



OPEN ACCESS

EDITED BY

James L. Stafford,
University of Alberta, Canada

REVIEWED BY

Sushil Kumar,
Indian Institute of Technology
Bombay, India
Shoutao Zhang,
Zhengzhou University, China

*CORRESPONDENCE

Christian A. Devaux
✉ christian.devaux@mediterranee-
infection.com

SPECIALTY SECTION

This article was submitted to
Comparative Immunology,
a section of the journal
Frontiers in Immunology

RECEIVED 08 July 2022

ACCEPTED 05 December 2022

PUBLISHED 23 December 2022

CITATION

Devaux CA, Pontarotti P, Nehari S and
Raoult D (2022) 'Cannibalism' of
exogenous DNA sequences: The
ancestral form of adaptive immunity
which entails recognition of danger.
Front. Immunol. 13:989707.
doi: 10.3389/fimmu.2022.989707

COPYRIGHT

© 2022 Devaux, Pontarotti, Nehari and
Raoult. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

'Cannibalism' of exogenous DNA sequences: The ancestral form of adaptive immunity which entails recognition of danger

Christian A. Devaux^{1,2*}, Pierre Pontarotti^{1,2}, Sephora Nehari¹
and Didier Raoult¹

¹Aix-Marseille University, Institut de recherche pour le développement (IRD), Assistance Publique Hôpitaux de Marseille (APHM), MEPHI, Institut Hospitalo-universitaire (IHU)-Méditerranée Infection, Marseille, France, ²Department of Biological Sciences, Centre National de la Recherche Scientifique, Centre National de la Recherche Scientifique (CNRS)-SNC5039, Marseille, France

Adaptive immunity is a sophisticated form of immune response capable of retaining the molecular memory of a very great diversity of target antigens (epitopes) as non-self. It is capable of reactivating itself upon a second encounter with an immunoglobulin or T-cell receptor antigen-binding site with a known epitope that had previously primed the host immune system. It has long been considered that adaptive immunity is a highly evolved form of non-self recognition that appeared quite late in speciation and complemented a more generalist response called innate immunity. Innate immunity offers a relatively non-specific defense (although mediated by sensors that could specifically recognize virus or bacteria compounds) and which does not retain a memory of the danger. But this notion of recent acquisition of adaptive immunity is challenged by the fact that another form of specific recognition mechanisms already existed in prokaryotes that may be able to specifically auto-protect against external danger. This recognition mechanism can be considered a primitive form of specific (adaptive) non-self recognition. It is based on the fact that many archaea and bacteria use a genome editing system that confers the ability to appropriate viral DNA sequences allowing prokaryotes to prevent host damage through a mechanism very similar to adaptive immunity. This is indistinctly called, 'endogenization of foreign DNA' or 'viral DNA predation' or, more pictorially 'DNA cannibalism'. For several years evidence has been accumulating, highlighting the crucial role of endogenization of foreign DNA in the fundamental processes related to adaptive immunity and leading to a change in the dogma that adaptive immunity appeared late in speciation.

KEYWORDS

adaptive immunity, DNA cannibalism, transposons, recombination activating gene (RAG), defective proviruses

Introduction

Exocannibalism (from Greek *exo-*, 'from outside' and cannibalism, 'to eat humans'), is the consumption of flesh outside one's close social group. Why introduce the concept of DNA 'cannibalism' to describe endogenization of foreign DNA? Over the past centuries, archaeologists and anthropologists have provided evidence of human cannibalism, a primitive behavior consisting of eating one's own kind. Although cannibalism may have been practiced for different reasons, one of the best-known forms of human cannibalism took place in the context of warfare, either to fully eradicate enemies in a show of anger or to acquire the powers of those who had been defeated (1). Cannibalism is currently a major human taboo, but it is surprisingly common in the animal kingdom and there is a lot of good reasons to eat your own kind (2–4). In this paper the reference to the concept of 'cannibalism' is only a metaphor, voluntarily chosen to image a situation of 'genetic eating of the enemy in a context of warfare'. It illustrates one way among others to evade dangers in the more global context of the fundamental belief that the immune system is designed to discriminate some self from some non-self, especially when the non-self is a pathogenic agents. 'Stricto sensu' when one evokes a phenomenon of capture by one species (e.g., humans) of DNA sequences relating to a foreign species (e.g., virus), the word 'cannibalism' is not appropriate insofar as cannibalism refers to events which occurs between individuals of the same species (consumption of conspecifics as food). It is however worth noting that once endogenized, certain DNA sequences can make it possible to fight against enemies belonging to the species from which they originate. By analogy with primitive human cannibalism, it can be considered that a 'form of cannibalism' existed as early as prokaryotes, which have managed to appropriate the 'strength of their aggressor' and 'fight' against its related organisms through 'cannibalism of enemy DNA sequences'. Seen under this prism, it is interesting to wonder about the role that DNA 'cannibalism' may have played in the conservation of specific defense mechanisms during speciation of both prokaryotes and eukaryotes, and on the mechanisms of DNA sequence transfers (e.g., horizontal transfer, vertical transfer).

Obviously, this concept of DNA 'cannibalism' may be surprising for immunologists who advocate: i) that the immune system's (IS) primary driving force is not to discriminate between self and non-self (complicated structures that often share features with self antigens) but that its function mainly entails recognition of danger; ii) that IS maintains a state of tolerance to pathogen load without compromising host fitness; iii) that the pathological outcome of the microorganism-host interaction is determined by the amount of damage to the host with the possible contributions of both the

microorganism and the host; and/or iv) that hereditary information is embedded in diverse physical forms (DNA, epigenetic methylation, RNA, proteins, symbionts) representing a continuum of evolutionary qualities (5–7). The outcome of microorganism-host interaction can be either beneficial or detrimental to the microorganism, to the host, or to both. Evolution of life can be guided by the microorganism-host interactions and is contingent on changing hereditary information relayed through time (inherited information). Ancient microorganism-host interactions in which bacteria were incorporated into a primordial host as organelles are likely to have resulted in the evolution of eukaryotes (8). Commensal bacteria maintaining symbiotic relationship with the host provide essential functions to their host such as nutrition, resistance to pathogen colonization and influence of the immune system, thereby being indispensable for the host's survival (9, 10). There is also evidence obtained in recombinant-activating gene (RAG) deficient mice suggesting that some microbes may benefit from the specific immune responses that they elicit (11). In the 'damage-response framework' model (6), a pathogen is defined as a microorganism that is capable of causing damage to a host. Many microorganisms can be considered dangerous, like viruses or bacteria secreting toxins, yet other similar microorganisms are not, like nonlytic viruses or bacteria secreting beneficial nutrient. If a virus infects a cell, replicates without harming the host cell, and moves on, it would not necessarily damage the host cell and there would be no need to eliminate it and it might even carry genes useful to the host cell (e.g. retroviruses jump in and out constantly, sometimes leaving bits of themselves behind). Archea and bacteria are constantly exposed to the invasion by elements such as phages or plasmids (12). In response they have evolved an arsenal of defense systems such as antiphage systems (frequently physically clustered in genomic 'defense islands') or defense systems specifically targeting foreign plasmids, that limit the intrusion of foreign elements (13, 14). The existence of a potential 'Pan immune system' where the effective arsenal of defense systems is not encoded by the genome of a single microorganism but rather by its pan-genome, has also been postulated (15). As far back in speciation as in the archaea and bacteria, endogenization of foreign DNA sequences was developed as a mechanism of adaptive immunity allowing the prokaryote to recognize and destroy a pathogenic agent (i.e., a bacteriophage) that would have previously infected it, or one of its ancestors. These mechanisms of DNA 'cannibalism' are probably much more numerous and useful than what has previously been described. In particular, it is currently well established that endogenous retroviruses (ERVs) are most likely the proviral remnants of ancestral virus infections. Amazingly, these ERVs form multigene families representing about 8%-10% of the human genome (16, 17). They have certainly contributed to the

evolution of most organisms in their complex ecosystems, and for some of them, to survive the pressure of pathogenic agents and their genetic drift (18). This view agrees with the endogenization-amplification theory, which postulates that on various occasions exogenous retroviruses have endogenized in animal genomes, with subsequent amplification (19). It has been estimated that the largest human endogenous retrovirus (ERV) family, HERV-K, consists of 1306 copies (20). Interestingly, Old World monkeys possess a few copies of HERV-K, whereas New World monkeys lack HERV-K, suggesting that HERV-K must have entered the human lineage after the split with the New World monkeys but before the split with the Old World monkeys (21). Since humans possess many more copies than Old World monkeys, this indicates that the HERV-K family must have greatly expanded after the human lineage split from Old World monkeys. Moreover, the study of the evolutionary history of HERV-W in humans (213 copies in the human genome), and ERV-W in primate lineages has revealed that closely related species share the same ERV elements at orthologous loci and that the number of shared elements reflects the degree of relatedness between species (22).

