From Molecular Entities to Competent Agents: Viral Infection-Derived Consortia Act as Natural Genetic Engineers

Günther Witzany

"To understand a sentence means to understand a language.
To understand a language means to be master of a technique"

(Ludwig Wittgenstein)

Abstract Endogenous viruses and defectives, transposons, retrotransposons, long terminal repeats, non-long 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 live habitat. This means that DNA is not solely a genetic storage medium that serves as an evolutionary protocol, but it is also a species-specific ecological niche. 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. They infect, insert, delete, some cut and paste, others copy and paste and spread within the genome. They change host genetic identities either by insertion, recombination or the epigenetic (re)regulation of genetic content, and co-evolve with the host and interact in a module-like manner. In this respect they play vital roles in evolutionary and developmental processes. In contrast to accidental point mutations, integration at various preferred sites is not a randomly occurring process but is coherent with the genetic content of the host; otherwise, important protein coding regions would be damaged, causing disease or even lethal consequences for the host organism. In contrast to "elements", "entities" and "systems", biological agents are capable of identifying sequence-specific loci of genetic text. They are masters of the shared technique of coherently identifying and combining nucleotides according contextual needs. This natural genetic engineering competence is absent in inanimate nature, and therefore represents a core capability of life.

G. Witzany (✉)
Telos – Philosophische Praxis, Vogelsangstraße 18c, 5111 Buermoos, Austria
e-mail: witzany@sbg.at

G. Witzany (ed.), *Viruses: Essential Agents of Life*, DOI 10.1007/978-94-007-4899-6_20,

1 Introduction

During the first five decades of molecular biology it was a common concept that DNA mainly stored the information from which protein-coded sequences are translated. In contrast to this, increasing amounts of knowledge suggested that the largest parts of the genome do not code for proteins but serve as regulatory elements.

Since Barbara McClintock it became obvious that there are DNA sequences that can move within the genomic content. Mobile genetic elements, transposable elements, genetic parasites and selfish DNA are some of the terms suggested in the attempt to find a correct molecular biological term for nucleotide sequences that move, insert, delete and change the genetic identity of host organisms. These "elements", "entities" and "parasites" now take center stage in discussions regarding regulatory elements in epigenetics and genetics, evolutionary novelties and the coordination of growth and development.

Although the abundance of different terms for these molecular structures and functions is increasing, no unifying perspective is available. Their origins are still in the dark, although some of them, a variety of ribozymatic structures, seem to date back to the early RNA world, and even RNA viruses and their defectives may be older than cellular life (Forterre 2005).

2 Coherent Adaptation from the Cell-First to Virus-First Perspective

2.1 Viral Competences Without Cellular Counterparts

Up until now it has been a mainstream assumption that viruses are escaped genetic elements of cells. Because they cannot replicate without cells, they must have originated later in evolution than the first cells. Increasing empirical data do not fit this picture but better fit the virus-first perspective (Forterre 2005; Villarreal 2005; Koonin et al. 2006). According to these data, RNA and DNA viruses have polyphyletic origins and represent a variety of features that are not present in cellular life (Villarreal and Witzany 2010).

Since viruses with RNA genomes are the only living beings that use RNA as a storage medium, they are considered to be remnants of an earlier RNA world that predated DNA (Forterre 2006; Villarreal 2005; Brüssow 2007; Koonin 2009). Negative-stranded RNA viruses have genome structures and replication patterns that are dissimilar to all known cell types. There is no known similarity between RNA-viral replicases and those of any known cell type. Furthermore there are no references to DNA-viruses having a cellular origin. Also, nucleo-cytoplasmic large DNA viruses such as Mimivirus have no known homologs in either viral or cellular genomes (Holmes 2011). Phylogenetic analyses point to an older time scale, as DNA-repairing proteins of DNA viruses do not have any counterparts in cellular life (Villarreal 2005).

