ISSN 0891-4168, Molecular Genetics, Microbiology and Virology, 2018, Vol. 33, No. 4, pp. 223–232. © Allerton Press, Inc., 2018.
Original Russian Text © R.N. Mustafin, 2018, published in Molekulyarnaya Genetika, Mikrobiologiya i Virusologiya, 2018, No. 4, pp. 182–190.

REVIEWS

Hypothesis on the Origin of Viruses from Transposons

R. N. Mustafin[a, b, *, **]

[a] *Bashkir State Medical University, Ufa, 450008 Russia*
[b] *Bashkir State University, Ufa, 450076 Russia*
**e-mail: ruji79@mail.ru*
***e-mail: rectorat@bashgmu.ru*

Received February 18, 2018; revised February 18, 2018; accepted April 7, 2018

Abstract—In this review, the role of transposons in the origin of viruses, losing the traits of their evolutionary precursors due to their high mutability, is considered. However, there are a number of common properties of viruses and transposons suggesting their phylogenetic relationship, including the ability to integrate into the host genome, specific activation in certain tissues, high degree of mutability, the existence of virophages propagating only in the presence of another virus (which is similar to non-autonomous transposons, for which the expression products of autonomous ones are required). Ideas about the emergence of viruses from transposons in evolution are discussed. The genomic elements exhibiting a dual nature of existence as mobile genetic elements and viruses are found: polintons, Tlr elements, and PLV viruses. It is assumed that a horizontal transfer (HT) of transposons, during the natural selection of which the elements possessing the virulence properties are preserved (while the genes required for the integration can mutate), is a key event required for turning into viruses. The horizontal transposon transfer, which is common in all representatives of living organisms, is accompanied by their variability required for the acquisition of new adaptive traits. In the course of evolution, the mechanisms of protection against viruses and transposons, including RNA interference, DNA methylation, and histone modification, began to be used to control the operation of the genomes, providing intercellular interactions; this explains the emergence of multicellular organisms. In the evolution of eukaryotes, the transposons have been used successfully for transformations of regulatory gene networks, as well as possible sources of new genes encoding both proteins and non-coding RNA, the fragmentary translation of which can produce short functional peptides. Thus, the products of transposon transcription and translation are the most important sources of evolutionary transformations; these mechanisms could be the basis for the evolution of viruses and the emergence of the fundamental properties of living organisms when they appear.

Keywords: protein-coding genes (PCG), horizontal transfer (HT), review, regulation, transposable elements (TE), evolution

DOI: 10.3103/S0891416818040067

INTRODUCTION

Transposons (transposable elements, TEs) are mobile genetic elements detectable in all living organisms. They contribute to the regulation of gene expression during their development and adaptation, as well as being the main sources of genetic variations in the evolution of genomes. For example, most primate-specific sequences regulating the gene expression originated from TEs [1]. They constitute a significant portion of most genomes of multicellular eukaryotes: 15–47% in insects, 35–69% in mammals, and up to 90% in plants [2]. In prokaryotes, TEs provide an opportunity for some genes of the basic set to move into the auxiliary category during evolution; and auxiliary ones, to the basic category [3]. In addition, mobile elements of bacteria with their movements inactivate the host genes or change their regulation, as well as induce all types of chromosomal rearrangements and provide the gene transfer between different individuals [4].

TEs form 2 classes: 1, retrotransposons; and 2, DNA transposons. Retro-TEs are divided into those carrying long terminal repeats (LTRs) (retroviruses: Ty1/Copia, Ty3/Gypsy, Bel/Pao, Dirs) and containing no LTR (nonLTR) retrotransposons: LINE and SINE [5]. Retro-TEs move to a new locus leaving the initial copy in the same place. The identity of the generated copies contribute to homologous recombination between TEs, which induces intragenomic chromosomal rearrangements (deletions, translocations, inversions, segmental duplications) and intragenic mutations (when introduced into the gene), which is considered as an engine of evolution [6]. DNA-TEs move by cutting out the original copy. Although it was demonstrated for some prokaryotic DNA-TEs (for example, Tn1, IS1, IS2, and IS4) that their transpositions cannot be associated with the exclusion from the places of their initial localization in plasmids or chromosome (replicative transpositions) [4].

Two subclasses are distinguished among DNA-TEs: cut/paste elements and helitrons (rolling circle elements) [2]. The cut/paste elements move using transposase (recombinase class enzyme) and the new place of integration is usually located close to the old place [7]. They contain terminal inverted repeats (TIRs) flanked by target site duplications (TSDs) of different lengths and compositions. The latter are sensitive to transposases, depending on the enzymatic properties of which DNA-TEs are attributed to different superfamilies. Thus, 5'-TA-3' TSD are typical for *Tc1/Mariner*; 5'-TTAA-3', for *piggiBack*; 8 bp TSD, for *hobo-Ac-Tam3*; 3 bp, for *PIF/Harbinger*; and 9–12 bp, for *Mutator/MuDR*. A new DNA-TE group (*Spy*), the members of which do not create TSDs during insertion was also detected. Instead of this, the Spy transpose exactly between the host 5'-AAA and TTT-3' nucleotides without duplication or modification of the target AAATTT sites [2]. In addition, non-autonomous DNA-TEs are distinguished. They include miniature inverted repeats (MITEs) and some of them contain terminal repeats that tend to be located in tandem arrays; in a number of cases, these arrays lead to the emergence of satellite DNA in the peri- and centromeric regions of the chromosomes [8]. The RelocaTE, RetroSeq, TEMP, TIF, and ITIS (Identification of Transposon Insertion Sites) algorithm systems were developed and implemented to determine the TE integration sites [9].

TEs constitute a much smaller portion in prokaryotic genomes, since the horizontal transfer (HT) of the genes by means of plasmids, integrons, and bacteriophages has the greatest significance in their evolution. Both DNA-TEs and retro-TEs are found in prokaryotes. The latter form the 17 main groups in bacteria: group II introns, retrons, diversity-generating retroelements (DGRs), Abi-like, CRISPR-Cas-associated, and group II-like retro-TEs, as well as 11 other groups of retroelements with unknown functions. Among them, DGRs, abortive-bacteriophage infections (Abis), and clustered regularly interspaced short palindromic repeats (CRISPRs) have the properties of protection against foreign elements [10]. Despite the efficiency of these protective systems, HT in prokaryotes is the main source of evolutionary variability. For eukaryotes, TEs gained importance in the transformation of genomes, constituting a significant part of their genomes; viruses are most probably the products of the evolution of TEs, when they acquire the ability to transfer horizontally. The diversity and distribution of viruses in eukaryotes indicates the significance of TEs in their origin. Subsequently, viruses lose the traits of an evolutionary relationship with their precursors due to their pronounced mutability.

