Supporting Information (SI) Appendix

Classification: Biological Sciences, Immunology, Evolution

Title: An amphioxus RAG1-like DNA fragment encodes a functional central domain of vertebrate core RAG1

Yanni Zhang[a,1], Ke Xu[a,1], Anqi Deng[a], Xing Fu[a], Anlong Xu[b,2] and Xiaolong Liu[a,2]

[a]State Key Laboratory of Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

[b]State Key Laboratory of Biocontrol, Department of Biochemistry, College of Life Sciences, Sun Yat-Sen (Zhongshan) University, Guangzhou 610275, China

[1]These authors contributed equally to this work

[2]Correspondence should be addressed to X.L. (liux@sibs.ac.cn) and A.X. (lssxal@mail.sysu.edu.cn)

Supporting Information Appendix

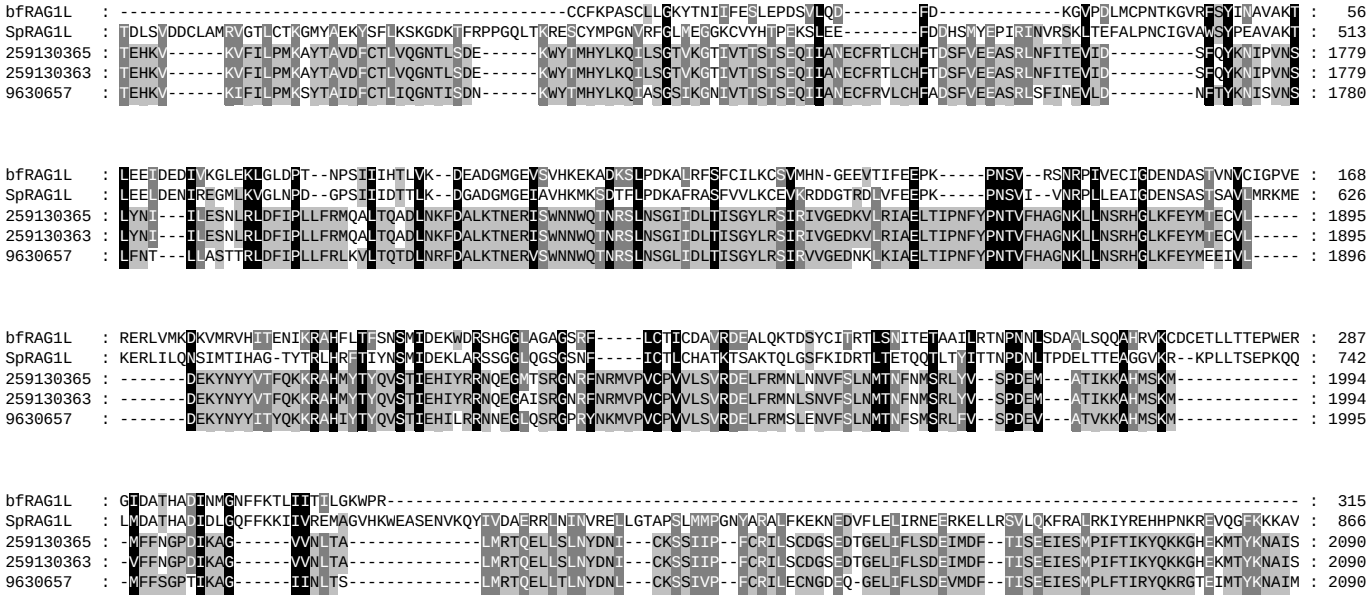


Fig. S1. Multiple alignment between bfRAG1L and virus proteins. Amino acids are shaded using conserved properties according to the color scheme in Genedoc. The GenBank GI numbers of proteins are indicated on the left side. Positions of amino acids of the corresponding proteins are indicated in the right side.

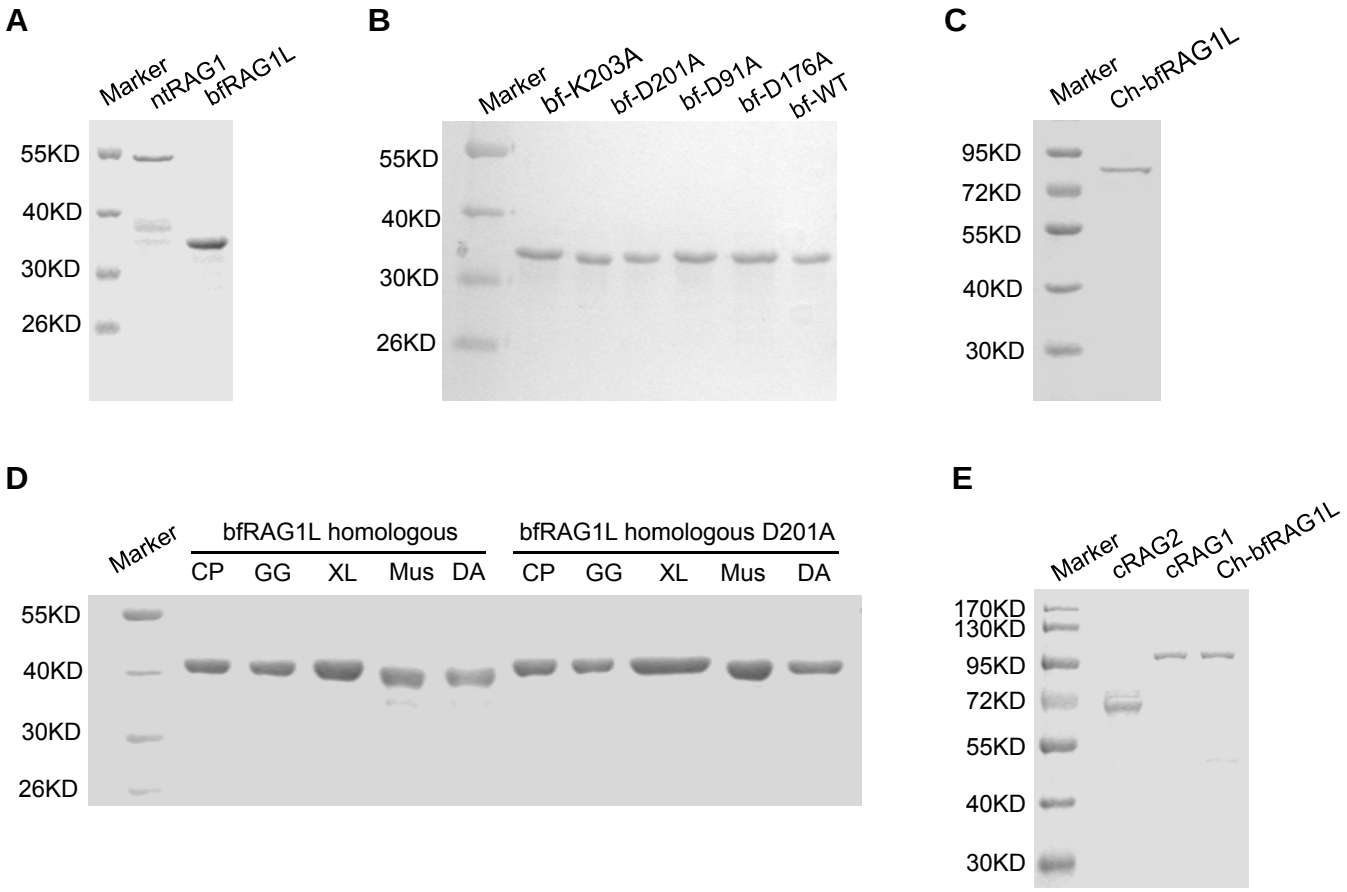


Fig. S2. Protein purification. Purification of bfRAG1L and ntRAG1 (*A*), point mutants of (*B*), Ch-bfRAG1L (*C*) or bfRAG1L-homologous or bfRAG1L-homologous D201A mutants (*D*) in other species expressing in *E.coli* strain BL21(DE3). (*E*) Protein purification of GST-cRAG2, cRAG1 and Ch-bfRAG1L individually expressing in HEK-293T cells. The purified proteins were determined by SDS/PAGE and Coomassie staining. CP, bull shark; DA, zebrafish; GG, chicken; Mus, mouse; XL, frog. Marker, the molecular weight of proteins are indicated. The results are representative of three (*A*-*C*) or two (*D* and *E*) independent experiments.

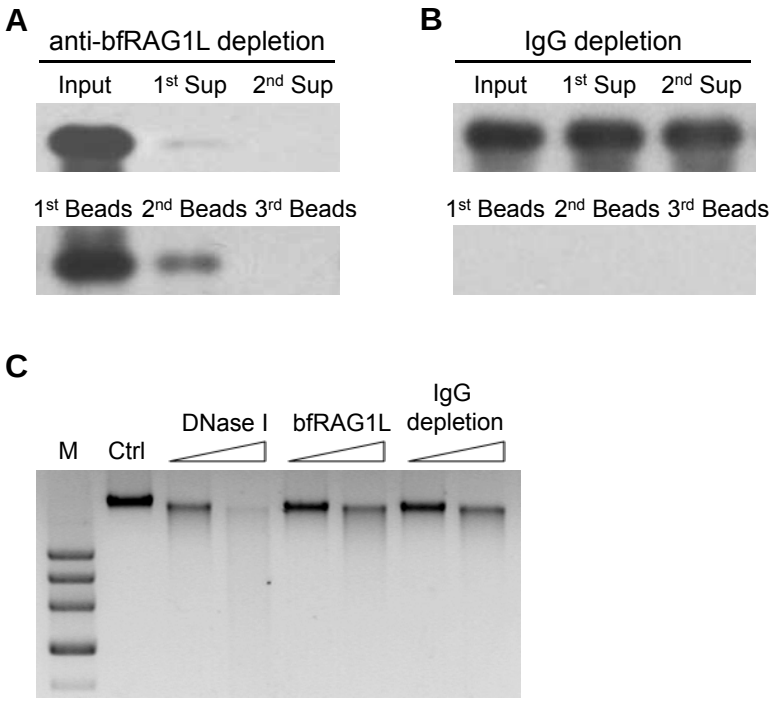


Fig. S3. Immunodepletion of bfRAG1L. Immunodepletion of bfRAG1L with anti-bfRAG1L serum (*A*) or mouse IgG (*B*). The supernatant (Sup) and beads were analyzed by SDS-PAGE and immunoblotted with anti-bfRAG1L antibody. Supernatant, purified bfRAG1L protein immunodepleted with anti-bfRAG1L serum. (*C*) bfRAG1L and mouse IgG-depleted bfRAG1L were used for detecting the nuclease activity. Ctrl, untreated genomic DNA; M, DNA marker. The results are representative of two independent experiments.