2.2 The Persistent Viral Life Strategy

In addition, the fact that viruses have two completely different life strategies seems to be of major importance: acute viruses that exhibit lytic action induce disease and even death, whereas the life strategy of persistent viruses implies compatible interactions with the host, either by integration into the host genome or within the cell plasma, and these types of virus are non-destructive throughout most life stages of the host (Villarreal 2007; Roossinck 2011).

The persistent lifestyle allows viruses to transmit complex viral phenotypes to the host organism (Hambly and Suttle 2005). Doing so enables the host to broaden its evolutionary potential, which may well lead to the formation of *de novo* nucleotide sequences, a new sequence order and therefore to new phenotypes (Frost et al. 2005). Some endogenized retroviruses are still active if expressed, such as the Human Endogenous Retrovirus (HERV) family playing crucial roles in the placentation of mammals, whilst others remain as defectives, such as *env*, *gag* and *pol*, and play co-opted roles in host gene regulation (Gao et al. 2003; Gimenez et al. 2009). Retroviruses identify transcription factor binding sites as being integration relevant (Felice et al. 2009) and non-retroviral RNA-viruses can also become integrated within vertebrate genomes (Klenerman et al. 1997; Geuking et al. 2009). Interestingly also persistent DNA viruses within mammalian genomes have been reported recently (Horie and Tomonaga 2011). Ribozymatic structures that autocatalyze and are active as ensembles, such as tRNAs, editosomes, spliceosomes and ribosomes, are also important modules of RNA viruses. As endogenized modules, they regulate the genetic expression of host organisms (Feschotte 2008; Witzany 2011a).

The natural genome editing competencies of viruses are most complex in bacteria, in which the complete nucleotide word order is largely determined, combined and recombined by viruses (Witzany 2011b). Hence, the main genomic novelties are found in the prokaryotic domain from where they originally evolved into higher life forms: probably all basic enzymatic variations originated therein (Villarreal 2005, 2009a), whereas later evolutionary novelties seem to be the result of a great variety of modified gene regulations (Hunter 2008).

2.3 Persistent Viral Settlers in All Cellular Genomes

Massive viral colonization occurred from the very beginning of cellular life, starting with the evolution of Bacteria and Archaea and, as recently suggested, Eukarya in parallel (Boyer et al. 2011; Forterre 2011). The formation of all kingdoms, their families, genera and species relies on the effects of multiple viral colonization events and results in diversified lineages and ultimately in the evolution of new species (Villarreal 2005).

Today, viruses are recognized as being the most abundant life form in the oceans. It is estimated that 10^{30} viruses live in the ocean and that 10^{23} viral infection events occur per second. They are the major source of mortality to all living agents in the

sea but, on the contrary, are also major settlers in the genomes of sea organisms that serve as immune functions against infections by closely related viruses (Suttle 2007; Villarreal 2011).

Increasing levels of complexity and diversity occur through variation, i.e., inheritable genetic innovation, new combinatorial patterns of genetic content and a variety of non-coding RNAs that serve as regulatory networks and modify the genomic content (Witzany 2009a; Domingo 2011). Transposable elements in cellular genomes are the most likely remnants of viral infection events (Goodier and Kazazian 2008; Villarreal 2009b; O'Donnell and Burns 2010). In addition, the repeat sequences of mobile genetic elements such as LINEs, SINEs, LTR-retroposons, non-LTR-retroposons and ALUs are clearly related to retroviruses, as are reverse transcriptases (Batzer and Deininger 2002; Eickbush 2002). Also, repeat sequences found in telomeres and centromeres are most likely to be of viral origin (Witzany 2008). There are strong indicators that, because of their repetitive sequences, the various non-coding RNAs are derived from retroviral infection events and currently act as modular tools for cellular needs (Weber 2006; Witzany 2009b).

3 Transposable Elements (TEs)

The crucial step in understanding mobile genetic elements was the insight by Barbara McClintock that control elements are able to change their chromosomal location (Hua-Van et al. 2011).