INTERCONVERSIONS OF VIRUSES AND TRANSPOSONS IN EVOLUTION

The distribution and composition of TEs and viruses are specific for different domains of the living, which suggests their interconversions in evolution, in which TEs are sources of viruses. Thus, DNA viruses constitute a larger portion in prokaryotes, and DNA transposons are also the most common TEs. The prevalence of RNA viruses is typical for eukaryotes, in the genomes of which retro-TEs are prevalent. TEs can be autonomous and non-autonomous; the latter propagate by means of the protein products of the former. The same properties are also typical for viruses; there are satellite viruses (virophages) reproducing by means of the factors of larger viruses [11].

It is assumed that exogenous retroviruses were formed from Ty3/Gypsy LTR-containing retro-TEs (their sequences are detected in the genomes of most living organisms) [12]. With regard to the origin of other viruses, there are no precise data proving their origin. The logic of the emergence of viruses in evolution to transmit information of adapting properties leads to the conclusion about their origin from TEs due to their high level of mutability. RNA viruses could emerge from retro-TEs, while DNA viruses could emerge from DNA-TEs. At the same time, the enzymes used for integration in the host genome could acquire other properties required for autonomous reproduction under the influence of a number of mutations. Due to this, originated from TE exogenous viruses lost the properties of integration in the host genome simultaneously with the acquisition of virulence and transmission to other cells and organisms. Due to mutations, many viruses finally lost the traits of TEs. At present, the intermediate evolutionary niches demonstrating the mechanisms of the emergence of viruses from TEs are found: the HT of TEs [13] and transmission of domesticated retro-TEs (emerged from Ty3/gypsy) in capsids from endogenous Arc proteins (similar to retroviral Gag) between neurons [14]. Further mutations of these intermediate forms can cause the formation of exogenous viruses containing no TE sequences in their composition; this is a logical completion of a series of their changes with the formation of new virus taxa characterized by a high level of diversity.

The emergence of viruses in eukaryotes is a side product of the evolutionary transformations of genomes using TEs. The ability of viruses that contain no genes encoding the enzymes required for insertion to integrate into the host genome supports this assumption. For example, the integration of an entire viral genome is possible in hepatitis B [15], a part of the genome is usually integrated in adenoviral and herpes viral infections, both the complete genome and its part can be integrated into the infection with oncoviruses [16].The tropism of viruses to certain tissues and integration in the genomes of their cells also confirm their origin from TEs, since the latter are characterized by the activation in specific tissues arranging the gene networks to perform tissue-specific functions by the cells [17–19]. In addition, TE genes and sequences have a high potential to form with their mutations the genes of functional RNA and their pro-

tein products, which is important for the evolution of viruses. The detection that TEs are the most important sources of non-coding RNA (ncRNA) genes [20, 21], regulatory sequences [22], and protein-coding genes (PCGs) of their hosts [22–24] is a confirmation of this potential.

Both viruses and TEs have an increased level of mutability. The endogenous retrovirus (ERV) sequences are susceptible to disorganization by the formation of mutations with the shift of the reading frame, stop codons, and deletions that affect their ability for replication and transposition [25]. Due to their enhanced mutability, many TEs lose their initial properties, including the ability for transpositions, and are co-opted by the hosts to form new adaptive traits. For example, many genes derived from TEs were detected by means of TE-specific domains; some of them form tandem clusters of the gene families [24]. The ERV PCG also provide the genetic material co-opted by the hosts. For example, the *Syncytin* genes in mammals is the best representation of the proven ERV gene product's domestication for cellular functions. The *Syncytin* genes are co-opted from the ERV *envelope* genes independently in different mammalian lines performing a physiological function in the placenta [26]. Numerous data about the origin of eukaryotic genes by TE domestication were obtained; they were described in lengthy reviews [22, 23]. In addition to domestication, TEs embedded in introns can be used for the emergence of new domains of existing PCGs by exonization. For example, the domains of LAP2α splicing isoforms in the *TMPO* and *ZNF451* genes typical for all vertebrates originated from ORF1 of the *DIRS1-like* retrotransposon. This splicing isoform *TMPO*(*LAP2α*) is used for new important cellular functions of the protein [27]. The exonization is associated with the ability of TE translation products to form DNA-binding domains and other spatial structures allowing the protein to acquire new functions. Besides intronic TEs, the use of insertions for the alternative 3'-UTR is a widespread phenomenon [22].

Cases of interconversion of viruses and TEs are described in the literature. The penetration of the *Myotis lucifugus* bat ERV in the genomes of more than 100 mammalian species was proven. Moreover, the same species of retroviruses move many times from the infectious pathogen to the genomic parasite (retro-TE) experiencing different dynamics of invasion in different hosts [28]. Viruses of the Caulimoviridae family of plants evolved from LTR-containing retro-TEs. Since these viruses are similar to retroviruses, they are united in the group of pararetroviruses that have a circular double-stranded DNA and are replicated in plants by means of intermediate RNA. There are species from this family that can integrate into the host genome and are designated as endogenous pararetroviruses (EPRV) [29]. In addition, TEs that are assumed to perform a function similar to RNA interference were found directly in viral genomes. Thus, the mobile element *s2m* was found in the composition of viruses from four families: Astroviridae, Caliciviridae, Picornaviridae, and Coronaviridae. This conservative 43 base pair of TE was for the first time found in 1997 at the 3'-end of viruses from the Astroviridae family [30]. In addition, many viruses like TEs also encode micro-RNA genes that play an important role in viral infection and their replication [31], which also indicates in favor of the origin of viruses from TEs. Virophages that according to genetic homology are an evolutionary link between double-stranded DNA viruses and eukaryotic DNA-TE *Maverick/Polinton* were also detected [32]. For example, rumen virophages (RVPs) encode the typical large protein of virophages, ATPase, and protease in combination with DNA polymerase typical for polintons. The RVP genomes represent a hybrid of virophages and polintons as a linear molecule with terminal inverted repeats, which is able to form infectious virions [33].

The *Polinton* family members are a good example of interconversions of TEs and viruses. Polintons inhabit the genomes of protists, fungi, and animals, including amoeba, *Phakopsora pachyrhizi*, hydra, sea anemones, nematodes, fruit flies, beetles, sea urchins, ascidians, fishes, lizards, frogs, and chickens. Similarly to known TEs, polintons exist in the form of autonomous and non-autonomous elements. It is assumed that they evolved from linear plasmids that acquired retroviral integrase approximately 1 billion years ago [34]. According to the results of phylogenetic analyses, it was demonstrated that polintons in evolution spread mostly vertically [35], although data about their HT were obtained [36]. The *Polinton* family is especially abundantly spread in the genomes of some protists. These self-synthesizing TEs encode their own DNA polymerase (DNAP), retrovirus-like integrase, capsid proteins, DNA packaging ATPase, and capsid maturation proteinase. Consequently, polintons are an alternative between the transposon and viral way of life. The comparative genomic analysis of polintons, virophages, polinton-like viruses (PLVs), and other viruses with a double-stranded DNA demonstrated that polintons could be ancestors of a wide spectrum of eukaryotic viruses, including adenoviruses and megaviruses [37]. In addition, the virus-like particles formed from the host RNA molecules (mRNA, rRNA, ncRNA, and TE) packed with the proteins of the RNA flock house virus (FHV) envelope were found in eukaryotes. The packaging of these host RNA allows a HT between the genes by eukaryotic genomes, which contain a viral pathogen [38].

