

Lab Note

Development of an *in vitro* TurbolD labeling assay for the detection of protoRAG-interacting proteins in the amphioxus

Qingyi Lu^{1,†}, Jie Xu^{1,†}, Junye Hong^{2,†}, Enfan Xiao¹, Qiuzhu Wei¹, Yuhe Sun¹, Zihan Zhao¹, Yuhang Zhang^{2,*}, and Guangrui Huang^{1,*}

¹School of Life Sciences, Beijing University of Chinese Medicine, Beijing 102488, China, and ²School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

[†]These authors contributed equally to this work.

*Correspondence address. E-mail: yuhang.zhang@sjtu.edu.cn (Y.Z.); hgr@bucm.edu.cn (G.H.)

Received 1 October 2025 Accepted 9 January 2026 Published 19 March 2026

The recombination-activating gene (RAG)-mediated V(D)J rearrangement mechanism is essential for generating the diversity of the adaptive immune system in jawed vertebrates. The RAG1 and RAG2 proteins form a tetrameric complex that recognizes and cleaves pairs of recombination signal sequences (RSSs) [1]. In jawed vertebrates, the N-terminal region of RAG1 and the C-terminal region of RAG2 are critical for the regulation of physiological function and enhancement of recombination accuracy [2]. Moreover, the intracellular activity of RAG1/2 is rigorously regulated and is thought to represent the aspect shaped by host coevolution. However, these regulatory mechanisms remain incompletely understood, impeding our deeper understanding of the physiological functions of RAG.

Previous studies have shown that jawed vertebrate RAG evolved from invertebrate RAG-like (RAGL) transposases and underwent key structural changes during evolution [3,4]. RAGL genes have been identified in multiple invertebrate species, but from *Branchiostoma belcheri* (Bb), the RAG-like protein in Bb amphioxus (BbRAG1L) and the RAG2-like protein in Bb amphioxus (BbRAG2L) are among the most extensively characterized to date. Lancelets (amphioxus) represent the most basal extant chordates (cephalo-chordates) that diverged from the other two chordate lineages (urochordates and vertebrates). Consequently, amphioxus has been considered a “living fossil” for studying the origins and evolution of adaptive immunity [5]. Homologs of the core structural domain of RAG1 were discovered in the amphioxus genome in 2005 [6], and the homologs were subsequently shown to exhibit recombinase activity in cooperation with murine-derived RAG2 [7]. A large number of transposon genes were then identified in the amphioxus genome [8], followed by the identification of target site duplication (TSD) and terminal inverted repeat (TIR) sequences in the amphioxus genome [5]. Notably, BbRAG2L lacks the C-terminal

region present in murine RAG2, including the acidic hinge and PHD domain. BbRAG1L coupled with BbRAG2L displays significant transposition activity *in vitro* and in human cell lines [5]. In contrast, vertebrate RAG exhibits limited transposon activity *in vitro* and extremely low transposon activity *in vivo* [2]. This leads to the question of how RAG lost its transposase activity during evolution. Structural comparison and functional experiments of BbRAGL (protoRAG) and vertebrate RAG have revealed that vertebrates undergo essential adaptive changes to eliminate the primitive transposase activity of protoRAG [2]. Research on BbRAGL has therefore provided strong support for the hypothesis that RAG originated from a domesticated transposon.

Given the similarity in both structure and function between BbRAG1L and vertebrate RAG and the fact that BbRAGL proteins are the only RAGL proteins that can currently be efficiently expressed in vertebrate cells, characterizing the binding partners of BbRAGL involved in amphioxus is of great significance for understanding the regulation of BbRAGL and its adaptation and evolution. Parallel comparison of BbRAGL and RAG interaction networks may further clarify how BbRAGL transitioned from a transposase to a recombinase during host coevolution and reveal general principles of vertebrate intracellular RAG regulation.

