



(86) Date de dépôt PCT/PCT Filing Date: 2005/05/23
(87) Date publication PCT/PCT Publication Date: 2006/08/10
(85) Entrée phase nationale/National Entry: 2006/10/25
(86) N° demande PCT/PCT Application No.: US 2005/017901
(87) N° publication PCT/PCT Publication No.: 2006/083275
(30) Priorité/Priority: 2004/05/21 (US60/573,563)

(51) Cl.Int./Int.Cl. *C07K 14/705* (2006.01)
(71) Demandeurs/Applicants:
THE UAB RESEARCH FOUNDATION, US;
BENAROYA RESEARCH INSTITUTE, US
(72) Inventeurs/Inventors:
PANCER, ZEEV, US;
COOPER, MAX D., US;
AMEMIYA, CHRIS, US;
GARTLAND, G. LARRY, US;
EHRHARDT, GOETZ, US
(74) Agent: MARKS & CLERK

(54) Titre : RECEPTEURS DE LYMPHOCYTES VARIABLES, POLYPEPTIDES ET ACIDES NUCLEIQUES ASSOCIES,
ET LEURS UTILISATIONS
(54) Title: VARIABLE LYMPHOCYTE RECEPTORS, RELATED POLYPEPTIDES AND NUCLEIC ACIDS, AND USES
THEREOF

(57) Abrégé/Abstract:

Disclosed are compositions and methods related to variable lymphocyte receptors (VLRs).



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 August 2006 (10.08.2006)

PCT

(10) International Publication Number
WO 2006/083275 A2

(51) International Patent Classification:
C07K 14/705 (2006.01)

(74) Agents: **MCKEON, Tina, W.** et al.; NEEDLE & ROSENBERG, P.C., SUITE 1000, 999 Peachtree Street, Atlanta, Georgia 30309-3915 (US).

(21) International Application Number:
PCT/US2005/017901

(22) International Filing Date: 23 May 2005 (23.05.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/573,563 21 May 2004 (21.05.2004) US

(71) Applicants (*for all designated States except US*): **THE UAB RESEARCH FOUNDATION** [US/US]; 1530 3RD AVENUE SOUTH, Ab 1120g, Birmingham, Alabama 35294-0111 (US). **BENAROYA RESEARCH INSTITUTE** [US/US]; 1201 Ninth Avenue, Seattle, Washington 98101 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **PANCER, Zeev** [IL/US]; 8 CHARLES PLAZA, Apartment 2105, Baltimore, Maryland 21201 (US). **COOPER, Max, D.** [US/US]; 3228 Carlisle Road, Birmingham, Alabama 35213 (US). **AMEMIYA, Chris** [US/US]; 2414 First Avenue, 706, Seattle, Washington 98121 (US). **GARTLAND, G., Larry** [US/US]; 3740 Valley Head Road, Birmingham, Alabama 35223 (US). **EHRHARDT, Goetz** [US/US]; 2251 HIGHLAND AVENUE SOUTH, Apt. 1109, Birmingham, Alabama 35205 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VARIABLE LYMPHOCYTE RECEPTORS, RELATED POLYPEPTIDES AND NUCLEIC ACIDS, AND USES THEREOF

(57) Abstract: Disclosed are compositions and methods related to variable lymphocyte receptors (VLRs).

WO 2006/083275 A2

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 73

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 73

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

VARIABLE LYMPHOCYTE RECEPTORS, RELATED POLYPEPTIDES AND NUCLEIC ACIDS, AND USES THEREOF

The application claims the benefit of U.S. provisional Application 60/573,563, filed May 21, 2004, which is incorporated herein by reference in its entirety.

5 This invention was made with government support under NIH/NIAID Grant AI39816 and HG02526-01 and NSF Grants MCB-0317460 and IBN-0321461. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

10 Adaptive immune responses in jawed vertebrates are initiated when antigens are recognized by specific lymphocyte receptors. Antigen receptor diversity is generated via recombination of variable, diversity and joining gene segments in the immunoglobulin (Ig) and T cell receptor (TCR) gene loci. This combinatorial rearrangement generates vast
15 repertoires of antibodies against unprocessed antigens and of TCRs that recognize antigen fragments presented within the cusp of major histocompatibility complex (MHC) class I and II molecules. Clonally diverse lymphocytes thus form the cornerstone of vertebrate adaptive immunity in the form of Ig bearing B cells and TCR bearing T cells that differentiate from stem cell precursors within primary hematopoietic tissues and the thymus. Cardinal elements of this recombinatorial immune system are conserved in all
20 jawed vertebrates and the multigene TCR and Ig loci are remarkably complex even in the most basal gnathostome representatives, sharks, skates, and rays (Rast et al., 1997; Flajnik and Kasahara, 2001; Flajnik, 2002).

