



Viruses and Evolution – Viruses First? A Personal Perspective

Karin Moelling^{1,2*} and Felix Broecker³

¹ Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland, ² Max Planck Institute for Molecular Genetics, Berlin, Germany, ³ Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

The discovery of exoplanets within putative habitable zones revolutionized astrobiology in recent years. It stimulated interest in the question about the origin of life and its evolution. Here, we discuss what the roles of viruses might have been at the beginning of life and during evolution. Viruses are the most abundant biological entities on Earth. They are present everywhere, in our surrounding, the oceans, the soil and in every living being. Retroviruses contributed to about half of our genomic sequences and to the evolution of the mammalian placenta. Contemporary viruses reflect evolution ranging from the RNA world to the DNA-protein world. How far back can we trace their contribution? Earliest replicating and evolving entities are the ribozymes or viroids fulfilling several criteria of life. RNA can perform many aspects of life and influences our gene expression until today. The simplest structures with non-protein-coding information may represent models of life built on structural, not genetic information. Viruses today are obligatory parasites depending on host cells. Examples of how an independent lifestyle might have been lost include mitochondria, chloroplasts, *Rickettsia* and others, which used to be autonomous bacteria and became intracellular parasites or endosymbionts, thereby losing most of their genes. Even *in vitro* the loss of genes can be recapitulated all the way from coding to non-coding RNA. Furthermore, the giant viruses may indicate that there is no sharp border between living and non-living entities but an evolutionary continuum. Here, it is discussed how viruses can lose and gain genes, and that they are essential drivers of evolution. This discussion may stimulate the thinking about viruses as early possible forms of life. Apart from our view “viruses first”, there are others such as “proteins first” and “metabolism first.”

OPEN ACCESS

Edited by:

Gustavo Caetano-Anollés,
University of Illinois
at Urbana-Champaign, United States

Reviewed by:

Guenther Witzany,
Telos-Philosophische Praxis, Austria
Herve Seligmann,
UMR MEPHI, Aix-Marseille Université,
France

*Correspondence:

Karin Moelling
moelling@molgen.mpg.de

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 04 November 2018

Accepted: 28 February 2019

Published: 19 March 2019

Citation:

Moelling K and Broecker F (2019)
Viruses and Evolution – Viruses First?
A Personal Perspective.
Front. Microbiol. 10:523.
doi: 10.3389/fmicb.2019.00523

Keywords: evolution, RNA world, ribozymes, origin of life, viruses

RIBOZYMES AND VIROIDS

The origin of life on Earth has recently regained attention following the discovery of exoplanets within possible habitable zones (Kasting et al., 1993). The astronomical number of expected exoplanets suggests that there is a high statistical chance that life has also evolved somewhere else. This possibility stimulates the thinking of how life started on the early Earth, which may help to extrapolate to other planets.

Life presumably started simple. This is plausible, but is an assumption. The smallest known bacteria are still rather large. One of the smallest known metabolically autonomous bacterial species is *Pelagibacter ubique* with about 1,400 genes (Giovannoni et al., 2005). Genome reduction of

Mycoplasma mycoides by systematic deletion of individual genes resulted in a synthetic minimal genome of 473 genes (Hutchison et al., 2016). Can one consider simpler living entities?

There are elements with zero genes that fulfill many criteria for early life: ribozymes, catalytic RNAs closely related to viroids. They were recovered *in vitro* from 10^{15} molecules (aptamers), 220 nucleotides in length, by 10 rounds of selection. Among the many RNA species present in this collection of quasispecies RNAs were catalytically active members, enzymatically active ribozymes. The sequence space for 220-mer RNAs is about 3×10^{132} (Eigen, 1971; Wilson and Szostak, 1999; Brackett and Dieckmann, 2006).

The selected ribozymes were able to replicate, cleave, join, and form peptide bonds. They can polymerize progeny chemically, allow for mutations to occur and can evolve. One molecule serves as catalyst, the other one as substrate. Replication of ribozymes was demonstrated in the test tube (Lincoln and Joyce, 2009). Ribozymes can form peptide bonds between amino acids (Zhang and Cech, 1997). Thus, small peptides were available by ribozyme activity. Consequently, an RNA modification has been proposed as peptide nucleic acid (PNA), with more stable peptide bonds instead of phosphodiester bonds (Zhang and Cech, 1997; Joyce, 2002). Replication of RNA molecules can be performed chemically from RNA without polymerase enzymes. In addition, deoxyribozymes can form from ribonucleotides (Wilson and Szostak, 1999). Thus, DNA can arise from RNA chemically, without the key protein enzyme, the reverse transcriptase.

An entire living world is possible from non-coding RNA (ncRNA) before evolution of the genetic code and protein enzymes. Ribozymes naturally consist of circular single-stranded RNAs (Orgel, 2004). They lack the genetic triplet code and do not encode proteins. Instead, they exhibit structural information by hairpin-loops that form hydrogen bonds between incomplete double strands, and loops free to interact with other molecules. They represent a quasispecies in which many species of RNA may form, such as ribozymes, tRNA-like molecules, and other ncRNAs. RNAs within such a pool can bind amino acids. Ninety different amino acids have been identified on the Murchison meteorite found in Australia, while on Earth only about 20 of them are used for protein synthesis (Meierhenrich, 2008). Where formation of ribozymes occurred on the early Earth is a matter of speculation. The hydrothermal vents such as black smokers in the deep ocean are possibilities where life may have started (Martin et al., 2008). There, temperature gradients and clay containing minerals such as magnesium or manganese are available. Pores or niches offer possibilities for concentration of building blocks, which is required for chemical reactions to occur. Interestingly, also ice is a candidate for ribozyme formation and chemical reactions. Ice crystals displace the biomolecules into the liquid phase, which leads to concentration, creating a quasicellular compartmentalization where *de novo* synthesis of nucleotide precursors is promoted. There, RNA and ribozymes can emerge, which are capable of self-replication (Attwater et al., 2010).

tRNA-amino acid complexes can find RNAs as “mRNAs.” Such interactions could have contributed to the evolution of the genetic code. This sequence of events can lead to primitive ribosome precursors. Ribozymes are the essential catalytic elements in ribosomes: “The ribosome is a ribozyme” (Cech,

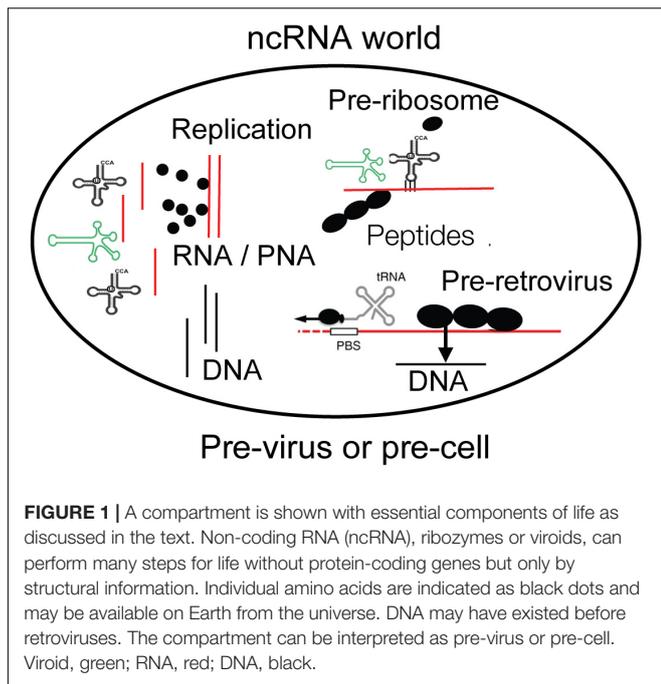
2000), supplemented with about a hundred scaffold proteins later during evolution. The proteins have structural functions and contribute indirectly to enzymatic activity. Are these ribosome-bound ribozymes fossils from the early Earth? Small peptides can be formed by ribozymes before ribosomes evolved, whereby single or dimeric amino acids may originate from the universe (Meierhenrich, 2008).

Small peptides with basic amino acids can increase the catalytic activity of ribozymes as shown *in vitro* (Müller et al., 1994). Such proteins are known as RNA-binding proteins from RNA viruses that protect the RNA genome, with motifs such as RAPRKKG of the nucleocapsid NCp7 of HIV (Schmalzbauer et al., 1996). Peptides can enhance the catalytic activity of ribozymes up to a 100-fold (Müller et al., 1994). Such peptides of RNA viruses serve as chaperones that remove higher ordered RNA structures, allowing for more efficient interaction of RNA molecules and increasing transcription rates of RNA polymerases (Müller et al., 1994). Ribonucleoproteins may have also been functionally important during the evolution of ribosomes (Harish and Caetano-Anolles, 2012).

These pre-ribosomal structures are also similar to precursor-like structures of retroviruses. Reverse transcription can be performed by ribozymes chemically. This action does not necessarily require a protein polymerase such as the reverse transcriptase. Similarly, deoxyribonucleotides can arise by removal of an oxygen without the need of a protein enzyme (a reductase) as today, and allow for DNA polymerization (Wilson and Szostak, 1999; Joyce, 2002). The same elements of the precursors for ribosomes are also building blocks of retroviruses, which may have a similar evolutionary origin (Moelling, 2012, 2013). tRNAs serve as primers for the reverse transcriptase, and the sequence of promoters of transposable elements are derived from tRNAs (Lander et al., 2001). The ribozymes developed into more complex self-cleaving group II introns with insertion of genes encoding a reverse transcriptase and additional proteins (Moelling and Broecker, 2015; Moelling et al., 2017) (**Figure 1**).