In recent years, proximity labeling techniques have greatly facilitated approaches to identifying RAG cofactors [9]. For example, RAG1 fused to biotin ligase RAG1 fused to a biotin ligase found in *Escherichia coli* (BirM) has been used to define dynamic interaction networks of RAG-associated proteins [9]. Inspired by these advances, we sought to apply a similar approach to identify the BbRAGL cofactors in amphioxus and build a regulatory network of BbRAGL that could shed light on its relationship with host evolution. However, endogenous BbRAGL expression is undetectable in the adult amphioxus (data not shown), and the generation of

transgenic amphioxus lines expressing proximity-based labeling proteins remains technically challenging. These constraints make it difficult to analyze native cofactors of endogenous BbRAGL *in vivo*. Because BbRAGL can be expressed in vertebrate cell lines, we instead developed an *in vitro* proximity-labeling strategy. We fused TurboID (TbID), an engineered biotin ligase, to the N-terminus of BbRAG1L or BbRAG2L (Figure 1A), purified the fusion proteins, and incubated them with protein lysates prepared from the hepatic cecum and colons of adult *B. belcheri* amphioxus *in vitro* to label the interacting proteins. The proteins were finally analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Figure 1B).

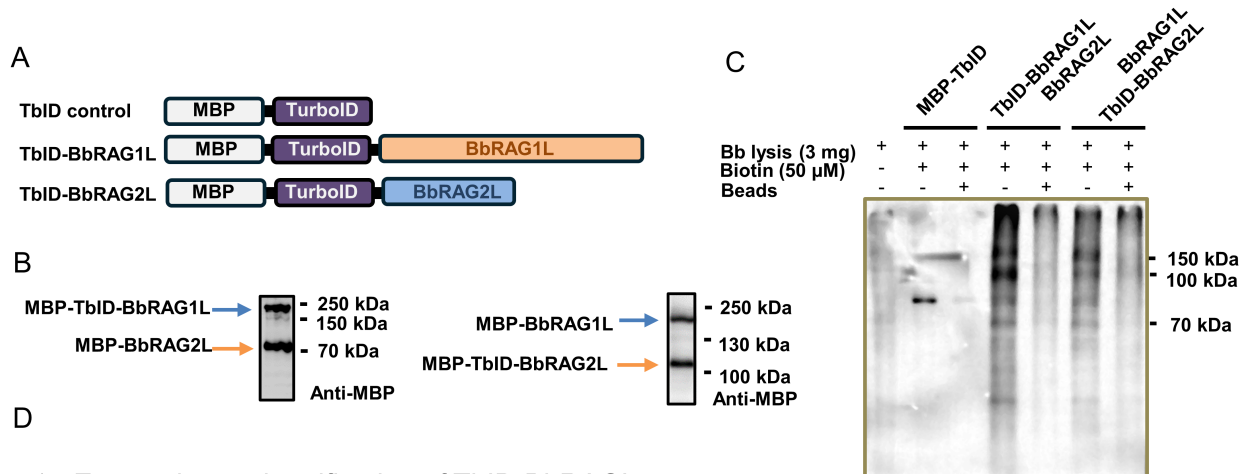
A pTT5 vector was used to express plasmids encoding maltose-binding protein (MBP) and TurboID with or without full-length BbRAG1L or BbRAG2L [2,5], yielding pTT5-MBP-TurboID, pTT5-MBP-TurboID-BbRAG1L, and pTT5-MBP-TurboID-BbRAG2L. Plasmids expressing MBP-BbRAG1L and MBP-BbRAG2L were also generated. Expi293F cells were cultured in serum-free medium (P82019; OPM Biosciences, Shanghai, China) in a humidified incubator with 8% CO₂ at 37°C and shaking at 120 rpm. pTT5-MBP-TurboID was transfected with PEI-MAX transfection reagent (24765-1; Polysciences, Warrington, USA) at a plasmid-to-PEI ratio of 1:3 (w/w). pTT5-MBP-TurboID-BbRAG1L was cotransfected with pTT5-MBP-BbRAG2L, or pTT5-MBP-TurboID-BbRAG2L was cotransfected with pTT5-MBP-BbRAG1L for the purpose of labeling different proteins to BbRAG1L or BbRAG2L. A total of 1–2 µg of plasmid DNA was added per milliliter of culture. Protein purification was performed as described previously [2]. After 3 days of culture, cells were collected by centrifugation (500 g) and frozen at –80°C. The cell pellets were resuspended in lysis buffer (25 mM Tris, pH 7.5, 1 M KCl, and 2 mM DTT) and disrupted by sonication, followed by centrifugation at 230,000 g for 1 h at 4°C (all subsequent steps were conducted at 4°C). The supernatants were mixed with pre-equilibrated dextrin beads (SA077100; Smart-Life Sciences, Changzhou, China) for 2 h with gentle rotation and then loaded onto a gravity-flow column. The beads were washed with 50–100 mL of lysis buffer, and bound proteins were eluted with 5 mL of elution buffer (25 mM Tris, pH 7.5, 1 M KCl, 2 mM DTT, and 80 mM maltose). The MBP-TurboID took more than 10 mL of elution buffer. The eluates were further purified by size-exclusion chromatography (SEC) on a Superdex 200 Increase 10/300 GL column equilibrated in 20 mM HEPES (pH 7.5), 0.5 M NaCl, and 2 mM DTT.