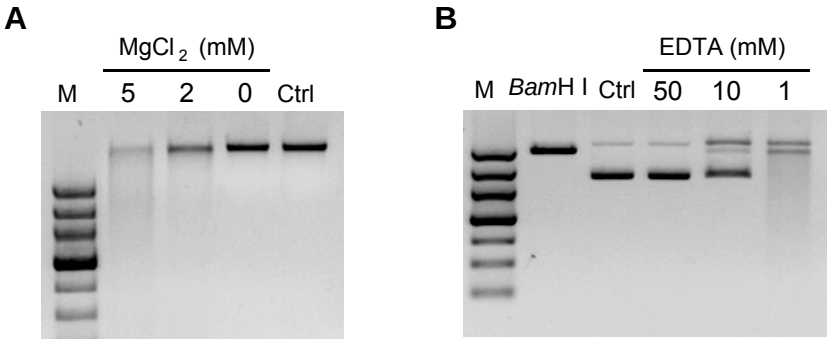


Fig. S4. The nuclease activity of bfRAG1L is metal ion-dependent. The nuclease activities of bfRAG1L on substrates genomic DNA from HEK-293T cells (*A*) or plasmid pBR322 (*B*) in the presence of Mg^{2+} or EDTA are shown. Ctrl, untreated genomic DNA (*A*) or untreated plasmid pBR322 (*B*). M, DNA marker. The results are representative of four (*A*) or three (*B*) independent experiments.

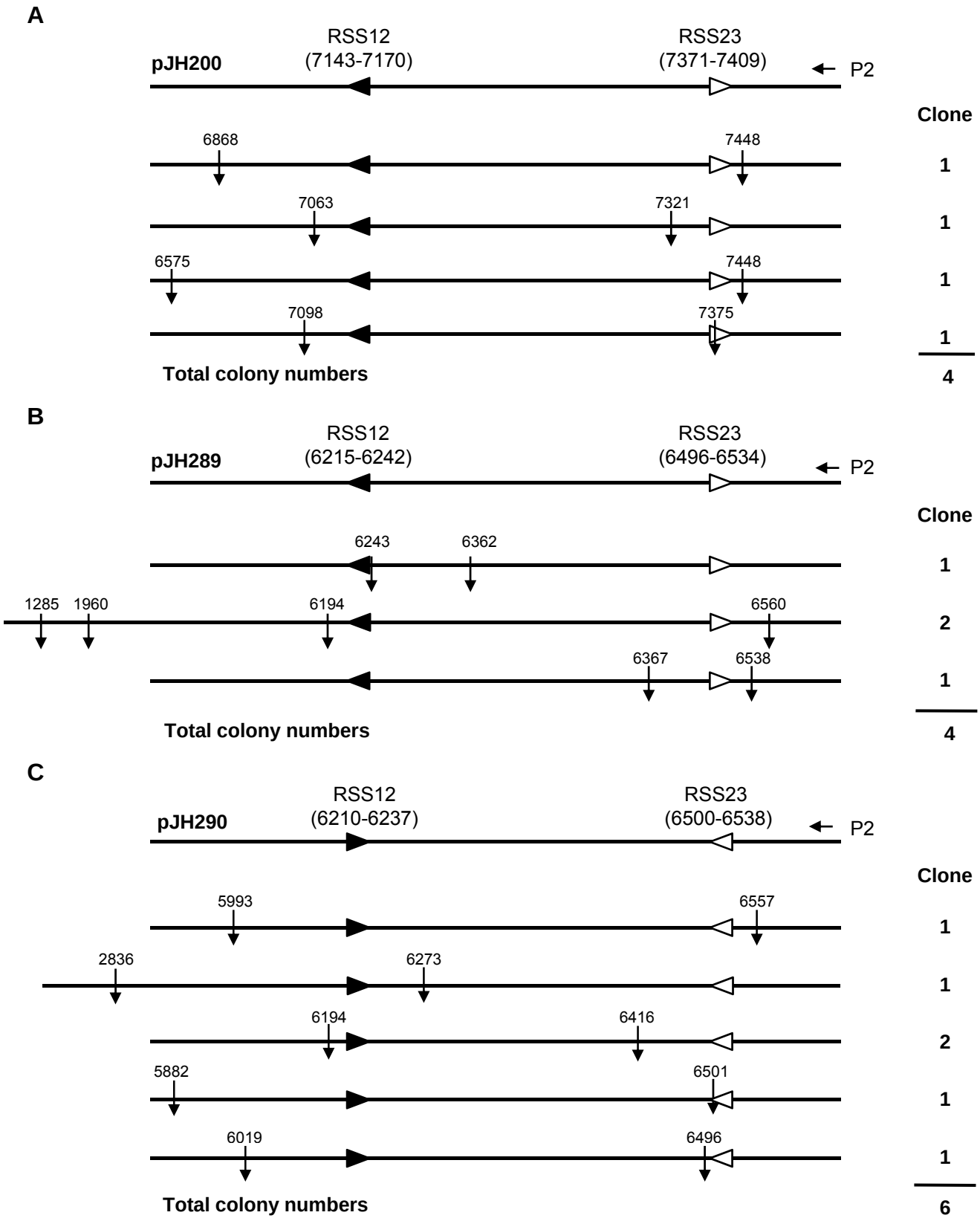


Fig. S5. Bacterial colony assay for analysis of bfRAG1L-introduced cleavage sites within plasmid substrates in the presence of mouse RAG2 in 293T cells. Schematic diagram of plasmid substrates pJH200 (*A*), pJH289 (*B*) or pJH290 (*C*). 12-RSS and 23-RSS are indicated as the black triangle and white triangle, respectively. P2 primer are used for sequencing (*SI Appendix*, Table S6). The downward arrows reflect the cut positions of nucleic acids with numbers as indicated. The sequences of all the clones on the LB agar plate were analyzed from three independent experiments. The total colony numbers are indicated below the sequences.

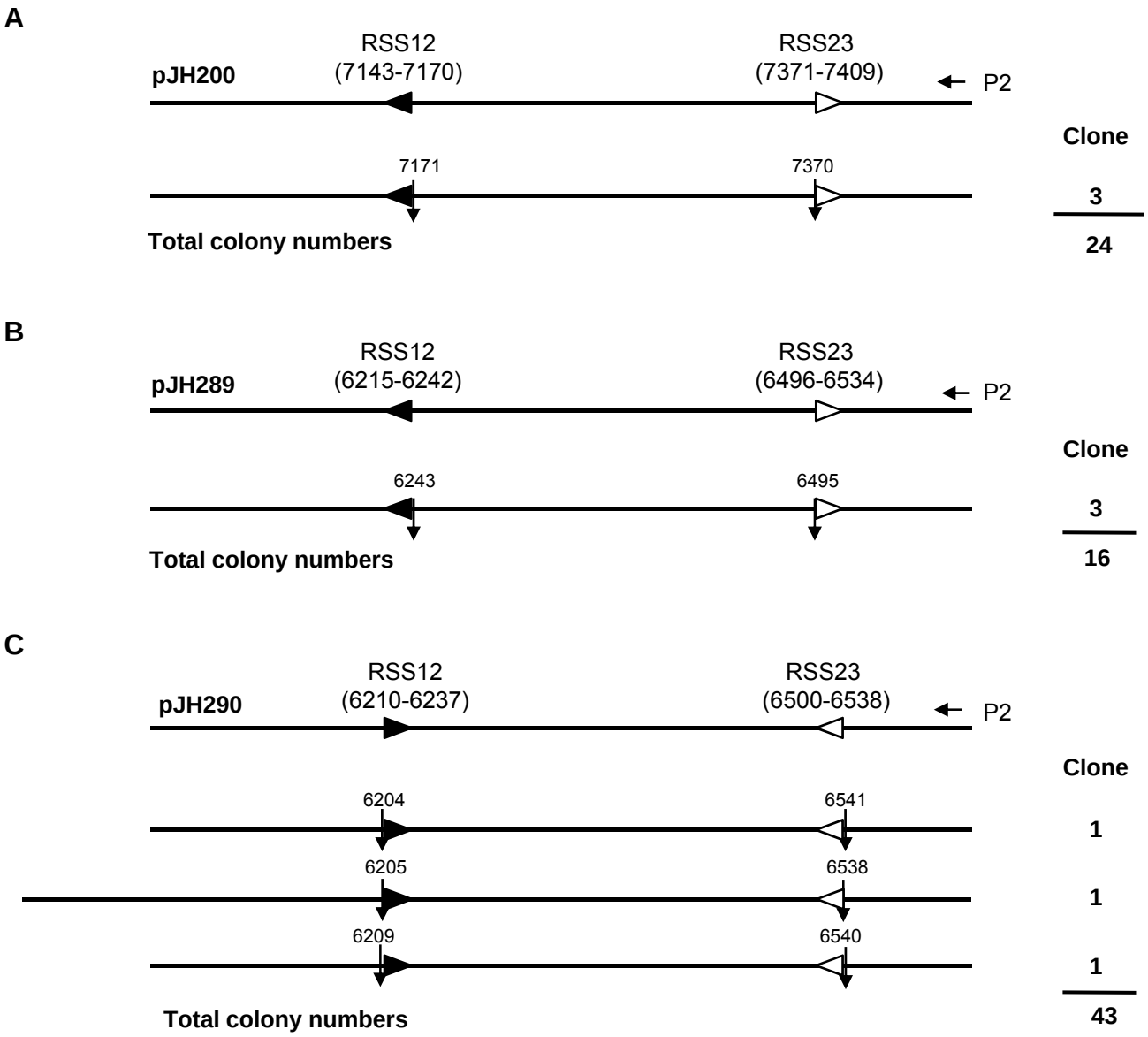


Fig. S6. Bacterial colony assay for analysis of cRAG1-introduced cleavage sites within plasmid substrates in the presence of mouse RAG2 in 293T cells. Schematic diagram of plasmid substrates pJH200 (*A*), pJH289 (*B*) or pJH290 (*C*). 12-RSS and 23-RSS are indicated as the black triangle and white triangle, respectively. P2 primer are used for sequencing (*SI Appendix*, Table S6). The downward arrows reflect the cut positions of nucleic acids with numbers as indicated. In all clones, three clones were selected from three independent experiments for sequences analysis. The total colony numbers are indicated below the sequences.

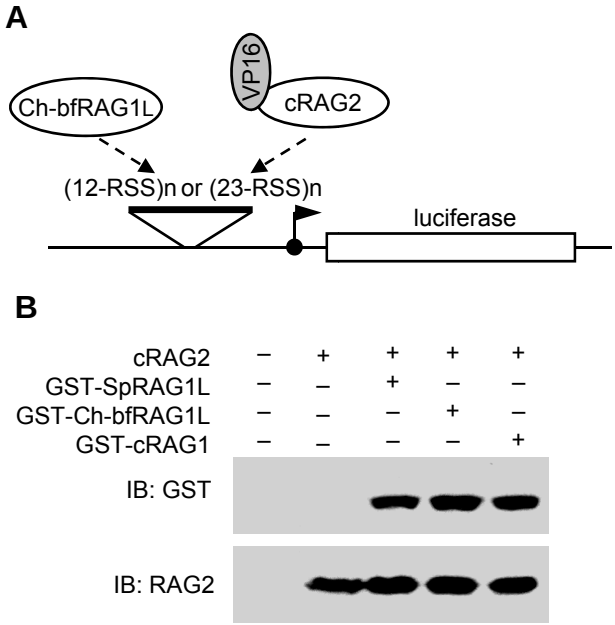


Fig. S7. Luciferase assay for detecting RSS recognition. (*A*) Schematic diagram of luciferase assay (1). Briefly, the inverted triangle represents the recombination signal sequences (RSSs) insertion of the upstream of luciferase promoter gene (black circle). Binding of RAG1 and RAG2-VP16 (open and shaded ovals) to RSSs could initiate the transcription of luciferase gene by VP16. (*B*) Extracts from HEK-293T cells that had been transfected with indicated plasmids were used to detect protein expression. The results are representative of three independent experiments.