It came as a surprise that the genomes of almost all species are rich in ncDNA, transcribed into ncRNAs but not encoding proteins, as evidenced, for instance, by the “Encyclopedia of DNA Elements” (ENCODE) project. ncDNA amounts to more than 98% of the human DNA genome (Deveson et al., 2017). Higher organisms tend to have more non-coding information, which allows for more complex modes of gene regulation. The ncRNAs are regulators of the protein-coding sequences. Highly complex organisms such as humans typically have a high number of ncRNA and regulatory mechanisms. ncRNA can range from close to zero in the smallest bacteria such as *Pelagibacter ubique* to about 98% in the human genome.

RNA viruses such as the retrovirus HIV harbor ncRNAs for gene regulation such as the *trans*-activating response element (TAR), the binding site for the Tat protein for early viral gene expression. Tat has a highly basic domain comprising mostly Lys and Arg residues, resembling other RNA binding proteins. ncRNA also serves on viral RNA genomes as ribosomal entry sites, primer binding sites or packaging signals. DNA synthesis depends on RNA synthesis as initial event, with RNA primers as starters for DNA replication, inside of cells as



well as during retroviral replication, proving a requirement of RNA (Flint, 2015).

The number of mammalian protein-coding genes is about 20,000. Surprisingly, this is only a fifth of the number of genes of bread wheat (Appels et al., 2018). Tulips, maize and other plants also have larger genomes, indicating that the number of genes does not necessarily reflect the complexity of an organism. What makes these plant genomes so large, is still an open question. Could the giant genomes possibly be the result of breeding of plants by farmers or gardeners?

According to Szostak there are molecules which appear like relics from the RNA world such as acetyl-CoA or vitamin B12, both of which are bound to a ribonucleotide for no obvious reason – was it “forgotten” to be removed? (Roberts and Szostak, 1997; Szostak et al., 2001; Szostak, 2011). Perhaps the connected RNA serves as structural stabilizer. Lipid vesicles could have formed the first compartments and enclosed ribozymes, tRNAs with selected amino acids, and RNA which became mRNA. Is this a pre-cell or pre-virus (Figure 1)?

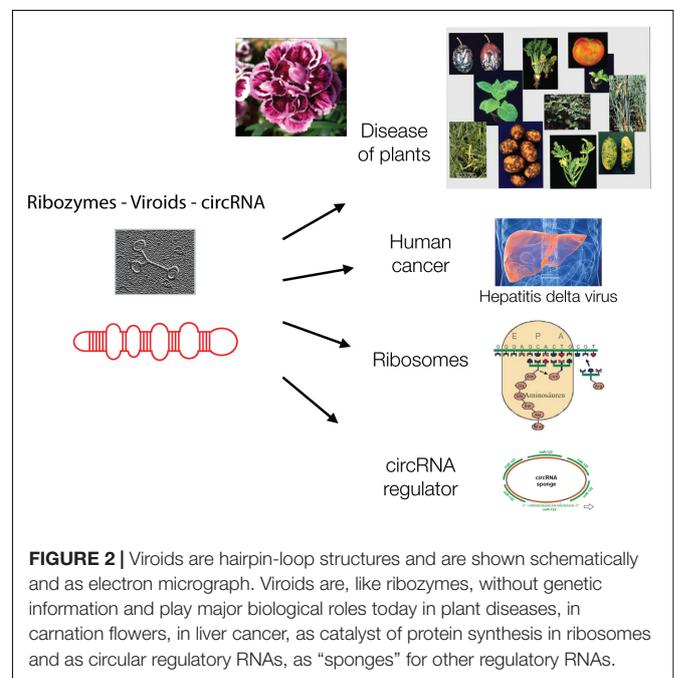
Patel et al. (2015) demonstrated that the building blocks of life, ribonucleotides, lipids and amino acids, can be formed from C, H, O, P, N, S in a “one pot” synthesis. This study can be regarded as a follow-up study of the classical Urey-Miller *in vitro* synthesis of biomolecules (Miller, 1953; Miller and Urey, 1959). Transition from the RNA to the DNA world was promoted by the formation of the reverse transcriptase. The enzyme was first described in retroviruses but it is almost ubiquitous and found in numerous cellular species, many of which with unknown functions (Simon and Zimmerly, 2008; Lescot et al., 2016). It is an important link between the RNA and the DNA worlds. The name reverse transcriptase is historical and irritating because it is the “real” transcriptase during the transition from the RNA to the DNA

world. Similarly, the ribonuclease H (RNase H) is an essential enzyme of retroviruses (Mölling et al., 1971). The RNase H turned out to be one of the five most frequent and ancient proteins (Ma et al., 2008) that belongs to a superfamily of more than sixty different unique representatives and 152 families with numerous functions (Majorek et al., 2014).

Some of the many tRNAs can become loaded with amino acids. There are viruses containing tRNA-like structures (TLS), resembling these early RNAs (Dreher, 2009). The TLS of these viruses typically bind to a single amino acid. TLS-viruses include plant viruses, such as Turnip yellow mosaic virus, in Peanut clump virus, Tobacco mosaic virus (TMV), and Brome mosaic virus. Only half a tRNA is found in Narnaviruses of fungi. The amino acids known to be components of tRNA-like viruses are valine, histidine and tyrosine. The structures were also designated as “mimicry,” enhancing translation (Dreher, 2009, 2010). They look like “frozen” precursor-like elements for protein synthesis. This combination of a partial tRNA linked to one amino acid can be interpreted as an evolutionary early step toward protein synthesis, trapped in a viral element.

Ribozymes are related to the protein-free viroids. Viroids are virus-like elements that belong to the virosphere, the world of viruses (Chela-Flores, 1994). Viroids lack protein coats and therefore were initially not designated as viruses but virus-like viroids when they were discovered in 1971 by Theodor Diener. He described viroids as “living fossils” (Diener, 2016) (Figure 2).

From infected potatoes, Diener isolated the Potato spindle tuber viroid (PSTVd) whose genome was about a 100-fold smaller than those of viruses known at that time. The viroids known today are ranging from 246 to 467 nucleotides. They contain circular single-stranded RNA, are protein-free and self-replicating with no genetic information, but only structural



information in the form of hairpin-loops (Riesner et al., 1979). They can generate copies of themselves in the appropriate environment. They were designated as the “frontiers of life” (Flores et al., 2014).

The knowledge of virus composition was based on TMV and its crystallization by Wendell Stanley in 1935 (Pennazio and Roggero, 2000). The genome of TMV is protein-coding single-stranded RNA of about 6,400 nucleotides that is enclosed by a rod-like protein coat. Viroids, in contrast, do not encode proteins and lack coats but they are closely related to viruses. Viroids can lose their autonomy and rely on host RNA polymerases to replicate, are capable of infecting plants and many are economically important pathogens. There are two families, the nucleus-replicating *Pospiviroidae* such as PSTVd and the chloroplast-replicating *Avsunviroidae* like the Avocado sunblotch viroid (ASBVd). Their replication requires host enzymes. Thus, autonomy is replaced by dependence on host enzymes and an intracellular lifestyle.

Most viroids are often enzymatically active ribozymes – yet they are examples that this trait can get lost as a result of changing environmental conditions. Loss of ribozyme activity is a functional, not a genetic loss. Only the nuclear variants, the *Pospiviroidae*, can lose their ribozyme activity and use the cellular RNase III enzyme for their replication. In contrast, the *Avsunviroidae* are still active hammerhead ribozymes. Thus, inside the nucleus of a host cell, the enzymatic RNA function can become unnecessary. Not genes, but a function, the catalytic activity, gets lost.

Viroids did apparently not gain genes but cooperated for a more complex lifestyle. For example, Carnation small viroid-like RNA (CarSV RNA) cooperates with a retrovirus and is accompanied by a homologous DNA generated by a reverse transcriptase. This enzyme presumably originates from a pararetrovirus of plants. Pararetroviruses package virus particles at a different stage during replication than retroviruses, the DNA, not the RNA. This unique combination between two viral elements has so far only been detected with CarSV in carnation flowers (Flores et al., 2005, 2014). Why did such a cooperation evolve – perhaps by breeding gardeners? RNA is sensitive to degradation; therefore, genetic increase and growth of the genome may not be favorable energetically – at least not in plants. Gain of function is, in this case, cooperation.

The circular RNA (circRNA) is related to ribozymes/viroids as a chief regulator of other regulatory RNAs, a “sponge” absorbing small RNAs. Micro RNAs (miRNAs) are post-transcriptional regulators that are affected by the presence of circRNAs. circRNAs were detected in human and mouse brains and testes as well as in plants. They can bind 70 conserved miRNAs in a cell and amount up to 25,000 molecules (Hansen et al., 2013). Their structure is reminiscent of catalytically active ribozymes.

There is an exceptional viroid that gained coding information and entered the human liver (Taylor, 2009). The viroid is known as hepatitis delta virus (HDV). It has the smallest genome of any known animal virus of about 1,680 nucleotides. It has properties typical of viroids, since it contains circRNA, forms similar hairpin-loops and replicates in the nucleus using host enzymes. Two polymerases have to redirect their specificity from DNA

to RNA to generate the HDV genome and antigenome. Both of them have ribozyme activity. In contrast to other ribozymes, HDV encodes a protein, the hepatitis delta antigen (HDVAg) that occurs in two forms, the small-HDVAg (24 kDa) supporting replication and the large-HDVAg (27 kDa) that helps virion assembly. The gene was presumably picked up from the host cell by recombination of HDV’s mRNA intermediate with a host mRNA. Transmission depends on a helper virus, the Hepatitis B virus (HBV), which delivers the coat (Taylor, 2009). Does packaging by a helper virus protect the genome and thereby allow for a larger viroid to exist?