During the preparation of amphioxus lysates, we dissociated the hepatic cecum and colons under a stereomicroscope. Their tissues were sonicated in TAP lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.5% NP40, 5% glycerol, and 1.5 mM MgCl₂). After centrifugation (12,000 g) at 4°C, the supernatants were collected and quantified with a Pierce BCA protein assay kit (A65453; Thermo Fisher Scientific, Waltham, USA).

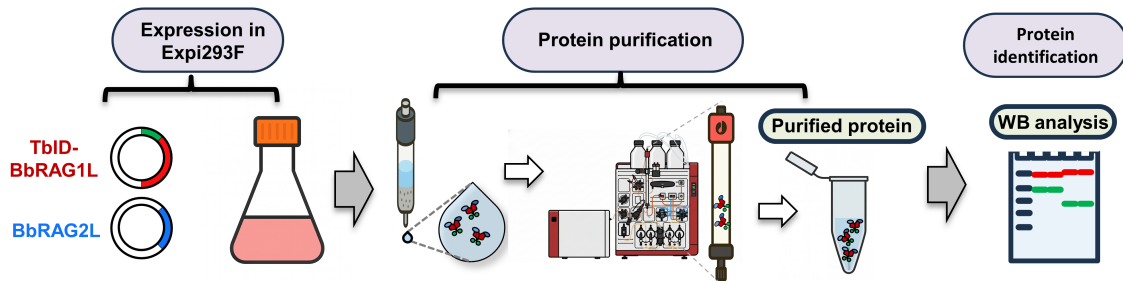
To label the BbRAGL-interacting proteins, amphioxus lysates (approximately 3 mg for each sample) were incubated with purified TurboID-tagged BbRAG1L or TurboID-tagged BbRAG2L proteins (approximately 100 µg for each sample) in 1 mL of reaction buffer (0.5 µM biotin, 20 mM Tris, pH 7.5, 1 mM ATP, and 5 mM MgCl₂) in a 1.5-mL EP tube at 37°C for 10 min with rotation. Reactions were stopped by placing the centrifuge tubes on ice. The lysates were transferred to 15-mL centrifuge tubes and incubated with streptavidin agarose beads (SM007005; Smart Life Sciences) for 3 h at 4°C with rotation to enrich the labeled proteins. Beads were washed

