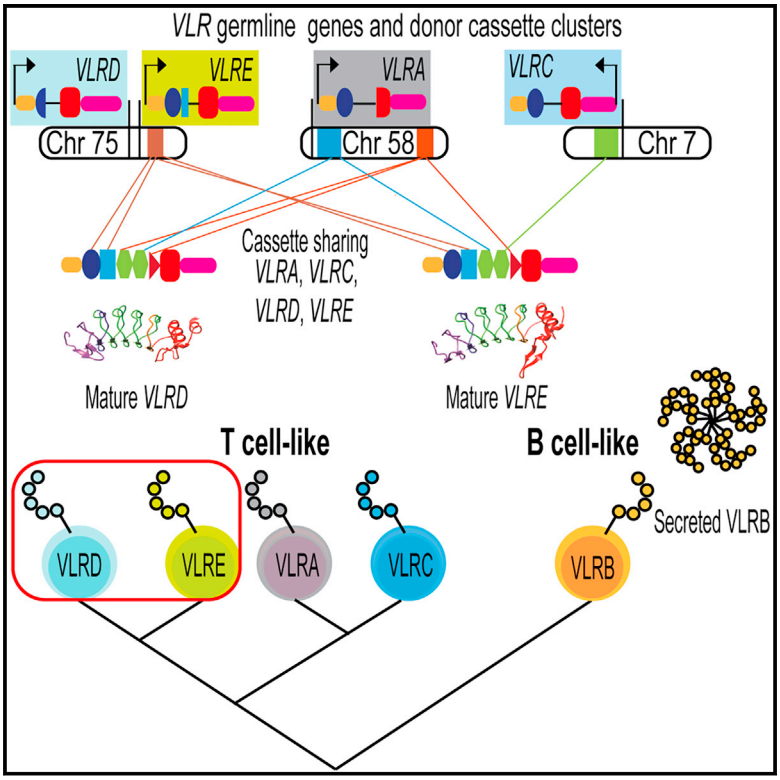
Cell Reports

Article

Evolution of two distinct variable lymphocyte receptors in lampreys: VLRD and VLRE

Graphical abstract



Authors

Sabyasachi Das, Thomas Boehm, Stephen J. Holland, ..., Ryan D. Heimroth, Masayuki Hirano, Max D. Cooper

Correspondence

sdas8@emory.edu (S.D.), boehm@ie-freiburg.mpg.de (T.B.), mdcoope@emory.edu (M.D.C.)

In brief

Das et al. report the discovery of two distinct variable lymphocyte receptors, VLRD and VLRE in lampreys, which possess an alternative adaptive immune system. The characterization of VLRD and VLRE provides valuable insights into the evolution of T-like lymphocytes in jawless vertebrates.

Highlights

- Two distinct VLRs, VLRD and VLRE, are found in lampreys
- VLRD and VLRE, phylogenetically close to VLRA and VLRC, expressed in T-like cells
- Donor cassettes shared inter-chromosomally among mature VLRA, VLRC, VLRD, and VLRE
- VLRD^{+} and VLRE^{+} cells may be part of the T cell arm of lamprey immunity

Das et al., 2023, Cell Reports *42*, 112933
August 29, 2023
https://doi.org/10.1016/j.celrep.2023.112933

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Evolution of two distinct variable lymphocyte receptors in lampreys: VLRD and VLRE

Sabyasachi Das,[1,2,*] Thomas Boehm,[3,4,*] Stephen J. Holland,[3] Jonathan P. Rast,[1,2] Francisco Fontenla-Iglesias,[1,2] Ryo Morimoto,[3] J. Gerardo Valadez,[1,2] Ryan D. Heimroth,[1,2] Masayuki Hirano,[1,2] and Max D. Cooper[1,2,5,*]

[1]Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA 30322, USA

[2]Emory Vaccine Center, Emory University, Atlanta, GA 30317, USA

[3]Department of Developmental Immunology, Max-Planck Institute of Immunobiology and Epigenetics, Stuebeweg 51, 79108 Freiburg, Germany

[4]Faculty of Medicine, University of Freiburg, Breisacher Str. 153, 79110 Freiburg, Germany

[5]Lead contact

*Correspondence: sdas8@emory.edu (S.D.), boehm@ie-freiburg.mpg.de (T.B.), mdcoope@emory.edu (M.D.C.)

https://doi.org/10.1016/j.celrep.2023.112933

SUMMARY

Jawless vertebrates possess an alternative adaptive immune system in which antigens are recognized by variable lymphocyte receptors (VLRs) generated by combinatorial assembly of leucine-rich repeat (LRR) cassettes. Three types of receptors, VLRA, VLRB, and VLRC, have been previously identified. VLRA- and VLRC-expressing cells are T cell-like, whereas VLRB-expressing cells are B cell-like. Here, we report two types of VLRs in lampreys, VLRD and VLRE, phylogenetically related to VLRA and VLRC. The germline *VLRD* and *VLRE* genes are flanked by 39 LRR cassettes used in the assembly of mature *VLRD* and *VLRE*, with cassettes from chromosomes containing the *VLRA* and *VLRC* genes also contributing to *VLRD* and *VLRE* assemblies. *VLRD* and *VLRE* transcription is highest in the triple-negative ($VLRA^-/VLRB^-/VLRC^-$) population of lymphocytes, albeit also detectable in $VLRA^+$ and $VLRC^+$ populations. Tissue distribution studies suggest that lamprey $VLRD^+$ and $VLRE^+$ lymphocytes comprise T-like sublineages of cells.

INTRODUCTION

Phylogenetic studies have revealed that two alternative forms of adaptive immune systems arose in vertebrates about 500 million years ago.[1,2] The extant jawed vertebrates generate their immunoglobulin domain-based B cell and T cell receptors for antigens through the recombination of different V-(D)-J gene segments.[3,4] Instead, the extant jawless vertebrates (lampreys and hagfishes) somatically assemble equally vast numbers of functional antigen receptors, called variable lymphocyte receptors (VLRs), through the addition of leucine-rich repeat (LRR)-encoding donor cassettes into incomplete germline genes.[5–9]

Three *VLR* genes (*VLRA*, *VLRB*, and *VLRC*) have previously been identified in lampreys and hagfishes.[5,10–13] The incomplete germline versions of these *VLR* genes have an intervening noncoding sequence interrupting their N-terminal and C-terminal coding sequences. These germline *VLR* genes are flanked by hundreds of genomic donor cassettes encoding different LRR motifs that are available as templates for the serial piece-wise and stepwise replacement of intervening sequences in the assembly of a mature *VLR* gene. The assembly occurs by a poorly understood gene conversion-like process.[7] The repertoire of anticipatory receptors generated via this combinatorial VLR assembly process in the jawless vertebrates is comparable in size to that of Ig domain-based antigen receptors of jawed vertebrates.[6,12]

Two cytidine deaminase (CDA) genes have been identified in the lamprey genome.[6,7,12,14] VLRB receptor assembly occurs in hematopoietic tissues and is dependent upon *CDA2* activity,[14] while VLRA and VLRC receptor assembly takes place in thymus-equivalent regions at the tips of gill folds and their proximal filaments and is associated with *CDA1* expression.[15] The assembled *VLR* genes are expressed in a monoallelic lineage-specific fashion.[7,16,17] VLRB is expressed by B-like cells that respond to antigen stimulation by proliferation and differentiation into plasma cells that secrete multivalent VLRB antibodies.[6,18,19] VLRA and VLRC are expressed by two T-like lineages that respectively resemble $TCR\alpha\beta^+$ and $TCR\gamma\delta^+$ cells in jawed vertebrates.[17,20]

Given that multiple lymphocyte sublineages have evolved in jawed vertebrates,[21,22] we sought evidence for similar complexity of the lymphocyte differentiation pathways in the jawless vertebrates. Here, through an extensive similarity search of genome sequences of six different lamprey species, we identify two previously unrecognized classes of lamprey *VLR* genes that we name *VLRD* and *VLRE*. Our characterization of these *VLR* genes indicates that they are expressed predominantly by lymphocytes that do not express VLRA, VLRB, or VLRC. In terms of sequence conservation, configuration of germline loci, donor LRR cassette sharing during assembly, and tissue distribution, we find that *VLRD* and *VLRE* are most closely related to *VLRA* and *VLRC*, thus defining two additional T cell-like sublineages in lampreys.

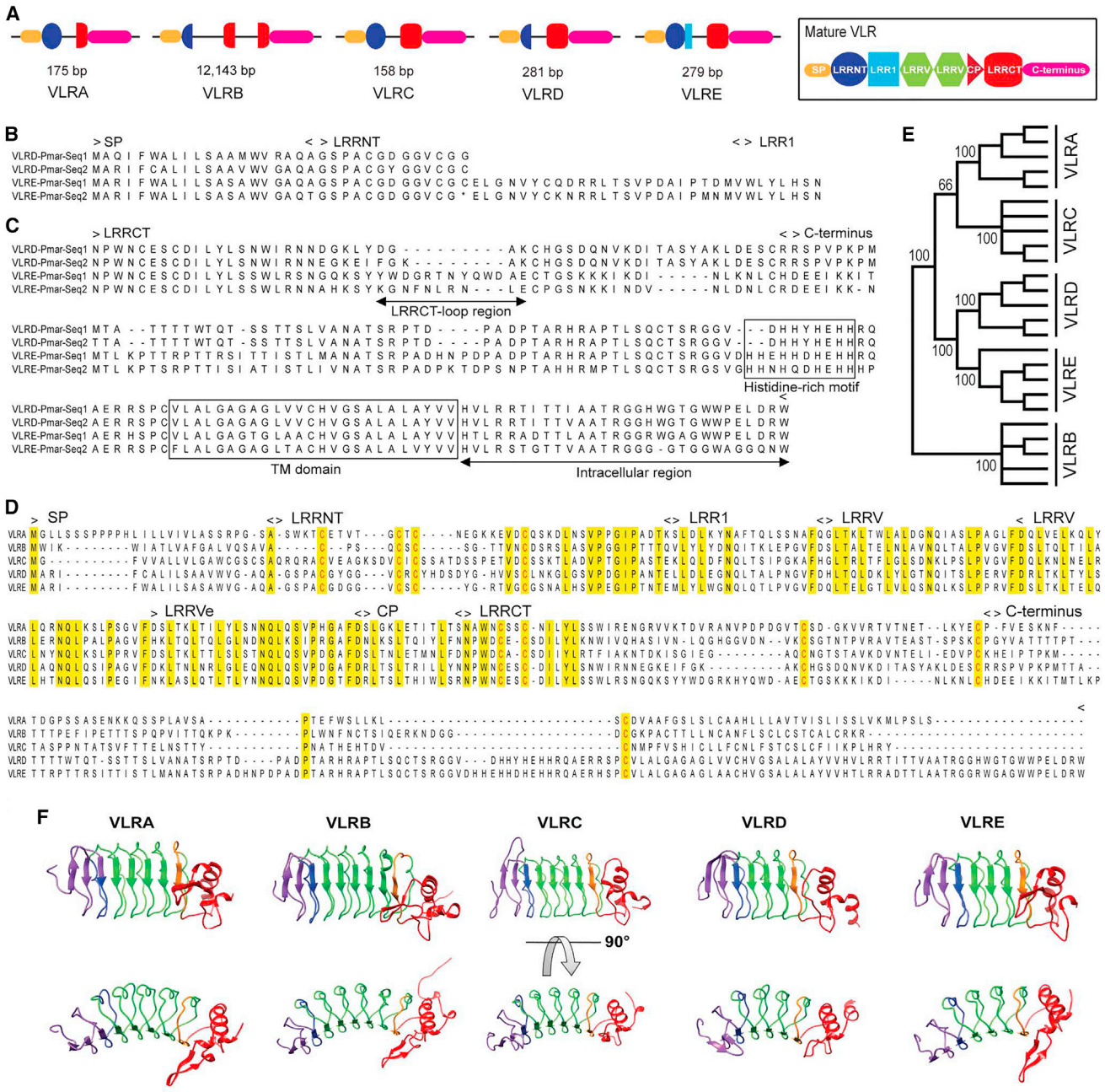


Figure 1. Newly identified VLRD and VLRE members of the VLR family

(A) Cartoon illustrating the configuration of incomplete germline genes of five VLR isotypes in sea lamprey. For comparison, a generic mature VLR configuration is illustrated in the box.

