



Guardian of the Genome: An Alternative RAG/Transib Co-Evolution Hypothesis for the Origin of V(D)J Recombination

Iryna Yakovenko¹, Jacob Agronin², L. Courtney Smith² and Matan Oren^{1*}

¹ Department of Molecular Biology, Ariel University, Ariel, Israel, ² Department of Biological Sciences, George Washington University, Washington, DC, United States

OPEN ACCESS

Edited by:

Brian Dixon,
University of Waterloo, Canada

Reviewed by:

Pierre Pontarotti,
Centre National de la Recherche
Scientifique (CNRS), France
Yuko Ota,
University of Maryland, Baltimore,
United States

*Correspondence:

Matan Oren
matanor@ariel.ac.il

Specialty section:

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

Received: 13 May 2021

Accepted: 05 July 2021

Published: 28 July 2021

Citation:

Yakovenko I, Agronin J, Smith LC and
Oren M (2021) Guardian of the
Genome: An Alternative RAG/Transib
Co-Evolution Hypothesis for the
Origin of V(D)J Recombination.
Front. Immunol. 12:709165.
doi: 10.3389/fimmu.2021.709165

The appearance of adaptive immunity in jawed vertebrates is termed the immunological ‘Big Bang’ because of the short evolutionary time over which it developed. Underlying it is the recombination activating gene (RAG)-based V(D)J recombination system, which initiates the sequence diversification of the immunoglobulins and lymphocyte antigen receptors. It was convincingly argued that the *RAG1* and *RAG2* genes originated from a single transposon. The current dogma postulates that the V(D)J recombination system was established by the split of a primordial vertebrate immune receptor gene into V and J segments by a *RAG1/2* transposon, in parallel with the domestication of the same transposable element in a separate genomic locus as the RAG recombinase. Here, based on a new interpretation of previously published data, we propose an alternative evolutionary hypothesis suggesting that two different elements, a *RAG1/2* transposase and a *Transib* transposon invader with RSS-like terminal inverted repeats, co-evolved to work together, resulting in a functional recombination process. This hypothesis offers an alternative understanding of the acquisition of recombinase function by RAGs and the origin of the V(D)J system.

Keywords: RAG1, RAG2, RSS, transposons, immunological big bang, terminal inverted repeats, guns for hire, adaptive immune system evolution

INTRODUCTION: V(D)J RECOMBINATION AND RAGS

An outstanding feature of the jawed vertebrates is their adaptive immune system that is capable of recombining gene segments to create a diverse repertoire of immunoglobulins (Igs) (1, 2) and T cell receptors (TCRs) (3). After the primary response to a specific pathogen, the adaptive immune system mounts an enhanced secondary response to subsequent encounters with that pathogen, which is the basis for immunological memory. Sequence variation of the antigen-binding sites in the Igs and TCRs is essential for the host immune system to recognize and destroy the extensive array of pathogens before any irreversible damage takes place. Hence, the mechanisms to diversify the Ig family repertoire are of significant advantage for the fitness of higher vertebrates.

Ig genes are composed of heavy and light chain subunits, and their unique antigen-binding structures are accordingly determined by the recombination of either two or three types of gene segments. The heavy chain variable domain is encoded by variable (V), diversity (D), and joining (J) segments, while the light chain variable domain lacks the D segment (4, 5). TCRs have an analogous heterodimer structure with protein chains containing variable domains that are similarly encoded by different combinations of either *VDJ* or *VJ* segments (6). The recombination of *VDJ* or *VJ* gene segments, commonly defined as *V(D)J* recombination, is facilitated by a complex of two enzymes that are encoded by the recombination activating genes (*RAG1* and *RAG2*) (7). The *RAG1/2* complex recognizes and binds to recombination signal sequences (RSSs) that flank the *V(D)J* gene segments (**Figure 1A**). RSSs function as terminal inverted repeats (TIRs) and are composed of semi-conserved

heptamer and nonamer sequences separated by a spacer region of either 12 or 23 base pairs (bp) (**Figure 2A**). According to the 12/23 rule, *V(D)J* recombination can only occur asymmetrically, based on the pairing of RSSs with 12 and 23 bp spacers (10). This system maintains the specificity of the recombination process and prevents the gene segments from recombining incorrectly and the *Ig* loci from recombining with other *Ig* or *TCR* loci. Recombination between gene segments is initiated early in the development of lymphocytes by the *RAG1/2* complex bound to the RSSs. The complex nicks DNA at the 5' end of the RSS heptamer at the junction with the coding gene segment. This allows the free 3' OH group to attack the phosphodiester bond on the opposite strand in a transesterification reaction that forms covalently closed hairpins at the ends of the coding segments and blunt-end double-strand breaks at the ends of the RSS heptamers (11). The RSS ends are ligated head-to-head to form the signal