three times at room temperature with washing buffer (20 mM Tris, pH 7.5, and 3 M urea). A small fraction of the supernatants was retained to validate the biotinylation efficiency by western blot analysis (Figure 1C). For on-bead digestion, the washed beads were incubated in 10 mM DTT prepared in 25 mM NH₄HCO₃ solution for 1 h at 37°C, followed by alkylation with 25 mM iodoacetamide for 20 min in the dark. After overnight tryptic digestion, the released peptides were desalted and purified using Pierce C18 Spin Tips and then reconstituted in 0.1% formic acid for LC-MS/MS analysis (Figure 1D). Mass spectrometry (MS)-based proteomics was performed using an Easy-nLC 1200 system coupled to a Q Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific) under standard conditions. Peptides were separated on a C18 reversed-phase analytical column (75 µm × 20 cm, 3 µm particle size) at a flow rate of 300 nL/min. Mobile phase A and phase B consisted of 0.1% formic acid soluted in water and 80% acetonitrile water solution, respectively. The mass spectrometer was operated in positive electrospray ionization (ESI+) mode using a data-dependent acquisition (DDA) strategy. Full MS scans and MS/MS scans were acquired in the Orbitrap. Raw data were processed using Proteome Discoverer software (version 3.0; Thermo Fisher Scientific). Spectra were searched against the amphioxus UniProt database. Protein abundances were normalized using a protein abundance-based normalization method. Statistical analysis of quantitative differences was performed using ANOVA. From three independent biological replicates, we identified proteins that were significantly enriched (adjusted *P* value < 0.05, fold change > 2) in the BbRAG1L and BbRAG2L groups relative to the TurboID control (Supplementary Table 1,2). In addition, 84 proteins in the BbRAG1L group and 84 in the BbRAG2L group were detected exclusively in TurboID-tagged BbRAGL samples but not in TurboID samples. The BbRAG1L and BbRAG2L groups shared 186 common proteins that were in proximal contact with BbRAG1L or BbRAG2L (Figure 2A). To map these amphioxus proteins to their vertebrate counterparts, we used BLAST to compare the 301 amphioxus protein sequences against the *Homo sapiens* database and identified 290 unique homologs (Figure 2B). Among these, 182 proteins were shared by the BbRAG1L and BbRAG2L groups (Figure 2C).

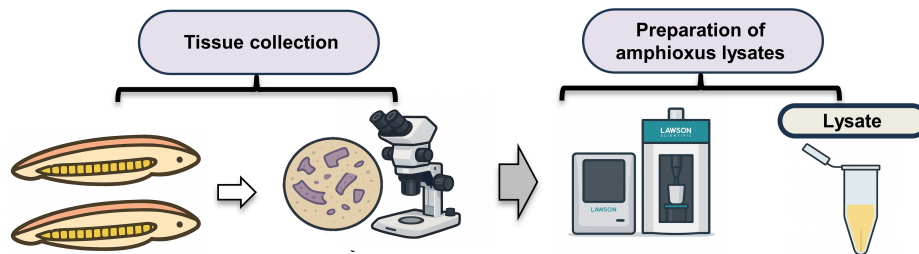
To delineate the potential functions of BbRAG1L/2L-interacting proteins, we performed Gene Ontology (GO) enrichment analyses using the clusterProfiler R package. False discovery rate (FDR) values were calculated using the Benjamini-Hochberg correction method to adjust for multiple hypothesis testing. GO terms with an adjusted *P* < 0.05 were considered statistically significant in this study. Enrichment results were visualized using the ggplot2 R package in R 4.4. In the biological process (BP), cellular component (CC) and molecular function (MF) categories, the top 20 terms are shown in Figure 2D, Figure 2E, and Figure 2F, respectively. Notably, “DNA ligation”, “V(D)J recombination” and “ligase activity” were significantly enriched, exhibiting substantial overlap with known clusters of RAG-interacting proteins. In particular, the identification of protein homologs associated with “DNA ligation” and “V(D)J recombination” indicates that RAGL had already established connections with proteins possessing this function even before RAGL acquired catalytic activity in V(D)J recombination. However, interpretation of the GO cellular component enrichment results should take the methodological limitations into full consideration. As shown in Figure 2E, we observed significant enrichment of nuclear-associated compartments (including the known RAG-associated proteins TOP1



