

Chapter 8

The Contribution of Viruses to Immune Systems



Felix Broecker

Abstract Cellular organisms have evolved a plethora of immune systems to defend against selfish genetic elements and pathogens, including viruses. Virus-derived and virus-related sequences (such as those of mobile genetic elements) constitute a substantial portion of the genomes of all life forms. Some of these sequences mediate resistance to viruses or virus-like parasites and are integral components of many immune systems acting against various invaders (viral and cellular), such as the Rag1/2 system of vertebrates, which generates antibody and T cell receptor diversity. Recently, intimate evolutionary relationships between viral and virus-like sequences and various cellular immune pathways of both pro- and eukaryotes have been uncovered. Here, I argue that the most basic—and likely evolutionarily the first—immune system may be the superinfection exclusion (SIE) mechanism, a phenomenon where one parasitic element (a virus or virus-like entity) prevents or restricts invasion of a compartment (e.g., a cell) by another parasitic element. The SIE mechanism is still a feature of many extant viruses and the related viroids, which are putative relics of an ancient RNA world that existed before the emergence of cells. During cellular evolution, various more complex immune mechanisms fully or partially derived from viruses and virus-like element have evolved. In this chapter, I summarize the current knowledge on the contribution of viral and virus-like elements to the evolution of various cellular immune systems. The emerging picture is that many of today’s cellular immune systems have evolved from simple SIE mechanisms to highly complex defense strategies, frequently involving viral or virus-like sequences.

8.1 Introduction

The idea that immune systems may have evolved from viral sequences is not new, and to my knowledge there are two pioneering scientists who mainly brought forward this idea. First, Eugene V. Koonin, who (amongst many other topics) studies

F. Broecker (✉)
Idorsia Pharmaceuticals Ltd., Allschwil, Switzerland

the evolution of viruses and virus-like elements, reported in 2008 that eukaryotic RNA interference (RNAi), a type of immune system that likely originated as an antiviral defense mechanism, “seems to have been pieced together from ancestral archaeal, bacterial and *phage* proteins” (my emphasis) (Shabalina and Koonin 2008), whereby phages are the viruses infecting prokaryotes. Many more scientific articles by Koonin’s group on the origin of immune systems from viral sequences have followed. Second, evolutionary scientist Luis P. Villarreal wrote, in 2011, a seminal article on ‘Viral Ancestors of Antiviral Systems’ (Villarreal 2011). The concepts outlined in this chapter are mainly based on Koonin’s and Villarreal’s ideas, and I complement these with my own ideas wherever possible. The person who initially stimulated my interest in this question, however, is virologist Karin Moelling, who already in 2006 recognized distinct similarities between the enzymes of the RNAi machinery and those involved in retroviral replication, suggesting a common evolutionary origin (Moelling et al. 2006). Finally, I wish to mention two more names, Emmanuelle Charpentier and Jennifer Doudna, who have been awarded the Nobel Prize in Chemistry “for the development of a method for genome editing” in 2020. This genome editing method is CRISPR-Cas, nowadays widely used for genetic manipulation of all kinds of genomes in laboratories worldwide. CRISPR-Cas, however, is not only a very useful tool in molecular biology, but also an evolutionarily ancient immune system of prokaryotes that acts against genetic parasites such as phages and plasmids. The reason why I mention CRISPR-Cas here is because its evolution involved various virus-like sequences (Koonin and Makarova 2017, explained in more detail in Sect. 8.6.2). Briefly, CRISPR-Cas could only evolve because of the existence of a virus-related genetic parasite, a transposon called *casposon* (Krupovic et al. 2014). CRISPR-Cas is therefore a perfect example of how virus-like sequences have been coopted by a host cell to exert anti-viral functions. But before going into more details, I like to provide a definition of a virus, which brings us to the idea of the *Greater Virus World*.

8.2 The Greater Virus World

What is a virus? Conventionally, it is defined as an infectious agent that is an obligate intracellular parasite. According to this textbook definition, a virus exists in two different states, an extracellular one with surface structures such as capsid or envelope proteins that allows it to move between organisms or cells, and an intracellular one during which new viruses are produced with the help of the cellular protein synthesis machinery. This classical definition has recently been challenged by sequence analyses of viruses and their evolutionary relatives, the mobile genetic elements (MGEs). First, there is evidence that viruses (or their evolutionary ancestors) were already diverse at the time of the last universal cellular ancestor (LUCA), with extensive evolution (likely from RNA to DNA viruses or virus-like entities) even before the existence of LUCA (Krupovic et al. 2020). This means that the ancestors of viruses have not originated as cellular parasites, since they may have

predated cells during evolution. Second, MGEs (which include (retro)transposons, plasmids and viroids) and *bona fide* viruses (cell-infecting genetic parasites with an intracellular and an extracellular state) share a number of hallmark genes, e.g., those involved in replication, which suggests that viruses have evolved from MGEs and vice versa multiple times throughout evolution (Koonin and Dolja 2014). ‘The Greater Virus World’, a term coined by Koonin and his colleague Valerian Dolja, thus includes both *bona fide* viruses and MGEs. In this article, I will adopt this broad definition of a virus and highlight the contribution of both, *bona fide* viruses and MGEs, to the evolution of the various immune systems.

8.3 Viruses as Drivers of Evolution

Parasite-host coevolution is a major aspect of the evolution of all (cellular and pre-cellular) life (Koonin 2016). Of note, the genetic diversity of the virosphere, which is the entirety of the viruses on Earth, is substantially higher than that of cellular life forms, and viral evolution typically occurs at much faster pace, especially in the case of RNA viruses that have limited proofreading mechanisms during replication of their genomes (Paez-Espino et al. 2016; Koonin and Dolja 2013). Consequently, large numbers of novel viruses are currently being detected by high-throughput sequencing (Wolf et al. 2020). Viruses frequently hijack cellular genes and vice versa, i.e., there is constant genetic exchange. Rather than simply being understood as solely disease-causing or detrimental agents, viruses and other parasitic MGEs are increasingly being recognized as entities that can provide benefits to the host, e.g., by protecting from superinfection and through exaptation of genetic material of the viral parasite for host functions (Koonin 2016; Broecker and Moelling 2019a). Moreover, a large portion of adaptations of cellular proteins are driven by the action of viruses; e.g., in an estimated 30% of all conserved mammalian proteins (Enard et al. 2016). All known life forms harbor such genetic parasites, and sequences originating from viruses and MGEs constitute large fractions of cellular genomes, up to 90% in some plant species and up to two thirds of the human genome (Koonin 2016; de Koning et al. 2011). Originally dismissed as ‘junk DNA’, it is now well-established that these sequences are frequently transcribed, provide promoters, enhancers, polyadenylation and splice sites and are thereby substantially involved in host gene regulation (Gogvadze and Buzdin 2009). Moreover, they contribute to the formation of new genes, either directly (e.g., the syncytin genes originating from retroviruses which will be described in Sect. 8.5.1) or indirectly through pseudogene formation mediated by retroelements, whose replication machinery can reverse-transcribe mRNAs of cellular genes and then re-insert the DNA copies into the genome. Some of the viral genes are involved in the various immune systems that have evolved in cellular life forms.

8.4 Immune Systems: An Evolutionary Perspective

The following sections contain a noncomprehensive overview of the involvement of viruses and virus-like elements in the evolution of various immune systems. I will start with an attempt to define the term *immune system*.

8.4.1 What Is an Immune System?

The emergence of identity, i.e., the ability of an entity (e.g., a cell) to discriminate self from non-self, has likely been a crucial step in the evolution from an inanimate predecessor world to the living world in which we exist (Villarreal and Witzany 2013). The immunological self/non-self model, proposed by Frank Macfarlane Burnet in 1949, applies the concept of identity to the immune system, stating that any foreign (*non-self*) element triggers an immune reaction of an organism, whereas any component of the organism itself (*self*) does not (Burnet and Fenner 1949). This is, based on today's knowledge, an oversimplification, since many self-structures are recognized and eliminated by the mammalian immune system (e.g., dead cells are eliminated by phagocytes, and (pre)tumor cells are frequently recognized and eliminated as well) and many non-self-structures are tolerated, e.g., the approximately 4×10^{13} bacterial cells that reside as commensals, with many beneficial functions in digestion and immune function, in the intestinal tract of every human being (Pradeu and Carosella 2006). Viruses are also frequently tolerated by the human immune system, including *bona fide* viruses such as herpesviruses (which in immunocompetent individuals are mostly symptomless), as well as the about 100,000 endogenous retroviruses (ERVs) or fractions thereof in the human genome that could be considered as foreign sequences. The ERVs are typically benign or beneficial, however, they can be abnormally activated in and thus potentially contribute to certain disease states, including cancer, and have consequently been described as 'the enemy within' (Wilkins 2010). These ERVs originated from germline cells infected with *bona fide* retroviruses, mostly genomic introductions that occurred millions of years ago, and these infected germline cells should have been eliminated by the immune system according to the self/non-self model (a virus-infected cell is typically eliminated by the immune system as it presents virus-encoded non-self structures on the surface). However instead, the genetic information of many ERVs has been fixed in the genomes of many species, including humans. The complex immune systems of the various species on earth, pro- and eukaryotic, single- and multicellular, in many cases consequently do not merely distinguish self from non-self, but rather harmless (or even useful) from harmful, which is a much more elaborate process. Moreover, the involvement of microbes in immune systems is increasingly becoming recognized, i.e., the ability of the host to "manage and exploit beneficial microbes to fend off nasty ones" (Travis 2009), which is the subject of this chapter when 'microbes' are specified to 'viruses'.

8.4.2 A Simple Immune System Based on RNA?

It is possible to design RNA molecules that can cleave other RNA molecules in a sequence-specific manner. This type of catalytic RNA is a so-called hammerhead ribozyme that forms Watson-Crick basepairs with the target RNA and then cleaves a specific phosphodiester bond of the target RNA. This idea is being investigated as a potential therapeutic approach against autoimmune diseases and cancer (Citti and Rainaldi 2005). The artificial ribozyme could constitute an immune system as defined in the previous paragraph, if the target RNA is a harmful one, e.g., a parasitic RNA. The ribozyme does not “blindly” discriminate self from non-self. It specifically eliminates those RNAs that have sufficient complementarity and tolerates all other RNAs. It can distinguish harmful from harmless, if the information (harmful vs. harmless) is encoded in the RNA sequence of the parasite.