(B) Alignment of the N-terminal coding region of two copies of germline *VLRD* and two copies of germline *VLRE* genes of sea lamprey designated as Seq1 and Seq2, respectively.

(C) Concordance and divergence of C-terminal coding regions of germline *VLRD* and *VLRE* genes in sea lamprey. Both copies of *VLRD* and *VLRE* encode a complete LRRCT domain and the C terminus stalk region, which includes a histidine-rich motif, a transmembrane domain, and a short intracellular region. Both *VLRE*s (Seq1 and Seq2) encode relatively large LRRCT loops.

(D) Sequence comparison between representatives of mature VLRA, VLRB, VLRC, VLRD, and VLRE in sea lamprey. Deduced amino acid sequences are shown in the alignment with conserved residues highlighted in yellow. Conservation of cysteines is indicated by red color. Minor differences in the germline gene-encoded C-terminal regions are observed between lamprey individuals.

(E) Phylogeny of five VLR isotypes in lampreys. The phylogenetic tree is constructed using five representative sequences for each VLR isotype in sea lamprey. Bootstrap supports values are shown for interior branches.

(legend continued on next page)

RESULTS

Identification of *VLR* genes in sea lampreys

When currently available genome sequences for sea lamprey (*Petromyzon marinus*) were scanned using complete *VLRA*, *VLRB*, and *VLRC* sequences derived from both lampreys and hagfish as queries in TBLASTN searches, we identified a unique LRR C-terminal (LRRCT) module. Extension of the 5′ and 3′ genomic regions flanking this sequence revealed a previously unknown incomplete germline *VLR*-like gene with an N-terminal coding region followed by a non-coding intervening sequence and C-terminal coding region. A subsequent similarity search conducted by comparing this new germline *VLR*-like sequence with the available sea lamprey genome sequences from different animals led to the identification of a total of four distinct germline *VLR*-like genes that are illustrated in Figure 1 (see also Table S1). Based on an analysis of germline gene configurations and sequence compositions of these four *VLR*-like sequences, we concluded that they comprise two distinct types of *VLR*, designated *VLRD* and *VLRE*, each of which has two distinct but closely related subtypes. In the sea lamprey, there is 90% nucleotide sequence identity between the *VLRD-Pmar-Seq1* and *VLRD-Pmar-Seq2* germline sequences, while 85% nucleotide sequence identity is found between the *VLRE-Pmar-Seq1* and *VLRE-Pmar-Seq2* sequences. The N-terminal coding regions of the germline *VLRD* genes encode the signal peptide (SP) and the 5′ portion of the LRR N-terminal (LRRNT) module, whereas the *VLRE* germline genes encode the SP, an entire LRRNT module, and a 5′ LRR1 module (Figures 1A, 1B, and S1). The C-terminal coding regions for both germline *VLRD* and *VLRE* genes encode complete LRRCT modules, stalk regions, transmembrane domains, and short cytoplasmic tails (Figures 1A, 1C, S1, and S2).

Germline transcripts of *VLRD* and *VLRE* were readily detectable in white blood cells from sea lampreys, although mature sequences of *VLRD* and *VLRE* could not be recovered in our initial analyses. However, an extensive transcriptome analysis of lymphocytes in the gill region yielded several partially assembled sequences that contained one each of LRRCT and CP modules, preceded by two to four LRRV modules. We then designed two amplification strategies to specifically enrich cDNAs of assembled transcript sequences at the expense of the more common germline transcripts. In the first approach, we used forward primers complementary to the regions encoding the N-terminal SPs and a reverse primer binding to sequences encoding the connecting peptide (CP). In a second approach, we employed a collection of 10 primers covering the nucleotide sequences encoding presumptive LRRNT cassettes and a reverse primer located in the regions encoding the invariant LRRCT segments (see Table S2). In this way, 60 unique mature *VLRD* and *VLRE* sequences were recovered from sea lamprey blood leukocytes.

VLRD and VLRE sequence analysis and phylogenetic characterization

The predicted VLRD and VLRE proteins of sea lamprey exhibit an SP of 20 residues, an LRRNT module of $\geq$37 residues, an LRR1 module of 18 residues, followed by two to eight distinct LRRV modules (each 24 residues in length), a 12-residue-long CP, a $\geq$52-residue-long LRRCT module, and a $\geq$128-residue C terminus stalk region with a unique histidine-rich motif, followed by a transmembrane (TM) domain and short cytoplasmic tail (Figure 1D). Notably, the SP regions, LRRCT modules, and the C-terminal regions of VLRD and VLRE have only weak sequence similarity to the corresponding sequences of VLRA, VLRB, and VLRC.

To examine the phylogenetic relationship of lamprey VLRs, we constructed an unrooted neighbor-joining tree using the conceptually translated sequences of five representative sequences of each *VLR* gene. The analysis focused on those regions of the molecules that could be reliably aligned: LRRNT, LRR1, terminal LRRV, CP, and LRRCT in addition to the invariant SP and stalk regions. The VLRD and VLRE sequences are clustered with VLRA and VLRC sequences in the tree, whereas VLRB sequences appear as an outgroup (Figure 1E). The VLRD sequences are separated from VLRE sequences in the phylogenetic tree, although the amino acid compositions of the histidine-rich motifs, TM domains, and the cytoplasmic tails of the C terminus stalk regions are very similar for VLRD and VLRE. Notably, the glycine- and alanine-rich TM domains of VLRD and VLRE are distinct from that of VLRA-TM and VLRC-TM domains (Figure S2).

The LRRNT and LRRCT modules of all of the previously identified VLRs in jawless vertebrates (VLRA, VLRB, and VLRC) contain four cysteine residues that form two sets of disulfide bridges.[19] The cysteine configuration of LRRNT modules of VLRA, VLRB, and VLRC isotypes corresponds to C1-X$_m$-C2-X-C3-X$_n$-C4 (where X stands for any amino acid other than cysteine; m and n stand for variable numbers of amino acids). By contrast, the spacing of cysteines in LRRCT modules varies; whereas VLRB and VLRC exhibit a C1-X-C2-X$_m$-C3-X$_n$-C4 signature, VLRA is notable for its C1-X-X-C2-X$_m$-C3-X$_n$-C4 signature. The two residues separating C1 and C2 of LRRCT of VLRD and VLRE (note the conserved glutamic acid and serine residues) resemble the corresponding cysteine configuration of VLRA; the close sequence relationship among VLRA, VLRD, and VLRE is supported by the shared configuration of the first three cysteines (C1-X$_5$-C2-X-C3) in their LRRNT regions (Figure 1D).

Modeling the three-dimensional structures of VLRD and VLRE indicates that both adopt a solenoid structure like other VLRs (Figure 1F). With respect to the highly variable insert that distinguishes VLRA and VLRB from VLRC,[13,23] we note that only VLRE has the potential to form a protruding loop of its LRRCT region (Figure 1F). Moreover, the LRRCT loop regions of VLRE sequence 1 and VLRE sequence 2 in sea lampreys vary significantly in lengths (12 and 9 residues for sequence 1 and sequence

(F) Comparison of the predicted 3D structures of VLRA (GenBank: ABO27114), VLRB (GenBank: QII89098), VLRC (GenBank: KC244052), VLRD (GenBank: OQ595160), and VLRE (GenBank: OQ595165). The LRRNT and LRR1 regions are shown in blue, and LRRVs are shown in green, whereas the CP and LRRCT regions are represented in red. The model is based on sequences that are truncated at the junction of the LRRCT and stalk regions. Note that VLRE possesses a protruding loop in the LRRCT region like that of VLRA and VLRB, whereas the homologous region is too short for VLRC and VLRD to form a protruding loop.

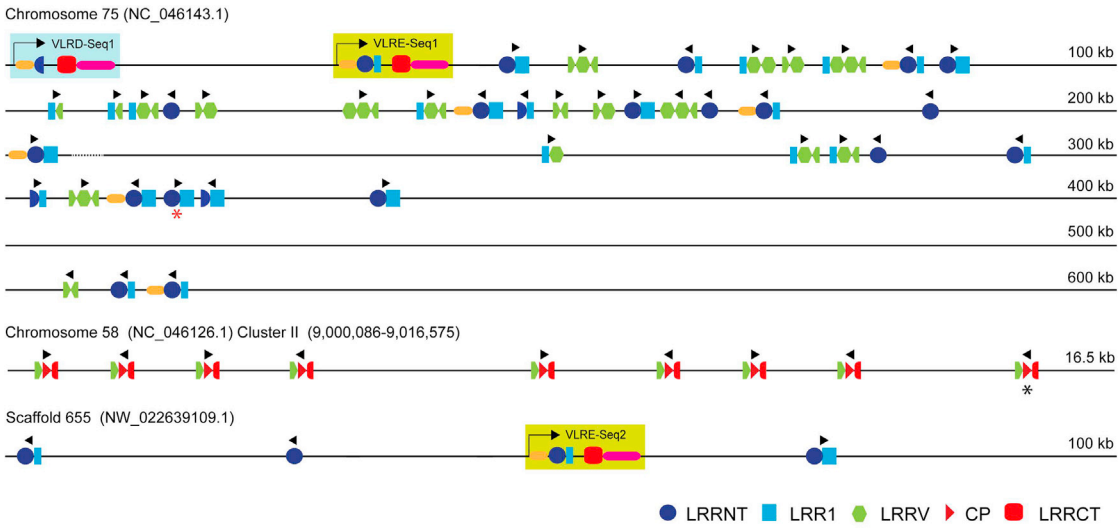


Figure 2. *VLRD/VLRE* locus organization in sea lamprey

A simplified map of the *VLRD/VLRE* locus is based on the current version of sea lamprey genome sequence (kPetMar1) in which one copy of germline *VLRD* and two copies of germline *VLRE* are found. The germline *VLRD-Seq1* (blue shading) and *VLRE-Seq1* (green shading) and associated 39 cassettes are located on chromosome 75. Cassettes are shown in proportion to genomic spacing, but icons are not to scale. The region shown is 600 kb downstream of the *VLRD-Seq1* start codon. The dotted line indicates an unresolved region of 36,030 nucleotides. A map of the nine CP-containing cassettes (*3′LRRV-CP-5′LRRCT*) on chromosome 58 used exclusively by *VLRD/VLRE*, except for one cassette (indicated by a black asterisk), which is used in assembled *VLRC*. The second germline *VLRE* gene (*VLRE-Seq2*) is located on scaffold 655, which also contains three donor cassettes. The arrowhead above the donor cassettes indicates the transcription orientation, whereas the red asterisk below indicates the presence of an internal stop codon in the genomic donor cassette.