In plants, viroids may not be able to become bigger possibly due to their sensitivity to degradation – but they cannot become much smaller either. Only a single viroid is known that is completely composed of protein-coding RNA with triplets (AbouHaidar et al., 2014). Viroids and related replicating RNAs are error-prone replicating units and the error frequency imposes a certain minimal size onto them, as they would otherwise become extinct. This mechanism has been described as “error catastrophe,” which prevents survival (Eigen, 1971, 2013). The viroids and related RNAs are the smallest known replicons. Smaller ones would become extinct in the absence of repair systems.

In summary, RNA can catalyze many reactions. Protein enzymes which may have evolved later have higher catalytic activities. Ribozymes are carriers of information, but do not require coding genes. Information is stored in their sequence and structure. Thus, replication of an initial RNA is followed by flow of information, from DNA to RNA to protein, as described the Central Dogma (Crick, 1968). Even an information flow from protein to DNA has been described for some archaeal proteins (Béguin et al., 2015). The DNA-protein world contains numerous ncRNAs with key functions. ncRNA may serve as a model compound for the origin of life on other planets. Hereby not the chemical composition of this molecule is of prime relevance, but its simplicity and multifunctionality. Furthermore, RNA is software and hardware in a single molecule, which makes it unique in our world. There are other scenarios besides the here discussed “virus-first,” such as “protein-first,” “metabolism-first” or the “lipid world” (Segré et al., 2001; Andras and Andras, 2005; Vasas et al., 2010; Moelling, 2012). Some of these alternative concepts were built on phylogenomics, the reconstruction of the tree of life by genome sequencing (Delsuc et al., 2005). Surprisingly, it was Sir Francis Crick, one of the discoverers of the DNA double-helix, who stated that he would not be surprised about a world completely built of RNA. A similar prediction was made by Walter Gilbert (Crick, 1968; Gilbert, 1986). What a vision! Our world was almost 50 years later defined as “RNA-protein” world (Altman, 2013). One can speculate our world was built of ribozymes or viroids, which means “viruses first.”

SPIEGELMAN’S MONSTER

ncRNAs appear as relics from the past RNA world, before DNA, the genetic code and proteins evolved. However, ncRNA is essential in our biological DNA world today. It is possible to

produce such ncRNA today in the test tube by loss of genic information from protein-coding RNA. This reduction to ncRNA was demonstrated *in vitro* with phage RNA. Phage Q β genomic RNA, 4,217 nucleotides in length, was incubated in the presence of Q β replicase, free nucleotides and salts, a rich milieu in the test tube. The RNA was allowed to replicate by means of the Q β replicase. Serial transfer of aliquots to fresh medium led to ever faster replication rates and reduction of genomic size, down to 218 nucleotides of ncRNA in 74 generations. This study demonstrated that, depending on environmental conditions, an extreme gene reduction can take place. This experiment performed in 1965 was designated as “Spiegelman’s Monster.” Coding RNA became replicating ncRNA (Spiegelman et al., 1965; Kacian et al., 1972)!

Manfred Eigen extended this experiment and demonstrated further that a mixture containing no RNA to start with but only ribonucleotides and the Q β replicase can under the right conditions in a test tube spontaneously generate self-replicating ncRNA. This evolved into a form similar to Spiegelman’s Monster. The presence of the replicase enzyme was still necessary in these studies. Furthermore, a change in enzyme concentration and addition of short RNAs or an RNA intercalator influenced the arising RNA population (Sumper and Luce, 1975; Eigen, 2013). Thus, the complexity of genomes depends on the environment: poor conditions lead to increased complexity and rich environments to reduced complexity.

The process demonstrated in this experiment with viral components indicates that reversion to simplicity, reduction in size, loss of genetic information and speed in replication can be major forces of life, even though this appears to be like a reversion of evolution. The experiment can perhaps be generalized from the test tube to a principle, that the most successful survivors on our planet are the viruses and microorganisms, which became the most abundant entities. Perhaps life can start from there again.

These studies raise the question of how RNA molecules can become longer, if the small polymers become smaller and smaller, replicate faster and outcompete longer ones. This may be overcome by heat flow across an open pore in submerged rocks, which concentrates replicating oligonucleotides from a constant feeding flow and selection for longer strands. This has been described for an increase from 100 to 1,000 nucleotides *in vitro*. RNA molecules shorter than 75 nucleotides will die out (Kreysing et al., 2015). Could a poor environment lead to an increase of complexity? This could be tested. Ribozymes were shown to grow in size by uptake of genes, as demonstrated for HDV (Taylor, 2009).

MICROBIOME IN THE HUMAN INTESTINE

An interesting recent unexpected example supporting the notion that environmental conditions influence genetic complexity, is the human gut microbiome. Its complexity increases with diverse food, while uniform rich food reduces its diversity and may lead to diseases such as obesity. Colonization of the human intestinal tract starts at birth. A few dozen bacterial

and viral/phage species are conserved between individuals (core sequences) as a stable composition (Broecker et al., 2016c, 2017). Dysbiosis has been observed in several chronic diseases and in obesity, a loss of bacterial richness and diversity. Nutrition under affluent conditions with sugar-rich food contributes to obesity, which results in a significant reduction of the complexity of the microbiome. This reduction is difficult to revert (Cotillard et al., 2013; Le Chatelier et al., 2013). The gut microbiome in human patients with obesity is reminiscent of the gene reduction described in the Spiegelman’s Monster experiment: reduction of genes in a rich environment.

The reduction of the complexity of the microbiome is in part attributed to the action of phages, which under such conditions, defined as stress, lyse the bacteria. Fecal microbiota transplantation can even be replaced by soluble fractions containing phages or metabolites from the donor without bacteria (Ott et al., 2017). Analogously, the most highly complex microbiomes are found in indigenous human tribes in Africa, which live on a broad variety of different nutrients. It is a slow process, though, to increase gut microbiota complexity by diverse nutrition. The obesity-associated microbiota that survive are fitter and more difficult to counteract. Urbanization and westernization of the diet is associated with a loss of microbial biodiversity, loss of microbial organisms and genes (Segata, 2015).

To understand the mechanism and driving force for genome reduction, deletion rates were tested by insertion of an indicator gene into the *Salmonella enterica* genome. The loss of the indicator gene was monitored by serial passage in rich medium. After 1,000 generations about 25% of the deletions caused increased bacterial fitness. Deletions resulted in smaller genomes with reduced or absence of DNA repair genes (Koskiniemi et al., 2012). Gene loss conferred a higher fitness to the bacteria under these experimental conditions.

MIMIVIRUSES – GAIN OR LOSS?

The recently discovered mimiviruses and other giant viruses are worth considering for understanding the evolution of life with respect to the contribution of viruses. Their hosts are, for example, *Acanthamoeba*, *Chlorella*, and *Coccolithus* algae (*Emiliana huxleyi*), but also corals or sponges as discussed more recently. Mimiviruses were first discovered in cooling water towers in Bradford, United Kingdom in 2003 with about 1,000 genes, most of which unrelated to previously known genes. Mimiviruses have received attention because they contain elements that were considered hallmarks of living cells, not of viruses, such as elements required for protein synthesis, tRNAs and amino acid transferases. The mimiviruses harbor these building blocks as incomplete sets not sufficient for independent protein synthesis as bacteria or archaea can perform, preventing them from leading an autonomous life (La Scola et al., 2003, 2008). They are larger than some bacteria. Giant viruses can be looked at as being on an evolutionary path toward a cellular organism. Alternatively, they may have evolved from a cellular organism by loss of

genetic information (Nasir and Caetano-Anolles, 2015). Giant viruses have frequently taken up genes from their hosts by horizontal gene transfer (HGT) (La Scola et al., 2008; Nasir and Caetano-Anolles, 2015; Colson et al., 2018). A graph on genome sizes shows that mimiviruses and bacteria overlap in size, indicating a continuous transition between viruses and bacteria and between living and non-living worlds (based on Holmes, 2011) (Figure 3). Other giant viruses, such as megaviruses, were discovered in the ocean of Chile with 1,120 genes. Most recently the Klosneuvirus was identified in the sewage of the monastery Klosterneuburg in Austria in 2017 with 1.57 million (mio) basepairs (Mitch, 2017). *Pithovirus sibericum* is the largest among giant viruses discovered to date with a diameter of 1.5 microns, a genome of 470,000 bp with 467 putative genes, 1.6 microns in length, and it is presumably 30,000 years old as it was recovered from permafrost in Siberia (Legendre et al., 2014). The smaller *Pandoraviruses* with 1 micron in length have five times larger genomes, 2,500,000 bp (Philippe et al., 2013) (Figure 3).

The giant viruses can even be hosts to smaller viruses, the virophages, reminiscent of bacteriophages, the viruses of bacteria. These virophages such as Sputnik are only 50 nm in size with 18,343 bp of circular dsDNA and 21 predicted protein-coding genes. They replicate in viral factories and consume the resources of the mimivirus, thereby destroying it. Some virophages can even integrate into the genome of the cellular host and can be reactivated when the host is infected by giant viruses. Thus, giant viruses suggest that viruses are close to living entities or may have been alive (La Scola et al., 2008; Fischer and Hackl, 2016).