1 - Expression and purification of TbID-BbRAGL



2 - Preparation of amphioxus cell lysate



3 - Interactome protein identification

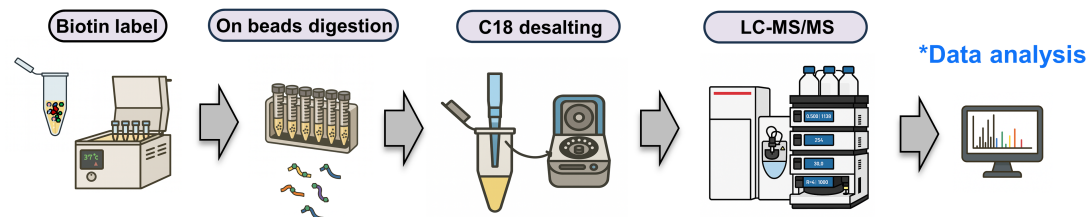


Figure 1. Workflow of the *in vitro* TurboID-based BbRAGL interactome assay (A) Schematic representation of the constructs used in the TurboID labeling assay, including MBP-TurboID (TbID control), MBP-TurboID-BbRAG1L, and MBP-TurboID-BbRAG2L fusion proteins. (B) Western blot analysis confirming the expression of BbRAG1L, BbRAG2L, and their respective TurboID fusion proteins in Expi293F cells. Anti-MBP antibodies were used to detect MBP-tagged proteins, with molecular-weight shifts indicating successful fusion to TurboID and BbRAGL. (C) Validation of *in vitro* TurboID labeling of interacting proteins in Bb lysates. TurboID-BbRAG1L and TurboID-BbRAG2L efficiently biotinylate proximal proteins in amphioxus lysates. Upon incubation with streptavidin beads, biotin-labeled proteins are selectively enriched on the beads, as shown by streptavidin blotting. (D) Stepwise workflow for the *in vitro* BbRAGL interactome assay. The procedure includes (1) expression and purification of TurboID-BbRAG1L and TurboID-BbRAG2L in Expi293F cells; (2) preparation of amphioxus tissue lysates from hepatic cecum and colons; and (3) proximity-dependent biotin labeling, on-bead tryptic digestion, peptide desalting, and LC-MS/MS analysis for identification of BbRAGL-interacting proteins.