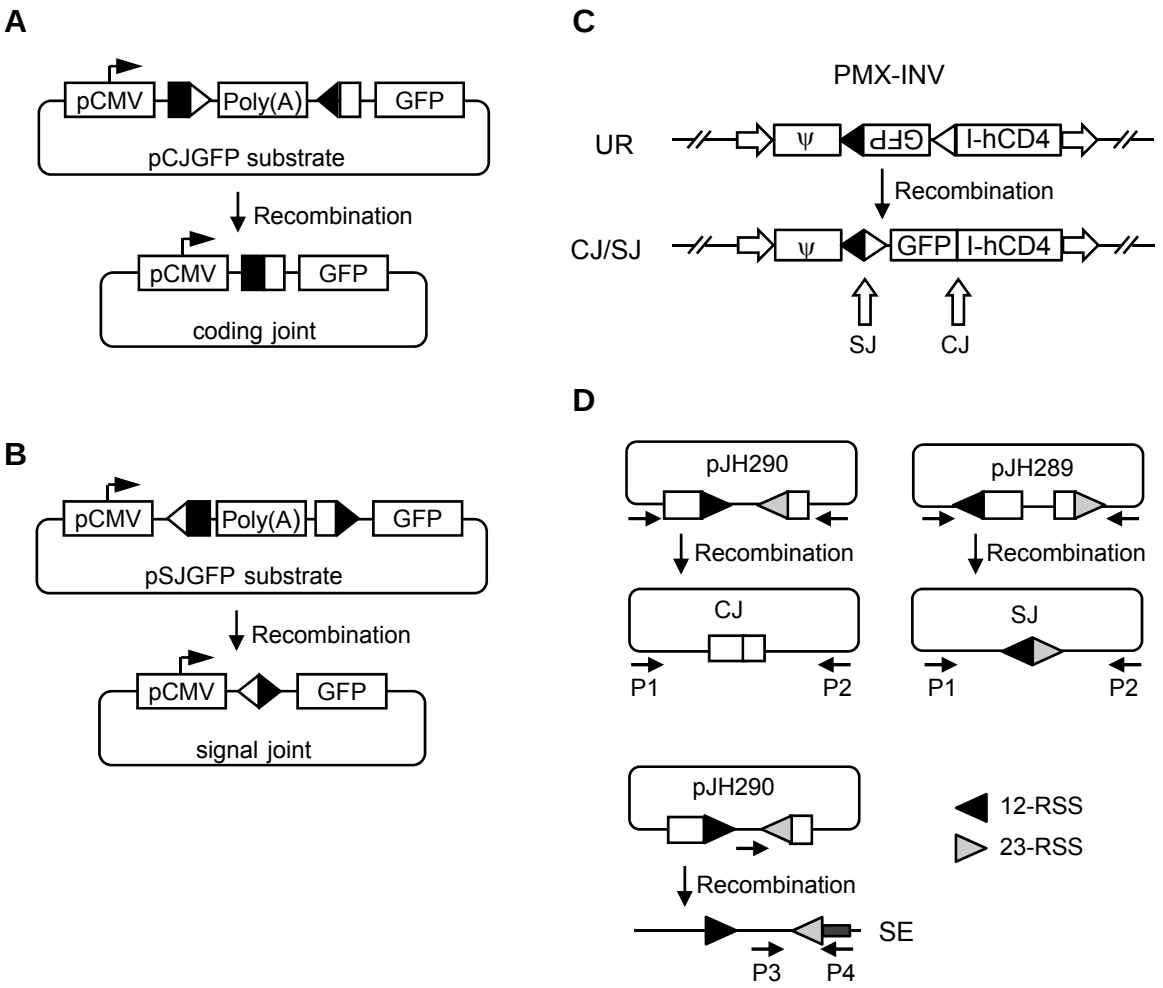


Fig. S8. Experimental schemes for measuring gene recombination. Experimental strategies with fluorescent reporter substrates (*A* and *B*) or integrated substrate pMX-INV (*C*) for detecting signal joint (SJ) and coding joint (CJ) (2, 3). Packaging sequence (ψ), GFP cDNA, IRES-hCD4 cDNA (I-hCD4) are indicated. (*D*) Experimental strategies with recombination substrates pJH290 and pJH289. 12-RSS and 23-RSS are indicated as the black triangle and shaded triangle, respectively. Linker DNA for detecting signal end (SE) formation in LMPCR assay is shown as a rectangle. Arrows reflect the orientation of primers (*SI Appendix*, Table S6) used for PCR to detect formation of coding joint (CJ), signal joint (SJ) and signal end (SE).

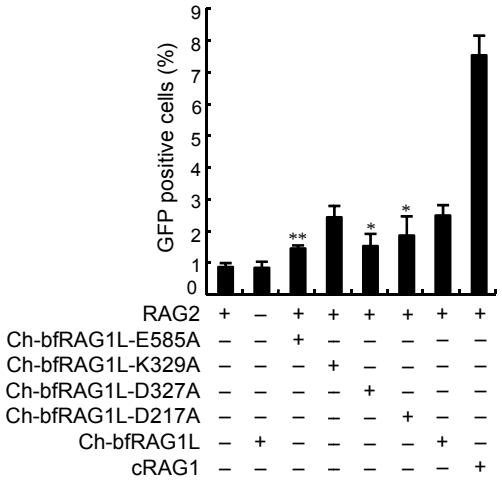


Fig. S9. The effects of mutations of Ch-bfRAG1L on gene recombination activities. HEK-293T cells were transfected with the indicated plasmids and pSJGFP reporter plasmid. The GFP level was subsequently measured by flow cytometry. The percentages of GFP^+ cells are shown (the means $\pm$ SD is calculated from triplicate experiments). Error bars indicate the standard deviations. The asterisks reflect the values significantly different from wild-type Ch-bfRAG1L and RAG2 complex (*P<0.05; **P<0.01, by the Student *t* test).

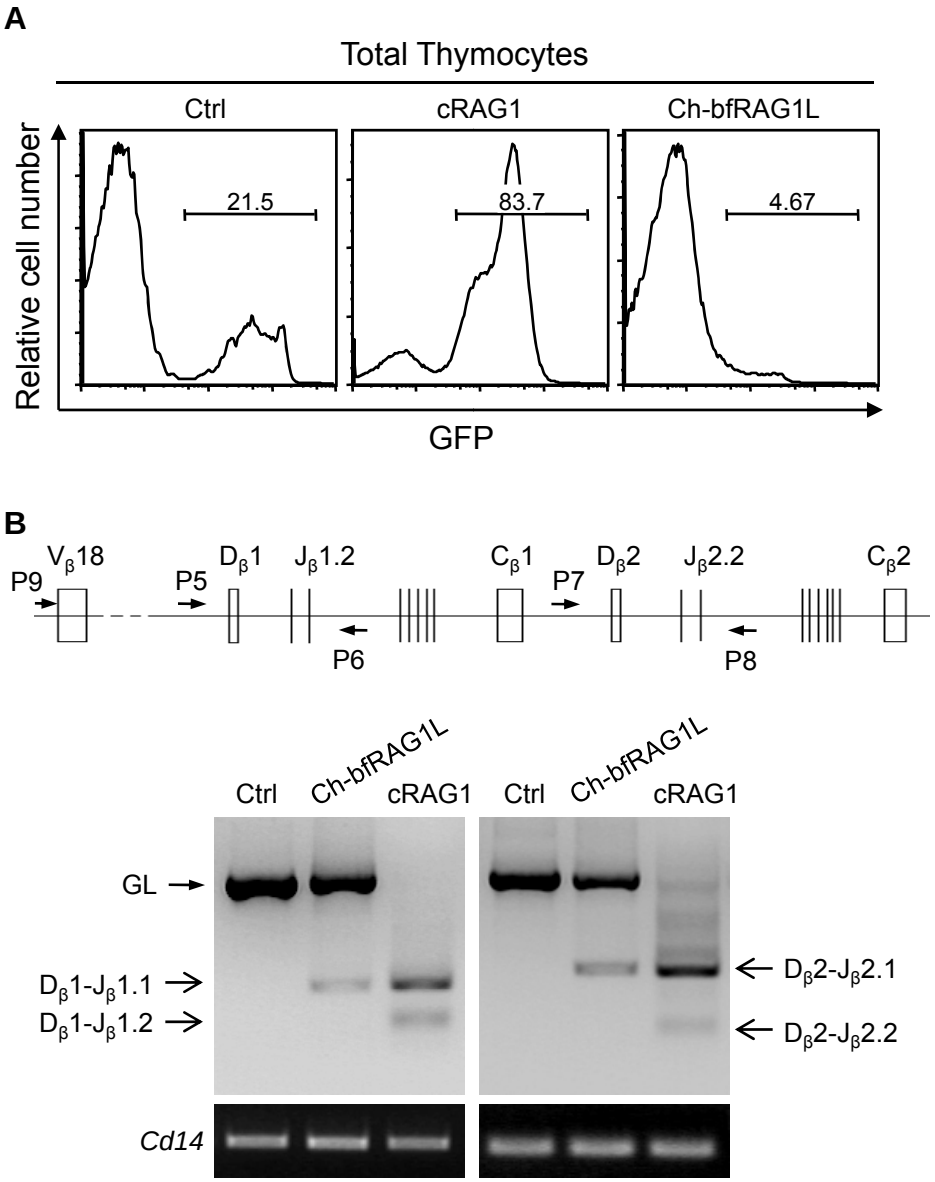


Fig. S10. Ch-bfRAG1L can introduce Tcrb V(D)J recombination in RAG1-deficient mice. (*A*) The percentages of GFP-positive thymocytes are shown. Ctrl, thymocytes only expressing GFP protein. (*B*) PCR analysis to detect the products of DJb recombination in total thymocytes. GL, germline. The recombination products were analyzed by the primers P5/P6, P7/P8 (*SI Appendix*, Table S6). The results are representative of three (*A*) or four (*B*) independent experiments.