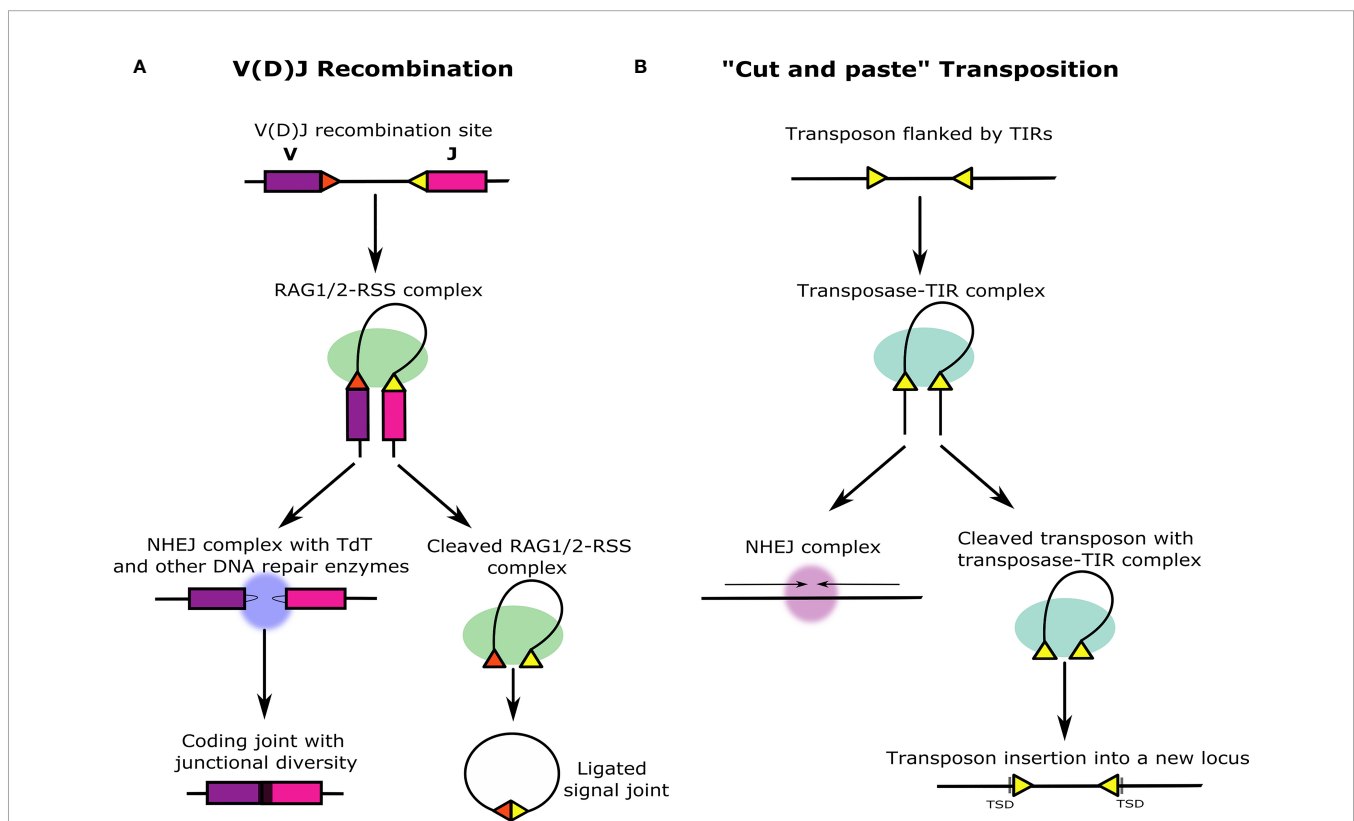


FIGURE 1 | Similarities and differences between *V(D)J* recombination and 'cut and paste' transposition. **(A)** *V(D)J* recombination occurs at the immune gene loci in differentiating lymphocytes during early T and B cell maturation stages. The *RAG1/2* protein complex (green) binds to two asymmetric RSSs (yellow and red triangles) flanking *V*, *D*, and *J* gene segments (in this illustration, the *D* segment is not shown). The DNA double helix bends and folds into the recombination synaptic complex based on the selected RSS pair. Next, *RAG1/2* introduces a nick at the intersection between each RSS and the coding gene segment that leads to the formation of closed DNA hairpins on the coding segments, and blunt, 5' phosphorylated RSS ends at the signal ends that remain associated with the *RAG1/2* complex and are ligated together forming a signal joint. The signal joint circle is deleted from the genome. Before ligation, the coding ends are subjected to further diversification by DNA repair enzymes together with TdT (blue) that generate junctional sequence diversity (black region between purple and pink gene segments). **(B)** 'Cut and paste' transposition starts similarly to *V(D)J* recombination with the transposase enzymes binding to the TIRs flanking the ends of the transposon (yellow triangles). Analogous to the beginning of the *V(D)J* recombination, the DNA double helix bends and folds into a transposition synaptic complex. The transposase makes double-stranded breaks in the DNA, and the transposon is entirely excised including the TIRs. The genomic location from which the transposon is excised is immediately ligated by NHEJ mechanism. Unlike the excised *V(D)J* signal joint circle that is lost from the genome, the excised transposon with the transposase-TIR complex creates a double-stranded break in a different region in the genome and integrates into the target site. This activity generates target site duplications (TSDs) on both sides of the integrated transposon that are formed similarly in *RAG* transposition events.

immunological ‘Big Bang’ because of the relatively short evolutionary time over which it evolved (14). It was speculated that two primary events were necessary for the emergence of an AIS - the appearance of *RAG* genes and two rounds of whole-genome duplications, that provided multiple copies of genes required for accelerated AIS evolution, including the *Ig* gene family. For example, four paralogous gene regions that encode the major histocompatibility complex (MHC) proteins and complement proteins, which are critical elements of the vertebrate AIS, are likely the result of such duplications [reviewed in (15)]. Because RAG-mediated V(D)J recombination is the fundamental process for jawed vertebrate combinatorial immunity, it became the focus of research on the evolution of the vertebrate AIS.

Almost since the discovery of the V(D)J recombination system, it was proposed that the structure of the loci encoding the antigen receptors might have been attained by the insertion of a particular ‘alien seed’ into a primordial vertebrate antigen receptor that split it into *V* and *J* segments (4). The term alien seed has been used to describe a possible ancient transposon that was inserted, of which only its TIR sequences (*i.e.*, the RSSs) have remained throughout evolution. This hypothesis was later extended to include the formation of a *D* segment as a result of the duplication events that led to formation of the VDJ segments clusters (16). It was also noted that the V(D)J recombination process bears fundamental features of DNA cut-and-paste transposition [*e.g.* (17)] (**Figure 1**). The RAG protein structure contains a DDE catalytic motif, critical for RAG1 function (18), which is also a key motif in certain integrases and transposases (18, 19). Moreover, RAG1 and RAG2 work together as a transposase *in vitro*, that is capable of excising a DNA segment bounded by RSSs and, in few cases, inserting it into a different location in the DNA (20, 21). Short DNA duplications that accompany these *in vitro* transpositions on both sides of the insertion are highly reminiscent of the ~ 5 bp target site duplications (TSD) that is a hallmark of DNA transposons (22). The function of the RAG recombinases, together with the tight physical linkage between vertebrate *RAG1* and *RAG2* genes, implied that the two genes might have acted once as a single transposon (23, 24) (*RAG1/2* transposon).

To explain the evolutionary appearance of the vertebrate RAG recombinase simultaneously with its V(D)J targets, it was suggested that the RAGs and the split antigen receptor event were derived from at least two insertions of the same *RAG1/2* transposon into the germline early in the evolution of the vertebrate lineage (16, 25–27). Another version of the transposon/split receptor gene hypothesis suggest that the original *RAG1/2* recombinase and the split of the immune receptor gene derived from different but similar mobile genetic elements that were originated from a common transposon ancestor [*e.g.* (28, 29)]. To gain credence for the theory, researchers searched invertebrate genomes for such a transposon with core regions encoding both *RAG1* and *RAG2* proteins. Indeed, the supposed missing link or an active ‘living transposon fossil’ was identified in the lancelet, *Branchiostoma belcheri*, and was termed ‘*protoRAG*’ (30). This finding seemingly enforced the existing transposon/split receptor gene hypothesis

(8), which has become the sole interpretation for the ample findings that have accumulated in this field. However, a few major questions regarding the process remained unanswered by the transposon/split receptor gene hypothesis. Perhaps the most important question is what were the evolutionary selection forces that drove the formation of this rather complicated mechanism? Despite this gap in understanding, no alternative hypothesis has been proposed.

RAG ORIGIN AND EVOLUTION

In parallel with investigations to define the functions encoded by the vertebrate *RAG* genes, efforts have been made to trace possible transposon relatives of the *RAG* genes that would shed light on their evolutionary origin. Kapitonov and Jurka (8) identified several DNA transposons of the Transib superfamily in invertebrates with predicted transposase cut-and-paste function. The transposase contains a functional core region of about 600 amino acids that is highly similar to the core region of vertebrate *RAG1* and includes the DDE motif essential for the RAG recombinase catalytic activity (8). Among all of the DDE transposases investigated, the *Transib* family shows the highest sequence similarity to *RAG1*. Furthermore, these transposons exhibit 5 bp TSDs (8) that are also observed upon experimental transpositions with *RAG1/2* (20). This led to the conclusion that the ancestry of the *RAG1* gene lies within the *Transib* transposon family (8). Although some *Transib* family members are similar to *RAG1*, none include an N terminal region that is present in *RAG1*. The clue for the origins of the N terminal region in *RAG1* came with the discovery of a transposable element (*N-RAG-TP*) in the sea slug, *Aplysia californica*, that is composed entirely of the *RAG1* N-terminal-like sequence (31). This finding led to the assumption that the complete *RAG1* structure was likely derived from the recombination between a *Transib* and the *N-RAG-TP* transposon. We suggest that the *N-RAG-TP* transposon was inserted into the 5′ end of the *Transib* sequence and not vice versa, keeping the original *Transib* TIRs on both sides of the recombined sequence without interrupting the transposase function (**Figure 3**).