CRISPR-Cas, the genome editing system that confers adaptive immunity to bacteria

On infection of bacteria, bacteriophages enter either a lytic cycle resulting in lysis of bacterial cells and horizontal transmission or a lysogenic cycle characterized by the integration of the prophage into the host genome and vertical transmission. Bacteria use the genome editing Clustered Regularly Interspaced Short Palindromic Repeats CRISPR-Cas, as an immune defense based on recognition of invading lytic bacteriophages. The CRISPR-Cas systems described in archaea and bacteria have been broadly divided into two classes: Class I (divided into types I, III and IV) and Class 2 (divided into types II, V and VI) (23–25). So, when attacked by a bacteriophage, the target bacterium reacts by ‘viral predation/DNA cannibalism’ as it captures small pieces of the viral DNA and inserts them into its own DNA in a particular pattern leading to generating segments known as CRISPR arrays. The CRISPR arrays allow the ‘educated’ bacterium to ‘retain the memory’ of the aggressor, like a genetic ‘fingerprint’ of the invader. These systems are usually composed of multiple CRISPR-associated genes (cas genes) on the bacterial chromosome and CRISPR array consisting of unique DNA spacers. The CRISPR-Cas system mediates its immune defense function through three distinct phases: first, the adaptation phase during which Cas proteins (e.g., Cas 1, Cas 2 sometimes associated with Cas 4) select and process DNA fragments from the invader and insert them into the CRISPR array of the host genome leading to the acquisition

of new spacers (it is the central element of the adaptive immune response device); second the expression phase during which the CRISPR locus is transcribed into a long pre-crRNA that is cut into different crRNAs corresponding to different exogenous fragments that combine with Cas proteins constituting a surveillance unit inside the bacterium; third, the interference if the bacterium subsequently encounters the same bacteriophage (or its close relatives) again, the crRNAs-Cas complex will bind to specific complementary regions of the exogenous virus DNA. The bacterium then uses an activate nuclease (e.g. class 1, Cas 3, Cas 9, Cas10) to cut the DNA apart, which disables the virus (26). It is likely the most ancestral form of acquired immunity found in most archaea (about 90%) and in many bacteria (about 40%) (27). However, it was also reported that only 0.83% of *Staphylococcus aureus* strains from different geographical regions have type IIA CRISPR-Cas, suggesting that this specific system may result from spontaneous horizontal gene transfer event (28). It has been proposed that the ‘ancestral CRISPR-Cas system’ appeared in thermophilic archaea, then spread to other lineages of archaea and bacteria through horizontal transfers (29). This system then evolved, both at the level of acquisition of immunity and of the genes coding for the proteins in this defense system, explaining why bacteriophage infection is a major driving force for the maintenance of CRISPR-Cas immune system and why there is a strong selective benefits of phage-encoded anti-CRISPR (Acr) genes for both the phage and the host under context where lysogeny is suppressed. It has been suggested that maladaptive type I CRISPR-Cas immune systems (cannot eliminate the invading bacteriophages due to imperfect matching of spacers to the integrated prophage) occur frequently in nature (30). The Cas proteins constituting the molecular machinery of adaptive immunity have nuclease, helicase, and polymerase-like activities. The diversification of the CRISPR-Cas systems is likely to be partly driven by their competitive coevolution with virus (e.g., phage)-encoded Acr proteins that interact with different CRISPR-Cas components, preventing Cas proteins from binding or cleaving phage DNA (24, 26).

The same mechanism was also reported by one of us (DR) in giant viruses, capable of integrating a sequence from a Zamilon virophage in an operon that was named ‘Mimivirus virophage resistance element’ (MIMIVIRE). Indeed, MIMIVIRE confers lineage A strains of Mimivirus resistance to Zamilon infectivity (31). One of us described this as the ‘Eat me cake’ theory, or ‘genetic cannibalism of the enemy’ theory (32). However, the fact that MIMIVIRE represents an adaptive immune system in giant viruses, remains a subject of debate mainly because the corresponding Zamilon sequence appears devoid of distinct flanking sequences that may serve as protospacer adjacent motifs (33). Anyway, cannibalization of ‘enemy DNA sequences’ likely plays a major role in the protection of populations, including humans, against foreign pathogenic agents.

From speciation to very elaborate adaptive immunity in humans

Through the work of immunologists, it was established that vertebrates, including humans, are remarkably equipped to respond to attacks by invading pathogens, including bacteria, viruses, fungi, and parasites. Defense mechanisms more or less specific for these pathogens have been classified into two main categories of immune responses: the 'nonspecific' innate immunity that takes minutes to hours to operate and does not retain a memory of previous responses, and the adaptive immunity that takes days to weeks before being able to operate and engages the pathogen with specificity and immune memory (Table 1). There are four main components of innate immunity, which consist of the activation of pre-existent mechanisms: 1) physical or anatomical barriers (e.g., skin, mucosa, tight junctions) and chemical barriers (e.g., mucus, lysozyme, defensin, reactive oxygen species); 2) phagocytosis (i.e., monocytes, macrophages, neutrophils); 3) blood proteins (e.g. activation of complement, lectins, fibrinogen), and 4) inflammation mediators (i.e., macrophages, mast cells, natural killer cells) and inflammation molecules (e.g., production of interleukins, cytokines and interferon). Activation of the innate immune system is initiated by soluble pattern recognition receptors (PRRs) that can be expressed on innate immune cells, bound to the extracellular matrix, or circulate in the blood as soluble molecules (34). By contrast, specific adaptive immunity depends upon the somatic diversification of antigen-receptor genes to generate vast repertoires of B- and T-cell receptors (35–38). Elegant studies from evolutionary immunologists have revealed that innate immunity can be found within the phylum of multicellular organisms (metazoans) that emerged as long as 600 million years ago followed by a remarkable diversification of metazoan species (including the vertebrate lineage) while adaptive immunity is only

evidenced with the gnathostomes (vertebrates with jaws) and evolved to work in concert with the innate immune system (39–42).

The mobile elements of the genomes at the origin of adaptive immunity

It has been proposed that the genes encoding the Cas1 and Cas9 proteins which are experts in cutting and integrating DNA fragments, originate from genes carried by transposons (43, 44). Comparative genomic analysis of Cas1 homologs that are not associated with CRISPR-Cas loci led to the discovery of a novel family of self-synthesizing transposons, the Casposons (e.g., Casposon-encoded Cas 1 or casposase) (24). It is striking to note that Variable/Diversity/Joining (V(D)J) recombination, which makes it possible to create a great diversity of immunoglobulins ($>10^{14}$ different paratopes from immunoglobulins also known as antibodies generated by B-cells) and T-cell receptors (T-cell receptors recognize peptide fragments presented by the major histocompatibility complex class I or class II molecules) in gnathostomes, is based on Rag (recombination activating gene) 1 and Rag2 proteins of the recombinase complex that found their origin in genes carried by a transposon.

Transposable elements (TEs) are major components of all vertebrate genomes that cause genomic instability (45). Transposons are mobile DNA segments (or MGE, for mobile genetic elements) that move from one location to another within the host genome. The retrotransposons also called Class I RNA transposons transpose *via* reverse transcription of an RNA intermediate copied back to the DNA form, while Class II DNA transposons move within the host genome *via* an 'excision-insertion' mechanism (46, 47).

TABLE 1 Examples of the outcomes of host defense against microorganisms.

Defense system	Host benefit	Retains the 'memory' of invader
In bacteria		
-Innate immunity		
Phage infection sensor (e.g., CapRelSJ46)	Yes	No
-Adaptive defense		
CRISPR-Cas adaptive immunity	Yes	Yes
In humans		
-Innate immunity		
Defensin	Yes	No
Interferon	Yes	No
-Adaptive defense		
B-cell response	Yes (most frequently)	Yes
Horizontal transfer of transposon	Yes (most frequently)	Yes (most frequently)
Endogenization and provirus 'cannibalism'	Yes (most frequently)	Yes (most frequently)

Although it was proposed that RAG1 could derive from a herpes virus recombinase or a retroviral nuclease (48, 49), it is more generally accepted that RAG1 derives from a Transib superfamily transposon present in an ancestor of jawed vertebrates that lived about 500 million years ago (50, 51). The ancestral RAG transposon consisting of recombination signal sequences flanking RAG1-like and RAG2-like genes was probably inserted into an exon of a jawed vertebrate ancestor Ig/TCR-like gene. Structural analysis has identified two jawed-vertebrate-specific adaptations that allowed the domestication of ancestral RAG transposase into a RAG recombinase (52). Retention of the RAG1 and RAG2 genes in gnathostomes and duplications of the V, D, and J segments likely built the ancestral recombinatorial immune system of vertebrates. Expression of the recipient gene occurred when the inserted transposon was excised by the Rag proteins and the two exon ends rejoined by double-strand DNA break repair factors. A second transposon insertion into the same exon could have again split this gene into V and D fragments to yield the tripartite V(D)J structure characteristic of the Ig/TCR variable-region genes. In this case it cannot be referred to as DNA ‘cannibalism’ but more simply the domestication of a transposon, except to put forward the unlikely hypothesis that it could be the fingerprint of an ancestral CRISPR-Cas cannibalism of RAG to defend a vertebrate against a transposon (53).