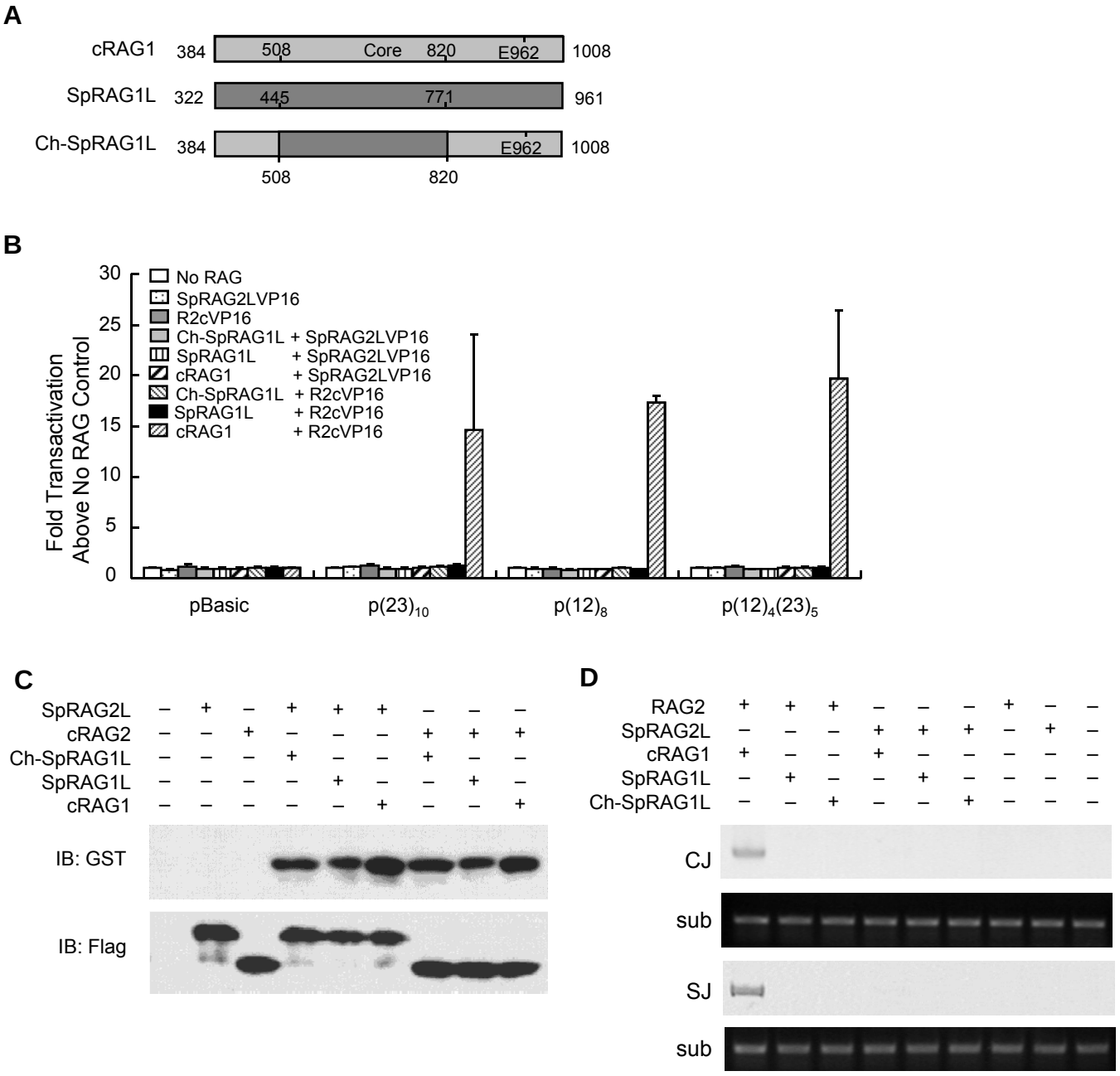
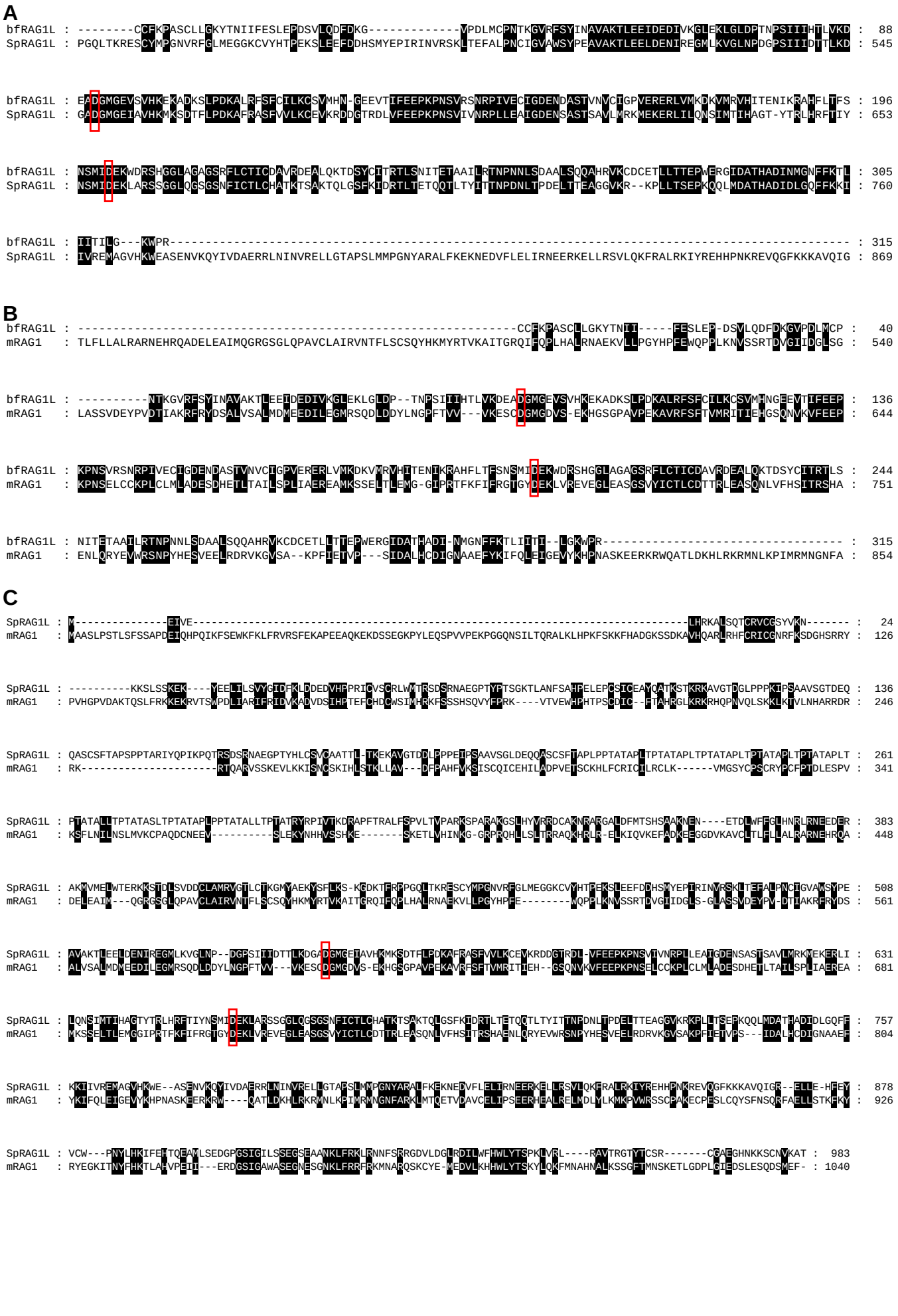


Fig. S11. The reconstitution of SpRAG1L into cRAG1-like protein as Ch-bfRAG1L. (*A*) Scheme of the constitution of Ch-SpRAG1L. (*B*) Ch-SpRAG1L failed to bind with RSS in the presence of Flag-tag RAG2 or SpRAG2L. The "fold transactivation" of luciferase activity was normalized against "no RAG control" with the indicated plasmids (the means ± SD is calculated from triplicate experiments). Error bars reflect the standard deviations. (*C*) The expression of transfected genes is shown. (*D*) Ch-SpRAG1L failed to generate recombination products in the presence of RAG2 or SpRAG2L. The results are representative of three independent experiments.