Polintons encode two capsid proteins, which allows them to lead a double lifestyle as TEs and viruses [39]. Evolutionary relationships of polintons with bacterial tectiviruses and linear mitochondrial plasmids were also demonstrated [40]. In addition, PLV viruses have been recently detected; they are similar to polintons and virophages in terms of the genome size and conservative morphogenetic modules, except that PLVs

contain no retroviral type integrase [33]. This indicates the role of TEs in the origin of viruses that change their properties in the course of evolution due to the selection of new adaptive traits emerged as a result of mutations. For example, some PLVs contain tyrosine recombinase (integrase) common for bacteria and bacteriophages, while certain PLVs integrate into the genomes of algae, showing a dual nature of existence as TEs and viruses [33]. Some DNA-TEs, for example, Tlr elements found in the genome of the *Tetrahymena thermophila* infusoria germline, also have such a dualism [41].

Certain ERVs are able to form virus-like particles with the infectious activity. Phylogenetic analysis demonstrated that exogenous retroviruses emerged from LTR-containing retro-TEs as a result of the acquisition of the *env* gene, the protein product of which allowed the formation of an infectious viral particle [42]. The structural similarity of the hepatitis B virus with retroviruses, as well as its ability to integrate into the host genome, also suggests its origin from TEs [43]. The origin of viruses from TEs is also indicated by the way that transposon became widespread in a similar way as interspecific pandemics during a HT. It was detected that germinal infiltrations of the *SPIN* DNA-TE family that gave birth to one of the largest bursts of DNA-TE activity (almost 100000 *SPIN* copies per haploid genome), originated 15–46 million years ago. The consensus *SPIN* sequences are identical by 96% throughout their length (2.9 kb) in the genomes of mouse rodents (rats/mice), bush babies, small brown bat, tenrec, opossum, and non-mammalian quadrupeds (anoles lizard and the African clawed frog) [44]. TEs are also the most important sources for a change in the gene network's regulatory system. ERVs deposit an extensive reservoir of prefunctional latent *cis*-regulatory elements (promoters, transcription factor binding sites) that are recruited in evolution in the composition of the normal regulatory network of adjacent host genes [26]. Many promoters and polyadenylation signals originate from the species-specific TEs. The ability of TEs to self-regulate (form protein products interacting with specific DNA sequences, the sources of which are the transposons themselves) [22] is also evidence of the origin of viruses from TEs as a prerequisite to autonomous existence and the acquisition of new functions during evolution. In this respect, it is possible to draw an analogy with the genomes of eukaryotes, during the evolution of which strategies for the emergence of new functional RNA and the products of their translations were developed for the emergence of the host PCGs.

TRANSPOSON–NON-CODING RNA–PEPTIDE STRATEGY

TEs are abundant sources of tightly controlled biochemically active non-coding elements such as non-coding RNA (ncRNA) and transcription factor binding sites. Multiple recent studies confirmed the idea that TEs are co-opted to regulate the genes of their hosts in all members of a living organism, and the waves of TE invasions in the organisms catalyzed the evolution of gene regulation networks [45]. TEs are the most important sources of micro-RNA [20] and are found in more than 83% of the functional domains of long non-coding RNA (lncRNA) [21]. LncRNA exhibit many functions independently as ribozymes, as riboswitches, or as a part of ribonucleoproteins (RNPs). LncRNA can interact with DNA-binding proteins, RNA polymerase, and histone-modifying complexes affecting the regulation of genome transcription [46]. Similarly to the functions as a part of the products of PCG translation, TE sequences in lncRNA domains are used to form the protein and nucleic acid recognition sites. Most lncRNA exons are organized from TE sequences; therefore, mature lncRNA molecules contain a combination of repeats creating a structure similar to the protein domains. Phylogenetic analysis demonstrated that some TE families naturally enrich (mainly ERV) or leave (nonLTR retroelements) lncRNA genes during evolution. In addition to exons, TEs organize the structures of promoters, splicing sites, and polyadenylation sites of lncRNA. According to the studies by Johnson et al., TEs are the basis for the development of the modular organization of lncRNA molecules as discrete domains, the combinations of which determine the lncRNA function. This is caused by the ability of TE fragments to form DNA-, RNA-, and protein-binding domains. The use of a TE mapping to predict lncRNA functions and to explain the evolution of lncRNA regulatory networks is assumed [21]. In addition to performing specific functions at the level of RNA molecules, lncRNA are able to be translated to the peptides [47–50] that undergo evolutionary selection based on their functional suitability for adaptation and can become sources for the emergence of new PCGs [51–53]. A number of recent studies demonstrated that lncRNA can contain open reading frames (ORFs), bind to ribosomes, and be translated. Several hundred functional peptides formed in this way were detected in Danio fish and humans [54], in mice [55], and in Arabidopsis [56]. Moreover, many lncRNA have a structure similar to mRNA: they are transcribed by RNA polymerase II, caped, and polyadenylated accumulating in the cytoplasm [57]; the plant pri-microRNA have the same properties. Due to the presence of short ORFs, processing to obtain 5'-caped and 3'-polyadenylated tails [58], the plants' pri-microRNA are also able to be translated with the production of functional peptides [59]. Such strategies could be developed in viruses, explaining their great variety and prevalence due to the high level of mutability with the emergence of new coding sequences from genes originating from TEs.