Solitary *RAG1* orthologues (*RAGIL* genes) that encode the N terminal domain are present in invertebrates, and vertebrate *RAG1* may function independently of *RAG2* (23, 32). Therefore, it was assumed that the *RAG1* and *RAG2* genes were joined together in the *RAG1/2* locus after the primordial *RAG1* gained the region encoding the N terminal domain (27) (**Figure 3**). While it is not essential for transposition, the *RAG2* protein has an important function in the precision of *RAG1/2* recombinase activity. The *RAG2* core consists of a six-bladed beta-propeller (WD40 repeat) that interacts with both *RAG1* and the DNA of the coding segment next to the heptamer (33). The non-core region of *RAG2* consists of a specialized zinc finger or plant homeodomain (PHD), which binds to the N terminal tail of methylated histone 3 (H3K4me3) and is required for correct interaction with the open chromatin during recombination (34). A functional PHD domain is also encoded by *RAG2* from the sea

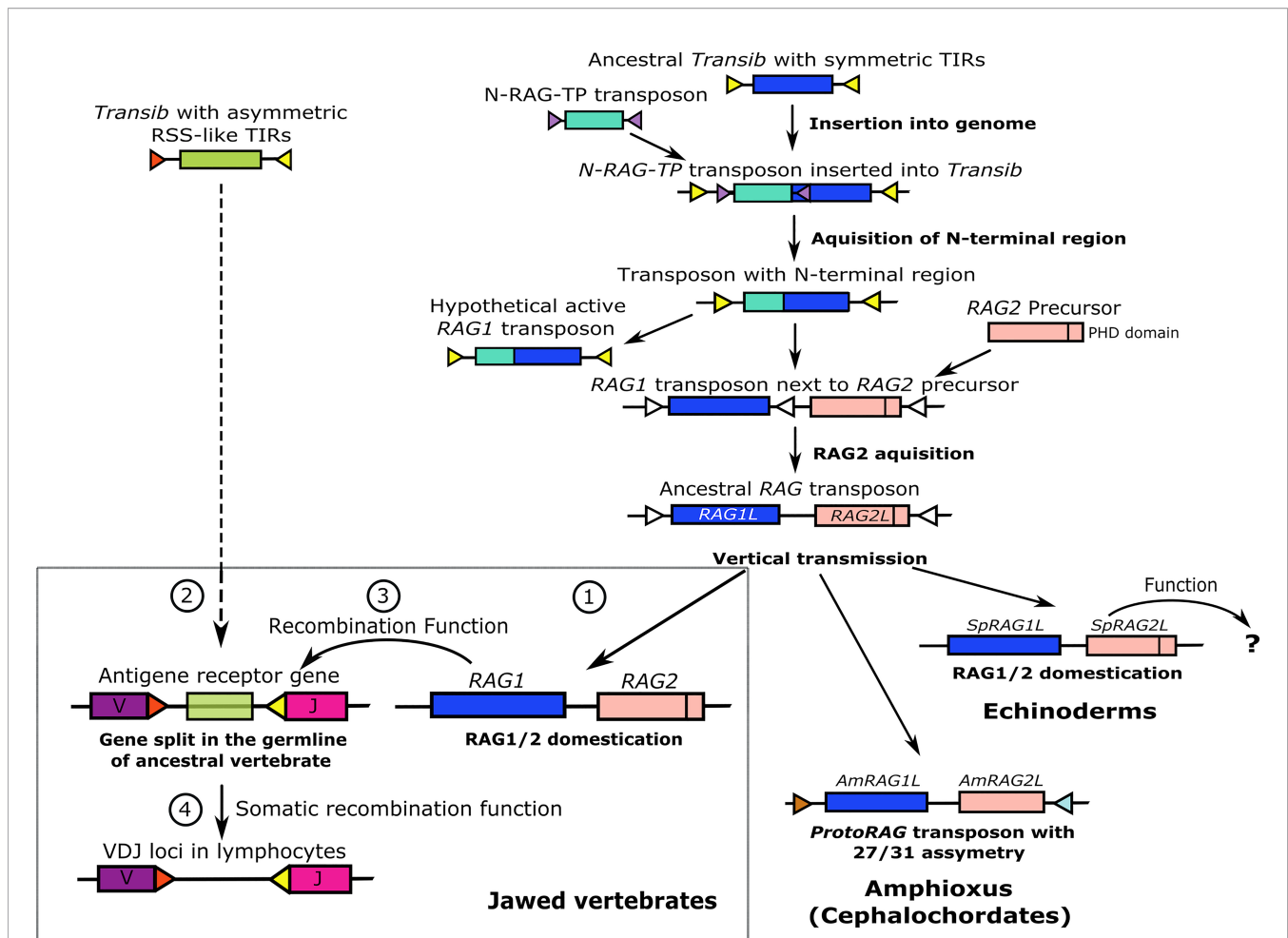


FIGURE 3 | A hypothesis for RAG evolution; an alternative to the transposon/split receptor gene model. The evolution of RAG genes begins as an ancestral *Transib* transposon with symmetric TIRs. Insertion of an *N-RAG-TP* transposon and recombination with the ancestral *Transib* gave rise to a new transposon with all the core features of an extant RAG1 gene, including identical TIR sequences and the N terminal region. An active RAG2 precursor GE was inserted close to or inside the RAG1L gene that became an ancestral hypothetical RAG1/2 transposon that was mainly further transmitted vertically. In amphioxus, the *protoRAG* transposon lost its PHD domain but maintained its transposition activity based on 27/31 asymmetric TIR spacers. In sea urchins, the RAG1/2L genes were domesticated and may have gained function, which is currently unknown. According to our guardian of the genome hypothesis (lower left box), RAG1/2 was also domesticated early in the jawed vertebrate lineage to protect vertebrate genomes from insertion and excision of harmful TEs (labeled as 1 in the box). Later, a *Transib* transposon with 12/23 asymmetric TIRs entered and survived inside a jawed vertebrate antigen receptor exon, splitting it into V and J segments (2 in the box). As an ex-transposon endonuclease, the RAG1/2 machinery was co-opted and trained to excise the asymmetric TIR-flanked *Transib* transposon threat in the V(D)J receptor loci. Fortunately, the excision process in the V(D)J loci provided an immunological advantage and therefore the RAG1/2 complex gained its V(D)J recombinase function in the germline (3 in the box). Eventually, under selective pressure, the V(D)J machinery was selected evolutionarily to work primarily in the lymphoid cell lineages (4 in the box).

urchin, *Strongylocentrotus purpuratus* (*SpRAG2*) (35), but is absent from the invertebrate amphioxus *protoRAG* transposon that does not have recombinase activity (30) (Figure 3).

The RAGL genes are found in multiple invertebrate deuterostomes, including echinoderms (36), cephalochordates (30), and hemichordates (acorn worms) (29) and were recently detected in protostomes, including oysters, mussels, ribbon worms, and even in the non-bilaterian cnidarians (9). These RAGL genes come in all flavors: RAG1/2 pairs with intact or partial TIRs or without TIRs, unlinked RAG1L genes and RAG2L genes, as well as RAG pseudogenes. However, despite considerable efforts, RAGL genes have not been identified in

jawless vertebrates or urochordates (tunicates) (9). This evolutionary gap may be the outcome of either horizontal gene transmission or loss in certain phylogenetic groups based on conventional vertical transmission (36). The horizontal transfer hypothesis, to its extreme, suggests that an ancient RAG transposon ‘jumped’ several times from some invertebrates into common ancestors of cephalochordates or jawed vertebrates and only later was domesticated in jawed vertebrates (37). However, while the amphioxus *protoRAG* transposon is likely to transpose *in vivo* (30), there is no evidence for inter-species transposition of the RAGL sequences. To the contrary, a recent analysis shows that the RAG sequence

phylogeny is gradual and directional rather than patchy (present in some clades and absent in others), supporting a more conventional evolutionary path that relies on vertical, not horizontal RAG gene transmission (9). This hypothesis suggests that the *RAG1/2L* pair was possibly present in its current form in most metazoan lineages and may have been lost in the lineages of tunicates and jawless vertebrates. In any case there is no evidence that the V(D)J recombination system arose at any time earlier than the vertebrate lineage. Nevertheless, the necessity of generating sequence diversity among immune genes in tunicates and jawless vertebrates is not doubted, and therefore, the absence of V(D)J recombination may have driven the evolution of the other diversification mechanisms in these phylogenetic groups. An example for such mechanism is the copy choice mechanism of the variable lymphocyte receptor (*VLR*) gene family in jawless vertebrates that generates significant sequence diversity in the encoded antigen receptors and relies on the function of activation-induced deaminases (AIDs) rather than RAGs (38–41). We would like to note that the answer to the RAG transmission question still remains, and it is possible that both transmission types took place.