A
bfRAG1L : --------CCFKPASCLLGKYTNIIFESLEPDSVLQDFDKG-------------VPDLMCPNTKGVRFSYINAVAKTLEEIDEDIVKGLEKLGLDPTNPSIIIHTLVKD : 88
SpRAG1L : PGQLTKRESCYMPGNVRFGLMEGGKCVYHTPEKSLEEFDDHSMYEPIRINVRSKLTEFALPNCIGVAWSYPEAVAKTLEELDENIREGMLKVGLNPDGPSIIIDTTLKD : 545
bfRAG1L : EADGMGEVSVHKEKADKSLPDKALRFSFCILKCSVMHNGEEVTIFEEPKPNSVRSNRPIVECIGDENDASTVNVCIGPVERERLVMKDKVMRVHITENIKRAHFLTFS : 196
SpRAG1L : GADGMGEIAVHKMKSDTFLPDKAFRASFVVLKCEVKRDDGTRDLVFEEPKPNSVIVNRPLLEAIGDENSASTSAVLVRKMEKERLILQNSIMTIHAGT-YTRLHRFTIY : 653
bfRAG1L : NSMIDEKWDRSHGGLAGAGSRFLCTICDAVRDEALQKTDSYCITRTLSNITETAAILRTNPNNLSDAALSQQAHRVKCDCETLLTTEPWERGIDATHADINMGNFFKTL : 305
SpRAG1L : NSMIDEKLARSSGGLQGSGSNFICTLCHATKTSAKTQLGSFKIDRTLTETQQTLTYITTNPDNLTPDELTTEAGGVKR--KPLLTSEPKQQLMDATHADIDLGQFFKKI : 760
bfRAG1L : IITILG---KWPR---------- : 315
SpRAG1L : IVREMAGVHKWEASENVKQYIVDAERRLNINVRELLGTAPSLMMPGNYARALFKEKNEDVFLELIRNEERKELLRSVLQKFRALRKIYREHHPNKREVQGFKKKAVQIG : 869
B
bfRAG1L : ----------CCFKPASCLLGKYTNII-----FESLEP-DSVLQDFDKGVPDLMCP : 40
mRAG1 : TLFLLALRARNEHRQADELEAIMQGRGSGLQPAVCLAIRVNTFLSCSQYHKMYRTVKAITGRQIFQPLHALRNAEKVLLPGYHPFEWQPPLKNVSSRTDVGIIDGLSG : 540
bfRAG1L : ----------NTKGVRFSYINAVAKTLEEIDEDIVKGLEKLGLDP--TNPSIIIHTLVKDEADGMGEVSVHKEKADKSLPDKALRFSFCILKCSVMHNGEEVTIFEEP : 136
mRAG1 : LASSVDEYPVDTIAKRFRYDSALVSALMDMEEDILEGMRSQDLDDYLNGPFTVV---VKESQDGMGDVS-EKHGSGPAVPEKAVRFSFTVMRITIEHGSQNVKVFEEP : 644
bfRAG1L : KPNSVRSNRPIVECIGDENDASTVNVCIGPVERERLVMKDKVMRVHITENIKRAHFLTFSNSMIDEKWDRSHGGLAGAGSRFLCTICDAVRDEALQKTDSYCITRTLS : 244
mRAG1 : KPNSELCCKPLCLMLADESDHETLTAILSPLIAEREAMKSSELTLEWG-GIPRTFKFIFRGTGVDEKLVREVEGLEASGSVYICTLCDTTRLEASQNLVFHSITRSHA : 751
bfRAG1L : NITETAAILRTNPNNLSDAALSQQAHRVKCDCETLLTTEPWERGIDATHADI-NMGNFFKTLITTI--LGKWPR---------- : 315
mRAG1 : ENLQRYEVWRSNPYHESVEELRDRVKGVSA--KPFIETVP---SIDALHCDIGNAAEFYKIFQLEIGEVYKHPNASKEERKRWQATLDKHLRKRMNLKPIMRMNGNFA : 854
C
SpRAG1L : M----------EIVE----------LHRKALSQTCRVCGSYVKN------- : 24
mRAG1 : KAASLPSTLSFSSAPDEIQHPQIKFSEWKFKLFRVRSFEKAPEEAQKEKDSSEGKPYLEQSPVVPEKPGGQNSILTQRALKLHPKFSKKFHADGKSSDKAVHQARLRHFCRICGNRFKSDGHSRRY : 126
SpRAG1L : ----------KKSLSSKEK----YEELILSVYGIDFKLDDEDVHPPRICVSCRLWYTRSDSRNAEGPTYPTSGKTLANFSAHPELEPCSICEAVQATKSTKRKAVGTDGLPPPKIPSAAVSGTDEQ : 136
mRAG1 : PVHGPVDAKTQSLFRKKEKRVTSMPDLIARIFRIDVKADVDSIHPTEFCHDCWSIMHRKFSSSHSQVYFPRK----VTVEWHPHTPSCDICI--FTAHRGLKRKRHQPNVQLSKKLKTVLNHARRDR : 246
SpRAG1L : QASCSFTAPSPPTARIYQPIKPQTRSDSRNAEGPTYHLCSVCAATTL-TKEKAVGTDDLPPPEIPSAAVSGLDEQQASCSFTAPLPPTATAPLTPTATAPLTPTATAPLTPTATAPLT : 261
mRAG1 : RK----------RTQARVSSKEVLKKISNCSKIHLSTKLLAV---DFPAHFVKSISCQICEHILADPVETSCKHLFCRICILRCLK------VMGSYCPSCRYPCFPTDLESPV : 341
SpRAG1L : PTATALLTPTATASLTPTATAPLPPTATALLTPTATRYRPIVTKDRAPFTRALFSPVLTVPARKSPARAKGSLHYYRRDCAKNRARGALDFMTSHSAAKNEN----ETDLWFPGLHNRLRNEEDER : 383
mRAG1 : KSFLNILNSLMVKCPAQDCNEEV----------SLEKYNHHVSSHKE-------SKETLVHINKG-GRPRQHLLSLTRAQKHRLR-ELKIQVKEFADKEEGGDVKAVCLTLFLLALRARNEHRQA : 448
SpRAG1L : AKVVMELWTERKKSTDLSVDDCLAMRVGTLCTKGMYAEKYSFLKS-KGDKTGRPPGQLTKRESCYMPGNVRFGLMEGGKCVYHTPEKSLEEFDDHSYYEPIRINVRSKLTEEALPNCIGVAWSYPE : 508
mRAG1 : DELEAIK---QGRGSGLQPAVCLAIRVNTFLSCSQYHKMYRTVKAITGRQIFQPLHALRNAEKVLLPGYHPFE--------VQPPLKNVSSRTDVGIIDGLS-GLASSVDEYPV-DTIAKRFRVDS : 561
SpRAG1L : AVAKTLEELDENIREGMLKVGLNP--DGPSIIIDTTLKDGADGMGEIAVHKMKSDTFLPDKAFRASFVVLKCEVKRDDGTRDL-VFEEPKPNSVIVNRPLLEAIGDENSASTSAVLMRKMEKERLI : 631
mRAG1 : ALVSALMDMEEDILEGMRSQDLDDYLNGPFTVV---VKESQDGMGDVS-EKHGSGPAVPEKAVRFSFTVMRITIEH--GSQNVKVFEEPKPNSELCCKPLCLMLADESDHETLTAILSPLIAEREA : 681
SpRAG1L : LQNSIMTIHAGTYTRLHRFTIYNSMIDEKLARSSGGLQGSGSNFICTLCHATKTSAKTQLGSFKIDRTLTETQQTLTYITTNPDNLTPDELTTEAGGVKRKPLLTSEPKQQLMDATHADIDLGQFF : 757
mRAG1 : KKSSELTLEMGGIPRTFKFIFRGTGYDEKLVREVEGLEASGSVYICTLCDTTRLEASQNLVFHSITRSHAENLQRYEVWRSNPYHESVEELRDRVKGVSAKPFIEIVPS---IDALHCDIGNAAEF : 804
SpRAG1L : KKIIVRFMAGVHKWE--ASENVKQYIVDAERRLNINVRELLGTAPSLMMPGNYARALFKEKNEDVFLELIRNEERKELLRSVLQKFRALRKIYREHHPNKREVQGFKKKAVQIGR--ELLE-HFEY : 878
mRAG1 : YKIFQLEIGEVYKHPNASKEERKRW----QATLDKHLRKRMNLKPIMRMNGNFARKLMTQETVDAVCELIPSEERHEALRELMDLVLKMKPVWRSSCPAKECPESLCQYSFNSQRFAELLSTKFKY : 926
SpRAG1L : VCW---PNYLHKIFEHTQEAMLSEDGPGSIGILSSEGSEAANKLFRKLRNNFSRRGDVLDGLRDILWFHWLYTSPKLVRL----RAVTRGTYTCSR-------CGAEGHNKKSCNVKAT : 983
mRAG1 : RYEGKITNYFHKTLALVPEIIL---ERDGSIGAWASEGNESGNKLFRRFRKMNARQSKCYE-MEDVLKHHWLYTSKYLQKFMNAHNALKSSGFTMNSKETLGDPLGIEDSLESQDSMEF- : 1040

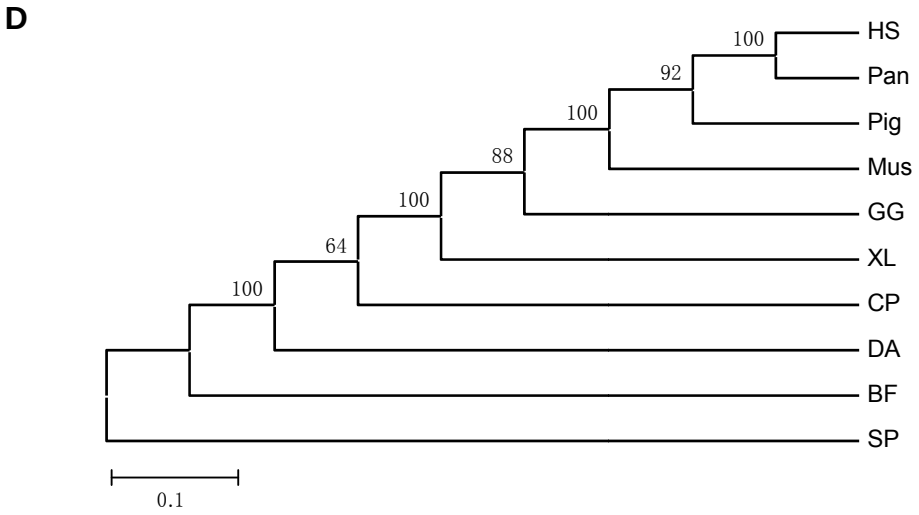


Fig. S12. Multiple alignment analysis. Sequences alignments of bfRAG1L with SpRAG1 (*A*), bfRAG1L with mouse RAG1 (*B*) or SpRAG1L with mouse RAG1 (*C*). The conserved DDE active site motif amino acids are shown in red boxes. The identical residues of bfRAG1L with SpRAG1L are indicated by the black shading. Numbers indicate amino acid positions. (*D*) The phylogenetic tree was constructed by MEGA molecular evolutionary genetic analysis software (4). The evolutionary distance was corrected using the scale depicted by the horizontal bar. Bootstrap values higher than 50% are shown at the nodes. Species abbreviations are as follows: BF, lancelet; CP, bull shark; DA, zebrafish; GG, chicken; HS, human; Mus, mouse; Pan, chimpanzee; SP, sea urchin; XL, frog.

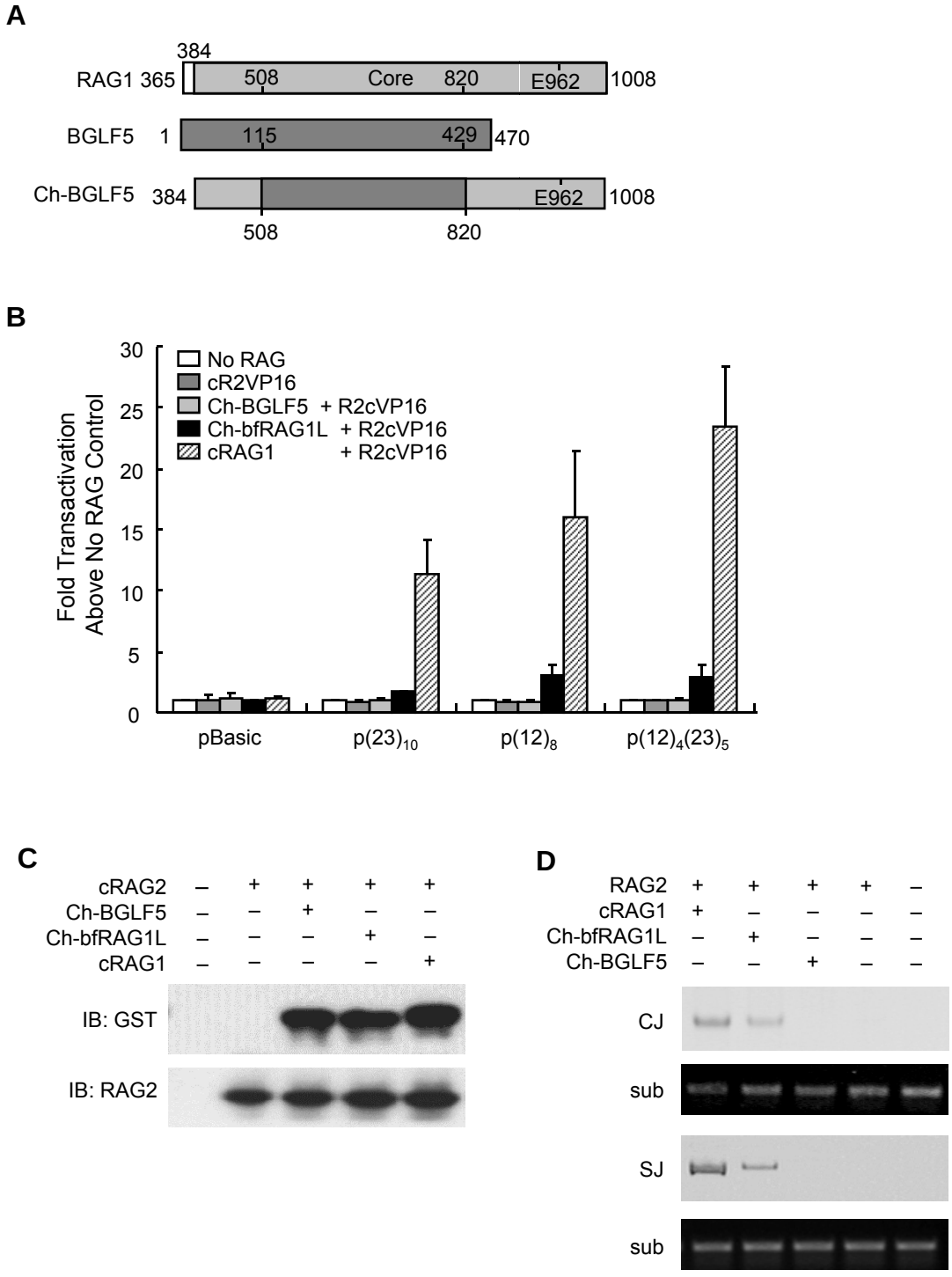


Fig. S13. The reconstitution of BGLF5 into cRAG1-like protein. (*A*) Scheme of the constitution of cRAG1-like protein Ch-BGLF5. (*B*) Ch-BGLF5 failed to bind with RSS in the presence of RAG2. The "fold transactivation" of luciferase activity was normalized against "no RAG control" with the indicated plasmids (the means ± SD is calculated from triplicate experiments). Error bars reflect the standard deviations. (*C*) The expression of transfected genes is shown. (*D*) Ch-BGLF5 failed to generate recombination products. The results are representative of three independent experiments.

