Virolution Can Help Us Understand the Origin of Life

Luis P. Villarreal

CONTENTS

20.1	Objective: A Virus First Overview	421
20.2	Some Definition and Other Problems	423
20.3	Current View: Individual Type Selection and Exapted Viral Genes	425
20.4	QS-C and Virolution	426
20.5	ENCODE Project: Viral *Junk* as *Essential* RNA Regulation	427
20.6	Fittest Individual Type Reconsidered	428
20.7	Corrective and Collective Power of Virus via Modern Quasispecies: QS-C	429
20.8	RNA World Reconsidered: Infectious Stem-Loops That Operate as QS-C Collectives	430
20.9	Agents Join Groups and Social Networks	432
20.10	Conclusions	435
Glossary		435
Review Questions		436
References		437

20.1 OBJECTIVE: A VIRUS FIRST OVERVIEW

A *virus first* perspective for understanding human evolution will likely seem counterintuitive or even preposterous to many readers (Villarreal, 2004). Surely, these most selfish and destructive agents (virus) cannot be proposed to have contributed substantially to the many complex features that make us human. Yet, developing and supporting this assertion is the objective of this chapter. Viruses are genetic parasites, often capable of transmission and dependent on their host for replication and/or maintenance. They are thus fundamentally able to interact with and contribute to host genetic (and epigenetic) content. It is this capacity that allows virus to be editors of host genetic content (Villarreal and Witzany, 2010; Witzany, 2006, 2009a). We know viruses to be agents of disease, often serious and even fatal. In what way can this capacity relate to the complexity needed to generate human capabilities? But viruses are also capable of colonizing and persisting in host genomes and becoming one with them. In so doing, they bring new and diffuse instruction sets to their

host that can promote new regulatory networks with new capacities. This process has been called *virolution*, virus-mediated evolution (Ryan, 2009). And it is the persisting viruses that are highly host specific which, usually sexually transmitted, also have the ability to differentially affect host survival. The relationship of persisting viruses to its host population has been proposed to contribute significantly to host survival and affect the tree of life (Villarreal, 1999, 2006, 2007, 2008; Villarreal and Ryan, 2011). This process is shown schematically in Figure 20.1.

Such a process is fundamentally synbiogenic (Pereira et al., 2012). Indeed, it will be asserted subsequently that persisting human viruses were likely involved in the Homo sapiens—Neanderthal evolutionary outcome.

Why would viruses promote novelty via the formation of complex networks able to contribute to host phenotypes? The currently accepted view is that viruses are simply providing an extended source of errors (diversity) that can occasionally become *exapted* by their host for host purposes. An infected individual host variant will survive and somehow adapt virus information for its own survival. Networks are then created from this information in stepwise series of selection events. The real answer, however, lies much deeper than is likely to be appreciated. Indeed, it relates directly to the earliest events in the evolution of life reaching all the way back into the RNA world. This world is characterized by consortial, cooperative, multifunctional, and transmissive

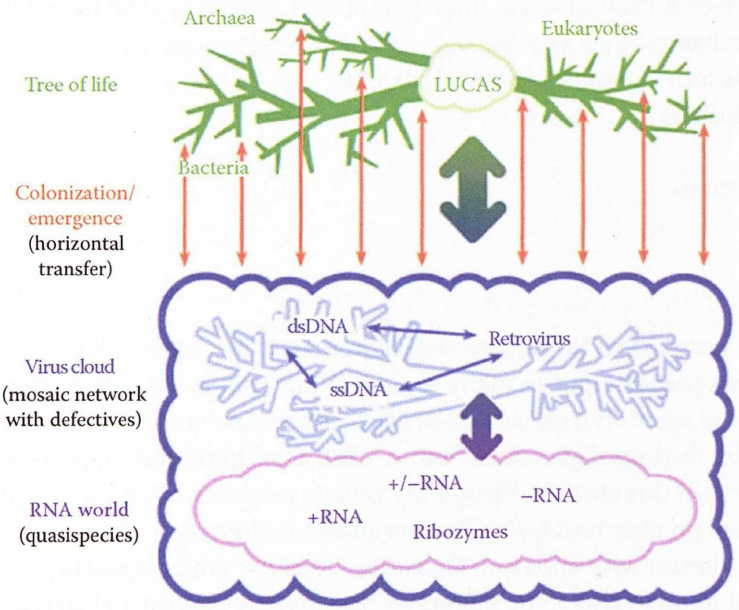

FIGURE 20.1 Overview of the viral cloud contribution to all the domains of the host evolutionary tree. Virus contribution to host is not occasional, but ongoing as shown. Also, such viral information is not due to *errors* as viruses are competent in host code and can be natural editors of host code. Viruses are thus an original and ongoing force for all life. (Reprinted from Villarreal, L.P. and Witzany, G., *J. Theor. Biol.*, 262, 698, 2010. With permission.)

RNA agents that operate in groups that can identify network membership and preclude nonmembers (immunity) (Villarreal and Witzany, 2013a,b).

We have long focused on the modest error-based genetic adaptations associated with neo-Darwinian selection and evolution. But whenever a host genome becomes colonized by nonancestral endogenous retroviruses (ERVs) and related elements that replicate via RNA, a quasispecies consortia (QS-C)-mediated process again applies to modify existing RNA societies that provided identity (control, immunity) superimposing new and often multiple uses of stem-loop RNAs that are now engaged in and providing new identity networks. This is a much more creative and punctuated process, able to promote complex regulatory shifts, but one that still essentially uses invasive stem-loop RNA agents.

20.2 SOME DEFINITION AND OTHER PROBLEMS

The term virus has a broad and almost instinctive meaning to many people with respect to disease. It is, however, worthwhile exploring a current definition of this term in order to employ it with greater consistency and precision. Since it is well known that many viruses can infect and exist within their host with no disease, clearly, disease cannot be a defining characteristic. Nor is uncontrolled replication a defining characteristic since many viruses have highly regulated replication cycles. Some do encode proteins involved in membrane synthesis. Some do not even encode their own capsid or membrane proteins, so this too cannot be a defining characteristic. But so far, no virus has been observed to code for a ribosome. Nor do they appear to encode many of the most fundamental metabolic proteins. Thus, viruses are fundamentally molecular entities that are parasitic to living systems (with ribosomes and energy production). But some viruses are parasitic to other viruses (thus parasitic to living systems plus virus), and most viruses can generate defective versions of themselves that are parasitic to the host system plus self-full virus. These situations can be very important for some specific viral lifestyles. Thus, our definition must be inclusive of all of these situations. I therefore propose the following characteristics for defining virus:

- A virus is a molecular genetic parasite.
- A virus must be competent in the instruction system of its host system.
- A virus must superimpose (edit) new instructions onto the host system (extending the code, bringing novelty, promoting symbiosis).
- Viral instructions must promote maintenance of the virus (i.e., self-identity compatible) which includes directed replication needed for either maintenance and/or transmission.
- Virus instructions can also simply include compelling the host to *maintain* the viral instruction set (persistence) and replicate it along with the host.
- Viral instructions must oppose (i.e., damage) competitive instruction sets (i.e., host immunity and/or virus competition).