- Transposable elements must be able to differentiate between self and non-self, which means being able to differentiate between endogenous and exogenous agents (Malone and Hannon 2009).
- Transposable elements are the major driving force in evolution because they are agents that produce variability (Wessler 2006; Shapiro 2011).
- Transposable element copies seem to be individuals, but TE families can be viewed as species and host genomes are their species-specific ecological niches (Le Rouzic et al. 2007; Venner et al. 2009).
- Transposable elements exist in every known eukaryotic, bacterial and archaeal genome. The key enzyme is reverse transcriptase, which is present in eukaryotic telomerases and mobile RNA agents such as retroviruses, group II introns and retroposons (Xiong and Eickbush 1988; Lambowitz and Zimmerly 2004; Eickbush and Jamburuthugoda 2008). Autonomous retroelements of eukaryotes are LTRs, LINEs, DIRS and PLEs. Non-autonomous retroelements are SINEs (which are derived from tRNAs and use LINEs to transpose). Class II elements transpose directly without RNA copy intermediates via cut and paste; some are coupled with host-replication. Some superfamilies are related and class II elements (Polintons, Mavericks, Helitrons) only transpose if one strain is cut on each side (Eickbush and Malik 2002; Kapitonov and Jurka 2008). In eukaryotes, LTR retroposons (Copia, BEL and Gypsy) integrate DNA copies via integrase into host genomes (Delelis et al. 2008). Ginger DNA transposons, with two

subgroups (Ginger 1 and Ginger2/Tdd), are prevalent in eukaryotes (Bao et al. 2010).

- Transposable elements can insert near or within genes and can alter or destroy the gene. Inactivation, spatiotemporal changes in expression, alternative splicing, changes in expression and changes in protein activity can result (Levin and Moran 2011).
- Transposable elements are a major factor in genome expansion. In eukaryotes, TEs are abundant in heterochromatin, centromeres and telomeres. In prokaryotes, TEs are the major reason for genomic variability (Villarreal 2012).
- Transposable elements are controlled by epigenetic markings:

3.1 Epigenetic Marking and Immunity by TEs

Epigenetic marking originally emerged to defend genomes against genetic invaders (Huda et al. 2010). Later on, these elements were used in all gene regulations, especially by higher metazoans and plants, to coordinate lineage-specific gene regulation in developmental processes such as parental imprinting, the cell cycle, germ line development and early embryogenesis (Xiao et al. 2008).

Also some kinds of epigenetic silencing of TEs is known as RNA interference, which uses short non-coding RNAs (e.g., siRNAs, microRNAs, piwiRNAs) (Slotkin and Martienssen 2007). Interestingly, RNAi is able to identify specific sequence orders (Witzany 2009b). Whereas small interfering RNAs are generated from exogenous dsRNAs that lead to the destruction of transcripts, piwiRNAs are derived from long transcripts of transposon-rich genomic sequences. They target repeat sequences, especially TEs, and silencing occurs by multiple coordinated steps such as amplification, RNA destruction, epigenetic modification and heterochromatin formation. The piwiRNAs are germline specific and serve as a genome defense against germline invasions. MicroRNAs are derived from endogenous RNA repeats and serve in a variety of gene expression regulations. They target ALU elements and regulate synaptic plasticity and memory (Bredy et al. 2011). Some viruses interfere with endogenous miRNAs to control host gene expression (Mahajan et al. 2008; Villarreal 2011). Some defense systems act to inactivate TEs through syntax error, found in fungi as repeat-induced point mutations (RIPs).

All eukaryotes share RNAi systems, as indicated by homologs of all three proteins that are part of RNAi (ARG family, Dicer, RdRP) and are found in all three kingdoms. Endogenous DNA is protected from degradation by methylation via restriction/modification modules. Interestingly, the clustered regulatory interspaced short palindromic repeats (CRISPRs) serve as a kind of adaptive immunity in bacteria: sequence parts from foreign mobile genetic elements such as phages and plasmids are integrated between CRISPR regions, where they are transcribed as small RNAs that guide protein complexes that target invading DNA (Marraffini and Sontheimer 2010). Eukaryotic RNAi and prokaryotic CRISPRs are not phylogenetically related, although they are both derived from consortia of infecting viruses (Villarreal 2011).

