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

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

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.
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. Cooper1,2,5,*

1Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA 30322, USA
2Emory Vaccine Center, Emory University, Atlanta, GA 30317, USA
3Department of Developmental Immunology, Max-Planck Institute of Immunobiology and Epigenetics, Stuebeweg 51, 79108 Freiburg, Germany
4Faculty of Medicine, University of Freiburg, Breisacher Str. 153, 79110 Freiburg, Germany
5Lead contact
*Correspondence: scdas@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) in the addition of leucine-rich repeat (LRR)-encoding donor cassettes. 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 non-coding 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.5,7,12,14 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 TCRαβ+ and TCRγδ+ 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 VLREs (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 ≥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 ≥52-residue-long LRRCT module, and a ≥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. The cysteine configuration of LRRNT modules of VLRA, VLRB, and VLRC isotypes corresponds to C1-Xm-C2-X-C3-Xn-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 VLRA and VLRC exhibit a C1-X-C2-Xm-C3-Xn-C4 signature, VLRA is notable for its C1-X-X-C2-Xm-C3-Xn-C4 signature. The two residues separating C1 and C2 of LRRCT of 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-Xm-C2-C-X-C3) in their LRRCT 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, 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
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.

### Genomic organization of the VLRD and VLRE loci

Using an iterative similarity search strategy, we mapped the genomic sequences of sea lamprey (kPetMar1).24,25 The genomic donor cassettes along with the incomplete germline gene; cassettes from both of these chromosomes are used in assembled VLRD genes, both germline 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 VLRA gene (Figure 2). In contrast to the situation for VLRA 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 genes used in VLRD and VLRE assemblies are located in two clusters (cluster I and cluster II of the VLRA locus) 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 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 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 58. Cassettes are shown in proportion to genomic spacing, but icons are not to scale. The region shown is 600 kb downstream of the chromosome 58 used exclusively by VLRD and genomic donor cassettes along with the incomplete germline gene; cassettes from both of these chromosomes are used in assembled VLRD sequences 1 and VLRD sequence 2 in sea lampreys, respectively.

On chromosome 58 (GenBank: NC_046126.1), the latter harboring the VLRA 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.

---

**Table S1**

<table>
<thead>
<tr>
<th>LRRNT</th>
<th>LRR1</th>
<th>LRRV</th>
<th>CP</th>
<th>LRRCT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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)
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 VLRA and these two reported VLR assemblies (or with 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.25,26,27

**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

---

**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.
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 like in the sea lamprey, genomic donor cassette sharing is also

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

To examine the cellular expression patterns of levels for VLRA+ lymphocytes are higher than those for VLRD and VLRE and cellular and tissue expression patterns of sequences in the European brook lamprey (Figure S6). To identify cells expressing mRNA transcripts of the identified 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
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, 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. 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. The lack of 5′LRRCT-LRRCTm cassettes for VLRD and VLRE in any of the lamprey genome sequences...
analyzed here indicates that except for the first five amino acid residues (which may be contributed by 3'LR-C-5'LLRCT cassette), 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.

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 with 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 both VLRA and VLRE transcripts.
(L) Positive control for HCR experiment showing VLRB+ cells dispersed in the typhlosole parenchyma, showing a VLRB+ cell (arrow) and VLRB low cell (arrowhead). The square in the bottom shows a magnification of one VLRB+ 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.
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 theVLRD and VLRE genes. The germline VLRA and VLRC 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 matureVLRA and VLRC assemblies. The genomic constellation strongly suggests that a local gene duplication event gave rise to VLRA and VLRC genes. Interestingly, genomic donor cassette usage among VLRA 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, VLRA, and VLRE assemblies (see Figures 3 and 4).

Another interesting trend for the incorporation of genomic donor cassettes into mature VLRA/VLRE assemblies is worthy of note. All of the 3-LRRV-CP-S-LRRCT cassettes that contribute to mature VLRA/VLRE assemblies are located in cluster II on chromosome 58, whereas LRRNT-, LRR1-, and LRRV1-encoding sequences originate from the VLRA/VLRE 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 two clusters on chromosome 58, with occasional contributions of LRRV-encoding genomic donor cassettes that contain the germline VLRA (cluster I of chromosome 58) and VLRC (chromosome 7) genes are used for VLRA, VLRC, VLRA, and VLRE assemblies (see Figures 3 and 4).

Our previous studies demonstrated that VLRA 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 αβ and γδ lineages of jawed vertebrates.\(^{17,20}\) Distinct cytidine deaminases appear to be responsible for the assembly of VLRA genes in lampreys; CDA2 has been shown to be required for VLRA 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 VLRA\(^+\) cells express CDA2.\(^{17,20}\) The CDA1 and VLRA mRNA were detected at gill thymus regions in lampreys.\(^{15}\) Here, we found high expression of VLRA and VLRC (in measurements that include both germline and assembled sequence transcripts) in the gill region (Figure 6), and HCR in situ data indicate that VLRC\(^+\) cells are located both in the thyroid area and in the gill filaments (Figure 7). Expression levels of VLRA and VLRC are particularly high in the triple-negative (VLRA\(^−\)/VLRB\(^−\)/VLRC\(^−\)) population of lamprey blood lymphocytes. Although VLRA/VLRC 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, VLRA, and VLRC loci. Whether two copies of VLRA and VLRC 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 VLRA- and VLRC-expressing cells appear to represent minor populations of specialized T-like lymphocytes. Notably, the VLRC\(^+\) 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 VLRA genes indicates an unprecedented complexity of lymphocyte lineages of jawless vertebrates. Our comparative VLRA sequence analyses and gene expression profiles align the VLRA- and VLRC-expressing cells within the T cell arm of lamprey immunity. This raises an interesting discrepancy between the VLRA 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 VLRA and VLRC 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 VLRA- and VLRC-specific monoclonal antibodies precludes the isolation of VLRA\(^+\) and VLRC\(^+\) lymphocytes, thus hindering the in-depth characterization of gene expression and other salient features of these cells. The assessment of mature VLRA and VLRC sequences poses challenges, as VLRA\(^+\) and VLRC\(^+\) cells are rare in the different developmental stages and immune states analyzed so far. The speculation that VLRA and VLRC are expressed on different T cell subsets is based primarily on genomic characterization of the VLRA and VLRC loci, donor cassette sharing among mature VLRA, VLRC, VLRA, and VLRC, phylogenetic analysis, and gene expression analysis. We are developing anti-VLRA and anti-VLRC 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 VLRA and VLRC 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
We support inclusive, diverse, and equitable conduct of research.

INCLUSION AND DIVERSITY

NovAb. However, studies reported in this manuscript are not related to lamprey antibodies for biomedical purposes, and J.P.R. is a consultant for M.D.C. is a cofounder and shareholder of NovAb, Inc., which produces lamprey antibodies for biomedical purposes; 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.

REFERENCES


**STAR METHODS**

**KEY RESOURCES TABLE**

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

(Continued on next page)
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, to retrieve additional VLR-like genes (see Table S1). To identify genomic donor cassettes, we used two rounds of BLASTN searches as described previously 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. 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., 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, VLRC 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 equipped with a AxioCam MRc (Zeiss).

Transmembrane domain and 3D structure prediction
Transmembrane domain was predicted by TMHMM and HMMTOP software. LRR domains are identified by SMART sequence analysis tool. The 3D structure prediction was conducted using AlphaFold, 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\textsuperscript{38} and also manually inspected. Neighbor-joining trees\textsuperscript{39} were constructed using the MEGA software (version 11) with the pairwise deletion option.\textsuperscript{37} The JTT matrix-based method\textsuperscript{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.