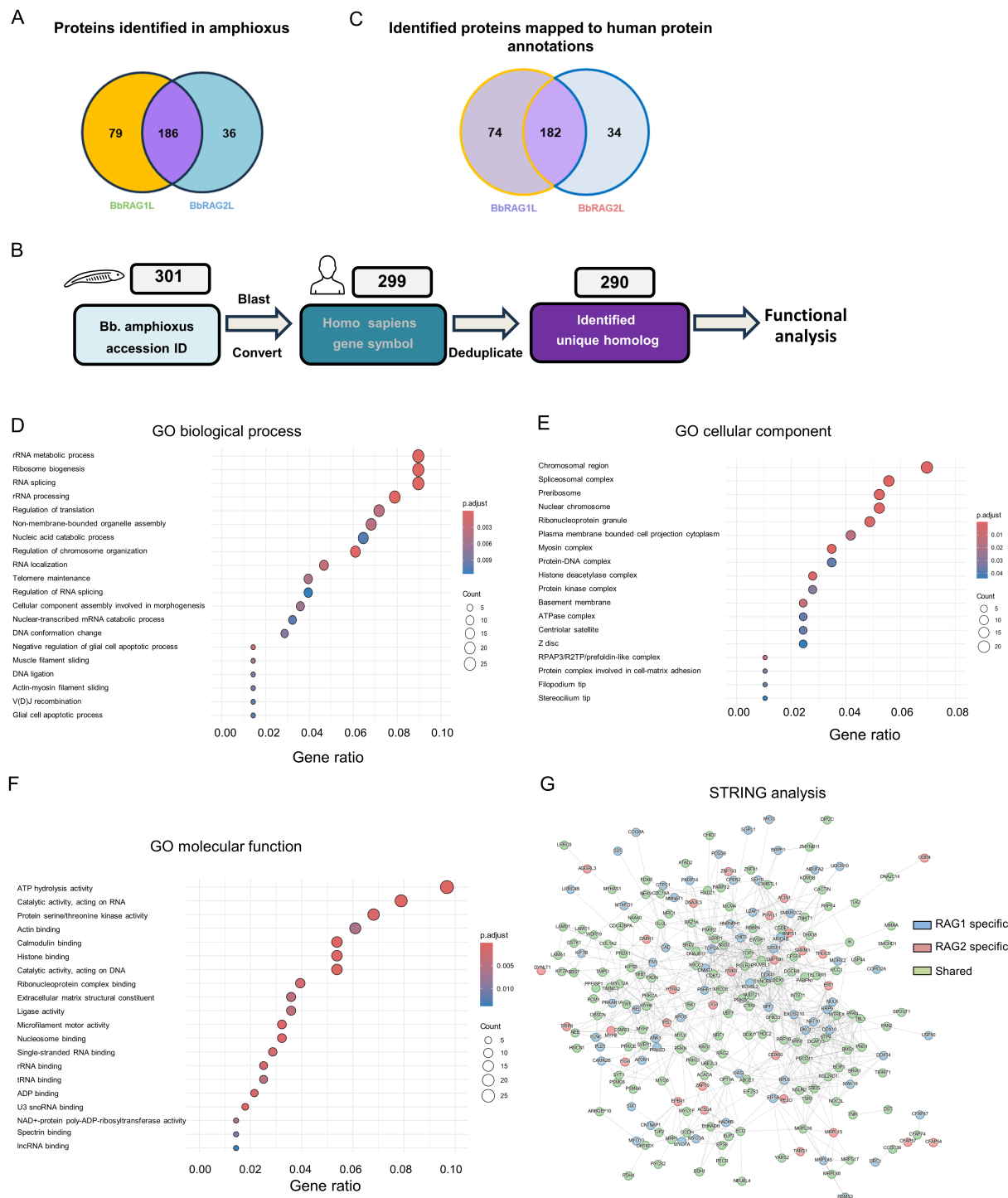


Figure 2. Identification of the functional annotations of amphioxus interactome proteins and their functional characterization (A) Venn diagram showing the overlap of amphioxus proteins biotinylated by BbRAG1L and BbRAG2L in the *in vitro* TurboID labeling assay. A total of 301 proteins were identified across the two groups, with 186 shared proteins and 79 or 36 uniquely associated with BbRAG1L or BbRAG2L, respectively. (B) Schematic workflow illustrating the conversion of amphioxus protein accession IDs into human gene symbols using BLAST, followed by deduplication to obtain 290 unique human homologs for downstream functional analysis. (C) Venn diagram of human homologs corresponding to BbRAG1L- and BbRAG2L-associated amphioxus proteins. Among the mapped homologs, 182 were shared between both groups, whereas 74 and 34 were uniquely associated with BbRAG1L and BbRAG2L, respectively. (D–F) GO enrichment analysis of BbRAG1L- and BbRAG2L-associated proteins identified by proximity labeling, categorized into Biological Process (D), Cellular Component (E), and Molecular Function (F). The top 20 significantly enriched terms are shown for each category. Dot size indicates the number of proteins associated with each term, and color reflects the adjusted *P* value. The x-axis shows the GeneRatio. (G) STRING-based protein–protein interaction network of BbRAG1L- and BbRAG2L-interacting proteins. Nodes are color-coded as BbRAG1L-specific (blue), BbRAG2L-specific (red), or shared (green). Only proteins with at least one detected interaction are displayed.

and PARP1) [9], as well as enrichment of cytoplasmic terms (e.g., “plasma membrane-bounded cell projection cytoplasm”, including FMR1). Because TurboID labeling was performed in amphioxus lysates, cellular compartmentalization cannot be preserved; thus, these GO annotations primarily reflect the known subcellular localizations of the identified proteins rather than subcellular localization features acquired by BbRAGL through their interactions.

