Evolution of variable lymphocyte receptor B antibody loci in jawless vertebrates

Sabyasachi Das[a,b,1], Jonathan P. Rast[a,b], Jianxu Li[a,b,2], Mitsutaka Kadota[c], John A. Donald[d], Shigehiro Kuraku[c,e,f], Masayuki Hirano[a,b], and Max D. Cooper[a,b,1]

[a]Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA 30322; [b]Emory Vaccine Center, Emory University, Atlanta, GA 30317; [c]Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research, 650-0047 Kobe, Japan; [d]School of Life and Environmental Sciences, Deakin University, Geelong, VIC 3220, Australia; [e]Molecular Life History Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics, 411-8540 Shizuoka, Japan; and [f]Department of Genetics, Graduate University for Advanced Studies, Sokendai, 411-8540 Shizuoka, Japan

Contributed by Max D. Cooper; received September 9, 2021; accepted November 10, 2021; reviewed by Thomas Boehm and Masanori Kasahara

Three types of variable lymphocyte receptor (VLR) genes, *VLRA*, *VLRB*, and *VLRC*, encode antigen recognition receptors in the extant jawless vertebrates, lampreys and hagfish. The somatically diversified repertoires of these VLRs are generated by serial stepwise copying of leucine-rich repeat (LRR) sequences into an incomplete germline *VLR* gene. Lymphocytes that express VLRA or VLRC are T cell–like, while VLRB-expressing cells are B cell–like. Here, we analyze the composition of the *VLRB* locus in different jawless vertebrates to elucidate its configuration and evolutionary modification. The incomplete germline *VLRB* genes of two hagfish species contain short noncoding intervening sequences, whereas germline *VLRB* genes in six representative lamprey species have much longer intervening sequences that exhibit notable genomic variation. Genomic clusters of potential *LRR* cassette donors, fragments of which are copied to complete *VLRB* gene assembly, are identified in Japanese lamprey and sea lamprey. In the sea lamprey, 428 *LRR* cassettes are located in five clusters spread over a total of 1.7 Mbp of chromosomal DNA. Preferential usage of the different donor cassettes for *VLRB* assemblage is characterized in our analysis, which reveals evolutionary modifications of the lamprey *VLRB* genes, elucidates the organization of the complex *VLRB* locus, and provides a comprehensive catalog of donor *VLRB* cassettes in sea lamprey and Japanese lamprey.

jawless vertebrates | variable lymphocyte receptor | immune system evolution | adaptive immunity | transposable elements

Both humoral and cellular arms of adaptive immunity are present in jawed and jawless vertebrates (1, 2). However, whereas antigen recognition in jawed vertebrates is mediated by immunoglobulin (Ig) domain-based B cell receptors and T cell receptors (TCRs), antigen recognition in jawless vertebrates is mediated by leucine-rich repeat (LRR)-based variable lymphocyte receptors (VLRs) (2–6). Three types of *VLR* genes (*VLRA*, *VLRB*, and *VLRC*) have been found in the extant jawless vertebrates, lampreys and hagfish (3, 7–10). The germline *VLR* genes are incomplete in that they possess sequences encoding only the invariant N- and C-terminal portions of mature *VLR* separated by noncoding intervening sequences. The germline *VLR* genes are flanked by hundreds of different LRR-encoding sequences and these serve as templates for serial piece-wise replacement of the intervening sequences during the multistep assembly of mature *VLR*s for lymphocyte expression (3, 4, 11–14). *VLR* gene assembly involves a gene conversion–like mechanism mediated by cytosine deaminases, CDA1 and CDA2, which are AID-like members of the AID/APOBEC family (4, 8, 11, 15, 16). The fully assembled *VLR* genes are expressed in a monoallelic and lineage-specific fashion by lymphocytes (17). VLRA and VLRC are expressed respectively by T-like cells that resemble the TCR$\alpha\beta^+$ and TCR$\gamma\delta^+$ cells in jawed vertebrates, whereas VLRB is expressed by B-like cells, which respond to antigen stimulation by proliferation and differentiation into mature plasma cells that secrete multivalent VLRB antibodies (17–19). Coincident with *CDA1* expression, the *VLRA* and *VLRC* receptor genes are assembled within thymus-equivalent ("thymoid") regions located in the tips and proximal filaments of the gill folds, whereas *VLRB* receptor gene assembly is dependent on *CDA2* expression by cells within gut-associated and kidney hematopoietic tissues (16, 17, 20).

Lampreys and hagfish last shared a common ancestor with each other ~480 million y ago (Ma) and with jawed vertebrates >500 Ma (21). This pivotal position in chordate phylogeny makes them informative representatives for studies of the evolution of adaptive immunity. The studies here are focused on a comparison of *VLRB* genes in two hagfish and six lamprey species to gain insight into their complex genomic structures and evolution.

Results

Size Differences in Agnathan VLR Germline Genes. Germline *VLRA*, *VLRB*, and *VLRC* genes were compared in two hagfish species, the Pacific hagfish (*Eptatretus stoutii*) and inshore

Significance

Jawless vertebrates, lampreys and hagfish, use three types of variable leucine-rich repeat proteins as antigen receptors: VLRB on B-like cells and VLRA or VLRC on T-like lymphocytes. In their incomplete germline status, the *VLRB* genes in different lamprey species uniquely possess long noncoding intervening regions, which contain different sets of transposable elements and potential donor cassettes. Phylogenetic analysis of the germline *VLR* genes reveals ongoing evolution over the last 250 million y. Comparative assessment of genomic sequences in two lamprey species yields a detailed map of their *VLRB* loci. The use of this map, together with an analysis of cassette usage in a large repertoire of expressed *VLRB*s, supports a fragmented Lego-like model of *VLRB* assembly.

Author contributions: S.D., J.P.R., M.H., and M.D.C. designed research; S.D., J.P.R., J.L., M.K., J.A.D., S.K., and M.H. performed research; M.K., J.A.D., and S.K. contributed new reagents/analytic tools; S.D., J.P.R., J.L., M.K., J.A.D., S.K., M.H., and M.D.C. analyzed data; and S.D., J.P.R., M.H., and M.D.C. wrote the paper.

Competing interest statement: M.D.C. is a cofounder and shareholder of NovAb, Inc., which produces lamprey antibodies for biomedical purposes, and J.P.R. is a consultant for NovAb. One of the reviewers (T.B.) and three of the authors (S.D., M.H., and M.D.C.) contributed to a review article published online in 2017 (14).

Reviewers: T.B., Max Planck Institute of Immunobiology and Epigenetics; and M.K., Hokkaido Daigaku Daigakuin Igaku Kenkyuin.

[1]To whom correspondence may be addressed. Email: mdcoope@emory.edu or sdas8@emory.edu.

[2]Present address: AnalyticsIQ, Inc., Atlanta, GA 30346.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116522118/-/DCSupplemental.

Published December 8, 2021.

hagfish (*Eptatretus burgeri*), and in six lamprey species. Among the six lamprey species, the sea lamprey (*Petromyzon marinus*, *Pm*), Japanese lamprey (*Lethenteron camtschaticum*), European brook lamprey (*Lampetra planeri*), and Far Eastern brook lamprey (*Lethenteron reissneri*) belong to the Petromyzoninae subfamily, whereas the pouched lamprey (*Geotria australis*) and the short-headed lamprey (*Mordacia mordax*) belong to the distantly related Geotrinae and Mordaciinae subfamilies, respectively, which diverged from the Petromyzoninae lineage ∼250 Ma (21). Although each of the germline *VLR* genes contains two segments encoding N- and C-terminal portions of the mature VLR protein, they are quite different from each other, most notably in the length of their noncoding intervening regions. Intervening regions of the *VLRA* and *VLRC* genes are relatively short in both hagfish and lampreys, ranging from 157 bp to 186 bp in length (Fig. 1*A* and *SI Appendix*, Table S1). The intervening regions of the hagfish *VLRB* genes are even shorter, being only 104 bp in the Pacific hagfish and 118 bp in the inshore hagfish, whereas the intervening sequences of the lamprey *VLRB* genes are much larger, ranging from 6,936 bp to 19,261 bp. Interestingly, in the Far Eastern brook lamprey, the length of the intervening sequences of the two complete copies of *VLRB* germline genes are very different in size.

Frequencies of VLR-Expressing Lymphocyte Populations in Agnathan Representatives. To compare the relative frequencies of the different VLR-expressing lymphocytes in lamprey and hagfish, we used previously reported monoclonal and polyclonal antibodies for sea lamprey (17, 19) and a new panel of monoclonal and polyclonal antibodies with specificity for Pacific hagfish VLRs, including two monoclonal anti-VLRA antibodies, two monoclonal anti-VLRC antibodies, and a monoclonal antibody and rabbit antiserum against VLRB, the specificities of which were verified by immunofluorescence analysis of transfectants (*SI Appendix*, Fig. S1*A*) and native VLRA$^+$, VLRC$^+$, and VLRB$^+$ cells in the blood of adult hagfish (*SI Appendix*, Fig. S1*B*). Our results indicate that VLRC$^+$ lymphocytes outnumber the VLRA$^+$ and VLRB$^+$ lymphocytes in hagfish, whereas the VLRB$^+$ lymphocytes are predominant in sea lampreys (*SI Appendix*, Fig. S1 *C* and *D*).