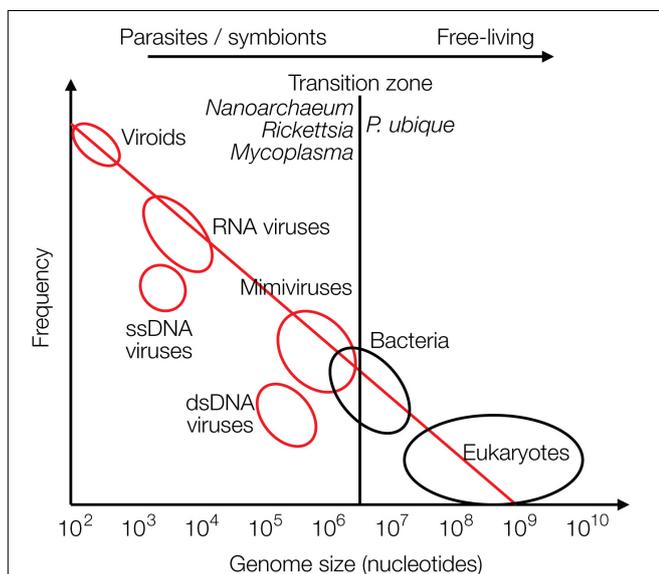


FIGURE 3 | Size distribution of viruses (red circles), free-living bacteria and eukaryotes (black circles) are shown relative to their frequencies. The transition zone between parasites or symbionts versus free-living species is indicated by a black line. The transition is not a sharp borderline as shown by the circles and as discussed in the text (modified from Holmes, 2011).

In biology it is common to distinguish between living and dead matter by the ability to synthesize proteins and replicate autonomously. The giant viruses may be considered as missing link between the two, because they harbor “almost” the protein synthesis apparatus. The transition from living to the non-living world is continuous, not separated by a sharp borderline (Figure 3).

Viruses are not considered alive by most of the scientific community and as written in textbooks, because they cannot replicate autonomously. Yet some of the giant viruses are equipped with almost all components of the protein synthesis machinery close to bacteria suggesting that they belong to the living matter (Schulz et al., 2017). The ribozymes may have been the earliest replicating entity. Perhaps also other viruses were initially more independent of the early Earth than they are today. As described in Figure 1 there may have been initially no major difference between an early virus or an early cell. Only later viruses may have given up their autonomous replication and became parasites – as has been described for some bacteria (see below).

Efforts have been made to identify the smallest living cell that is still autonomously replicating. Among the presumably smallest naturally occurring bacteria is *Pelagibacter ubique* of the SAR11 clade of bacteria (Giovannoni, 2017), which was discovered in 1990. It is an alpha-proteobacterium with 1,389 genes present ubiquitously in all oceans. It can reach up to 10^{28} free living cells in total and represents about 25% of microbial plankton cells. Very little of its DNA is non-coding. It harbors podophage-type phages, designated as “pelagiphage” (Zhao et al., 2013). This small bacterium was designated as the most common organism on the planet. Why is it so successful? This autonomous bacterium is smaller than some parasitic giant viruses. Craig Venter, who first succeeded in sequencing the human genome, tried to minimize the putative smallest genome of a living species, from *Mycoplasma mycoides*, a parasitic bacterium that lives in ruminants (Gibson et al., 2008, 2010). His group synthesized a genome of 531,000 bp with 473 genes, 149 of them (32%) with unknown functions (Hutchison et al., 2016). Among the smallest parasitic living organisms is *Nanoarchaeum equitans*. It is a thermophile archaeon which lives at 80°C and at pH 6 with 2% salt (Huber et al., 2003). Its genome has a size of 490,000 bp and encodes 540 genes. *N. equitans* is an obligate symbiont of a bigger archaeon, *Ignicoccus* riding on it as on a horse, hence the name (Huber et al., 2003).

The world of viruses covers a range of three logs in size of their genomes: from zero genes to about 2,500 genes amounting to about 2,500,000 bp of DNA. The zero-gene viroids are about 300 bases in length (Figure 3).

The virosphere is the most successful reservoir of biological entities on our planet in terms of numbers of particles, speed of replication, growth rates, and sequence space. There are about 10^{33} viruses on our planet and they are present in every single existing species (Suttle, 2005).

There is no living species without viruses! Viruses also occur freely in the oceans, in the soil, in clouds up to the stratosphere and higher, to at least 300 km in altitude. They populate the human intestine, birth canal, and the outside of the body as

protective layer against microbial populations. Microbes contain phages that are activated during stress conditions such as lack of nutrients, change in temperatures, lack of space and other changes of environmental conditions.

RETROVIRUSES AS DRIVERS OF EVOLUTION

One of the most earth-shaking papers of this century was the publication of the human genome sequence (Lander et al., 2001). About half, possibly even two-thirds of the sequence are composed of more or less complete endogenous retroviruses (ERVs) and related retroelements (REs) (de Koning et al., 2011). REs amplify via copy-and-paste mechanisms involving a reverse transcriptase step from an RNA intermediate into DNA. In addition, DNA transposable elements (TEs) move by a cut-and-paste mechanism. The origin of REs is being discussed as remnants of ancient retroviral germline infections that became evolutionarily fixed in the genome. About 450,000 human ERV (HERV) elements constitute about 8% of the human genome consisting of hallmark retroviral elements like the *gag*, *pol*, *env* genes and flanking long terminal repeats (LTR) that act as promoters (Lander et al., 2001). Howard Temin, one of the discoverers of the reverse transcriptase, in 1985 already described endogenous retrovirus-like elements, which he estimated to about 10% of the human and mouse genome sequence (Temin, 1985). The actual number is about 45% as estimated today (Lander et al., 2001). In some genes such as the *Protein Kinase Inhibitor B (PKIB)* gene we determined about 70% retrovirus-related sequences (Moelling and Broecker, 2015). Is there a limit? Could it have been 100%? Retroviruses are estimated to have entered the lineage of the mammalian genome 550 million years ago (MYA) (Hayward, 2017). Older ERV sequences may exist but are unrecognizable today due to the accumulation of mutations.

ERVs undergo mutations, deletions or homologous recombination events with large deletions and can become as short as solo LTR elements, which are a few hundred bp in length – the left-overs from full-length retroviral genomes of about 10,000 bp. The LTR promoters can deregulate neighboring genes. Homologous recombination events may be considered as gene loss or gene reduction events. It is the assumption that the ERVs, which were no longer needed for host cell defense, were no longer selected for by evolution and consequently deleted as unnecessary consumers of energy.

Eugene Koonin points out that infection and integration are unique events occurring at a fast pace, while loss and gene reduction may take much longer time frames (Wolf and Koonin, 2013).

A frequent gene reduction of eukaryotic genomes is the loss of the viral envelope protein encoded by the *env* gene. Without a coat, retroviruses can no longer leave the cell and infect other cells. They lose mobility and become obligatory intracellular elements. Helper viruses can supply envelope proteins in *trans* and mobilize the viruses. TEs or REs can be regarded as examples of coat-free intracellular virus relics – or could it have been the other way round, perhaps precursors of full-length retroviruses?

These elements can be amplified intracellularly and modify the host genomes by integration with the potential danger of gene disruption and genetic changes. REs can lead to gene duplications and pseudogene development, with one copy for stable conservation of acquired functions and the other one for innovations (Cotton and Page, 2005). Such duplications constitute large amounts of mammalian genomes (Zhang, 2003). Retroviruses have an RNase H moiety duplication, one of which serves as a catalytically inactive linker between the RT polymerase and the enzymatically active RNase H (Xiong and Eickbush, 1990; Malik and Eickbush, 2001; Moelling and Broecker, 2015; Moelling et al., 2017). This gene duplication dates back to 500 mio years (Cotton and Page, 2005).

Gene duplications are a common cause of cancer, which often occurs only in the genome of the cancer cell itself, less affecting offsprings. Myc, Myb, ErbB2, Ras, and Raf are oncogenes amplified in diverse types of human cancers (Vogelstein and Kinzler, 2002). The ability of retroviruses to integrate makes them distinct from endosymbionts which stay separate. Yet the net result is very similar, acquisition of new genetic information, which is transmitted to the next generation, if the germline is infected and endogenization of the virus occurred.

Viral integration is not limited to eukaryotic cells but also a mechanism in prokaryotes for maintenance of the lysogenic state of phages inside bacteria.

Also, for other eukaryotic viruses such as HBV, the envelope surface antigen BHsAg can be deleted, which leads to an obligatory intracellular life style for the virus, which especially in the presence of HCV promotes cancer (Yang et al., 2016).

HIV has been shown to rapidly lose one of its auxiliary genes, *nef*, originally for negative factor. The gene was lost within a rather low number of passages of the virus grown under tissue culture conditions by selection for high virus titer producing cells. Deletion of *nef* resulted in a significant increase of the virus titer in culture – hence the name. The *nef* gene product was of no need inside tissue culture cells, rather it was inhibitory for replication. However, it is essential for pathogenicity in animals, and subsequently *nef* was reinterpreted as “necessary factor” (Flint, 2015).

Also, the human hosts of HIV can lose a significant terminal portion of a seven transmembrane receptor in lymphocytes, the primary target cell for HIV entry and for virus uptake. This molecule, the CCR5 cytokine receptor is truncated by 32 carboxy-terminal amino acids (CCR5-Δ32), disabling the receptor functionally. The allele frequency of the mutant CCR5-Δ32 mutant is about 10% in the European population, making these people resistant to HIV infections (Solloch et al., 2017). This gene loss in Europeans has been shown to make the individuals resistant not only against HIV infection but also against malaria. This may have been the selective pressure in the past before HIV/AIDS arose. No side effect for humans lacking this gene has been described (Galvani and Slatkin, 2003).