3.2 TEs Sometimes Adapt Co-opted and Also Exonize

An interesting aspect of evolutionary processes is co-opted adaptation, where the host genome uses TE-encoded functions for purposes other than those originally served. This means that either a complete protein is adapted or only the domain. We know this from the telomeric retroelements HetA and TART, which act telomerase-like, i.e., serve as telomerase to complete chromosome ends. Transposable elements carry sequences into regions that are relevant for regulation, coding or intronic functions. There they may be responsible for changes in functions such as expression, alternative splicing, transcription, start and – very important – termination (Zhsang and Saier 2009). Both classes of TEs can be recruited for cellular functions and thereby lose their mobile features; later on they can be identified fixed in populations as intact open reading frames (Volff 2006), or they are fixed in repetitive sequences that protect chromosome centers (centromeres) and ends (telomeres), inticating related origins (Witzany 2008).

The different levels of co-opted adaptation of TEs by the host genomes may lead to new regulations of prevalent genes or even to new genes (Schmitz and Brosius 2011). If the proteins encoded by TEs are not required, a host genome can use the TE sequences for other purposes that can be beneficial for host genomes, such as non-coding sequences with special open reading frames, or protein-regulated binding sites (Kim and Pritvjard 2007).

Interestingly, a proportion of former TEs are found in exons relevant for protein building (Dixon et al. 2007). The role of exonized TEs is well known in alternative splicing. Also, transcription factor binding sites and other promoter regions are derived from TE sequences (Bourque et al. 2008).

3.3 Non-repeat vs. Repeat Nucleotide Sequences

Transposable elements share repeat sequences as essential parts of their identity (Jurka et al. 2007). This is an important feature because non-repeat sequences are the most relevant part of the protein coding sequences of translational mRNAs, which are a coherent protein coding line-up of exons in which all intronic sequences are spliced out (Shapiro and Sternberg 2005). But repeat sequences are relevant to all vital cellular processes and major players in natural genetic engineering processes, such as:

- transcription (promoters, enhancers, silencers, transcription attenuation, terminators, and regulatory RNAs);
- post-transcriptional RNA processing (mRNA targeting, RNA editing);
- translation (enhancement of SINE, mRNA translation);
- DNA replication (origins, centromeres, telomeres, meiotic pairing and recombination);
- localization and movement, chromatin organization (heterochromatin, nucleosome positioning elements, epigenetic memory, methylation, epigenetic imprinting and modification);

- error correction and repair (double-strand break repair by homologous recombination, methyl-directed mismatch repair) and
- DNA restructuring (antigenic variation, phase variation, genome plasticity, uptake and integration of laterally transferred DNA, chromatin diminuation, VDJ recombination, and immunoglobulin class switching) (Sternberg and Shapiro 2005).

3.4 The Largest Family of TEs: ALU Repeats

The Alu repeat family is the largest family of mobile genetic elements in the human genome (Batzer and Deininger 2002; Stoddard and Belfort 2010); Alu repeats contain recognition sites for restriction enzymes. The Alu elements are found in introns and are derived from 7SL RNA genes, which form part of the ribosome complex. The mobilization of Alu elements requires amplification by reverse transcription of an Alu-derived RNA polymerase III transcript. The Alu elements do not have an open reading frame and therefore need some long interspersed nucleotiode elements (LINEs). Insertions of Alu may have positive and negative effects: they can alter the transcription of a gene by changing the methylation status of its promoter. Homologous recombination between dispersed Alu elements can result in genetic exchanges, duplications, deletions and translocations, and 25 % of all simple repeats in primate genomes, including microsatellites, are associated with Alu repeats (Smalheiser and Torvik 2006); Alu repeats are important for alterations in sequence content. The methylation levels of Alu vary in different tissues at different times throughout development. Furthermore, Alu elements act as global modifiers of gene expression through variations in their own methylation status (Batzer and Deininger 2002). Their expression increases as a response to cellular stress and viral and translation inhibition.