There is also abundant evidence for adaptive immunity in the jawless vertebrates, lamprey and hagfish, the only surviving descendents from the early vertebrate radiation
25 (Forey and Janvier, 1993). Humoral and cell mediated types of immunologic responses have been reported for these agnathans. For example, lampreys produce specific circulating agglutinins in response to primary antigenic stimulation, make higher agglutinin levels after booster immunization (Finstad and Good, 1964; Marchalonis and Edelman, 1968; Litman et al., 1970; Pollara et al., 1970; Good et al., 1972; Hagen et al.,
30 1985), reject second set skin allografts at an accelerated rate (Finstad et al., 1964; Perey et

al., 1968; Good et al., 1972; Fujii and Hayakawa, 1983) and exhibit delayed type hypersensitivity reactions (Finstad and Good, 1964; Good et al., 1972). Agnathan adaptive immune responses have been attributed to cells that morphologically resemble the lymphocytes found in the lympho-hematopoietic tissues and blood of jawed
5 vertebrates (Finstad and Good, 1964; Finstad et al., 1964; Perey et al., 1968; Cooper, 1971; Piavis and Hiatt, 1971; Good et al., 1972; Kilarski and Plytycz, 1981; Zapata et al., 1981; Fujii, 1982; Fujii and Hayakawa, 1983; Ardavin and Zapata, 1987; Mayer et al., 2002a). Like their mammalian counterparts, lamprey lymphocytes are more irradiation sensitive than other blood cell types (Good et al., 1972), aggregate and proliferate in
10 response to antigenic stimulation (Finstad and Good, 1964; Cooper, 1971; Piavis and Hiatt, 1971), and express transcription factors that are involved in mammalian lymphocyte differentiation, such as PU.1/Spi-B and Ikaros (Haire et al., 2000; Shintani et al., 2000; Anderson et al., 2001; Mayer et al., 2002b). Surprisingly, however, Ig, TCR, and MHC genes have not been previously identified in jawless vertebrates or in the
15 genome sequence of the invertebrate urochordate *Ciona intestinalis* (Azumi et al., 2003). The present invention relates to a novel lymphocyte receptor and nucleic acids that encode a novel lymphocyte receptor.

SUMMARY OF THE INVENTION

20 In accordance with the purposes of this invention, as embodied and broadly described herein, this invention, in one aspect, relates to polypeptides comprising a novel lymphocyte receptor or fragments thereof. The invention further relates to nucleic acids that encode the lymphocyte receptors or fragments. Further provided are methods of making and using the polypeptides and nucleic acids. Such uses include a
25 broad range of purification, therapeutic and diagnostic methods.

Additional advantages of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended
30 claims. It is to be understood that both the foregoing general description and the

following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

5 BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate several embodiments of the invention and together with the description, serve to explain the principles of the invention.

Figure 1 shows lamprey leukocytes and VLRs. Figure 1a shows a light scatter
 10 analysis of blood leukocytes before and after immunostimulation with antigen/mitogen cocktail. Figure 1b shows sorted immunostimulated leukocytes: small lymphocytes (R1) large lymphocytes (R2) or myeloid cells (R3). Wright-Giemsa stain, 100X. Scale bar = 10 μ m. Figure 1c shows virtual Northern blots of VLR and GAPDH (control). Amplified cDNA from tissues or sorted cells from hematopoietic organs and blood of
 15 immunostimulated and unstimulated larvae are shown. Figure 1d shows a VLR stick model: signal peptide, N-terminal LRR, nine LRRs, connecting peptide, C-terminal LRR, threonine-proline rich stalk, GPI-anchor and hydrophobic tail (Clone 12.26, 417 residues, AY577974). Figure 1e shows the cell surface expression of epitope-tagged VLR and Fc γ RIIb (control) expressed in mouse thymoma cells, treated with (+PLC) or without (-
 20 PLC) bacterial GPI-phospholipase C. Figure 1f shows a 3D model of VLR diversity region viewed in two rotations (clone 12.26).

Figure 2 shows a survey of VLR diversity in two lamprey larvae. Alignment of 20 diversity regions PCR amplified from lymphocytes. PCR primers were located in regions conserved in all VLR sequences: signal peptide 5' to LRRNT and near 3' of LRRCT.
 25 Donor animals and clone numbers are indicated. The locations of LRR motifs are also indicated. Black: 100% identity; gray: 60-99%; white: 60%. Sequences 1.3-2.10 correspond herein to SEQ ID NOs:1-20, respectively.

Figure 3 shows an assessment of VLR protein diversity in 13 individual larvae. Genetic distance dendrogram of 112 VLR diversity regions from cDNA and genomic
 30 PCR clones. Larvae numbers and clone numbers (e.g., 6.20 = donor 6, clone 20) are

indicated in red for immunostimulated (N = 27) and green for unstimulated (N = 41) donors. Asterisk (*) indicates clones derived from single cell isolates (N = 12), including two VLRs from one isolate (9.16S, 9.16L); and clones derived from a control 10-cell pool are denoted 10C (N = 4). *Mature VLR* sequences derived from genomic DNA are in blue
 5 (N = 28; blood #10,12; carcass #11, 13). The mean diversity for the entire set is 1.36 ± 0.03 , ranging 0.28-0.54 within the groups of sequences from 13 individuals.