Viruses have been proven to be drivers of evolution (Villarreal and Witzany, 2010), including the human genome, which by at least 45% is composed of sequences related to retroviruses. In addition, endogenized retroviruses supplied the syncytin genes that are essential for the development of the mammalian placenta,

and allowed the growth of embryos without its rejection by the maternal immune system (Dupressoir et al., 2012). Thus, the same property which causes immunodeficiency in HIV-infected patients and leads to AIDS causes syncytia formation, cell fusion after infection by a retrovirus. Viruses have also been proposed to be at the origin of the evolution of adaptive immunity (Villarreal, 2009). Thus, viruses shaped genomes by supplying essential genes and mechanisms.

ENDOGENIZATION OF VIRUSES

Endogenization of retroviruses has occurred in the mammalian genomes for at least 550 mio years (Hayward, 2017). If the integrated ERVs did not provide any selective advantage, they deteriorated and accumulated mutations with loss of function. This was directly proven by reconstruction of an infectious retrovirus from the consensus sequence of 9 defective endogenous virus sequences, designated as Phoenix. The virus was expressed from a constructed synthetic DNA clone in cell culture and formed virus particles identified by high resolution microscopic analysis (Dewannieux and Heidmann, 2013).

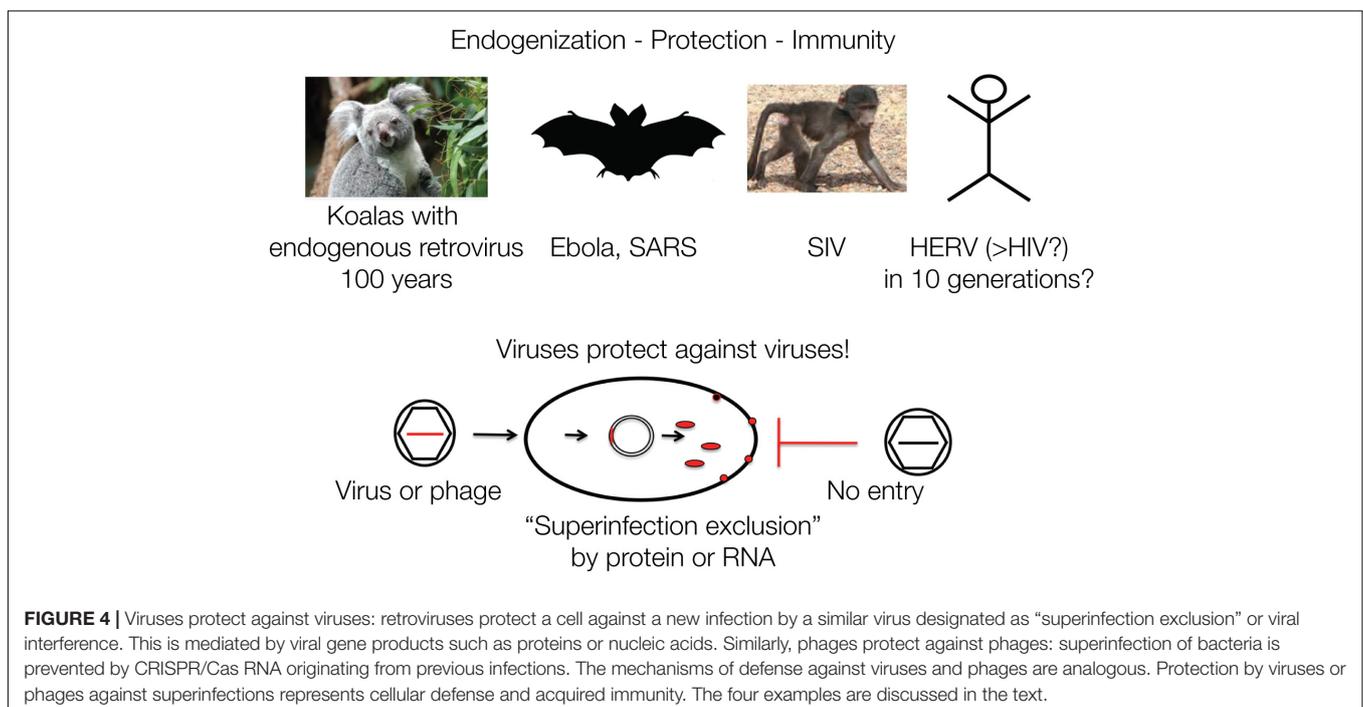
The koalas in Australia are currently undergoing endogenization of a retrovirus (koala retrovirus, KoRV) in “real time” and demonstrate possible consequences for immunity. In the early 1900s, some individuals were transferred to islands, including Kangaroo Island, close to the Australian mainland for repopulation purposes, as koalas were threatened to become extinct. Today, the majority of the koala population is infected by KoRV, which is closely related to the Gibbon ape leukemia virus (GALV). Yet, koalas isolated on Kangaroo Island are KoRV negative, which allows dating the introduction of

KoRV into the koala population to about one hundred years ago. Many of the infected koalas fell ill and died, yet some populations became resistant within about 100 years, corresponding to about 10 generations. The koalas likely developed resistance due to the integrated DNA proviruses. The retrovirus is transmitted as exogenous as well as endogenous virus, similar to the Jaagsiekte sheep retrovirus (JSRV), whereby the endogenized viruses protect with a viral gene product, such as Env, against *de novo* infections by “superinfection exclusion” (Tarlinton, 2012). The contribution of retroviruses to the antiviral defense is striking, since all retroviral genes have analogous genes in the siRNA/RNAi defense mechanism of eukaryotic cells (Moelling et al., 2006).

VIRUSES PROTECT AGAINST VIRUSES

Retroviruses can protect against infection by other related viruses, for example, by expressing Env proteins that block cell-surface receptors (Villarreal, 2011). A comparable mechanism protects bacterial cells against DNA phages, by integrated phage DNA fragments that are transcribed into mRNA and hybridize to incoming new DNA phages and thereby lead to their destruction by hybrid-specific nucleases, CRISPR/Cas immunity (Charpentier and Doudna, 2013). It is often not realized that immunity acquisition in bacteria and mammalian cells follow analogous mechanisms (Figure 4).

Integration of retroviruses normally occurs in somatic cells after infection as an obligatory step during the viral life cycle. Infection of germline cells can lead to transmission to the next generation and ultimately result in inherited resistance. Endogenized retroviruses likely caused resistance



to the exogenous counterparts. Similarly, resistance to Simian Immune Deficiency virus (SIV) in some monkey species may be explained by endogenization (Li et al., 2017, 2018). In the case of phages and their prokaryotic hosts the mechanism is described as CRISPR/Cas, which follow analogous principles of “endogenization” of incoming genetic material for subsequent exclusion.

One may speculate that HIV may also eventually become endogenized into the human genome. There is some evidence that HIV can infect human germline cells and can be transmitted to the embryonic genome (Wang et al., 2011). How long this may take is not known – 10 generations?

The loss of function of ERVs can occur by mutations, deletions of the *env* or other genes and ultimately all coding genes by homologous recombination, leaving behind only one LTR. The number of retrovirus-like elements add up to about 450,000, corresponding to 8% of the human genome (Lander et al., 2001; Cordaux and Batzer, 2009). The promoter regions were analyzed for their contribution to cancer by activating neighboring genes – as a consequence of a former retrovirus infection. Indeed, activated cellular genes by “downstream promotion” were identified in animal studies with activation of the *myc* gene as one of many examples, leading to chronic, not acute development of cancer (Ott et al., 2013). As a general mechanism for human cancer today the LTRs are, however, not identified as a major culprit. Most of the ERVs we find today have been integrated during evolution in introns or other regions where their presence is relatively harmless. Did the other ones result in death of the carriers which disappeared? The effects of LTRs on the expression levels of neighboring host genes was studied with the endogenous human virus, HERV-K, as a possible cause of cancer, but this appears not to be a general phenomenon (Broecker et al., 2016b). As shown for the koalas, ERVs can confer immunity to viral infections (Feschotte and Gilbert, 2012).

A related ERV, HERV-H, was shown to produce an RNA that keeps early embryonic cells pluripotent and even revert adult cells to regain pluripotency (Grow et al., 2015). Thus, the role of ERVs may be more complex than we presently know.

Transposable elements and REs that lost the ability of cellular transmission by deletion of the coat protein majorly contribute to genetic complexity of host cells. They are “locked” inside the cells and are major drivers of the increase of genetic complexity (Cordaux and Batzer, 2009). One could speculate that these intracellular elements are replication-incompetent retroviruses lacking coats (Lander et al., 2001). Bats transmit viruses such as Ebola and SARS coronavirus without suffering from disease (Beltz, 2018). Even RNA viruses such as Bornaviruses have been shown to integrate by illegitimate reverse transcription, possibly also supplying immunity against superinfection (Katzourakis and Gifford, 2010).