In 2006, M. Levine et al. [60] detected the formation of new PCGs in drosophila due to the selection of possible variants of lncRNA expression. Researchers

used the genomes of *Drosophila melanogaster* and related species to conduct a genome-wide search for new *D. melanogaster* genes that emerged from the non-coding DNA. As a result, 5 such genes mainly expressed in the testicles were described. In 2008, J. Cai et al. [61] found a new PCG in the *Saccharomyces cerevisiae*, which was formed from a non-coding DNA sequence. In 2012, C. Xie et al. described 24 hominoid-specific PCGs generated de novo from lncRNA genes [62]. Further studies demonstrated that the emergence of new proteins from lncRNA is common in eukaryotes and plays an important role in evolution. For example, in 2014, when studying the *Mus musculus*, *Homo sapiens*, *Danio rerio*, *D. melanogaster*, *A. thaliana*, and *S. cerevisiae*, J. Ruiz-Orera et al. [51] found that most of the lncRNA expressed in the cells of six different species is associated with ribosomes. The researchers tested the hypothesis that lncRNA can act as a source for the synthesis of new peptides. Moreover, it was found that lncRNA showed a similar coding potential with evolutionary young protein-coding sequences, indicating that they play an important role in the de novo evolution of the proteins [51]. Recent genome-wide studies allowed to detect hundreds of functional micropeptides that can be generated during lncRNA translation [54, 63]. Thus, micropeptide mioregulin (MLN), which is formed during the translation of skeletal muscle-specific RNA annotated as lncRNA, was identified. This lncRNA in human contains 138 nucleotide ORFs and consists of 3 exons with a total length of 16.5 kb. The MLN micropeptide consists of 46 amino acid residues and forms a single transmembrane alpha-helix, which interacts with the sarcoplasmic reticulum Ca^{2+} ATPase (SERCA) on the membrane of the sarcoplasmic network and regulates the movement of calcium ions [48]. In addition, the micropeptide consisting of 34 amino acid residues called the dwarf open reading frame (DWORF), which is localized on the membrane of the sarcoplasmic network, where the activity of SERCA is enhanced. The *Dworf* gene transcript, DWORF micropeptide, is annotated as NONMMUG026737 lncRNA in mice and LOC10050737 lncRNA in humans. In the mouse genome, the *Dworf* gene is transcribed from a 2.8 kb locus on chromosome 3 with the formation of two transcription isoforms 300 nucleotides in length. In humans, the LOC10050737 lncRNA forms only a single isoform and is transcribed from the 4.5 kb region on chromosome 3 [50]. In mammals, the peptides generated by the lncRNA translation have a regulatory effect on the expression of the target genes [49]. In plants, a number of peptides developed during the lncRNA translation were also detected: ENOD40, IPS1, LDMAR, COOLAIR, and OLDAIR [47]. Moreover, lncRNA and pri-micro RNA are also able to be translated into functional peptides [49], while TEs are important sources of the emergence of microRNA genes [20]. Thus, the system TE–ncRNA–proteins is a universal strategy for the evolutionary transformations of eukaryotic genomes and can also be used by viruses that originated from TEs. This principle of functional RNA selection as a source of the protein can explain not only the origin of viruses and their evolution but also the origin of life, when the molecules of functional RNA, the translation products of which contributed to the self-regulation and the emergence of new functions, were selected from mobile elements for millions of years. The most optimal combinations of RNA and protein interactions, contributing to the optimal adaptation, persisted in the course of evolution.

HORIZONTAL TRANSPOSON TRANSFER

In 1984, R.B. Khesin [4] in the book "Inconstancy of the Genome" wrote that a significant portion of hereditary variability is caused by transposable elements. According to Khesin, the discovery and study of various TEs makes us to take a fresh look at a number of position of evolutionary theory. He assumed that parallel variability can be explained by the effect of the elements with similar site-specific integration, while some cases of convergence are based on the transfer of identical genes between different organisms. Moreover, the author explained some qualitative leaps in historical development (especially metabolism rearrangements) by the gene transfer between distant species [4]. The HT common in prokaryotes (81% of their genes are involved in HTs [64]) is also an important source of genomic diversity in eukaryotes. The HT of transposons should logically be present in eukaryotes due to the abundance of TEs in their genomes and their inherent TE mobility. According to the results of phylogenetic analyses, DNA-TE HT was found among vertebrates and invertebrates 45–15 million years ago. The events of the TEs' HT were described in different members of Tetrapoda, as well as blood sucking insects (*Rhodnius prolixus*), as a vector. In 2013, C. Gilbert et al. found the HT of the TEs of the OposCharlie1 (OC1) family, which penetrated into the common ancestor of the *Dasyuridae* family approximately 17 million years ago, in the Tasmanian devil [65]. A HT was for the first time demonstrated in 1990 by S. Daniels et al. on drosophila [66]. Subsequently, it was found in a number of works that almost all the main TE types are capable of HTs in a large number of eukaryotes [67]. The HT of TEs in new genomes is considered as an important strength of the control of genomic variations and biological innovations. In addition, HT plays an important role in the persistence of TEs in eukaryotic genomes. In the studies by H. Zhang et al. [67], the examples of the repeated HT of three *Chapaev* TE families in a large number of animal species, including mammals, reptiles, jaw fish, lampreys, and insects, were demonstrated. Multiple comparisons of *Chapaev* TEs in these species detected extremely high sequence identity levels (79–99%), which is incompatible with vertical evolution, taking into account the deep divergence in the

time separating these hosts. The discontinuous distribution between species and the absence of a purifying selection affecting these TEs indicate that they independently and horizontally transfer between species [67]. The prevalence of TEs in eukaryotes suggests this phenomenon as an intermediate evolutionary link in the emergence of viruses. Viral interspecific transmission is a serious threat to humans and animals. Most human viral diseases are zoonotic [28]. Interspecific transmission during the HT of TEs and viruses is a common property among them and supports the assumption of viruses originating from TEs.

The RAG1 and RAG2 genes form a recombinase complex required for V(D)J recombination, which generates a diversity of immunoglobulins and T-cell receptors [68]. V(D) recombination is a variation of the main transposition principle and is mediated by transposase descendants. The structure of inverted repeats and the right/left asymmetry of recombination signal sequences (RSSs) makes us recall terminal repeats of insertion sequences (ISs) in prokaryotes [69]; therefore, the *RAG1* gene was initially attributed to the integrase gene family (*INT*) in bacteria, while the *RAG2* gene was considered a homolog of the *IHF* gene, which regulates the work of integrase [70]. However, the *SpRag1L* and *SpRag2L* genes, homologous to the *RAG1* and *RAG2* genes in vertebrates by sequences and by genomic organization, were detected in further studies in the *Strongylocentrotus purpuratus* purple sea urchin [71]. The cutting and assembly activity similar to the *RAG* transposase was detected in vitro in the Transib transposase in the *Helicoverpa zea* insect, indicating their phylogenetic kinship and origin from a common ancestor [72]. It is generally accepted that the presence of RAG in the genomes of jaw vertebrates and other lines is a result of a horizontal TE transfer. In all jaw vertebrates, the *RAG* genes are located close to the *NWC* gene, the promoter of which exhibits a bidirectional activity, which contributes to the preservation of RAG-TE in the host genome. This is caused by the fact that the *NWC* gene is evolutionarily conservative and is located upstream from the *RAG2* gene, convergently transcribing with it. The *NWC* gene plays an important role in the ontogenesis of multicellular organisms, which is indicated by the presence of its orthologs in a number of invertebrate species such as sea urchin (*Strongylocentrotus purpuratus*), sea anemone (*Nematostella vectensis*), starfish (*Asterina pectininera*), sea mussel (*Mytilus edulis*), sea snail (*Lottia gigantean*), stony coral (*Acropora millepora*), and lamellar *Trichoplax adhaerens*. Due to the important role of the *NWC* gene and its evolutionary conservatism, the successful insertion of the *RAG* genes close by contributed to their preservation in the selection process and use for the hosts' needs [68]. The *AdLINE3* retro-TE, which turned out to be a member of the RTE cluster initially identified in nematodes, was detected in the genome of wild peanuts. The RTE elements were found in 82 plant species (from angiosperms to algae). In flowering plants, RTEs were obtained from the An-RTE family showing a significant identity of DNA sequences with the retro-TE of 42 animal species from 4 phylums. The phylogenetic analysis of animal and plants demonstrated the HT of An-RTE from ancient aphid or ancestral arthropods into angiosperms. Note that some An-RTEs were recruited as coding sequences of functional genes involved in metabolic or other biochemical processes in plants [73]. Due to the accumulation of data about a large number of new transposon HT cases in eukaryotes, the first HT database focused on the HT of TEs between eukaryotes (Horizontal Transposon Transfer DataBase (HTT-DB), http://lpa.saogabriel.unipampa.edu.br:8080/httdatabase/) was created [13].