4 The Ribosome Acts as a Ribozyme

After pre-translational, translational and post-translational RNA editing (by editosomes), followed by alternative splicing (by spliceosomes), the next crucial step is the correct recognition of the initiation codon of messenger RNA (by ribosomes) (Witzany 2011a). Here, identification of the precise start site for reading the message is crucial for successful decoding (Benelli and Londei 2009). Ribosomes are composed of two-thirds RNA and one-third protein. Ribosomes are assembled into a functional complex. As it is understood today, ribosomal proteins are useful in stabilizing. Only RNAs are found around the catalytic site of the ribosome, with no ribosomal proteins (Belousoff et al. 2010). This means that the ribosome serves as a good example of the co-opted adaptation of a ribozyme (Moore and Steitz 2006).

5 The Two Halves of tRNA

Interestingly, tRNAs did not evolve to serve in protein synthesis first. As demonstrated by Maizels et al. (1999), they represent a composition of formerly different components, with one half serving to mark single-stranded RNA for replication in the RNA world, whereas the lower half of the tRNA was a later acquisition. As demonstrated in nanoarchaeota (Randau and Söll 2008), the various tRNA species are encoded as two half genes, one encoding the conserved T-loops and 3′ acceptor stem, the other encoding the D-stem and the 5′ acceptor stem subunit. In nanoarchaeota, the CCA sequence (which is important in tRNAs for protein synthesis in nearly all cellular life) is not encoded in tRNA genes but is added post-transcriptionally by an enzyme (Xiong and Steitz 2004). It seems that the evolution of protein synthesis is coupled with a variety of older genetic agents and seems to be another example of co-opted adaptation. In agreement with these findings are investigations which demonstrated that pre-tRNAs act in self-cleavage, which is clearly a ribozymatic reaction independent of translation (Phizicky 2005; Wegrzyn and Wegrzyn 2008).

6 Recombination of the RNA Virus

Recombination rates represent different modes of viral genome organization. If recombination occurs in a single genetic segment it is called RNA recombination. The recombination of whole genomic sequences is called reassortment. Copy choice recombination is where RNA polymerase mediates viral replication switches from the donor template to the acceptor template while remaining bound to the nascent nucleic acid chain, thereby generating an RNA sequence with mixed ancestry (Simon-Loriere and Holmes 2011).

Non-homologous recombination can also occur between different genomic regions and non-related RNAs. Defective viruses with long genome deletions compete with fully functional viruses for cellular resources. Reassortment is restricted to viruses that possess segmented genomes and involves the packaging of segments with a different ancestry into a single virion. In complementation, a defective virus can parasitize a fully functional virus that is infecting the same cell. Then, the defective virus can restore its own fitness by borrowing the proteins of the functional virus (Simon-Loriere and Holmes 2011).

Genetic damage (e.g. oxidative stress) is the driving force behind recombination because it forces reverse transcriptase to seek alternative and functional templates. From this perspective, recombination is part of repair. Recombination can then be viewed as a by-product of genome organization. Gene segmentation helps to differentiate transcriptional subunits that can serve as parts of complementation. In this perspective, viruses with segmented genomes serve as a major source of genetic novelty.

7 Context, Not Syntax Determines Meaning

Interacting consortia of endogenous viruses and their defectives serve as actual key regulators in host cells. They cooperate as complementary tools and act as major sources of "variation", i.e. adaptational genetic change and genetic innovation in host organisms. The ability of all these viral-derived agents to identify correct sequence sites for insertion, deletion, reintegration, recombination, repair and translation initiation, as well as inhibition, supports the argument of Manfred Eigen, that the genetic nucleotide sequences of living organisms represent language-like structures and features. Eigen persisted in trying to understand this not metaphorically but literally with molecular syntax and semantics (Witzany 2010).