2, respectively) and in amino acid compositions (Figure 1C). In contrast to VLRE, there are only three residues (DGA) and four residues (FGKA) in the homologous LRRCT region of VLRD sequence 1 and VLRD sequence 2 in sea lampreys, respectively (Figure 1C).

Genomic organization of the *VLRD* and *VLRE* loci

Using an iterative similarity search strategy, we mapped the genomic donor cassettes along with the incomplete germline *VLRD* and *VLRE* genes in the newly available chromosome-scale genomic sequences of sea lamprey (kPetMar1).[24,25] The germline *VLRD*-sequence 1 and *VLRE* sequence 1 are located only ~35.2 kb apart on chromosome 75, whereas the germline *VLRE* sequence 2 is located elsewhere on scaffold 655 (Figure 2; Table S1). The *VLRD* sequence 2 could not be identified in the current version of the genome assembly but is present in contig22334 of an earlier version of the sea lamprey genome (Petromyzon_marinus-7.0) (Table S1). The *VLRD* and *VLRE* germline genes are immediately upstream of 39 donor cassettes (21 are LRRNT-encoding, most of which include either a complete or a partial LRR1-encoding region) that are spread over a 520-kb region of chromosome 75 (GenBank: NC_046143.1) (Figure 2; Table S3). These cassettes appear to be used exclusively in the assembly of mature *VLRD* and *VLRE* genes, since they were not found in the available collections of mature sequences of the three previously identified *VLR*s. Notably, the nine CP region cassettes (*3′LRRV-CP-5′LRRCT* cassettes encoding a mere five amino acid residues of the LRRCT portion), which are potentially dedicated to *VLRD* and *VLRE* assemblies, are located in a small genomic region of less than 20 kb (cluster II) on chromosome 58 (GenBank: NC_046126.1); this chromosome also harbors the *VLRA* germline gene in a second donor cassette cluster at the opposite end (cluster I, see Figure 3). Only one germline-encoded LRRCT region is present for each of *VLRD* and *VLRE* (Figure 2). Thus, in contrast to the situation with VLRA and VLRC,[11,12] the overall diversity of the C-terminal CP-LRRCT segment is very limited in VLRD and VLRE sequences. Interestingly, one of the nine *3′LRRV-CP-5′LRRCT* cassettes was found in a cDNA from an assembled *VLRC* gene (Figure 2). In contrast to the situation for *VLRD* genes, both germline *VLRE* genes encode a complete LRRNT module (Figures 1A and 2), which may be modified by insertion of sequences of LRRNT module-encoding donor cassettes. The LRRNT regions of the VLRD and VLRE proteins therefore may vary in size depending on which *LRRNT* cassettes serve as donors.

In addition to 18 LRRV-encoding cassettes found on chromosome 75, other *LRRV* cassettes used in *VLRD* and *VLRE* assemblies are located in two clusters (cluster I and cluster II of the *VLRA* locus) on chromosome 58 and in a cluster on chromosome 7 (GenBank: NC_046075.1), the latter harboring the *VLRC* germline gene; cassettes from both of these chromosomes are used for *VLRA* and *VLRC* assemblies as well (Figure 3). Two currently unplaced scaffolds also contain donor cassettes that are used in *VLRD* and *VLRE* assembly: scaffold 785 (GenBank: NW022639236.1) and scaffold 655 (GenBank: NW022639109.1), the latter of which encodes a second *VLRE* germline gene.

We found that mature *VLRD* and *VLRE* share identical LRR-encoding modules (Figures 3 and 4), much like donor cassette sharing between *VLRA* and *VLRC* assemblies.[26] Shared use of donor cassettes is most pronounced for the LRRV modules,

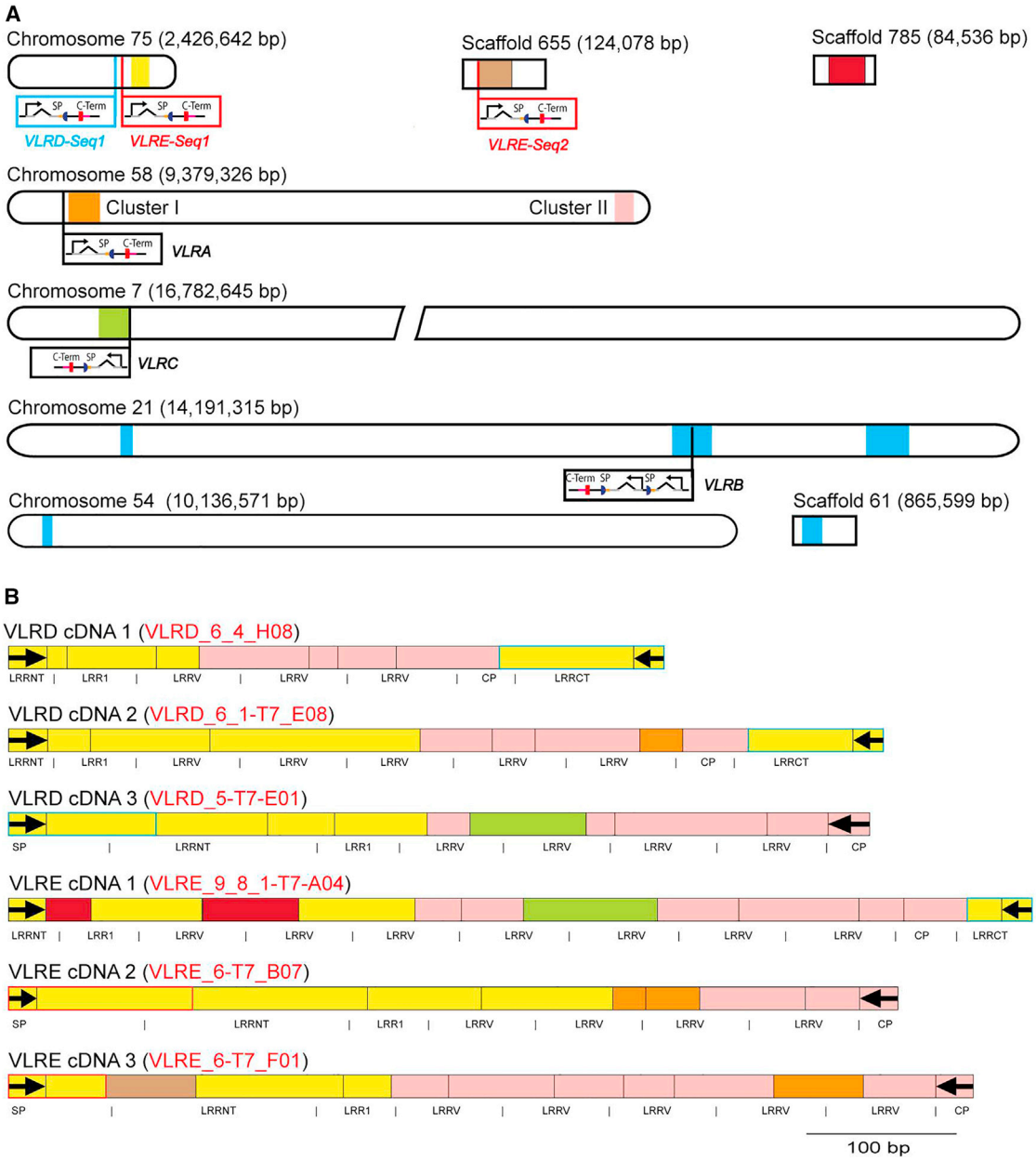


Figure 3. Contribution of genomic donor cassettes to mature *VLRD* and *VLRE* assemblies in sea lamprey

(A) Cartoon illustrating genomic donor cassette usage in mature *VLRD* and *VLRE* assemblies; cassettes are contributed from six clusters located in disparate genomic regions (genome assembly kPetMar1.pri). Germline *VLRD/VLRE* genes on chromosome 75 are indicated by boxes next to their genomic positions (VLRD, blue; VLRE, red) flanking a cluster of 39 donor cassettes (yellow box). Cassettes from two genomic clusters on chromosome 58 (cluster I [orange box] and II [pink box]) also contribute to the *VLRD/VLRE* assemblies. Note that the *VLRA* germline gene is located in cluster I. Several additional LRRV-encoding cassettes used in *VLRD/VLRE* assemblies are located in a cluster on chromosome 7 (green box), which contains the germline *VLRC* gene. Two unplaced scaffolds (top right) also contain donor cassettes used in *VLRD/VLRE* assembly: scaffold 785 (GenBank: NW022639236.1, dark red box) and 655 (GenBank: NW022639109.1, brown box), which encodes a second *VLRE* germline gene. The donor cassettes for *VLRB* are encoded in five clusters on chromosomes 21, chromosome 54, and on unplaced scaffold 61 (blue rectangles). These dedicated *VLRB* donor cassettes do not contribute to *VLRD* or *VLRE* assemblies but are shown for comparison. (B) Donor cassette contributions from the six genomic clusters mapped to six representative *VLRD/VLRE* cDNA amplicon sequences. Colors indicating the putative origin of assembled cDNA sequence correspond to the chromosomal clusters as in (A). Germline *VLRD/VLRE* contributions are shown with blue (*VLRD*) or red (*VLRE*) outlines. Primer sequences used to amplify *VLRD/VLRE* cDNAs are indicated with black arrows. In each case, the N-terminal donor cassettes are contributed from the cluster on chromosome 75 flanking the *VLRD/VLRE* genes (yellow boxes), with rarer contributions from LRRNT-encoding cassettes on the two unplaced scaffolds (scaffold 785 [dark red boxes] and scaffold 655 [brown box]). More C-terminal cassettes up to and including the CP donor cassette are

(legend continued on next page)