- These viral instructions may subvert (colonize) opposing or competing instruction sets so as to maintain a coherent viral instruction system.
- The simplicity of RNA virus instructions requires that they be a coherent consortia of diverse RNA instructions (QS-C).

In addition to these defining characteristics, I would propose that the original *viral* instruction systems were simple stem-loop RNA replicators, as proposed for the RNA world (Briones et al., 2009). These parasitic replicators were able to transmit and occupy (ligate) their other RNA stem loops, including their own QS. Such self-invasion promotes the emergence of more complex functions (e.g., ribozymes and a consortial ribosome). RNA viruses still depend on these stem-loop instruction agents for basic identity and replication. The host (DNA) has become a habitat for these RNA societies.

There are other important problems involving definitions that should be mentioned. These include the terms networks and systems. Although popular use of these terms is consistent with the way I hope to apply them, the real problem, however, relates to attempts to mathematically define these terms so that calculation-based approaches can be applied to them. For example, a network stems from the concept of a net, with knots (nodes) connected to each other in binary links. This can be described mathematically. Similarly, formal complex systems posit a mathematical foundation for defining systems (von Bertalanffy and LaViolette, 1981). However, in the context of diverse but coherent RNA agents (QS-C), it is not possible to mathematically set either the potential interactions or the nature of these interactions for a single RNA entity as it will have conditional and context- and history-dependent activities (uses) within the population. This issue will be expanded further later. However, it does compel us to use the terms *networks* and *systems* in a less-defined (but popular) way. The concept of network in particular will be important for our discussions as it will relate directly to group identity that will require the specification of network membership characteristics. For an RNA agent, being a member of a network relates directly to its identity markers (often stem loops).

I will often consider the issues of group identity and group behavior as these are proposed to provide the foundations of social mechanisms. I will seek to define a network from the perspective of a consortium of RNA agents and apply the strategies of these diffuse transmissive agents to explain the creation of new networks and the editing of existing ones. However, it will be very difficult to think about and communicate these consortial or social issues. This is not because they are so inherently complex, but more because they are fundamentally interactive (social) phenomena that resist a linear explanation. For example, assigning a single function (or fitness) to an entity that is part of a consortium or network will inherently restrict out thinking about how the entity must function in the context of a consortium. A social system will have individual agents (such as RNA stem loops) that will fundamentally have multiple (often opposing) activities and uses. This is most apparent in the study of viral QS-C presented later. A system or network must always have this feature. We need to think socially, not serially.

I will present the case that immunity mechanisms are very the same as group identity mechanisms and that both operate using various strategies (such as addiction modules) that can destroy nonidentity and nongroup membership, while supporting group membership.

One more note of caution to keep in mind before further developing this (multi) line of reasoning. The types of RNA stem-loop-mediated changes we see in recent human evolution look hauntingly similar to those I proposed for the very early events in the evolution of life (see next chapter). It seems that from the very origins of life to recent genomic changes in human DNA, RNA societies have been acquired from infectious events and are mediating identity and social phenomena. It should therefore be considered that changes in RNA stem-loop composition may best define membership for all living systems and provide new identity systems in all domains of life, including virus.

20.3 CURRENT VIEW: INDIVIDUAL TYPE SELECTION AND EXAPTED VIRAL GENES

The development of neo-Darwinian thinking in the 1930s stems directly from the foundation that natural selection acting on variation (mostly from replication errors) in individuals selects for the survival of fittest type variant. Thus, the variation in offspring originates from the direct ancestor to the selected individual. However, when *nonancestor* virus-derived genes are seen to occur in host genomes, it is typically reasoned that such genes simply represent another form of variation (errors) that was also somehow associated with individual survival. The surviving host individual was then able to adapt (exapt) these genes for its own purpose and survival. This explanation still invokes a central role for individuals. What then results are various scenarios, such as kin selection, tit for tat, or arms race ideas involving a serial one-upmanship and linear process of selection. Any networks that emerge will then need to stem from this same serial process. The process is not prone to punctuated bursts in evolution nor is it particularly prone to rapid emergence of complexity or novelty. Also, any associative or group behavior that emerges, such as altruistic behavior, will similarly stem indirectly from individual survival, as described by various kin selection or game theory models. This view has been well accepted for numerous decades, and many current evolutionary biologists no longer question its basic tenets. Some even like to think of this as laws of evolution. But this is a view that emerged well before we understood the broad and ancient prevalence of virus. In the last several decades, analysis of comparative genomics and metagenomic sequencing of numerous habitats has shown us that virus-derived sequences dominate in all habitats so far evaluated (see Koonin, 2011). The term virosphere has been introduced to describe this vast cloud of genetic information. And within the genomes of organisms, virus-derived information is almost always the most dynamic component of host DNA for all domains of life. Much of this virus-derived host DNA, however, has long been seen to play no useful role; it was junk DNA that was the product of selfish replicators (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Yet, much of this *junk* was clearly viral derived, and often, its expression was associated with reproductive tissue (Ono et al., 1985). More recently, such *junk* has seemed much more important for the functioning of the organism (Volff, 2006).

But it is still basically seen as *exapted* stuff, put to some inadvertent good host use following individual selection. But the existence of a vast virosphere should compel us to think differently about virus-derived information in host genomes. All domains of life must survive in this ancient, unrelenting, and extremely adaptable virosphere. How can life thrive in this situation? In the following, we will look at the viruses themselves for an answer to this question. And from this question, we should rethink some of the tenets of evolutionary biology.

20.4 QS-C AND VIROLUTION

We should now ask: what is virolution and how does it affect host evolution? Can it provide some core, essential function needed for life? And more fundamentally, does virolution operate with additional principles, such as consortial group identities (QS-C) that can colonize and transform host, which fundamentally promote networks and complexity? If so, can these principles help us better understand the origin of life or provide insight into the origin of our human social capacities? Later, I attempt to summarize a large body of evidence from many domains of research that I feel help us better define virolution and the origin of host complexity. This will provide a core theme that will link all these diverse studies with the role of stem-loop RNA in viral and host identity. This role of QS-C in virolution is presented schematically in Figure 20.2.