ENDOSYMBIOSIS

There are two prominent events that significantly contributed to the success of life and the formation of cells. Both of them are associated with gene reduction. This phenomenon may play a

role for the evolution of viruses from autonomous to parasitic lifestyles. In the 1960s Lynn Margulis proposed an extracellular origin for mitochondria (Margulis, 1970, 1993). An ancestral cell, perhaps an archaeon, was infected by an anaerobic bacterium, which gave rise to mitochondria. Similarly, cyanobacteria formed the chloroplasts in modern plant cells. Mitochondria arose around 1.45 billion years ago (BYA) (Embley and Martin, 2006). Mitochondria and chloroplasts are the most striking examples for a change in lifestyle from autonomous bacteria to endosymbionts. This transition is often considered as extremely rare and a hallmark of evolution of life on our planet. However, there are many other obligate intracellular parasites such as *Rickettsia*, *Chlamydia trachomatis*, *Coxiella burnetii* (the causative agent of Q fever), *Mycobacterium leprae*, *M. tuberculosis*, and *M. mycoides* (Beare et al., 2006).

The change of lifestyle of the endosymbionts in the two cases of mitochondria and chloroplasts is striking. Both of them drastically reduced their genetic make-up. Mitochondria contain less than 37 genes, left from the original about 3,000 genes. Is endogenization of retroviruses, the ERVs, which are integrated into germline cells, related to endosymbiosis? Are these endosymbionts models for the transition from autonomous lifestyle to a parasitic life- which may have taken place with viruses?

A more recent typical example for a reductive evolution are *Rickettsia*. These bacteria were assumed for some time to be viruses because of their obligatory intracellular parasitic existence. *Rickettsia* have evolved from autonomously replicating bacteria. Reductive evolution of endosymbionts can yield bacteria with tiny genomes on the expense of autonomous extracellular life. Their genomes are 1.11 mio bp in length with about 834 protein-coding genes, and loss of 24% by reductive evolution (Ogata et al., 2001). *Rickettsia* may have some relationship with cyanobacteria, which are considered as the major symbionts.

Can one speculate that viruses may have been autonomous entities initially? Viroids may have undergone transition from autonomy to parasites, just as shown for mitochondria, chloroplasts or *Rickettsia*? To which extent have viruses been autonomous and independent of cellular metabolisms originally – and contributed to the origin of cells? Could they only later have lost their autonomy and become parasitic?

VIRUSES AND ONCOGENES

Viruses are minimalistic in their composition and must have undergone stringent gene reductions (Flint, 2015). How small can their genomes become? Most coding RNA viruses still contain regulatory elements, ncRNA at the 3' and 5' terminal regions for ribosomal entry, protein synthesis, transcriptional regulation, and others.

A subgroup of retroviruses is an interesting example in respect to simultaneous loss and gain of genetic information. The oncogenic retroviruses or tumorviruses can recombine with cellular genes which under the promoters of retroviruses can become oncogenes and drivers of cancer. About a hundred

oncogenes have been selected for in the laboratories and studied over decades for understanding the molecular mechanisms of cancer. Selection for growth advantages of the host cells led to the discovery of the fastest growth-promoting oncogenes we know today, such as Ras, Raf, ErbB or Myc, which are in part successful targets for anticancer drugs (Moelling et al., 1984).

These oncogenes were in most cases taken up by the retroviruses at the expense of structural (*gag*), replicating (*pol*) or envelope (*env*) genes, and are often expressed as fusion proteins with Gag. Thus, oncogenic retroviruses are obligatory intracellular defective viruses and were selected for in the laboratory by researchers for the oncogenes with the most potent growth promoting ability. They need the supply of replicatory genes in *trans* from co-infecting helper viruses to infect other cells (Flint, 2015). Retroviruses are able to pick up cellular genes, transfer and integrate them into neighboring cells. Some strains of Rous sarcoma virus maintain replication competent when carrying the cell-derived *src* (for sarcoma) oncogene encoding a protein of 536 amino acids that apparently can fit into the retroviral particle along with the full-size viral genome (Broecker et al., 2016a). Spatial reasons may have influenced the formation of oncogenic retroviruses and limited their size and thereby led to their defective phenotypes.

There are indications that the uncontrolled activity of (retro)transposons in germline cells can result in diseases such as male infertility – presumably by “error catastrophe,” caused by too many transposition events. In mammals, piRNAs tame transposon activity by means of the RNase H activity of PIWI proteins during spermatogenesis (Girard et al., 2006).

Only a minority of viruses are pathogens; most of them do not cause diseases. On the contrary, they are most important as drivers of evolution, as transmitters of genetic material, as innovative agents. In particular, the RNA viruses are the most innovative ones. Some of them are pathogenic and dangerous, such as HIV or influenza virus, or viroids in plants. RNA viruses are able to change so rapidly that the host immune system is unable to counteract the infection. Pathogenicity arises when environmental conditions change, for instance, when a virus enters a new organism or species.

Increase of cellular complexity by viruses is an important feature of evolution. Such major evolutionary changes are recently taken as arguments against the evolutionary theory by Charles Darwin who considered gradual changes, small increments by mutations as the main basis for selection and evolution. New criticism is addressing this thinking, considering larger changes as evolutionary drivers. Such changes arise by many complex phenomena such as endosymbiosis, infection by prokaryotes, viruses and fungi, recombination of genes, HGT, infections, sex. Dramatic changes such as endosymbiosis or pathogen infections extend Darwin’s concept of evolution.

CONCLUSION

There are numerous examples for the contribution of viruses to the evolution of life since at least as long as 550 MYA

(Hayward, 2017). But genetic noise through random mutations does not allow us to go back to the origin of life. It may not be impossible that the earliest compartment was indistinguishable, either a pre-cell or a pre-virus. By analogy one may speculate that at some point autonomous viruses gave up independence for an obligatory intracellular life – as has been described for mitochondria and chloroplasts but also intracellular bacteria such as *Rickettsia*. This speculation is based on the concept that early life must have started simple and with high genetic variability and then became more complex. But complexity can be given up for a less energy consuming lifestyle with small genomes and high speed of replication (Moelling, 2012, 2013). Therefore, the question may be repeated: “Are viruses our oldest ancestors?” Some fossil life can be partially reproduced *in vitro* by Spiegelman’s Monster and Eigen’s follow-up experiments, explaining the great surviving potential of simple ncRNA.

Viruses can be pathogens, but their recognition as primarily causing diseases is wrong. This notion is based on the history of viruses in medicine, as explained in a book entitled “Viruses: More Friends Than Foes” (Moelling, 2017). The scenario described here focuses on viruses as drivers of evolution.

The early RNA world gained interest 20–30 years ago as evidenced by the references provided above. Surprisingly, there are scientists who still believe in the “pansperm hypothesis” and think that retroviruses are of extraterrestrial origin (Steele et al., 2018). The recent interest in the origin of life arose from the newly discovered exoplanets whose number increases daily – and which may be as numerous as 10^{25} . Thus, pure statistics make some people believe that there is extraterrestrial life.

The extraterrestrial life is mimicked in laboratories on Earth with many assumptions – perhaps this overview stimulates some thinking. The discussion presented here should be taken as concept about simple replicating and evolving entities possibly arising from different building blocks in other environments, with structure being more relevant than sequence.

AUTHOR CONTRIBUTIONS

Both authors wrote the manuscript and approved the final version.

FUNDING

FB was financially supported by a fellowship of the German Academy of Sciences Leopoldina.

ACKNOWLEDGMENTS

Dedicated in memoriam to Manfred Eigen (1927–2019) with gratitude for invitations to KM to many inspiring Winter Seminars in Klosters, Switzerland. We would also like to thank Prof. Peter Palese for his generous support.