THE 12/23 RECOMBINATION ASYMMETRY AND THE ‘TRANSIB SEED’

One of the critical elements in the V(D)J recombination system is the unique 12/23 bp asymmetric RSSs, that are equivalent to transposon TIRs. The 12/23 asymmetry rule appears to be more essential for the accurate execution of V(D)J recombination than the spacer sequences themselves that are highly variable (42). The 12/23 RSS asymmetry is only recognized when the RAG1 recombinase is complexed with RAG2 in jawed vertebrates. Although mouse RAG1 alone is capable of mediating V(D)J recombination in the absence of RAG2, its recombination efficiency is reduced significantly. Moreover, RAG1 alone does not show a preference for asymmetric 12/23 RSSs compared to symmetric 12/12 RSSs (23). Consequently, the possibility should be considered that the ancestral RAG1 gene might have originated from an ancestral Transib transposon with symmetric TIRs (Figure 3), similar to some existing *Transib* transposons (Figure 2B, alignment 1). Furthermore, recombinant sea urchin SpRAG1 (rSpRAG1) and rHzTransib, a functional *Transib* transposons from the corn earworm moth, *Helicoverpa zea* (43), both mediate V(D)J recombination through mouse 12/23 RSS sequences when combined with mouse RAG2. However, when rSpRAG1 is co-expressed with sea urchin recombinant rSpRAG2, it does not show recombination activity (23).

The above observations may suggest that V(D)J recombination, which relies on asymmetric 12/23 TIRs, dependent on the more recent evolution of RAG2 in the vertebrate lineage (23). Additionally, many of the *RAGL* transposons in early bilaterians, if they have transposase activity, probably rely on the ~15-17 bp TIRs that include a

heptamer with an A-rich region (termed transposon region 2; TR2) and show no preference for spacer asymmetry (Figure 2C) (9). The only asymmetric invertebrate TIRs found are those that flank the *protoRAG* in amphioxus with 27/31 bp spacer asymmetry, which is very different from that of the V(D)J RSSs (44). This gap challenges the scenario in which the antigen receptor gene was split into V and J segments by an ancestral hypothetical RAG transposon as proposed by the current hypothesis.

Remarkably, the only known transposition that incorporates 12/23 asymmetry other than RAG-based V(D)J recombination is found in some *Transib* transposon family members (Figure 2B, alignment 2) (8). This type of TIR asymmetry surrounds few *Transib* transposons and have almost perfect 12 or 23 ± 1 bp spacers with heptamer and nonamer sequences of which some are identical to those of the canonical RSSs in higher vertebrates (Figure 2B alignment 3). The probability that such sequence match between the RSS and *Transib* TIR termini occurred by chance is less than 10⁻³ (8). When including the match in the asymmetric spacer sizes of the 5' and 3' TIRs, the resemblance seems too great to be a coincidence. We propose that this unique asymmetrical TIR structure has a functional role not only for the RSS mediated recombination, but also in *Transib* transposition as is known for other TIR transposons (44, 45). In light of the above, we suggest that the alien seed, which entered the primordial antigen receptor gene in the jawed vertebrate lineage, was not an ancestral RAG transposon as the current hypothesis suggests but may have been a *Transib* transposon with asymmetric 12/23 RSS-like TIRs similar to currently existing *Transib* transposons. This transposon insertion scenario is supported by the argument that the *Transib* transposons are smaller than the RAG genes and are more likely to have horizontal transposition capabilities based on their phylogenetic distribution (22). It is noteworthy that the *Transib* transposons are absent from the genomes of all vertebrates (8). This fact seems to challenge our suggested hypothesis. However, as discussed below, evolutionary forces involved in RAG domestication may explain, at least in part, the absence of *Transib* transposons from vertebrate genomes.

TRANSPOSON RECRUITMENT FOR HOST DEFENSE

Selfish mobile genetic elements (MGEs), including viruses and transposons or TEs, frequently invade genomes of organisms from all kingdoms (46). Most TEs are site-specific nucleases that function by incorporating their DNA into the genome (47, 48). Therefore, TEs may impact genomic integrity and cellular function, and thus may be harmful or lethal to the host [reviewed in (49)]. To resist an assault by TEs, successful life forms have developed complex defense mechanisms such as DNA methylation of the transposon-containing regions and sequence specific RNA degradation of TE transcripts (50, 51). Some organisms have recruited TEs to combat other TEs. In these cases, TEs have been incorporated into host genomes and domesticated or co-opted to function in host defense. The ‘guns

for hire' theory (52) addresses the question of how TEs could become the guardians against invaders through gradual mutations, immobilization, and domestication. Although domesticated TEs are widespread across many taxa, it remains unclear how TE domestication starts and proceeds (45, 53, 54). The CRISPR-Cas adaptive immune system of archaea and some bacteria is an example of co-option of TEs in prokaryotes for defense against bacteriophage infection (55). The adaptation module plus both varieties of effector modules in CRISPR-Cas are considered to have evolved from either transposons or transposon-like modules (56). Other examples for domesticated TEs for protection and/or development include the following: a) Recruitment of the *PiggyBac* transposon in ciliates, which removes TE-associated non-coding internal eliminated sequences (IES) (57, 58). b) Bacterial XerC/XerD and archaeal XerA recombinases that were potentially recruited from a range of protective MGEs and function in site specific resolution of circular chromosome dimers (59, 60). c) Piwi-interacting RNAs (piRNAs) that are encoded by small TEs and are present of many eukaryotes including mammals. The piRNAs associate with a PIWI nuclease mostly in germline cells that functions in cleavage and silencing of complementary TE transcripts (61). d) Finally, *gag*, *pol* and *env* genes in humans originated from endogenous retroviruses and have been co-opted to combat other retroviruses [reviewed in (62)]. Hence, the proteins encoded by TEs should be viewed as potential offensive weapons of the host in the evolutionary arms race between hosts and their DNA parasites.

RAG DOMESTICATION AND THE ALTERNATIVE RAG/TRANSIB CO-EVOLUTION HYPOTHESIS

How the *RAG* transposon was domesticated to function as a recombinase has remained a major question for understanding the origin of V(D)J recombination (16). The current transposon/split receptor gene hypothesis does not provide an explanation for the domestication of the *RAG1/2* transposon prior to the formation of the whole V(D)J system. It assumes that both processes occurred simultaneously. This concept can be challenging because of the separate locations of the *RAG1/2* locus itself (chromosome 11 in humans), versus the *RAG* recombinase targets - the *VJ* loci of the kappa and lambda light chain genes (chromosomes 2 and 22 in humans) and the *VDJ* locus of the immunoglobulin heavy chain gene (chromosome 14 in humans). According to the current dogma, the insertion of the hypothetical *RAG* transposon into the ancestral vertebrate immune receptor gene happened only once (followed by whole genome duplications) within the timeframe of the V(D)J system appearance. This scenario seems to be unlikely considering that no other traces of *RAG* transposons or their TIRs have been identified elsewhere in higher vertebrate genomes. Furthermore, the presence of *RAG1/2* genes and their possible domestication in invertebrates are considered to have

occurred much earlier than their domestication in ancestral jawed vertebrates, for example in purple sea urchin (28).

To address these gaps, we suggest an alternative hypothesis to explain the formation of the V(D)J recombination system in the vertebrate lineage through a more gradual evolutionary process, during which separate *RAG* and *Transib* elements co-evolved to work together. We propose that a prequel to the current *RAG* recombinase function was its initial domestication for the purpose of host protection. According to the 'guns for hire' hypothesis (52) and in similar to domesticated TEs listed above, the *RAG1/2* complex, encoded by a domesticated ex-transposon, may have been initially co-opted to protect the host against transposons that could jeopardize the integrity of the genome. As a protein complex with a core that is encoded by a hypothetical immobilized transposon, the ancestral *RAG1/2* transposase, with only relatively minor modifications to its amino acid sequence, could have later acquired its current recombinase function.