The DNA-TEs of the *Tc1-mariner* superfamily (*Bari1* and *Bari3*) were used in the experimental conditions to analyze the promoter activity of the 5'-terminal sequence capable of controlling the reporter gene transcription even in phylogenetically distantly related organisms. Using these transposons, HT mimicry was studied in a wide range of hosts (flies, human, yeasts, and bacteria). Based on the data obtained, it was suggested that the *Bari* TE family developed *diffuse promoters*, which contributed to their distribution among all kinds of living organisms due to the HT [64]. Actually, the TEs' HT can be considered as a decisive process in maintaining and distributing TEs in the genomes of all eukaryotes. A significant volume of HT of transposons was registered in drosophila, since it is an outstanding model of evolutionary genetics, while the large volume of genomic data makes it especially suitable for the development and application of reliable statistical approaches for the detection of TEs' HT. Reliable information about TEs' HT was also obtained in a wide range of other eukaryotic species in invertebrates, vertebrates, plants, and their parasites. Since TEs make up a significant part of the nuclear genome in multicellular eukaryotes and are an important source of genetic variations in evolution, TEs' HT should be considered as a key source of eukaryotic genome variability [74].

INTERACTION OF TRANSPOSONS AND VIRUSES

Viruses can act as TE vectors. The detection of *Chapaev* TEs in the *Bracovirus* members demonstrated that these viruses can act as a possible vector for the horizontal distribution of the *Chapaev* TE [67]. TEs continue to be registered as part of viruses but their contribution to the evolution of the viral genome remains largely unexplored. The miniature inverted-repeat transposable element (MITE) family (called *Submariner*) was detected as a part of the giant *Pandoravirus salinus* virus. The DNA-TEs associated with the *Submariner* in the genomes of the *Acanthamoeba castellanii*, which is the host of Pandoravirus and contains the coding sequence residues for the Tc1/mari-

ner transposase, have been found. This indicates the wide-ranging distribution of the *Submariner* MITE that can move in the Pandoravirus host (amoeba). Ten out of 30 MITE in the composition of Pandoravirus were localized within the coding regions of the predicted genes, while the others were close to the genes, indicating the role of these TEs in providing viruses with genetic novelty [75].

LTR-containing retro-TEs are involved in the antiviral protection of the host genomes, causing a restriction of exogenous retroviruses. This property can be used to treat viral infections, particularly acquired immunodeficiency. A significant part of vertebrate genomes consists of ERVs. Some of the gene products encoded by ERVs and other retro-TEs can perform protective functions for the host against viral infections. In particular, it was demonstrated that the products of the *env* genes act as restriction factors against the related exogenous retroviruses in chickens, sheep, mice, and cats. The presence of such mechanisms in genomes of other organisms, including humans, is assumed [26].

Besides the protection from exogenous viruses, TEs can contribute to the integration of viral sequences in the host genomes, which often leads to chronic disease [15, 16]. This property indicates the evolutionary relationship and possible origin of viruses from TEs. For example, the phylogenetic interrelation with retroviruses has been proved relative to the hepatitis B virus [43]. In addition, viruses can contribute to the activation of TEs, which once again indicates their phylogenetic relationship and the presence of common properties and sequences. It was demonstrated that infection with human cytomegalovirus (HCMV) induces the HERV transcriptional activity in certain cell types [76].

TE promoters often show the activity regulated in space and time, which depends on the cell type or on the effect of environmental information such as stress or infection. At the same time, the expression of many TEs is limited by different stages of gametogenesis and early embryogenesis in plants and animals. Some TEs also show tightly regulated activity in the somatic tissues of a variety of organisms. At the same time, TEs have the ability to integrate non-randomly in the genomes (many TEs developed the mechanisms of integration in genomic areas that maximize the probability of their distribution [45]). The use of TEs for the integration of viruses in the genome indicates both the possible origin of viruses from TEs and the prospects of the use of this process in genetics. In this respect, the search for TEs that are able to integrate into strictly defined genome loci is important; this is promising for their use in gene therapy using viral vectors. For example, the proteins encoded in humans by the evolutionarily conservative DNA-TE-derived *PGBD5* gene can induce the same type of locus-specific DNA transpositions in human cells by cutting and pasting. DNA transpositions catalyzed by the PGBD5 protein in human cells occur on a genome-wide scale with the precise removal and preferred insertion in the TTAA region. The obvious preservation of the *PGBD5* transposition activity suggests that genomic remodeling contributes to its biological function [1]. Although most non-LTR TEs are introduced into the host genome almost randomly, some of them are inserted in specific sequences within the target sites. Based on their structural and phylogenetic features, non-LTR TE are classified into two large groups: (1) the elements encoding restriction enzyme-like endonuclease (RLEs); and (2) the elements encoding apurine/apyrimidine endonucleases (APEs). All members of the first group include site-specific elements. Out of 20 members of the second group, only Tx1 and R1 contain site-specific elements. The targets of site-specific non-LTR TEs are usually located within the gene cluster with many copies such as the rRNA cluster genes or repeating genomic sequences such as telomeric repeats. It is assumed that site-specific TE insertions are associated with the selection of the transposition variants that least damage the host genome. The site and sequence specificity varies even in closely related non-LTR TEs and changes during evolution. Thus, for RLE elements, DNA binding motifs affect the specificity of the insertion of these transposons in certain genome regions. The highly specialized integration properties of these site-specific non-LTR TEs make them ideal alternative tools for the delivery of genes specific for sequences, especially for therapeutic purposes in human diseases [77].