RAG genes and the repertoire of paratopes

In humans, the receptors for non-self epitopes (e.g. immunoglobulin paratopes) are composed of the so-called constant regions and the so-called variable regions that consists of V, D, and J segments. The variable regions are at the origin of the diversity of the paratopes. The V(D)J recombination can be divided into three steps. It is initiated by the Rag1 and Rag2 complex which introduces a double-strand DNA break (dsDNAb) at the border between V and D segments or D and J segments and their respective recombination signal sequences (RSS) that flank each gene segment and that are composed of conserved heptamer and nonamer sequences separated by a poorly conserved spacer sequence of either 12 or 23 bp, creating hairpin-sealed coding ends and blunt signal ends. This DNA lesion is repaired using the non-homologous end-joining (NHEJ) pathway DNA repair machinery. SNM1C/Artemis, which is recruited and phosphorylated by the Ku/DNA-PK complex, opens the DNA hairpins through its endonuclease activity (required for cleavage of hairpin intermediates). The XRCC4/Cernunnos/DNA-LigaseIV complex finally seals coding and signal joins (54–56). This highlight the crucial role of the transfer of the RAG transposons for the generation of the pathway leading to the

paratopes repertoire in humans. Domestication of a transposon to give rise to the Rag1-Rag2 recombinase and V(D)J recombination, was a pivotal event in the evolution of adaptive immune system of jawed vertebrates (52). It was recently hypothesized that an ancestral form of the current RAG recombinase function may have been domesticated for the purpose to protect the host against transposons that could jeopardize the integrity of the genome (53).

Adaptive immunity in multicellular organisms before gnathostomes with jaws: that’s the question

All extant jawed vertebrates can rearrange antibody and T-cell receptor gene segments. The common ‘use it or lose it’ theory (57) assumes that the size of genomes is not infinitely expandable, and species have to make choices. Accordingly, during evolution a ‘useful mechanism’ is generally preserved but can be eliminated from a species if it becomes useless or can be replaced by a more efficient mechanism. So, should we truly consider, as is currently the case, that there was no adaptive immunity in multicellular organisms before gnathostomes? that is all the more surprising since we now know that the first forms of adaptive immunity could be found in unicellular organisms that were present on our planet a billion years ago. Would this adaptive immunity of the first unicellular organisms have lost its usefulness with the establishment of a more effective innate immunity in metazoans or was unfavorable to the survival of multicellular organisms? Vertebrates have always been colonized by bacteria and other microorganisms (microbiota) that perform essential functions for the survival of their host (58). Under these conditions, horizontal transfer of DNA sequences is possible. Because CRISPR comes from bacteria that live on or infect humans, an important fraction of the human population has built up immunity to the Cas9 from bacteria over time (59). Yet for multicellular organisms, cells taken individually are sometimes the target of aggression, such as a viral attack that could lead to an integration of the genome of infectious agents into the host genome. That can be fatal to the host but sometimes gives the host a better adaptive value. Viruses can be exquisite vectors for shuttling foreign DNA into eukaryotic cells and favoring DNA endogenization. This is also true for vertebrates without jaws. Moreover, it has been reported that in the chicken, immunoglobulin light chain (IgL) diversity is generated by recombination between a single functional variable (VL) and joining (JL) gene segments and subsequent somatic diversification of the rearranged VL region (60, 61). The IgL loci of different birds (such as the quail, duck, pigeon, or turkey) consist of a family of VL elements but undergo a single major rearrangement event similar to that observed in chickens. In contrast, several rearrangements have been observed in the

Muscovy duck locus the genome of which contains 2 functional VL segments (VL1 and VL5) and 3 VL pseudogenes critical for antibody diversity suggesting that combinatorial IgL diversity has evolved independently in some avian species (62). Interestingly, the RAG1 and RAG2 genes have been found in the chicken but only RAG-2 mRNA expression occurs in B-cells undergoing antibody diversification by gene conversion (63). The study of *Carcharhinus* sharks genome also found the RAG-1 gene in this species with similarities to integration host factor genes of the bacterial site-specific recombination system (64). Thus, the evolution of the ancestral adaptive immune system may have been initiated by a transfer of microbial site-specific recombinases.

Considering these arguments, it is worth noting that the sequence encoding proteins *SpRag1L* and *SpRag2L* closely related to RAG1/RAG2 has been identified in an invertebrate, the purple sea urchin (an echinoderm) (65). This could represent one of the missing links, considering that major DNA elements required to achieve adaptive immunity preexisted in speciation before gnathostomes. Other echinoderms also appear to carry adjacent RAG1 and RAG2-like genes pair in their genome (51).

There is also an adaptive immunity system in agnathans but it is not achieved using RAG system. Indeed, Cyclostomes, which contain two phyla (hagfishes and lamprey) have their own adaptive immune system based on variable lymphocyte receptors (VLRs) the genes of which diversify somatically through gene conversion (66).

Horizontal transfer of transposon

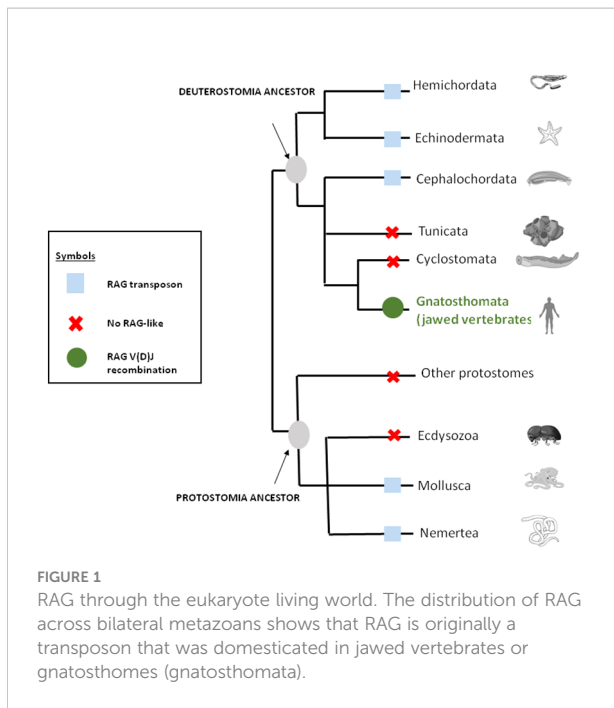
Several families of repeat ‘active’ mobile genetic elements and ‘fossil’ of previously mobile genetic elements can be found within vertebrate genomes, having been part of our molecular evolution (67–69). In humans they are composed of the long terminal repeat (LTRs) retroelements including the human endogenous retroviruses (HERV) and the retrotransposons, the non-LTR retroelements including the non-autonomous short interspersed elements (SINEs such as Alu and MIR lacking reverse transcriptase) and the long interspersed elements (LINEs such as LINE1), and the DNA transposons. If we can now explain how the system was built that makes it possible to generate adaptive immunity in humans, we do not yet have all the elements necessary to determine when and how the transfer of this genetic information could have occurred. At least seven major classes of DNA transposons are found in the human genome, with some reflecting ancient eukaryotic elements such as the ‘mariner’ sequence (70). Horizontal transmission of a transposon from fish to human has been reported for a member of the Tc1/mariner superfamily of transposons (71). Hundreds of cases of horizontal transfer into new genomic background have been described in multicellular eukaryotes, suggesting that this process has been a major force propelling eukaryotic

genome evolution (72). It has been suggested that prokaryotic transposable elements may be delivered to eukaryotic hosts, such as the IS5-like insertion sequence transferred into a bdelloid rotifer (73), or the *Merlin* superfamily related to bacterial IS1016 insertion sequences transferred to diverse animal genomes as well as humans (74). This could explain the patchy distribution of some eukaryotic DNA transposons that are phylogenetically related to bacterial insertion sequences. Yet the precise mechanisms by which the transposable element can be transferred from a prokaryote to humans remains largely unknown. It can include potential infectious vectors as suggested by the discovery of a *Bov-B*-derived short interspersed element (a non-LTR retrotransposon) from *Echis ocellatus* reptiles integrated into the genome of the taterapox virus (TATV), and orthopoxvirus that replicates in a West African rodent (75). LTR retrotransposons can make their own virus-like particles, and sometimes they encode envelope-like proteins (76), that may confer infectious properties facilitating transfer. A recent study has demonstrated the impressive extent of horizontal gene transfer (HGT) between 201 eukaryotic and 108,842 viral taxa with the identification of 1,333 candidates for virus-to-eukaryote transfers, 4,807 candidates for eukaryote-to-virus transfers, and 600 transfers with unknown directionality, altogether affecting 2,841 distinct protein families (77). Thus, the involvement of microorganisms and viruses to explain the ‘cannibalism’ of small DNA sequences is easy to imagine.