However, Manfred Eigen failed because he shared the common opinion of the early 1970s, that syntax order in natural languages/codes determines meaning of a given sequence. As we know since Ludwig Wittgenstein ("The meaning of a word is its use within a language") (Witzany 2010), it is the context (pragmatics) in which the living agent is concretely interwoven that determines the meaning (function) of a given sequence. Living agents that use natural languages/codes are able to invent *de novo* sign-sequences as well as reuse sequence parts in novel contextual set ups: natural languages/codes emerge through a consortium of interacting living agents that share a limited number of signs (signaling molecules, symbols) und use them according to combinatorial, context-sensitive and content-coherent rules. With this limited number of signs (characters) and limited number of rules, an identical sequence can even have contradictory semantics (meanings) depending on the situational context in which a sequence-bearing organism is involved.

The most striking example of this adaptive ability is epigenetics. For example, under harmful stress situations or changing environmental conditions, epigenetic marking can change. As reported for plants, such stressful situations can reactivate the genomic sequences of grand and great-grand parents if the genetic features of the parents are not sufficient to react appropriately to the stressful situation (Lolle et al. 2002; Pearson 2005). The fact that retroposons are stress-inducible elements is not only reported in plants, but they can also become active during mammalian maternal stress, which acts during early fetal life and can induce non-Mendelian-inherited epigenetic traits (Huda and Jordan 2009; Huda et al. 2010).

8 Conclusion

Viruses represent the most abundant source of nucleic acids on earth and each cellular organism is infected by multiple viruses and RNA agents of viral origin. The genome ecosphere for competing viral settlers is a rather limited resource. It is most likely that there is no nucleic acid sequence space to be free or unsettled.

Three novel core concepts suggest a fundamental change in our view on life: (i) viruses and ribozymatic interactions predate the evolution of cellular life; (ii) that

are the agents of (epi)genetic invention, recombination, repair and regulation in cellular life; (iii) these agents are able to coherently combine the molecular syntax of nucleic acid language according to contextual needs. This contradicts the most prominent paradigmatic core concept of neo-darwinsm, that chance mutations represent the selection-relevant reason for variation.

The change from a mechanistic view of molecular biology on nucleic acid sequences as random assemblies of physical entities to an agent-based perspective on genetic texts as the result of complex viral-driven natural genetic engineering seems to be on the horizon. Investigations can now focus on action and interaction motifs of persistent viral consortia with their hosts rather than solely on physical and chemical properties. Agent-driven natural genome editing of genetic text sequences is completely absent in inanimate nature. Therefore, the borderline between life and non-life is not only metabolism but the emergence of natural genome editing.

References

Bao W, Kapitonov VV, Jurka J (2010) Ginger DNA transposons in eukaryotes and their evolutionary relationships with long terminal repeat retrotransposons. Mob DNA 1:3

Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3:370–380

Belousoff MJ, Davidovich C, Zimmerman E et al (2010) Ancient machinery embedded in the contemporary ribosome. Biochem Soc Trans 38:422–427

Benelli D, Londei P (2009) Begin at the beginning: evolution of translational initiation. Res Microbiol 160:493–501

Bourque G, Leong B, Vega VB et al (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18:1752–1762

Boyer M, Madoui MA, Gimenez G, La Scola B et al (2011) Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4th domain of life including giant viruses. PLoS One 5(12):e15530

Bredy TW, Lin Q, Wei W, Baker Andresen D et al (2011) Mirco RNA regulation of neural plasticity and memory. Neurobiol Learn Mem 96:89–94

Brüssow H (2007) The quest for food a natural history of eating. Springer, New York

Delelis O, Carayon K, Saib A, Deprez E et al (2008) Integrase and integration: biochemical activities of HIV-I integrase. Retrovirol 5:114

Dixon RJ, Eperon IA, Samani NJ (2007) Complementary intron sequence motifs associated with human exon repetition: a role for intragenic, inter-transcript interactions in gene expression. Bioinformatics 23:150–155

Domingo E (2011) Paradoxical interplay of viral and cellular functions. Viruses 3:272–277

Eickbush TH (2002) Repair by retrotransposition. Nat Genet 31:126–127

Eickbush TH, Jamburuthugoda VK (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134:221–234