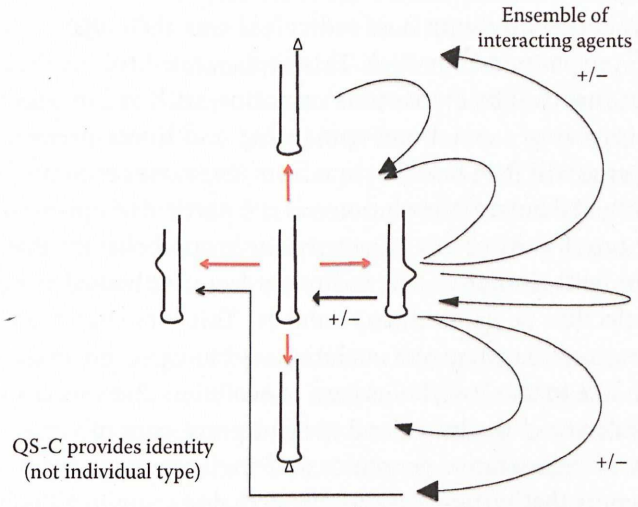

FIGURE 20.2 The crucial difference of QS-C with former QS concepts (fittest type—mutant spectra) is the basically consortial organization of functional RNA ensembles. Shown earlier are the possible consortial interactions (black arrows) of just one diversified RNA stem loop. These multiple activities (shown as +/−) preclude individual fitness definitions but require emergence and adaptation of group membership identities. Defectives with similar subviral RNA (stem-loop groups) remain relevant in both evolutionary and developmental processes. As a result of this basic evolutionary process of RNA stem-loop consortial building, we can look at the emergence of de novo identities. (Reprinted from Villarreal, L.P. and Witzany, G., *World J. Biol. Chem.*, 4, 71, 2013. With permission.)

As shown, such virolution is what promotes the creation of new *systems*, not serial selection from errors. But this looks like errors since most of the instructions are subviral. Virus, the ultimate and nearly invisible selfish agents, has finally taught us about the power of consortia. It is a big lesson and it applies to all levels of life. But why would a consortium of viral agents act to promote complexity? The answer I will develop later is that it is for the sake of superimposing group identity and group survival. The QS-C has to incorporate a new viral identity onto the host. This colonization will also clearly affect host survival in its extant virosphere. The virosphere matters for the success of all life. Such a colonizing event must promote the survival of this information and new viral identity/ecology that results. This is a very different perspective than that of selfish individual type selection. And this view suggests we adopt a new philosophy of biology in which collective behaviors are core for all life forms. And although virolution supports various forms of multilevel selection, it does not conflict with traditional individual type selection that emerged with DNA-based cells and virus. But whenever infectious sets of RNA-based replicating agents successfully colonize a host, they will again bring to bear the creative, cooperative, and distributed power of QS-C selection to their host. This is a most ancient process that still operates on DNA, using DNA as a stable habitat (Villarreal and Witzany, 2013a,b). I will now examine both the earliest events (RNA world) and most recent events (human-specific evolution) from this virolution perspective. One fundamental theme will apply throughout this examination that will be introduced both early and later in this chapter. This is the fundamental role of stem-loop RNA structures in the identity and function of infectious forms of RNA. When I focus on human-specific features, for example, such as our reproductive or social brain changes, it will be from the perspective of a role of infectious forms of stem-loop RNAs. The RNAs have multiple regulatory capacities that lead to a better understanding of RNA cascades and networks, which are the products of or promoted by serial colonization of virus (and often provide antiviral activity). These regulatory stem-loop RNAs will mostly occupy introns, 3′ untranslated region (UTR) and some 5′ promoter regions. We will also see that older identity/regulatory systems become subjected to manipulation (repurposed) or elimination following successful colonization.

20.5 ENCODE PROJECT: VIRAL *JUNK* AS *ESSENTIAL* RNA REGULATION

For many years, molecular biologist assumed that the complex RNA expression patterns observed by various techniques (such as hybridization kinetics) in the mammalian brain were due to the expression of many genes, which was expected for such a complex organ (see Chaudhari and Hahn, 1983). However, comparative genomics has made clear that gene transcription differs little between human and great apes (Khaitovich et al., 2004). Indeed, total gene numbers differ remarkably little between the simplest animals (*Caenorhabditis elegans*) and humans. But by far, the biggest differences between human and chimpanzee genomes were due to insertion and deletions (indels) (Mills et al., 2006; Watanabe et al., 2004; Wetterbom et al., 2006). The great majority of these indels are the result of retrotransposon activity of various types (ERVs, long terminal repeats [LTRs], long interspersed nuclear elements [LINEs], and alus being most numerous). Of these, the alu elements and transcripts are particularly active and affect RNA editing and intron splicing

in the human genome (Sakate et al., 2007). In addition, they are frequently involved in epigenetic control and can emerge or expand rapidly in genomes (Zeh et al., 2009). Such a large-scale retroposon colonization would seem to pose a highly genotoxic situation for the human genome, an idea that seems supported by genomic analysis (Keightley et al., 2005, 2006). And yet this noncoding DNA is species specific (Toder et al., 2001), evolving quickly in humans (Bird et al., 2007), but also appears to be under very strong selective constraints (Bejerano et al., 2004; Bush and Lahn, 2005). This seems problematic in several ways: this is an inherently destructive event that should seldom result in novel or complex phenotype, plus it is both rapidly changing between species yet sometimes highly conserved. Indeed, this high rate of change was previously used to argue for the idea that it must be junk DNA. Yet, these are the changes that must be addressed and included to explain the emergence of the large and social human brain (Chapter 21). How then can we understand the origin of the most complex organ known (human brain) in the context of such massive introduction of errors? Clearly, we cannot. But perhaps the concept of *errors* is itself in error as implied earlier. Indeed, a major correction in our thinking has emerged from the ENCODE project. This project is a consortium of researchers that has sought to characterize all the RNA transcribed from the human genome, including RNA that is not cytoplasmic polyadenylated mRNA but is noncoding RNA (Mattick, 2005). It is now quite clear that most of this *junk* is transcribed and that 95% of the transcripts are from repeated sequences that were retrotransposed (Mattick, 2010,2011; Mattick et al., 2010). These transcripts include a previously poorly studied class of long noncoding RNA (lncRNA) (see Khalil et al., 2009). Furthermore, these noncoding transcripts appear particularly relevant to human brain and cognitive development and evolution (Barry and Mattick, 2012; Mattick and Mehler, 2008; Mehler and Mattick, 2007). Additionally, long-term memory also seems to use noncoding RNA (Mercer et al., 2008). These observations have led Mattick to propose that genetic programming in higher organisms (including human) has been misunderstood for 50 years (Mattick, 2001). Regulatory RNA derived from retrotransposons is key to eukaryotic complexity, compelling us to abandon the concept of selfish junk DNA. But in this realization, we also come to realize this regulatory RNA is operating mostly as stem-loop RNAs that have complex, multilevel, and even opposing functions. It is clearly operating and evolving as a network. But networks of stem-loop RNAs are also thought to have been crucial for the origin of RNA-based life. Could it be that the creative power of societies of stem-loop RNAs involved in the origin of life is still at work during recent human evolution? Let us further evaluate this idea.