In nature, hammerhead ribozymes are found in viroids, which are virus-related, protein-free infectious agents consisting of highly structured, circular non-coding RNA. Viroids are possible remnants of the ancient RNA world thought to have existed before the evolution of DNA or proteins (Diener 1989; Flores et al. 2014). In the ancient RNA world, a primordial RNA-based immune system could have been constituted by a viroid that eliminates another viroid via ribozymatic cleavage *in trans* (Table 8.1). Although known natural viroids are generally self-cleaving, they can be modified relatively easily to yield *trans*-cleaving derivatives (Jimenez et al. 2015), suggesting that *trans*-cleaving ribozymes may have existed or may still exist naturally. However, this example is merely a molecule acting against other molecules. If we add a cell-like structure or compartment, perhaps an early primordial cell in the RNA world, such a *trans* cleaving ribozyme might be beneficial to that cell by protecting against other, parasitic RNAs. (As a side note, parasites inevitably occur within evolving life forms (Koonin et al. 2017). No life exists without parasites, and even parasites often have parasites—there will be an example of a virus infecting another virus below in Sect. 8.6.1). Thus, the cell or compartment might benefit from hosting such a parasite-cleaving RNA. The parasite-cleaving RNA might itself be a parasite of the cell (i.e., a viroid-like structure), which however may be tolerated since its presence provides a net benefit to the cell. This hypothetical primordial immune system highlights an important phenomenon that may be at the origin of various immune systems; superinfection exclusion (SIE). This phenomenon is defined as an infection by a virus or viroid that protects against superinfection by the same or a different virus or viroid (Ziebell and Carr 2010). (I designated Sect. 8.5 below for a more detailed discussion of SIE.) If the first virus or viroid is asymptomatic or causing mild symptoms only but protects against superinfection by a more virulent virus or viroid then there is a net benefit for the cell. In plants, SIE has been reported for both *bona fide* viruses as well as viroids. Coming back to the abovementioned RNA-based immunity against RNA infectious agents in the present-day world: The mechanism of action of SIE by extant hammerhead viroids is likely not based on RNA cleavage (since natural hammerhead ribozymes can only cleave *in cis*) but on RNA silencing mechanisms provided by the host cell

Table 8.1 Examples of immune systems whose evolution involved viruses or virus-like elements including: ERVs (endogenous retroviruses); piRNA (Piwi-interacting RNA); SIE (superinfection exclusion); and siRNA (small interfering RNA)

Immune system	Organisms	Function	Involvement of viruses/virus-like sequences
Viroid/ribozyme-based immunity (speculative)	Pre-cellular life or early cells (RNA world?)	A catalytic RNA destroys other (potentially parasitic) RNAs in a sequence-dependent manner	Known ribozymes are self-cleaving but can be modified easily to <i>trans</i> -cleaving ones that may exist or may have existed naturally (Jimenez et al. 2015)
Restriction-modification	Prokaryotes	The host genome is methylated at a specific sequence, a restriction endonuclease cuts unmethylated (invading) DNA of that sequence	Phages and MGEs mediate horizontal gene transfer of restriction-modification systems (Furuta et al. 2010; Murphy et al. 2013)
Prophages	Prokaryotes	SIE mediated by prophages or prophage-derived genes against exogenous phages	Prophages or prophage-derived genes mediate resistance to phage infection by various mechanisms (Bondy-Denomy et al. 2016)
CRISPR-Cas	Prokaryotes	Adaptive immune system of prokaryotes; a piece of invading genetic information is integrated into the host genome and is used to guide nucleases to invaders with homologous sequences	At least four mobile genetic elements contributed to the evolution of CRISPR-Cas systems (Koonin and Makarova 2017). The transposon called casposon is the origin of all CRISPR-Cas systems
Argonaute-based targeting	Prokaryotes	Genomes of invaders are cleaved into small fragments that are used to guide Argonaute nucleases to invaders with sequence complementarity, inducing their degradation	Argonaute proteins and the retroviral reverse transcriptase-RNase H proteins share structural and functional properties, suggesting a common evolutionary origin (Moelling et al. 2006)
piRNA	Eukaryotes	piRNAs are transcribed from piRNA clusters and guide Argonaute proteins to invaders with sequence complementarity that are consequently degraded	Most piRNA sequences originate from transposons and mediate transposon suppression, especially in germline cells (Iwasaki et al. 2015). Some piRNA sequences originate from viruses with putative anti-viral activity, mainly in insects (Ophinni et al. 2019)
RNA interference	Eukaryotes	RNA interference has likely evolved as an antiviral mechanism (Shabalina and Koonin 2008). It acts by processing	The RNA interference and retroviral replication machineries share structural and functional homologies that

(continued)

Table 8.1 (continued)

Immune system	Organisms	Function	Involvement of viruses/virus-like sequences
		invading RNAs into small RNAs (siRNAs) that then guide the RNA-induced silencing complex to invading RNAs with sequence complementarity, which induces the degradation of the invading RNA	suggest a common evolutionary origin (Moelling et al. 2006). The ‘endo-siRNA’ pathway uses siRNAs from sense-antisense pairs of RNAs from transposons to suppress transposon activity in mouse brain and embryonic stem cells (Nandi et al. 2016; Berrens et al. 2017)
Endogenous retroviruses	Eukaryotes	Genes derived from endogenous retroviruses mediate SIE against exogenous retroviruses	Various examples of coopted retroviral <i>env</i> and <i>gag</i> genes with antiviral activity have been described (see Sects. 8.5.1 and 8.5.2)
Interferon system	Eukaryotes	The interferon system is part of the antiviral innate immune system of animals (metazoans)	ERVs have been coopted to provide transcription factor binding sites to various interferon-stimulated genes and thereby are involved in regulating antiviral responses (Chuong et al. 2016; Ito et al. 2017)
Antibodies and T cell receptors	Eukaryotes	In jawed vertebrates, antibodies and T-cell receptors are the basis of adaptive immunity	The Rag1/2 system that generates the diversity of antibodies and T-cell receptors as well as the recognition sites for recombination originate from a <i>Transib</i> transposon (Kapitonov and Koonin 2015; Huang et al. 2016)
Mucosal immunity	Eukaryotes	In animals (metazoans), phages adhering to mucosal surfaces mediate immunity against bacterial infections	The mucus layers of metazoans and bacteriophages may have co-evolved such that phages adhere to mucus via immunoglobulin-like proteins and thus provide a protective barrier against invading bacteria (Barr et al. 2013)

(Kovalskaya and Hammond 2014). However, it is possible to express hammerhead ribozymes targeting a pathogenic viroid to “immunize” plant cells against disease. Thus, the hypothetical primordial immune system described above is not as far-fetched and might exist or have existed in nature.

8.4.3 Innate and Adaptive Immunity

The term “immune system” commonly refers to the eukaryotic, or more specific, the mammalian immune system (prokaryotic immune systems will be described below in Sects. 8.5.7 and 8.6.2). The mammalian immune system can be subdivided into two arms, the innate and the adaptive immune system. Examples for the former include the well-described toll-like receptors (TLRs) which are a group of pattern-recognition receptors (PRRs) that recognize microbial/viral/fungal structures such as lipopolysaccharide, double-stranded RNA, zymosan, etc. (Mahla et al. 2013). From an evolutionary perspective, the TLR protein family is over 700 million years old and found throughout the eumetazoan clade (all multicellular organisms except sponges and placozoa, the simplest known animals), a group which includes diverse animals such as squids, jellyfish, mammals, annelids, as well as insects, to name a few (Leulier and Lemaitre 2008). These types of receptors have been first described in the model organism *Drosophila melanogaster* (fruit fly) in 1985 (Anderson et al. 1985a, b), by the group of Nobel Prize Awardee Christiane Nüsslein-Volhard. (As a side note, the toll receptors of *D. melanogaster* are mainly involved in embryonic development and not in immunity and have thus evolved functional divergence in different species (Kambris et al. 2002)). Other PRRs include RIG-I-like receptors (RLRs), NOD-like receptors (NLRs) and C-type lectin receptors (CLRs) (Mahla et al. 2013). The activation of PRRs leads to complex antimicrobial responses. Like TLRs, RLRs, NLRs and CLRs are found in both invertebrate and vertebrate genomes; homologs of NLRs can also be found in plants, suggesting an even earlier evolutionary origin (Lange et al. 2011; Zou et al. 2009; Sattler et al. 2012; Jones et al. 2016). Another type of innate immune system that has likely evolved as a defense against viruses and MGEs is RNA interference (RNAi), which acts by small interfering RNAs (siRNAs) and is found not only in animals but also in plants and fungi and is therefore evolutionarily more ancient than PRR-based systems (Obbard et al. 2009; Shabalina and Koonin 2008; Ge and Zamore 2013). In animals, a variation of RNAi, using PIWI-interacting RNAs (piRNAs) is used to silence transposons in germline cells to ensure fertility (Ge and Zamore 2013).

Several hundred million years after the emergence of the abovementioned innate immune systems passed until about 450–500 million years ago a new type of immune system evolved, the adaptive immune system. Remarkably, comparable adaptive immune systems evolved in parallel at least twice, in jawless vertebrates (agnathans) as well as in jawed vertebrates (gnathostomes). The prerequisites for such an adaptive immune system are (1) a molecular machinery that allows for the rearrangement of germline-encoded antigen-receptor genes and (2) a dedicated repertoire of cells, each of which expresses a different antigen receptor, e.g., B cells expressing specific immunoglobulins (Igs) (Bayne 2003). The recombination machineries involved in diversifying the antigen-receptors in vertebrates allow for the generation of theoretically over 10^{14} receptors with different specificities, a number which is impossible to be encoded in any genome (e.g., mammalian genomes encode only about 20,000 genes). In jawed vertebrates, these diverse

receptors are Igs expressed on B cells and T cell receptors (TCRs) expressed on T cells. Recognition of a specific structure of a pathogen by an antigen-receptor triggers clonal amplification of that cell, its differentiation and, in the case of B cells, production of antibodies with the same antigen binding specificity (Cooper and Alder 2006). As will be discussed below the mechanism that diversifies Igs and TCRs, V(D)J recombination, relies on an ancient transposon that allows for the key step to occur, genomic recombination. Agnathans have immune cells that share similarities with B and T cells but antigen receptor diversity (which is in the same theoretical order of magnitude as the one achieved by V(D)J recombination) is not generated by recombination, but instead by gene conversion (Boehm et al. 2012).

8.5 Viruses Against Viruses: Superinfection Exclusion, a Simple Immune System

Superinfection exclusion (SIE) is the ability of a first viral infection to restrict secondary viral infections of the same cell. In most cases, SIE prevents infections by the same virus that caused the first infection, or closely related ones, but in some cases (e.g., in the case of the virophages described below in Sect. 8.6.1) also restricts infection by genetically non-related viruses. Importantly, SIE can be regarded as a simple type of immune system. This phenomenon was discovered in the 1920s in tobacco. Tobacco plants infected with a non-virulent variant of the Tobacco mosaic virus (TMV) were shown to be protected against a more virulent TMV isolate (McKinney 1929). This example highlights that the host (here, the tobacco plant) can benefit from SIE under certain conditions if: (1) The first virus infecting the cell exerts little or no fitness cost to the host (it is relatively benign); (2) The first virus establishes a latent infection that is not cleared by the host's immune system (otherwise, the benefit would only be transient) and (3) The infection confers protection against one or several viruses that are more virulent than the first virus. If we now imagine that the tobacco plant evolves such that the genome of the benign TMV strain becomes stably inherited to the following plant generations (e.g., by genomic integration of the whole TMV genome, or of the gene(s) that mediate the resistance to virulent TMV strains), then the result would be an inheritable immune system (to my knowledge, endogenous TMV elements have not been reported, so this remains a thought experiment, but the sequences of many other plant viruses are found in the genomes of various plant species, with possible antiviral functions including the generation of small RNAs used for the RNAi machinery (Chu et al. 2014; da Fonseca et al. 2016)).