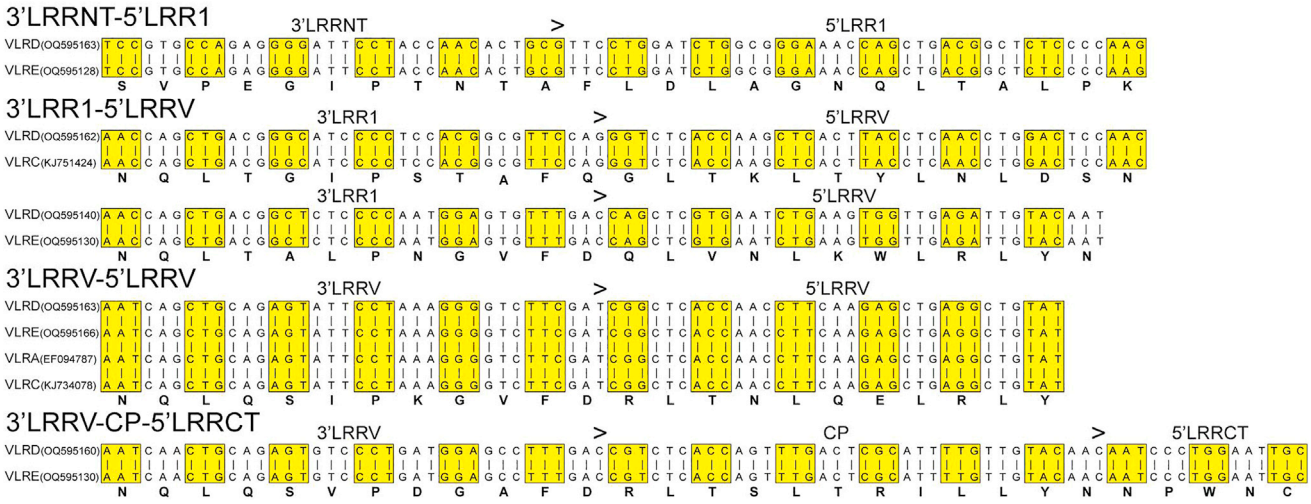


Figure 4. Genomic donor cassettes shared for *VLRD* and *VLRE* assemblies

Cassettes encoding the 3′LRRNT-5′LRR1, 3′LRR1-5′LRRV, and 3′LRRV-CP-5′LRRCT regions are shared between *VLRD* and *VLRE* assemblies, while cassettes encoding 3′LRRV-5′LRRV region are frequently shared between *VLRA*, *VLRC*, *VLRD*, and *VLRE*. One example of each cassette sharing category is shown. Alternative codons are highlighted. The GenBank accession numbers are given in parenthesis.

although the cassettes corresponding to the 3′LRRNT-5′LRR1, 3′LRR1-5′LRRV, 3′LRRV-5′LRRV, and 3′LRRV-CP-5′LRRCT regions are also shared between mature *VLRD* and *VLRE*. This phenomenon supports the close evolutionary and functional relationship of *VLRD* and *VLRE* genes. Interestingly, whereas the LRRV module-encoding cassettes on chromosome 75 are dedicated to *VLRD/VLRE* assemblies, those on chromosomes 7 and 58 are sometimes shared between *VLRA*, *VLRC*, *VLRD*, and *VLRE* assemblies (Figures 3, S3, and S4); by contrast, cassette sharing between *VLRB* and these two reported *VLR* assemblies (or with *VLRA* and *VLRC*) was never observed. These observations indicate that VLRB represents a functionally distinct branch of lamprey antigen receptors.

As no genomic donor *5′LRRCT-LRRCTm* cassette (encoding the 5′ region and the middle of the LRRCT domain) was found for *VLRD* and *VLRE* (except for five amino acids including the first cysteine residue at 5′LRRCT region, which can also be a part of *3′L-C-5′LRRCT* cassette), a major portion of the LRRCT module is never shared between VLRD and VLRE. Our present results are concordant with previous findings indicating that the LRRCT regions are unique for each of the different VLR isotypes.[5,11,12,26,27]

VLRD and *VLRE* in different lamprey species

Sequences homologous to the *P. marinus* germline *VLRD* and *VLRE* genes were also found in five additional lamprey species: European brook lamprey, Japanese lamprey, Far Eastern brook lamprey, Western brook lamprey, and Pacific lamprey (Table S1). The identification of two *VLRD* and two *VLRE* germline copies with closely related sequences (designated sequences 1 and 2) in five of the six lamprey species examined suggests that *VLRD* and *VLRE* are multicopy genes in lampreys. In the current version of the Western brook lamprey genome assembly, however, we could find only one copy each of germline *VLRD* and *VLRE* genes (in scaffolds 2,692 and 95, respectively). For Pacific lampreys,[28] we identified one copy each of the *VLRD* and *VLRE* in the reference male genome (ETRm_v1) and one copy of *VLRD* and two copies of *VLRE* in the reference female genome (ETRf_v1). This observation suggests that the diversification of VLRD and VLRE genes is associated with speciation events that have occurred relatively recently.[29]

Our comparison of *VLRD* and *VLRE* sequences from all six lamprey species revealed a clear separation into clusters for *VLRD* and *VLRE* sequences, respectively (Figure 5). However, in some instances, the orthologous relationships between sequence 1 and sequence 2 for both *VLRD* and *VLRE* could not be resolved, possibly due to either independent duplication or due to partial homogenization of *VLRD* and/or *VLRE* genes in certain lamprey lineages. As observed for sea lamprey, the N-terminal coding regions of the germline *VLRD* genes of other lamprey species encode an SP and a 5′LRRNT module, whereas the N-terminal coding region of the *VLRE* germline genes encode the SP, the entire LRRNT module, and the 5′LRR1 module (Figure S5A). The cysteine configurations in the LRRNT and LRRCT regions for both VLRD and VLRE are conserved in all lamprey species (Table S4). As expected, the differences in LRRCT modules (including the differential lengths in the LRRCT loop regions) and the C-terminal regions between VLRD and VLRE are

invariably contributed from chromosome 58 (pink and orange boxes), with some contributions from donor cassettes on chromosome 7 (green boxes). Each box section indicated along the cDNA sequence separated with vertical black lines represents a contiguous sequence potentially derived from a single cassette contribution. The transcript regions encoding the signal peptide and LRR motifs are indicated below each transcript. The precise genomic locations of donor cassette sequence matches used to construct these diagrams are shown in Figures S3 and S4.

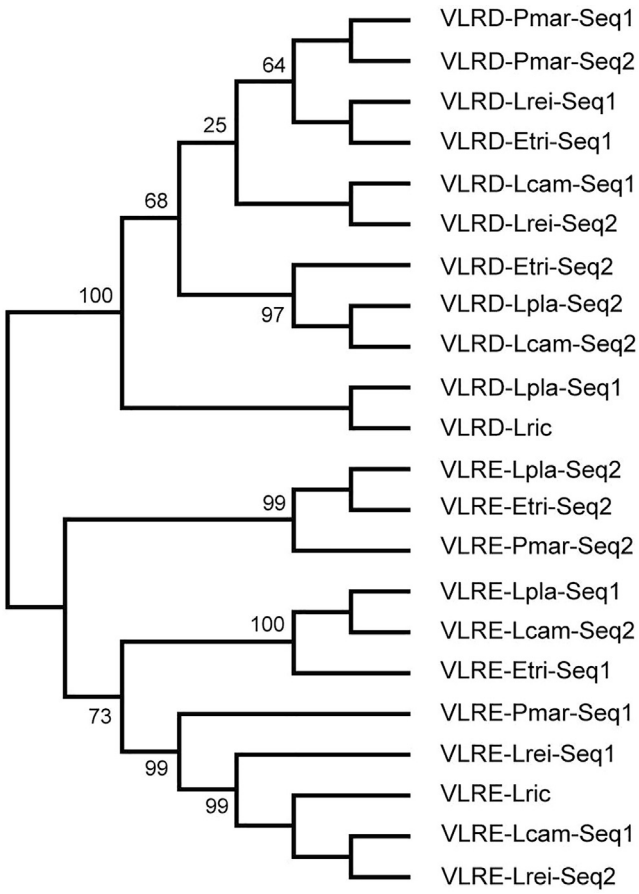


Figure 5. Phylogenetic comparison of VLRD and VLRE of six lamprey species

Amino acid sequences of the N-terminal coding region and C-terminal coding regions from germline genes are used to construct the phylogenetic tree because mature VLRD and VLRE sequences are currently unavailable for all lamprey species. Bootstrap values are shown for interior branches.

preserved across lamprey species (Figure S5B). It is notable that, like in the sea lamprey, genomic donor cassette sharing is also evident in our analysis of the assembled *VLRA*, *VLRC*, and *VLRD* sequences in the European brook lamprey (Figure S6).

Cellular and tissue expression patterns of *VLRD* and *VLRE*

To examine the cellular expression patterns of *VLRD* and *VLRE*, we isolated sea lamprey lymphocyte populations using mouse monoclonal antibodies against VLRA, VLRB, or VLRC, the specificities of which were confirmed by testing against a panel of VLRA-, VLRB-, and VLRC-expressing transfectants.[6,17,18,20] Real-time RT-PCR analysis indicated that the highest expression levels (these include both germline and assembled transcripts) of *VLRD* and *VLRE* are found in the triple-negative ($VLRA^-$/$VLRB^-$/$VLRC^-$) population of lymphocytes; *VLRD* and *VLRE* expression levels for $VLRA^+$ lymphocytes are higher than those for the $VLRC^+$ population of cells, whereas $VLRB^+$ lymphocytes were consistently negative (Figure 6). When the germline and assembled sequences were analyzed separately using specific primers, the expression levels of germline transcripts for *VLRD* and *VLRE* are noticeably higher compared to the assembled transcripts.

We also examined the expression of *VLRD* and *VLRE* in different tissues of sea lampreys and European brook lampreys. These two *VLR* genes were found to have similar expression patterns in both of these lamprey species, with relatively high transcript levels being noted in the gills and intestine-typhlosole region compared with relatively low levels of expression in skin, blood, and kidneys (Figures 6 and S7).