20.6 FITTEST INDIVIDUAL TYPE RECONSIDERED

RNA viruses have long been recognized as distinct agents from their host cells in that they were the sole survivors of the RNA world that still used RNA as a genetic molecule. That they could replicate so readily and be characterized in the laboratory made them ideal systems to study variation in RNA replication. The variation was considered to result mostly from copy errors of a low-fidelity polymerase. And since viruses could be *cloned* (plaqued), they apparently adhered to the concept of individual fittest type selection. Since it was realized early on that RNA replications at the dawn of life in the RNA would also replicate

with high error rates, this seemed to present a problem for the origin of life and the origin of the genetic code. It was from this perspective in the 1970s that Eigen developed the QS equations to explain the quantitative behavior of RNA populations that were generated via errors of the master fittest individual type template (Eigen, 1971). The basic assumptions were then that there was a master fittest type RNA template that would generate a cloud of RNA progeny due to copy errors but that this cloud would have certain overall behaviors (such as error threshold). Many more theoretical papers followed this early publication by Eigen, by his colleagues, and numerous others. And in the following decades, a large number of laboratory studies by RNA virologist sought to evaluate and measure various aspects of QS theory. It became very clear that the QS behavior of RNA viruses was very important for understanding clinical outcomes of human infections. And indeed, some of the insights of QS theory were observed, such as error threshold. The concepts of variation of the master fittest type became entrenched during this period as there seemed to be no conflict with more traditional neo-Darwinian selection. Thus, a consensus took hold for what QS is or means associated with master fittest type.

20.7 CORRECTIVE AND COLLECTIVE POWER OF VIRUS VIA MODERN QUASISPECIES: QS-C

In the ensuing several decades, many laboratory observations were made that indicated more complex collective behaviors for viral QS than were predicted by the QS equations. Two of the more active laboratories were those of John Holland and Esteban Domingo (see Domingo et al., 2008). The most recent compilation of these studies outlines many of the collective behaviors that have been made with QS (Domingo et al., 2012). In my opinion, the culmination study that most clearly reported that QS have more complex collective behaviors was the study from the Andino group of poliovirus pathogenesis in a mouse model in which diversity and cooperation were key to viral fitness (Vignuzzi et al., 2006). The QS collectives have distinct and measurable fitness. They can compete with and exclude related populations. They have minority populations that are crucial for overall fitness (Briones and Domingo, 2008; Briones et al., 2006). They can display heterogeneity important for fitness that is not observed in the consensus type (Borderia et al., 2012). They can suppress their own replication through lethal defection (Grande-Perez et al., 2005). They can be composed of members that can complement and interfere with replication of the collective, and many of these features can be observed in clinical infections such as humans with hepatitis C virus (Domingo and Gomez, 2007). Thus, QS are collectives that have positive and negative interacting members that are bound together for a combined fitness that depends on diversity (Arbiza et al., 2010; Lauring and Andino, 2010; Ojosnegros et al., 2011). It is thus ironic in that it is from the viruses, the most selfish of all genetic entities; we experimentally observe the characteristics of cooperative, collective behavior. And it was the *fittest type* assumptions of Eigen that generated QS equations and theory that stimulated the development of this modern collective QS view. But we are left with a conceptual contradiction. Modern QS observations do not depend on the master fittest type, and the consensus sequence may not predict to the fitness of the diverse collective. Diversity itself seems crucial. Such dynamic diversity allows a population of otherwise

rather simple agents (such as HIV-1) to defeat a highly complex and evolved system of adaptive and innate immunity in their human host. If such infections were limited to the fittest type individuals, they would fail to overcome such a complex system. Not only can the power of QS defeat our human immune system, it has also largely defeated our combined human technology by frustrating the development of effective vaccines for 30 years. All this impressive biological competence comes from a small and *simple* virus! I therefore submit a modification to QS terminology to incorporate this collective and cooperative feature. The term QS-C will indicate a *collective* of *cooperative* character to the population. That way, the original term, QS, can still apply to fittest type models.

With this clarification, it should become apparent that all RNA replicators (especially simple ones) will have high rates of diversity generation (not error). In addition, all genetic entities that replicate via RNA will also be prone to QS-C (collective) behaviors. These behaviors will include both cooperative and competitive interactions, even within the same individual molecule. RNA, however, is not simply providing a syntax for genetic information. It is more than a code. It can also provide structure (stem loop), identity (stem loops, 5′, 3′ ends), and functional (ribozyme) activity. And it can be dynamic (e.g., pseudoknots) and responsive to the environment (riboswitches). Because of this much extended capacity relative to DNA, RNA can be considered as a more active entity, with behaviors that make it able to function as an *agent* to affect its own activity and survival (Witzany, 2009b, 2011). In that light, DNA can be considered as a habitat for various RNAs. It was from this perspective that Witzany and I proposed that DNA should be considered as a habitat for these active RNA agents (Villarreal and Witzany, 2013a,b). But this discussion of simple RNA replicators suggests that the concept of QS-C should also apply to the ideas and experiments concerning the *RNA world* hypothesis. Yet curiously, very little *RNA world* research has addressed any issues regarding QS (see Altman, 1989; Gesteland et al., 1999), let along the more modern QS-C idea. As many are starting to think that life originated in a cooperating situation (see Holmes, 2012), it is worth briefly considering if the QS-C concept will provide a different scenario for the origin of life.

20.8 RNA WORLD RECONSIDERED: INFECTIOUS STEM-LOOPS THAT OPERATE AS QS-C COLLECTIVES

To evaluate the QS-C and infectious perspective on the RNA world hypothesis, I will apply and explore the RNA agent concept introduced earlier to the role of stem-loop ribozymes in the origin of life. The main objective is to incorporate the historically absent QS-C and parasitic perspective (with its inherent feature for group fitness) into the process that creates RNA societies. I will not explore early chemical evolution that might have led to the emergence of RNA molecules but will instead assume that RNA has come into existence and follow its features from this perspective. One immediate consequence of this perspective is that we will be focused on collective features of RNA populations and will thus evaluate the chemical consequence of ribozyme QS societies, not individual replicators. This foundation immediately creates a situation in which *systems* of molecules with multiple behaviors will have the primary role in promoting the origin of life. It will also be important early on to consider how these systems maintain coherence (group identity, presented later), as this is an essential feature.