Table S1. Significance of similarities between bfRAG1L and virus proteins.

Virus proteins			PSI-BLAST E (first iteration)
Gi 74136753	recombination activating protein 1-like	Strongylocentrotus purpuratus	7e-74
Gi 259130365	RNA dependent RNA polymerase	Tensaw virus	1e-05
Gi 259130363	RNA dependent RNA polymerase	Tensaw virus	2e-05
Gi 9630657	hypothetical protein BUNVsLgp1	Bunyamwera virus	3e-05

bfRAG1L protein was as a query for searching against all the GenBank proteins excluding jawed vertebrates proteins. The columns 1-2 listed all the GenBank proteins excluding jawed vertebrates proteins producing significant alignments with E-values better than threshold in the first iteration in PSI-BLAST searches and their GI numbers are indicated (*SI Materials and Methods*). Column 3 listed the names of species. Column 4 indicated the E-values of the best matches between bfRAG1L and the corresponding proteins (E-value <3e-05).

Table S2. The sequence analysis of the bfRAG1L-introduced cleavage sites within the plasmid pBR322.

Cleavage site of bfRAG1L on pBR322

3′	5′	3′	5′
ACAGATGTAGGTGTTCCACA	----------	GGGTAGCCAGCAGCATCCTG
CAACCCCGCCAGCCTAGCCG	----------	GGTCCTCAACGACAGGAGCA
CGGAAACCGAAGACCATTCA	----------	TGTTGTTGCTCAGGTCGCAG
GCGTATCGGTGATTCATTCT	----------	GCTAACCAGTAAGGCAACCC
GATGCGGTATTTTCTCCTTA	----------	CGCATCTGTGCGGTATTTCA
CAGAGGTTTTCACCGTCATC	----------	ACCGAAACGCGCGAGGCAGC
CGCAGGAAAGAACATGTGAG	----------	CAAAAGGCCAGCAAAAGGCC
GTCTGCCTCCCCGCGTTGCG	----------	TCGCGGTGCATGGAGCCGGG
CCCGGCTAGGCTGGCGGGGT	----------	TGCCTTACTGGTTAGCAGAA
AGACATCTGTGAATCGCTTC	----------	ACGACCACGCTGATGAGCTT
AGGGCGCGTCAGCGGGTGTT	----------	GGCGGGTGTCGGGGCGCAGC
ACGGATGGCATGACAGTAAG	----------	AGAATTATGCAGTGCTGCCA
CAATTAATAGACTGGATGGA	----------	GGCGGATAAAGTTGCAGGAC
AGACATCTGTGAATCGCTTC	----------	ACGACCACGCTGATGAGCTT
AGGGCGCGTCAGCGGGTGTT	----------	GGCGGGTGTCGGGGCGCAGC
ACGAGCGTGACACCACGATG	----------	CCTGCAGCAATGGCAACAAC
GCGGTATCATTGCAGCACTG	----------	GGGCCAGATGGTAAGCCCTC
CCATATGCGGTGTGAAATAC	----------	CGCACAGATGCGTAAGGAGA
CCGCTTCCTCGCTCACTGAC	----------	TCGCTGCGCTCGGTCGTTCG
CGCTCATCGTCATCCTCGGC	----------	ACCGTCACCCTGGATGCTGT
CGCTGGAGCGCGACGATGAT	----------	CGGCCTGTCGCTTGCGGTAT
CAGCGCTCTGGGTCATTTTC	----------	GGCGAGGACCGCTTTCGCTG
CGATGATCGGCCTGTCGCTT	----------	GCGGTATTCGGAATCTTGCA
CAAGAATTCTCATGTTTGAC	----------	AGCTTATCATCGATAAGCTT
CAGGCACCGTGTATGAAATC	----------	TAACAATGCGCTCATCGTCA
GCTTCCATTCAGGTCGAGGT	----------	GGCCCGGCTCCATGCACCGC
ATAAATCGCCGTGACGATCA	----------	GCGGTCCAGTGATCGAAGTT

9/27	have C nucleotide at 5′	33.3%
6/27	have G nucleotide at 5′	22.2%
5/27	have A nucleotide at 5′	18.5%
7/27	have T nucleotide at 5′	25.9%
5/27	have C nucleotide at 3′	18.5%
11/27	have G nucleotide at 3′	40.7%
6/27	have A nucleotide at 3′	22.2%
5/27	have T nucleotide at 3′	18.5%

Dashed line indicated the cut position with 5'-phosphorylated and 3'-OH end as shown. Sequences were analyzed from two independent experiments.

Table S3. Sequence analysis of coding joints (*A*), signal joints (*B*) and signal ends (*C*).

A. Coding joints

Ch-bfRAG1L + RAG2

left coding end	right coding end	del(bp)	homol	clones
CCAAGCTTGGGCTGCAGGTCGAC	GGATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG			
CCAAGCTTGGGCTGCAGGTCGAC	----------AAAATG	48	0	1/50
CCAAGCTTGGGCTGCAGGTC**GA**	------TCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	4	2	4/50
CCAAGCTTGGGCTGCAG**G**	----------GGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	14	1	2/50
CCAAGCTT	----------CGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	22	0	1/50
CCAAGCTTGGGCTGCAGGT**CG**	----------GGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	11	2	2/50
CCAAGCTTGGGCTGCAGG**TC**	----------CCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	8	2	8/50
CCAAGCTTGGG**C**	----------CCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	16	1	1/50
CCAAGCTTGGGCTGCAG	----------TCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	9	0	1/50
CCAAGCTTGGGCTGCAGG	---*ATCCCT*------GGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	13	0	1/50
CCAAGCTTGGGCTGCAGGTCGAC	---GATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	1	0	1/50
CCAAGCTTGGGCTGCAGGTC	----------CCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	7	0	1/50
CCAAGCTTGGGCT**G**	----------ATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	11	1	1/50
CCAAGCTTGGGCTGCAGG**TC**	----------AGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	18	2	2/50
CCAAGCTTGGGCTGCAGGTCGAC**GT**	-------CGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	7	0	1/50
CCAAGCTTGGGCTGC	---------GGGATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	8	0	1/50
CCAAGCTTGGGCTGCAGGT**C**	----------CCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	9	1	1/50
CCAAGCTTGGGCTGCA**GG**	---------ATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	7	2	1/50
CCAAGCTTGGGCTGCAGGTCGAC**G-**	---TCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	3	0	1/50
CCAAGCTTGGGCTGCAGGT**C**	----------GGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	11	1	3/50
CCAAGCTTGGGCTGCA	----------CCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	11	0	1/50
CCAAGCTT	--------*AATTCCCTCCCC*---------ATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	27	0	2/50
CCAAGCTT	---*AAT*---CCACTGTGGGATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	15	0	1/50
CCAAGCTTGGGCTGCAGGTCGACCACAG	----------GGAAGCTAAAATG	41	0	1/50
CCAAGCTTGGGCTGCA**G**	----------ATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	8	1	1/50
CCAAGCTTGGGCTGCAGGT	-------GATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	5	0	3/50
CCAAGCTTGGGCTGCAGG	----------CCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	9	0	1/50
CCAAGCTTGGGCTGCAG-CTACCACTGTGGGATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAA		6	0	1/50
CCAAGCTTGGGCTGCAGG	--*GTC*---GATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	6	0	1/50
CCAAGCTTGGGCTGCAGGTCGA**C**	---------CGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	7	1	1/50
CCAAGCTTGGGCTGCAGGTCGAC	----------GCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	21	0	1/50
CCAAGCTTGGGCTGCAGGTCGAC	---GATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	1	0	1/50
CCAAGCTTGGGCTGCAGGT	--------ATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	6	0	1/50

27/50 have microhomologies (**bold underlined**) < 3nt: 54.0%
50/50 have deletions: 100%
20/50 have deletions > 10nt: 40.0%
5/50 have nucleotide insertion (*italic*): 10.0%
2/50 have P nucleotides (**bold**): 4.0%

Table S3A continued

cRAG1 + RAG2

left coding end	right coding end	del(bp)	homol	clones
CCAAGCTTGGGCTGCAGGTCGAC	GGATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG			
CCAAGCTTGGGCTGCAGGTCGAC	GGATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	0	0	1/37
CCAAGCTTGGGCTG**C**----------	CCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	14	1	2/37
CCAAGCTTGGGCTGCAGG**TC**----------	CCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	8	2	6/37
CCAAGCTTGGGCTGC----------	GATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	9	0	1/37
CCAAGCTTGGGCTGCA**GG**----------	ATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	7	2	4/37
CCAAGCTTGGGCTGCAGGTC**G**------	GATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	3	1	3/37
CCAAGCTTGGGCTGCAGGTC**GA**-------	TCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	4	2	3/37
CCAAGCTTGGGCTGCA**G**----------	ATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	8	1	2/37
CCAAGCTTGGGCTGC**A**----------	TCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	10	1	1/37
CCAAGCTTGGGCTGCAGGT**C**----------	CCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	9	1	1/37
CCAAGCTTGGGCTGC----------	CCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	12	0	1/37
CCAAGCTTGGGCTGCAGGTCG----------	CCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	8	0	1/37
CCAAGCTTGGGCTGCAGGT----------	GGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	14	0	1/37
CCAAGCTTGGGCTGCAGGTCGA**C**--------	CCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	5	1	1/37
CCAAGCTTGGGCTGCAGGTC----------	CCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	7	0	1/37
CCAAGCTTGGGCTGCAGGT**CG**----------	GGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	11	2	1/37
CCAAGCTTGGGCTGCAGGTCGAC**GT**-----	CCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	4	0	2/37
CCAAGCTTGGGCTGCAGG----------	CCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	11	0	1/37
CCAAGCTTGGGCTGCAGGT----------	ATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	6	0	1/37
CCAAGCTTGGGCTGCAG----------	TCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	9	0	1/37
CCAAGCTTGGGCTGCAGGTCGAC----	GATCCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	1	0	1/37
CCAAGCTTGGGCTGCAGGTCGAC-------	CCCCGGGGATCAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG	4	0	1/37

24/37 have microhomologies (**bold underlined**) < 3nt: 64.9%
36/37 have deletions: 97.3%
6/37 have deletions > 10nt: 16.2%
0/37 have nucleotide insertion (*italic*): 0%
2/37 have P nucleotides (**bold**): 5.4%