RNA *in situ* hybridization analysis of sea lamprey tissues

To identify cells expressing mRNA transcripts of the identified *VLR*s in immune-related tissues, frozen sections of the gills (including the thymus-equivalent regions in the tips of the gill folds), the epipharyngeal ridge, kidney, typhlosole, intestine, and skin (Figure 7) were examined by hybridization chain reaction (HCR) *in situ* imaging.[30,31] Due to the high nucleotide sequence similarity between invariant regions (including 3′ UTR) of the identified *VLR* genes, we were able to design a set of 20 specific probes for one *VLRE* gene and nine probes for one *VLRD* gene of sea lampreys. We also designed a set of 20 probes specific for *VLRA* transcripts and specific probes for *VLRB* transcripts as relevant positive controls. Use of the *VLRD* gene probes failed to reveal positive cells, perhaps because of the very low basal expression level of *VLRD* (see Figure 6). However, $VLRE^+$ cells were identified in all of the tissues tested; they exhibited a punctate staining pattern similar to that seen for $VLRA^+$ T-like cells and notably different from the highly abundant *VLRB* transcripts for some of the $VLRB^+$ cells (Figures 7F–7L). $VLRE^+$ cells were abundant in the epipharyngeal ridge, gill, and intestine but not in the kidney and typhlosole (Figures 7F–7H). In the epipharyngeal ridge, $VLRE^+$ cells were located either near the basement membrane of the epithelium or close to the apical surface (Figure 7G). Similarly, in the intestine, $VLRE^+$ cells were located in close proximity to the basement membrane of the epithelial cells (Figure 7F). Notably, $VLRE^+$ cells were scattered within the thymus-equivalent region of the gill fold tips and the adjacent gill filaments (Figure 7H). The *VLRA*- and *VLRE*-expressing lymphocytes could be assigned to three different categories based on their distinct expression patterns: (1) $VLRA^+/VLRE^-$, (2) $VLRA^+/VLRE^+$, and (3) $VLRA^-/VLRE^+$ (Figures 7J and 7K). In the gill filaments, thymoid region, epipharyngeal ridge, and intestine, we found a mixture of these three lymphocyte populations. However, in the typhlosole and kidneys, the vast majority of $VLRA^+$ cells in the typhlosole and kidneys were $VLRE^-$. *VLRB*-expressing cells were especially abundant in the typhlosole (Figure 7L) as expected from the results of previous studies of $VLRB^+$ cell distribution.[6,15,18]

DISCUSSION

The discovery of additional *VLRD* and *VLRE* genes, in addition to the previously defined *VLRA*, *VLRB*, and *VLRC* genes, supports the notion that the VLR system in lampreys is evolutionarily dynamic. The two identified additions to the VLR gene family

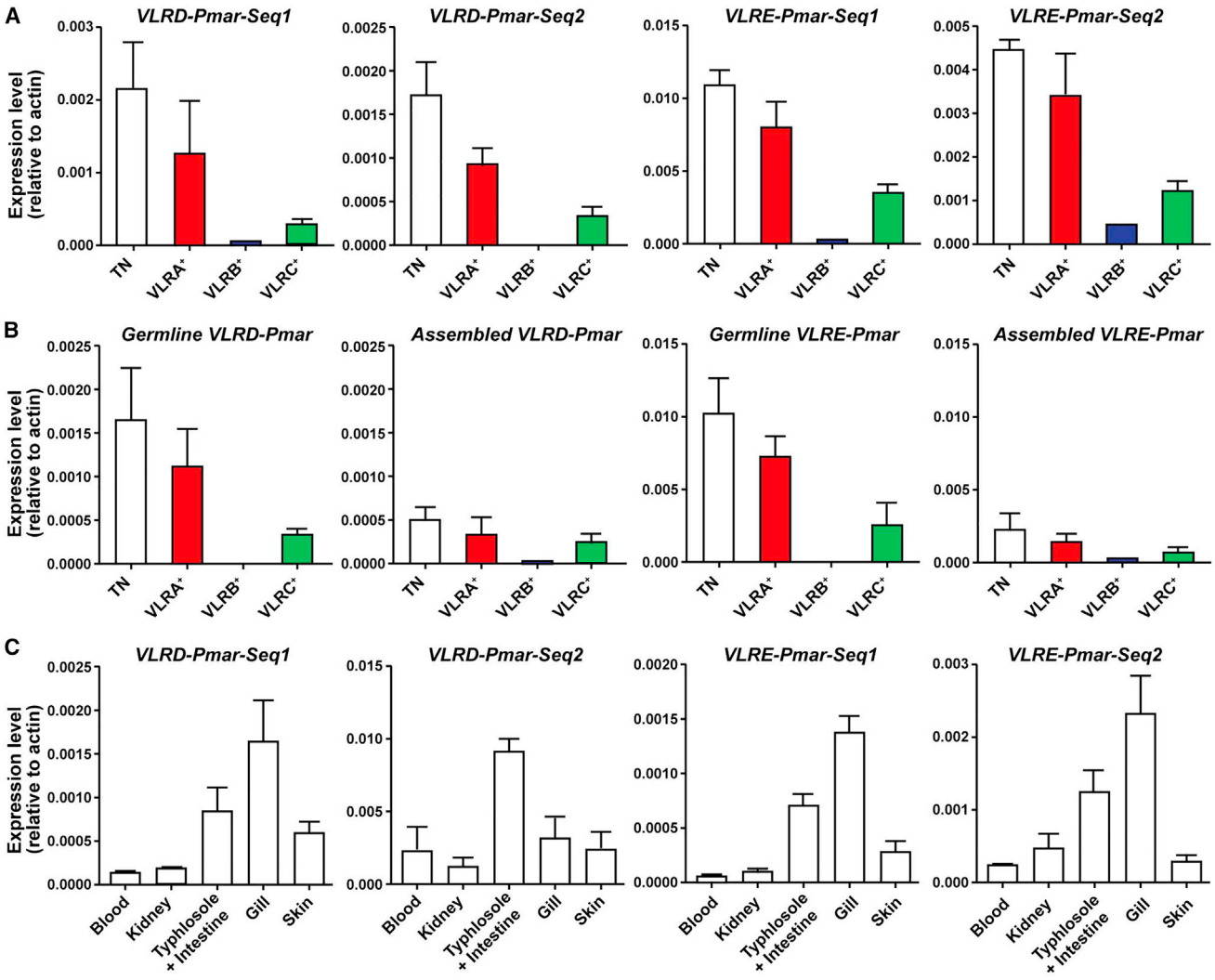


Figure 6. Cellular and tissue distribution of *VLRD* and *VLRE* in sea lampreys

(A) Expression of two duplicate copies (sequence 1 and sequence 2) of *VLRD* and *VLRE* genes for different lymphocyte populations. TN represents triple-negative ($VLRA^-/VLRB^-/VLRC^-$) lymphocyte population.

(B) Cellular distribution of germline and assembled *VLRD* and *VLRE* genes in sea lampreys. TN represents triple-negative ($VLRA^-/VLRB^-/VLRC^-$) lymphocyte population. Bars indicate standard error of mean for at least three lamprey larvae in each experiment.

(C) Tissue expression profiles for *VLRD* and *VLRE*. Transcripts are analyzed by real-time RT-PCR with *beta-actin* as control for both cellular and tissue distribution analyses. Bars indicate standard error of mean for three lamprey larvae in each experiment.

described here are distinguished from the other known VLRs by several unique features. Both VLRD and VLRE receptors share a histidine-rich motif in the C-terminal stalk region that is not found in any other VLR. The conservation of histidine-rich motif near the transmembrane domain of VLRD and VLRE in different lamprey species suggests that this motif could have specialized structural or functional roles. Moreover, the sequence compositions of the transmembrane and cytoplasmic tail regions also distinguish *VLRD/E* from *VLRA/C*. Modeling studies predict that the LRRCT region of VLRE forms a protruding loop similar to those seen in VLRA and VLRB, whereas the LRRCT portions of both VLRC and VLRD lack this loop (Figure 1). Since the highly variable loop of VLRB receptors is often involved in antigen binding,[32,33] it seems likely that VLRD and VLRE engage antigen in different ways.

An interesting dichotomy of the five currently known VLRs is noteworthy with respect to the diversity in the C-terminal LRR region. For VLRA, VLRB, and VLRC, the first five amino acid residues of LRRCT domain are encoded by the *3′LRRV-CP-5′LRRCT* cassettes.[26,27,34] In the case of VLRA and VLRB, the many *5′LRRCT-LRRCTm* cassettes contribute to substantial diversity in the LRRCT domain; this feature is less prominent in VLRC assemblies, since the sea lamprey and Japanese lamprey genomes harbor only two *5′LRRCT-LRRCTm* cassettes for VLRC sequences.[11,34] The lack of *5′LRRCT-LRRCTm* cassettes for *VLRD* and *VLRE* in any of the lamprey genome sequences

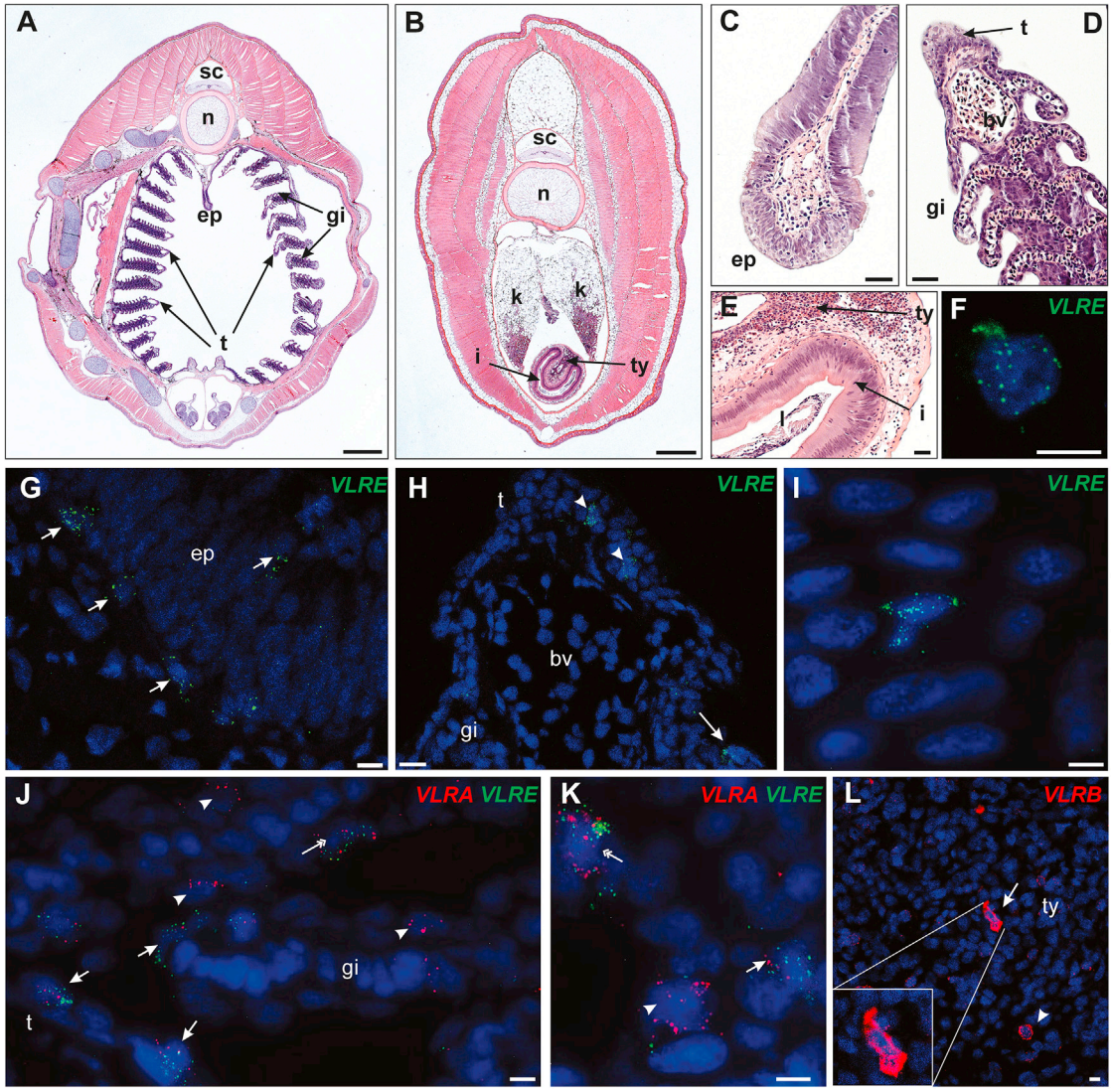


Figure 7. Photomicrographs of naive larvae cross-sections stained with hematoxylin and eosin or counterstained with DAPI (blue) and stained with HRC specific probes for *VLRE*, *VLRA*, and *VLRB* transcripts

(A) Immune-related organs in an anterior animal section showing the epipharyngeal ridge (ep), gills (gi), and thymoid area (t).