Indeed, a basic and continuing theme will be that a core function of stem-loop RNAs is to provide molecular identity through all of evolution, including recent human evolution. This identity theme will persist throughout this chapter and be frequently reintroduced. The idea is then that individual members of stem-loop RNA societies were collectively able to invade (ligate into) each other to form a more stable and capable (ribozyme active) consortia with emergent, transformative, and unpredictable abilities. These collectives would lead to the origin of various ribosome and RNA cellular societies (still linked to its stem-loop tRNA origin). Such a scenario also introduces the basic role of cooperation in the origin of life. It does not, however, eliminate competition, preclusion, and extinction that are also inherent features of QS-C behaviors. Furthermore, the identity and transmissive role for stem-loop RNAs sets the early (precellular) foundation for the origin of viruses whose emergence will further drive host evolution via colonization. The cooperative and parasitic features of QS-C will also promote the early participation of peptides in the identity and evolution of the ribonucleoprotein (RNP) complex society. The maintenance of these RNA societies as a coherent collective will generally be mediated by addiction modules (counteracting functions that recognize each other and harm nonrecognized partners) that underlie group identity and immunity in all living systems. With this foundation, the emergence of genes, DNA, cells, and individual fittest type selection can all be derived. But the emergence of DNA and cells and Darwinian evolution does not terminate the central role for transmissive RNA societies in the evolution of life. DNA becomes a habitat for these stem-loop *identity* RNAs, and it is from this perspective that I will subsequently examine recent events in human evolution. One issue should already be clear. This scenario posits that collective and cooperative behaviors were and remain essential for the emergence of living complexity. QS-C then provides a conceptual foundation for the study of cooperating chemical networks in which RNA mixtures of self-replicating ribozymes can form highly cooperative and dynamic autocatalytic cycles (Vaidya et al., 2012). Let us now put this into the perspective of virolution.

On the origin of the RNA world, short RNA oligomers formed by chemical processes needed to become longer RNAs able to perform template-based catalysis. It has been proposed that the initial chemical formation of hairpin-like RNAs (stem loops) could provide ribozyme activity following a ligation-based modular evolution that would yield ribozyme autocatalysis (Briones et al., 2009). Indeed, later, I present a series of studies that support this modular view. But according to the parameters of QS-C evolution, for a consortium of RNA stem-loop replicators to survive, they must form a coherent population. They must share their identity and survival. The recognition of the stem-loop sequence itself by catalytic agents could provide such common identity. Alternatively, chemical markers or initiators of catalysis could also mark the common population for priming or replication. Thus, it is very interesting that the smallest ribozyme so far reported consists of just five nucleotides able to catalyze aminoacylation of the 3′ end (Yarus, 2011). The addition of an amino acid to an RNA molecule has many interesting chemical implications. A ribozyme has rather limited chemical potential compared to proteins. This is mostly due to proton disassociation constant of various amino acid moieties that are not close to pH neutrality. Thus, amino acids are much more capable as chemical catalyst for this reason. Without the participation of amino acids, ribozymes must attain complex folds, often with some dynamic

character (pseudoknots) to be effective catalyst allowing them to cleave and ligate RNA. Given this chemical advantage, we might expect that RNA evolution was greatly facilitated (but not coded) by peptides that contribute catalytically. In addition, such a modified RNA would likely also provide a chemical marker that could distinguish this RNA population. Indeed, this molecular identity idea is developed later as a way to better understand the origin of tRNA and its role in initiating replication of so many RNA viruses as well as how this chemical marker could promote the symbolic genetic code.

A good starting point for the accumulation of complexity seems to be hairpin ribozymes whose activity can be controlled by external effectors (Muller et al., 2012). Structural variation in these ribozymes allows progeny RNA to have different functions from their parental RNAs. The objective is to replicate RNA with RNA which hairpin ribozymes can perform via a sequence of ligation reactions that produce a longer ribozyme (Cheng and Unrau, 2010). Along these lines, two short hairpin RNAs can catalyze their own ligation to form larger RNA constructs (Gwiazda et al., 2012). Thus, we see interactions that promote more complex progeny. However, for a fully active ribozyme, complex RNA folding is needed. And such folding is cooperative (Behrouzi et al., 2012). Folded ribozymes can also interact with other small molecules promoting their function as riboswitches (Ferre-D'Amare, 2011). This includes amino acids that could promote either catalytic control or group identity marking. And the ribozyme folds can also be dynamic and context sensitive as seen in pseudoknots (Perreault et al., 2011). But ribozymes can also be invasive, including self-invasive (Kumar and Joyce, 2003). Thus, stem-loop RNAs have many behaviors that would allow them to function as a mixture of agents involved in their own recognition and synthesis. Of particular interest is their ability to self-ligate as this could promote the emergence of RNA societies with self-identity. We can also think of tRNA as stem-loop RNA with various functions and histories. Indeed, it appears that tRNAs evolved from two separate hairpins (Dick and Schamel, 1995), in which each of the stem loops interacts with a different ribosomal RNA subunit (presented later). This is a very interesting observation from an RNA society perspective. The invasive nature of intron ribozymes (endonuclease) also applies to tRNA from archaea, but here four distinct specificities are known (Fujishima et al., 2011). This very much resembles an identity system in which introns are marking central cellular (self) agents (tRNAs) for group identity but should destroy similar tRNAs (viral, other cellular, etc.) lacking the intron marking. It is thus also interesting that tRNAs with various linked amino acids themselves have been proposed to have originated before the translation system as genomic 3′ tags needed for RNA ribozyme replication (Rodin et al., 2011; Sun and Caetano-Anolles, 2008). This early function can also be explained as having served as a tag for group identity and could better explain the polyphyletic nature of the origin of tRNA (Di Giulio, 2013).

20.9 AGENTS JOIN GROUPS AND SOCIAL NETWORKS

The QS-C perspective allows us to consider the role of stem-loop RNAs in the origin of the RNA world in which the action of individual agents can cooperate and be combined into a more capable collective action of a population. Thus, the origin of spontaneous cooperating networks of stem-loop RNA replicators (Vaidya et al., 2012) can be understood from this perspective. However, I will use the term network to include some distinct features,

specifically network membership. To designate this situation, I apply the term social network to distinguish networks that have no membership criteria. Networks can be either open or closed. Biological networks almost always have a closed feature to them. Basically, for a network to be coherent and able to act collectively, it must limit membership to promote coordination. Otherwise, it is simply a collection of uncoordinated agents, and there will be no selection for maintaining the network existence. If we are examining a network composed of stem-loop RNAs, it will be necessary for the individual RNAs to have some feature or behaviors that maintains membership (such as replication and recognition). This requires interaction. If only one type of RNA is supported (e.g., high-fidelity replication), there can be no complementation and complex function (i.e., ribozyme) for the collective. A diversity of behavior and type will be essential. Recall, however, that these RNAs act as agents in which various (multiple) behaviors will be possible even for the same sequence. This means there are diversity of interaction as well as diversity of type. Thus, overall interaction of an RNA agent with the collective must promote coherence and continued existence. What then are the features that promote continued existence (selection) for a network? This does not require that only positive (e.g., replication) interactions be supported. Negative interactions, including interference, will also be needed. For example, highly efficient runaway replicons would overtake a QS collective and yield only one RNA type. Thus, the QS would lose complementing functionality and would also consume all substrates if they were not regulated. This situation presents a problem in those habitats with limited substrates (likely a very common state). Therefore, some level of self-regulation (negation) in the collective would promote the survival of the collective, especially if these RNAs could interact with the substrate in a regulatory (riboswitch) manner. That efficient replicators become susceptible to parasitic replicators would provide an inherently spontaneous process of self-regulation. Yet, the collective will still need to promote replication when it is favored. Accordingly, it becomes important for members of the collective to be subjected to both positive and negative self-regulation via RNA–RNA interactions. However, here too, there must be some limits to self-regulation as the collective cannot tolerate overly active self-regulating members that will extinguish the collective. Thus, we see that being a successful member of a collective has many (and multiple) behaviors associated with it. On top of that, as a QS-C replicates, these features will drift with time in a dynamic manner. In this context, we can see that a random RNA stem loop or a stem-loop RNA from a different QS collective would likely not be coherent with the other members of a particular QS. A QS society is generally rather specific for its members. Group selection has already started. Indeed, many experiments with RNA viruses infecting humans and animals have shown that a particular QS will exclude other QS of the same virus. And such society membership is also time dependent in that the serial passage of the same viral QS will usually result in subsequent QS that precludes prior individual members of the QS. This behavior has often been called a *Red Queen* behavior, but such a classical neo-Darwinian view does not incorporate or acknowledge the issue of group membership. The membership view, on the other hand, allows us to understand the maintenance of minority types in the collective since these members can provide a needed but complementing catalytic control. Thus, a QS society is a network that will naturally promote the emergence