B. Signal joints

Ch-bfRAG1L + RAG2

left signal end	right signal end	del(bp)	homol	clones
GGTTTTTGTTCCAGTCTGTAGCACTGTG	CACAGTGGTAGTACTCCACTGTCTGGCTGT			
GGTTTTTGTTCCAGTCTGTAGCACTGTG	CACAGTGGTAGTACTCCACTGTCTGGCTGT	0	0	37/50
GGTTTTTGTTCCAGTCTGTAGCACTGT**G**----------TGGTAGTACTCCACTGTCTGGCTGT		5	1	1/50
GGTTTTTGTTCCAGTCTGTAGCACTGTGGTC---*GCACAG*--CACAGTGGTAGTACTCCACTGTCTGGCTGT		0	0	1/50
GGTTTTTGT----------*ACAGCCAG*----------ACAGTGGTAGTACTCCACTGTCTGGCTGT		20	0	1/50
GGTTTTTGTTCCAGTCTGTA--------*AGA*-----GGATCCCACAGTGGTAGTACTCCACTGTCTGGCTGT		8	0	2/50
GGTTTTTGTTCCAGTCTGTAGCACTGTGGTCGACTCTCATCG----------CACTGTCTGGCTGT		16	0	1/50
----------(*+60bp*)----------GTACTCCACTGTCTGGCTGT		58	0	1/50
GGTTTTTGTTCCAGTCTGTAGCACTG-------*CG*-------CACAGTGGTAGTACTCCACTGTCTGGCTGT		2	0	1/50
GGTTTTTGTTCCAGTCTGTAGCACTGTGGTCGAC----------TACTCCACTGTCTGGCTGT		11	0	1/50
----------*AATTCCCTCCCCAG*------CACAGTGGTAGTACTCCACTGTCTGGCTGT		48	0	1/50
GGTTTTTGTTCCAGTCTGTAGCACTGTGGTCG----------CACAGTGGTAGTACTCCACTGTCTGGCTGT		0	0	1/50
GGTTTTTGTTCCAGTCTGTAGCACTG----------AGTACTCCACTGTCTGGCTGT		11	0	1/50
GGTTTTTGTTCCAGTCTGTAG----------GT		35	0	1/50

37/50 have correct signal joint structures: 74.0%
1/50 have microhomologies (**bold underlined**) < 3nt: 2.0%
11/50 have deletions: 22.0%
7/50 have deletions > 10nt: 14.0%
7/50 have nucleotide insertion (*italic*): 14.0%

cRAG1 + RAG2

left signal end	right signal end	del(bp)	homol	clones
GGTTTTTGTTCCAGTCTGTAGCACTGTG	CACAGTGGTAGTACTCCACTGTCTGGCTGT			
GGTTTTTGTTCCAGTCTGTAGCACTGTG	CACAGTGGTAGTACTCCACTGTCTGGCTGT	0	0	34/36
GGTTTTTGTTCCAGTCTGTAGCACTGTG---*T*----CACAGTGGTAGTACTCCACTGTCTGGCTGT		0	0	1/36
GGTTTTTGTTCCAGTCTGTAGCACTG----------CACAGTGGTAGTACTCCACTGTCTGGCTGT		2	0	1/36

34/36 have correct signal joint structures: 94.4%
0/36 have microhomologies (**bold underlined**) < 3nt: 0%
1/36 have deletions: 2.8%
0/36 have deletions > 10nt: 0%
1/36 have nucleotide insertion (*italic*): 2.8%

C. Signal ends

Ch-bfRAG1L + RAG2

signal end	linker	del(bp)	clones
GGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGTG	CACTTCAGATCTCCCGAGTCA		
GGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGTG	CACTTCAGATCTCCCGAGTCA	0	5/40
GGTTTTTGTACAGCCAGACAGTGGAG----------CACTTCAGATCTCCCGAGTCA		13	6/40
GGTTTTTGTACAGCCAGACA----------CACTTCAGATCTCCCGAGTCA		19	5/40
GGTTTTTGTACAGCCAGACAGTGGAGT----------CACTTCAGATCTCCCGAGTCA		12	5/40
GGTTTTTGTACAGCCAGACAG----------CACTTCAGATCTCCCGAGTCA		18	1/40
GGTTTTTGTACAGCCAGAC----------CACTTCAGATCTCCCGAGTCA		20	4/40
GGTTTTTGTACAGCCAGACAGTGGA----------CACTTCAGATCTCCCGAGTCA		14	6/40
GGTTTTTGTACAGCCAG----------CACTTCAGATCTCCCGAGTCA		22	2/40
GGTTTTTGTACAGCCAGACAGTG----------CACTTCAGATCTCCCGAGTCA		16	2/40
GGTTTTTG----------CACTTCAGATCTCCCGAGTCA		31	1/40
GGTTTTTGTACAGCCAGACAGTGG----------CACTTCAGATCTCCCGAGTCA		15	1/40
----------CACTTCAGATCTCCCGAGTCA		50	1/40
GGTTTTTGTACAGCCA----------CACTTCAGATCTCCCGAGTCA		23	1/40

5/40 have correct signal end structure: 12.5%
35/40 have deletions: 87.5%
35/40 have deletions > 10nt: 87.5%

cRAG1 + RAG2

signal end	linker	del(bp)	clones
GGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGTG	CACTTCAGATCTCCCGAGTCAC	0	25/28
GGTTTTTGTACAGCCAGACAGTGGA----------CACTTCAGATCTCCCGAGTCAC		14	1/28
GGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGT------CACTTCAGATCTCCCGAGTCAC		1	2/28

25/28 have correct signal end structures: 89.3%
2/28 have deletions: 7.1%
1/28 have deletions > 10nt: 3.6%

The top of each of sequences was shown with precise junctions. bp, base pair; Homology-mediated recombination, bolded line; Deletion, dashed line; Insertions, italic; P nucleotides, bold. Sequences were analyzed from three independent assays with 10 to 20 clones in each experiment.

Table S4. Sequence analysis of coding joints formed in D-Jb.

A. Dβ1-Jβ1.1 recombination

Ch-bfRAG1L+ RAG2

left coding end	right coding end	del(bp)	homol	clones
GGGACAGGGGGC	CAAACACAGAAGTCTTCTTTGGTAAA			
GGGACAGG--------*TT*----------CACAGAAGTCTTCTTTGGTAAA		8	0	17
GGGAC------*TGGGGTGCAGG*-----AACACAGAAGTCTTCTTTGGTAAA		9	0	19
GGGACAGGGGG----*GCAGG*-------AACACAGAAGTCTTCTTTGGTAAA		3	0	1
GGGACAGG------*CT*----------CACAGAAGTCTTCTTTGGTAAA		8	0	1
GGGACAGGGG--------*TCAGAG*----------GTCTTCTTTGGTAAA		13	0	1
GGGACAGGG----*AACG*---------AAACACAGAAGTCTTCTTTGGTAAA		4	0	1
GGGACAGGG-----*TG*----------AACACAGAAGTCTTCTTTGGTAAA		5	0	1
GGGACA-------*AAAAT*-------CAAACACAGAAGTCTTCTTTGGTAAA		6	0	1

cRAG1 + RAG2

left coding end	right coding end	del(bp)	homol	clones
GGGACAGGGGGC	CAAACACAGAAGTCTTCTTTGGTAAA			
GGGACA---------*AT*----------AACACAGAAGTCTTCTTTGGTAAA		8	0	3
GGGACAGGGGG----*GA*----------CACAGAAGTCTTCTTTGGTAAA		5	0	2
GGGACAGGG-------*TGGG*---------CACAGAAGTCTTCTTTGGTAAA		7	0	2
GGGACAGGGGG---*AGGGG*---**TTG**CAAACACAGAAGTCTTCTTTGGTAAA		1	3	2
GGGACAGGG-------*TG*------**G**CAAACACAGAAGTCTTCTTTGGTAAA		3	1	3
GGGACAGGGG----*TAGGG*----------CACAGAAGTCTTCTTTGGTAAA		6	0	1
GGGACAGGG----*AGGAG*----------CACAGAAGTCTTCTTTGGTAAA		7	0	1
GGGACAGGGGG-----*GGTGG*--------CACAGAAGTCTTCTTTGGTAAA		7	0	1
GGGACAGGGGGC----------AAACACAGAAGTCTTCTTTGGTAAA		1	0	1
GGGACAGGG-----*TGAC*----------CAGAAGTCTTCTTTGGTAAA		9	0	1
GGGACA---------*ATCA*----**TG**CAAACACAGAAGTCTTCTTTGGTAAA		6	2	2
GGGACAGG----------CAAACACAGAAGTCTTCTTTGGTAAA		4	0	1

B. Dβ2-Jβ2.1 recombination

Ch-bfRAG1L + RAG2

left coding end	right coding end	del(bp)	homol	clones
GGGACTGGGGGGGC	TAACTATGCTGAGCAGTTCTTCGGAC			
GGGACTGGGGGGG-------*TGA*---------GCTGAGCAGTTCTTCGGAC		8	0	34
GGGACTGGGGGGG-----*TGAGCC*----------GAGCAGTTCTTCGGAC		11	0	1
GGGACTGGGGGGG---*TGAGCTGAGCG*----------GTTCTTCGGAC		16	0	1

cRAG1 + RAG2

left coding end	right coding end	del(bp)	homol	clones
GGGACTGGGGGGGC	TAACTATGCTGAGCAGTTCTTCGGAC			
GGGACTGGGGGGG-----*TA*----**A**TAACTATGCTGAGCAGTTCTTCGGAC		1	0	9
GGGACTGGGGGGGC**G**----------CTATGCTGAGCAGTTCTTCGGAC		3	1	2
GGGACTGGGGGGGC**GC**---*TG*---------TGCTGAGCAGTTCTTCGGAC		6	1	1
GGGACTGGGGGGG----*GTAATAG*----CTATGCTGAGCAGTTCTTCGGAC		4	2	1
GGGACTGGGGG------*AAC*------AACTATGCTGAGCAGTTCTTCGGAC		4	0	5
GGGACTGGGGGGG----------ACTATGCTGAGCAGTTCTTCGGAC		3	0	2
GGGACTGGGG----------AACTATGCTGAGCAGTTCTTCGGAC		5	0	1
GGGACTGGGGGGGC----------TATGCTGAGCAGTTCTTCGGAC		5	0	1
GGGACTGGGGG-------*AGA*---------ATGCTGAGCAGTTCTTCGGAC		8	0	1

The top of each of sequences was shown with precise junctions. bp, base pair; Deletion, dashed line; Insertions, italic. Sequences were analyzed from two independent assays with 10 to 20 clones in each experiment.

Table S5. Sequence analysis of coding joints formed in V18-DJb.

A. V18-DJβ1 recombination

cRAG1 + RAG2

left coding end	right coding end	del(bp)	homol	clones
TTTCTGTAGTTCTAGAGA	GGGACAGGGGGC			
TTTCTGTAGTTCTAGAGA----*TCTTT*------GGGACAGGG--*AT*		3	0	1
TTTCTGTAGTTCTAGAGA----*TCGGTGGGAT*--------GGGGC		7	0	4
TTTCTGTAGTTCTAGAG---------------GGGACAG-----*CT*		6	0	6

B. V18-DJβ2 recombination

Ch-bfRAG1L + RAG2

left coding end	right coding end	del(bp)	homol	clones
TTTCTGTAGTTCTAGAGA	GGGACTGGGGGGGC			
TTTCTGTAGTTCTAGAG--------------------GGGGG--- *TTG*		10	0	9
TTTCTGTAGTTCTAGAG--------------------GGGG----*ATTG*		11	0	1
TTTCTGTAGTTCTAGAG--------------------GGGGG---*TTGTATA*		10	0	1

The top of each of sequences was shown with precise junctions. bp, base pair; Deletion, dashed line; Insertions, italic. Sequences were analyzed from two independent assays with 5 clones in each experiment.

Table S6. PCR primers and linker primers.