Figure 4 shows VLR genome blots of restriction-enzyme digested DNA that were hybridized with VLR N-terminal or C-terminal probes. Figure 4(a) shows blots of three lampreys (blood DNA #10,12; carcass #13) Only animal 13 showed a polymorphic
 10 *Bam*HI pattern. Figure 4b shows a genome spread of erythrocytes pooled from 10 lampreys. Pulse-field blot hybridization shows matching patterns for both probes, with an additional 350 kb *Not*I N-terminal band corresponding to a 5' *gVLR* duplication.

Figure 5 shows the genomic organization of the VLR locus. Figure 5a shows motifs identified in a 57 kb *gVLR* contig (AY577941) melded from clones PAC16 (44 kb)
 15 and PAC3 (33 kb) that overlap over 20 kb. Dashed lines represent PAC inserts; red bars indicate N-terminal and C-terminal probes. Figure 5b that PAC4 (58 kb, AY577942) aligns with the *gVLR* contig over 11.7 kb (nt 45,882-57,609). Cassettes of 1-3 LRRs are positioned in forward or reverse orientations: eight in the *gVLR* contig and 17 in PAC4. Figure 5c shows LR-PCR analysis of the *gVLR*. DNA from blood (#10) or body carcass
 20 (#13) amplified with primers *gVLR*.F1+*gVLR*.R1 (indicated in Figure 5a and Figure 5e). PAC16 amplicon served as control. The ~20 kb band corresponds to the germline VLR and the ~8 kb band corresponds to mature VLRs. Figure 5d shows lymphocyte specific rearrangement of mature VLRs. LR-PCR from sorted pools of 100 lymphocytes or erythrocytes. The ~14 kb band corresponds to the germline VLR and the ~1 kb band
 25 corresponds to mature VLRs that were amplified only from lymphocyte DNA. Figure 5e shows an illustration of an 8 kb mature VLR amplicon.

Figure 6 shows the multiple alignment of 22 VLR proteins predicted from EST clones (single pass 5' sequence, some incomplete C-termini). Black: full identity; yellow 80-99%; green: 60-79%; white <60%. The amino acid sequences for
 30 LyEST3090-LyEST5266 correspond to SEQ ID NOs:21-42, respectively.

Figure 7 shows an ORF of a representative VLR (cDNA clone LyEST2913, AY578059). The start methionine is at nt 118-120 and the stop codon at nt 937-939. Nucleotide sequence conserved in exons 2 and 4 of the germline VLR are colored red; the diverse 5' LRRCT corresponding to exon 3 is colored green. Structural motifs are indicated above the protein sequence; GPI cleavage site is colored blue. The amino acid sequence shown corresponds to SEQ ID NO:43, and the nucleic acid sequence shown corresponds to SEQ ID No:156.

Figure 8 shows the multiple alignment of 112 VLR diversity regions PCR amplified from 13 lampreys. Genomic and RT-PCR clones from immunostimulated and unstimulated lampreys. Unstimulated animals: animal designated #1-4 (N=41), sorted single lymphocytes from animal designated #8 (N=4) and clones from a pool of 10 cells from animal designated #8. 10C (N=4); Immune stimulated animals: from animals designated #5-7 (N=27) and sorted single lymphocytes from animal designated #9 (N=8) including one isolate with two VLRs (9.16S, 9.16L); *Mature VLRs*: larval genomic DNA extracted from blood designated #10-13 (N=28) or carcass (#11, 13). Black: 80-100% identity; yellow 60-79%; green: 40-59%; white <40%. From the top of the alignment, the amino acid sequence for 1.1 corresponds to SEQ ID NO:13, amino acid sequences 7.27-4.7 correspond to SEQ ID NOs:45-52, amino acid sequence 1.5 corresponds to SEQ ID NO:12, amino acid sequence 4.14 corresponds to SEQ ID NO:54, amino acid sequence 1.7 corresponds to SEQ ID NO:8, amino acid sequence 3.15 corresponds to SEQ ID NO:56, amino acid sequence 2.1 corresponds to SEQ ID NO:5, amino acid sequence 2.2 corresponds to 10, amino acid sequence 2.7 corresponds to SEQ ID NO:11, amino acid sequences 4.8-6.22 correspond to SEQ ID NOs:60-65, amino acid sequences 2.4 corresponds to SEQ ID NO:3, amino acid sequence 1.8 corresponds to SEQ ID NO:2, amino acid sequences 7.3-6.21 correspond to SEQ ID NOs:68-72, amino acid sequence 1.2 corresponds to SEQ ID NO:5, amino acid sequence 2.14 corresponds to SEQ ID NO:6, amino acid sequence 3.7 corresponds to SEQ ID NO:75, amino acid sequence 1.6 corresponds to SEQ ID NO:7, amino acid sequence 5.3 corresponds to SEQ ID NO:77, amino acid sequence 10.1 corresponds to SEQ ID NO:78, amino acid sequence 2.14 corresponds to SEQ ID NO:4, amino acid