Superinfection exclusion is not restricted to plant cells but occurs widely in prokaryotic as well as eukaryotic (single-cell or multicellular organisms) systems (Broecker and Moelling 2019b). Prokaryotic SIE mechanisms will be described below and they have a broad presence. SIE has also been described for several human pathogenic viruses, including vaccinia, measles, hepatitis C, West Nile,

influenza virus and others (Birukov and Meyers 2018). Another example is human immunodeficiency virus (HIV). One of the essential steps in the replication cycle of HIV (as for all retroviruses) is the integration of the (reverse-transcribed) genome into the genome of the host cell (forming a so-called provirus). This is an important feature, as an integrated retroviral genome (or a part thereof) can be inherited by the next host generation (which, however, requires infection of germline cells, see below). It is important to note that HIV mediates SIE, and it has been shown that the virus does so by expressing an accessory protein, Nef, which downregulates the receptor for HIV, CD4, and one coreceptor, CCR5 (Michel et al. 2005). Thereby, an HIV-infected cell is less likely to become superinfected by other HIV particles.

In 2014, an article was published claiming that HIV was ‘*en route* to endogenization’ (Colson et al. 2014). Peripheral blood mononuclear cells (PBMCs) of one described patient (who tested negative for the protective CCR5- Δ 32 genomic mutation) harbored defective HIV-1 proviruses. These PBMCs could not be (super)infected with the identical HIV-1 strain *in vitro*, which suggests that the HIV-1 proviruses conferred resistance to infection. Although most proviruses had premature stop codons, some of them showed intact open reading frames. The presence of apparently protective HIV-1 proviruses suggested that HIV-1 proviruses could potentially mediate SIE. However, whether the virus is being *endogenized* is a more complex question. Endogenization is a two-step process whereby viral genetic information becomes part of a host’s genome. First, the viral genome needs to be integrated into the genome of the host cell (this occurs during the normal retroviral lifecycle but can also occur ‘accidentally’ for other viruses, see below). In multicellular organisms (such as humans) it is necessary that genomic integration occurs in germline cells, as genomic integrations in somatic cells will not be inherited. Second, the viral genome in its entirety, or parts thereof, become fixed in the population. The human genome, for example, contains about 700,000 endogenous retroviruses (complete or fragments thereof), which constitute up to 8% of the genomic sequence, compared to ~20,000 human genes which account for ~2% of the total coding sequence of the genome (Belshaw et al. 2004; Escalera-Zamudio and Greenwood 2016). In the HIV example described above, the patient had proviruses in their PBMCs (i.e., in somatic cells) but not in germline cells, which would be required for vertical transmission and thus, endogenization. Therefore, only a germline infection with HIV-1 may confer inheritable resistance against HIV-1 induced disease at the population level. It is debated if HIV is able to infect germline cells (spermatozoa and oocytes) and whether integrated proviruses can be transmitted vertically (Baccetti et al. 1994, 1998; Bagasra et al. 1994; Barboza et al. 2004; Cardona-Maya et al. 2009, 2011; Nuovo et al. 1994), and in general the more complex retroviruses (genus lentivirus, of which HIV is an example) appear to enter the germline much less frequently than do simple retroviruses (e.g., all of the 100,000 known HERVs in the human genome are derived from simple retroviruses, and not a single one from a lentivirus). Endogenous lentiviruses have been described, however, in the genomes of other species, such as rabbits, non-human primates, weasels, colugos, ferrets and bats (Katzourakis et al. 2007; Gifford et al. 2008; Gilbert et al. 2009; Han and Worobey 2012, 2015; Cui and Holmes 2012; Jebb

et al. 2020). It thus appears to be theoretically possible that HIV-1 may indeed become endogenized at some point in the future. The propensity for endogenization of simple retroviruses, however, appears to be substantially higher, given that the vast majority of known ERVs are derived from simple retroviruses. For these simple ERVs there are several examples for SIE described in the literature, as summarized in the following section.

8.5.1 Endogenous Retrovirus-Mediated Immunity in Eukaryotes: The Envelope Protein

A virus is endogenized when it enters the genome of a germline cell, is transmitted vertically to the next generations of host and becomes fixed in the host population (Fig. 8.1a). In eukaryotes, retroviruses are by far the most frequently endogenized viruses, as during their replication there is an obligatory step during which the retroviral genome is integrated into the host's genome. These ERVs, or parts thereof,

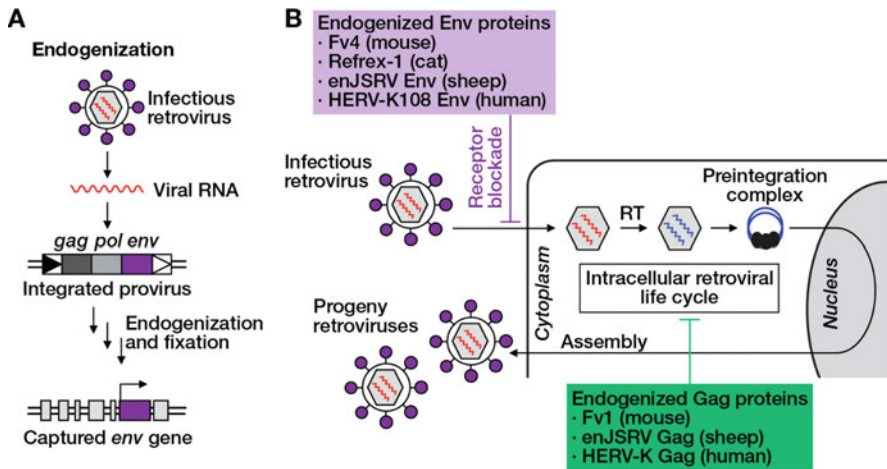


Fig. 8.1 Endogenous retrovirus-mediated antiviral immunity. (a) Schematic of the process of endogenization of a retrovirus. A retrovirus infects a germline cell. The viral RNA is reverse-transcribed into a DNA copy and integrated into the cellular genome, forming a provirus. If this provirus is beneficial or at least not detrimental to the survival of the cell and the organism, it can become endogenized and eventually fixed in the population. Most of the provirus can decay over time due to random mutations, but certain beneficial genes (here, the retroviral *env* gene is used as an example) can be captured and exert novel functions, such as serving as an antiviral defense mechanism. (b) Examples of endogenous retrovirus-derived genes that serve as immune defense against retroviruses. Env proteins can act by blocking receptors (termed a receptor blockade) to prevent infection of a cell. Gag proteins have been shown to inhibit the retroviral replication at various steps during the intracellular life cycle. Details on the indicated examples in mouse, cat, sheep and humans are described in the text. Viral RNA is depicted by red wavy lines, viral DNA by blue wavy lines. RT, reverse transcriptase

can mediate resistance to infections by *bona fide* retroviruses. Of note, the genomes of eukaryotes harbor large amounts of ERV sequences; about 8% of the human genome sequence originates from retroviruses (Gifford and Tristem 2003). Most of the human ERVs (HERVs) have invaded the ancestral genome many millions of years ago, and none of the HERV sequences identified in the human genome have any known infectious counterparts still existing; these viruses are most likely extinct except for their endogenous remains. Intensively studied examples of endogenized viral genes are the syncytins which originate from retroviral *env* genes (Lavialle et al. 2013). Specific *env* genes have been independently exapted from different retroviral proviruses at least seventeen times during evolution. Some of these genes are used as syncytins or otherwise functionally related genes which are critical for placentation in mammals and some viviparous lizard species (Cornelis et al. 2017; Imakawa and Nakagawa 2017). Through their immunosuppressive domain (ISD), syncytins likely contribute to the prevention of maternal immune rejection of the fetus via various mechanisms, including the inhibition of leukocytes and the suppression of pro-inflammatory cytokines (Cianciolo et al. 1985; Haraguchi et al. 1995, 1997, 2008).

In addition to syncytins, other retroviral *env* genes have been endogenized that do not exert immunosuppressive, but instead anti-retroviral functions (Fig. 8.1b). For example, the mouse *Friend virus susceptibility 4 (Fv4)* gene confers resistance of mice to murine leukemia viruses (MuLVs) (Suzuki 1975). Physically, *FV4* is a truncated MuLV-like provirus containing the 3' portion of the *pol* gene and the entire *env* gene (Ikeda et al. 1985). The *env*-encoded protein binds to the cellular receptor used by MuLV and thereby prevents infection by the exogenous retrovirus, a process referred to as receptor blockade, a variant of SIE. Another captured *env* gene in mice, *resistance to MCF (Rmcf)*, mediates resistance to mink cell focus-inducing (MCF) viruses and MuLVs, likely also via receptor blockade (Hartley et al. 1983; Brightman et al. 1991; Jung et al. 2002).

In cats, the *refrex-1* gene confers resistance to feline leukemia virus-D (Ito et al. 2013). It is a truncated retroviral *env* gene that contains the putative receptor-binding domain but lacks a C-terminal portion due to a premature stop codon.

Env-mediated interference has also been demonstrated in human cells. The Env protein encoded by a human endogenous retrovirus of the HERV-K(HML-2) family has been shown to interfere with HIV-1 production in vitro (Terry et al. 2017). It is a full-length Env protein that, compared to the consensus, ancestral HERV-K(HML-2) Env, has four mutations that appear to be required for inhibiting HIV-1. Interestingly, HERV-K(HML-2) expression in T cells is activated by HIV-1 infection (Gonzalez-Hernandez et al. 2012), which suggests that expression of Env (and Gag, see Sect. 8.5.2 below) may have evolved as an inducible mechanism of protection against exogenous retroviruses. Another example of an antiviral Env protein in human cells is encoded by a HERV-T provirus (Blanco-Melo et al. 2017). Its expression confers resistance to a reconstructed infectious HERV-T virus (as the virus is extinct) via receptor blockade in vitro. In addition, *Suppressyn*, a truncated *env* gene expressed by a HERV-F element with a known role in placental

development, may restrict infection by exogenous retroviruses (Malfavon-Borja and Feschotte 2015).

8.5.2 Endogenous Retrovirus-Mediated Immunity in Eukaryotes: The Gag Protein

Gag is another retroviral gene that has been captured by mammalian hosts for immune defense against exogenous retroviruses (Fig. 8.1b). The best studied *gag*-derived restriction factor is *Friend virus susceptibility 1 (Fv1)* of mice (Best et al. 1996), which inhibits murine leukemia virus (MuLV) at a stage between entry and proviral integration. The Fv1 protein interacts with the retroviral capsid protein in the preintegration complex of MuLV (Best et al. 1996). *FV1* originates from a MERV-L *gag* gene. In sheep, enJSRV-expressed Gag protein inhibits exogenous JSRV at a late stage of the retroviral life cycle during viral assembly (Palmarini et al. 2004). In human cells, a HERV-K(HML-2) Gag protein inhibits HIV-1 release and reduces infectivity of progeny HIV-1 virions (Monde et al. 2017).