P1	5'-TCACACAGGAAACAGCTATGACCATG-3'
P2	5'-GGGATATATCAACGGTGGTATATCCAGTG-3'
P3	5'-TCGGCAGGTCTCCAGTAATGA-3'
P4	5'-CGGTGACTCGGGAGATCTGA-3'
P5	5'-ACCTATGGGAGGGTCCTTTTTTGTATAAAG-3'
P6	5'-TCCATCCCACACCCAAAGACCC-3'
P7	5'-GATGTAACATTGTGGGGACTGGG-3'
P8	5'-CTCTGAATATCTTCCTCCCATGGTACC-3‘
P9	5'-ATGCAAGGCCTGGAGACAGCAGTATC
P10	5'-ATTGGATCCTGCTGCTTTAAGCCAGCAAGCTGT-3‘
P11	5'-ATTTGCGGCCGCGCACCTGGGCCACTTCCCTAATATA-3‘
FM11	5'-CACTTCAGATC-3'
FM25	5'-GCGGTGACTCGGGAGATCTGAAGTG-3'

SI Materials and Methods

Mouse. RAG1$^{-/-}$ mice were purchased from the Jackson Laboratory and subsequently maintained in the SPF facility. All animal experiments were approved by the institutional animal use committee of the Shanghai Institutes for Biological Sciences (Chinese Academy of Sciences).

Cell Lines and Transfection. HEK-293T cells maintained in DMEM (Gibco-BRL) supplemented with 10% FBS (fetal bovineserum) at 37°C under 5% CO2. Transfections were performed using calcium phosphate precipitation. *Raji* cells were grown in RPMI 1640 (Gibco-BRL) with 10% FBS, non-essential amino acids, and penicillin-streptomycin. *Ragi* cells were electroporated with the Gene Pulser Xcell eukaryotic system (BIO-RAD) at 260 v/965 μF.

Antibodies. Anti-bfRAG1L serum was generated by immunizing mice with purified recombinant GST-bfRAG1L fusion protein that had been expressed in *E. coli* BL21 (DE3). Antibodies to GST and RAG2 were from Santa Cruz. Antibody to Flag was from Sigma.

Sequence Analysis. The PSI-BLAST (5) was used for analysis of the similarity between bfRAG1L and all the GenBank proteins, excluding jawed vertebrates proteins, with E-value cut-off of 5. Multiple alignments of bfRAG1L and virus proteins were produced by COBALT (6). Shading and refinement of the sequence alignment were performed in Genedoc (7). Phylogenetic trees were constructed by using MEGA2 (4).

Recombinant Protein Purification. The *bfRAG1L* gene was generated by PCR with primer sets P10 and P11 (*SI Appendix*, Table S6) from genomic DNA of the amphioxus *Branchiostoma floridae* (GenBank Trace Archive identification numbers 543943730, 538583629). The genes encoding ntRAG1, bfRAG1L mutants, Ch-bfRAG1L and bfRAG1L-like truncations of RAG1 from other species were cloned into vector pET28A and were expressed in *E. coli* BL21 (DE3). The bacterial culture was sonicated in Ni-lysis buffer containing 20 mM imidazole. After centrifugation, the supernatant was passed through pre-equilibrated Ni-Sepharose resin (GE Healthcare) and stored at -20°C. The proteins were eluted after thorough

washing, and the purity was determined by SDS/PAGE and Coomassie staining. In hairpin formation experiment, cRAG1, cRAG2 and Ch-bfRAG1L were cloned into pEBG vector and transiently expressed in HEK-293T cell line. Purification protocol was performed as described before (8).

Immunodepletion. The polyclonal antibodies to bfRAG1L and normal mouse IgG were cross-linked to Protein A/G beads (Santa Cruz). The Ni-Sepharose resin-purified protein bfRAG1L was mixed with antibody-treated Protein A/G beads in IP buffer at 4°C. Supernatant after centrifugation was incubated once again with treated-Protein A/G beads. This process was repeated three times. The supernatant was concentrated and substituted with elution buffer using a 10 kDa ultracentrifugal filter (Millipore).

Nuclease Activity Assay. DNase activity was detected using supercoiled plasmid pBR322 (Promega) or genomic DNA (gDNA) from HEK-293T cells as substrates. The substrate (0.25 μg) was incubated with 0.1 μg or 0.2 μg of purified protein as indicated at 37°C for 8 hr in buffer containing 20 mM Tris-HCl, pH 8.5, 10 mM $MgCl_2$, 1 mM dithiothreitol and 100 mM KCl. For RNase activity, the same amount of purified protein was incubated with RNA from HEK-293T at 37°C for 3 hr in 1 × RT buffer (Invitrogen) supplemented with 5 mM $MgCl_2$. The digested products were resolved using 1% agarose gels. For metal ion-dependent activity, the same nuclease assays were executed in the presence of different concentrations of Mg^{2+} or EDTA. The religation reaction was carried out as follows: the linear plasmid pBR322 (digested by *Bam*H I) was mixed with purified bfRAG1L or DNase I at 37°C. The nucleic acids were extracted using phenol/chloroform, precipitated with ethanol and dissolved in ddH_2O. Thereafter, all preparations were religated overnight at 16°C and analyzed using 1% agarose gel. Sequence analysis of the cleavage sites within substrate plasmid pBR322 was performed as follows: after resolution using 1% agarose gel, the digested products were ligated to the blunt ends of pCDNA3.1 and transformed into *E.coli* DH5α for subsequent sequence analysis.

Immunoprecipitation and Immunoblotting. The lysates of HEK-293T cells transfected with pEBG-cRAG2 and different types of RAG1 were incubated with anti-Flag agarose at 4°C. Ethidium bromide (10 mg/ml) was included in the binding

reaction to exclude nonspecific DNA-mediated interactions. After washing with IP buffer, the precipitates were eluted and analyzed by immunoblotting.

Confocal Laser-Scanning Immunofluorescence Microscopy. HEK-293T cells were transfected with plasmids as indicated to determine the subcellular localization of the transfection products. Cells were grown on coverslips and fixed with 4% (w/v) PFA for 15 min. After washing, DAPI (4', 6-diamidino-2-phenylindole•2HCl) (Sigma) was used for 5 min to stain the nuclei. After washing the cells, images were acquired with a TCS SP2 confocal microscope.

Bacterial Colony Assay. The plasmid substrates pJH200, pJH289 or pJH290 (0.7 μg) and bfRAG1L (1.5 μg) and RAG2 (1.5 μg) expression vectors were cotransfected into HEK-293T cells. After 48 hours, plasmid DNA were recovered from the cells by alkaline lysis, treated with *Dpn* I and were transformed into *E. coli* DH5α and spread on an LB agar plate containing 100 μg/ml Amp and 11 μg/ml Cam. The colonies were used for sequence analysis.

Luciferase Assay. HEK-293T cells were cultured and transfected with 0.5 ng of Renilla luciferase, 250 ng of reporter substrate, 1.5 μg of each type of RAG1 expression vector and 1.5 μg of cRAG2-VP16 or Flag-SpRAG2L-VP16 or 3 μg of pEBG-GST as a control. After 48 hours, the luciferase activities were detected using the Promega dual-luciferase assay system. The activity of each sample was caculated relative to the RAG-free control. All experiments were performed in triplicate.

Nucleotide Sequence Analysis of Recombiantion products. Cell-based PCR assays were carried out as described previously (9). The recombination substrates pJH290 or pJH289 (0.7 μg) and different types of RAG1 (1.5 μg) and RAG2 (1.5 μg) were cotransfected into HEK-293T cells. After 48 hours, the recombinant products were isolated using alkaline lysis and were analyzed by PCR with primer sets P1 and P2 (*SI Appendix*, Table S6) for coding and signal joint detection, respectively. For the LMPCR assay, the recovered DNA was ligated with linker by annealing FM11 to FM25 (*SI Appendix*, Table S6) overnight at 16°C and analyzed with primers P3 and P4 (*SI Appendix*, Table S6). PCR products were isolated using 2% agarose gel or 12% polyacrylamide, subjected to TA cloning (TaKaRa), and transformed into *E. coli*

DH5α for sequence analysis.

Detection of Coding and Signal Joint Formation by Fluorescent Reporter Substrates. To detect V(D)J recombination, HEK-293T cells were transfected with plasmids pCJGFP or pSJGFP (0.7 μg) and different types of RAG1(1.5 μg) and RAG2 (1.5 μg), or an equal amount of an empty vector as indicated using a previously described protocol (2).

Detection of Hairpin Formation by Oligonucleotide Cleavage Assay. Prenicked 12-RSS substrate was made by annealing the digoxigenin-labeled oligonucleotide DAR42 to DAR40 and DG10 as described previously (10). Cleavage conditions were based on previously described (8). Specially, purified 2 μg each of RAG complex were incubated with prenicked 12-RSS and unlabeled 23-RSS (DG61/62) substrates at 37°C for 1 hr. Products were resovled on a 20% denaturing polyacrylamide gel containing 7 M urea and electroblotted onto the nylon membrane. After UV-crosslinking, the membrane was immunodetected with anti-digoxigenin AP conjugate and visualized with chemiluminescence substrate.

Measurement of V(D)J Recombination by Chromosomally Integrated Retroviral Substrate. The pMX-INV retroviral recombination substrate was integrated into *Raji* cell genomic DNA, as previously described (3). Cells were electroporated with each type of RAG1 and RAG2 expression vectors as indicated and grown for 48 hours after transfection. GFP-positive cells were measured as described above.

Murine Retroviral Bone Marrow Transplantation. Retrovirus-mediated gene transfer was performed as described previously (11). In brief, bone marrow cells from $RAG1^{-/-}$ mice were isolated following injection of 5-fluorouracil. After cultured in prestimulation medium containing 50 ng/ml rmSCF, 6 ng/ml rmIL-3 and 10 ng/ml rmIL-6 (PeproTech) for 24 h, bone marrow cells were resuspended with pMCs-GFP-based retroviral supernatant in the presence of 8 μg/ml polybrene (Sigma) and the same cytokines as mentioned above, and they were then centrifuged at 1500 ×*g* for 1.5 h at 32°C. The same process was repeated the following day. The recovered bone marrow cells were injected into the tail vein of 200 cGy-irradiated recipient $RAG1^{-/-}$ mice. Total thymocytes and bone marrow cells from these recipient $RAG1^{-/-}$

mice were used for cell staining and detection of endogenous gene rearrangement after 4 weeks as described previously (9, 12).

Statistical Analysis. All data are expressed as the mean ± SD. The results were analyzed using Student *t* test. *P* values < 0.05 were considered statistically significant.

References

1. Difilippantonio MJ, et al. (1996) RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* 87:253-262.
2. Corneo B, et al. (2007) Rag mutations reveal robust alternative end joining. *Nature* 449:483-486.
3. Bredemeyer AL, et al. (2006) ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* 442:466-470.
4. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244-1245.
5. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
6. Papadopoulos JS, Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23:1073-1079.
7. Nicolas KB, Nicolas HB Jr, Deerfield DW II (1997) GeneDoc: Analysis and Visualization of Genetic Variation. EMBNEW News 4:14.
8. Spanopoulou E, et al. (1996) The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* 87:263-276.
9. Wang X, et al. (2008) Regulation of Tcrb recombination ordering by c-Fos-dependent RAG deposition. *Nat Immunol* 9:794-801.
10. McBlane JF, et al. (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 83:387-395.
11. Persons DA, et al. (1997) Retroviral-mediated transfer of the green fluorescent protein gene into murine hematopoietic cells facilitates scoring and selection of transduced progenitors in vitro and identification of genetically modified cells in vivo. *Blood* 90:1777-1786.
12. Dudley DD, et al. (2003) Impaired V(D)J recombination and lymphocyte development in core RAG1-expressing mice. *J Exp Med* 198:1439-1450.