Genomic Donor Cassettes in the Lamprey Germline VLRB. The much larger intervening regions in the lamprey *VLRB* germline genes, relative to those of hagfish *VLRB* genes and to all of the *VLRA* and *VLRC* genes in agnathans, led us to focus initially on the genomic organization of *VLRB* loci in lampreys. A notable feature of the lamprey *VLRB* germline genes is the presence of *LRR* donor cassettes in the intervening region (Fig. 1*A*). The pouched lamprey (Geotrinae subfamily) and lampreys belonging to the Petromyzoninae subfamily have at least one *5′LRRCT* cassette located between long stretches of noncoding intervening sequence (encoding the 5′ and middle portion of the mature LRRCT region and designated *inv-5′LRRCT*) (Fig. 1 *A* and *B*). *P. marinus*, *L. reissneri*, and *L. planeri* have one *inv-5′LRRCT*, whereas *L. camtschaticum* has two such intervening cassettes that share 97.5% nucleotide sequence identity (designated as A and B in Fig. 1*A*). In contrast, the germline *VLRB* intervening sequence of the short-headed lamprey (Mordaciinae subfamily) contains no *inv-5′LRRCT* cassette; instead, it includes a multiplex cassette encoding a composite LRRNT-LRR1-LRRV–N-terminal LRRV region.

Two of the lamprey species studied here have either a partial or a complete duplication of their germline *VLRB* genes. The partial *VLRB* gene in the sea lamprey genome, which was previously described (3), is located 214 kb downstream of the intact germline *VLRB* in scaffold NC_046089.1; it contains a 5′UTR and N-terminal coding region that is identical in nucleotide sequence to the germline *VLRB* gene and a *5′LRRCT* cassette (designated as cassette B) that is closely related to the *VLRB* germline *inv-5′LRRCT* (designated cassette A; 91.7% nucleotide identity). This duplicate is present also in the alternative (allelic) assembly of the sea lamprey genome (kPetMar1.alt, scaffold JAAIYF01001022.1) in the same configuration as in the primary genome assembly. Although mis-assemblies that incorporate portions of different haplotypes in tandem are possible for these complex loci, the presence of this same duplicate sequence in multiple animals, considerable differences in sequence outside of small regions of identity in the signal peptide and invariant portion of the LRRNT, and the similar organization of the haplotype assemblies support the presence of the duplication in the configuration reported in the current genome assembly.

The genome of the Far Eastern brook lamprey (22) includes a complete duplication of the germline *VLRB* gene. This duplicate copy also contains one *inv-5′LRRCT* cassette (Fig. 1*A*). The N-terminal coding regions of these two copies (encoding signal peptide and 5′LRRNT) have 100% nucleotide identity, whereas the *inv-5′LRRCT* cassettes and C-terminal coding regions (encoding the invariant 3′LRRCT and stalk region) have 97.5% and 99.6% nucleotide sequence identities, respectively. The Far Eastern brook lamprey duplicate falls in a region of an inverted partial duplication of the cassette cluster containing the germline *VLRB* gene that is found also in the Japanese lamprey, although in this case the germline gene is not present in this duplicated region and has presumably been lost.

Phylogenetic Analyses of Germline *VLRB* Gene Components. We conducted phylogenetic analyses based on the N- and C-terminal coding regions and the inv-*5′LRRCT* cassettes of the lamprey *VLRB* genes. In the unrooted phylogenetic trees of both N-terminal (Fig. 1*C*) and C-terminal (Fig. 1*D*) coding regions, all Petromyzoninae subfamily lampreys (Japanese lamprey, European brook lamprey, Far Eastern brook lamprey, and sea lamprey) are clustered together, whereas short-headed lamprey (Mordaciinae subfamily) and pouched lamprey (Geotrinae subfamily) cluster in a more distant clade. In the phylogenetic tree based on *inv-5′LRRCT* cassettes, the pouched lamprey *inv-5′LRRCT* cassette is distantly related from those of the Petromyzoninae subfamily lampreys (Fig. 1*E*). Both N-terminal and *inv-5′LRRCT* cassette phylogenies suggest that the partial *VLRB* duplication event was specific for the sea lamprey lineage. Similarly, the phylogenetic analysis of N- and C-terminal coding regions indicate the species-specific duplication of germline *VLRB* in Far Eastern brook lamprey. High sequence identity in the N- and C-terminal coding regions, as well as in the *inv-5′LRRCT* cassettes between the two copies of germline *VLRB* genes, supports this scenario. Although not supported by high bootstrap values, the phylogenetic analysis suggests the possibility that duplication of the *inv-5′LRRCT* cassette in the Japanese lamprey germline *VLRB* gene may have occurred prior to the divergence of some members of the genus *Lethenteron* and that an *inv-5′LRRCT* duplicate was subsequently lost in *L. reissneri*.

Usage of inv-5′LRRCT Cassettes. We sought to determine the usage of the *inv-5′LRRCT* cassettes in the European brook, Japanese lamprey, and sea lamprey, wherein extensive assembled cDNA sequences are available. By analyzing the mature fully-assembled *VLRB* sequences, we found that the *inv-5′LRRCT* cassettes serve as donor cassettes (Fig. 2). Sequences of the two *inv-5′LRRCT* cassettes in the Japanese lamprey and of the single *inv-5′LRRCT* cassettes in sea lamprey and European brook lamprey are all identifiable in mature *VLRB* transcript sequences. The *inv-5′LRRCT* sequence of the partial *VLRB* duplicated region in sea lamprey (*Pm* cassette B) was also identifiable in mature *VLRB* sequences. Each of the different *inv-5′LRRCT*