As mentioned above, genes closely related to RAG1 and RAG2 found in echinoderm (51, 65) could correspond to a primitive Rag1/2 but their relationship to transposons is uncertain since this gene pair apparently lacks the terminal invert repeat (TIR) and target site duplication (TSD) signatures of transposons. One of us (PP) reported evidence suggesting that the RAG transposon was active through the deuterostomes evolution and is still active in several lineages (78). (Figure 1) This could represent one of the missing links considering that major DNA elements required to achieve adaptive immunity preexisted in speciation before gnathostomes. More recently, RAG-like (RAG-L) transposons were found in protostomes including oysters and mussels, suggesting their ancient bilaterian origin (79) The genetic mechanism of domestication of RAG is shown in Figure 2.

Endogenization and/or provirus DNA ‘cannibalism’

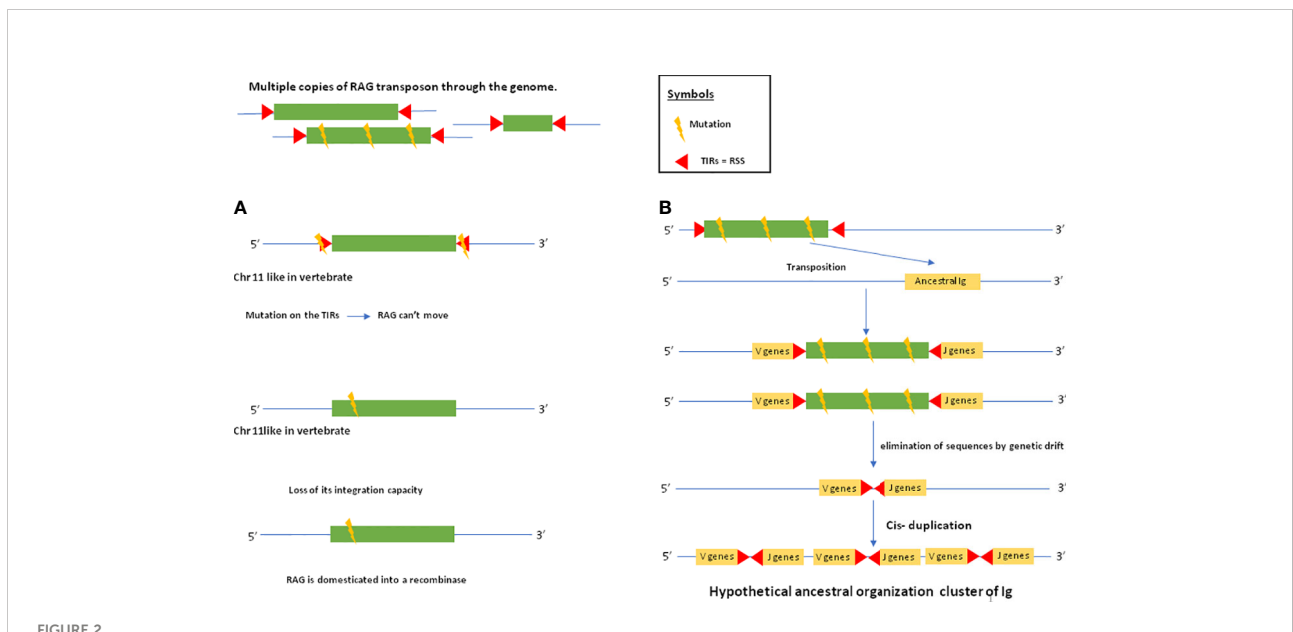
Another form of DNA ‘cannibalism’ is the endogenization of retroviruses. Although infection of humans by retroviruses such as the human T lymphotropic virus (HTLV) or human immunodeficiency virus (HIV) generally has a deleterious effect on the host as a consequence of provirus integration and



provirus gene expression, the discovery that some retroviral genes integrated into the DNA of germ-line cells (followed by their vertical transmission from mother to child) can ensure

essential biological functions has changed our view about the integration of exogenous viruses and their relationship to adaptive immunity. A common and elegant example of this mechanism can be illustrated by the synthesis of syncytin, an endogenized retroviral envelope protein involved in human placental morphogenesis (80, 81). The comparison of the sequences of syncytin genes in different mammals (primates, ruminants, rodents, etc.) has shown that they come from different retroviral integrations in the genome of the ancestors of each lineage at different times of radiation (diversification) of mammals. The oldest syncytin gene known to date is the syncytin-Car1 gene, found in all carnivore species studied. Its presence would date from at least 80 million years ago (Myr) The integration of the human syncytin 1 gene, conserved in hominoids, would date from 30 Myr while the syncytin 2 gene, present in all primates except prosimians, would be older (45 Myr) The murine syncytin A and B genes would have been integrated more than 25 Myr ago. Thus, retroviral infections have made it possible, on multiple occasions during evolution, to confer a selective advantage on infected individuals of different species, by promoting cell fusion and the formation of a syncytium at the fetomaternal interface and hence embryo development (82, 83).

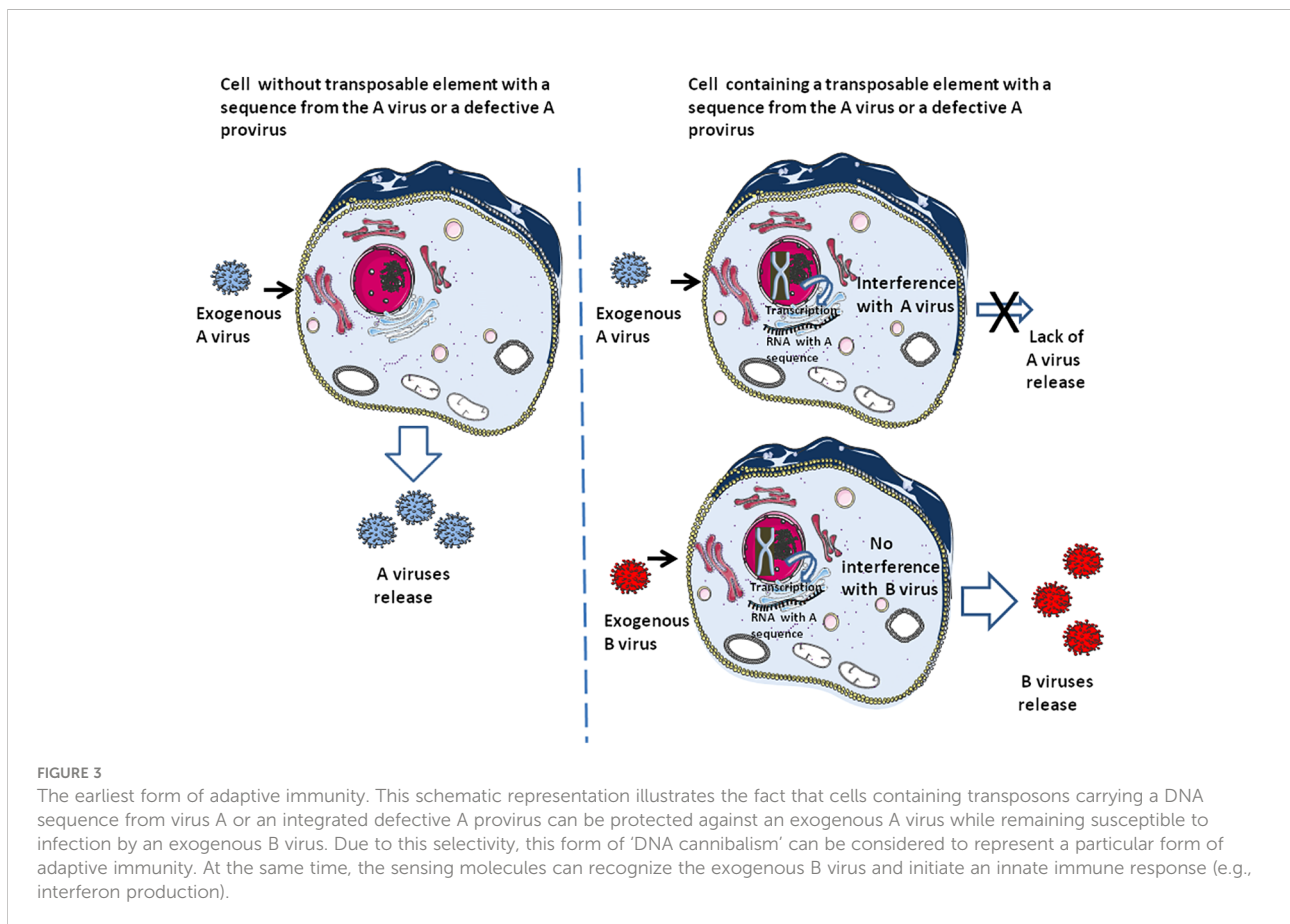
Since retroviral DNA ‘cannibalism’ has been demonstrated, it can be questioned whether retroviruses that have recently emerged and spread in human populations could be