(B) Immune-related organs at the level of the anterior gut showing the kidneys (k), typhlosole (ty), and intestine (i).

(C–E) Magnification of the mucosal-related tissues: epipharyngeal ridge, gill, and intestine, respectively.

(F) Detail of a *VLRE*$^{+}$ cell in the intestine, showing a characteristic dotted pattern.

(G) *VLRE*$^{+}$ cells (arrow) associated to the epithelial cells of the epipharyngeal ridge.

(H) *VLRE*$^{+}$ cells are located both in the thymoid (t) area (arrowhead) and in the gill filaments (arrow).

(I) *VLRE*$^{+}$ cell located within the epithelial cells of the skin.

(J and K) Double staining with *VLRA* (red) and *VLRE* (green) HCR probes in the gill. The arrows indicate cells expressing mainly VLRE transcripts, while arrowheads indicate cells expressing mainly *VLRA* transcripts. Double arrows indicate cells expressing both *VLRA* and *VLRE* transcripts.

(L) Positive control for HCR experiment showing *VLRB*$^{+}$ cells dispersed in the typhlosole parenchyma, showing a *VLRB*^{hi} cell (arrow) and *VLRB*^{low} cell (arrowhead). The square in the bottom shows a magnification of one *VLRB*^{hi} cell. The "bv," "n," "sc," and "l" stand for blood vessel, notochord, spinal cord, and lumen, respectively. Scale bars for (A) and (B), 1 mm; (C), (D), and (E), 50 μm; (G) and (H), 10 μm; and for (F), (I), (J), (K), and (L), 5 μm.

analyzed here indicates that except for the first five amino acid residues (which may be contributed by *3′LR-C-5′LRRCT* cassettes), the LRRCT domains of VLRD and VLRE are encoded by their respective germline genes. Hence, with respect to the paucity of LRRCT diversity in mature sequences, VLRD and VLRE group together with VLRC. Collectively, the presence or absence of certain structural features suggests the hybrid nature of VLRD and VLRE when compared with VLRA and VLRC. The observation that, despite common features, clear sequence differences exist between the two copies of VLRD and VLRE in their LRRCT domains suggests that this further diversification is functionally important.

The availability of a high-quality updated version of sea lamprey genome sequence (kPetMar1)[24,25] has allowed us to define the genomic structure and the repertoire development of the *VLRD* and *VLRE* genes. The germline *VLRD* and *VLRE* genes are located in close genomic proximity, flanked by 39 LRRNT-, LRR1-, and LRRV-encoding donor cassettes (chromosome 75 in sea lamprey), which are exclusively shared among mature *VLRD* and *VLRE* assemblies. The genomic constellation strongly suggests that a local gene duplication event gave rise to *VLRD* and *VLRE* genes. Interestingly, genomic donor cassette usage among *VLR* genes is not restricted to cassettes in *cis* configuration to the germline gene; indeed, cassette sharing was previously observed for *VLRA* and *VLRC*,[26] which reside on different chromosomes in sea lamprey genome assemblies.[27] The strongest support for cassette sharing in *trans* comes from our present observation that a subset of LRRV-encoding genomic donor cassettes located on chromosomes that contain the germline *VLRA* (cluster I of chromosome 58) and *VLRC* (chromosome 7) genes are used for *VLRA*, *VLRC, VLRD*, and *VLRE* assemblies (see Figures 3 and 4).

Another interesting trend for the incorporation of genomic donor cassettes into mature *VLRD*/*VLRE* assemblies is worthy of note. All of the *3-LRRV-CP-5-LRRCT* cassettes that contribute to mature *VLRD*/*VLRE* assemblies are located in cluster II on chromosome 58, whereas LRRNT-, LRR1-, and LRRV1-encoding sequences originate from the *VLRD/E* flanking cassettes on chromosome 75. The sequences encoding the CP region, LRRVe, and most LRRV modules are invariably contributed by two clusters on chromosome 58, with occasional contributions of LRRV sequences from chromosome 7. These features imply the presence of chromosome-scale positioning mechanisms in the template-mediated assembly process of VLRD/E. This conclusion is reinforced by the finding that, as has been observed for mature *VLRA* and *VLRC*,[26] no VLRB-encoding cassettes are incorporated into the assembled *VLRD* and *VLRE* sequences. In humans, rare *trans*-locus rearrangements have also been observed for TCR genes,[35] whereas TCR/Ig chimeric genes are only known from leukemias.[36]

Our previous studies demonstrated that *VLRB* is expressed by B-like lymphocytes, whereas *VLRA* and *VLRC* genes are expressed by two different types of T-like cells, respectively akin to the $\alpha\beta$ and $\gamma\delta$ lineages of jawed vertebrates.[17,20] Distinct cytidine deaminases appear to be responsible for the assembly of *VLR* genes in lampreys; CDA2 has been shown to be required for VLRB assembly but not for VLRA and VLRC assembly,[14] and it has been assumed, but not yet proven, that CDA1, the second cytidine deaminase, is responsible for the assembly of VLRA and VLRC. However, VLRA$^+$ cells and VLRC$^+$ express *CDA1* preferentially, whereas VLRB$^+$ cells express *CDA2*.[17,20] The *CDA1* and *VLRA* mRNA were detected at gill thymoid regions in lampreys.[15] Here, we found high expression of *VLRD* and *VLRE* (in measurements that include both germline and assembled sequence transcripts) in the gill region (Figure 6), and HCR *in situ* data indicate that *VLRE*$^+$ cells are located both in the thymoid area and in the gill filaments (Figure 7). Expression levels of *VLRD* and *VLRE* are particularly high in the triple-negative (VLRA$^-$/VLRB$^-$/VLRC$^-$) population of lamprey blood lymphocytes. Although *VLRD*/*VLRE* expression is also detectable in the VLRA$^+$ and VLRC$^+$ populations of lymphocytes, it is absent in VLRB$^+$ lymphocytes. This suggests the possibility of a shared transcriptional regulation among the *VLRA*, *VLRC*, *VLRD*, and *VLRE* loci. Whether two copies of *VLRD* and *VLRE* sequences are co-expressed or define distinct populations of lymphocytes will be interesting to examine at the single-cell level. In view of the scarcity of retrievable mature sequences and overall low levels of gene expression, however, the VLRD- and VLRE-expressing cells appear to represent minor populations of specialized T-like lymphocytes. Notably, the *VLRE*$^+$ cells are found mainly in tissues that are in contact with the environment, such as the gills, epipharyngeal ridge, and intestine, rather than in systemic organs, like typhlosole and kidney, thereby hinting a barrier protective role in lampreys.

In conclusion, the discovery of two additional *VLR* genes indicates an unprecedented complexity of lymphocyte lineages of jawless vertebrates. Our comparative VLR sequence analyses and gene expression profiles align the VLRD- and VLRE-expressing cells within the T cell arm of lamprey immunity. This raises an interesting discrepancy between the VLRB antibody producing B-like lineage and the T-like lineages, of which there appear to be four or more distinct types. Future functional characterization of the cells that express the different versions of VLRD and VLRE promises to yield fresh insight into the evolution of T-like pathways of lymphocyte differentiation in jawless vertebrates.

Limitations of the study

The current lack of VLRD- and VLRE-specific monoclonal antibodies precludes the isolation of VLRD$^+$ and VLRE$^+$ lymphocytes, thus hindering the in-depth characterization of gene expression and other salient features of these cells. The assessment of mature *VLRD* and *VLRE* sequences poses challenges, as VLRD$^+$ and VLRE$^+$ cells are rare in the different developmental stages and immune states analyzed so far. The speculation that VLRD and VLRE are expressed on different T cell subsets is based primarily on genomic characterization of the *VLRD* and *VLRE* loci, donor cassette sharing among mature *VLRA*, *VLRC*, *VLRD*, and *VLRE*, phylogenetic analysis, and gene expression analysis. We are developing anti-VLRD and anti-VLRE reagents to enhance exploration of the evolution and diversification of T-like lymphocytes in jawless vertebrates and the roles of cytidine deaminases in the assembly of the *VLRD* and *VLRE* genes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Lamprey species
- METHOD DETAILS
  - Lamprey genome and transcriptome analysis

- Flow cytometric analysis and cell sorting
- Genomic PCR and cloning
- Quantitative real-time PCR
- Hybridization chain reaction
- Transmembrane domain and 3D structure prediction
- Sequence alignment and phylogenetic trees

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112933.

ACKNOWLEDGMENTS

This study was supported by National Institutes of Health grants R01AI072435, R35GM122591, and GM108838, National Science Foundation grants 1655163 and 1755418, the Georgia Research Alliance, and the Max Planck Society. We thank the Emory University Integrated Cellular Imaging Core of the Winship Cancer Institute for help with confocal images and R.E. Karaffa II and K. Fife (the Emory University School of Medicine Flow Cytometry Core) for help with cell sorting. We also thank the Stowers Institute for Medical Research for granting us access to the genomic and transcriptomic data of multiple lamprey species from SIMRbase.

AUTHOR CONTRIBUTIONS

S.D., T.B., J.P.R., M.H., and M.D.C. designed research; S.D., S.J.H., J.P.R., F.F.-I., R.M., J.G.V., R.D.H., and M.H. performed research; all authors analyzed data; S.D., T.B., J.P.R., M.H., and M.D.C. wrote the paper.