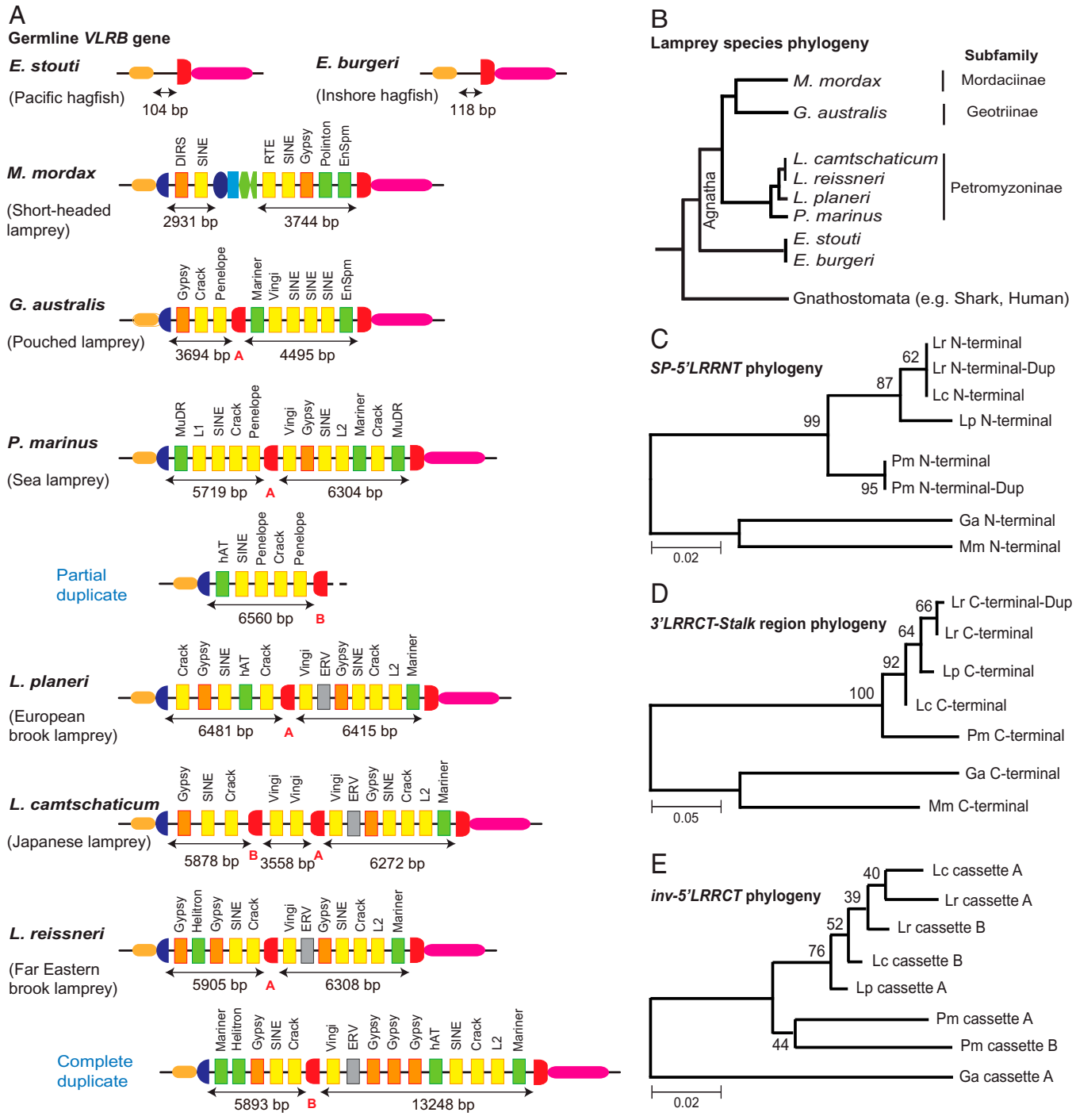

Fig. 1. Comparative analysis of incomplete germline *VLRB* genes in two hagfish and six lamprey species. (*A*) Genomic configurations of germline *VLRB* genes. The relatively large *VLRB* intervening sequences in the lampreys contain multiple TE sequences, the classification of which is indicated by color coding: orange (LTR retrotransposon), yellow (non-LTR retrotransposon), gray (endogenous retrovirus), and green (DNA transposon). Except for the germline *VLRB* in *M. mordax* (short-headed lamprey), which possesses a multiplex cassette encoding LRRNT-LRR1-LRRV–N-terminal LRRV, the *VLRB* genes in other lamprey species all contain a *5′LRRCT* cassette. However, *L. camtschaticum* (Japanese lamprey) has two *5′LRRCT* cassettes (shown as A and B) in the intervening region, whereas *P. marinus* (sea lamprey) and *L. reissneri* (Far Eastern brook lamprey) have another *inv-5′LRRCT* cassette (cassette B) in the partial *VLRB* duplicate and in the complete *VLRB* duplicate germline gene, respectively. This partial duplicate of *VLRB* in the sea lamprey is located downstream of the germline gene (see Fig. 4) as is the complete duplicate of *VLRB* in the Far Eastern brook lamprey. The figure is not drawn to scale. (*B*) The phylogenetic relationships of the eight representative agnathan species (Cyclostomata) are adapted from refs. 21, 38, and 39. The closer phylogenetic relationship between the two Southern Hemisphere lamprey lineages is suggested but remains to be validated with genome-wide dataset (38). (*C*) Phylogenetic tree based on germline *VLRB* N-terminal coding regions. (*D*) Phylogenetic tree of germline *VLRB* C-terminal coding regions, and (*E*) Phylogenetic tree of the intervening region *5′LRRCT* cassettes of germline *VLRB* genes. The VLRB phylogenetic trees (*C*–*E*) were constructed by the neighbor-joining method. The scientific names of the organisms are denoted with a two-letter code at the end of each branch: Ga, *G. australis*; Lc, *L. camtschaticum*; Lr, *L. reissneri*; Lp, *L. planeri*; Mm, *M. mordax*; Pm, *P. marinus*. The numbers next to each node indicate bootstrap confidence values after 1,000 replications.

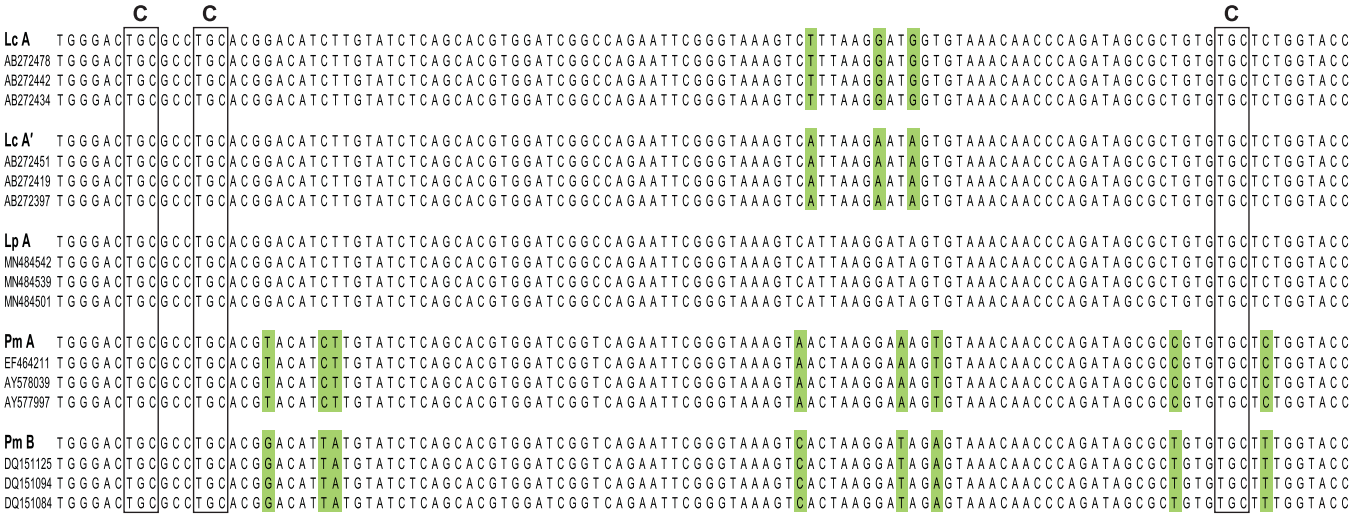

Fig. 2. Usage of *5′LRRCT* donor cassettes located in the intervening sequence of the germline *VLRB* gene in Japanese lamprey (Lc), European brook lamprey (Lp), and sea lamprey (Pm). The Lc *VLRB* gene contains two *inv-5′LRRCT* cassettes (designated as A and B), whereas the Pm *VLRB* gene has one *inv-5′LRRCT* cassette (designated as cassette A) and, in addition, a partially duplicated *VLRB* sequence together with an *inv-5′LRRCT* cassette (designated as cassette B). The genomic donor cassette sequence is shown in each upper panel and three examples of mature *VLRB* with their accession numbers are shown immediately below. Nucleotide sequence mismatches between duplicate copies of *inv-5′LRRCT* cassettes in Japanese lamprey and sea lamprey are highlighted in green. The codons encoding the conserved LRRCT cysteine residues are boxed for orientation.

cassettes can thus contribute to the generation of *VLRB* diversity in the lampreys.

Transposable Element Sequences in Germline VLRB Genes. In search of additional clues to the evolution of the germline *VLRB* intervening sequences of lampreys and hagfish, we used the Repbase reference collection of repeat elements (23) for their comparative analysis. While the hagfish *VLRB* intervening sequences did not yield hits in our similarity search against available repeats in Repbase, the *VLRB* intervening regions in all lampreys analyzed were found to contain sequences from several categories of transposable elements (TEs) (Fig. 1*A*). Among these, non–long terminal repeat (non-LTR) retrotransposons are most abundant, although LTR retrotransposons, DNA transposons, and endogenous retroviruses are also found. The diversity and richness of TEs in the lamprey *VLRB* intervening sequences implies the occurrence of multiple species-specific modifications in this region. A close association of an EnSpm DNA transposon to the C-terminal coding region is conserved in the pouched lamprey and the short-headed lamprey, whereas the close association of the Mariner-type DNA transposon sequence to the C-terminal coding region as well as the close association between the Vingi non-LTR retrotransposon and the *inv-5′LRRCT* cassette are conserved in European brook, Japanese, Far Eastern brook, and sea lampreys.

Overall Organization of the VLRB Locus in Japanese Lamprey and Sea Lamprey. The genomic sequence data for the Japanese lamprey and sea lamprey were used for detailed analysis of *VLRB* locus organization and its potential evolutionary dynamics. Using an iterative similarity search strategy (*Materials and Methods*), we determined the minimal number and physical map position of potential *LRR* donor cassettes flanking the incomplete germline *VLRB* gene (Fig. 3 and *SI Appendix*, Table S2). In addition to the germline *VLRB* gene, 384 different potential donor cassettes were identified in 22 genomic scaffolds and 91 contigs in the Japanese lamprey genome assembly (*SI Appendix*, Table S2). Almost half of the *LRR* donor cassettes (168) are located in two scaffolds (scaffold00524, which contains the germline *VLRB* gene, and scaffold00104) (Fig. 3). Unlike the *VLRC* locus, which includes only five types of *LRR* cassettes as defined by domain composition (12, 13), the donor cassettes in the *VLRB* locus of Japanese lamprey includes at least 33 categories of donor cassettes based on LRR composition (*SI Appendix*, Table S2). Most of these putative donor cassettes are composed of multiple LRR modules, some of which encode repeating units of the same type of modular subunits (e.g., LRRV-LRRV-LRRV) while others encode multiple types of subunits (e.g., LRRNT-LRR1-LRRV). The largest such multiplex cassette in Japanese lamprey encodes a total of six LRRV subunits (*SI Appendix*, Table S2).