REFERENCES

- AbouHaidar, M. G., Venkataraman, S., Golshani, A., Liu, B., and Ahmad, T. (2014). Novel coding, translation, and gene expression of a replicating covalently closed circular RNA of 220 nt. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14542–14547. doi: 10.1073/pnas.1402814111
- Altman, S. (2013). The RNA-protein-world. *RNA* 19, 589–590. doi: 10.1261/rna.038687.113
- Andras, P., and Andras, C. (2005). The origin of life - the 'protein interaction world' hypothesis: protein interactions were the first form of self-reproducing life and nucleic acids evolved later as memory molecules. *Med. Hypotheses* 64, 678–688. doi: 10.1016/j.mehy.2004.11.029
- Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A., and Holliger, P. (2010). Ice as a protocellular medium for RNA replication. *Nat. Commun.* 1:76. doi: 10.1038/ncomms1076
- Beare, P. A., Samuel, J. E., Howe, D., Virtaneva, K., Porcella, S. F., and Heinzen, R. A. (2006). Genetic diversity of the Q fever agent, *Coxiella burnetii*, assessed by microarray-based whole-genome comparisons. *J. Bacteriol.* 188, 2309–2324. doi: 10.1128/JB.188.7.2309-2324.2006
- Béguin, P., Gill, S., Charpin, N., and Forterre, P. (2015). Synergistic template-free synthesis of dsDNA by *Thermococcus nautili* primase PolpTN2, DNA polymerase PolB, and pTN2 helicase. *Extremophiles* 19, 69–76. doi: 10.1007/s00792-014-0706-1
- Beltz, L. A. (2018). *Bats and Human Health: Ebola, SARS, Rabies and Beyond*. Hoboken, NJ: Wiley-Blackwell.
- Brackett, D. M., and Dieckmann, T. (2006). Aptamer to ribozyme: the intrinsic catalytic potential of a small RNA. *ChemBiochem* 7, 839–843. doi: 10.1002/cbic.200500538
- Broecker, F., Hardt, C., Herwig, R., Timmermann, B., Kerick, M., Wunderlich, A., et al. (2016a). Transcriptional signature induced by a metastasis-promoting c-Src mutant in a human breast cell line. *FEBS J.* 283, 1669–1688. doi: 10.1111/febs.13694
- Broecker, F., Horton, R., Heinrich, J., Franz, A., Schweiger, M. R., Lehrach, H., et al. (2016b). The intron-enriched HERV-K(HML-10) family suppresses apoptosis, an indicator of malignant transformation. *Mob. DNA* 7:25. doi: 10.1186/s13100-016-0081-9
- Broecker, F., Klumpp, J., and Moelling, K. (2016c). Long-term microbiota and virome in a Zürich patient after fecal transplantation against *Clostridium difficile* infection. *Ann. N. Y. Acad. Sci.* 1372, 29–41. doi: 10.1111/nyas.13100
- Broecker, F., Russo, G., Klumpp, J., and Moelling, K. (2017). Stable core virome despite variable microbiome after fecal transfer. *Gut Microbes* 8, 214–220. doi: 10.1080/19490976.2016.1265196
- Cech, T. R. (2000). Structural biology. The ribosome is a ribozyme. *Science* 289, 878–879.
- Charpentier, E., and Doudna, J. A. (2013). Biotechnology: rewriting a genome. *Nature* 495, 50–51. doi: 10.1038/495050a
- Chela-Flores, J. (1994). Are viroids molecular fossils of the RNA world? *J. Theor. Biol.* 166, 163–166. doi: 10.1006/jtbi.1994.1014
- Colson, P., Levasseur, A., La Scola, B., Sharma, V., Nasir, A., Pontarotti, P., et al. (2018). Ancestrality and mosaicism of giant viruses supporting the definition of the fourth TRUC of microbes. *Front. Microbiol.* 9:2668. doi: 10.3389/fmicb.2018.02668
- Cordaux, R., and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* 10, 691–703. doi: 10.1038/nrg2640
- Cotillard, A., Kennedy, S. P., Kong, L. C., Prifti, E., Pons, N., Le Chatelier, E., et al. (2013). Dietary intervention impact on gut microbial gene richness. *Nature* 500, 585–588. doi: 10.1038/nature12480
- Cotton, J. A., and Page, R. D. (2005). Rates and patterns of gene duplication and loss in the human genome. *Proc. Biol. Sci.* 272, 277–283. doi: 10.1098/rspb.2004.2969
- Crick, F. H. (1968). The origin of the genetic code. *J. Mol. Biol.* 38, 367–379. doi: 10.1016/0022-2836(68)90392-6
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375. doi: 10.1038/nrg1603
- Deveson, I. W., Hardwick, S. A., Mercer, T. R., and Mattick, J. S. (2017). The dimensions, dynamics, and relevance of the mammalian noncoding transcriptome. *Trends Genet.* 33, 464–478. doi: 10.1016/j.tig.2017.04.004
- Dewannieux, M., and Heidmann, T. (2013). Endogenous retroviruses: acquisition, amplification and taming of genome invaders. *Curr. Opin. Virol.* 3, 646–656. doi: 10.1016/j.coviro.2013.08.005
- Diener, T. O. (2016). Viroids: "living fossils" of primordial RNAs? *Biol. Direct* 11:15. doi: 10.1186/s13062-016-0116-7
- Dreher, T. W. (2009). Role of tRNA-like structures in controlling plant virus replication. *Virus Res.* 139, 217–229. doi: 10.1016/j.virusres.2008.06.010
- Dreher, T. W. (2010). Viral tRNAs and tRNA-like structures. *Wiley Interdiscip. Rev. RNA* 1, 402–414. doi: 10.1002/wrna.42
- Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta* 33, 663–671. doi: 10.1016/j.placenta.2012.05.005
- Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523. doi: 10.1007/BF00623322
- Eigen, M. (2013). *From Strange Simplicity to Complex Familiarity*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198570219.001.0001
- Embley, T. M., and Martin, W. (2006). Eukaryotic evolution, changes and challenges. *Nature* 440, 623–630. doi: 10.1038/nature04546
- Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* 13, 283–296. doi: 10.1038/nrg3199
- Fischer, M. G., and Hackl, T. (2016). Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature* 540, 288–291. doi: 10.1038/nature20593
- Flint, S. J. (2015). *Principles of Virology*, 4th Edn. Washington, DC: ASM Press. doi: 10.1128/9781555819521
- Flores, R., Gago-Zachert, S., Serra, P., Sanjuán, R., and Elena, S. F. (2014). Viroids: survivors from the RNA world? *Annu. Rev. Microbiol.* 68, 395–414. doi: 10.1146/annurev-micro-091313-103416
- Flores, R., Hernández, C., Martínez de Alba, A. E., Daròs, J. A., and Di Serio, F. (2005). Viroids and viroid-host interactions. *Annu. Rev. Phytopathol.* 43, 117–139. doi: 10.1146/annurev.phyto.43.040204.140243
- Galvani, A. P., and Slatkin, M. (2003). Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15276–15279. doi: 10.1073/pnas.2435085100
- Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., et al. (2008). Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319, 1215–1220. doi: 10.1126/science.1151721
- Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi: 10.1126/science.1190719
- Gilbert, W. (1986). Origin of life: the RNA world. *Nature* 319:618. doi: 10.1038/319618a0
- Giovannoni, S. J. (2017). SAR11 bacteria: the most abundant plankton in the Oceans. *Ann. Rev. Mar. Sci.* 9, 231–255. doi: 10.1146/annurev-marine-010814-015934
- Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309, 1242–1245. doi: 10.1126/science.1114057
- Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202. doi: 10.1038/nature04917
- Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522, 221–225. doi: 10.1038/nature14308
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993
- Harish, A., and Caetano-Anolles, G. (2012). Ribosomal history reveals origins of modern protein synthesis. *PLoS One* 7:e32776. doi: 10.1371/journal.pone.0032776