8.5.3 Evolution of Retrovirus-Mediated Immunity in Real Time?

Jaagsiekte sheep retrovirus (JSRV) provides an example of a recent or ongoing endogenization (Armezzani et al. 2014). The youngest identified endogenous elements (enJSRV) were integrated into the sheep genome only 200 years ago, and exogenous, infectious JSRV is still circulating. The sheep genome contains about 27 enJSRV sequences, of which 16 harbor intact *env* genes. enJSRV Env protein likely exerts syncytin-like functions during placentation (Dunlap et al. 2006) and has been shown to prevent infection by exogenous JSRV via receptor blockade (Spencer et al. 2003). Another example of an endogenization in real-time is currently occurring in koalas. Like JSRV, koala retrovirus (KoRV) co-exists in both endogenous and exogenous form (Tarlinton et al. 2006). While endogenous KoRV elements can be identified in the genomes of most koalas, there is substantial inter-individual variation in the integration sites and extensive regional variation, indicative of an ongoing endogenization process. It has been speculated that Env (or other proteins) expressed by endogenous KoRV elements may provide protection against exogenous KoRV infections, analogous to the examples described above (Sarker et al. 2020). In favor of this hypothesis, full-length *env* mRNA appears to be expressed by many endogenous KoRV elements (Tarlinton et al. 2017). However conversely, it has been suggested that koalas show in utero expression of KoRV antigens and those antigens get tolerized as the developing immune system recognizes them as harmless self structures. Consequently, their immune systems may be unable to mount an

immune response against exogenous KoRV. In favor of the latter hypothesis, koalas with mostly intact integrated KoRV proviruses are often unable to mount antibody responses against KoRV antigens, even after vaccination, in contrast to animals with less intact KoRV proviruses that can generate antibodies (Tarlinton et al. 2017; Olagoke et al. 2019). Thus, at the current stage integrated and vertically transmitted KoRV elements may either have a beneficial or detrimental effect on the population, and it will take more time for evolutionary selection to resolve this, perhaps by having protection-providing KoRV genes become fixed in the population, as seen for JSRV. It is estimated that KoRV first entered the koala population only between 100 and 200 years ago (Greenwood et al. 2018), whereas the initial infection of the sheep genome with JSRV likely occurred 5–7 million years ago (Armezzani et al. 2014).

8.5.4 Superinfection Exclusion by Other Endogenized Eukaryotic Viruses

Although ERVs constitute the vast majority of endogenous viruses, mammalian genomes also harbor numerous sequences derived from, e.g., *Borna-*, *Filo-*, *Parvo-*, *Circo-*, *Rhabdo-* and *Herpesviridae* (Belyi et al. 2010a, b; Horie et al. 2010; Katzourakis and Gifford 2010; Aswad and Katzourakis 2014). Many integrated (non-reverse transcribing) RNA virus-derived sequences likely originate from reverse transcription and integration via either the replication machinery of retroelements or by nonhomologous recombination (Suzuki et al. 2014). These non-retroviral endogenous viral sequences are frequently referred to as endogenous viral elements (EVEs).

Among these EVEs are the negative sense ssRNA Borna disease virus (BDV) sequences (Belyi et al. 2010b). Interestingly, species that contain genomic BDV sequences (e.g., primates, rats, mice and squirrels) are relatively resistant to infection with exogenous BDVs. In contrast, highly susceptible species like horses, sheep and cattle can develop fatal encephalitis upon BDV infection and do not have detectable BDV sequences in their genomes.

There is also experimental evidence for protection by endogenous BDV sequences. The ground squirrel genome harbors an endogenous bornavirus-like nucleoprotein (itEBLN) sequence with 77% amino acid similarity to circulating infectious BDV (Fujino et al. 2014). The itEBLN RNA binds to the ribonucleoprotein of infectious BDV and is incorporated into virions. This appears to inhibit viral trafficking and cell-to-cell spread (Kim et al. 2020).

Like squirrels, humans usually do not develop Borna disease. Only few anecdotal cases of fatal BDV-induced encephalitis have been reported (Hoffmann et al. 2015). Seven human endogenous bornavirus-like nucleoprotein (hsEBLN) elements are expressed at the RNA level (Sofuku et al. 2015); hsEBLN-2 is also known to be expressed as protein (Ewing et al. 2007). In primate and rodent genomes, EBLNs are

enriched within piRNA clusters (Parrish et al. 2015). These EBLNs express functional piRNAs that are antisense to the BDV nucleoprotein mRNA and are expressed in testes. However, whether they mediate inhibition of BDV infection in these or other cells is unknown. Since piRNAs are also known to be expressed in some somatic cell types such as neurons (Lee et al. 2011), EBLN sequences may protect from brain BDV infection. This could at least partially explain why species with endogenous EBLN sequences are relatively resistant to BDV-induced encephalitis. In addition, EBLNs may protect from Borna disease by inducing immune tolerance through in utero EBLN protein expression (Horie 2017). This tolerization of the immune system may limit the possible pathogenicity associated with anti-nucleoprotein immune responses that develop during BDV infection. The BDV nucleoprotein is a major target for cytotoxic T cell responses (Stitz et al. 1993; Planz and Stitz 1999). Thus, tolerance to nucleoprotein may protect against BDV-induced encephalitis which mostly results from immune-mediated inflammation. In addition, EBLN RNAs may act as antisense RNAs to the BDV genome (Horie 2017). It is important to note that EBLN sequences also are found in the genomes of various other species, including whales, birds and lamprey (Kobayashi et al. 2016; Hyndman et al. 2018), where they might also exert antiviral functions.

Aedes mosquitoes are important vectors for human pathogenic flaviviruses such as Dengue and Zika virus. Their genomes contain various endogenous flaviviral sequences (Suzuki et al. 2017). Small RNAs like piRNAs and siRNAs, known to play an important role in antiviral defense in insects (Cullen et al. 2013), are produced from these endogenous viruses and might play a role in antiviral defense (see Sect. 8.5.5).

8.5.5 piRNA-Guided CRISPR-Cas-Like Immunity in Eukaryotes Based on Endogenous Viral Sequences

CRISPR-Cas immunity of prokaryotes will be described in more detail below, but I like to mention it here as there are some interesting similarities between this immune system and piRNA-mediated immunity of eukaryotes. Briefly, CRISPR-Cas is an inheritable immune system that requires three steps. First, fragments of foreign (e.g., viral) DNA or reverse-transcribed RNA are captured and integrated as spacers into specialized genomic regions called CRISPR loci or CRISPR array. Second, the spacers are transcribed and processed into small RNAs (crispr RNAs/crRNAs). Third, the crRNAs guide a nuclease (Cas) to complementary DNA/RNA upon re-exposure of the invader. Cas then inactivates the invading DNA/RNA. It has recently been suggested that the eukaryotic piRNA system may exert analogous functions in some species, especially in insects (Ophinni et al. 2019). Here, foreign viral RNA is reverse-transcribed and preferentially inserted into piRNA clusters (analogous to the CRISPR loci), then transcribed and processed into small RNAs (piRNAs, analogous to crRNAs). These virus-derived piRNAs then guide a nuclease

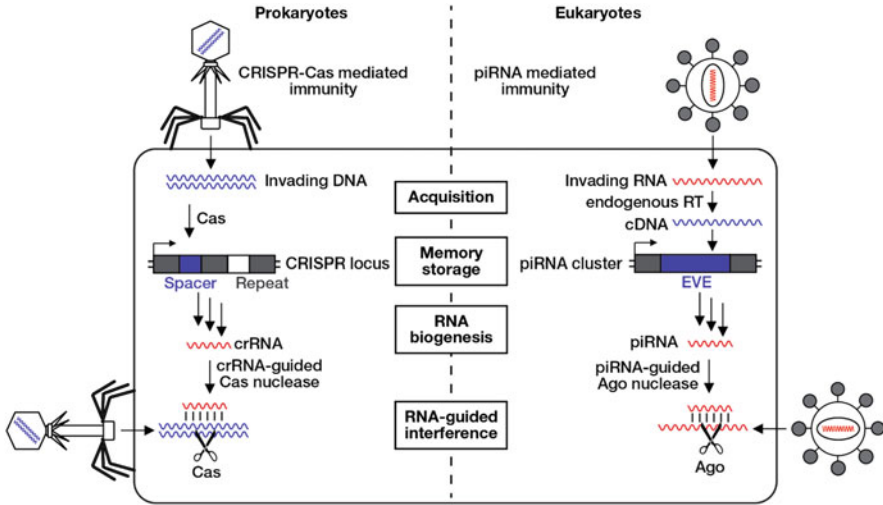


Fig. 8.2 Similarities between CRISPR-Cas immunity in prokaryotes (left) and piRNA-mediated immunity in eukaryotes (right). *Left*: A portion of a phage (top) DNA (or reverse-transcribed RNA in the case of RNA phages) invading the prokaryotic cell is inserted via a Cas protein into a CRISPR locus in the prokaryotic genome. The newly inserted spacer is transcribed along with the CRISPR locus and the transcript is processed into a small RNA (crRNA) that guides another Cas nuclease to an invading DNA (or reverse-transcribed RNA) of another invading phage (bottom) based on sequence complementarity. The Cas nuclease cleaves the invading nucleic acid (indicated by scissors), thus providing immunity against the newly infecting phage. *Right*: In eukaryotes, a similar mechanism is constituted by the piRNA machinery. An invading RNA (of an RNA virus, top) is reverse-transcribed by an endogenous reverse transcriptase and then integrated into a piRNA cluster. The piRNA cluster DNA is transcribed and processed into small RNAs (piRNAs) that guide an Argonaute nuclease to the RNA of a newly infecting RNA virus (right) based on sequence complementarity. The Argonaute protein cleaves the newly invading RNA, thus mediating immunity to the RNA virus. DNA molecules are represented by blue wavy lines, RNA molecules by red wavy lines. Ago, Argonaute; EVE, endogenous viral element; RT, reverse transcriptase. Figure modified from Ophinni et al. (2019)

(Argonaute, analogous to Cas) to the viral RNA, which inactivates the invader (Fig. 8.2).

Such virus-derived piRNAs, specific for *Drosophila X* virus and other RNA viruses, have been first identified in a *D. melanogaster* cell line (Wu et al. 2010). However, to date silencing activity in cells of *D. melanogaster* has only been observed against transposons and endogenous retroviruses (the canonical function of piRNAs), but not against exogenous viruses. In contrast, the link between viral piRNAs and activity against exogenous viruses is stronger in *Aedes* mosquitoes, which show expression of piRNAs derived, for example, from *Toga*-, *Flavi*-, *Bunya*- and *Reoviridae* (all are RNA virus families) in germline cells and somatic tissues. piRNAs are known to be able to guide Argonaute nucleases to complementary RNAs, which induces their degradation. Knockdown of proteins involved in amplification of piRNAs reduced viral piRNA expression and enhanced Dengue virus

replication in *Aedes*-derived cells (Miesen et al. 2016). Viral piRNAs have been discovered in other arthropods as well, including whiteflies, and some of these piRNA sequences have been shown to also target DNA viruses (Ophinni et al. 2019). In mammals, the only identified example of endogenous virus-derived piRNAs are those originating from EBLNs, as described in the previous section. It thus appears that the antiviral activity of piRNAs is likely more pronounced in arthropods than in mammals.

