Evolution of RAG transposon unveiled

Xiao Ding[1] & Chen Dong[2*]

[1]*School of Life Sciences, Tsinghua University, Beijing 100084, China;*
[2]*Institute for Immunology and School of Medicine, Tsinghua University, Beijing 100084, China*

One hallmark of the adaptive immune system evolved from jawed vertebrates is the generation of a highly diversified repertoire of antigen receptors. RAG proteins are key recombinase in mammalian lymphoid cells. A recent study has identified a prototype of RAG in Lancelet (Huang et al., 2016).

In non-lymphoid cells, the Ig and TCR genetic loci consist of an array of widely separated variable (V), diversity (D), and joining (J) gene segments. While in mature lymphocytes, functional Ig chains and TCR chains are assembled via a DNA rearrangement process known as V (D) J recombination. Random recombination, imprecise junction and somatic hypermutation contribute to the extraordinary diverse system that recognizes myriad antigens. V (D) J recombination is dependent on the cooperative activity of RAG1 and RAG2. V, D, and J gene segments are flanked by specific sequences named recombination signal sequences (RSS), which contain well-organized heptamer-spacer-nonamer sequence motif, where the spacer is either 12 or 23 bps long. RAG1/2 can introduce double-strand breaks at these RSS sites with the restriction of "12/23 rules", and then leads to the antigen receptor segments to joint via the NHEJ (nonhomologous end-joining) pathway (Schatz and Swanson, 2011).

How did the adaptive immune system arise is a fascinating topic. The evolution of RAG proteins coincided with the appearance of adaptive immune system in jawed vertebrates, which makes the topic shift to asking how RAG emerged. "RAG transposon" hypothesis was first put for-

ward by Sakano and Tonegawa in 1970s, they found the inverted repeat stem structures formed between the 12RSS and 23RSS, which is reminiscent of the terminal inverted repeats (TIRs) in a transposon (Sakano et al., 1979). Indeed, RAG proteins are found to have transposition activity of RSS-flanked DNA *in vitro* (Agrawal et al., 1998). *Transib* transposon family identified in invertebrate have the RAG1-like domain and similar target sites, which is considered to be the ancestor of the core region in RAG1, whereas RAG2 has little sequence similarity to any known transposases, making it elusive how dose RAG2 evolve. A closely linked *RAG1-like* and *RAG2-like* gene pairs were identified in sea urchin but they lack the apparent transposon signatures like TIRs and TSDs (Flajnik and Kasahara, 2010). Thus, there is no definitive evidence to support the RAG transposon hypothesis so far.

In the latest issue of *Cell*, Shengfeng Huang et al. reported the *ProtoRAG* DNA transposon family in Lancelet, a chordate that diverged from vertebrates 550 million years ago (Huang et al., 2016). *ProtoRAG* transposon encodes intron-containing *RAG1-like* and *RAG2-like* gene pair flanked by 5-bp TSDs, which is the characteristic of *Transib* transposon family. Like RSS in vertebrate RAG, the TIRs in *ProtoRAG* transposon contain consensus heptamer and nonamer sequences, and the heptamer sequence resembles the RSS heptamer. At protein level, bbRAG1L shares 29% sequence identity with vertebrate RAG1 and acidic residues at active sites are conserved. However, unlike that in sea urchin and vertebrate, bbRAG2L in Lancelet lacks the entire C-terminal PHD domain.

Using a nick-hairpin mechanism, vertebrate RAG1/2 proteins perform the endonuclease activity on 12RSS/

*Corresponding author (email: chendong@tsinghua.edu.cn)

23RSS substrates during V (D) J recombination *in vivo*, and the cleaved RSSs are joined precisely. By transfecting bbRAG1L/2L with artificial transposon flanking by *ProtoRAG* TIRs to 293T cells, the authors found bbRAGL1/L2 can perform TIR-specific endonuclease activity *ex vivo* and generate host DNA joints with small deletions or short insertions. Furthermore, bbRAGL1/L2 is capable to mediate intermolecular transposition in human cells, and generate precise TIR-TIR joints (TTJs) although at a very low frequency. *In vitro* cleavage study revealed that bbRAG1L/2L applies the similar nick-hairpin mechanism, which is also dependent on HMGB1 protein and divalent cations. Notably, bbRAGL1/L2 can't recognize RSS sites and vertebrate RAG can't recognize *ProtoRAG* TIRs either, indicating that these two RAG machineries are no longer compatible during evolution.

This study discovered that *ProtoRAG* in Lancelet is the long-sought RAG transposon, which provided strong evidence for the transposon origin of RAG. However, more questions remain unanswered. For example, as indicated by the authors themselves, *RAG1/2-like* genes in *ProtoRAG* transposon are interrupted by introns, which is quite distinctive from typical transposon genes, while vertebrate RAG genes have no introns. Up to now, no *RAG1/2-like* genes are identified in jawless vertebrates. How RAG transposon was domesticated by jawed vertebrate is still elusive.

Another important question is RAG2's origin. Vertebrate RAG2 consists of six Kelch-like repeats and a PHD domain, which lacks similarity to any known transposons. So far in all cases *RAG2-like* gene is always found to link with *RAG1-like* gene, no individual *RAG2-like* gene is identified. The *ProtoRAG* gene discovered by Shengfeng Huang et al. provided the evidence that *RAG2-like* gene is linked with *RAG1-like* gene within the RAG transposon, which supports the concept that *RAG2* was an integral component of an ancestral Rag transposon. However, bbRAG2L protein lacks the C-terminal PHD domain, which is reported to suppress the transposase activity (Elkin et al., 2003; Tsai and Schatz, 2003). This is consistent with the data that bbRAG1L/2L prefers to mediate transposition rather than TIR-TIR joints *in vivo*, indicating the possibility that the acquirement of PHD domain in vertebrate RAG2 protein may switch RAG function from transposition to recombination.