endogenized to provide adaptative immunity to the host against the exogenous retrovirus. Despite the fact that the HTLV-2 genome has not been found integrated into the germinal cells, the low pathogenic HTLV-2 has been considered to be a good candidate for viral endogenization (84, 85). The question has been addressed regarding the endogenization of HIV (86) and this attractive hypothesis requires further clinical attention and biological investigation. It was also found that human testicular germ cells can support HIV entry and integration, which could be endogenized in the future (87). Evidence has been reported in animals indicating that this process is currently ongoing with the KoRV retrovirus causing lymphoid neoplasia and immunosuppressive (AIDS-like) disease in koalas (88–90). Following the endogenization process of KoRV-B, the ‘cannibalized KoRV-B DNA’ dam transmission of KoRV-B was demonstrated by the observation that the progeny of koalas are no longer susceptible to the virus (91, 92). Similar observations have been reported for murine leukemia virus (MLV) and mouse mammary tumor virus (MMTV) endogenous in mice (93). Endogenization of MMTV-like elements has been reported in genomes of American pikas (94) or endogenous jaagsiekte sheep retrovirus (enJSRV) in ovine and caprine genomes (95).

DNA ‘cannibalism’ of provirus

Defective viral genome (DVG) that produce defective interfering particles containing a fraction of the viral genome are only able to replicate in the presence of a helper virus and can interfere with the replication of homologous infectious viruses containing their full-length genome (96–100). It is thought that DVG may function in the establishment of virus persistence (101, 102). This defective provirus can also be considered as a process related to DNA ‘cannibalism’, conferring either innate immunity (i.e., through induction of interferon), specific adaptive immunity to the host cell against the homologous infectious virus (or the host species when present in the germinal cells), or both (Figure 3). This ‘DNA cannibalism’ could be regarded as a provirus DNA ‘cannibalism’ since in this case it is an infectious virus from which part of the proviral genome has been deleted that is endogenized. Regarding the human retrovirus, HTLV-1 is known to generate numerous defective viruses, and within such defective viruses a majority had only one Long terminal repeat (LTR), and the 5’LTR was preferentially deleted in about 40-50% of lymphoma-type adult T leukaemia/



lymphoma (103). As reported by one of us (CD), the HTLV-1 HBZ protein, encoded by an anti-sense mRNA, was found expressed in ATL cells carrying defective provirus and is associated with progression to lymphoma (104). Regarding HIV, defective proviruses are produced in large quantities during natural infection. It was found that defective particles do not interfere with virus production from proviral DNA but rather inhibit re-infection, thereby limiting the spread of infection and progression of the disease by reducing the yield of infectious virus (105). Moreover, HIV-1 infectivity and core assembly were reported to be altered due to the interference of HIV-1 Gag formation by HERV-K Gag particles (106). Consequently, post-suicide provirus DNA cannibalism and HERV interference could contribute to generating specific adaptive immunity against infectious viruses.

DNA ‘cannibalism’ of viral sequences from non-retroviral origin

So far, we have discussed mechanisms involving exogenous retroviruses that have been endogenized or are in the process of endogenizing. But what about viruses that do not belong to the retrovirus family? The main example is the human herpesvirus 6 (HHV-6). HHV-6 infects most people during childhood and can reactivate later in life to cause disease. Both HHV-6A and HHV-6B are, in rare cases, found integrated in leukocyte chromosomes of immunocompetent individuals (107). Sequences homologous to HHV-6 have been found integrated within the genome of about 1% of humans, but it is not clear how this could have occurred. The subtelomeric integration of an ancient HHV6A variant on chromosome 22q (chr22q) in East Asians affects PIWI-interacting RNAs (piRNAs) known to block germline integration of transposons (108).

Adeno-associated virus (AAV), a nonautonomous parvovirus, can establish latency through site-specific genome integration into human chromosome 19 at the *AAVS1/Mbs85* locus involving the AAV2 Rep68 protein (109, 110). Interesting observations have been reported regarding parvoviruses that suggest these viruses have frequently invaded the germ lines of diverse animal species, including mammals, fish, birds and arthropods (111). The identification of orthologous endogenous parvovirus sequences in the genomes of humans and other mammals suggests that parvoviruses have coexisted with mammals for at least 98 Myr. Furthermore, the expression of some of the endogenized parvoviral genes in eukaryotic organisms suggests that these viral genes might have a beneficial function in the host. This indicates that the process of cannibalism is not restricted to retroviruses.

Discussion

Thus, two apparently very different adaptive immune systems, the ‘old’ CRISPR-Cas system of archaea and the ‘recent’ V(D)J segments rearrangement leading to immunoglobulin and T-cell receptors would have been produced from DNA sequences encoded by transposons (112). However, the precise mechanisms by which transposable elements can be transferred to humans remain largely unknown. It may include potential infectious vectors or be based on the ability of LTR retrotransposons to make their own virus-like particles, which may confer infectious properties facilitating transfer. Thus, the involvement of microorganisms and viruses to explain DNA ‘cannibalism’ of small sequences from exogenous infectious pathogens is currently easier to imagine. The very high percentage of endogenous retroviruses in the human genome suggests that these sequences are likely molecular witnesses of ancestral infections and there is no obvious reason why this process might not reproduce itself in the future with emerging retroviruses or even other human pathogens.

Adaptive immunity has been well characterized in vertebrates with jaws during the past decade. With recent advances in immunology, next generation peptidomics, and genomics it is becoming evident that the T cell receptor (TCR) repertoires of an individual is shaped both by self and non-self antigens (immunopeptides). The impact of the variability in the self-immunopeptidome on thymic selection could explain differences in the TCR repertoire of different individuals, despite an identical HLA type (113, 114). The immunopeptidome includes highly variable sequences such as transposable elements, LINE-1, and endogenous retroviruses (115). The difference in the TCR repertoire driven by the self-immunopeptidome could be important in the initiation of autoimmune diseases, cancers and the immune response to pathogens (116, 117). A primitive form of adaptive immunity we called ‘DNA cannibalism’ established in prokaryotes to combat bacteriophages, has also been widely documented. However, there was an apparent gap in between the identification of adaptive immunity in prokaryotes and vertebrates with jaws. Further investigation of these mechanisms in different species progressively allows filling in the gap and arriving at the conclusion that there has been no gap during speciation regarding adaptive immunity. The investigation of the RAG1/RAG2 in humans leads to the discovery of equivalent systems in birds, sharks, and echinoderms and highlights a process of domestication of a transposons during speciation.

There is also evidence that independent events of DNA ‘cannibalism’ in different species may have been functionally convergent in the history of endogenization. If one assumes that adaptive immunity is a general mark of living organisms, it

remains to be seen how and when the gene responsible for this function were transferred to the human genome. Arguably the best model for this is the RAG system that has been present in different species during speciation and originate from transposons although it presents homology to microbial integrase. The provirus DNA ‘cannibalism’ found in human appear quite similar to the specific adaptive immunity conferred by ‘DNA cannibalism’ of bacteriophages, as it allows human cells to fight infectious viruses that are homologous to endogenized defective proviruses. However, we know almost nothing about the fact that this process is random or governed by genetics laws.

Until now it has been thought that adaptive immunity should be regarded as a very sophisticated process of recent acquisition complementing innate immunity. In the light of recent data, it has become clear that this is not the case, and that processes leading to various forms of adaptive immunity have existed for a very long time and represent an effective tool both in the specific fight against pathogens and as an active component in the evolution of species.

Author’s note

We wish to recall that the first references to the word ‘cannibalism’ used as a metaphor intended to image a situation which is not cannibalism *stricto sensu*, are the work of the philosophers Michel Foucault (1926-1984) and Claude Levy Strauss (1908-2009), who held the chair of social anthropology at the College de France and developed the concept of ‘cannibal democracy’.