8.5.6 Do Endogenous Viruses Render Bats Resistant to Viral Infections?

Bats have been associated with a number of zoonotic viral diseases and constitute an important reservoir for diverse viruses, including members of the families *Flavi*-, *Rhabdo*- and *Bunyaviridae* (Olival et al. 2017). The bat immune system is unique in its ability to tolerate viral infections which are typically lethal to other mammalian species (Banerjee et al. 2020). It has been suggested that endogenous viruses may play a role in the bats' immune tolerance to viruses (Skirmuntt et al. 2020). Bat genomes contain numerous ERVs and sequences originating from *Borna*-, *Filo*-, *Parvoviridae* and others. Interestingly, to date only one exogenous bat-specific retrovirus has been identified in Australian bats, the Hervey pteropid gammaretrovirus (HPG) that is related to KoRV, suggesting that bats may have transmitted this virus to koalas (Hayward et al. 2020). It could be speculated that the apparently very few exogenous bat retroviruses, which stand in contrast to the large diversity of ERVs, are the result of ERV-mediated immunity (Skirmuntt et al. 2020). Interestingly, no HPG-related ERVs have been identified in bat genomes, suggesting that HPG only recently entered bat populations and has either not (yet) been endogenized or is currently in the process of endogenization (Hayward et al. 2020). It remains speculative whether the bat endogenous viruses mediate immunity against their exogenous counterparts. In favor of this hypothesis, the diversity of endogenous viral sequences is higher than in most other mammals (Jebb et al. 2020) and bats are highly immune to many viral infections, including the *Filoviridae* Ebola- and Marburg virus (the bat genome contains endogenous Filovirus elements with intact open reading frames for nucleoprotein and VP35 protein (Skirmuntt et al. 2020)).

8.5.7 Superinfection Exclusion by Endogenized Viruses in Prokaryotes

Superinfection exclusion is not limited to eukaryotes but is also widespread in prokaryotes. Prokaryotic genomes contain up to 10–20% of prophage-derived

sequences, and the presence or absence of specific prophage sequences can contribute to prokaryotic interstrain variability (Canchaya et al. 2003). There is often a fitness cost to the host associated with prophage integration (Iranzo et al. 2017). However, prophage sequences can also have beneficial effects, such as increasing virulence, which could extend the ecological range of the bacterium (Canchaya et al. 2004).

One example of prokaryotic SIE is the prophage-encoded Tip protein that suppresses expression of type IV pili on the surface of its host, *Pseudomonas aeruginosa*, with little or no fitness cost to the bacterium (Chung et al. 2014). The type IV pili are common entry receptors for phages. Consequently, Tip expression mediates resistance to various phages (Bondy-Denomy et al. 2016). Indeed, prophage-encoded SIE seems to be relatively broad, as only three prophages make *P. aeruginosa* resistant to at least 30 different phages. Another well-described example for prophage-mediated immunity in bacteria is the *sie₂₀₀₉* gene expressed by the lactococcal Tuc2009 prophage (McGrath et al. 2002). The Sie₂₀₀₉ protein localizes to the bacterial membrane and likely inhibits phage DNA injection. Various other phage-encoded genes that mediate resistance to phage infection have been identified in numerous bacterial hosts (Bondy-Denomy et al. 2016; McGrath et al. 2002); it thus appears that prophage-mediated SIE is a common mechanism of immunity in prokaryotes.

8.6 The Enemy of My Enemy Is My Friend: Harnessing Viruses for Complex Immune Systems

Any organism—unicellular or multicellular—is constantly exposed to a plethora of microorganisms, most benign, some potentially harmful, some opportunistic. Cellular organisms can internalize specific viruses and use them as a weapon against other viruses (or cellular pathogens). Genomic integration allows for the process of endogenization (sometimes referred to as domestication), whereby a given antiviral property of a protein (or any other useful function) may be preserved or even enhanced. Other portions of the integrated viruses may degenerate through accumulation of deleterious mutations over time or by mechanisms of genomic deletion in the absence of positive selection pressure. In the following sections I will highlight three examples of complex immune systems whose evolution involved viruses or virus-like elements. I will start by a relatively simple (yet more complex than SIE described above) example of a virus-based inheritable immune system in the protist *Cafeteria roenbergensis* (a unicellular eukaryote), followed by the more complex adaptive immune system of prokaryotes (CRISPR-Cas), and, lastly, the V(D)J recombination system that diversifies antibodies and T cell receptors in vertebrates, arguably the most complex immune system we know.

8.6.1 *A Small Virus Against a Giant Virus in the Protist Cafeteria roenbergensis: An Adaptive Immune System at the Population Level*

The marine protist *Cafeteria roenbergensis*, a single-celled eukaryote, is infected by a giant virus named *Cafeteria roenbergensis* virus (CroV). The CroV virus is a member of the nucleocytoplasmic large DNA viruses (NCLDV) group with a ~730,000 bp dsDNA genome. After entering the protist host via phagocytosis, CroV replicates in cytoplasmic viral factories, which are nucleus-like structures in which DNA replication and transcription occur. Viral protein synthesis and virion assembly takes place on the outside of the viral factories (Bell 2020). The infection is fatal to *C. roenbergensis* and, consequently, it has been suggested that CroV may play a role in regulating the protist's population in marine ecosystems (Fischer et al. 2010).

Cafeteria roenbergensis is also host to another, much smaller dsDNA virus termed Mavirus with a ~19,000 bp genome, which can only replicate in the presence of CroV (Fischer and Hackl 2016). Mavirus infection has no known negative consequence for *C. roenbergensis*. As its replication depends on another virus (CroV), Mavirus is designated as a 'virophage' analogous to the bacteria-infecting bacteriophages. The currently known virophages (family *Lavidaviridae*) all have in common their dependence on an NCLDV for replication (Mougari et al. 2019; Fischer 2020). Simultaneous co-infection with Mavirus protects *C. roenbergensis* from CroV-induced lysis by inhibiting the replication of CroV through a yet unknown mechanism. Interestingly, Mavirus can integrate into the protist's genome (forming 'provirophages'), likely due to the presence of a retroviral-like integrase that is packaged into the virion, as well as nuclear localization signals (Born et al. 2018). The provirophages remain transcriptionally silent until they are induced by a CroV superinfection. Notably, activation of the integrated Mavirus genomes does not prevent the lysis of the initial *C. roenbergensis* cell, likely because Mavirus activation occurs only at a late stage of CroV replication. It does, however, induce the massive production of Mavirus particles by the CroV viral factories. Consequently, the lysed cell releases Mavirus particles (along with CroV particles) into the environment. Neighboring *C. roenbergensis* cells are subsequently co-infected with CroV and Mavirus, and Mavirus inhibits CroV replication and cell lysis. This protects the protist's population, at the expense of losing the cell originally infected with CroV by lysis. It is an inheritable immune system with a twist, in which not the direct descendants of the provirophage-carrying cell are protected (as is the case, for example, for CRISPR-Cas immunity described below in Sect. 8.6.2). Instead, immunity is inherited at the population level (via the release of Mavirus particles and 'vaccination' of neighboring cells), and protection from CroV occurs in an altruistic manner. There is also a potential benefit to Mavirus; infection of *C. roenbergensis* and genomic integration may increase the frequency of interaction with the CroV host (Mougari et al. 2019).

It is still not known how wide-spread this type of virophage-mediated immunity is. Recently, it was shown that co-infection with specific virophages rescued *Acanthamoeba castellanii* (an amoebal species) from being lysed by amoeba infecting NCLDVs and reduced the production of NCLDV particles (Mougari et al. 2019, 2020). However, there is currently no evidence for protection mediated by integrated proviropages as is the case for *C. roenbergensis*. In contrast, integrated virophages are found in other organisms such as the unicellular green alga *Bigelowiella natans* (Koonin and Krupovic 2016). There, most of the 38 identified elements were transcriptionally active, and six represented complete proviropages (Blanc et al. 2015). Moreover, polintons, virophage-related transposons, are found in the genomes of various eukaryotes and likely originate from exogenous viruses (Koonin and Krupovic 2018). The proviropages found in the *B. natans* genome and the virophage-like polinton sequences may have been recruited as a defense against yet to be identified (or extinct) NCLDVs. Moreover, a number of novel virophages and their associated NCLDV hosts have been identified recently and become available for experimental testing (Fischer 2020; Xu et al. 2020; Gulino et al. 2020). The identified inheritable immune system of *C. roenbergensis* may therefore just be the tip of an iceberg (Koonin and Krupovic 2016).

8.6.2 CRISPR-Cas Immunity Largely Originates from Viruses and Mobile Genetic Elements

The CRISPR-Cas system is a form of prokaryotic adaptive immunity which utilizes a collection of DNA fragments (or reverse-transcribed RNA) of infecting phages or plasmids that are integrated as spacers into a designated region of the host genome, the CRISPR locus or CRISPR array (Hille et al. 2018). These spacers provide an immunological memory that protects the cell from invaders with similar sequences. Moreover, since the spacer is stably integrated into the CRISPR locus, immunity is passed on to future generations. As such, the spacers bear witness to the types of phages that are or have been infecting a certain prokaryotic species or strain (similar to ERVs that are indicators of past or ongoing retroviral infections in eukaryotes). CRISPR loci are found in about 50% of bacterial and 90% of archaeal genomes, with an average of three and five loci per genome, respectively (Grissa et al. 2007; Hille et al. 2018). The transcribed CRISPR locus RNA (termed pre-crRNA) is processed into smaller crRNAs that guide sequence-specific cleavage of complementary invading nucleic acids by Cas effector nucleases, which resembles the action of viral and transposon-derived piRNAs transcribed from piRNA clusters in eukaryotes (described above in Sect. 8.5.5).

Six types of CRISPR-Cas systems (I through VI) and various subtypes have been described that differ, for example, in the type of nuclease that mediates target cleavage, and whether DNA or RNA invaders are targeted (Koonin and Makarova 2017). The general steps, however, are identical in all systems; (1) adaptation (i.e.,

spacer acquisition), (2) expression (i.e., crRNA generation) and (3) interference (i.e., target nucleic acid cleavage). Interestingly, a minimum of four different MGEs were involved in CRISPR-Cas evolution (Koonin and Makarova 2017). First (and most importantly), all six systems originate from a transposon called casposon which utilizes Cas1 nuclease for DNA integration (Krupovic et al. 2014). Thus, the crucial invention for prokaryotic adaptive immunity to evolve was the ability to modify the sequence of the host's genome (interestingly, the same ability is at the origin of adaptive immunity of jawed vertebrates, as described in Sect. 8.6.3). Second, Cas2 nuclease and RNase domains of the HEPN family found in several Cas proteins likely originate from toxin-antitoxin modules, which can be regarded as MGEs as they are typically mobilized by plasmids (Koonin and Krupovic 2015; Koonin and Makarova 2017). Third, various Type III systems exapted a reverse transcriptase from an MGE (a mobile group II intron), which allows for spacer acquisition from RNA invaders (e.g., RNA phages). Fourth, the RuvC domains of Type II and V effector Cas nucleases (Cas9 and Cas12, respectively) likely originated from DNA transposons. Functional CRISPR-Cas systems can also be encoded by phages (Seed et al. 2013), suggesting that phages may mediate horizontal gene transfer of this type of immune system. CRISPR-Cas acquired immunity can in theory be transmitted across thousands of microbial generations (Weinberger et al. 2012), although phage evasion by mutation typically occurs within a few generations in coevolution studies (Westra et al. 2019). There is no known mechanism that can discriminate harmful from beneficial invaders (the former could be a lytic bacteriophage, the latter a plasmid conferring antibiotic resistance; both will be equally targeted by CRISPR-Cas). Thus, a CRISPR-Cas locus that provides an evolutionary disadvantage (as it prevents, for example, resistance to antibiotics) might be counter-selected, which might explain the existence of prokaryotes without any CRISPR loci (Westra et al. 2019).