sequence 1.3 corresponds to SEQ ID NO:1, amino acid sequences 6.16-7.26 correspond to SEQ ID NOs:81-119, amino acid sequence 2.15 corresponds to SEQ ID NO:14, amino acid sequence 2.8 corresponds to SEQ ID NO:17, amino acid sequences 5.6-7.33 correspond to SEQ ID NOs:122-125, amino acid sequence 1.10 corresponds to SEQ ID NO:19, amino acid sequence 2.10 corresponds to SEQ ID NO:20, amino acid sequence 1.4 corresponds to SEQ ID NO:15, amino acid sequences 12.19-4.3 correspond to SEQ ID NOs:129-132, amino acid sequence 1.9 corresponds to SEQ ID NO:16, amino acid sequences 5.5 -3.3 correspond to SEQ ID NOs:134-144, amino acid sequence 2.13 corresponds to SEQ ID NO:18, and amino acid sequences 3.6-3.9 correspond to SEQ ID NOs:146-155.

Figure 9 shows the evolutionarily conserved agnathan VLRs. VLR amino acid sequences representing the Inshore hagfish (*Eptatretus burgeri*), Pacific hagfish (*E. stoutii*), Sea lamprey (*Petromyzon marinus*; GenBank accession AY577946), American brook lamprey (*Lampetra appendix*) and Northern brook lamprey (*Ichthyomyzon fossor*). Blue shade: 100% identity; yellow: 60-99%; green: 40-59%; red: hydrophobic tail region.

Figure 10 shows the genetic distance among Pacific hagfish VLR diversity regions (LRRNT to LRRCT). Proteins predicted from PCR amplified lymphocyte-like cDNA clones, or blood genomic PCR amplicons from five animals. Scale bars represent 5% amino acid divergence. A. VLR-A (N=139). B. VLR-B (N=70). Green: unstimulated; red: immunostimulated; blue: genomic mature VLR; asterisk – related sequences.

Figure 11 shows the hagfish VLR gene loci. Figure 11A shows the Pacific hagfish VLR-A. Figure 11B shows the Inshore hagfish VLR-A. Figure 11C shows the Pacific hagfish VLR-B. Figure 11D shows the Inshore hagfish VLR-B. Sequence of inserts from four BAC clones, with uncaptured gaps marked. Location of VLR germline genes and flanking cassettes, in reverse or forward orientation, is indicated in kilobases (graphics are out of scale). GenScan gene predictions indicated in blue: an unrelated LRR gene upstream from the Pacific hagfish germline VLR-A gene and two flanking transposase ORFs in the Inshore hagfish VLR-A and Pacific hagfish VLR-B loci.

Figure 12 shows the Agnathan VLR genes, transcripts and phylogeny. Figure 12A shows a schematic presentation of germline and mature VLR genes of Pacific hagfish and Sea lamprey. Colored bars indicate coding regions; size in nucleotides; positions of PCR primers (Table 5) used to amplify hagfish VLR are indicated by arrows and labeled F (forward) R (reverse). Figure 12B shows Pacific hagfish VLRs PCR amplified from lymphocyte-like transcripts (RT-PCR) or blood genomic DNA. Agarose gel image; molecular weight marker indicated on the left (kilobases); position of germline and mature VLR amplicons indicated on the right. Figure 12C shows the phylogenetic analysis of agnathan VLRs. Neighbor Joining tree of hagfish and lamprey VLR proteins (same sequences as in Fig. 9); bootstrap values are indicated. Scale bar represents 10% amino acid divergence. Figure 12D shows a model for the evolution of agnathan VLR.

Figure 13 shows the Genetic distance among Inshore hagfish VLR diversity regions (LRRNT to LRRCT). Proteins were predicted from leukocyte cDNA clones, or mature VLR amplicons from genomic DNA of three animals. Scale bars represent 5% amino acid divergence. A. VLR-A (N=66). B. VLR-B (N=18). Red: hagfish #7; green: #8; blue: genomic mature VLR from hagfish #9.

20

DETAILED DESCRIPTION

A lymphocentric search was initiated for primordial elements of the vertebrate immune system in the sea lamprey, *Petromyzon marinus*, a modern representative of the oldest vertebrates. An earlier analysis of transcripts expressed by lymphocyte-like cells from lamprey hematopoietic tissues identified several homologs of immune system molecules (Mayer et al., 2002a; Uinuk-Ool et al., 2002; Uinuk-Ool et al., 2003), but none of the cardinal Ig superfamily receptor elements employed by jawed vertebrates for specific adaptive immunity were identified. Reasoning that activated lymphoblasts present in the blood stream were more likely to express the genes involved in adaptive responses, the present study began with a survey of the transcriptome of blood lymphocytes from

30

immunostimulated lamprey larvae. This search revealed a novel type of highly variable lymphocyte receptors which are described here.