These different categories of donor cassettes are intermingled in the *VLRB* locus with varying spacing, except for the regions sharing potential duplication events, which exhibit repetitive congruent spacing (Fig. 3 and *SI Appendix*, Table S2). Sequence analysis of donor cassettes in these scaffolds suggests duplication events involving single or multiple types of donor cassettes, as denoted in Fig. 3. The authenticity of these presumptive duplication events is supported by phylogenetic analysis, the high degree of sequence similarity for the duplicated cassette pairs and their flanking sequences (≥95%), and by the congruent genomic orientation of their constituent cassettes. For example, a block duplication comprising two consecutive *LRRNT-LRR1-LRRV-5′LRRV* multiplex cassettes is identifiable upstream from the germline *VLRB* gene in scaffold00524 (Fig. 3). Scattered events of short tandem duplication of regions carrying one or two cassettes are also noted in the Japanese lamprey *VLRB* locus.

Of special note, sequence from a *5′LRRCT* cassette located in an intron of a flanking *carbohydrate sulfotransferase 14*–like (*CHST14*-like) gene (Fig. 3 and *SI Appendix*, Fig. S2) is used in two mature *VLRB* sequence (accession nos. AB275693 and AB272468), while transcriptome data indicates expression of the *CHST14*-like gene. This intronic *5′LRRCT* cassette is also found in the predicted *CHST14*-like gene in the sea lamprey (discussed below) and the Far Eastern brook lamprey genome (accession no. PRJNA558325). High nucleotide sequence identity (96%) is observed between the intronic *5′LRRCT* cassette of *CHST14*-like gene in sea lamprey and Japanese lamprey.

To gain a broader view, we identified a total of 428 VLRB-encoding genomic donor cassettes in a comprehensive map of the sea lamprey *VLRB* locus. Most of the cassettes (n = 358)

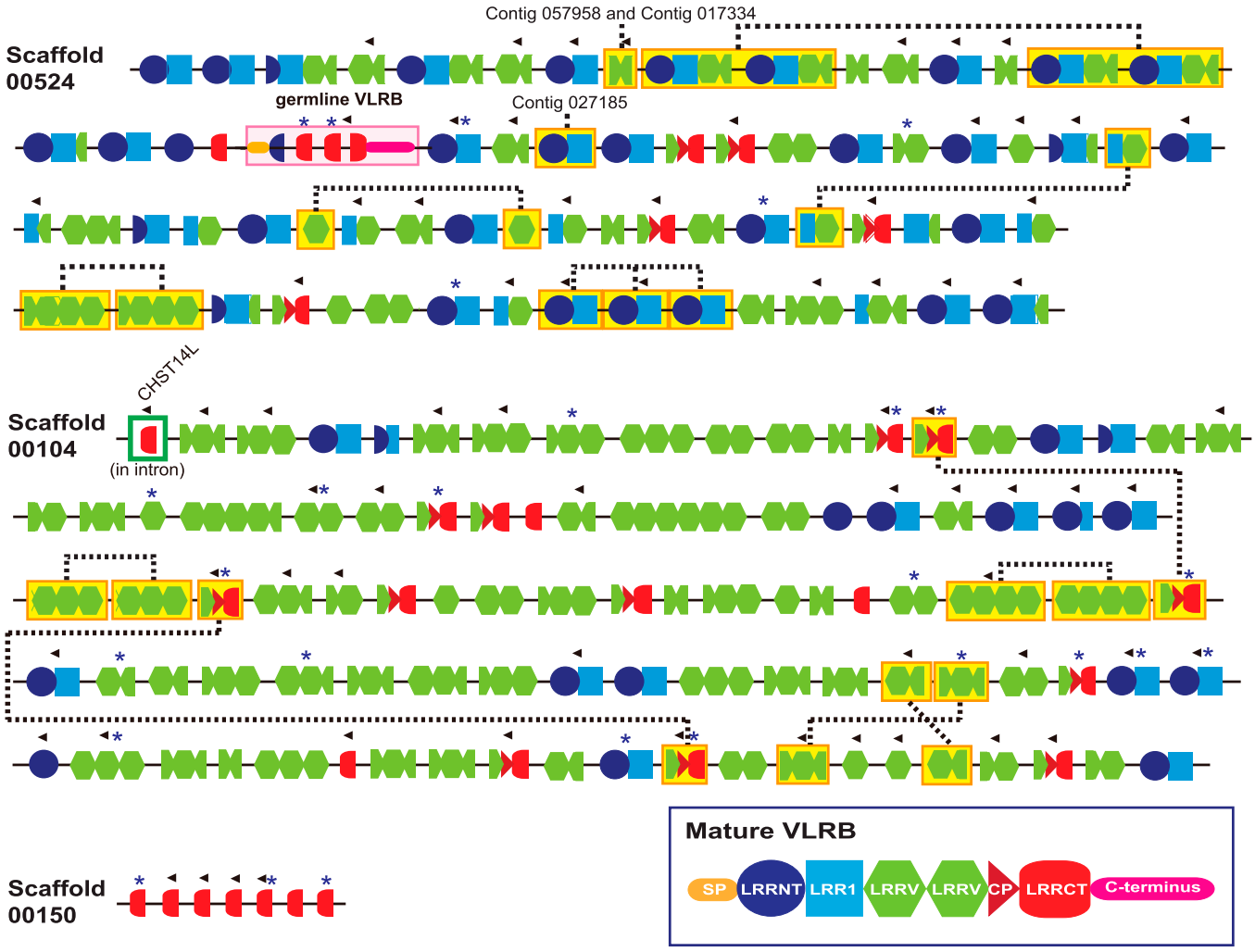

Fig. 3. *VLRB* locus organization in the Japanese lamprey. The incomplete germline *VLRB* gene (red box in Scaffold00524) and 384 donor cassettes are distributed in 22 genomic scaffolds and 91 contigs (*SI Appendix*, Table S2). Three representative *VLRB* scaffolds are shown here because of their length and large number of constituent *VLRB* donor cassettes. Presumptive duplication events for VLRB-encoding donor cassettes (indicated by background yellow colors) are connected by dotted black line. Frequently used donor cassettes are indicated by an asterisk. Arrowheads above individual donor cassettes indicate their reverse orientation relative to other donor cassettes in a particular scaffold. In scaffold00104, a *5′LRRCT* cassette is within the intron of the *CHST14* gene. The fragmented genome assembly precludes determination of the relative order and orientation of the scaffold. The illustrated components are not drawn to scale and regions of unresolved sequence are ignored. A cartoon of an assembled *VLRB* is shown in the black box at the bottom right).

are found on the same scaffold as the germline *VLRB* gene (National Center for Biotechnology Information [NCBI] accession no. NC_046089.1) in two major clusters (clusters I and II) and one smaller cluster (cluster III) (Fig. 4 *A* and *B* and *SI Appendix*, Table S3). Clusters I and II contain 167 and 178 donor cassettes, respectively, whereas cluster III contains an array of 13 *5′LRRCT* cassettes. The germline *VLRB* gene and its partial duplicate (Fig. 1) are in cluster II in the same transcription orientation, whereas the *CHST14*-like gene with an intronic *5′LRRCT* cassette (Fig. 3) is located next to cluster I (Fig. 4 *A* and *B*). Notably the three clusters of sea lamprey scaffold NC_046089.1 correspond to the three Japanese lamprey scaffolds detailed in Fig. 3 (also indicated in Fig. 4*A*). Similarly, these three genomic clusters are found at similar positions in a recent chromosomal level assembly of the Japanese lamprey (Scaffold MU249593.1 in assembly GCA_018977245.1) (24), and of the Far Eastern brook lamprey, as evident in the dot plot comparison (Fig. 4*C* and *SI Appendix*, Fig. S4). Two additional *VLRB* cassette clusters are present in the sea lamprey, one on a smaller, unplaced scaffold (cluster IV, scaffold NW_022639073.1) with 55 *5′LRRCT* cassettes and eight *3′LRRV-CP* cassettes, and the other on a separate large scaffold (cluster V; NC_046122.1) with seven nearly identical *5′LRRCT* (*SI Appendix*, Fig. S3 *D* and *E* and Table S3).