The current V(D)J recombination process may also be regarded as a failed transposition (**Figure 1A**). A functional and structural analysis of *RAG* and *BbRAGL*, revealed a two tier mechanism for domestication and loss of transposition capability in the vertebrate *RAG* recombinase (63, 64). The first tier is based on a single, highly conserved Arg848 in *RAG1*. When the *RAG1* Arg848 was replaced experimentally with the invertebrate *RAG1L* equivalent of a Met848, the *RAG* transposition activity was significantly increased, whereas the reciprocal switch (replacing Met848 in *BbRAGL* with Arg) had the opposite outcome. The second tier is based on the acidic hinge domain within the first 1–383 aa of vertebrate *RAG2*. Removing the acidic hinge results in a significant increase in *RAG* transposition *in vivo*. Together, *RAG1* Arg848 and the *RAG2* acidic hinge suppress *RAG*-mediated transposition *in vivo* by more than 1,000 fold (63). We suggest that at least some of these evolutionary changes might have been involved in the initial domestication of the *RAG1/2* transposase as the guardian of the genome.

Based on the almost perfect sequence identity between some *Transib* TIRs and the RSSs (**Figure 2B**), we assume that some of the early invading transposons that were recognized by the guardian *RAG1/2* complex were members of the *Transib* family, from which *RAG1* is thought to have originated. As a hypothetical guardian of genome, the early-domesticated *RAG1/2* complex may have neutralized the *Transib/Transib*-like transposons by excision and prevention of their reintegration into vertebrate genomes. This notion is consistent with the absence of *Transib* family transposons from all vertebrate genomes that have been analyzed instead of sequenced to the date (8, 9). Following this line of thought, the extant *Ig* family loci with the V(D)J segments are the only places in the genome where the *Transib* transposons survived by purifying selection based on the evolutionary advantage in their specific locations that provided increased immune receptor sequence diversity. How the *Transib* transposons inside *Ig* sequences initially avoided the surveillance of a guardian *RAG1/2* is one of the questions that remains open in our hypothesis. It may be that the 12/23 asymmetry of the *Transib* invader has prevented its initial recognition and its full excision from the germline, a

malfunxion that was later corrected in the contemporary recombination mechanism. The original *Transib* TIR sequences (that became the RSSs) are accordingly the functional descendants of this once neutralized transposon, of which the sequence has been modified to the point of retaining no identifiable similarity to any TE.

Other proteins and enzymes associated with the RAGs during V(D)J recombination may have co-evolved to assist the initial domesticated RAG function as a guardian of the genome. One of these enzymes is TdT, which adds the random N nucleotides to the coding DNA ends after the hairpins are nicked and opened by Artemis/PKc (65). The benefit of this process is the initiation of randomized junctional diversity within the third complementarity-determining region 3 (CDR3) of the Ig and TCR chains that is the key for antigen binding. However, this process is also very wasteful because the junctional diversification process often results in frameshifts that translate into missense or truncated proteins, further supporting the notion that V(D)J recombination was not the original function of the RAG1/2 complex. When considering the RAG complex as a genome guardian, this random sequence diversification process may have been selected for during evolution because it changes the sequence of the original site from which the intervening transposon was excised, providing a protective advantage by preventing its reintegration into the same location in the genome. Supposedly, later in the evolution this junctional diversification process was established to generate immune receptor gene diversity.

The *Ig* gene family structure across different vertebrate species may provide additional clues to the later stages of V(D)J recombination evolution. Although V(D)J recombination is present in all jawed vertebrates, the structure of the *Ig* and *TCR* loci differ significantly among different classes of vertebrates. Two major structures of *Ig/TCR* loci are either as a translocon or a cluster. The *Ig* genes in mammals, birds, reptiles, and amphibians bear the more common translocon configuration in which tandem duplications of *V*, *D* (if present) and *J* gene segments are grouped in each locus [reviewed in (15)]. On the contrary, cartilaginous fish such as sharks, skates, and rays have an *Ig* gene cluster configuration in which each cluster is composed of one or two *V* gene segments, one or a few *D* gene segments (if present), a *J* gene segment, and an *Ig* constant region exon. The entire locus is made up of many repeats of such clusters or miniloci (15, 66, 67) where V(D)J recombination occurs within a cluster and not between clusters, resulting in a limited combinatorial diversity (68). Some of the *Ig* clusters in cartilaginous fish contain fully pre-joined (germline-joined) *VJ* or *VDJ* gene segments or partially pre-joined *VD-J* combinations indicating the possible activity of RAG machinery in the germline (69, 70). The pre-joined *Ig* genes are expressed during early developmental stages, and thus may provide protection to the progeny of this animal group (71). Few cases of pre-joined *VDJ* segments are present in teleost fish (72) and even in one case of a mammal (73, 74). Therefore, it is plausible that the RAG mechanism may have acted originally in the germline, supporting the suggested hypothetical genome guardian function. Perhaps later in the evolution of the

ancestral cartilaginous or of bony fish, this activity became sufficiently advantageous and was shifted to the lymphocyte lineage of the soma (Figure 3).

Taken together, we propose that the early evolution of the V(D)J recombination system may have evolved in four gradual steps (Figure 3, box). In the first step, an immobilized *RAG1/2* transposon in an ancestral vertebrate was co-opted to guard the genome from other invading TEs. In the second step, a distinct *Transib* transposon with asymmetric TIRs (the *Transib* seed), was inserted into an antigen receptor gene in an ancestral vertebrate, splitting an exon into *V* and *J* segments. In the third step, the *RAG1/2* complex gained its current recombinase function in the germline as has been speculated for some cartilaginous fish. The new feature could have been transferred to the offspring to provide an initial immunological advantage. In the fourth step, the RAG recombination mechanism was selected to function mostly in the somatic lymphocyte cell lineage in higher vertebrates because of the benefits that it imparted to the AIS through *Ig* family sequence diversification.

We suggest that the initial *RAG1/2* domestication as a genome guardian might have occurred soon after the origin of the vertebrate lineage. As no substantial evidence exists for our hypothesis, significant experimental work will need to be undertaken to support it. For example, one approach may be to test the possible protective function of the vertebrate RAGs against experimentally induced *Transib* transposon intrusions into the germline genome of model vertebrates such as mouse or zebrafish.

RAG DOMESTICATION IN INVERTEBRATES

While the biological functions of the vertebrate RAGs are known, the functions of invertebrate *RAGL* genes are yet unclear. As mentioned above, multiple pairs of *RAG1/2L* and numerous solitary *RAGIL* genes have been identified in invertebrates, including a variety of deuterostomes and even several protostomes (9). The *RAG1/2L* gene pairs in invertebrates are always positioned tail to tail as in vertebrate genomes. They may be incomplete, mutated into pseudogenes, or have the potential to encode intact, full-length proteins (9, 36). The TIRs (RSSs) flanking these gene pairs, when present, have a conserved heptamer CAC sequence, but otherwise show rather weak sequence similarity to the vertebrate RSSs, and none follow the 12/23 spacer asymmetry (Figure 2C). Some TIR-bearing *RAG1/2L* gene pairs that have been identified to date are also flanked with TSD sequences adjacent to the TIRs suggesting that they may maintain transposition activity.