8.6.3 At the Heart of Adaptive Immunity in Jawed Vertebrates Is an Ancient Mobile Genetic Element

In contrast to the prokaryotic adaptive immune system, CRISPR-Cas, immunological memory in multicellular vertebrates is restricted to somatic cells, more specifically, to dedicated cells of the immune system (B and T cells). This memory is therefore not inherited to the next generation. In jawed vertebrates, the diversity of Igs/antibodies and TCRs is generated by a process called V(D)J recombination, in which variable (V), diversity (D) and joining (J) gene segments are recombined. Further antibody diversification is then achieved by somatic hypermutation, whereby random mutations are introduced, and the B cells subjected to a selection process for increased binding affinity to the antigen (Kapitonov and Koonin 2015). In contrast to CRISPR-Cas, elaborate mechanisms have evolved that can differentiate between harmful and harmless, which, for example, renders the adaptive

immune system able to mount a response against pathogenic bacteria, but at the same time it tolerates the normal, healthy microbiota of the gastrointestinal tract, the lung, the skin, and other organs. The tolerance to the healthy microbiota is the result of a co-evolution of the hosts and the microbes. Ancestors unable to tolerate the microbes did not survive, similarly those microbes that were not able to evade the immune responses also died. In addition, even though there is no inheritance of the adaptive immunity of jawed vertebrates, there is an evolutionary selection for those individuals that can mount a protective immune response against deadly pathogens.

The ability to produce diversity of antibodies and TCRs in jawed vertebrates developed around 450–500 million years ago (Kapitonov and Koonin 2015). As for CRISPR-Cas, the ability to modify the genome sequence has been the crucial event at the evolutionary origin of adaptive immunity. Both the Rag1 and Rag2 proteins that mediate V(D)J rearrangement by recombining V, D and J segments are encoded by a single genomic locus. They originate from a DNA transposon called *Transib* that today is found in the starfish, oyster and sea urchin genomes, but not anymore in those of jawed vertebrates, where it went extinct (Kapitonov and Koonin 2015). An active *Transib* transposon encoding Rag1 and Rag2-like proteins was recently discovered in the lancelet genome, and its terminal inverted repeats (the sequences flanking the transposon) share similarities with the recombination signal sequences that are recognized by the Rag1/2 complex (Huang et al. 2016). Thus, not only the genes encoding the proteins required for V(D)J recombination but also the recognition sequences for the recombination to occur originate from a transposon.

8.7 Discussion

Viruses have traditionally been regarded mainly as disease-causing agents (hence the name *virus*, Latin for poison). Yet, viruses and their relatives, the MGEs, have also majorly contributed to the evolution of cellular organisms by introducing, mobilizing and amplifying genetic material. The recruitment of sequences from viruses, transposons and other MGEs for pro- and eukaryotic immune systems appears to be strikingly common, as illustrated by the examples presented above. Additional immune systems that have evolved with the involvement of viral or virus-like sequences are discussed briefly below (see also Table 8.1).

One is the prokaryotic restriction-modification (RM) mechanism. The RM systems consist of both a restriction endonuclease that cleaves invading DNA (e.g., of a phage) at a specific sequence motif as well as a methylase that masks that motif in the prokaryotic genome via DNA methylation. The motifs are typically short and are thereby present in many invading DNAs. Thus, RM systems can be regarded as a prokaryotic innate immune system. These RM systems are present in ca. 90% of prokaryotic genomes (Murphy et al. 2013), can be mobilized by phages and frequently co-localize with transposon-derived genes and may be flanked by inverted repeats and target site duplications, with those flanking structures seemingly characteristic of transposons (Naderer et al. 2002; Furuta et al. 2010; Makarova et al.

2011; Takahashi et al. 2011). In addition, transposons can carry functional RM systems (Khan et al. 2010), indicating that these defense systems may have evolutionarily originated from transposons.

Another prokaryotic innate immune system involves Argonaute proteins that act as RNA- or DNA-guided nucleases to cleave invading RNA or DNA (Swarts et al. 2014; Koonin and Krupovic 2015). Like the Argonaute proteins involved in eukaryotic small RNA-guided defense mechanisms, prokaryotic Argonaute proteins share striking structural and functional similarities with the retroviral reverse transcriptase-RNase H proteins (Moelling et al. 2006), suggesting a common evolutionary ancestry.

In both pro- and eukaryotes, antisense transcripts from viral and MGE sequences may act by forming double-stranded RNA complexes with invading RNAs when there is sufficient sequence complementarity, which may induce degradation of the invading RNA (Broecker and Moelling 2019b).

In eukaryotes, interferons are mediators of innate immunity, especially during viral infections. It has been shown that transposons and ERVs have been specifically co-opted by the host to provide enhancers and binding sites for transcription factors for interferon-stimulated genes (Chuong et al. 2016; Ito et al. 2017). Transposons and ERVs have thus significantly shaped the regulation of the interferon response

As the last example, I would like to mention an interesting interaction that happens at the mucosal surfaces of animals (metazoans). Some phages express immunoglobulin-like domains on their surface that bind to mucin glycoproteins expressed in mucosal tissues such as the gastrointestinal tract or lungs (Barr et al. 2013). Thereby, phages are specifically enriched in mucus and protect the underlying epithelium from invading bacteria. The phages also benefit by an increased frequency of interaction with their target bacteria in what can be seen as a symbiotic relationship. Thus, phages enriched in mucosae serve as a non-host derived immune system against bacterial infections.

The fact that genomes of virtually all cellular organisms harbor large numbers of MGEs and viral sequences suggests that yet unknown functionalities will likely be identified in the future. For example, it was recently discovered that prokaryotic defense islands, genomic regions involved in various immune mechanism, are enriched with transposon sequences whose potential functions remain to be determined (Doron et al. 2018; Koonin 2018).

Of note, immune defense is not the only function of endogenized viruses and MGEs. For example, deleting all prophages in *Escherichia coli* results in various fitness deficits, including increased susceptibility to antibiotics and osmotic stress, and causes deficits in growth and biofilm formation (Wang et al. 2010). In eukaryotes, transposons and ERVs do not only modulate the interferon response, but also play roles, amongst others, in cell differentiation, stem cell pluripotency and embryogenesis (Chuong et al. 2016). Thus, viral and virus-like sequences have adopted multifaceted roles, not only for immune defense, and have been a major driving force in the evolution of cellular life.

I would like to end with a comment on the current SARS-Coronavirus-2 pandemic. As viral evolutionist Aris Katzourakis mentioned via Twitter with respect to

the analysis of bat genomes for virus-derived sequences (Skirmuntt et al. 2020), there is “no endogenous coronavirus so far” (Tweet by @ArisKatzourakis, May 22, 2020)—even in bats. To my knowledge endogenous coronaviruses have not been reported in any species to date, although there is *in vitro* evidence that coronavirus RNA can be reverse transcribed and integrated into the genome of human cells (Zhang et al. 2020). However, there seems to be an interplay between SARS-Coronavirus-2 and HERVs in humans. It was shown that various HERV families are upregulated in the lungs of SARS-Coronavirus-2 infected people (Kitsou et al. 2020). However, it remains to be determined whether the upregulated HERVs confer protection, contribute to pathophysiology or if this is simply a by-stander effect.

Acknowledgements I like to express my gratitude to Prof. Karin Moelling (Max Planck Institute for molecular Genetics Berlin) for ongoing support and for stimulating and fostering my interest in the topic of the viral origin of immune systems, among many others. I also like to acknowledge Dr. Günther Witzany who kindly invited me to his exciting symposia *DNA Habitats and its RNA Inhabitants* (2014) and *Evolution—Genetic Novelty/Genomic Variations by RNA Networks and Viruses* (2018) both held in Salzburg, Austria, where many of the ideas described in this chapter have been developed.

References

- Anderson KV, Bokla C, Nüsslein-Volhard C (1985a) Establishment of dorsal-ventral polarity in the *Drosophila* embryo: the induction of polarity by the Toll gene product. *Cell* 42:791–798
- Anderson KV, Jürgens G, Nüsslein-Volhard C (1985b) Establishment of dorsal-ventral polarity in the *Drosophila* embryo: genetic studies on the role of the Toll gene product. *Cell* 42:779–789
- Armezzani A, Varela M, Spencer TE et al (2014) “Ménage à Trois”: the evolutionary interplay between JSRV, enJSRVs and domestic sheep. *Viruses* 6:4926–4945
- Aswad A, Katzourakis A (2014) The first endogenous herpesvirus, identified in the tarsier genome, and novel sequences from primate rhadinoviruses and lymphocryptoviruses. *PLoS Genet* 10: e1004332
- Baccetti B, Benedetto A, Burrini A et al (1994) HIV-particles in spermatozoa of patients with AIDS and their transfer into the oocyte. *J Cell Biol* 127:903–914
- Baccetti B, Benedetto A, Collodel G et al (1998) The debate on the presence of HIV-1 in human gametes. *J Reprod Immunol* 41:41–67
- Bagasra O, Farzadegan H, Sehshamma T et al (1994) Detection of HIV-1 proviral DNA in sperm from HIV-1-infected men. *AIDS* 8:1669–1674
- Banerjee A, Baker ML, Kulcsar K et al (2020) Novel Insights Into Immune Systems of Bats. *Front Immunol* 11:26
- Barboza JM, Medina H, Doria M et al (2004) Use of atomic force microscopy to reveal sperm ultrastructure in HIV-patients on highly active antiretroviral therapy. *Arch Androl* 50:121–129
- Barr JJ, Auro R, Furlan M et al (2013) Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci USA* 110:10771–10776
- Bayne CJ (2003) Origins and evolutionary relationships between the innate and adaptive arms of immune systems. *Integr Comp Biol* 43:293–299
- Bell PJJ (2020) Evidence supporting a viral origin of the eukaryotic nucleus. *Virus Res* 289:198168
- Belshaw R, Pereira V, Katzourakis A et al (2004) Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci USA* 101:4894–4899