As in Japanese lamprey, most sea lamprey cassettes are multiplex in that they consist of multiple subunits (*SI Appendix*, Table S3). Eighty-three sea lamprey cassettes encode at least a partial LRRNT region, often together with an LRR1 and LRRV encoding sequence. Two-hundred and nineteen cassettes contain exclusively *LRRV* elements, encoding regions as small as a partial LRRV up to one encoding eight and a half contiguous LRRV motifs. Thirty cassettes encode CP sequences, almost always downstream of a contiguous *3′LRRV* element and upstream of a few codons overlapping *5′LRRCT* sequence. Exceptions to these multiplex cassettes are the 90 cassettes that encode N-terminal LRRCT, which are always found individually and never fused to other modules (although small stretches of CP sequence are found at their 5′ end). Most of these encode the complete variable portion of the VLRB LRRCT (75 of 90 cassettes, up to the third cysteine in the domain, including the variable loop region

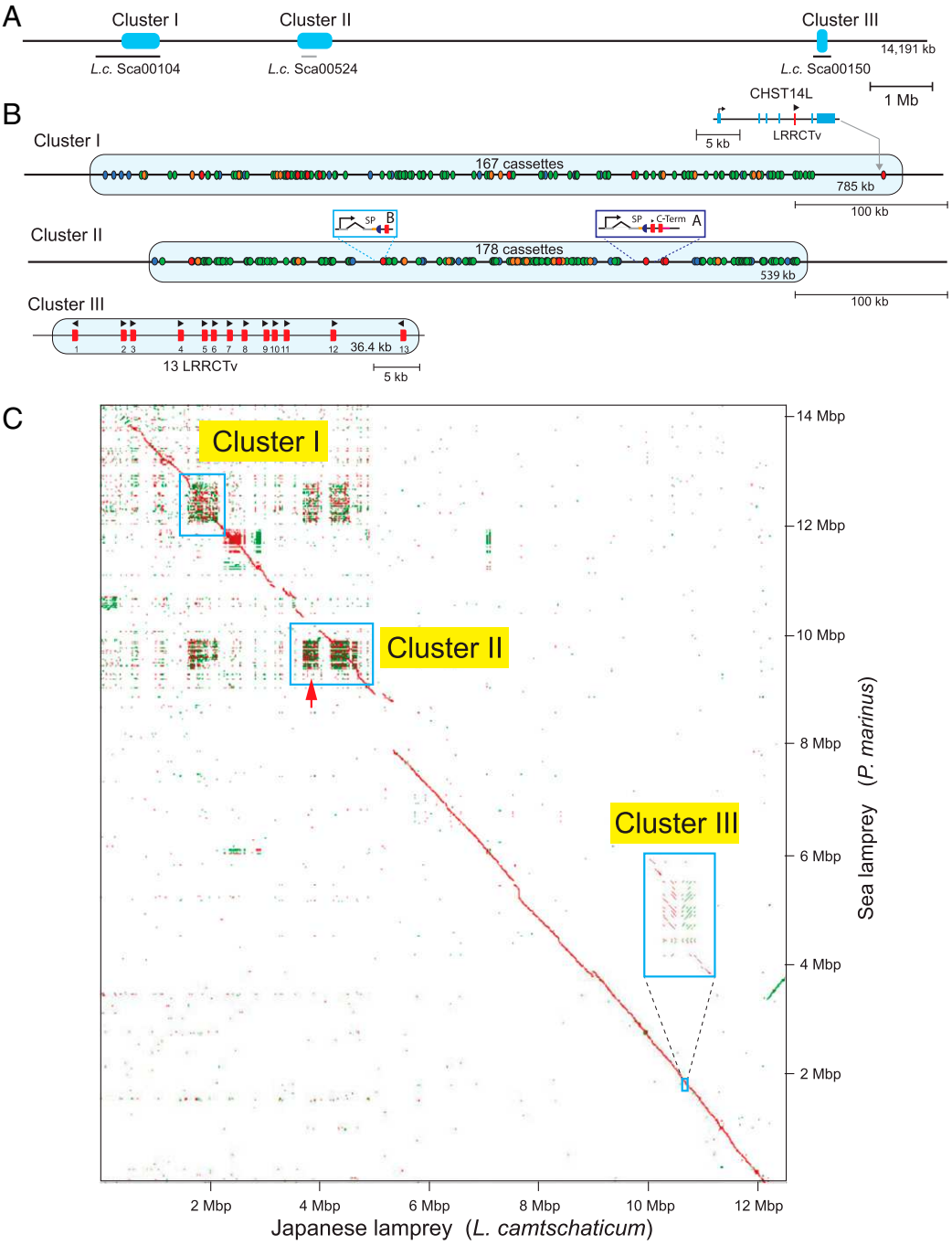

Fig. 4. Genomic map of the sea lamprey (*P. marinus*) *VLRB* locus. (*A*) *VLRB* components are shown in same orientation as the *VLRB* germline gene and in reverse complement to the *VLRB* encoding genomic reference sequence (NCBI accession no. NC_046089.1). *VLRB* donor cassettes in three regions of the chromosome are listed in the same orientation as the core germline *VLRB* gene: cluster I, a region with 167 cassettes spanning from 12.0 to 12.6 Mbp in the NC_046089.1 sequence; cluster II, spanning from 9.3 to 9.9 Mbp in the NC_046089.1 sequence, contains the germline *VLRB* gene, the partial duplication of the germline *VLRB,* and 178 donor cassettes; cluster III, a group of 13 *5′LRRCT* cassettes at 1.5 Mbp in the NC_046089.1 sequence. *VLRB* cassettes are distributed in the regions indicated with blue shading. The positions corresponding to the segments of the three Japanese lamprey scaffolds diagrammed in Fig. 3 (Lc_Sca00104, Lc_Sca00524, Lc_Sca00150) are shown as lines under the chromosome map corresponding to the three homologous sea lamprey genomic segments. Sequence for the entire locus is defined except for two small unresolved regions of 1,539 bp and 17,802 bp located well outside of the donor cassette clusters. (*B*) An expanded view of the *VLRB* regions containing clusters I to III shown in light blue shading with cassettes indicated with colored ovals and associated scale bars. The germline *VLRB* gene and its partial duplicate (see SP, signal peptide and C-term, C-terminal labels) are in the same transcription orientation in cluster II. A map of the sea lamprey CHST14L gene with a *5′LRRCT* cassette encoded in an intron (red rectangle) is shown on the right. An arrow indicates its position flanking the *VLRB* cluster I, as it does also in *L. camtschaticum* Sca00104 (Fig. 3). Arrowheads indicate transcriptional orientation. (*C*) A dot plot of alignments between sea lamprey sequence containing *VLRB* (NC_046089.1) and the homologous *VLRB* encoding sequence from the recent Japanese lamprey, *L. camtschaticum*, assembly (scaffold MU249593.1). Clusters I, II, and III are positioned similarly in genome assemblies from both species. Donor cassettes are restricted to the regions encompassed by the boxes in both species. A duplication and inversion of part of cluster II is evident in the Japanese lamprey (red arrow). An expanded view of cluster III is shown. Matching among different cassettes within clusters is evident in the checkered array of alignment marks within the three boxed cluster regions. Green marks indicate forward matches, red marks indicate reverse complement matches relative to the GenBank sequences.

that is often important in antigen binding) (14), although 11 cassettes encode more truncated N-terminal LRRCT sequences. Notably, the tandem *5′LRRCT* clusters (III, IV, V) contain a total of 76 *5′LRRCT* cassettes, 73 of which possess a loop-encoding region.

Sea lamprey *VLRB* locus clusters I and II extend across long stretches of DNA (785,355 bp and 538,636 bp, respectively) and contain similar numbers of cassettes, but they differ from each other in donor cassette compositions. For example, cluster II contains eight LRRNT-LRR1-LRRV$_n$ encoding cassettes, whereas cassettes with this composition are absent in cluster I. Cluster II also contains twice the number of *LRRNT*-containing cassettes (n = 55) than are found in cluster I (n = 25). A more detailed genomic map of the sea lamprey *VLRB* locus is shown in *SI Appendix*, Fig. S3. In most cases, donor cassettes are closely related to other cassettes within their cluster of origin, suggesting that the cluster compositions are dominated by internal duplications as opposed to whole-scale cluster duplication. There are some notable exceptions with highly similar cassettes being found in both cluster I and cluster II (e.g., the *LRRV-CP* cassettes I-100 and II-116 are identical in nucleotide sequence).

Several additional characteristics in the overall organization of the sea lamprey *VLRB* locus are notable. Donor cassette density is low in the vicinity of the incomplete germline *VLRB* gene and the partial duplicate. These "desert regions" include ∼14 kb immediately upstream and 31 kb downstream of the germline *VLRB* gene, and 17 kb upstream and 18 kb downstream of the partially duplicated germline *VLRB* gene. This spacing characteristic is also evident in the Far Eastern brook lamprey and Japanese lamprey *VLRB* loci. Another notable feature of the *VLRB* locus is the monotypic clustering of the *5′LRRCT* cassettes: the majority of these (75 of 90) are found in three clusters of related cassettes, which are located outside of the major clusters I and II that contain only eight and six *5′LRRCT* cassettes, respectively. Cassette orientation in the *5′LRRCT* clusters (clusters III, IV, and V) is biased, with strings of up to 20 tandem cassettes in the same orientation. It is notable that non-VLRB genes are apparently completely excluded from the total of 1.7 Mbp of cassette cluster DNA, although they can be relatively dense in the regions flanking the clusters.

Trends in VLRB Donor Cassette Usage. Cassette usage was assessed by searching a set of unique *VLRB* cDNA sequences for identifiable donor cassette sequences. When the frequency of matches was corrected for duplicated genomic cassettes (i.e., those with identical or nearly identical nucleotide sequences), a wide range in the frequency of usage of the different donor cassettes was evident in both Japanese lamprey and sea lamprey. Certain cassettes within each *LRR* sequence category were frequently used as templates, while usage of some of the others was never observed in our dataset of mature *VLRB* sequences (Fig. 3 and *SI Appendix*, Fig. S3).