Author contributions

CD wrote the first draft of the manuscript. SN designed the figures. All authors contributed to the article and approved the submitted version.

References

- Lindenbaum S. Cannibalism. In: Stein F, editor. *The Cambridge encyclopedia of anthropology*. Cambridge: Cambridge University Press (2021). doi: 10.29164/21cannibalism
- Culot L, Liedo-Ferrer Y, Hoelscher O, Munoz Lazo FJJ, Huynen MC, Heymann EW. Reproductive failure, possible maternal infanticide, and cannibalism in wild moustached tamarins, *Saguinus mystax*. *Primates* (2011) 52:179–86. doi: 10.1007/s10329-011-0238-6
- Crossland MR, Hearnden MN, Pizzato L, Alford RA, Shine R. Why be a cannibal? The benefits to cane toad, *Rhinella marina* [= *Bufo marinus*], tadpoles of consuming conspecific eggs. *Anim Behav* (2011) 82(4):775–82. doi: 10.1016/j.anbehav.2011.07.009
- Dorward LJ. New record of cannibalism in the common hippo, *Hippopotamus amphibius* (Linnaeus 1758). *Afr J Ecol* (2015) 53(3):385–7. doi: 10.1111/aje.12197
- Matzinger P. Tolerance, danger, and the extended family. *Annu Rev Immunol* (1994) 12:991–1045. doi: 10.1146/annurev.iy.12.040194.005015
- Casadevall A, Pirofski LA. The damage-response framework of microbial pathogenesis. *Nat Rev Microbiol* (2003) 1:17–24. doi: 10.1038/nrmicro732
- Skern-Mauritzen R, Mikkelsen TN. The information continuum model of evolution. *BioSystems* (2021) 209:104510. doi: 10.1016/j.biosystems.2021.104510
- Margulis L. Symbiosis and evolution. *Sci Am* (1971) 225:48–57. doi: 10.1038/scientificamerican0871-48
- Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell* (2014) 157(1):121–41. doi: 10.1016/j.cell.2014.03.011
- Devaux CA, Mezouar S, Mege J-L. The e-cadherin cleavage associated to pathogenic bacteria infections can favor bacterial invasion and transmigration, dysregulation of the immune response and cancer induction in humans. *Front Microbiol* (2019) 10:2598. doi: 10.3389/fmicb.2019.02598
- Khan AA, Yurkovetskiy L, O’Grady K, Pickard JM, de Pooter R, Antonopoulos DA, et al. Polymorphic immune mechanisms regulate commensal repertoire. *Cell Rep* (2019) 29:541–50. doi: 10.1016/j.celrep.2019.09.010

Funding

This work was supported by the French Government under the « Investissements d’avenir » (Investments for the Future) programme managed by the Agence Nationale de la Recherche (ANR, FR: National Agency for Research), (reference: Méditerranée Infection 10-IAHU-03) by DR and the annual budget allocation from Aix-Marseille Université and IRD to the MEPHI laboratory.

Acknowledgments

The Figure 3 was designed using the Servier Medical Art supply of images available under a Creative Commons CC BY 3.0 license.

Conflict of interest

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

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

12. Cazares A, Figueroa W, Cazares D. Diversity of microbial defence systems. *Nat Rev Microbiol* (2022) 20:191. doi: 10.1038/s41579-022-00702-3
13. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* (2018) 359:eaar4120. doi: 10.1126/science.aar4120
14. Zhang T, Tamman H, Wallant KC, Kurata T, LeRoux M, Srikant S, et al. Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* (2022) 612:132–140. doi: 10.1038/s41586-022-05444-z
15. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Com* (2022) 13:2561. doi: 10.1038/s41467-022-30269-9
16. De Parseval N, Heidmann T. Human endogenous retroviruses: from infectious elements to human genes. *Cytogenet Genome Res* (2005) 110(1-4):318–32. doi: 10.1159/000084964
17. Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet* (2006) 7:149–73. doi: 10.1146/annurev.genom.7.080505.115700
18. Kojima S, Yoshikawa K, Ito J, Nakagawa S, Parrish NF, Horie M, et al. Virus-like insertions with sequence signatures similar to those of endogenous nonretroviral RNA viruses in the human genome. *Proc Natl Acad Sci USA* (2021) 118(5):e2010758118. doi: 10.1073/pnas.2010758118
19. Jaenisch R. Endogenous retroviruses. *Cell* (1983) 32:5–6. doi: 10.1016/0092-8674(83)90491-9
20. Belshaw R, Katzourakis A, Paces J, Burt A, Tristem M. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol Biol Evol* (2005) 22:814–7. doi: 10.1093/molbev/msi088
21. Steinhuber S, Brack M, Hunsmann G, Schwelberger H, Dierich MP, Vogetseder W. Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. *Hum Genet* (1995) 96:188–92. doi: 10.1007/BF00207377
22. Grandi N, Cadeddu M, Blomberg J, Mayer J, Tramontano E. HERV-W group evolutionary history in non-human primates: Characterization of ERV-W orthologs in catarrhini and related ERV groups in platyrrhini. *BMC Evol Biol* (2018) 18:6. doi: 10.1186/s12862-018-1125-1
23. Mojica FJM, Rodriguez-Valera F. The discovery of CRISPR in archaea and bacteria. *FEBS J* (2016) 283:3162–9. doi: 10.1111/febs.13766
24. Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-cas systems. *Curr Opin Microbiol* (2017) 37:67–78. doi: 10.1016/j.mib.2017.05.008
25. Chaudhuri A, Halder K, Datta A. Classification of CRISPR/Cas system and its application in tomato breeding. *Theoret Appl Genet* (2022) 135:367–87. doi: 10.1007/s00122-021-03984-y
26. Mohanraju P, Saha C, van Baaren P, Louwen R, Staals RHJ, van der Oost J. Alternative functions of CRISPR-cas systems in the evolutionary arms race. *Nat Rev Microbiol* (2022) 20:351363. doi: 10.1038/s41579-021-00663-z
27. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* (2010) 327:167. doi: 10.1126/science.1179555
28. Kamruzzaman M, Yan A, Castro-Escarpulli G. Editorial: CRISPR-cas systems in bacteria and archaea. *Front Microbiol* (2022) . 13:887778. doi: 10.3389/fmicb.2022.887778
29. Makarova KS, Aravind L, Wolf YI, Koonin EV. Unification of cas protein families and a simple scenario for the origin and evolution of CRISPR-cas systems. *Biol Direct* (2011) 6:38. doi: 10.1186/1745-6150-6-38
30. Rollie C, Chevaliereau A, Watson BNJ, Chyou TY, Fradet O, McLeod I, et al. Targeting of temperate phages drives loss of type I CRISPR-cas systems. *Nature* (2020) 578(7793):149–53. doi: 10.1038/s41586-020-1936-2
31. Levasseur A, Bekliz M, Chabrière E, Pontarotti P, La Scola B, Raoult D. MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature* (2016) 513:249–52. doi: 10.1038/nature17146
32. Raoult D. The 'Eat me cake' theory, or genetic cannibalism of the enemy: A cause of vanishing antimicrobial resistance. *Int J Antimicrob Agents* (2018) 52(4):441–442. doi: 10.1016/j.ijantimicag.2018.08.018ff
33. Claverie JM, Abergel C. CRISPR-cas-like system in giant viruses: why MIMIVIRE is not likely to be an adaptive immune system. *Virologica Sin* (2016) 31(3):193–6. doi: 10.1007/s12250-016-3801-x
34. Takeuchi O, Akira S. Pattern recognition receptors and inflammation. *Cell* (2010) 140:805–20. doi: 10.1016/j.cell.2010.01.022
35. Hozumi N, Tonegawa S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci USA* (1976) 73(10):3628–32. doi: 10.1073/pnas.73.10.3628
36. Tonegawa S. Somatic generation of antibody diversity. *Nature* (1983) 302:575–81. doi: 10.1038/302575a0
37. Hedrick SM, Cohen DI, Nielsen EA, Davis MM. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *J Immunol* (1984) 308:149–53. doi: 10.1038/308149a0
38. Malissen M, McCoy C, Blanc D, Trucy J, Devaux C, Schmitt-Verhulst AM, et al. Direct evidence of chromosomal inversion during T-cell receptor rearrangements. *Nature* (1986) 319(6048):28–33. doi: 10.1038/319028a0
39. Litman GW, Anderson MK, Rast JP. Evolution of antigen binding receptors. *Annu Rev Immunol* (1999) 17:109–47. doi: 10.1146/annurev.immunol.17.1.109
40. Flajnik MF, Kasahara M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* (2001) 15:351–62. doi: 10.1016/S1074-7613(01)00198-4
41. Cooper MD, Alder MN. The evolution of adaptive immune systems. *Cell* (2006) 124:815–22. doi: 10.1016/j.cell.2006.02.001
42. Smith NC, Rise ML, Christian SL. A comparison of the innate and adaptive immune systems in cartilaginous fish, ray-finned fish, and lobe-finned fish. *Front Immunol* (2019) 10:2292. doi: 10.3389/fimmu.2019.02292
43. Chylinski K, Makarova KS, Charpentier E, Koonin EV. Classification and evolution of type II CRISPR-cas systems. *Nucleic Acids Res* (2014) 42(10):6091–105. doi: 10.1093/nar/gku241
44. Krupovic M, Makarova KS, Forterre P, Prangisvili D, Koonin EV. Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-cas immunity. *BMC Biol* (2014) 12:36. doi: 10.1186/1741-7007-12-36
45. Etchegaray E, Naville M, Volff JN, Haftek-Terreau Z. Transposable element-derived sequences in vertebrate development. *Mobile DNA* (2021) 12:1. doi: 10.1186/s13100-020-00229-5
46. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* (2007) 8:973–82. doi: 10.1038/nrg2165
47. Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in repbase. *Nat Rev Genet* (2008) 9:411–2. doi: 10.1038/nrg2165-cl
48. Dreyfus DH. Paleo-immunology: evidence consistent with insertion of a primordial herpes virus-like element in the origins of acquired immunity. *PLoS One* (2009) 4:e5778. doi: 10.1371/journal.pone.0005778
49. Zhang Y, Xu K, Deng A, Fu X, Xu A, Liu X. An amphioxus RAG1-like DNA fragment encodes a functional central domain of vertebrate core RAG1. *Proc Natl Acad Sci USA* (2014) 111:397–402. doi: 10.1073/pnas.1318843111
50. Huang S, Chen Z, Yan X, Yu T, Huang G, Yan Q, et al. Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat Commun* (2014) 5:5896. doi: 10.1038/ncomms6896
51. Kapitonov VV, Koonin EV. Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biol Direct* (2015) 10:20. doi: 10.1186/s13062-015-0055-8
52. Zhang Y, Cheng TC, Huang G, Lu Q, Surleac MD, Mandell JD, et al. Transposon molecular domestication and the evolution of the RAG recombinase. *Nature* (2019) 569(7754):79–84. doi: 10.1038/s41586-019-1093-7
53. Yakovenko I, Agronin J, Smith LC. And Oren m guardian of the genome: An alternative RAG/Transib Co-evolution hypothesis for the origin of V(D)J recombination. *Front Immunol* (2021) 12:709165. doi: 10.3389/fimmu.2021.709165
54. Soulas-Sprauel P, Rivera-Munoz P, Malivert L, Le Guyader G, Abramowski V, Revy P, et al. V(D)J and immunoglobulin class switch recombinations: A paradigm to study the regulation of DNA end-joining. *Oncogene* (2007) 26:7780–91. doi: 10.1038/sj.onc.1210875
55. Karim MF, Liu S, Laciak AR, Volk L, Koszelak-Rosenblum M, Lieber MR, et al. Structural analysis of the catalytic domain of Artemis endonuclease/SNM1C reveals distinct structural features. *J Biol Chem* (2020) 295:12368–77. doi: 10.1074/jbc.RA120.014136
56. Gan T, Wang Y, Liu Y, Schatz DG, Hu J. RAG2 abolishes RAG1 aggregation to facilitate V(D)J recombination. *Cell Rep* (2021) 37(2):109824. doi: 10.1016/j.celrep.2021.109824
57. Moran NA. Microbial minimalism: Genome reduction in bacterial pathogens. *Cell* (2002) 108(5):583–6. doi: 10.1016/S0092-8674(02)00665-7
58. Devaux CA, Raoult D. The microbiological memory, an epigenetic regulator governing the balance between good health and metabolic disorders. *Front Microbiol* (2018) 9:1379. doi: 10.3389/fmicb.2018.01379
59. Wagner DL, Peter L, Schmuck-Henneresse M. Cas9-directed immune tolerance in humans—a model to evaluate regulatory T cells in gene therapy? *Gene Ther* (2021) 28:549–59. doi: 10.1038/s41434-021-00232-2
60. Ratcliffe MJH, Härtle S. B cells, the bursa of fabricius and the generation of antibody repertoires. In: *Avian immunology*. Toronto, Canada: Elsevier (2014). p. 65–89. doi: 10.1016/B978-0-12-396965-1.00004-2