Eickbush TH, Malik HS (2002) Evolution of retrotransposons. In: Craig N, Craigie R, Gellert M, Lambowitz A (eds) Mobile DNA II. American Society of Microbiology Press, Washington, DC, pp 1111–1144

Felice B, Cattoglio C, Cittaro D, Testa A et al (2009) Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS One 4:e4571

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9:397–405

Forterre P (2005) The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie 87:793–803
Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5–16
Forterre P (2011) A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. Res Microbiol 162:77–91
Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732
Gao X, Havecker ER, Baranov PV et al (2003) Translational recoding signals between gag and pol in diverse LTR retrotransposons. RNA 9:1422–1430
Geuking MB, Weber J, Dewannieux M, Gorelik ME et al (2009) Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. Science 323:393–396
Gimenez J, Montgiraud C, Oriol G et al (2009) Comparative methylation of ERVWE1/syncytin-1 and other human endogenous retrovirus LTRs in placenta tissues. DNA Res 16:195–211
Goodier JL, Kazazian HH (2008) Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 135:23–35
Hambly E, Suttle CA (2005) The virosphere, diversity and genetic exchange within phage communities. Curr Opin Microbiol 8:444–450
Holmes EC (2011) What does virus evolution tell us about virus origins? J Virol 85:5247–5251
Horie M, Tomonaga K (2011) Non-retroviral fossils in Vertebrate genomes. Viruses 3:1836–1848
Hua-Van A, Le Rouzic A, Boutin TS, Filee J et al (2011) The struggle for life of the genome's selfish architects. Biol Direct 6:19
Huda A, Jordan IK (2009) Epigenetic regulation of mammalian genomes by transposable elements. Ann N Y Acad Sci 1178:276–284
Huda A, Mariño-Ramírez L, Jordan IK (2010) Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. Mob DNA 25:2
Hunter P (2008) The great leap forward major evolutionary jumps might be caused by changes in gene regulation rather than the emergence of new genes. EMBO Rep 9:608–611
Jurka J, Kapitonov VV, Kohany O, Jurka MV (2007) Repetitive sequences in complex genomes: structure and evolution. Annu Rev Genomics Hum Genet 8:241–259
Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in repbase. Nat Rev Genet 9:411–412
Kim SY, Pritvjard JK (2007) Adaptive evolution of conserved noncoding elements in mammals. PLoS Genet 3:e147
Klenerman P, Hengartner H, Zinkernagel RM (1997) A non-retroviral RNA virus persists in DNA form. Nature 390:298–301
Koonin EV (2009) On the origin of cells and viruses: primordial virus world scenario. Ann N Y Acad Sci 1178:47–64
Koonin EV, Senkevich TG, Dolja VV (2006) The ancient virus world and evolution of cells. Biol Direct 1:29
Lambowitz AK, Zimmerly S (2004) Mobile group II introns. Annu Rev Genet 38:1–35
Le Rouzic A, Dupas S, Capy P (2007) Genome ecosystem and transposable elements species. Gene 390:214–220
Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12:615–627
Lolle SJ, Victor JL, Young JM et al (2002) Genome-wide non-mendelian inheritance of extragenomic information in Arabidopsis. Nature 434:505–509
Mahajan VS, Drake A, Chen J (2008) Virus-specific host miRNAs: antiviral defenses or promotors of persistent infection? Trends Immunol 30:1–7
Maizels N, Weiner AM, Yue D et al (1999) New evidence for the genomic tag hypothesis: archaeal CCA-adding enzymes and tRNA substrates. Biol Bull 196:331–334
Malone CD, Hannon GJ (2009) Small RNAs as guardians of the genome. Cell 136:656–668
Marraffini A, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nature 11:181–190