The present invention may be understood more readily by reference to the following detailed description of preferred embodiments of the invention and the
5 Examples included therein and to the Figures and their previous and following description.

Before the present compounds, compositions, articles, devices, and/or methods are disclosed and described, it is to be understood that this invention is not limited to specific synthetic methods, specific recombinant biotechnology methods unless
10 otherwise specified, or to particular reagents unless otherwise specified, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

As used in the specification and the appended claims, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise.
15 Thus, for example, reference to “a pharmaceutical carrier” includes mixtures of two or more such carriers, and the like.

Ranges may be expressed herein as from “about” one particular value, and/or to “about” another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly,
20 when values are expressed as approximations, by use of the antecedent “about,” it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint.

“Optional” or “optionally” means that the subsequently described event or
25 circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

As used herein, “polypeptide,” “protein,” and “peptide” are used interchangeably to refer to amino acid sequences.

The invention relates to a variable lymphocyte receptor (VLR), which is a polypeptide capable of somatic rearrangement, which comprises 1-12 leucine rich repeats and which can function in adaptive immunity.

The invention provides an isolated polypeptide comprising an N-terminal leucine rich repeat (LRRNT), one or more leucine rich repeats (LRRs) (referred to herein as the internal LRRs), a C-terminal leucine rich repeat (LRRCT), and a connecting peptide, wherein the connecting peptide comprises an alpha helix. The length of the polypeptide can comprise as few as about 130 amino acids or as many as about 225 amino acids. Examples of the general structure and specific sequences of the polypeptides and encoding nucleic acids are shown in Figures. Furthermore numerous examples of various regions (including the signal peptide, LRRNT, LRR, LRRCT, connecting peptide, stalk and hydrophobic tails) can be found in Figures.

Optionally the connecting peptide is located on the N-terminal side of the LRRCT, and more specifically located between the internal LRR and the LRRCT. The connecting peptide can be linked to an internal LRR and the LRRCT. Thus disclosed herein are polypeptides comprising a LRRNT, one or more internal LRRs, a connecting peptide, and a LRRCT, in that order. Also disclosed are polypeptides, wherein the internal LRR region between the LRRNT and the LRRCT comprises 1, 2, 3, 4, 5, 6, 7, 8, or 9 leucine rich repeats, with LRR 1 located adjacent to or close to the LRRNT. As used herein LRRs 1, 2, 3, 4, 5, 6, 7, 8, or 9 are considered to run from the LRRNT to the LRRCT consecutively. Thus disclosed herein are polypeptides comprising a LRRNT, 1, 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, or 1-9 LRRs, a connecting peptide, and a LRRCT, in that order.

Leucine rich repeats (LRRs) are short sequence motifs typically involved in protein to protein interactions, wherein the LRRs comprise multiple leucine residues. LRRs contain leucine or other aliphatic residues, for example, at positions 2, 5, 7, 12, 16, 21, and 24. However, it is understood and herein contemplated that the leucine or other aliphatic residues can occur at other positions in addition to or in the place of residues at positions 2, 5, 7, 12, 16, 21, and 24. For example, a leucine can occur at position 3 rather than position 2. It is also understood that structurally, the motifs form β -sheet structures.

Thus, for example, a disclosed polypeptide comprising a LRRNT, 5 LRR, a LRRCT, and a connecting peptide would comprise 7 β -sheet structures and the alpha helix of the connecting peptide.

It is understood that the length and sequence of each LRR can vary from the other
5 LRRs in the polypeptide as well as from the LRRNT and LRRCT. For example, one embodiment of the present invention are polypeptides comprising a LRRNT, 1-9 LRR, a connecting peptides, and a LRRCT, wherein the first internal LRR is LRR1, and wherein LRR1 comprises less than about 20 amino acids. Also disclosed are polypeptides, wherein LRR1 comprises about 18 amino acids. Optionally, the polypeptide further
10 comprises LRR2-9, wherein LRR2-9 are less than about 25 amino acids each. Also disclosed are polypeptides, wherein LRR2-9 comprise about 24 amino acids each. LRR 1-9 can be the same or different from each other in a given polypeptide both in length and in specific amino acid sequence.