CONCLUSIONS

Pathways of the emergence of exogenous viruses from transposons, as well as the possibilities of their interconversion during the course of evolution, are described in the review article. Data on the peculiarities of the elements able to exist both as viruses and as transposons are presented. It is assumed that they represent an intermediate link in evolution between TEs and viruses. The emergence of viral particles from transposons and their subsequent evolution are caused by the high degree of mutability of their sequences; changes in their sequences contribute to the acquisition of new properties, including adaptive properties. This TE property also contributed to their use by the host genomes to generate ncRNA and protein genes, which explains the widespread presence of TEs in all living organisms, especially in multicellular eukaryotes. There are a number of common properties of TEs and viruses that confirm their phylogenetic relationship. In addition, some giant viruses can contain retroelements. Autonomous transposons that not contain LTRs contribute to the transfer of non-autonomous transposons, as well as presumably to the integration of certain exogenous viruses to the genome. LTR-containing retroelements can be involved in the

host's antiviral protection causing the restriction of exogenous retroviruses. This property can be used to treat viral infections, especially acquired immunodeficiency. The need to study interactions of viruses and transposons is associated with the use of exogenous viruses for gene therapy, since the application of viral vector systems can cause undesirable effects from transposons (their imbalance, as is well known, can induce carcinogenesis). In addition, the positive effects of the interaction of viral vector systems can be used to modulate the functioning of the organism, particularly, in the fight against aging.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Henssen, A.G., Henaff, E., Jiang, E., et al., Genomic DNA transposition induced by human PGBD5, *Elife,* 2015, vol. 4, p. e10565. https://doi.org/10.7554/eLife.10565.
2. Han, M.J., Xu, H.E., Zhang, H.H., Feschotte, C., and Zhang, Z., Spy: A new group of eukaryotic DNA transposons without target site duplications, *Genome Biol. Evol.,* 2014, vol. 6, no. 7, pp. 1748–1757. https://doi.org/10.1093/gbe/evu140.
3. Ravin, N.V. and Shestakov, S.V., Genome of prokaryotes, *Vavilovskii Zh. Genet. Sel.,* 2013, vol. 17, pp. 972–984.
4. Khesin, R.B., *Nepostoyanstvo genoma* (Nonconstancy of Genome), Moscow: Nauka, 1984.
5. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., and Schulman, A.H., A unified classification system for eukaryotic transposable elements, *Nat. Rev. Genet.,* 2007, vol. 8, no. 12, pp. 973–982.
6. Ayarpadikannan, S., Lee, H.E., Han, K., and Kim, H.S., Transposable element-driven transcript diversification and its relevance to genetic disorders, *Gene,* 2015, vol. 558, no. 2, pp. 187–194.
7. Ma, B., Li, T., Xiang, Z., and He, N., MnTEdb, a collective resource for mulberry transposable elements, *Database,* 2015, vol. 27, p. bav004. https://doi.org/10.1093/database/bav004.
8. Luchetti, A., TerMITEs: Miniature inverted-repeat transposable elements (MITEs) in the termite genome (*Blattodea: Termitoidae*), *Mol. Genet. Genomics,* 2015, vol. 290, no. 4, pp. 1499–1509. https://doi.org/10.1007/s00438-015-1010-1.
9. Jiang, C., Chen, C., Huang, Z., Liu, R., and Verdier, J., ITIS, a bioinformatics tool for accurate identification of transposon insertion sites using next-generation sequencing data, *BMC Bioinf.,* 2015, vol. 16, p. 72. https://doi.org/10.1186/s12859-015-0507-2.
10. Toro, N. and Nisa-Martinez, R., Comprehensive phylogenetic analysis of bacterial reverse transcriptases, *PLoS One,* 2014, vol. 9, p. e114083. https://doi.org/10.1371/journal.pone.0114083.
11. Koonin, E.V., Dolja, V.V., and Krupovic, M., Origins and evolution of viruses of eukaryotes: The ultimate modularity, *Virology,* 2015, vols. 479–480, pp. 2–25. https://doi.org/10.1016/j.virol.2015.02.039
12. Skalka, A.M., Retroviral DNA transposition: Themes and variations, *Microbiol. Spectrum,* 2014, vol. 2, no. 5, p. MDNA300052014.
13. Dotto, B.R., Carvalho, E.L., da Silva, A.F., Dezordi, F.Z., Pinto, P.M., Campos, T.L., Rezende, A.M., and Wallau, G.D.L., HTT-DB: New features and updates, *Database (Oxford),* 2018, vol. 1. https://doi.org/10.1093/database/bax102.
14. Pastuzyn, E.D., Day, C.E., Kearns, R.B., Kyrke-Smith, M., Taibi, A.V., McCormick, J., Yoder, N., Belnap, D.M., Erlendsson, S., Morado, D.R., Briggs, J.A.G., Feschotte, C., and Shepherd, J.D., The neuronal gene Arc encodes a repurposed retrotransposon gag protein that Mediates intercellular RNA Transfer, *Cell,* 2018, vol. 172, nos. 1–2, pp. 275–288. https://doi.org/10.1016/j.cell.2017.12.024.
15. Tarocchi, M., Polvani, S., Marroncini, G., and Galli, A., Molecular mechanism of hepatitis B virus-induced hepatocarcinogenesis, *World J. Gastroenterol.,* 2014, vol. 20, no. 33, pp. 11630–11640.
16. Speiseder, T., Nevels, M., and Dobner, T., Determination of the transforming activities of adenovirus oncogenes, *Methods Mol. Biol.,* 2014, vol. 1089, pp. 105–115.
17. Notwell, J.H., Chung, T., Heavner, W., and Bejerano, G., A family of transposable elements co-opted into developmental enhancers in the mouse neocortex, *Nat. Commun.,* 2015, vol. 6, p. 6644.
18. Pavlicev, M., Hiratsuka, K., Swaggart, K.A., Dunn, C., and Muglia, L., Detecting endogenous retrovirus-driven tissue-specific gene transcription, *Genome Biol. Evol.,* 2015, vol. 7, no. 4, pp. 1082–1097. https://doi.org/10.1093/gbe/evv049.
19. Ito, J., Suqimoto, R., Nakaoka, H., Yamada, S., Kimura, T., Hayano, T., and Inoue, I., Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses, *PLoS Genet.,* 2017, vol. 13, no. 7, p. e1006883.
20. Gim, J., Ha, H., Ahn, K., Kim, D.S., and Kim, H.S., Genome-wide identification and classification of microRNAs derived from repetitive elements, *Genomics Inf.,* 2014, vol. 12, pp. 261–267.
21. Johnson, R. and Guigo, R.J., The RIDL hypothesis: Transposable elements as functional domains of long noncoding RNAs, *RNA,* 2014, vol. 20, pp. 959–976.
22. Feschotte, C., Transposable elements and the evolution of regulatory networks, *Nat. Rev. Genet.,* 2008, vol. 9, pp. 397–405.
23. Zdobnov, E.M., Campillos, M., Harrington, E.D., Torrents, D., and Bork, P., Protein coding potential of retroviruses and other transposable elements in vertebrate genomes, *Nucleic Acids Res.,* 2005, vol. 33, pp. 946–954.
24. Hoen, D.R. and Buraeu, T.E., Discovery of novel genes derived from transposable elements using integrative genomic analysis, *Mol. Biol. Evol.,* 2015, vol. 32, no. 6, pp. 1487–1506. https://doi.org/10.1093/molbev/msv042.
25. Le Dantec, C., Vallet, S., Brooks, W.H., and Renaudineau, Y., Human endogenous retrovirus group E and

its involvement in diseases, *Viruses,* 2015, vol. 7, no. 3, pp. 1238–1257.