- Belyi VA, Levine AJ, Skalka AM (2010a) Sequence from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. *J Virol* 84:12458–12462
- Belyi VA, Levine AJ, Skalka AM (2010b) Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/Marburgvirus sequences in vertebrate genomes. *PLoS Pathog* 6: e1001030
- Berrens RV, Andrews S, Spensberger D et al (2017) An endosRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell Stem Cell* 21:694–703
- Best S, Le Tissier P, Towers G et al (1996) Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* 382:826–829
- Biryukov J, Meyers C (2018) Superinfection exclusion between two high-risk human papillomavirus types during a coinfection. *J Virol* 92:e01993–e01917
- Blanc G, Gallot-Lavallée L, Maumus F (2015) Provirophages in the *Bigelowiella* genome bear testimony to past encounters with giant viruses. *Proc Natl Acad Sci USA* 112:E5318–E5326
- Blanco-Melo D, Gifford RJ, Bieniasz PD (2017) Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. *eLife* 6:e22519
- Boehm T, McCurley N, Sutoh Y et al (2012) VLR-based adaptive immunity. *Annu Rev Immunol* 30:203–220
- Bondy-Denomy J, Qian J, Westra ER et al (2016) Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* 10:2854–2866
- Born D, Reuter L, Mersdorf U et al (2018) Capsid protein structure, self-assembly, and processing reveal morphogenesis of the marine virophage mavirus. *Proc Natl Acad Sci USA* 115:7332–7337
- Brightman BK, Li QX, Trepp DJ et al (1991) Differential disease restriction of Moloney and Friend leukemia viruses by the mouse Rcmf gene is governed by the viral long terminal repeat. *J Exp Med* 174:389–396
- Broecker F, Moelling K (2019a) What viruses tell us about evolution and immunity: beyond Darwin? *Ann N Y Acad Sci USA* 1447:53–68
- Broecker F, Moelling K (2019b) Evolution of immune systems from viruses and transposable elements. *Front Microbiol* 10:51
- Burnet FM, Fenner F (1949) *The production of antibodies*, 2nd edn. Macmillan, Melbourne, Australia
- Canchaya C, Proux C, Fournous G et al (2003) Prophage genomics. *Microbiol Mol Biol Rev* 67:238–276
- Canchaya C, Fournous G, Brüßow H (2004) The impact of prophages on bacterial chromosomes. *Mol Microbiol* 53:9–18
- Cardona-Maya W, Velilla P, Montoya CJ et al (2009) Presence of HIV-1 DNA in spermatozoa from HIV-positive patients: changes in the semen parameters. *Curr HIV Res* 7:418–424
- Cardona-Maya W, Velilla PA, Montoya CJ et al (2011) In vitro human immunodeficiency virus and sperm cell interaction mediated by the mannose receptor. *J Reprod Immunol* 92:1–7
- Chu H, Jo Y, Cho WK (2014) Evolution of endogenous non-retroviral genes integrated into plant genomes. *Curr Plant Biol* 1:55–59
- Chung IY, Jang HJ, Bae HW et al (2014) A phage protein that inhibits the bacterial ATPase required for type IV pilus assembly. *Proc Natl Acad Sci USA* 111:11503–11508
- Chuong EB, Elde NC, Feschotte C (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351:1083–1087
- Cianciolo GJ, Copeland TD, Oroszlan S et al (1985) Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* 230:453–455
- Citti L, Rainaldi G (2005) Synthetic hammerhead ribozymes as therapeutic tools to control disease genes. *Curr Gene Ther* 5:11–24
- Colson P, Ravaux I, Tamalet C et al (2014) HIV infection en route to endogenization: two cases. *Clin Microbiol Infect* 20:1280–1288

- Cooper MD, Alder MN (2006) The evolution of adaptive immune systems. *Cell* 124:815–822
- Cornelis G, Funk M, Vernochet C et al (2017) An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. *Proc Natl Acad Sci USA* 114:E10991–E11000
- Cui J, Holmes EC (2012) Endogenous lentiviruses in the ferret genome. *J Virol* 86:3383–3385
- Cullen BR, Cherry S, tenOever BR (2013) Is RNA interference a physiologically relevant innate immune response in mammals? *Cell Host Microbe* 14:374–378
- da Fonseca GC, de Oliveira LFV, de Moraes GL et al (2016) Unusual RNA plant virus integration into the soybean genome leads to the production of small RNAs. *Plant Sci* 246:62–69
- de Koning AP, Gu W, Castoe TA et al (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7:e1002384
- Diener TO (1989) Circular RNAs: relics of precellular evolution? *Proc Natl Acad Sci USA* 86:9370–9374
- Doron S, Melamed S, Ofir G et al (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359:eaar4120
- Dunlap KA, Palmarini M, Varela M (2006) Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proc Natl Acad Sci USA* 103:14390–14395
- Enard D, Le C, Gwennap C et al (2016) Viruses are a dominant driver of protein adaptation in mammals. *eLife* 5:e12469
- Escalera-Zamudio M, Greenwood AD (2016) On the classification and evolution of endogenous retrovirus: human endogenous retroviruses may not be ‘human’ after all. *APMIS* 124:44–51
- Ewing RM, Chu P, Elisma F et al (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* 3:89
- Fischer MG (2020) The virophage family Lavidaviridae. *Curr Issues Mol Biol* 40:1–24
- Fischer MG, Hackl T (2016) Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature* 540:288–291
- Fischer MG, Allen MJ, Wilson WH et al (2010) Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc Natl Acad Sci USA* 107:19508–19513
- Flores R, Gago-Zachert S, Serra P et al (2014) Viroids: survivors from the RNA world? *Annu Rev Microbiol* 68:395–414
- Fujino K, Horie M, Honda T et al (2014) Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. *Proc Natl Acad Sci USA* 111:13175–13180
- Furuta Y, Abe K, Kobayashi I (2010) Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res* 38:2428–2443
- Ge DT, Zamore PD (2013) Small RNA-directed silencing: the fly finds its inner fission yeast? *Curr Biol* 23:R318–R320
- Gifford R, Tristem M (2003) The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* 26:291–315
- Gifford RJ, Katzourakis A, Tristem M et al (2008) A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci USA* 105:20362–20367
- Gilbert C, Maxfield DG, Goodman SM et al (2009) Parallel germline infiltration of a lentivirus in two Malagasy lemurs. *PLoS Genet* 5:e1000425
- Gogvadze E, Buzdin A (2009) Retroelements and their impact on genome evolution and functioning. *Cell Mol Life Sci* 66:3727–3742
- Gonzalez-Hernandez MJ, Swanson MD, Contreras-Galindo R et al (2012) Expression of human endogenous retrovirus type K (HML-2) is activated by the Tat protein of HIV-1. *J Virol* 86:7790–7805
- Greenwood AD, Ishida Y, O’Brien SP et al (2018) Transmission, evolution, and endogenization: lessons learned from recent retroviral invasions. *Microbiol Mol Biol Rev* 82:e00044–e00017

- Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172
- Gulino K, Rahman J, Badri M et al (2020) Initial mapping of the New York City Wastewater Virome. *mSystems* 5:e00876-19
- Han GZ, Worobey M (2012) Endogenous lentiviral elements in the weasel family (Mustelidae). *Mol Biol Evol* 29:2905–2908
- Han GZ, Worobey M (2015) A primitive endogenous lentivirus in a colugo: insights into the early evolution of lentiviruses. *Mol Biol Evol* 32:211–215
- Haraguchi S, Good RA, James-Yarish M et al (1995) Differential modulation of Th1- and Th2-related cytokine mRNA expression by a synthetic peptide homologous to a conserved domain within retroviral envelope protein. *Proc Natl Acad Sci USA* 92:3611–3615
- Haraguchi S, Good RA, Cianciolo GJ et al (1997) Immunosuppressive retroviral peptides: immunopathological implications for immunosuppressive influences of retroviral infections. *J Leukoc Biol* 61:654–666
- Haraguchi S, Good RA, Day-Good NK (2008) A potent immunosuppressive retroviral peptide: cytokine patterns and signaling pathways. *Immunol Res* 41:46–55
- Hartley JW, Yetter RA, Morse HC III (1983) A mouse gene on chromosome 5 that restricts infectivity of mink cell focus-forming recombinant murine leukemia viruses. *J Exp Med* 158:16–24
- Hayward JA, Tachedjian M, Kohl C (2020) Infectious KoRV-related retroviruses circulating in Australian bats. *Proc Natl Acad Sci USA* 117:9529–9536
- Hille F, Richter H, Wong SP et al (2018) The biology of CRISPR-Cas: backward and forward. *Cell* 172:1239–1259
- Hoffmann B, Tappe D, Höper D et al (2015) A variegated squirrel bornavirus associated with fatal human encephalitis. *N Engl J Med* 373:154–162
- Horie M (2017) The biological significance of bornavirus-derived genes in mammals. *Curr Opin Virol* 25:1–6
- Horie M, Honda T, Suzuki Y et al (2010) Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* 463:84–87
- Huang S, Tao X, Yuan S et al (2016) Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell* 166:102–114
- Hyndman TH, Shilton CM, Stenglein MD et al (2018) Divergent bornaviruses from Australian carpet pythons with neurological disease date the origin of extant Bornaviridae prior to the end-Cretaceous extinction. *PLoS Pathog* 14:e1006881
- Ikeda H, Laigret F, Martin MA et al (1985) Characterization of a molecularly cloned retroviral sequence associated with FV-4 resistance. *J Virol* 55:768–777
- Imakawa K, Nakagawa S (2017) The phylogeny of placental evolution through dynamic integrations of retrotransposons. *Prog Mol Biol Transl Sci* 145:89–109
- Iranzo J, Cuesta JA, Manrubia S et al (2017) Disentangling the effects of selection and loss bias on gene dynamics. *Proc Natl Acad Sci USA* 114:E5616–E5624
- Ito J, Watanabe S, Hiratsuka T et al (2013) Refrex-1, a soluble restriction factor against feline endogenous and exogenous retroviruses. *J Virol* 87:12029–12040
- Ito J, Sugimoto R, Nakaoka H et al (2017) Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet* 13:e1006883
- Iwasaki YW, Siomi MC, Siomi H (2015) PIWI-interacting RNA: its biogenesis and functions. *Annu Rev Biochem* 84:405–433
- Jebb D, Huang Z, Pippel M et al (2020) Six reference-quality genomes reveal evolution of bat adaptations. *Nature* 583:578–584
- Jimenez RM, Polanco JA, Lupták A (2015) Chemistry and biology of self-cleaving ribozymes. *Trends Biochem Sci* 40:648–661
- Jones JDG, Vance RE, Dangl JL (2016) Intracellular innate immune surveillance devices in plants and animals. *Science* 354:aaf6395