The terminal LRRs, designated LRRNT and LRRCT, are typically longer than
15 each internal LRR. The LRRNT and LRRCT comprise invariant regions (regions that have little variation relative to the rest of the polypeptide as compared to similar variable lymphocyte receptors). The variable regions provide the receptors with specificity, but the invariant regions and general structural similarities across receptors help maintain the protective immunity functions. The polypeptide can comprise an LRRNT, wherein the
20 LRRNT comprises less than about 40 amino acids. Thus the LRRNT optionally comprises the amino acid sequence CPSQCSC (SEQ ID NO: 157), CPSRCSC (SEQ ID NO: 307), CPAQCSC (SEQ ID NO: 308), CPSQCLC (SEQ ID NO: 309), CPSQCPC (SEQ ID NO: 310), NGATCKK (SEQ ID NO: 311), or NEALCKK (SEQ ID NO: 312) in the presence or absence of one or more conservative amino acid substitutions.
25 Also disclosed are polypeptides comprising a LRRCT, wherein the LRRCT is less than about 60 amino acids, and optionally 40-60 amino acids in length. In particular, specifically disclosed are polypeptides, wherein the LRRCT comprises the amino acid sequence TNTPVRAVTEASTSPSKCP (SEQ ID NO:158), SGKPVRSIICP (SEQ ID NO: 313), SSKAVLDVTEEEAAEDCV (SEQ ID NO: 314), or
30 QSKAVLEITEKDAASDCV (SEQ ID NO: 315) in the presence or absence of

conservative amino acid substitutions.

As with all peptides, polypeptides, and proteins, it is understood that substitutions in the amino acid sequence of the LRRCT and LRRNT can occur that do not alter the nature or function of the peptides, polypeptides, or proteins. Such substitutions include
5 conservative amino acid substitutions and are discussed in greater detail below.

The disclosed compositions can also comprise a connecting peptide. Typically such peptides are short peptides less than 15 amino acids in length and comprise an alpha helix. Thus, for example, specifically disclosed are connecting peptides of 10, 11, 12, 13, 14, and 15 amino acids in length comprising an alpha helix. It is understood that the
10 connecting peptide serves to link structural components of the polypeptide. It is further understood that the connecting peptide of the polypeptide can be linked to the LRRCT.

The polypeptides of the invention can comprise soluble or membrane bound forms. Many mechanisms exist that allow a polypeptide to be soluble or membrane bound. For example, a polypeptide missing a transmembrane domain can be secreted
15 directly by a cell. Alternatively, a polypeptide can comprise a glycosyl-phosphatidyl-inositol (GPI) anchor which maintains the polypeptide on a membrane surface. Therefore, disclosed herein are polypeptides comprising a GPI anchor. Other mechanisms for maintaining a polypeptide bound to a surface are known in the art. For example, the polypeptide may be bound to a hydrophobic layer through single or multi-pass
20 transmembrane regions that form covalent interactions with the lipid bilayer of the membrane. Alternatively, the polypeptide may be bound to the surface through noncovalent interactions with surface proteins.

The polypeptides of the invention can be surface bound polypeptides. Trafficking to the cell surface can be conducted by means of a signal peptide which provides a
25 indicator to the intracellular transport machinery to deliver the polypeptide to the surface of a cell. Thus it is a further embodiment of the invention that the polypeptides of the invention comprise a signal peptide of the N-terminal of the polypeptide.

It is understood and herein contemplated that the polypeptides can comprise a hydrophobic tail.

The polypeptide can comprise a stalk region. The stalk region comprises a threonin-proline rich region and is optionally present in the membrane bound form of the polypeptide, along with the GPI anchor and the hydrophobic tail.

Examples of polypeptides of the invention include those comprising amino acid sequences of SEQ ID NOs: 1-43, 45-52, 54, 56, 60-65, 68-72, 75, 77-78, 81-119, 122-125, 129-132, 134-144, and 146-155. Sequences include GenBank Accession Numbers AY577941-AY578059 and CK988414-CK988652. Those sequences comprising the amino acid sequences of SEQ ID NOs: 1-20 represent examples of full length VLRs. The sequence comprising the amino acid sequence of SEQ ID NO: 43 is an example of a full length VLR with the signal peptide. Additional full length VLRs and fragments thereof comprising the amino acid sequences can be found in the figures. Based on the structure taught herein for the polypeptides of the invention, it will be understood that these sequences are examples of a genus of polypeptides. It is understood that the invention includes full length VLRs and fragments thereof.

Disclosed are the components to be used to prepare the disclosed compositions as well as the compositions themselves to be used within the methods disclosed herein.

These and other materials are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that while specific reference of each various individual and collective combinations and permutation of these compounds may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a particular polypeptide is disclosed and discussed and a number of modifications that can be made to a number of polypeptides are discussed, specifically contemplated is each and every combination and permutation of polypeptides and the modifications that are possible unless specifically indicated to the contrary. Thus, if a class of molecules A, B, and C are disclosed as well as a class of molecules D, E, and F and an example of a combination molecule, A-D is disclosed, then even if each is not individually recited each is individually and collectively contemplated meaning combinations, A-E, A-F, B-D, B-E, B-F, C-D, C-E, and C-F are considered disclosed. Likewise, any subset or combination of these is also disclosed. Thus, for example, the sub-group of A-E, B-F, and C-E

would be considered disclosed. This concept applies to all aspects of this application including, but not limited to, steps in methods of making and using the disclosed compositions. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific
5 embodiment or combination of embodiments of the disclosed methods.