- Hayward, A. (2017). Origin of retroviruses: when, where, and how? *Curr. Opin. Virol.* 25, 23–27. doi: 10.1016/j.coviro.2017.06.006
- Holmes, E. C. (2011). What does virus evolution tell us about virus origins? *J. Virol.* 85, 5247–5251. doi: 10.1128/JVI.02203-10
- Huber, H., Hohn, M. J., Stetter, K. O., and Rachel, R. (2003). The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Res. Microbiol.* 154, 165–171. doi: 10.1016/S0923-2508(03)00035-4
- Hutchison, C. A. III, Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. doi: 10.1126/science.aad6253
- Joyce, G. F. (2002). The antiquity of RNA-based evolution. *Nature* 418, 214–221. doi: 10.1038/418214a
- Kacian, D. L., Mills, D. R., Kramer, F. R., and Spiegelman, S. (1972). A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. *Proc. Natl. Acad. Sci. U.S.A.* 69, 3038–3042. doi: 10.1073/pnas.69.10.3038
- Kasting, J. F., Whitmire, D. P., and Reynolds, R. T. (1993). Habitable zones around main sequence stars. *Icarus* 101, 108–128. doi: 10.1006/icar.1993.1010
- Katzourakis, A., and Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet.* 6:e1001191. doi: 10.1371/journal.pgen.1001191
- Koskiniemi, S., Sun, S., Berg, O. G., and Andersson, D. I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet.* 8:e1002787. doi: 10.1371/journal.pgen.1002787
- Kreysing, M., Keil, L., Lanzmich, S., and Braun, D. (2015). Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat. Chem.* 7, 203–208. doi: 10.1038/nchem.2155
- La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299:2033. doi: 10.1126/science.1081867
- La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104. doi: 10.1038/nature07218
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* 500, 541–546. doi: 10.1038/nature12506
- Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., et al. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4274–4279. doi: 10.1073/pnas.1320670111
- Lescot, M., Hingamp, P., Kojima, K. K., Villar, E., Romac, S., Veluchamy, A., et al. (2016). Reverse transcriptase genes are highly abundant and transcriptionally active in marine plankton assemblages. *ISME J.* 10, 1134–1146. doi: 10.1038/ismej.2015.192
- Li, H., Li, L., Liu, L. R., Omange, R. W., Toledo, N., Kashem, M. A., et al. (2018). Hypothetical endogenous SIV-like antigens in Mauritian cynomolgus macaques. *Bioinformatics* 14, 48–52. doi: 10.6026/97320630014048
- Li, H., Nykoluk, M., Li, L., Liu, L. R., Omange, R. W., Soule, G., et al. (2017). Natural and cross-inducible anti-SIV antibodies in Mauritian cynomolgus macaques. *PLoS One* 12:e0186079. doi: 10.1371/journal.pone.0186079
- Lincoln, T. A., and Joyce, G. F. (2009). Self-sustained replication of an RNA enzyme. *Science* 323, 1229–1232. doi: 10.1126/science.1167856
- Ma, B. G., Chen, L., Ji, H. F., Chen, Z. H., Yang, F. R., Wang, L., et al. (2008). Characters of very ancient proteins. *Biochem. Biophys. Res. Commun.* 366, 607–611. doi: 10.1016/j.bbrc.2007.12.014
- Majorek, K. A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K., et al. (2014). The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.* 42, 4160–4179. doi: 10.1093/nar/gkt1414
- Malik, H. S., and Eickbush, T. H. (2001). Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res.* 11, 1187–1197. doi: 10.1101/gr.185101
- Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven, CT: Yale University Press.
- Margulis, L. (1993). *Symbiosis in Cell Evolution*, 2nd Edn. New York, NY: W.H. Freeman and Co.
- Martin, W., Baross, J., Kelley, D., and Russell, M. J. (2008). Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* 6, 805–814. doi: 10.1038/nrmicro1991
- Meierhenrich, U. (2008). *Amino Acids and the Asymmetry of Life*. Berlin: Springer-Verlag.
- Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science* 117, 528–529. doi: 10.1126/science.117.3046.528
- Miller, S. L., and Urey, H. C. (1959). Organic compound synthesis on the primitive Earth. *Science* 130, 245–251. doi: 10.1126/science.130.3370.245
- Mitch, L. (2017). Giant viruses found in Austrian sewage fuel debate over potential fourth domain of life. *Science* doi: 10.1126/science.aal1005
- Moelling, K. (2012). Are viruses our oldest ancestors? *EMBO Rep.* 13:1033. doi: 10.1038/embor.2012.173
- Moelling, K. (2013). What contemporary viruses tell us about evolution: a personal view. *Arch. Virol.* 158, 1833–1848. doi: 10.1007/s00705-013-1679-6
- Moelling, K. (2017). *Viruses, More Friends than Foes*. Toh Tuck: World Scientific Press.
- Moelling, K., and Broecker, F. (2015). The reverse transcriptase-RNase H: from viruses to antiviral defense. *Ann. N. Y. Acad. Sci.* 1341, 126–135. doi: 10.1111/nyas.12668
- Moelling, K., Broecker, F., Russo, G., and Sunagawa, S. (2017). RNase H as gene modifier, driver of evolution and antiviral defense. *Front. Microbiol.* 8:1745. doi: 10.3389/fmicb.2017.01745
- Moelling, K., Heimann, B., Beimling, P., Rapp, U. R., and Sander, T. (1984). Serine- and threonine-specific protein kinase activities of purified gag-mil and gag-raf proteins. *Nature* 312, 558–561. doi: 10.1038/312558a0
- Moelling, K., Matskevich, A., and Jung, J. S. (2006). Relationship between retroviral replication and RNA interference machineries. *Cold Spring Harb. Symp. Quant. Biol.* 71, 365–368. doi: 10.1101/sqb.2006.71.010
- Mölling, K., Bolognesi, D. P., Bauer, H., Büsen, W., Plassmann, H. W., and Hausen, P. (1971). Association of viral reverse transcriptase with an enzyme degrading the RNA moiety of RNA-DNA hybrids. *Nat. New Biol.* 234, 240–243. doi: 10.1038/newbio234240a0
- Müller, G., Strack, B., Dannull, J., Sproat, B. S., Surovoy, A., Jung, G., et al. (1994). Amino acid requirements of the nucleocapsid protein of HIV-1 for increasing catalytic activity of a Ki-ras ribozyme in vitro. *J. Mol. Biol.* 242, 422–429. doi: 10.1006/jmbi.1994.1592
- Nasir, A., and Caetano-Anolles, G. (2015). A phylogenomic data-driven exploration of viral origins and evolution. *Sci. Adv.* 1:e1500527. doi: 10.1126/sciadv.1500527
- Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., et al. (2001). Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science* 293, 2093–2098.
- Orgel, L. E. (2004). Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* 39, 99–123. doi: 10.1080/10409230490460765
- Ott, G., Rosenwald, A., and Campo, E. (2013). Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification. *Blood* 122, 3884–3891. doi: 10.1182/blood-2013-05-498329
- Ott, S. J., Waetzig, G. H., Rehman, A., Moltzau-Anderson, J., Bharti, R., Grasis, J. A., et al. (2017). Efficacy of sterile fecal filtrate transfer for treating patients with clostridium difficile infection. *Gastroenterology* 152, 799–811.e7. doi: 10.1053/j.gastro.2016.11.010
- Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., and Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* 7, 301–307. doi: 10.1038/nchem.2202
- Pennazio, S., and Roggero, P. (2000). The discovery of the chemical nature of tobacco mosaic virus. *Riv. Biol.* 93, 253–281.
- Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., et al. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341, 281–286. doi: 10.1126/science.1239181
- Riesner, D., Henco, K., Rokohl, U., Klotz, G., Kleinschmidt, A. K., Domdey, H., et al. (1979). Structure and structure formation of viroids. *J. Mol. Biol.* 133, 85–115. doi: 10.1016/0022-2836(79)90252-3
- Roberts, R. W., and Szostak, J. W. (1997). RNA-peptide fusion for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12297–12302. doi: 10.1073/pnas.94.23.12297
- Schmalzbauer, E., Strack, B., Dannull, J., Guehmann, S., and Moelling, K. (1996). Mutations of basic amino acids of NCP7 of human immunodeficiency virus type 1 affect RNA binding in vitro. *J. Virol.* 70, 771–777.

- Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356, 82–85. doi: 10.1126/science.aal4657
- Segata, N. (2015). Gut microbiome. Westernization and the disappearance of intestinal diversity. *Curr. Biol.* 25, R611–R613. doi: 10.1016/j.cub.2015.05.040
- Segré, D., Ben-Eli, D., Deamer, D. W., and Lancet, D. (2001). The lipid world. *Orig. Life Evol. Biosph.* 31, 119–145. doi: 10.1023/A:1006746807104
- Simon, D. M., and Zimmerly, S. (2008). A diversity of uncharacterized reverse transcriptases in bacteria. *Nucleic Acids Res.* 36, 7219–7229. doi: 10.1093/nar/gkn867
- Solloch, U. V., Lang, K., Lange, V., Böhme, I., Schmidt, A. H., and Sauter, J. (2017). Frequencies of gene variant CCR5-Δ32 in 87 countries based on next-generation sequencing of 1.3 million individuals sampled from 3 national DKMS donor centers. *Hum. Immunol.* 78, 710–717. doi: 10.1016/j.humimm.2017.10.001
- Spiegelman, S., Haruna, I., Holland, I. B., Beaudreau, G., and Mills, D. (1965). The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 54, 919–927. doi: 10.1073/pnas.54.3.919
- Steele, E. J., Al-Mufti, S., Augustyn, K. A., Chandrajith, R., Coghlan, J. P., Coulson, S. G., et al. (2018). Cause of cambrian explosion - terrestrial or cosmic? *Prog. Biophys. Mol. Biol.* 136, 3–23. doi: 10.1016/j.pbiomolbio.2018.03.004
- Sumper, M., and Luce, R. (1975). Evidence for de novo production of self-replicating and environmentally adapted RNA structures by bacteriophage Qbeta replicase. *Proc. Natl. Acad. Sci. U.S.A.* 72, 162–166. doi: 10.1073/pnas.72.1.162
- Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361. doi: 10.1038/nature04160
- Szostak, J. W. (2011). An optimal degree of physical and chemical heterogeneity for the origin of life? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2894–2901. doi: 10.1098/rstb.2011.0140
- Szostak, J. W., Bartel, D. P., and Luisi, P. L. (2001). Synthesizing life. *Nature* 409, 387–390. doi: 10.1038/35053176
- Tarlington, R. E. (2012). “Koala retrovirus endogenisation in action,” in *Viruses: Essential Agents of Life*, ed. G. Witzany (Berlin: Springer), 283–291. doi: 10.1007/978-94-007-4899-6_14
- Taylor, J. M. (2009). Replication of the hepatitis delta virus RNA genome. *Adv. Virus Res.* 74, 103–121. doi: 10.1016/S0065-3527(09)74003-5
- Temin, H. M. (1985). Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. *Mol. Biol. Evol.* 2, 455–468. doi: 10.1093/oxfordjournals.molbev.a040365
- Vasas, V., Szathmari, E., and Santosa, M. (2010). Lack of evolvability in self-sustaining autocatalytic networks: a constraint on the metabolism-first path to the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1470–1475. doi: 10.1073/pnas.0912628107
- Villarreal, L. P. (2009). The source of self: genetic parasites and the origin of adaptive immunity. *Ann. N. Y. Acad. Sci.* 1178, 194–232. doi: 10.1111/j.1749-6632.2009.05020.x
- Villarreal, L. P. (2011). Viral ancestors of antiviral systems. *Viruses* 3, 1933–1958. doi: 10.3390/v3101933
- Villarreal, L. P., and Witzany, G. (2010). Viruses are essential agents within the roots and stem of the tree of life. *J. Theor. Biol.* 262, 698–710. doi: 10.1016/j.jtbi.2009.10.014
- Vogelstein, B., and Kinzler, K. W. (2002). *The Genetic Basis of Human Cancer*, 2nd Edn. New York, NY: McGraw-Hill, 116.
- Wang, D., Li, L. B., Hou, Z. W., Kang, X. J., Xie, Q. D., Yu, X. J., et al. (2011). The integrated HIV-1 provirus in patient sperm chromosome and its transfer into the early embryo by fertilization. *PLoS One* 6:e28586. doi: 10.1371/journal.pone.0028586
- Wilson, D. S., and Szostak, J. W. (1999). In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* 68, 611–647. doi: 10.1146/annurev.biochem.68.1.611
- Wolf, Y. I., and Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *Bioessays* 35, 829–837. doi: 10.1002/bies.201300037
- Xiong, Y., and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362. doi: 10.1002/j.1460-2075.1990.tb07536.x
- Yang, W. T., Wu, L. W., Tseng, T. C., Chen, C. L., Yang, H. C., Su, T. H., et al. (2016). Hepatitis B surface antigen loss and hepatocellular carcinoma development in patients with dual hepatitis B and C infection. *Medicine* 95:e2995. doi: 10.1097/MD.0000000000002995
- Zhang, B., and Cech, T. R. (1997). Peptide bond formation by in vitro selected ribozymes. *Nature* 390, 96–100. doi: 10.1038/36375
- Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298. doi: 10.1016/S0169-5347(03)00033-8
- Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., et al. (2013). Abundant SAR11 viruses in the ocean. *Nature* 494, 357–360. doi: 10.1038/nature11921

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Moelling and Broecker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.