Of particular interest are the *SpRAG1L* and *SpRAG2L* gene pair in the purple sea urchin, *S. purpuratus*. The *SpRAG1/2L* genes lack associated TIR sequences but have complete open reading frames, suggesting purifying selection to preserve function. The *SpRAG1/2L* genes produce full-length proteins with all major functional domains similar to the vertebrate RAGs including the RAG1 core with the DDE catalytic center (27, 36), and a complete RAG2

structure including a functional PHD (35). The *SpRAG1/2L* genes are co-expressed during early developmental stages and in immune cells or coelomocytes, and the expressed proteins form a stable complex with each other. Furthermore, SpRAG1L binds to the vertebrate RSSs with 12 bp spacers in the presence of vertebrate high mobility group (HMG) proteins (36). The sea urchin *SpRAG1/2L* genes contain multiple exons, while both vertebrate RAG genes consist of a single exon [reviewed in (75)]. The acquisition of introns may enable alternative splicing (76) as was predicted for *SpRAGL* genes (36) that might also point to some kind of functionality. In addition to the *SpRAG1/2L* gene homologs, other purple sea urchin genes, which encode repair enzymes essential in vertebrates for V(D)J recombination, including *Artemis* and *TdT*, are also expressed in adult coelomocytes and some other tissues (77). In other words, sea urchins have all basic components required for a functional vertebrate-like recombination mechanism.

Although the *RAG1/2L* gene pairs among different sea urchin species are located in non-syntenic regions of the genomes (8, 24), they appear to be either silenced or domesticated, and therefore are not likely to transpose. Together, sea urchin RAG1L and RAG2L proteins may form a vertebrate-like RAG1/2 complex that possibly functions as an endonuclease or even as a recombinase of sea urchin immune gene(s), or perhaps has some other unknown function. In the scenario where the sea urchin RAG1/2L serves as a recombinase, given that sea urchin genomes do not have any *V(D)J*-like sequence segments, we hypothesize that the proteins encoded by the *SpRAG1/2L* genes may act on different immunological targets. One of the plausible hypothetical targets may be the *SpTransformer* (*SpTrf*) immune gene family (formerly known as *Sp185/333*) in *S. purpuratus* (78), which appears to be subjected to somatic diversification in individual coelomocytes that includes gene deletions and duplications (79). Interestingly, each of the *SpTrf* genes is surrounded by GA short tandem repeats (80, 81), which may serve as TIR-like signal sequences for RAGL recombination-like activity. The function of the sea urchin RAGL proteins will be noteworthy for future investigations. A crucial step in this direction will be the identification of the genomic targets for RAG1/2L activity in sea urchins.

CONCLUSION

In the constant arms race between host and pathogens, an advantageous and permanent invention and refinement of

REFERENCES

1. Tonegawa S. Somatic Generation of Antibody Diversity. *Nature* (1983) 302 (5909):575–81. doi: 10.1038/302575a0
2. Sakano H, Maki R, Kurosawa Y, Roeder W, Tonegawa S. Two Types of Somatic Recombination Are Necessary for the Generation of Complete Immunoglobulin Heavy-Chain Genes. *Nature* (1980) 286(5774):676–83. doi: 10.1038/286676a0
3. Hedrick SM, Cohen DI, Nielsen EA, Davis MM. Isolation of cDNA Clones Encoding T Cell-Specific Membrane-Associated Proteins. *Nature* (1984) 308 (5955):149–53. doi: 10.1038/308149a0
4. Sakano H, Hüppi K, Heinrich G, Tonegawa S. Sequences at the Somatic Recombination Sites of Immunoglobulin Light-Chain Genes. *Nature* (1979) 280(5720):288. doi: 10.1038/280288a0
5. Tonegawa S. Complete Immunoglobulin Heavy-Chain Genes. *Nature* (1980) 286:14. doi: 10.1038/286657a0
6. Yanagi Y, Yoshikai Y, Leggett K, Clark SP, Aleksander I, Mak TW. A Human T Cell-Specific cDNA Clone Encodes a Protein Having Extensive Homology to Immunoglobulin Chains. *Nature* (1984) 308(5955):145–9. doi: 10.1038/308145a0
7. Oettinger MA, Schatz DG, Gorka C, Baltimore D. RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V (D) J

defensive or offensive immune mechanisms is inherently costly. In some cases, it can be a double-edged sword when the immune defense turns to various forms of autoimmunity caused by imperfections and complexities of V(D)J recombination and self/non-self recognition. While the widely accepted transposon/split receptor gene hypothesis may explain many of the molecular adaptations of the RAG complex in its current recombinase function, it has not been clear how the RAGs were selected for this function in the jawed vertebrate lineage. The recruitment of pre-formed alien components of selfish Mobile genetic elements (MGEs) opens the door for understanding the fast and cost-effective incremental improvements in host defense that may have led to the appearance of a new immune diversification mechanism. Accordingly, we postulate that the ancestral vertebrate RAG transposon underwent a domestication process in response to pathogenic TE pressure and became a guardian of the genome against other MGE invasions, which was a prequel to its exaptation for its current recombination function. In contrast to the currently accepted scenario, we suggest that the alien seed that invaded the ancestral *Ig* receptor gene and split it into *V* and *J* segments was not a RAG transposon, but rather a *Transib* transposon with RSS-like asymmetric TIR sequences. This scenario may provide an explanation for the evolutionary forces that gave rise to the V(D)J recombination system.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.imgt.org>.

AUTHOR CONTRIBUTIONS

IY, MO, and LCS contributed to the development of the hypothesis. IY, MO, LCS, and JA wrote the article. IY made the figures. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Michele Britton for the editing work.