26. Malfavon-Borja, R. and Feschotte, C., Fighting fire with fire: Endogenous retrovirus envelopes as restriction factors, *J. Virol.,* 2015, vol. 89, no. 8, pp. 4047–4050. https://doi.org/10.1128/JVI.03653-14.
27. Abascal, F., Tress, M.L., and Valencia, A., Alternative splicing and co-option of transposable elements: The case of TMPO/LAP2α and ZNF451 in mammals, *Bioinformatic,* 2015, vol. 31, no. 14, pp. 2257–2261. https://doi.org/10.1093/bioinformatics/btv132.
28. Zhuo, X. and Feschotte, C., Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages, *PLoS Pathog.,* 2015, vol. 11, no. 11, p. e1005279. https://doi.org/10.1371/journal.ppat.1005279.
29. Llorens, C., Munoz-Pomer, A., Bernad, L., Botella, H., and Moya, A., Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees, *Biol. Direct,* 2009, vol. 4, pp. 41–72.
30. Tengs, T., Kristoffersen, A.B., Bachvaroff, T.R., and Jonassen, C.M., A mobile genetic element with unknown function found in distantly related viruses, *Virol. J.,* 2013, vol. 10, p. 132.
31. Kanai, A., Virus, phage, transposon and their regulatory small non-coding RNAs, *Uirusu,* 2011, vol. 61, no. 1, pp. 25–34.
32. Fischer, M.G. and Suttle, C.A., A virophage at the origin of large DNA transposons, *Science,* 2011, vol. 332, no. 6026, pp. 231–234.
33. Yutin, N., Shevchenko, S., Kapitonov, V., Krupovic, M., and Koonin, E.V., A novel group of diverse Polinton-like viruses discovered by metagenome analysis, *BMC Biol.,* 2015, vol. 13, p. 95.
34. Kapitonov, V.V. and Jurka, J., Self-synthesizing DNA transposons in eukaryotes, *Proc. Natl. Acad. Sci. U. S. A.,* 2006, vol. 103, no. 12, pp. 4540–4545.
35. Haapa-Paananen, S., Wahlberg, N., and Savilahti, H., Phylogenetic analysis of Maverick/Polinton giant transposons across organisms, *Mol. Phylogenet. Evol.,* 2014, vol. 78, pp. 271–274.
36. Dupuy, C., Periquet, G., Serbielle, C., Bezier, A., Louis, F., and Drezen, J.-M., Transfer of a chromosomal Maverick to endogenous bracovirus in a parasitoid wasp, *Genetiva,* 2011, vol. 139, no. 4, pp. 489–496.
37. Krupovic, M. and Koonin, E.V., Self-synthesizing transposons: Unexpected key players in the evolution of viruses and defense systems, *Curr. Opin. Microbiol.,* 2016, vol. 31, pp. 25–33.
38. Routh, A., Domitrovic, T., and Johnson, J.E., Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus, *Proc. Natl. Acad. Sci. U. S. A.,* 2012, vol. 109, no. 6, pp. 1907–1912.
39. Koonin, E.V., Krupovic, M., and Yutin, N., Evolution of double-stranded DNA viruses of eukaryotes: From bacteriophages to transposons to giant viruses, *Ann. N. Y. Acad. Sci.,* 2015, vol. 1341, pp. 10–24. https://doi.org/10.1111/nyas.12728.
40. Krupovic, M. and Koonin, E.V., Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution, *Nat. Rev. Microbiol.,* 2015, vol. 13, no. 2, pp. 105–115.
41. Krupovic, M., Bamford, D.H., and Koonin, E.V., Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses, *Biol. Direct,* 2014, vol. 9, pp. 6–13. https://doi.org/10.1186/1745-6150-9-6.
42. Xiong, Y. and Eickbush, T.H., Origin and evolution of retroelements based upon their reverse transcriptase sequence, *EMBO J.,* 1990, vol. 9, pp. 3353–3362.
43. Miller, R.H. and Robinson, W.S., Common evolutionary origin of hepatitis B virus and retroviruses, *Proc. Natl. Acad. Sci. U. S. A.,* 1986, vol. 83, no. 8, pp. 2531–2535.
44. Pace, J.K., Gilbert, C., Clark, M.S., and Feschotte, C., Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods, *Proc. Natl. Acad. Sci. U. S. A.,* 2008, vol. 105, no. 44, pp. 17023–17028. https://doi.org/10.1073/pnas.0806548105.
45. Chuong, E.B., Elde, N.C., and Feschotte, C., Regulatory activities of transposable elements: from conflicts to benefits, *Nat. Rev. Genet.,* 2017, vol. 18, no. 2, pp. 71–86.
46. Long, Y., Wang, X., Youmans, D.T., and Cech, T.R., How do IncRNAs regulate transcription, *Sci. Adv.,* 2017, vol. 3, p. eaao2110. https://doi.org/10.1126/sciadv.aao2110.
47. Zhang, J., Mujahid, H., Hou, Y., Nallamilli, B.R., and Peng, Z., Plant long ncRNAs: A new frontier for gene regulatory control, *Am. J. Plant Sci.,* 2013, vol. 4, ID:32139.
48. Anderson, D.M., Anderson, K.M., Cang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., and Olson, E.N., A micropeptide encoded by a putative long noncoding RNA regulates muscle performance, *Cell,* 2015, vol. 160, pp. 595–606.
49. Lv, S., Pan, L., and Wang, G., Commentary: Primary transcripts of microRNAs encode regulatory peptides, *Front. Plant Sci.,* 2016, vol. 7, p. 1436.
50. Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T., Cannon, S.C., Houser, S.R., Bassel-Duby, R., and Olson, E.N., A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle, *Science,* 2016, vol. 351, pp. 271–275.
51. Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M., Long non-coding RNAs as a source of new peptides, *Elife,* 2014, vol. 3, p. e03523. https://doi.org/10.7554/eLife.03523.
52. Pang, Y., Mao, C., and Liu, S., Encoding activities of non-coding RNAs, *Theronostics,* 2018, vol. 8, no. 9, pp. 2496–2507.
53. Ransohoff, J.D., Wei, Y., and Khavari, P.A., The functions and unique features of long intergenic non-coding RNA, *Nat. Rev. Mol. Cell Biol.,* 2018, vol. 19, no. 3, pp. 143–157.
54. Bazzini, A.A., Johnstone, T.G., Christino, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., and Giraldez, A.J., Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation, *EMBO J.,* 2014, vol. 33, no. 9, pp. 981–993.