A recent paper reported RAG1 can mediate V (D) J recombination in the absence of RAG2, however, recombination by RAG1 alone does not follow the "12/23 rules". More interestingly, *Transib* and RAG1-like protein in purple sea urchin can mediate 12/23-dependent DNA recombination with the help of RAG2 (Carmona et al., 2016), indi-
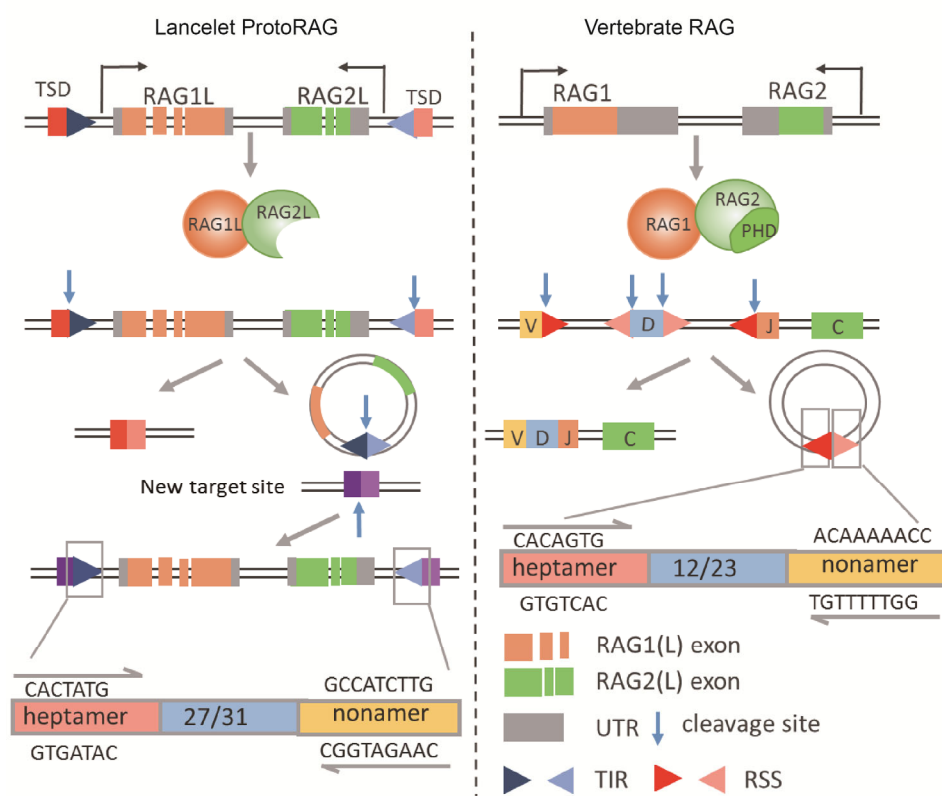

Figure 1   Structure and function of Lancelet ProtoRAG and vertebrate RAG. bbRAG1L/2L in Lancelet is encoded by tail-to-tail oriented, intron-containing *RAG1-like* and *RAG2-like* genes, flanked by a pair of terminal inverted repeats (TIRs). bbRAG1L/2L protein lacks the C-terminal PHD domain in RAG2 and prefers to mediate TIR-dependent DNA cleavage and transposition. The heptamer sequence recognized by bbRAG1L/2L resembles the counterpart in vertebrate RSS while the nonamers are not related. ProtoRAG thus represents a molecular ''living fossil'' of the long-sought RAG transposon.

cating that RAG2 is required for establishing the "12/23 rules". Whether exchange the *RAG2-like* gene in *ProtoRAG* with vertebrate *RAG2* can mediate more efficient DNA recombination and TIR-TIR joints in a 12/23-dependent manner is an open question.

ProtoRAG represents a molecular ''living fossil'' of the long-sought RAG transposon, which broadens our understanding of adaptive immune system evolution. Nonetheless, there are still some fantastic questions about RAG transposon domestication, such as the evolution gap in jawless vertebrates, the acquirement of PHD domain and the origin and the function of RAG2 protein. Follow-up studies will likely provide new vision of RAG origination and adaptive immunity generation.

Compliance and ethics   *The author(s) declare that they have no conflict of interest.*

Huang, S., Zhang, Y., Tao, X., Yuan, S., Li, P., Beilinson, H.A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H., Petillon, L.Y., Liu, X., Chen, S., Schatz, D.G., and Xu, A. (2016). Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. Cell 166, 102–114.

Schatz, D.G. and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu Rev Genet 45, 167–202.

Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S. (1979). Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature 280, 288–294.

Agrawal, A., Eastman, Q.M., and Schatz, D.G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394, 744–751.

Flajnik, M.F. and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11, 47–59.

Elkin, S.K., Matthews, A.G., and Oettinger, M.A. (2003). The C-terminal portion of RAG2 protects against transposition *in vitro*. EMBO J 22, 1931–1938.

Tsai, C.L. and Schatz, D.G. (2003). Regulation of *RAG1/RAG2*-mediated transposition by GTP and the C-terminal region of *RAG2*. EMBO J 22, 1922–1930.

Carmona, L.M., Fugmann, S.D., and Schatz, D.G. (2016). Collaboration of *RAG2* with *RAG1*-like proteins during the evolution of V(D)J recombination. Genes Dev 30, 909–917.