- Jung YT, Lyu MS, Buckler-White A et al (2002) Characterization of a polytropic murine leukemia virus proviral sequence associated with the virus resistance gene *Rmcf* of DBA/2 mice. *J Virol* 76:8218–8224
- Kambris Z, Hoffmann JA, Imler JL et al (2002) Tissue and stage-specific expression of the tolls in *Drosophila* embryos. *Gene Expr Patterns* 2:311–317
- Kapitonov VV, Koonin EV (2015) Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biol Direct* 10:20
- Katzourakis A, Gifford RJ (2010) Endogenous viral elements in animal genomes. *PLoS Genet* 6:e1001191
- Katzourakis A, Tristem M, Pybus OG et al (2007) Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci USA* 104:6261–6265
- Khan F, Furuta Y, Kawai M, Kaminska KH, Ishikawa K, Bujnicki JM, Kobayashi I (2010) A putative mobile genetic element carrying a novel type IIF restriction-modification system (PluTI). *Nucleic Acids Res* 38:3019–3030
- Kim KS, Yamamoto Y, Nakaoka S et al (2020) Modeling Borna disease virus in vitro spread reveals the mode of antiviral effect conferred by an endogenous Bornavirus-like element. *J Virol* 94:e01204–e01220
- Kitsou K, Kotanidou A, Paraskevis D et al (2020) Upregulation of human endogenous retroviruses in bronchoalveolar lavage fluid of COVID-19 patients. medRxiv. <https://doi.org/10.1101/2020.05.10.20096958>
- Kobayashi Y, Horie M, Nakano A et al (2016) Exaptation of bornavirus-like nucleoprotein elements in afrotherians. *PLoS Pathog* 12:e1005785
- Koonin EV (2016) Viruses and mobile elements as drivers of evolutionary transitions. *Philos Trans R Soc Lond B Biol Sci* 371:20150442
- Koonin EV (2018) Hunting for treasure chests in microbial defense islands. *Mol Cell* 70:761–762
- Koonin EV, Dolja VV (2013) A virocentric perspective on the evolution of life. *Curr Opin Virol* 3:546–557
- Koonin EV, Dolja VV (2014) Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol Mol Biol Rev* 78:278–303
- Koonin EV, Krupovic M (2015) Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet* 16:184–192
- Koonin EV, Krupovic M (2016) Virology: a parasite's parasite saves host's neighbors. *Nature* 540:2014–2205
- Koonin EV, Krupovic M (2018) The depths of virus exaptation. *Curr Opin Virol* 31:1–8
- Koonin EV, Makarova KS (2017) Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome Biol Evol* 9:2812–2825
- Koonin EV, Wolf YI, Katsnelson MI (2017) Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. *Biol Direct* 12:31
- Kovalskaya N, Hammond RW (2014) Molecular biology of viroid-host interactions and disease control strategies. *Plant Sci* 228:48–60
- Krupovic M, Makarova KS, Forterre P et al (2014) Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol* 12:36
- Krupovic M, Dolja VV, Koonin EV (2020) The LUCA and its complex virome. *Nat Rev Microbiol* 18:661–670
- Lange C, Hemmrich G, Klostermeier UC et al (2011) Defining the origins of the NOD-like receptor system at the base of animal evolution. *Mol Biol Evol* 28:1687–1702
- Lavialle C, Cornelis G, Dupressoir A et al (2013) Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci* 368:20120507
- Lee EJ, Banerjee S, Zhou H et al (2011) Identification of piRNAs in the central nervous system. *RNA* 17:1090–1099
- Leulier F, Lemaitre B (2008) Toll-like receptors—taking an evolutionary approach. *Nat Rev Genet* 9:165–178

- Mahla RS, Reddy MC, Prasad DV et al (2013) Sweeten PAMPs: role of sugar complexed PAMPs in innate immunity and vaccine biology. *Front Immunol* 4:248
- Makarova KS, Wolf YI, Snir S, Koonin EV (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol* 193:6039–6056
- Malfavon-Borja R, Feschotte C (2015) Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. *J Virol* 89:4047–4050
- McGrath S, Fitzgerald GF, van Sinderen D (2002) Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. *Mol Microbiol* 43:509–520
- McKinney HH (1929) Mosaic diseases in the Canary Islands, West Africa and Gibraltar. *J Agric Res* 39:557–578
- Michel N, Allespach I, Venzke S et al (2005) The Nef protein of human immunodeficiency virus establishes superinfection immunity by a dual strategy to down-regulate cell-surface CCR5 and CD4. *Curr Biol* 15:714–723
- Miesen P, Joosten J, van Rij RP (2016) PIWIs Go Viral: arbovirus-derived piRNAs in vector mosquitoes. *PLoS Pathog* 12:e1006017
- Moelling K, Matskevich A, Jung JS (2006) Relationship between retroviral replication and RNA interference machineries. *Cold Spring Harb Symp Quant Biol* 71:365–368
- Monde K, Terasawa H, Nakano Y et al (2017) Molecular mechanisms by which HERV-K Gag interferes with HIV-1 Gag assembly and particle infectivity. *Retrovirology* 14:27
- Mougari S, Sahmi-Bounsiar D, Levasseur A et al (2019) Virophages of giant viruses: an update at eleven. *Viruses* 11:733
- Mougari S, Chelkha N, Sahmi-Bounsiar D et al (2020) A virophage cross-species infection through mutant selection represses giant virus propagation, promoting host cell survival. *Commun Biol* 3:248
- Murphy J, Mahony J, Ainsworth S et al (2013) Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Appl Environ Microbiol* 79:7547–7555
- Naderer M, Brust JR, Knowle D, Blumenthal RM (2002) Mobility of a restriction-modification system revealed by its genetic contexts in three hosts. *J Bacteriol* 184:2411–2419
- Nandi S, Chandramohan D, Fioriti L et al (2016) Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. *Proc Natl Acad Sci USA* 133:12697–12702
- Nuovo GJ, Becker J, Simsir A et al (1994) HIV-1 nucleic acids localize to the spermatogonia and their progeny. A study by polymerase chain reaction in situ hybridization. *Am J Pathol* 144:1142–1148
- Obbard DJ, Gordon KHJ, Buck AH et al (2009) The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci* 364:99–115
- Olagoke O, Quigley BL, Eiden MV et al (2019) Antibody response against koala retrovirus (KoRV) in koalas harboring KoRV-A in the presence or absence of KoRV-B. *Sci Rep* 9:12416
- Olival KJ, Hosseini PR, Zambrana-Torrel C et al (2017) Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646–650
- Ophinni Y, Palatini U, Hayashi Y et al (2019) piRNA-guided CRISPR-like immunity in eukaryotes. *Trends Immunol* 40:998–1010
- Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA et al (2016) Uncovering Earth's virome. *Nature* 536:425–430
- Palmarini M, Mura M, Spencer TE (2004) Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation. *J Gen Virol* 85:1–13
- Parrish NF, Fujino K, Shiromoto Y et al (2015) piRNAs derived from ancient viral processed pseudogenes are transgenerational sequence specific immune memory in mammals. *RNA* 21:1691–1703
- Planz O, Stitz L (1999) Borna disease virus nucleoprotein (p40) is a major target for CD8⁺-T-cell-mediated immune response. *J Virol* 73:1715–1718
- Pradeu T, Carosella ED (2006) The self model and the conception of biological identity in immunology. *Biol Philos* 21:235–252

- Sarker N, Fabijan J, Owen H et al (2020) Koala retrovirus viral load and disease burden in distinct northern and southern koala populations. *Sci Rep* 10:263
- Sattler S, Ghadially H, Hofer E (2012) Evolution of the C-type lectin-like receptor genes of the DECTIN-1 cluster in the NK gene complex. *ScientificWorldJournal* 2012:931386
- Seed KD, Lazinski DW, Calderwood SB et al (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* 494:489–491
- Shabalina SA, Koonin EV (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* 23:578–587
- Skirmuntt EC, Escalera-Zamudio M, Teeling EC et al (2020) The potential role of endogenous viral elements in the evolution of bats as reservoirs for zoonotic viruses. *Annu Rev Virol* 7:103–119
- Sofuku K, Parrish NF, Honda T et al (2015) Transcription profiling demonstrates epigenetic control of non-retroviral RNA virus-derived elements in the human genome. *Cell Rep* 12:1548–1554
- Spencer TE, Mura M, Gray CA et al (2003) Receptor usage and fetal expression of ovine endogenous betaretroviruses: implications for coevolution of endogenous and exogenous retroviruses. *J Virol* 77:749–753
- Stitz L, Bilzer T, Richt JA et al (1993) Pathogenesis of Borna disease. *Arch Virol Suppl* 7:135–151
- Suzuki S (1975) FV-4: a new gene affecting the splenomegaly induction by Friend leukemia virus. *Jpn J Exp Med* 45:473–478
- Suzuki Y, Kobayashi Y, Horie M et al (2014) Origin of an endogenous bornavirus-like nucleoprotein element in thirteen-lined ground squirrels. *Genes Genet Syst* 89:143–148
- Suzuki Y, Frangeul L, Dickson LB et al (2017) Uncovering the repertoire of endogenous flaviviral elements in *Aedes* mosquito genomes. *J Virol* 91:e00571–e00517
- Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J (2014) The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* 21:743–753
- Takahashi N, Ohashi S, Sadykov MR, Mizutani-Ui Y, Kobayashi I (2011) IS-linked movement of a restriction-modification system. *PLoS One* 6:e16554
- Tarlinton RE, Meers J, Young PR (2006) Retroviral invasion of the koala genome. *Science* 442:79–81
- Tarlinton RE, Sarker N, Fabijan J et al (2017) Differential and defective expression of Koala Retrovirus reveal complexity of host and virus evolution. *bioRxiv*. <https://doi.org/10.1101/211466>
- Terry SN, Manganaro L, Cuesta-Dominguez A et al (2017) Expression of HERV-K108 envelope interferes with HIV-1 production. *Virology* 509:52–59
- Travis J (2009) Origins. On the origin of the immune system. *Science* 324:580–582
- Villarreal LP (2011) Viral ancestors of antiviral systems. *Viruses* 3:1933–1958
- Villarreal LP, Witzany G (2013) The DNA habitat and its RNA inhabitants: at the dawn of RNA sociology. *Genomics Insights* 6:1–12
- Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK (2010) Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* 1:147
- Weinberger AD, Sun CL, Pluciński MM et al (2012) Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput Biol* 8:e1002475
- Westra ER, van Houte S, Gandon S et al (2019) The ecology and evolution of microbial CRISPR-Cas adaptive immune systems. *Philos Trans R Soc Lond B Biol Sci* 374:20190101
- Wilkins AS (2010) The enemy within: an epigenetic role of retrotransposons in cancer initiation. *Bioassays* 32:856–865
- Wolf YI, Silas S, Wang Y et al (2020) Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol* 5:1262–1270
- Wu Q, Luo Y, Lu R et al (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci USA* 107:1606–1611

- Xu S, Zhou L, Liang X et al (2020) Novel cell-virus-virophage tripartite infection system discovered in the freshwater lake Dishui Lake in Shanghai, China. *J Virol* 94:e00149–e00120
- Zhang L, Richards A, Khalil A et al (2020) SARS-CoV-2 RNA reverse-transcribed and integrated into the human genome. *bioRxiv*. <https://doi.org/10.1101/2020.12.12.422516>
- Ziebell H, Carr JP (2010) Cross-protection: a century of mystery. *Adv Virus Res* 76:211–264
- Zou J, Chang M, Nie P et al (2009) Origin and evolution of the RIG-I like RNA helicase gene family. *BMC Evol Biol* 9:85