Moore PB, Steitz TA (2006) The roles of RNA in the synthesis of protein. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA world, 3rd edn. Cold Spring Harbor Laboratory Press, New York, pp 257–285
O'Donnell KA, Burns KH (2010) Mobilizing diversity: transposable element insertions in genetic variation and disease. Mob DNA 1:21
Pearson H (2005) Cress overturns textbook genetics. Nature 434:351–360
Phizicky EM (2005) Have tRNA, will travel. Proc Natl Acad Sci USA 102:11127–11128
Randau L, Söll D (2008) Transfer RNA genes in pieces. EMBO Rep 9:623–628
Roossinck MJ (2011) The good viruses: viral mutualistic symbiosis. Nat Rev Microbiol 9:99–108
Schmitz J, Brosius J (2011) Exonization of transposed elements: a challenge and opportunity for evolution. Biochimie 93:1928–1934
Shapiro JA (2011) Evolution: a view from the 21st century. Financial Times Prentice Hall, New York
Shapiro JA, Sternberg R (2005) Why repetitive DNA is essential to genome function. Biol Rev 80:1–24
Simon-Loriere E, Holmes EC (2011) Why do RNA viruses recombine. Nat Rev Microbiol 9:617–626
Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8:272–285
Smalheiser NR, Torvik VI (2006) Alu elements within human mRNAs are probable microRNA targets. Trends Genet 22:532–536
Sternberg R, Shapiro JA (2005) How repeated retroelements format genome function. Cytogenet Genome Res 110:108–116
Stoddard B, Belfort M (2010) Social networking between mobile introns and their host genes. Mol Microbiol 78:1–4
Suttle CA (2007) Marine viruses – major players in the global ecosystem. Nat Rev Microbiol 5:801–812
Venner S, Feschotte C, Biemont C (2009) Transposable elements dynamics: towards a community ecology of the genome. Trends Genet 25:317–323
Villarreal LP (2005) Viruses and the evolution of life. American Society for Microbiology Press, Washington, DC
Villarreal LP (2007) Virus-host symbiosis mediated by persistence. Symbiosis 44:1–9
Villarreal LP (2009a) The source of self: Genetic parasites and the origin of adaptive immunity. Ann NY Acad Sci 1178:194–232
Villarreal LP (2009b) Origin of group identity: Viruses, addiction and cooperation. Springer, New York
Villarreal LP (2011) Viral ancestors of antiviral systems. Viruses 3:1933–1958
Villarreal LP (2012) Viruses and host evolution: virus-mediated self identity. In: Lopez-Larrea C (ed) Self and non-self. LandesBioScience/Springer, Austin
Villarreal LP, Witzany G (2010) Viruses are essential agents within the roots and stem of the tree of life. J Theor Biol 262:698–710
Volff JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28:913–922
Weber MJ (2006) Mammalian small nucleolar RNAs Are mobile genetic elements. PLoS Genet 2:1984–1997
Wegrzyn G, Wegrzyn A (2008) Is tRNA only a translation factor or also a regulator of other processes? J Appl Genet 49:115–122
Wessler S (2006) Eukaryotic transposable elements: teaching old genomes new tricks. In: Caporale L (ed) The implicit genome. Oxford University Press, New York, pp 139–162
Witzany G (2008) The viral origins of telomeres, telomerases and their important role in eukaryogenesis and genome maintenance. Biosem 2:191–206
Witzany G (ed) (2009a) Natural genetic engineering and natural genome editing. Wiley Blackwell, Boston

Witzany G (2009b) Non-coding RNAs: persistent viral agents as modular tools for cellular needs. Ann N Y Acad Sci 1178:244–267
Witzany G (2010) Biocommunication and natural genome editing. Springer, Dordrecht
Witzany G (2011a) The agents of natural genome editing. J Mol Cell Biol 3:181–189
Witzany G (ed) (2011b) Biocommunication in soil microorganisms. Springer, Heidelberg
Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319:1527–1530
Xiong Y, Eickbush TH (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements and mitochondrial introns. Mol Biol Evol 5:675–690
Xiong Y, Steitz TA (2004) Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template. Nature 430:640–645
Zhsang Z, Saier MH (2009) A mechanism of transposon-mediated directed mutation. Mol Microbiol 74:29–43