We also analyzed the identified proteins using STRING (Figure 2G). Among the 290 candidate proteins, 237 showed at least one interaction with other identified proteins, and clear clustering patterns were observed. These results indicate that, even as an exogenous transposon, BbRAGL has already established extensive connections with host cellular pathways. These observations further suggest that the V(D)J recombination function mediated by RAG was probably formed after the ancestral RAGL proteins had adapted to the host and developed extensive interaction networks with the host. In particular, BbRAG2L contains only the core domain compared with mouse RAG2, lacking two key functional modules: the acidic hinge and the PHD domain. This observation implies that both the C-terminal regulatory functions of RAG2 in directing RAG genomic targeting [10] and the transposition-suppressive role of the acidic hinge [2] were acquired in an evolution-dependent manner. The physiological role of RAG in vertebrates is therefore unlikely to emerge abruptly. Instead, it is more likely shaped by continuous minor adjustments in the functions of related proteins or domains during coevolution.

In summary, we established an *in vitro* TurboID-based workflow to identify RAGL-interacting proteins in invertebrates and obtained a substantial set of candidate interactors in amphioxus. Although this strategy cannot fully preserve the influence of spatial protein distribution on their interactions with BbRAGL, to our knowledge, this study provides the first interaction networks of the invertebrate proteins BbRAG1L and BbRAG2L. Mapping the BbRAGL interactome not only sheds light on the molecular mechanisms underlying the co-evolution of RAG and host cells but also provides an important reference for understanding the fundamental principles that shape antibody-based adaptive immunity. This study lays an important foundation for future in-depth investigations of RAGL function in amphioxus and for reconstructing how RAG acquired its role in adaptive immunity during evolution.

Supplementary Data

Supplementary data are available at *Acta Biochimica et Biophysica*

Sinica online.

Data Availability

The raw proteomic data reported in this paper have been deposited in the Open Archive for Miscellaneous Data (OMIX) under OMIX ID OMIX014726.

Funding

This work was supported by the grant from the National Key R&D Program of China (No. 2022YFE0132100).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Teng G, Schatz DG. Regulation and evolution of the RAG recombinase. *Adv Immunol* 2015, 128: 1–39
2. Zhang Y, Cheng TC, Huang G, Lu Q, Surleac MD, Mandell JD, Pontarotti P, *et al.* Transposon molecular domestication and the evolution of the RAG recombinase. *Nature* 2019, 569: 79–84
3. Liu C, Zhang Y, Liu CC, Schatz DG. Structural insights into the evolution of the RAG recombinase. *Nat Rev Immunol* 2022, 22: 353–370
4. 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: 909–917
5. Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, Zhang Y, *et al.* Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell* 2016, 166: 102–114
6. Kapitonov VV, Jurka J, Nemazee D. RAG1 core and V(D)J recombination signal sequences were derived from transib transposons. *PLoS Biol* 2005, 3: e181
7. Zhang Y, Xu K, Deng A, Fu X, Xu A, Liu X. An amphioxus RAG1-like DNA fragment encodes a functional central domain of vertebrate core RAG1. *Proc Natl Acad Sci USA* 2014, 111: 397–402
8. Huang S, Chen Z, Yan X, Yu T, Huang G, Yan Q, Pontarotti PA, *et al.* Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat Commun* 2014, 5: 5896
9. Brecht RM, Liu CC, Beilinson HA, Khitun A, Slavoff SA, Schatz DG. Nucleolar localization of RAG1 modulates V(D)J recombination activity. *Proc Natl Acad Sci USA* 2020, 117: 4300–4309
10. Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. The *in vivo* pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* 2010, 141: 419–431