- Recombination. *Science* (1990) 248(4962):1517–23. doi: 10.1126/science.2360047
8. Kapitonov VV, Jurka J. RAG1 Core and V (D) J Recombination Signal Sequences Were Derived From Transib Transposons. *PLoS Biol* (2005) 3(6):e181. doi: 10.1371/journal.pbio.0030181
 9. 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) 11:1–20. doi: 10.1186/s13100-020-00214-y
 10. van Gent DC, Ramsden DA, Gellert M. The RAG1 and RAG2 Proteins Establish the 12/23 Rule in V (D) J Recombination. *Cell* (1996) 85(1):107–13. doi: 10.1016/S0092-8674(00)81086-7
 11. McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, et al. Cleavage at a V (D) J Recombination Signal Requires Only RAG1 and RAG2 Proteins and Occurs in Two Steps. *Cell* (1995) 83(3):387–95. doi: 10.1016/0092-8674(95)90116-7
 12. Kallenbach S, Doyen N, d'Andon MF, Rougeon F. Three Lymphoid-Specific Factors Account for All Junctional Diversity Characteristic of Somatic Assembly of T-Cell Receptor and Immunoglobulin Genes. *Proc Natl Acad Sci* (1992) 89(7):2799–803. doi: 10.1073/pnas.89.7.2799
 13. Nielsen SC, Boyd SD. Human Adaptive Immune Receptor Repertoire Analysis—Past, Present, and Future. *Immunol Rev* (2018) 284(1):9–23. doi: 10.1111/immr.12667
 14. Bernstein RM, Schluter SF, Bernstein 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* (1996) 93(18):9454–9. doi: 10.1073/pnas.93.18.9454
 15. Flajnik MF, Kasahara M. Origin and Evolution of the Adaptive Immune System: Genetic Events and Selective Pressures. *Nat Rev Genet* (2010) 11(1):47–59. doi: 10.1038/nrg2703
 16. Hsu E, Lewis SM. The Origin of V (D) J Diversification. In: *Molecular Biology of B Cells*. Elsevier (2015). p. 133–49. doi: 10.1016/B978-0-12-397933-9.00009-6
 17. Roth DB, Craig NL. VDJ Recombination: A Transposase Goes to Work. *Cell* (1998) 94(4):411–4. doi: 10.1016/S0092-8674(00)81580-9
 18. Landree MA, Wibbenmeyer JA, Roth DB. Mutational Analysis of RAG1 and RAG2 Identifies Three Catalytic Amino Acids in RAG1 Critical for Both Cleavage Steps of V (D) J Recombination. *Genes Dev* (1999) 13(23):3059–69. doi: 10.1101/gad.13.23.3059
 19. Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. The RAG Proteins and V (D) J Recombination: Complexes, Ends, and Transposition. *Annu Rev Immunol* (2000) 18(1):495–527. doi: 10.1146/annurev.immunol.18.1.495
 20. Hiom K, Melek M, Gellert M. DNA Transposition by the RAG1 and RAG2 Proteins: A Possible Source of Oncogenic Translocations. *Cell* (1998) 94(4):463–70. doi: 10.1016/S0092-8674(00)81587-1
 21. Agrawal A, Eastman QM, Schatz DG. Transposition Mediated by RAG1 and RAG2 and its Implications for the Evolution of the Immune System. *Nature* (1998) 394(6695):744. doi: 10.1038/nri2944
 22. Kapitonov VV, Jurka J. Molecular Paleontology of Transposable Elements in the *Drosophila Melanogaster* Genome. *Proc Natl Acad Sci* (2003) 100(11):6569–74. doi: 10.1073/pnas.0732024100
 23. Carmona LM, Fugmann SD, Schatz DG. Collaboration of RAG2 With RAG1-Like Proteins During the Evolution of V (D) J Recombination. *Genes Dev* (2016) 30(8):909–17. doi: 10.1101/gad.278432.116
 24. Kapitonov VV, Koonin EV. Evolution of the RAG1-RAG2 Locus: Both Proteins Came From the Same Transposon. *Biol Direct* (2015) 10(1):20. doi: 10.1186/s13062-015-0055-8
 25. Schatz DG. Antigen Receptor Genes and the Evolution of a Recombinase. In: *Seminars in Immunology*. Elsevier (2004). p. 245–56. doi: 10.1016/j.smim.2004.08.004
 26. Thompson CB. New Insights Into V (D) J Recombination and its Role in the Evolution of the Immune System. *Immunity* (1995) 3(5):531–9. doi: 10.1016/1074-7613(95)90124-8
 27. Carmona LM, Schatz DG. New Insights Into the Evolutionary Origins of the Recombination-Activating Gene Proteins and V (D) J Recombination. *FEBS J* (2017) 284(11):1590–605. doi: 10.1111/febs.13990
 28. Fugmann SD. The Origins of the Rag Genes—From Transposition to V (D) J Recombination. In: *Seminars in Immunology*. Elsevier (2010). p. 10–6. doi: 10.1016/j.smim.2009.11.004
 29. Poole JRM, Huang SF, Xu A, Bayet J, Pontarotti P. The RAG Transposon Is Active Through the Deuterostome Evolution and Domesticated in Jawed Vertebrates. *Immunogenetics* (2017) 69(6):391–400. doi: 10.1007/s00251-017-0979-5
 30. Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, et al. Discovery of an Active RAG Transposon Illuminates the Origins of V (D) J Recombination. *Cell* (2016) 166(1):102–14. doi: 10.1016/j.cell.2016.05.032
 31. Panchin Y, Moroz LL. Molluscan Mobile Elements Similar to the Vertebrate Recombination-Activating Genes. *Biochem Biophys Res Commun* (2008) 369(3):818–23. doi: 10.1016/j.bbrc.2008.02.097
 32. Montaudouin C, Boucontet L, Mailhé-Lembezat M-P, Mariotti-Ferrandiz M-E, Louise A, Six A, et al. Endogenous TCR Recombination in TCR Tg Single RAG-Deficient Mice Uncovered by Robust *In Vivo* T Cell Activation Selection. *PLoS One* (2010) 5(4):e10238. doi: 10.1371/journal.pone.0010238
 33. Callebaut I, Mornon J-P. (D) J Recombination Activating Protein RAG2 Consists of a Six-Bladed Propeller and a PHD Fingerlike Domain, as Revealed by Sequence Analysis. *Cell Mol Life Sci CMLS* (1998) 54(8):880–91. doi: 10.1007/s000180050216
 34. Matthews AG, Kuo AJ, Ramón-Maiques S, Han S, Champagne KS, Ivanov D, et al. RAG2 PHD Finger Couples Histone H3 Lysine 4(D) J Recombination. *Nature* (2007) 450(7172):1106–10. doi: 10.1038/nature06431
 35. Wilson DR, Norton DD, Fugmann SD. The PHD Domain of the Sea Urchin RAG2 Homolog, SpRAG2L, Recognizes Dimethylated Lysine 4 in Histone H3 Tails. *Dev Comp Immunol* (2008) 32(10):1221–30. doi: 10.1016/j.dci.2008.03.012
 36. Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. An Ancient Evolutionary Origin of the Rag1/2 Gene Locus. *Proc Natl Acad Sci* (2006) 103(10):3728–33. doi: 10.1073/pnas.0509720103
 37. Sniezewski L, Janik S, Laszkiewicz A, Majkowski M, Kisielow P, 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) 81:105–15. doi: 10.1016/j.dci.2017.11.013
 38. Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, et al. Variable Lymphocyte Receptors in Hagfish. *Proc Natl Acad Sci* (2005) 102(26):9224–9. doi: 10.1073/pnas.0503792102
 39. Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic Diversification of Variable Lymphocyte Receptors in the Agnathan Sea Lamprey. *Nature* (2004) 430(6996):174–80. doi: 10.1038/nature02740
 40. Nagawa F, Kishishita N, Shimizu K, Hirose S, Miyoshi M, Nezu J, et al. Hashimoto S-I: Antigen-Receptor Genes of the Agnathan Lamprey Are Assembled by a Process Involving Copy Choice. *Nat Immunol* (2007) 8(2):206–13. doi: 10.1038/ni1419
 41. Boehm T. Design Principles of Adaptive Immune Systems. *Nat Rev Immunol* (2011) 11(5):307–17. doi: 10.1038/nri2944
 42. Lee AI, Fugmann SD, Cowell LG, Ptaszek LM, Kelsoe G, Schatz DG. A Functional Analysis of the Spacer of V (D) J Recombination Signal Sequences. *PLoS Biol* (2003) 1(1):e1. doi: 10.1371/journal.pbio.0000001
 43. Hencken CG, Li X, Craig NL. Functional Characterization of an Active Rag-Like Transposase. *Nat Struct Mol Biol* (2012) 19(8):834–6. doi: 10.1038/nsmb.2338
 44. Tao X, Yuan S, Chen F, Gao X, Wang X, Yu W, et al. Functional Requirement of Terminal Inverted Repeats for Efficient ProtoRAG Activity Reveals the Early Evolution of V (D) J Recombination. *Natl Sci Rev* (2020) 7(2):403–17. doi: 10.1093/nsr/nwz179
 45. Dyda F, Chandler M, Hickman AB. The Emerging Diversity of Transpososome Architectures. *Q Rev Biophysics* (2012) 45(4):493. doi: 10.1017/S0033583512000145
 46. Broecker F, Moelling K. Evolution of Immune Systems From Viruses and Transposable Elements. *Front Microbiol* (2019) 10:51. doi: 10.3389/fmicb.2019.00051
 47. Arthur A, Sherratt D. Dissection of the Transposition Process: A Transposon-Encoded Site-Specific Recombination System. *Mol Gen Genet MGG* (1979) 175(3):267–74. doi: 10.1007/BF00397226
 48. Rice PA, Baker TA. Comparative Architecture of Transposase and Integrase Complexes. *Nat Struct Biol* (2001) 8(4):302–7. doi: 10.1038/86166