When the most highly used LRRNT-LRR1– and LRRCT-encoding cassettes were compared with the rarely used ones in both Japanese lamprey and sea lamprey, we found that high sequence conservation is the rule for the highly utilized cassettes (*SI Appendix*, Figs. S5–S7). Both LRRNT-LRR1– and LRRCT-encoding cassettes possess three ultraconserved regions: one that is characterized by overlapping sequence positions with the germline *VLRB* gene and two other regions that provide matches with the subsequently incorporated donor cassettes. In general, these regions are more conserved in commonly used cassettes, although matches of the region corresponding to the germline gene can be similar in both highly used and rarely used cassettes. We also observed that some of the rarely utilized LRRNT-LRR1–encoding cassettes have insertions in the LRRNT-encoding region (*SI Appendix*, Figs. S5 and S6) and that some of the infrequently used cassettes also have either an internal stop codon or highly divergent sequences in their 5′ or 3′ regions (*SI Appendix*, Fig. S3 and Tables S2 and S3).

Interestingly, the most frequently used *LRRNT*-containing cassettes are primarily located in cluster II (n = 15) of sea lamprey, while a minority are in cluster I (n = 2). The exclusively *LRRV*-containing cassettes that are most frequently used are evenly split within cluster I (n = 21) and cluster II (n = 22). For the CP-containing cassettes, the most frequently used donors come from cluster II (n = 6), although cluster I also encodes two CP cassettes (I-12 and I-100) (*SI Appendix*, Fig. S3*A*) with moderate prevalence just below the 20% cutoff used here. Notably, all of the top quintile of the most frequently used *5′LRRCT* cassettes are located in clusters III (n = 13) and IV (n = 5), with cluster III *5′LRRCT* cassettes being used most frequently. Thus, while donor cassettes from all clusters are used, preferential usage related to cluster position is evident when cassette-type is taken into consideration.

Our analysis indicates that either partial or the full sequences of donor cassettes can contribute to the assembly of mature *VLRB* genes (Fig. 5*A*). In both cases, short stretches of nucleotide sequence similarity are evident between donor and acceptor sequences. In some instances, we noted that multiple regions of a single cassette donor may be used for *VLRB* gene assembly. We also found donor cassettes that have internal in frame stop codons and frameshifts, although these are rare (e.g., of the 428 identified sea lamprey cassettes, 5 have in frame stop codons and 3 have frameshifts) (*SI Appendix*, Fig. S3). However, our analysis indicates that these types of cassettes can still be functional in that partial sequences of the *LRR* cassettes can contribute to mature *VLRB* assemblage (Fig. 5*B*).

Discussion

The genomic complexity of lamprey *VLRB* loci rivals that of jawed vertebrate immunoglobulin loci, with hundreds of donor cassettes encoded over large stretches of DNA, both in the vicinity of the incomplete germline gene and in more distant clusters. Here we characterize the structure of these loci from three viewpoints: 1) we compare the genomic organization of the germline *VLRB* gene among six lamprey and two hagfish species to reveal stark phylogenetic differences in gene structure and infer trends among lamprey lineages; 2) we catalog the structure and distribution of donor cassettes in two of these lamprey species to reveal conserved aspects of cassette diversity and local clustering; and 3) we develop a comprehensive map of the sea lamprey *VLRB* locus to reveal larger scale patterns in cluster distributions. Our phylogenetic analysis identifies species differences in some aspects of *VLRB* genomic structure, but also conserved features of cassette organization, clustering, and usage.

The intervening regions of germline *VLRB* genes in lampreys are considerably larger than their counterparts in hagfish, and they are also much larger than germline *VLRA* and *VLRC* genes in both lamprey and hagfish. This is due in part to the insertion of numerous TEs (or remnants of TEs) into the intervening sequences of the lamprey germline *VLRB* gene and to the presence of donor cassettes in the intervening sequence. Interestingly, only the germline *VLRB* genes in lampreys have been expanded, leaving the *VLRA*, *VLRC*, and hagfish *VLRB* genes in a presumably ancestral short-intervening region state. A plausible explanation is that a TE (or DNA-carrying TE sequence) was inserted into the intervening sequence in a common lamprey ancestor. Once there, it was fixed in the ancestral population and subsequent duplications and insertions of additional TEs led to further lineage-specific modification of the lamprey germline *VLRB* genes. This series of changes could have been initiated by a single insertion event in a common lamprey ancestor. There may also be constraints on the upper size of the lamprey *VLRB* intervening

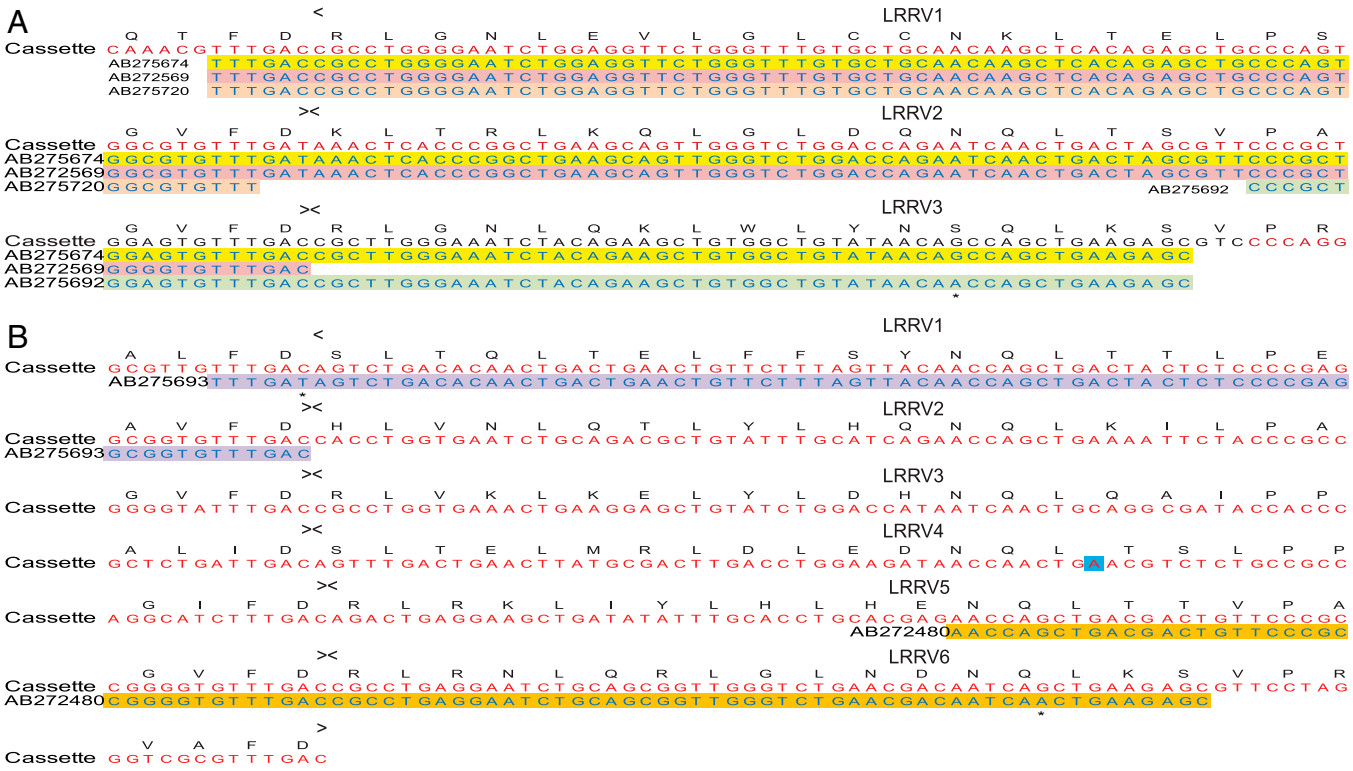

Fig. 5. Multiple segments of *LRR* cassettes may contribute to *VLRB* assembly in the Japanese lamprey. Two representative multiplex *LRRV* cassettes are shown: the donor cassette (*A*) potentially encode three modules (i.e., LRRV-LRRV-LRRV), whereas the one in *B* encodes six LRRV modules in two different reading frames (a single base pair insertion is indicated in blue). Assembled mature *VLRB* sequences, numbered and highlighted in different colors, are shown below the genomic donor cassette along with their accession numbers. The hypothetical translation of the putative *LRR* donor cassette is shown above the cassette sequence.

sequence since it ranges only about threefold from smallest to largest (*SI Appendix*, Table S1).

