization and satellite DNA distribution. *Cytogenet Cell Genet*. 1987;44:53–7.
80. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA Editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol*. 2004;2: e391.
81. Smith JM, Wynne-Edwards VC. Group selection and kin selection. *Nature*. 1964;201:1145–7.
82. Smith JM. Group Selection. *Quart Rev Biol*. 1976;51:277–83.
83. Witzany G. Reduction of biological phenomena. Deficits of systems theory and the alternatives. In: Schmitz W, editor. *Sign Processes in Complex Systems*. New York, NY: Mouton de Gruyter; 2002:303–7.
84. Van de Vijver G. No genetics without epigenetics? No biology without systems biology? On the meaning of a relational viewpoint in a complex account of living systems. *Ann NY Acad Sci*. 2009;1178:305–17.
85. Witzany G, editor. *Biocommunication in Soil Microorganisms*. Heidelberg, Germany: Springer; 2011.
86. Witzany G, editor. *Biocommunication of Fungi*. Dordrecht, Netherlands: Springer; 2012.
87. Witzany G, Baluska F, editors. *Biocommunication of Plants*. Berlin, Germany: Springer; 2012.
88. Witzany G, editor. *Biocommunication of Animals*. Dordrecht, Netherlands: Springer. In press.
89. Schuster P, Fontana W, Staddler PF, Hofacker IL. From sequences to shapes and back: a case study in RNA secondary structures. *Proc R Soc Lond B*. 1994;255:279–84.

90. Domingo E, Martin V, Perales C, Grande-Perez A, Garcia-Arriaza J, Arias A. Viruses as quasispecies: biological implications. *Curr Top Microbiol Immunol*. 2006;299:51–82.
91. Witzany G. The Agents of Genomic Creativity. In: Witzany G, editor. *Biosemiotics in Transdisciplinary Contexts*. Helsinki, Finland: Umweb; 2007:295–301.
92. Witzany G. The agents of natural genome editing. *J Mol Cell Biol*. 2011;3: 181–9.
93. Villarrreal LP, Witzany G. Viruses are essential agents within the roots and stem of the tree of life. *J Theor Biol*. 2010;262:698–710.
94. Kobayashi I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res*. 2001;29:3742–56.
95. Witzany G. From the "logic of the molecular syntax" to molecular pragmatism. Explanatory deficits in Manfred Eigen's concept of language and communication. *Evol Cog*. 1995;1:148–68.
96. Witzany G, Baluška F. Life's code script does not code itself. The machine metaphor for living organisms is outdated. *EMBO Rep*. 2012;13:1054–6.