of membership. And as noted, defective interfering agents can also contribute to membership control. As I have previously proposed (Villarreal, 2012), group membership can also be promoted by the combined action of toxic agents linked to antitoxic agents. A common version of a toxic agent is an endonuclease that will cleave sequences that are recognized (as foreign). The antitoxin in this case prevents the action of the endonuclease (e.g., via a bound protein or methylated base, dsRNA with another molecule, altered RNA fold). In this light, the endonuclease and ligation activities of stem-loop ribozymes are particularly interesting. A stem-loop ligase could provide a mechanism to recognize nonmember stem-loop RNAs and destroy them by ligation. Recall, however, that serial ligation can also be used to copy a stem-loop RNA. But such a situation has several very interesting implications. One of the problems with a society of stem-loop RNAs is that to attain their combined function, they need precise physical molecular placement relative to one another. This would normally require a high concentration dependence to counteract diffusion. By ligation, however, we could build a society of stem-loop RNAs that have covalently placed the various stem loops in the correct functional (or dynamic/regulatory) context and have lost their concentration dependence. It seems likely that such a process would involve an invasive self-colonizing stem-loop RNAs that result in one molecular entity with a common identity function. This would generate one entity that evolved from the ligation of a mixed set of stem-loop agents that now have a highly enhanced (collective) functional capacity. This collective would also have a highly enhanced capacity for persistence as it need not continually replicate individual stem-loop RNA agents to maintain its membership. The collective, however, would still need to oppose nonmember or other parasite participation. Additionally, a collective might attain a conditional (regulated) replication capacity if it incorporates stem-loop RNA riboswitches. It is by such a process that we can now consider the origin of the ribosome.

Membership is thus crucial for living networks (systems) to emerge. In examining the literature relevant to QS, the RNA world, and RNA network formation, we can indeed find some experimental evidence that supports QS and the spontaneous emergence of RNA networks. But almost completely lacking from such experiments is any evaluation of the membership issue. For example, QS-like behavior has been observed with in vitro RNA replicator studies (Arenas and Lehman, 2010). Nonenzymatic template (peptide)–directed autocatalytic systems can show network behavior (Dadon et al., 2012). And communities of RNA ribozyme replicator sets can also show lateral evolution (Hordijk and Steel, 2012). Also rule-based computing simulation has been applied to similar systems in an effort to understand the emergence of parasites and antiparasites (Jalasvuori et al., 2010). Along these lines, the hypercycle kinetic model was proposed to be a system of cross-catalyzing RNA replicators that depend on cooperation for growth, but this is not a collective autocatalytic system as proposed earlier (Szathmáry, 2013). But hypercycles as proposed are not able to tolerate parasites, let along depend on them for development. Yet, the biggest problem of all such studies is that there is no assumption regarding the basic importance of network or group membership. Without this network membership concept and its attending strategies and mechanisms, authentic collective action does not emerge. Systems do not develop. The dynamic nature of network membership and collective action poses many unsolved problems for existing theory.

For example, how is the multipotential of an individual RNA to be evaluated within the QS-C if we cannot specify all the other interactions and how they change with time? We cannot apply our current ideas of fitness to this individual RNA as the historical and population context is key. Network membership needs to be prominently considered if we are to understand the origin of the ribosome and the genetic code. As will be further outlined later, replicator identity marking via 3′ aminoacylated of a stem-loop RNAs appears most able to explain the origin of a tRNA-mediated genetic code. This is a big difference in our conceptual stance. For in contrast to Darwinian evolution, network members will generally have distinct ancestral histories. These members will mostly originate from separate parasitic lineages that were able to penetrate defenses and join the network (sometimes in mixtures). They do not need to descend from one individual or even be from the same type of agent (virus, transposon, intron, intene, etc.). From this perspective, we can understand why the two halves of tRNA have distinct evolutionary histories, yet tRNA is a core agent for the evolution of life. Thus, neither the amino acid-based (peptide) ancestors nor the RNA-based ancestors need a common origin to participate in a symbiogenic network. QS-C theory supports such a network process. And we will continue to apply the QS-C perspective for the rest of this chapter. In the following, network membership will provide the basis for examining noncoding RNA-based regulation needed for multicellular complexity (Lozada-Chavez et al., 2011).

20.10 CONCLUSIONS

The application of virolution overall to issues regarding the origin of life can provide us with a very distinct and new perspective on how living systems emerge from chemical replicators. The emergence of cooperative QS-C thinking from the more accepted QS equations of Manfred Eigen, based on individual type selection, has provided a conceptual foundation from which collective action of RNA agents can now be understood. As group membership becomes a basic criterion for the emergence of living systems, we also start to understand why the history and context of the RNA society become crucial for social survival and function. History and context dependence also lead to the emergence of symbolic code in living systems. Indeed, this QS-C thinking can also provide us with a transition point between the chemical world of RNA replicators and the living world of RNA agents that must belong to their respective society. The power of a consortium to solve complex, multilevel problems that can even use opposing and minority functions becomes evident. This power, which promoted the emergence of the RNA world, did not become extinct with the emergence of DNA-based life. As we will see subsequently, the consortial action of *parasitic* RNA stem-loop societies can also help us understand the emergence of our large social human brain.

GLOSSARY

ERVs: Endogenous retroviruses. Partial or complete sequences derived from retroviruses that have become part of the host genome.

Exapted: An evolutionary theory that proposes that DNA (genes) from other organisms, such as parasites, can become part of the host DNA following natural selection.

Genetic parasites: Genetic agents, such as viruses and transposable elements, that colonize and use host systems for their own maintenance.

Group identity: The capacity for a group of living and/or genetic agents to recognize other members of the same group.

LINEs: Long interspersed nuclear elements. Genetic agents, distinct from retroviruses, which derive from a type of retrotransposon, able to replicate from RNA via DNA.

LTRs: Long terminal repeats. The characteristic sequences found on both ends of a retrovirus genome.