A short intervening sequence thus seems the likely ancestral *VLR* state, given the small size similarity for the hagfish germline *VLRB* genes and those of the germline *VLRA* and *VLRC* genes in all the agnathan representatives examined. The much longer intervening sequences found in all lamprey *VLRB* genes likely evolved after divergence of the lamprey and hagfish lineages and were then followed by ongoing lineage-specific changes. Our comparative analysis of intervening sequences between six lamprey species and phylogenetic analysis of the germline N-terminal coding region, C-terminal coding region, and the *inv-5′LRRCT* cassette of lamprey germline *VLRB* support this scenario. The presence of a multiplex *LRRNT*-containing cassette within the germline *VLRB* gene in short-headed lamprey and the presence of one or more *inv-5′LRRCT* cassettes within the germline *VLRB* genes in members of the Petromyzoninae and Geotrinae lamprey subfamilies suggests that the *inv-5′LRRCT* cassette appeared in a common ancestor of Petromyzoninae and Geotrinae subfamilies after the divergence of Mordaciinae subfamily lampreys. Alternatively, the introduction of an *LRRNT*-containing multiplex cassette may have taken place within the Mordaciinae lineage. Our results imply a more recent duplication event comprising a genomic region of the germline *VLRB* gene that contains the N-terminal coding region, several TEs, and an *inv-5′LRRCT* cassette in sea lamprey, whereas in Far Eastern brook lamprey a complete duplication of germline *VLRB* was followed (or preceded) by expansion of the intervening sequence through insertion and duplication of TEs. The presence of two *inv-5′LRRCT* cassettes in the Japanese lamprey *VLRB* intervening sequence further suggests that the germline configuration of the incomplete *VLRB* gene is evolving independently in different lamprey lineages through lineage-specific transposon activities or other forms of DNA insertion. Different types of TEs have thus accumulated over time in the intervening region of the lamprey germline *VLRB* gene and their evolutionary modification of the *VLRB* locus likely continues in lampreys.

Analysis of the overall organization of the *VLRB* locus in three lamprey species (sea lamprey, Japanese lamprey, and Far Eastern brook lamprey) reveals that the incomplete germline *VLRB* genes of each are surrounded by numerous types of genomic donor cassettes. Most of the donor cassettes are in three homologous genomic clusters in these lamprey species. Multiple genomic inversions in the *VLRB* locus are evident within these clusters, and reorganization is especially evident in cluster II, which contains the germline *VLRB* gene. This is evident in the inability to align the sequences on the diagonal and inversions indicated by orthogonal alignments (Fig. 4*C*). The diversity in the incomplete germline *VLRB* gene, and sequence divergence and inversions evident in species comparisons of the cassette clusters, demonstrate the genomic plasticity of lamprey *VLRB* loci. Examples of recent *LRR* donor cassette duplication and subsequent diversification are also evident in the *VLRB* locus. The overall genomic organization and the evolutionary dynamics of the agnathan *VLRB* locus roughly resemble the genomic organization and the mode of evolution of immunoglobulin elements in rearranging Ig and TCR loci (25, 26), despite the basic structural differences between the LRR-based antigen receptors in jawless vertebrates and the Ig domain-based antigen receptors in jawed vertebrates. The gene cassette deserts surrounding the *VLRB* germline genes could reflect

necessary space for chromatin-level control of assembly. This may indicate spacing constraints in the *VLRB* loci analogous to those required for class switching. From a genomic point of view, the incomplete germline *VLRB* gene represents an equivalent signature of the Ig constant region gene segment, whereas the clusters of different types of *LRR* cassettes exemplify the functional correlates of the clusters of variable (V), diversity (D), and joining (J) gene segments of *Ig/TCR* locus in jawed vertebrates.

Most *VLRB* donor cassettes are multiplex units that encode more than one LRR module, unlike *VLRC* donor cassettes in lampreys, which exhibit a low complexity in terms of cassette structure. Whether the fused *VLRB* donor cassettes emerge from small abortive assemblies taking place outside of the germline gene or through some other mechanism may be elucidated through future comparisons of more closely related species (e.g., *L. camtschaticum* vs. *L. reissneri*) where orthologous cassette sequence can more easily be identified. The extent of *VLRB* donor cassette multiplicity also exceeds that for *VLRA* (13). Because *VLRB*-encoding cassettes are incorporated into mature *VLRB*s either as a full-length sequence or as fragments, the potential diversity in the lamprey *VLRB* repertoire may be much greater than the previous estimate of 10^{14} (8). It is notable that, like the biased usage of immunoglobulin heavy-chain variable region gene segments (*IGHV*) in humans and mice (27, 28), *VLRB*-encoding cassettes are used in mature *VLB* genes with varying prevalence. However, in contrast to the biased usage of proximal *IGHV* genes in mice (28), we could not detect usage bias related to donor cassette proximity to the incomplete germline *VLRB*. For example, frequently used *LRRV*-containing cassettes are evenly spaced throughout clusters I and II, whereas distant clusters can contain high-frequency *5′LRRCT* cassettes. The *inv-5′LRRCT* donor cassettes in sea lamprey are used with moderate frequency in mature *VLRB*s, whereas the more distant *5′LRRCT* cassettes in cluster III (8 Mbp from the germline *VLRB* gene) are used with highest frequency. Our analysis suggests that the sequence conservation is also a determinant of biased usage of donor cassettes in *VLRB* assembly (*SI Appendix*, Figs. S5–S7).

Inspection of a genomic map for the sea lamprey *VLRB* locus reveals that most donor cassettes are present in clusters located on the primary *VLRB* chromosome far away from the *VLRB* germline gene, with minor cassette contributions from donor cassettes in separate locations. Neither transcriptional orientation is obviously favored for highly utilized and moderately utilized donor cassettes, although biased orientation is evident in the *5′LRRCT* cassette clusters. In the process of *VLRB* assembly, donor DNA sequences from an average of 5 to 10 template cassettes must be brought from relatively large genomic distances into close juxtaposition with the recipient *VLRB* gene sequence during the template copying process. Sequences of donor cassettes from two or three of these distantly spaced clusters are often incorporated into a single *VLRB* gene assembly. These features imply that the template mediated assembly process involve chromosome-scale positioning mechanisms to control the precisely ordered interaction of distant donor cassette DNA with the germline *VLRB* gene. In some cases, as for the high-prevalence cluster III *5′LRRCT* cassettes, contributions can be highly biased toward more distant regions of the *VLRB* locus. Other aspects of the assembly process, such as the use of cluster I and cluster II *LRRV* cassettes, display little bias relative to genomic position. The extent to which lamprey *VLRB* assembly shares mechanisms with the chromosomal processes that coordinate *V(D)J* recombination, especially those that integrate the mechanics of Ig gene assembly with the wider gene regulatory networks that control of B cell development, will be an interesting area of future investigation.

Materials and Methods

Animals. Sea lamprey larvae and adults were captured in the wild and housed in our institutional animal facility. All experiments for the sea lamprey and mice were approved by the Institutional Animal Care and Use Committee at Emory University. The samples of the short-headed lamprey and pouched lamprey were collected under the Animal Ethics Project ID B29-2016 and the fisheries permit General Research RP1278 in Australia. Pacific hagfish (30- to 60-cm long) were purchased from Marinus, maintained at 14 to 17 °C in artificial sea water, and fed at 2-wk intervals with homogenized beef liver (29).

Protein Expression and Antibody Production in Hagfish. We used previously described methods for hagfish VLRA, VLRB, and VLRC recombinant protein expression and specific antibody production (17, 19). Briefly, *VLRA/B/C* cDNAs isolated from hagfish lymphocytes were cloned in-frame with the constant region of human IgG1 (IgG1-Fc). The VLR–IgG1-Fc fusion proteins were expressed in HEK-293T cells and purified from tissue culture supernatants by protein A chromatography (GE Healthcare). Monoclonal anti-VLRA antibodies were produced by immunization of BALB/c mice with VLR–IgG1-Fc proteins emulsified in TiterMax Gold adjuvant. Lymphocytes from the draining lymph nodes of immunized mice were fused with the Ag8.653 myeloma cell line using PEG-1500 (Roche). Two VLRA-specific hybridoma clones 1A2 (IgG1) and 1A4 (IgG2a), two VLRC-specific hybridoma clones 6D5 and 8A1 (both are IgG1), and a VLRB-specific hybridoma clone were identified by staining of HEK-293T cells transfected with a plasmid expressing the extracellular domain of VLRs. Anti-VLRB polyclonal antisera were produced by immunization of rabbits with the VLRB–IgG1-Fc proteins (PickCell Laboratories BV).

Flow Cytometric Analysis and Sorting of VLRA, VLRB, and VLRC Lymphocytes. Leukocytes isolated from lamprey and hagfish blood and tissues were prepared for flow cytometry, as described previously (17, 19). In brief, leukocytes were stained with anti-VLRA mouse monoclonal antibody (6D5), anti-VLRB rabbit polyclonal serum (R509), and anti-VLRC mouse monoclonal antibody (1A2) and matched secondary reagents. Flow cytometric analysis was performed on CyAn ADP (Dako) or Accuri C6 (BD Biosciences) flow cytometers. $VLRA^+$, $VLRB^+$, $VLRC^+$, and VLR triple-negative cells in the lymphocyte gate were sorted on a BD FACS Aria II (BD Bioscience) for genomic and quantitative RT–PCR analysis. The purity of the sorted cells was $96.5 \pm 1.5\%$ ($VLRA^+$), $96.1 \pm 1.4\%$ ($VLRB^+$), $98.7 \pm 1.3\%$ ($VLRC^+$), and $97.9 \pm 1.4\%$ (triple-negative).