The polypeptides of the invention have a desired function. The polypeptides as described herein selectively bind an antigen or an agent, much as an antibody selectively binds an antigen or agent. The polypeptides optionally are variable lymphocyte receptors (naturally occurring or non-naturally occurring) or fragments or variants thereof. The
10 term “variable lymphocyte receptors” is used herein in a broad sense and, like the term “antibody” includes various versions having various specificities. The polypeptides are tested for their desired activity using the *in vitro* assays described herein, or by analogous methods, after which their therapeutic, diagnostic or other purification activities are tested according to known testing methods.

The polypeptide of the invention can bind an extracellular agent (e.g., a pathogen) or antigen. Agents or antigens can include but are not limited to peptides, polypeptides, lipids, glycolipids, and proteins. Agents or antigens can originate from a variety of sources including but not limited to pathogenic organisms. The binding to an agent or antigen is understood to be selective. By “selectively binding” or “specifically binding” is
20 meant that it binds one agent or antigen to the partial or complete exclusion of other antigens or agents. By “binding” is meant a detectable binding at least about 1.5 times the background of the assay method. For selective or specific binding such a detectable binding can be detected for a given antigen or agent but not a control antigen or agent. Thus, disclosed are polypeptides that selectively bind, for example, a viral, bacterial,
25 fungal, or protozoan antigen or agent.

Thus specifically disclosed are polypeptides, wherein the polypeptide binds an agent, wherein the agent is a pathogenic agent. Also disclosed are polypeptides of the invention that selectively binds a pathogenic agent, wherein the pathogen is a virus. Many viruses are known to exist. Thus, the virus can be selected from the group of
30 viruses consisting of Herpes simplex virus type-1, Herpes simplex virus type-2,

Cytomegalovirus, Epstein-Barr virus, Varicella-zoster virus, Human herpesvirus 6, Human herpesvirus 7, Human herpesvirus 8, Variola virus, Vesicular stomatitis virus, Hepatitis A virus, Hepatitis B virus, Hepatitis C virus, Hepatitis D virus, Hepatitis E virus, Rhinovirus, Coronavirus, Influenza virus A, Influenza virus B, Measles virus, Polyomavirus, Human Papillomavirus, Respiratory syncytial virus, Adenovirus, Cocksackie virus, Dengue virus, Mumps virus, Poliovirus, Rabies virus, Rous sarcoma virus, Yellow fever virus, Ebola virus, Marburg virus, Lassa fever virus, Eastern Equine Encephalitis virus, Japanese Encephalitis virus, St. Louis Encephalitis virus, Murray Valley fever virus, West Nile virus, Rift Valley fever virus, Rotavirus A, Rotavirus B, Rotavirus C, Sindbis virus, Simian Immunodeficiency virus, Human T-cell Leukemia virus type-1, Hantavirus, Rubella virus, Simian Immunodeficiency virus, Human Immunodeficiency virus type-1, and Human Immunodeficiency virus type-2.

Also disclosed are polypeptides of the invention, wherein the pathogen is a bacterium. Many bacteria are known to exist. Specifically contemplated and herein disclosed are polypeptides that selectively bind a pathogen, wherein the pathogen is a bacterium selected from the list of bacteria consisting of *M. tuberculosis*, *M. bovis*, *M. bovis* strain BCG, BCG substrains, *M. avium*, *M. intracellulare*, *M. africanum*, *M. kansasii*, *M. marinum*, *M. ulcerans*, *M. avium* subspecies *paratuberculosis*, *Nocardia asteroides*, other *Nocardia* species, *Legionella pneumophila*, other *Legionella* species, *Salmonella typhi*, other *Salmonella* species, *Shigella* species, *Yersinia pestis*, *Pasteurella haemolytica*, *Pasteurella multocida*, other *Pasteurella* species, *Actinobacillus pleuropneumoniae*, *Listeria monocytogenes*, *Listeria ivanovii*, *Brucella abortus*, other *Brucella* species, *Cowdria ruminantium*, *Chlamydia pneumoniae*, *Chlamydia trachomatis*, *Chlamydia psittaci*, *Coxiella burnetii*, other Rickettsial species, *Ehrlichia* species, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Bacillus anthracis*, *Escherichia coli*, *Vibrio cholerae*, *Campylobacter* species, *Neisseria meningitidis*, *Neisseria gonorrhea*, *Pseudomonas aeruginosa*, other *Pseudomonas* species, *Haemophilus influenzae*, *Haemophilus ducreyi*, other *Haemophilus* species, *Clostridium tetani*, other *Clostridium* species, *Yersinia enterocolitica*, and other *Yersinia* species.