49. Jangam D, Feschotte C, Betrán E. Transposable Element Domestication as an Adaptation to Evolutionary Conflicts. *Trends Genet* (2017) 33(11):817–31. doi: 10.1016/j.tig.2017.07.011
50. Mette M, Aufsatz W, van der Winden J, Matzke M, Matzke A. Transcriptional Silencing and Promoter Methylation Triggered by Double-Stranded RNA. *EMBO J* (2000) 19(19):5194–201. doi: 10.1093/emboj/19.19.5194
51. Matzke M, Mette M, Matzke A. Transgene Silencing by the Host Genome Defense: Implications for the Evolution of Epigenetic Control Mechanisms in Plants and Vertebrates. *Plant Mol Biol* (2000) 43(2):401–15. doi: 10.1007/978-94-011-4183-3_20
52. Koonin EV, Makarova KS, Wolf YI, Krupovic M. Evolutionary Entanglement of Mobile Genetic Elements and Host Defence Systems: Guns for Hire. *Nat Rev Genet* (2020) 21(2):119–31. doi: 10.1038/s41576-019-0172-9
53. Sinzelle L, Izsvak Z, Ivics Z. Molecular Domestication of Transposable Elements: From Detrimental Parasites to Useful Host Genes. *Cell Mol Life Sci* (2009) 66(6):1073–93. doi: 10.1007/s00018-009-8376-3
54. Oliver KR, Greene WK. Transposable Elements: Powerful Facilitators of Evolution. *Bioessays* (2009) 31(7):703–14. doi: 10.1002/bies.200800219
55. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* (2012) 337(6096):816–21. doi: 10.1126/science.1225829
56. 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
57. Bischerour J, Bhullar S, Wilkes CD, Régner V, Mathy N, Dubois E, et al. Six Domesticated PiggyBac Transposases Together Carry Out Programmed DNA Elimination in Paramecium. *Elife* (2018) 7:e37927. doi: 10.7554/eLife.37927
58. Maurer-Alcalá XX, Nowacki M. Evolutionary Origins and Impacts of Genome Architecture in Ciliates. *Ann New York Acad Sci* (2019) 1447(1):110. doi: 10.1111/nyas.14108
59. Castillo F, Benmohamed A, Sztamari G. Xer Site Specific Recombination: Double and Single Recombinase Systems. *Front Microbiol* (2017) 8:453. doi: 10.3389/fmicb.2017.00453
60. Midonet C, Das B, Paly E, Barre F-X. XerD-Mediated FtsK-Independent Integration of TLC ϕ Into the Vibrio Cholerae Genome. *Proc Natl Acad Sci* (2014) 111(47):16848–53. doi: 10.1073/pnas.1404047111
61. Luo S, Lu J. Silencing of Transposable Elements by piRNAs in Drosophila: An Evolutionary Perspective. *Genomics Proteomics Bioinf* (2017) 15(3):164–76. doi: 10.1016/j.gpb.2017.01.006
62. Blanco-Melo D, Gifford RJ, Bieniasz PD. Co-Option of an Endogenous Retrovirus Envelope for Host Defense in Hominid Ancestors. *Elife* (2017) 6:e22519. doi: 10.7554/eLife.22519
63. 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. doi: 10.1038/s41586-019-1093-7
64. Zhang Y, Corbett E, Wu S, Schatz DG. Structural Basis for the Activation and Suppression of Transposition During Evolution of the RAG Recombinase. *EMBO J* (2020) 39(21):e105857. doi: 10.15252/embj.2020105857
65. Landau NR, Schatz DG, Rosa M, Baltimore D. Increased Frequency of N-Region Insertion in a Murine Pre-B-Cell Line Infected With a Terminal Deoxynucleotidyl Transferase Retroviral Expression Vector. *Mol Cell Biol* (1987) 7(9):3237–43. doi: 10.1128/mcb.7.9.3237-3243.1987
66. Daggfeldt A, Bengtén E, Pilsström L. A Cluster Type Organization of the Loci of the Immunoglobulin Light Chain in Atlantic Cod (*Gadus Morhua* L.) and Rainbowtrout (*Oncorhynchus Mykiss* Walbaum) Indicated by Nucleotide Sequences of cDNAs and Hybridization Analysis. *Immunogenetics* (1993) 38(3):199–209. doi: 10.1007/BF00211520
67. Hsu E. Assembly and Expression of Shark Ig Genes. *J Immunol* (2016) 196(9):3517–23. doi: 10.4049/jimmunol.1600164
68. Hinds K, Litman G. Major Reorganization of Immunoglobulin VH Segmental Elements During Vertebrate Evolution. *Nature* (1986) 320(6062):546–9. doi: 10.1038/320546a0
69. Kokubu F, Litman R, Shambloft M, Hinds K, Litman G. Diverse Organization of Immunoglobulin VH Gene Loci in a Primitive Vertebrate. *EMBO J* (1988) 7(11):3413–22. doi: 10.1002/j.1460-2075.1988.tb03215.x
70. Lee SS, Fitch D, Flajnik MF, Hsu E. Rearrangement of Immunoglobulin Genes in Shark Germ Cells. *J Exp Med* (2000) 191(10):1637–48. doi: 10.1084/jem.191.10.1637
71. Rumpfelt LL, Avila D, Diaz M, Bartl S, McKinney EC, Flajnik MF. A Shark Antibody Heavy Chain Encoded by a Nonsomatically Rearranged VDJ Is Preferentially Expressed in Early Development and Is Convergent With Mammalian IgG. *Proc Natl Acad Sci* (2001) 98(4):1775–80. doi: 10.1073/pnas.98.4.1775
72. Ventura-Holman T, Lobb CJ. Structural Organization of the Immunoglobulin Heavy Chain Locus in the Channel Catfish: The IgH Locus Represents a Composite of Two Gene Clusters. *Mol Immunol* (2002) 38(7):557–64. doi: 10.1016/S0161-5890(01)00075-X
73. Wang X, Olp JJ, Miller RD. On the Genomics of Immunoglobulins in the Gray, Short-Tailed Opossum *Monodelphis Domestica*. *Immunogenetics* (2009) 61(8):581–96. doi: 10.1007/s00251-009-0385-8
74. Wang X, Miller RD. Recombination, Transcription, and Diversity of a Partially Germline-Joined VH in a Mammal. *Immunogenetics* (2012) 64(9):713–7. doi: 10.1007/s00251-012-0627-z
75. Smith LC, Arizza V, Hudgell MAB, Barone G, Bodnar AG, Buckley KM, et al. Echinodermata: The Complex Immune System in Echinoderms. In: *Advances in Comparative Immunology*. Springer (2018). p. 409–501. doi: 10.1007/978-3-319-76768-0_13
76. Avgan N, Wang JJ, Fernandez-Chamorro J, Weatheritt RJ. Multilayered Control of Exon Acquisition Permits the Emergence of Novel Forms of Regulatory Control. *Genome Biol* (2019) 20(1):1–13. doi: 10.1186/s13059-019-1757-5
77. Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH. Gene Structure in the Sea Urchin *Strongylocentrotus Purpuratus* Based on Transcriptome Analysis. *Genome Res* (2012) 22(10):2079–87. doi: 10.1101/gr.139170.112
78. Smith LC, Lun CM. The SpTransformer Gene Family (Formerly Sp185/333) in the Purple Sea Urchin and the Functional Diversity of the Anti-Pathogen Rsptransformer-E1 Protein. *Front Immunol* (2017) 8:725. doi: 10.3389/fimmu.2017.00725
79. Oren M, Rosental B, Hawley TS, Kim G-Y, Agronin J, Reynolds CR, et al. Individual Sea Urchin Coelomocytes Undergo Somatic Immune Gene Diversification. *Front Immunol* (2019) 10:1298. doi: 10.3389/fimmu.2019.01298
80. Oren M, Hudgell MAB, D'Allura B, Agronin J, Gross A, Podini D, et al. Short Tandem Repeats, Segmental Duplications, Gene Deletion, and Genomic Instability in a Rapidly Diversified Immune Gene Family. *BMC Genomics* (2016) 17(1):1–19. doi: 10.1186/s12864-016-3241-x
81. Miller CA, Buckley KM, Easley RL, Smith LC. An Sp185/333 Gene Cluster From the Purple Sea Urchin and Putative Microsatellite-Mediated Gene Diversification. *BMC Genomics* (2010) 11(1):1–19. doi: 10.1186/1471-2164-11-575

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.

Copyright © 2021 Yakovenko, Agronin, Smith and Oren. 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.