55. Ingolia, N.T., Lareau, L.F., and Weissman, J.S., Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes, *Cell,* 2011, vol. 147, no. 4, pp. 789–802.

56. Juntawong, P., Girke, T., Bazin, J., and Bailey-Serres, J., Translational dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis, Proc. Natl. Acad. Sci. U. S. A.,* 2014, vol. 111, no. 1, pp. 203–212.

57. Van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P.B., de Bruijn, E., Hao, W., Maclnnes, A.W., Cuppen, E., and Simonis, M., Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes, *Genome Biol.,* 2014, vol. 15, no. 1, p. 6.

58. Fang, X. and Qi, Y., RNAi in plants: An argonaute-centered view, *Plant Cell,* 2016, vol. 28, pp. 272–285.

59. Couzigou, J.M., Lauressergues, D., Becard, G., and Comier, J.P., miRNA-encoded peptides (miPEPs): A new tool to analyze the role of miRNAs in plant biology, *RNA Biol.,* 2015, vol. 12, pp. 1178–1180.

60. Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H.A., and Begun, D.J., Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression, *Proc. Natl. Acad. Sci. U. S. A.,* 2006, vol. 103, pp. 9935–9939.

61. Cai, J., Zhao, R., Jiang, H., and Wang, W., De novo origination of a new protein-coding gene in *Saccharomyces cerevisiae, Genetics,* 2008, vol. 179, pp. 487–496.

62. Xie, C., Zhang, Y.E., Chen, J.Y., Liu, C.J., Zhou, W.Z., Li, Y., Zhang, M., Zhang, R., Wei, L., and Li, C.Y., Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs, *PLoS Genet.,* 2012, vol. 8, p. e1002942.

63. Andrews, S.J. and Rothnagel, J.A., Emerging evidence for functional peptides encoded by short open reading frames, *Nat. Rev. Genet.,* 2014, vol. 15, no. 3, pp. 193–204. https://doi.org/10.1038/nrg3520.

64. Palazzo, A., Caizzi, R., Viggiano, L., and Marsano, R.M., Does the promoter constitute a barrier in the horizontal transposon transfer process? Insight from Bari transposons, *Genome Biol. Evol.,* 2017, vol. 9, no. 6, pp. 1637–1645.

65. Gilbert, C., Waters, P., Feschotte, C., and Schaack, S., Horizontal transfer of OC1 transposons in the Tasmanian devil, *BMC Genomics,* 2013, vol. 14, p. 134. https://doi.org/10.1186/1471-2164-14-134.

66. Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., and Chovnick, A., Evidence for horizontal transmission of the P transposable element between *Drosophila species, Genetics,* 1990, vol. 124, no. 2, pp. 339–355.

67. Zhang, H., Feschotte, C., Han, M., and Zhang, Z., Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals, *Genome Biol. Evol.,* 2014, vol. 6, no. 6, pp. 1375–1386. https://doi.org/10.1093/gbe/evu112.

68. Sniezewski, L., Janik, S., Laszkiewicz, A., Majkowski, M., Kisielow, P., and Cebrat, M., The evolutionary conservation of the bidirectional activity of the NWC gene promoter in jawed vertebrates and the domestication of the RAG transposon, *Dev. Comp. Immunol.,* 2018, vol. 81, pp. 105–115. https://doi.org/10.1016/j.dci.2017.11.013.

69. Fugmann, S.D., The origins of the Ras genes—from transposition to V(D)J recombination, *Semin. Immunol.,* 2010, vol. 22, no. 1, pp. 10–16. https://doi.org/10.1016/j.smim.2009.11.004.

70. Bernstein, R.M., Schluter, S.F., Bernstein, H., and Marchalonis, J.J., Primordial emergence of the recombination activating gene 1 (RAG1): Sequence of the complete shark gene indicates homology to microbial integrases, *Proc. Natl. Acad. Sci. U. S. A.,* 1996, vol. 93, no. 18, pp. 9454–9459.

71. Fugmann, S.D., Messier, C., Novack, L.A., Cameron, R.A., and Rast, J.P., An ancient evolutionary origin of the Rag 1/2 gene locus, *Proc. Natl. Acad. Sci. U. S. A.,* 2006, vol. 103, no. 10, pp. 3728–3733.

72. Hencken, C.G., Li, X., and Craig, N.L., Functional characterization of an active Rag-like transposase, *Nat. Struct. Mol. Biol.,* 2012, vol. 19, no. 8, pp. 834–836. https://doi.org/10.1038/nsmb.2338.

73. Gao, D., Chu, Y., Xia, H., Xu, C., Heyduk, K., Abernathy, B., Ozias-Akins, P., Leebens-Mack, J.H., and Jackson, S.A., Horizontal transfer of non-LTR retrotransposons from arthropods to flowering plants, *Mol. Biol. Evol.,* 2018, vol. 35, no. 2, pp. 354–364. https://doi.org/10.1093/molbev/msx275.

74. Venner, S., Miele, V., Terzian, C., Biemont, C., Daubin, V., Feschotte, C., and Pontier, D., Ecological networks to unravel the routes to horizontal transposon transfer, *PLoS Biol.,* 2017, vol. 15, no. 2, p. e2001536. https://doi.org/10.1371/journal.pbio.2001536.

75. Sun, C., Feschotte, C., Wu, Z., and Mueller, R.L., DNA transposons have colonized the genome of the giant virus *Pandoravirus salinus, BMC Biol.,* 2015, vol. 13, p. 38. https://doi.org/10.1186/s12915-015-0145-1.

76. Assinger, A., Yaiw, K.C., Gottesdorfer, I., Leib-Mosch, C., and Soderberg-Naucler, C., Human cytomegalovirus (HCMV) induces human endogenous retrovirus (HERV) transcription, *Retrovirology,* 2013, vol. 10, p. 132. https://doi.org/10.1186/1742-4690-10-132.

77. Fujiwara, H., Site-specific non-LTR retrotransposons, *Microbiol. Spectrum,* 2015, vol. 3, no. 2, p. MDNA3-0001-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0001-2014.

Translated by A. Barkhash