Statistical Analysis. Statistical significance for lymphocyte counts was determined by a two-sample Student's *t* test and Fisher's exact test.

Analysis of VLRB Loci. To identify the genomic donor cassettes for *VLRB*s, we conducted BLASTN searches against two assemblies of the Japanese lamprey genome (LetJap1.0 [GenBank assembly GCA_000466285.1] and a recent assembly IMCB_Ljap_1.0 [GenBank assembly GCA_018977245.1]) using an E-value of $\leq 10^{-5}$, as described previously for analysis of VLRA and VLRC genes (13). For the sea lamprey genome sequence (30, 31), we analyzed the latest iteration of the germline sequence (kPetMar1). For Japanese lamprey and sea lamprey, we used 248 (accession nos.: AB272387–AB272577, AB275668–AB275724) and 504 (accession nos.: AY577943–AY578057, AY578059, DQ150997–DQ151384) assembled *VLRB* sequences present in GenBank as queries, respectively. Sequence profiles for each of the different motifs (LRRNT, LRR1, LRRV, connecting peptide [CP], LRRCT-variable portion) were built using HMMBUILD and used to search all genomic open reading frames ≥30 nucleotides using HMMSEARCH (v3.3, http://hmmer.org). For analysis of hagfish (*E. stoutii* and *E. burgeri*) *VLR* genomic regions, we used three BAC clones (accession nos. AY965679, AY965680, and AY965681). To identify the open reading frame of the putative genomic donor cassettes, we used an approach described previously (13). Boundaries of donor cassettes were determined based on the protein domain characterization in the SMART database (32). The incomplete germline *VLR* genes of the short-headed lamprey (*M. mordax*, GenBank accession nos.: OK032591–OK032593) and pouched lamprey (*G. australis*, GenBank accession nos.: OK032594–OK032596) were cloned using the reference of genome sequences. The germline *VLR* sequences from European brook lamprey (accession nos. KJ744044, MN484596, KC247680) and Far Eastern brook lamprey (BioProject no. PRJNA558325) were obtained from the NCBI.

For assessing the frequency of cassette usage in sea lamprey, we carried out an ungapped BLASTN search of the 504 complete cDNA sequences described above with each of the 428 *VLRB* donor cassette sequences requiring a 97% match and an E-value of 10^{-20}. Matches for each cassette were enumerated and the totals were corrected for duplicate cassettes by dividing the total count by the number of near identical cassettes (≥97% nucleotide identity). Usage frequency was then calculated as the corrected match total divided by total match for each of four cassette categories: *LRRNT*-containing,

LRRV-only, *CP*-containing, and *LRRCT*-containing. The top 20% of cassettes ranked by usage frequency within each category (top quintile) were considered frequently used.

Bioinformatic Analyses. To identify classes of TEs, the CENSOR software tool was used that screens query sequences against a reference collection of repeat elements (33). Putative duplications of genomic donor cassettes were characterized by the methods previously described (13). Sequences were aligned by MUSCLE (34) and manually adjusted where necessary. Neighbor-joining (35) trees were constructed by MEGA software (v5) (36) and the support for each node of the phylogenetic tree was tested by 1,000 bootstrap replications. The genomic dot plot was generated with the YASS genomic similarity search tool (37) with an E-value threshold of 10^{-60} and repetitive sequence masking.

Data Availability. The sequences reported in this paper have been deposited in the GenBank database (accession nos. OK032591–OK032596).

ACKNOWLEDGMENTS. This work was supported by NIH Grants R01AI072435, R35GM122591, and GM108838; National Science Foundation Grants 1655163 and 1755418; and the Georgia Research Alliance.

1. G. W. Litman, M. K. Anderson, J. P. Rast, Evolution of antigen binding receptors. *Annu. Rev. Immunol.* **17**, 109–147 (1999).
2. M. D. Cooper, M. N. Alder, The evolution of adaptive immune systems. *Cell* **124**, 815–822 (2006).
3. Z. Pancer *et al.*, Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430**, 174–180 (2004).
4. M. N. Alder *et al.*, Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* **310**, 1970–1973 (2005).
5. M. N. Alder *et al.*, Antibody responses of variable lymphocyte receptors in the lamprey. *Nat. Immunol.* **9**, 319–327 (2008).
6. B. R. Herrin, M. D. Cooper, Alternative adaptive immunity in jawless vertebrates. *J. Immunol.* **185**, 1367–1374 (2010).
7. Z. Pancer *et al.*, Variable lymphocyte receptors in hagfish. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9224–9229 (2005).
8. I. B. Rogozin *et al.*, Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* **8**, 647–656 (2007).
9. J. Kasamatsu *et al.*, Identification of a third variable lymphocyte receptor in the lamprey. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14304–14308 (2010).
10. J. Li, S. Das, B. R. Herrin, M. Hirano, M. D. Cooper, Definition of a third VLR gene in hagfish. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15013–15018 (2013).
11. F. Nagawa *et al.*, Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nat. Immunol.* **8**, 206–213 (2007).
12. S. Das *et al.*, Organization of lamprey variable lymphocyte receptor C locus and repertoire development. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6043–6048 (2013).
13. S. Das *et al.*, Genomic donor cassette sharing during VLRA and VLRC assembly in jawless vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14828–14833 (2014).
14. T. Boehm *et al.*, Evolution of alternative adaptive immune systems in vertebrates. *Annu. Rev. Immunol.* **36**, 19–42 (2018).
15. S. J. Holland *et al.*, Expansions, diversification, and interindividual copy number variations of AID/APOBEC family cytidine deaminase genes in lampreys. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3211–E3220 (2018).
16. R. Morimoto *et al.*, Cytidine deaminase 2 is required for *VLRB* antibody gene assembly in lampreys. *Sci. Immunol.* **5**, eaba0925 (2020).
17. M. Hirano *et al.*, Evolutionary implications of a third lymphocyte lineage in lampreys. *Nature* **501**, 435–438 (2013).
18. B. R. Herrin *et al.*, Structure and specificity of lamprey monoclonal antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2040–2045 (2008).
19. P. Guo *et al.*, Dual nature of the adaptive immune system in lampreys. *Nature* **459**, 796–801 (2009).
20. B. Bajoghli *et al.*, A thymus candidate in lampreys. *Nature* **470**, 90–94 (2011).
21. S. Kuraku, S. Kuratani, Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zool. Sci.* **23**, 1053–1064 (2006).
22. T. Zhu *et al.*, Chromosome-level genome assembly of *Lethenteron reissneri* provides insights into lamprey evolution. *Mol. Ecol. Resour.* **21**, 448–463 (2021).
23. J. Jurka, Repeats in genomic DNA: Mining and meaning. *Curr. Opin. Struct. Biol.* **8**, 333–337 (1998).
24. Y. Nakatani *et al.*, Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. *Nat. Commun.* **12**, 4489 (2021).
25. S. Das, M. Nozawa, J. Klein, M. Nei, Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. *Immunogenetics* **60**, 47–55 (2008).
26. S. Das, U. Mohamedy, M. Hirano, M. Nei, N. Nikolaidis, Analysis of the immunoglobulin light chain genes in zebra finch: Evolutionary implications. *Mol. Biol. Evol.* **27**, 113–120 (2010).
27. P. Kraj *et al.*, The human heavy chain Ig V region gene repertoire is biased at all stages of B cell ontogeny, including early pre-B cells. *J. Immunol.* **158**, 5824–5832 (1997).
28. B. A. Malynn, G. D. Yancopoulos, J. E. Barth, C. A. Bona, F. W. Alt, Biased expression of JH-proximal VH genes occurs in the newly generated repertoire of neonatal and adult mice. *J. Exp. Med.* **171**, 843–859 (1990).
29. D. S. Linthicum, W. H. Hildemann, Immunologic responses of Pacific hagfish. 3. Serum antibodies to cellular antigens. *J. Immunol.* **105**, 912–918 (1970).
30. J. J. Smith *et al.*, The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* **50**, 270–277 (2018).
31. J. J. Smith *et al.*, Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* **45**, 415–421, 421e1-2 (2013).
32. J. Schultz, F. Milpetz, P. Bork, C. P. Ponting, SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857–5864 (1998).
33. O. Kohany, A. J. Gentles, L. Hankus, J. Jurka, Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and censor. *BMC Bioinformatics* **7**, 474 (2006).
34. R. C. Edgar, MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
35. N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
36. K. Tamura *et al.*, MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
37. L. Noé, G. Kucherov, YASS: Enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* **33**, W540–W543 (2005).
38. S. Kuraku, Insights into cyclostome phylogenomics: Pre-2R or post-2R. *Zool. Sci.* **25**, 960–968 (2008).
39. I. C. Potter, H. S. Gill, C. B. Renaud, D. Haoucher, "The taxonomy, phylogeny, and distribution of lampreys" in *Lampreys: Biology, Conservation and Control*, M. F. Docker, Ed. (Springer, Dordrecht, 2015), pp. 35–73.