61. Garcia P, Wang Y, Viallet J, Macek Jilkova Z. The chicken embryo model: A novel and relevant model for immune-based studies. *Front Immunol* (2021) 12:791081. doi: 10.3389/fimmu.2021.791081
62. McCormack WT, Carlson LM, Tjoelker LW, Thompson CB. Evolutionary comparison of the avian IgL locus: combinatorial diversity plays a role in the generation of the antibody repertoire in some avian species. *Int Immunol* (1989) 1(4):332–41. doi: 10.1093/intimm/1.4.332
63. Carlson LM, Oettinger MA, Schatz DG, Masteller EL, Hurley EA, McCormack WT, et al. Selective expression of RAG-2 in chicken B cells undergoing immunoglobulin gene conversion. *Cell* (1991) 64:201–8. doi: 10.1016/0092-8674(91)90221-j
64. Bernstein RM, Schluter SF, Berstein H, Marchalonis JJ. Primordial emergence of the recombination activating gene 1 (RAG1): Sequence of the complete shark gene indicates homology to microbial integrases. *Proc Natl Acad Sci USA* (1996) 93:9454–9. doi: 10.1073/pnas.93.18.9454
65. Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. An ancient evolutionary origin of the Rag1/2 gene locus. *Proc Natl Acad Sci USA* (2006) 103:3728–33. doi: 10.1073/pnas.0509720103
66. Morales Poole JR, Pagani J, Pontarotti P. Convergent evolution of the adaptive immune response in jawed vertebrates and cyclostomes: An evolutionary biology approach based study. *Dev Comp Immunol* (2017a) 75:120–6. doi: 10.1016/j.dci.2017.02.011
67. Nelson PN, Hooley P, Roden D, Ejtehadi D, Rylance P, Warren P, et al. Human endogenous retroviruses: transposable elements with potential? *Clin Exp Immunol* (2004) 138:1–9. doi: 10.1111/j.1365-2249.2004.02592.x
68. Böhne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff JN. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res* (2008) 16:203–15. doi: 10.1007/s10577-007-1202-6
69. Porubsky D, Höps W, Ashraf H, Hsieh PH, Rodriguez-Martin B, Yilmaz F, et al. Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* (2022) 185(11):1986–2005. doi: 10.1016/j.cell.2022.04.017
70. Robertson HM, Lamp DJ. Recent horizontal transfer of a mariner transposable element among and between diptera and neuroptera. *Mol Biol Evol* (1995) 12(5):850–62. doi: 10.1093/oxfordjournals.molbev.a040262
71. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. Molecular reconstruction of *Sleeping beauty*, a *Tc1*-like Transposon from Fish, and Its Transposition in Human Cells. *Cell* (1997) 91(4):501–10. doi: 10.1016/S0092-8674(00)80436-5
72. Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* (2010) 25(9):537–46. doi: 10.1016/j.tree.2010.06.001
73. Gladyshev EA, Arkhipova IR. A single-copy IS5-like transposon in the genome of a bdelloid rotifer. *Mol Biol Evol* (2009) 8:1921–99. doi: 10.1093/molbev/msp104
74. Feschotte C. *Merlin*, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS 1016 insertion sequences. *Mol Biol Evol* (2004) 21:1769–80. doi: 10.1093/molbev/msh188
75. Piskurek O, Okada N. Poxviruses as possible vectors for horizontal transfer of retrotransposons from reptiles to mammals. *Proc Natl Acad Sci USA* (2007) 104:12046–51. doi: 10.1073/pnas.0700531104
76. Malik HS, Henikoff S, Eickbush TH. Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* (2000) 10:1307–18. doi: 10.1101/gr.145000
77. Irwin NAT, Pitis AA, Richards TA, Keeling PJ. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat Microbiol* (2022) 7:327–36. doi: 10.1038/s41564-021-01026-3
78. Morales Poole JR, Huang SF, Xu A, Bayet J, Pontarotti P. The RAG transposon is active through the deuterostome evolution and domesticated in jawed vertebrates. *Immunogenetics* (2017b) 69(6):391–400. doi: 10.1007/s00251-017-0979-5
79. Martin EC, Vicari C, Tsakou-Ngouafo L, Pontarotti P, Petrescu AJ, Schatz DG. Identification of RAG-like transposons in protostomes suggests their ancient bilaterian origin. *Mobile DNA* (2020) 11:17. doi: 10.1186/s13100-020-00214-y
80. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* (2000) 403(6771):785–9. doi: 10.1038/35001608
81. Blaise S, de Parseval N, Béné L, Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci USA* (2003) 100:13013–8. doi: 10.1073/pnas.2132646100
82. Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, et al. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci* (2013) 368(1626):20120507. doi: 10.1098/rstb.2012.0507
83. Chuong EB. The placenta goes viral: Retroviruses control gene expression in pregnancy. *the placenta goes viral: Retroviruses control gene expression in pregnancy. PLoS Biol* (2018) 16(10):e300028. doi: 10.1371/journal.pbio.300028
84. Casseb J, Janini LM, Barros LI, Luciano K, Lopes R, Paiva AM. Is the human T-cell lymphotropic virus type 2 in the process of endogenization into the human genome? *J Virus Erad* (2020) 6(4):100009. doi: 10.1016/j.jve.2020.100009
85. Casseb J, Lopes LR. Reflection about the ancient emergence of HTLV-2 infection. *AIDS Res Hum Retrovir* (2022). doi: 10.1089/aid.2022.0019
86. Colson P, Ravaux I, Tamalet C, Glanzova O, Baptiste E, Chabrière E, et al. HIV Infection en route to endogenization: two cases. *Clin Microbiol Infect* (2014) 20(12):1280–8. doi: 10.1111/1469-0691.12807
87. Mahé D, Matsuali G, Deleage C, Alvarenga RLLS, Satie AP, Pagliuzza A. Potential for virus endogenization in humans through testicular germ cell infection: the case of HIV. *J Virol* (2020) 94(24):e01145–20. doi: 10.1128/JVI.01145-20
88. Tarlinton RE, Meers J, Young PR. Retroviral invasion of the koala genome. *Nature* (2006) 442:79–81. doi: 10.1038/nature04841
89. Denner J, Young PR. Koala retroviruses: characterization and impact on the life of koalas. *Retrovirology* (2013) 10:108. doi: 10.1186/1742-4690-10-108
90. Zheng HQ, Pan Y, Tang S, Pye GW, Stadler CK, Vogelneist L, et al. Koala retrovirus diversity, transmissibility, and disease associations. *Retrovirology* (2020) 17:34. doi: 10.1186/s12977-020-00541-1
91. Xu W, Stadler CK, Gorman K, Jensen N, Kim D, Zheng H, et al. An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proc Natl Acad Sci USA* (2013) 110(28):11547–52. doi: 10.1073/pnas.1304704110
92. Quigley BL, Ong VA, Hanger J, Timms P. Molecular dynamics and mode of transmission of koala retrovirus as it invades and spreads through a wild Queensland koala population. *J Virol* (2018) 92(5):e01871–17. doi: 10.1128/JVI.01871-17
93. Stocking C, Kozak CA. Murine endogenous retroviruses. *Cell Mol Life Sci* (2008) 65(21):3383–98. doi: 10.1007/s00018-008-8497-0
94. Lemos de Matos A, de Sousa-Pereira P, Lissovsky AA, van der Loo W, Melo-Ferreira J, Cui J, et al. Endogenization of mouse mammary tumor virus (MMTV)-like elements in genomes of pikas (*Ochotona* Sp.). *Virus Res* (2015) 210:22–6. doi: 10.1016/j.virusres.2015.06.021
95. De Martini JC, Carlson JO, Leroux C, Spencer T, Palmarini M. Endogenous retroviruses related to jaagsiekte sheep retrovirus. *Curr Top Microbiol Immunol* (2003) 275:117–37. doi: 10.1007/978-3-642-55638-8_5
96. Huang AS, Baltimore D. Defective viral particles and viral disease processes. *Nature* (1970) 226(5243):325–7. doi: 10.1038/226325a0
97. Molenkamp R, Rozier BC, Greve S, Spaan WJ, Snijder EJ. Isolation and characterization of an arterivirus defective interfering RNA genome. *J Virol* (2000) 74(7):3156–65. doi: 10.1128/jvi.74.7.3156-3165.2000
98. Makino S, Yokomori K, Lai MM. Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. *J Virol* (1990) 64:6045–53. doi: 10.1128/jvi.64.12.6045-6053.1990
99. Snijder EJ, den Boon JA, Horzinek MC, Spaan WJ. Characterization of defective interfering RNAs of Berne virus. *J Gen Virol* (1991) 72(Pt7):1635–43. doi: 10.1099/0022-1317-72-7-1635
100. Vignuzzi M, Lopez CB. Defective viral genomes are key drivers of the virus–host interaction. *Nat Microbiol* (2019) 4:1075–87. doi: 10.1038/s41564-019-0465-Y
101. Xu J, Sun Y, Li Y, Ruthel G, Weiss SR, Raj A, et al. Replication defective viral genomes exploit a cellular pro-survival mechanism to establish paramyxovirus persistence. *Nat Commun* (2017) 8(1):799. doi: 10.1038/s41467-017-00909-6
102. Manzoni TB, Lopez CB. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. *Future Virol* (2018) 13(7):493–504. doi: 10.2217/fvl-2018-0021
103. Tamiya S, Matsuoka M, Etoh K, Watanabe T, Kamihira S, Yamaguchi K, et al. Two types of defective human T-lymphotropic virus type I provirus in adult T-cell leukemia. *Blood* (1996) 88(8):3065–73. doi: 10.1182/blood.V88.8.3065.bloodjournal8883065
104. Mesnard JM, Barbeau B, Devaux CA. HBZ, a new important player in the mystery of adult T-cell leukemia. *Blood* (2006) 108:3979–82. doi: 10.1182/blood-2006-03-007732
105. Bernier R, Tremblay M. Homologous interference resulting from the presence of defective particles of human immunodeficiency virus type 1. *J Virol* (1995) 69(1):291–300. doi: 10.1128/jvi.69.1.291-300.1995
106. Monde K, Terasawa H, Nakano Y. Molecular mechanisms by which HERV-K gag interferes with HIV-1 gag assembly and particle infectivity. *Retrovirology* (2017) 14:27. doi: 10.1186/s12977-017-0351-8
107. Ward KN, Leong HN, Nacheva EP, Howard J, Atkinson CE, Davies NWS, et al. Human herpesvirus 6 chromosomal integration in immunocompetent