Matagenomic: The study of populations of genomes as they occur in particular habitats.

Network membership: The capacity of a network to recognize allowed members from nonmembers.

Networks: A system of agents (or elements) that interact in a coherent fashion, usually associated with regulation.

Pseudoknots: The ability of a stem-loop-like RNA to fold itself into two distinct and dynamic conformations.

Quasispecies: A population of related genetic agents that are derived by variation during replication.

Riboswitches: A dynamic RNA fold structure that interacts with a regulator (such as a small molecule) to change its conformation and function.

Stem-loop RNA: Small regions of RNA that can fold back on themselves to form base pairs.

Synbiogenic: The generation of genetic novelty via the stable interaction with a symbiotic organism.

tRNA: Transfer RNA; the small clover-leaf-shaped RNA that binds to the ribosome and is responsible for providing the correct amino acid in the triplet genetic code.

UTR: Untranslated region. Sequences of DNA that are regulatory and do not code for protein synthesis.

Virolution: A term coined by Frank Ryan to describe host evolution mediated by the action (selection, protection, integration) of viruses.

Virosphere: The extended virus composition of a biological habitat.

REVIEW QUESTIONS

1. What is meant by the term *virosphere*?
2. What is meant by the term *virolution*? Explain how virolution promotes evolutionary novelty and why is this process more efficient than the serial selection from errors.
3. Explain *quasispecies* and QS-C. How do these two differ?
4. What are the individual agents in QS-C? How do they recognize each other? How is membership in the consortium determined and regulated?
5. Explain how the fitness of QS-C depends on diversity of its members.
6. Show the application of the concept of the ribozyme QS societies to the RNA world hypothesis. Why is it essential that one considers the QS societies rather than the individual replicators?

7. Spontaneous cooperating networks of stem-loop RNA replicators are at the origins of the RNA world. Discuss the concepts of networks versus *social* networks that have membership criteria. Why is membership important?

8. A QS society is specific for its members. Provide experimental data that illustrate this.

REFERENCES

Altman, S., 1999. The RNA world. *Scholar* 90812: 130421.

Arbiza, J., Mirazo, S., Fort, H., 2010. Viral quasispecies profiles as the result of the interplay of competition and cooperation. *BMC Evol. Biol.* 10, 137.

Arenas, C.D., Lehman, N., 2010. Quasispecies-like behavior observed in catalytic RNA populations evolving in a test tube. *BMC Evol. Biol.* 10, 80.

Barry, G., Mattick, J.S., 2012. The role of regulatory RNA in cognitive evolution. *Trends Cogn. Sci.* 16, 497–503.

Behrouzi, R., Roh, J.H., Kilburn, D., Briber, R.M., Woodson, S.A., 2012. Cooperative tertiary interaction network guides RNA folding. *Cell* 149, 348–357.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. *Science* 304, 1321–1325.

Bird, C.P., Stranger, B.E., Liu, M., Thomas, D.J., Ingle, C.E., Beazley, C., Miller, W., Hurles, M.E., Dermitzakis, E.T., 2007. Fast-evolving noncoding sequences in the human genome. *Genome Biol.* 8, R118.

Borderia, A.V., Lorenzo-Redondo, R., Pernas, M., Casado, C., Alvaro, T., Domingo, E., Lopez-Galindez, C., 2012. Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. *PLoS ONE* 5, e10319.

Briones, C., de Vicente, A., Molina-Paris, C., Domingo, E., 2006. Minority memory genomes can influence the evolution of HIV-1 quasispecies in vivo. *Gene* 384, 129–138.

Briones, C., Domingo, E., 2008. Minority report: Hidden memory genomes in HIV-1 quasispecies and possible clinical implications. *AIDS Rev.* 10, 93–109.

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

Bush, E.C., Lahn, B.T., 2005. Selective constraint on noncoding regions of hominid genomes. *PLoS Comput. Biol.* 1, e73.

Chaudhari, N., Hahn, W.E., 1983. Genetic expression in the developing brain. *Science* 220, 924–928.

Cheng, L.K.L., Unrau, P.J., 2010. Closing the circle: Replicating RNA with RNA. *Cold Spring Harb. Perspect. Biol.* 2, a002204.

Dadon, Z., Wagner, N., Cohen-Luria, R., Ashkenasy, G., 2012. Reaction networks, in: P.A. Gale and J.W. Steed (eds.), *Supramolecular Chemistry*. John Wiley & Sons, Ltd., New York.

Dick, T.P., Schamel, W.A., 1995. Molecular evolution of transfer RNA from two precursor hairpins: Implications for the origin of protein synthesis. *J. Mol. Evol.* 41, 1–9.

Di Giulio, M., 2013. A polyphyletic model for the origin of tRNAs has more support than a monophyletic model. *J. Theor. Biol.* 318, 124–128.

Domingo, E., Gomez, J., 2007. Quasispecies and its impact on viral hepatitis. *Virus Res.* 127, 131–150.

Domingo, E., Parrish, C.R., Holland, J.J., 2008. *Origin and Evolution of Viruses*, 2nd edn. Academic Press, San Diego, CA.

Domingo, E., Sheldon, J., Perales, C., 2012. Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev.* 76, 159–216.

Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603.

Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–522.

Ferre-D'Amare, A.R., 2011. Use of a coenzyme by the glmS ribozyme–riboswitch suggests primordial expansion of RNA chemistry by small molecules. *Philos. Trans. R. Soc. Lond.* 366, 2942–2948.

Fujishima, K., Sugahara, J., Miller, C.S., Baker, B.J., Di Giulio, M., Takesue, K., Sato, A., Tomita, M., Banfield, J.F., Kanai, A., 2011. A novel three-unit tRNA splicing endonuclease found in ultrasmall Archaea possesses broad substrate specificity. *Nucleic Acids Res.* 39, 9695–9704.

Gesteland, R.F., Cech, T., Atkins, J.F., 1999. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA, Cold Spring Harbor Monograph Series, 37.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Grande-Perez, A., Lazaro, E., Lowenstein, P., Domingo, E., Manrubia, S.C., 2005. Suppression of viral infectivity through lethal defection. *Proc. Natl. Acad. Sci. USA* 102, 4448–4452.

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

Holmes, B., 2012. First life may have survived by cooperating. *New Scientist* 216, 10.

Hordijk, W., Steel, M., 2012. A formal model of autocatalytic sets emerging in an RNA replicator system. arXiv:1211.3473.

Jalasvuori, M., Jalasvuori, M.P., Bamford, J.K.H., 2010. Dynamics of a laterally evolving community of ribozyme-like agents as studied with a rule-based computing system. *Orig. Life Evol. Biosph.* 40, 319–334.

Keightley, P.D., Lercher, M.J., Eyre-Walker, A., 2005. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* 3, e42.

Keightley, P.D., Lercher, M.J., Eyre-Walker, A., 2006. Understanding the degradation of hominid gene control. *PLoS Comput. Biol.* 2, e19 (author reply e26).

Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S. et al., 2004. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* 14, 1462–1473.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K. et al., 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* 106, 11667–11672.

Koonin, E.V., 2011. *The Logic of Chance: The Nature and Origin of Biological Evolution.* FT Press, Upper Saddle River, NJ.

Kumar, R.M., Joyce, G.F., 2003. A modular, bifunctional RNA that integrates itself into a target RNA. *Proc. Natl. Acad. Sci. USA* 100, 9738–9743.

Lauring, A.S., Andino, R., 2010. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6, e1001005.

Lozada-Chavez, I., Stadler, P.F., Prohaska, S.J., 2011. "Hypothesis for the modern RNA world": A pervasive non-coding RNA-based genetic regulation is a prerequisite for the emergence of multicellular complexity. *Orig. Life Evol. Biosph.* 41, 587–607.

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

Mattick, J.S., 2005. The functional genomics of noncoding RNA. *Science* 309, 1527–1528.

Mattick, J.S., 2010. The central role of RNA in the genetic programming of complex organisms. *Acad. Bras. Cienc.* 82, 933–939.

Mattick, J.S., 2011. The central role of RNA in human development and cognition. *FEBS Lett.* 585, 1600–1616.

Mattick, J.S., Mehler, M.F., 2008. RNA editing, DNA recoding and the evolution of human cognition. *Trends Neurosci.* 31, 227–233.

Mattick, J.S., Taft, R.J., Faulkner, G.J., 2010. A global view of genomic information—Moving beyond the gene and the master regulator. *Trends Genet.* 26, 21–28.

Mehler, M.F., Mattick, J.S., 2007. Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol. Rev.* 87, 799–823.

Mercer, T.R., Dinger, M.E., Mariani, J., Kosik, K.S., Mehler, M.F., Mattick, J.S., 2008. Noncoding RNAs in long-term memory formation. *Neuroscientist* 14, 434–445.

Mills, R.E., Luttig, C.T., Larkins, C.E., Beauchamp, A., Tsui, C., Pittard, W.S., Devine, S.E., 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* 16, 1182–1190.

Muller, S., Appel, B., Krellenberg, T., Petkovic, S., 2012. The many faces of the hairpin ribozyme: Structural and functional variants of a small catalytic RNA. *IUBMB Life* 64, 36–47.

Ojosnegros, S., Perales, C., Mas, A., Domingo, E., 2011. Quasispecies as a matter of fact: Viruses and beyond. *Virus Res.* 162, 203–215.

Ono, M., Toh, H., Miyata, T., Awaya, T., 1985. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: Close evolutionary relationship of type A particle gene to types B and D oncovirus genes. *J. Virol.* 55, 387–394.

Orgel, L.E., Crick, F.H., 1980. Selfish DNA: The ultimate parasite. *Nature* 284, 604–607.

Pereira, L., Rodrigues, T., Carrapiço, F., 2012. A symbiogenic way in the origin of life, in: J. Seckbach (ed.), *Genesis—In The Beginning, Cellular Origin, Life in Extreme Habitats and Astrobiology.* Springer, Dordrecht, the Netherlands, pp. 723–742.

Perreault, J., Weinberg, Z., Roth, A., Popescu, O., Chartrand, P., Ferbeyre, G., Breaker, R.R., 2011. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput. Biol.* 7, e1002031.

Rodin, A.S., Szathmary, E., Rodin, S.N., 2011. On origin of genetic code and tRNA before translation. *Biol. Direct* 6, 14.

Ryan, F., 2009. *Virolution.* Collins, London, U.K.

Sakate, R., Suto, Y., Imanishi, T., Tanoue, T., Hida, M., Hayasaka, I., Kusuda, J., Gojobori, T., Hashimoto, K., Hirai, M., 2007. Mapping of chimpanzee full-length cDNAs onto the human genome unveils large potential divergence of the transcriptome. *Gene* 399, 1–10.

Sun, F.J., Caetano-Anolles, G., 2008. The origin and evolution of tRNA inferred from phylogenetic analysis of structure. *J. Mol. Evol.* 66, 21–35.

Szathmáry, E., 2013. On the propagation of a conceptual error concerning hypercycles and cooperation. *J. Syst. Chem.* 4, 1–4.

Toder, R., Grutzner, F., Haaf, T., Bausch, E., 2001. Species-specific evolution of repeated DNA sequences in great apes. *Chromosome Res.* 9, 431–435.

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

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., Andino, R., 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439, 344–348.

Villarreal, L., Ryan, F., 2011. Viruses in host evolution: General principles and future extrapolations. *Curr. Top. Virol.* 9, 79–89.

Villarreal, L.P., 1999. DNA virus contribution to host evolution, in: E. Domingo, R.G. Webster, and J.J. Holland (eds.), *Origin and Evolution of Viruses.* Academic Press, San Diego, CA, pp. 391–420.

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

Villarreal, L.P., 2006. How viruses shape the tree of life. *Future Virol.* 1, 587–595.

Villarreal, L.P., 2007. Virus–host symbiosis mediated by persistence. *Symbiosis* 44, 1–9.

Villarreal, L.P., 2008. The widespread evolutionary significance of viruses, in: E. Domingo, C.R. Parrish, and J.J. Holland (eds.), *Origin and Evolution of Viruses.* Academic Press, London, U.K.

Villarreal, L.P., 2012. The addiction module as a social force, in: G. Witzany (ed.), *Viruses: Essential Agents of Life.* Springer Science+Business Media, Dordrecht, the Netherlands.

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

Villarreal, L.P., Witzany, G., 2013a. Rethinking quasispecies theory: From fittest type to cooperative consortia. *World J. Biol. Chem.* 4, 71–82.

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

Volff, J.-N., 2006. Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* 28, 913–922.

von Bertalanffy, L., LaViolette, P.A., 1981. *A Systems View of Man*. Westview Press, Boulder, CO.

Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T.D., Toyoda, A., Kuroki, Y., Noguchi, H. et al., 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* 429, 382–388.

Wetterbom, A., Sevov, M., Cavelier, L., Bergstrom, T.F., 2006. Comparative genomic analysis of human and chimpanzee indicates a key role for indels in primate evolution. *J. Mol. Evol.* 63, 682–690.

Witzany, G., 2006. Natural genome-editing competences of viruses. *Acta Biotheor.* 54, 235–253.

Witzany, G., 2009a. A perspective on natural genetic engineering and natural genome editing. Introduction. *Ann. N.Y. Acad. Sci.* 1178, 1–5.

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

Witzany, G., 2011. The agents of natural genome editing. *J. Mol. Cell Biol.* 3, 181–189.

Yarus, M., 2011. The meaning of a minuscule ribozyme. *Philos. Trans. R. Soc. Lond.* 366, 2902–2909.

Zeh, D.W., Zeh, J.A., Ishida, Y., 2009. Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays* 31, 715–726.