DECLARATION OF INTERESTS

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. However, studies reported in this manuscript are not related to lamprey antibodies.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: April 7, 2023
Revised: June 20, 2023
Accepted: July 18, 2023
Published: August 4, 2023

REFERENCES

1. Boehm, T., Hirano, M., Holland, S.J., Das, S., Schorpp, M., and Cooper, M.D. (2018). Evolution of Alternative Adaptive Immune Systems in Vertebrates. Annu. Rev. Immunol. *36*, 19–42. https://doi.org/10.1146/annurev-immunol-042617-053028.
2. Cooper, M.D., and Alder, M.N. (2006). The evolution of adaptive immune systems. Cell *124*, 815–822. https://doi.org/10.1016/j.cell.2006.02.001.
3. Flajnik, M.F., and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. *11*, 47–59. https://doi.org/10.1038/nrg2703.
4. Trancoso, I., Morimoto, R., and Boehm, T. (2020). Co-evolution of mutagenic genome editors and vertebrate adaptive immunity. Curr. Opin. Immunol. *65*, 32–41. https://doi.org/10.1016/j.coi.2020.03.001.
5. Pancer, Z., Amemiya, C.T., Ehrhardt, G.R.A., Ceitlin, J., Gartland, G.L., and Cooper, M.D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature *430*, 174–180. https://doi.org/10.1038/nature02740.
6. Alder, M.N., Rogozin, I.B., Iyer, L.M., Glazko, G.V., Cooper, M.D., and Pancer, Z. (2005). Diversity and function of adaptive immune receptors in a jawless vertebrate. Science *310*, 1970–1973. https://doi.org/10.1126/science.1119420.
7. Nagawa, F., Kishishita, N., Shimizu, K., Hirose, S., Miyoshi, M., Nezu, J., Nishimura, T., Nishizumi, H., Takahashi, Y., Hashimoto, S.i., et al. (2007). Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. Nat. Immunol. *8*, 206–213. https://doi.org/10.1038/ni1419.
8. Hirano, M., Das, S., Guo, P., and Cooper, M.D. (2011). The evolution of adaptive immunity in vertebrates. Adv. Immunol. *109*, 125–157. https://doi.org/10.1016/B978-0-12-387664-5.00004-2.
9. Das, S., Li, J., Hirano, M., Sutoh, Y., Herrin, B.R., and Cooper, M.D. (2015). Evolution of two prototypic T cell lineages. Cell. Immunol. *296*, 87–94. https://doi.org/10.1016/j.cellimm.2015.04.007.
10. Pancer, Z., Saha, N.R., Kasamatsu, J., Suzuki, T., Amemiya, C.T., Kasahara, M., and Cooper, M.D. (2005). Variable lymphocyte receptors in hagfish. Proc. Natl. Acad. Sci. USA *102*, 9224–9229. https://doi.org/10.1073/pnas.0503792102.
11. Kasamatsu, J., Sutoh, Y., Fugo, K., Otsuka, N., Iwabuchi, K., and Kasahara, M. (2010). Identification of a third variable lymphocyte receptor in the lamprey. Proc. Natl. Acad. Sci. USA *107*, 14304–14308. https://doi.org/10.1073/pnas.1001910107.
12. Rogozin, I.B., Iyer, L.M., Liang, L., Glazko, G.V., Liston, V.G., Pavlov, Y.I., Aravind, L., and Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat. Immunol. *8*, 647–656. https://doi.org/10.1038/ni1463.
13. Li, J., Das, S., Herrin, B.R., Hirano, M., and Cooper, M.D. (2013). Definition of a third VLR gene in hagfish. Proc. Natl. Acad. Sci. USA *110*, 15013–15018. https://doi.org/10.1073/pnas.1314540110.
14. Morimoto, R., O'Meara, C.P., Holland, S.J., Trancoso, I., Souissi, A., Schorpp, M., Vassaux, D., Iwanami, N., Giorgetti, O.B., Evanno, G., and Boehm, T. (2020). Cytidine deaminase 2 is required for VLRB antibody gene assembly in lampreys. Sci. Immunol. *5*, eaba0925. https://doi.org/10.1126/sciimmunol.aba0925.
15. Bajoghli, B., Guo, P., Aghaallaei, N., Hirano, M., Strohmeier, C., McCurley, N., Bockman, D.E., Schorpp, M., Cooper, M.D., and Boehm, T. (2011). A thymus candidate in lampreys. Nature *470*, 90–94. https://doi.org/10.1038/nature09655.
16. Kishishita, N., Matsuno, T., Takahashi, Y., Takaba, H., Nishizumi, H., and Nagawa, F. (2010). Regulation of antigen-receptor gene assembly in hagfish. EMBO Rep. *11*, 126–132. https://doi.org/10.1038/embor.2009.274.
17. Hirano, M., Guo, P., McCurley, N., Schorpp, M., Das, S., Boehm, T., and Cooper, M.D. (2013). Evolutionary implications of a third lymphocyte lineage in lampreys. Nature *501*, 435–438. https://doi.org/10.1038/nature12467.
18. Alder, M.N., Herrin, B.R., Sadlonova, A., Stockard, C.R., Grizzle, W.E., Gartland, L.A., Gartland, G.L., Boydston, J.A., Turnbough, C.L., Jr., and Cooper, M.D. (2008). Antibody responses of variable lymphocyte receptors in the lamprey. Nat. Immunol. *9*, 319–327. https://doi.org/10.1038/ni1562.
19. Herrin, B.R., Alder, M.N., Roux, K.H., Sina, C., Ehrhardt, G.R.A., Boydston, J.A., Turnbough, C.L., Jr., and Cooper, M.D. (2008). Structure and specificity of lamprey monoclonal antibodies. Proc. Natl. Acad. Sci. USA *105*, 2040–2045. https://doi.org/10.1073/pnas.0711619105.
20. Guo, P., Hirano, M., Herrin, B.R., Li, J., Yu, C., Sadlonova, A., and Cooper, M.D. (2009). Dual nature of the adaptive immune system in lampreys. Nature *459*, 796–801. https://doi.org/10.1038/nature08068.
21. Dorshkind, K., and Montecino-Rodriguez, E. (2007). Fetal B-cell lymphopoiesis and the emergence of B-1-cell potential. Nat. Rev. Immunol. *7*, 213–219. https://doi.org/10.1038/nri2019.

22. Bluestone, J.A., Mackay, C.R., O'Shea, J.J., and Stockinger, B. (2009). The functional plasticity of T cell subsets. Nat. Rev. Immunol. *9*, 811–816. https://doi.org/10.1038/nri2654.

23. Kanda, R., Sutoh, Y., Kasamatsu, J., Maenaka, K., Kasahara, M., and Ose, T. (2014). Crystal structure of the lamprey variable lymphocyte receptor C reveals an unusual feature in its N-terminal capping module. PLoS One *9*, e85875. https://doi.org/10.1371/journal.pone.0085875.

24. Timoshevskaya, N., Eşkut, K.I., Timoshevskiy, V.A., Robb, S.M.C., Holt, C., Hess, J.E., Parker, H.J., Baker, C.F., Miller, A.K., Saraceno, C., et al. (2023). An improved germline genome assembly for the sea lamprey Petromyzon marinus illuminates the evolution of germline-specific chromosomes. Cell Rep. *42*, 112263. https://doi.org/10.1016/j.celrep.2023.112263.

25. Smith, J.J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M.C., Parker, H.J., Cook, M.E., Hess, J.E., Narum, S.R., Lamanna, F., et al. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. Nat. Genet. *50*, 270–277. https://doi.org/10.1038/s41588-017-0036-1.

26. Das, S., Li, J., Holland, S.J., Iyer, L.M., Hirano, M., Schorpp, M., Aravind, L., Cooper, M.D., and Boehm, T. (2014). Genomic donor cassette sharing during VLRA and VLRC assembly in jawless vertebrates. Proc. Natl. Acad. Sci. USA *111*, 14828–14833. https://doi.org/10.1073/pnas.1415580111.

27. Das, S., Rast, J.P., Li, J., Kadota, M., Donald, J.A., Kuraku, S., Hirano, M., and Cooper, M.D. (2021). Evolution of variable lymphocyte receptor B antibody loci in jawless vertebrates. Proc. Natl. Acad. Sci. USA *118*, e2116522118. https://doi.org/10.1073/pnas.2116522118.

28. Hess, J.E., Smith, J.J., Timoshevskaya, N., Baker, C., Caudill, C.C., Graves, D., Keefer, M.L., Kinziger, A.P., Moser, M.L., Porter, L.L., et al. (2020). Genomic islands of divergence infer a phenotypic landscape in Pacific lamprey. Mol. Ecol. *29*, 3841–3856. https://doi.org/10.1111/mec.15605.

29. Brownstein, C.D., and Near, T.J. (2023). Phylogenetics and the Cenozoic radiation of lampreys. Curr. Biol. *33*, 397–404.e3. https://doi.org/10.1016/j.cub.2022.12.018.

30. Dirks, R.M., and Pierce, N.A. (2004). Triggered amplification by hybridization chain reaction. Proc. Natl. Acad. Sci. USA *101*, 15275–15278. https://doi.org/10.1073/pnas.0407024101.

31. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753. https://doi.org/10.1242/dev.165753.

32. Han, B.W., Herrin, B.R., Cooper, M.D., and Wilson, I.A. (2008). Antigen recognition by variable lymphocyte receptors. Science *321*, 1834–1837. https://doi.org/10.1126/science.1162484.

33. Velikovsky, C.A., Deng, L., Tasumi, S., Iyer, L.M., Kerzic, M.C., Aravind, L., Pancer, Z., and Mariuzza, R.A. (2009). Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen. Nat. Struct. Mol. Biol. *16*, 725–730. https://doi.org/10.1038/nsmb.1619.

34. Das, S., Hirano, M., Aghaallaei, N., Bajoghli, B., Boehm, T., and Cooper, M.D. (2013). Organization of lamprey variable lymphocyte receptor C locus and repertoire development. Proc. Natl. Acad. Sci. USA *110*, 6043–6048. https://doi.org/10.1073/pnas.1302500110.

35. Lipkowitz, S., Stern, M.H., and Kirsch, I.R. (1990). Hybrid T cell receptor genes formed by interlocus recombination in normal and ataxia-telangiectasis lymphocytes. J. Exp. Med. *172*, 409–418. https://doi.org/10.1084/jem.172.2.409.

36. Baer, R., Chen, K.C., Smith, S.D., and Rabbitts, T.H. (1985). Fusion of an immunoglobulin variable gene and a T cell receptor constant gene in the chromosome 14 inversion associated with T cell tumors. Cell *43*, 705–713. https://doi.org/10.1016/0092-8674(85)90243-0.

37. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. *38*, 3022–3027. https://doi.org/10.1093/molbev/msab120.

38. Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics *2*, 2.3. https://doi.org/10.1002/0471250953.bi0203s00.

39. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

40. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

41. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580. https://doi.org/10.1006/jmbi.2000.4315.

42. Tusnády, G.E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. Bioinformatics *17*, 849–850. https://doi.org/10.1093/bioinformatics/17.9.849.

43. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

44. Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. *28*, 231–234. https://doi.org/10.1093/nar/28.1.231.

45. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

46. Holland, S.J., Berghuis, L.M., King, J.J., Iyer, L.M., Sikora, K., Fifield, H., Peter, S., Quinlan, E.M., Sugahara, F., Shingate, P., et al. (2018). Expansions, diversification, and interindividual copy number variations of AID/APOBEC family cytidine deaminase genes in lampreys. Proc. Natl. Acad. Sci. USA *115*, E3211–E3220. https://doi.org/10.1073/pnas.1720871115.

47. Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. *8*, 275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-VLRA (R110)	Guo P et al.[20]	N/A
Mouse monoclonal anti-VLRB (4C4)	Alder et al.[6]	N/A
Mouse monoclonal anti-VLRC (3A5)	Hirano et al.[17]	N/A
Bacterial and virus strains		
Chemically Competent *E. coli*	Invitrogen	Cat# K287520
JM109 Competent cells	Promega	Cat# L2001
Biological samples		
Blood, Kidney, Typhlosole, Intestine, Gill, and Skin from sea lamprey (*P. marinus*) larvae	M. D. Cooper's lab	N/A
Blood, Kidney, Typhlosole, Intestine, and Gill from European brook lamprey (*L. planeri*) larvae	T. Boehm's lab	N/A
Chemicals, peptides, and recombinant proteins		
20X SSC Buffer	Thermo Fisher Scientific	Cat#AM9763
UtraPureTM Distilled Water	Thermo Fisher Scientific	Cat#10977015
Tween® 20	Sigma-Aldrich	P9416-50ML
PBS, 1X	Corning	21-040-CV
TRICAINE-S	Syndel	N/A
Percoll® PLUS	Cytiva	17-5445-01
HCR™ Amplifiers	Molecular instruments, Inc	N/A
HCR™ probe hybridization buffer	Molecular instruments, Inc	N/A
HCR™ probe wash buffer	Molecular instruments, Inc	N/A
HCR™ probe amplification buffer	Molecular instruments, Inc	N/A
VECTASHIELD Antifade Mounting Medium with DAPI	Vector Laboratories	H-1200-10
Critical commercial assays		
Zero Blunt™ TOPO™ PCR Cloning Kit	Invitrogen	Cat# K287520
Nucleospin® Gel and PCR Clean-up	Machery-Nagel	740609.50
pGEMT Easy	Promega	Cat# A1360
Trizol	Thermo Fisher Scientific	Cat# 15596026
QIAquick Gel Extraction Kit	Qiagen	Cat# 28704
QIAprep Spin Miniprep Kit	Qiagen	Cat# 27104
Q5 2x Mastermix	New England Biolabs	Cat# M0492S
Superscript III Reverse Transcription system	Invitrogen	Cat# 18080051
RNase-Free DNAse Set	Qiagen	Cat#79254
SYBR Green PCR Master Mix	Applied Biosystems	Cat# 4309155
Deposited data		
cDNA sequences	This paper	GenBank accession numbers: OQ595148-OQ595181
Genomic DNA sequences	This paper	GenBank accession numbers: OQ604520-OQ604523
Experimental models: Organisms/strains		
Sea lamprey (*P. marinus*)	Lamprey Service (Michigan, USA)	N/A
European brook lamprey (*L. planeri*)	March (Breisgau, Germany)	N/A
Oligonucleotides		
PCR primers, see Table S2	IDT	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
VLRA probes	Molecular instruments, Inc	PRN524
VLRB probes	Molecular instruments, Inc	PRP929
VLRD probes	Molecular instruments, Inc	PRN525
VLRE probes	Molecular instruments, Inc	PRN526
Software and algorithms		
MEGA software (version 11) package	Tamura et al.[37]	https://www.megasoftware.net/
CLUSTALW	Thompson et al.[38]	https://www.megasoftware.net/
Neighbor-joining phylogenetic tree building algorithm	Saitou and Nei[39]	https://www.megasoftware.net/
BLAST search algorithm	Altschul et al.[40]	https://blast.ncbi.nlm.nih.gov/Blast.cgi https://www.ensembl.org/Multi/Tools/Blast
Prism	GraphPad	https://www.graphpad.com/
TMHMM	Krogh et al.[41]	https://services.healthtech.dtu.dk/services/TMHMM-2.0/
HMMTOP	Tusnady and Simon[42]	http://www.enzim.hu/hmmtop/index.php
ImageJ	Schneider et al.[43]	https://ImageJ.nih.gov/ij/
MACSQuant Analyzer	Miltenyi Biotec	N/A
SMART	Schultz et al.[44]	http://smart.embl-heidelberg.de/
AlphaFold	Jumper et al.[45]	https://alphafold.ebi.ac.uk/
PyMOL Molecular Graphics System, Version 2.0	Schrödinger, LLC	https://pymol.org/2/
Other		
Sea lamprey, Japanese lamprey, Far eastern brook lamprey, Western brook lamprey, Pacific lamprey genome assemblies (See Table S1)	National Center for Biotechnology Information and SIMRbase	https://www.ncbi.nlm.nih.gov/genome/ https://simrbase.stowers.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Max D. Cooper (mdcoope@emory.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The cDNA and genomic sequences generated in the present study are publicly available in the GenBank database of the National Center for Biotechnology Information (NCBI) under the accession numbers OQ595148-OQ595181 and OQ604520-OQ604523. Microscopy data reported in this paper will be shared by the lead contact upon request.
- This paper does not report the original code.
- Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Lamprey species

Larvae (outbred, 8–15 cm long, age 3–4 years) of sea lamprey (*Petromyzon marinus*) and European brook lamprey (*Lampetra planeri*) were purchased from local suppliers and maintained in sand-lined aquariums at 18 °C. Animals are immature at this stage and sex could not be determined for all specimens. All experiments were performed in accordance with the relevant guidelines and regulations and approved by the Institutional Animal Care and Use Committee at Emory University and the Review Committee of the Max-Planck Institute.

METHOD DETAILS

Lamprey genome and transcriptome analysis

Previously described VLR sequences (VLRA, VLRB and VLRC) were used as queries for TBLASTN search against sea lamprey genome sequence. The extension of a genomic hit that contains a unique LRRCT region revealed a germline VLR-like gene. In the next step, another round of TBLASTN search was conducted using the amino acid sequences of C-terminal coding region of the newly identified VLR-like gene as query against sea lamprey, Japanese lamprey, Far Eastern brook lamprey, Western brook lamprey, and Pacific lamprey genome sequences, as well as against the available transcriptome sequences of sea lamprey and European brook lamprey,[46] to retrieve additional VLR-like genes (see Table S1). To identify genomic donor cassettes, we used two rounds of BLASTN searches as described previously[34] against sea lamprey genome sequence (kPetMar1) using 50 mature sequences as queries for the first-round similarity search.

Flow cytometric analysis and cell sorting

Leukocytes isolated from sea lamprey blood were stained for examination by immunofluorescence flow cytometry as described previously.[17] Briefly, buffy coat leukocytes from blood were stained with primary antibodies including rabbit anti-VLRA polyclonal serum (R110), mouse anti-VLRB mAb (4C4), mouse anti-VLRC mAb (3A5) and their matched secondary antibodies. Cells were gated using forward scatter-A (FSC-A) vs. side scatter-A (SSC-A) (lymphocytes), FSC-A vs. FSC-H (singlets), and negative LIVE/DEAD Aqua (Invitrogen) staining (live cells). Flow cytometric analysis was performed on a MACSQuant Analyzer (Miltenyi Biotec) and VLRA$^+$, VLRB$^+$, VLRC$^+$, VLR triple-negative (TN) cells were sorted on BD FACS Aria II (BD Bioscience) for real-time PCR analysis. The purity of the sorted cells was >90%.

Genomic PCR and cloning

Genomic DNA was extracted from the whole blood of lamprey larvae using the DNeasy kit (QIAGEN). Primers used for genomic PCR are listed in Table S2. PCR products were cloned with the Zero Blunt TOPO PCR Cloning Kit (Invitrogen) and then sequenced.

Quantitative real-time PCR

Different tissues from lamprey larvae were dissected and extracted for RNA isolation using RNeasy kits with on-column DNA digestion by DNase I (QIAGEN). First-strand cDNA was synthesized with random hexamer primers by Superscript IV (Invitrogen). Quantitative real-time PCR was conducted using SYBR Green on 7900HT ABI Prism (Applied Biosystems) and all samples were run in three replicates. The data were analyzed using one-way repeated measures by ANOVA performed with GraphPad Prism. The values for *VLR* genes were normalized to the expression of β-actin. Primers used in this analysis are listed in Table S2.

Hybridization chain reaction

Hybridization chain reaction (HCR) was performed as described by Choi et al.,[31] with slight modification. Sets of probes, hairpins, hybridization buffer, amplification buffer and wash buffer were purchased from Molecular Instruments, Inc. (USA). Briefly, fresh frozen sections of lamprey larvae were fixed in 4% paraformaldehyde in PBS at 4°C for 15 min and dehydrated with ethanol. After washing thrice with PBS, sections were pre-incubated with hybridization buffer at room temperature (RT) for 10 min. Slides were incubated overnight at 37°C with the probe sets of *VLRA*, *VLRB*, *VLRD* and *VLRE* diluted at 10 nM in hybridization buffer. Excess probes were removed by serial incubations of 30 min at 37°C with wash buffer 100%, 75%, 50% and 25% in 5X Saline Sodium Citrate buffer (SSCT; Thermo Fisher Scientific) 0.1% Tween 20. After the final incubation of 30 min at 37°C in SSCT, sections were incubated with the pre-amplification buffer for 30 min at RT. Six pmol of each pair of hairpins were independently snap cooled by heating at 95°C for 90 s, allowed to cool for 30 min to room temperature, and diluted at 40 nM in amplification buffer at 37°C. The probe solution was added to the samples and incubated overnight at RT. Samples were then washed twice in 5X SSCT for 30 min and 5 min at RT. Slides were mounted with Antifade Mounting Medium with DAPI (Vector Laboratories). All incubation steps were carried out in a humidified chamber. Images were captured with a Leica SP8 confocal microscope or an Axiovert 200M equiped with a AxioCam MRc (Zeiss).

Transmembrane domain and 3D structure prediction

Transmembrane domain was predicted by TMHMM[41] and HMMTOP[42] software. LRR domains are identified by SMART sequence analysis tool.[44] The 3D structure prediction was conducted using AlphaFold,[45] an artificial intelligence (AI) system available at EMBL's European Bioinformatics Institute (https://alphafold.ebi.ac.uk/) and visualized by PyMOL software (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Sequence alignment and phylogenetic trees

Sequences were aligned with CLUSTALW program[38] and also manually inspected. Neighbor-joining trees[39] were constructed using the MEGA software (version 11) with the pairwise deletion option.[37] The JTT matrix-based method[47] was used to compute the evolutionary distances.

QUANTIFICATION AND STATISTICAL ANALYSIS

Student's t-test was used for statistical analysis. For phylogenetic trees the reliability of branching patterns was assessed by bootstrap resampling with 1000 replications.