patients results in high levels of viral DNA in blood, sera, and hair follicles. *J Clin Microbiol* (2006) 44(4):1571–4. doi: 10.1128/JCM.44.4.1571-1574.2006

108. Liu X, Kosugi S, Koide R, Kawamura Y, Ito J, Miura H, et al. Endogenization and excision of human herpesvirus 6 in human genomes. *PLoS Genet* (2020) 16(8):e1008915. doi: 10.1371/journal.pgen.1008915

109. Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, et al. Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci USA* (1990) 87:2211–5. doi: 10.1073/pnas.87.6.2211

110. Weitzman MD, Kyostio SRM, Kotin RM, Owens RA. Adeno-associated virus (AAV) rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc Natl Acad Sci USA* (1994) 91:5808–12. doi: 10.1073/pnas.87.6.2211

111. Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, et al. Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. *J Virol* (2011) 85(19):9863–76. doi: 10.1128/JVI.00828-11

112. Koonin EV, Krupovic M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet* (2015) 16:184–92. doi: 10.1038/nrg3859

113. Vizcaino JA, Kubiniok P, Kovalchik KA, Ma Q, Duquette JD, Mongrain I. The human immunopeptidome project: A roadmap to predict and treat immune diseases. *Mol Cell Proteomics* (2020) 19(1):31–49. doi: 10.1074/mcp.R119.001743

114. Pontarotti P, Abi-Rached L, Yeh JH, Paganini J. Self-peptidome variation shapes individual immune responses. *Trends Genet* (2021) 37(5):414–420. doi: 10.1016/j.tig.2020.10.001

115. Saini SK, Due Orskov A, Bjerregaard AM, Unnikrishnan A, Holmberg-Thyden S, Borch A, et al. Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers. *Nat Com* (2020) 11:5660. doi: 10.1038/s41467-020-19464-8

116. Mustelin T, Ukadike KC. How retroviruses and retrotransposons in our genome may contribute to autoimmunity in rheumatological conditions. *Front Immunol* (2020) 11:593891. doi: 10.3389/fimmu.2020.593891

117. Müller MD, Holst PJ, Nielsen KKN. A systematic review of expression and immunogenicity of human endogenous retroviral proteins in cancer and discussion of therapeutic approaches. *Int J Mol Sci* (2022) 23:1330. doi: 10.3390/ijms23031330