Also disclosed are polypeptides of the invention that selectively bind a pathogen, wherein the pathogen is a protozoan or other parasite. Many parasitic infections are known to exist. Specifically contemplated and herein disclosed are polypeptides that selectively bind a pathogen, wherein the pathogen is a parasitic
5 infection selected from the group consisting of *Toxoplasma gondii*, *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, other *Plasmodium* species., *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, other *Leishmania* species., *Schistosoma mansoni*, other *Schistosoma* species., and *Entamoeba histolytica*.

Also disclosed are polypeptides of the invention that selectively bind a
10 pathogen, wherein the pathogen is a fungus. Many fungi are known to exist. Specifically contemplated and herein disclosed are polypeptides, wherein the pathogen is a fungi selected from the group fungi consisting of *Candida albicans*, *Cryptococcus neoformans*, *Histoplasma capsulatum*, *Aspergillus fumigatus*, *Coccidioides immitis*, *Paracoccidioides brasiliensis*, *Blastomyces dermatidis*, *Pneumocystis carinii*, *Penicillium*
15 *marneffi*, and *Alternaria alternata*.

The polypeptide of can also selectively bind to toxins. Herein “toxins” refer to any chemical or biological agent that effectively destroys any cell that it (the toxin) contacts. Notable examples of toxins include ricin, pertussis toxin, sarin, bacterial endotoxin, toxic shock syndrome toxin 1, cholera toxin, and snake venom toxins. Thus, specifically
20 discloses are polypeptides that bind to a toxin.

The polypeptides described herein can be modified and varied so long as the desired function is maintained. It is understood that one way to define any known variants and derivatives or those that might arise, of the disclosed genes and proteins herein is through defining the variants and derivatives in terms of homology to specific
25 known sequences. For example SED ID NO: 1 sets forth a particular amino acid sequence of the polypeptide encoded by any number of nucleic acids of the invention. Specifically disclosed are variants of these and other genes and proteins herein disclosed which have at least, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 percent homology to the stated
30 sequence. Those of skill in the art readily understand how to determine the homology

of two proteins or nucleic acids, such as genes. For example, the homology can be calculated after aligning the two sequences so that the homology is at its highest level.

In general, it is understood that one way to define any known variants and derivatives or those that might arise, of the disclosed genes and proteins herein, is through defining the variants and derivatives in terms of homology to specific known sequences. This identity of particular sequences disclosed herein is also discussed elsewhere herein. In general, variants of genes and proteins herein disclosed typically have at least, about 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent homology to the stated sequence or the native sequence. Those of skill in the art readily understand how to determine the homology of two proteins or nucleic acids, such as genes. For example, the homology can be calculated after aligning the two sequences so that the homology is at its highest level.

Another way of calculating homology can be performed by published algorithms. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection.

The same types of homology can be obtained for nucleic acids by for example the algorithms disclosed in Zuker, M. *Science* 244:48-52, 1989, Jaeger et al. *Proc. Natl. Acad. Sci. USA* 86:7706-7710, 1989, Jaeger et al. *Methods Enzymol.* 183:281-306, 1989 which are herein incorporated by reference for at least material related to nucleic acid alignment. It is understood that any of the methods typically can be used and that in certain instances the results of these various methods may differ, but the skilled artisan understands if identity is found with at least one of these methods, the sequences would be said to have the stated identity, and be disclosed herein.

For example, as used herein, a sequence recited as having a particular percent homology to another sequence refers to sequences that have the recited homology as calculated by any one or more of the calculation methods described above. For example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using the Zuker calculation method even if the first sequence does not have 80 percent homology to the second sequence as calculated by any of the other calculation methods. As another example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using both the Zuker calculation method and the Pearson and Lipman calculation method even if the first sequence does not have 80 percent homology to the second sequence as calculated by the Smith and Waterman calculation method, the Needleman and Wunsch calculation method, the Jaeger calculation methods, or any of the other calculation methods. As yet another example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using each of calculation methods (although, in practice, the different calculation methods will often result in different calculated homology percentages).

Protein variants and derivatives are well understood to those of skill in the art and in can involve amino acid sequence modifications. For example, amino acid sequence modifications typically fall into one or more of three classes: substitutional, insertional or deletional variants. Insertions include amino and/or carboxyl terminal fusions as well as intrasequence insertions of single or multiple amino acid residues. Insertions ordinarily will be smaller insertions than those of amino or carboxyl terminal fusions, for example, on the order of one to four residues. Immunogenic fusion protein derivatives, such as those described in the examples, are made by fusing a polypeptide sufficiently large to confer immunogenicity to the target sequence by cross-linking in vitro or by recombinant cell culture transformed with DNA encoding the fusion. Deletions are characterized by the removal of one or more amino acid residues from the protein sequence. Typically, no more than about from 2 to 6 residues are deleted at any

one site within the protein molecule. These variants ordinarily are prepared by site specific mutagenesis of nucleotides in the DNA encoding the protein, thereby producing DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture. Techniques for making substitution mutations at
5 predetermined sites in DNA having a known sequence are well known, for example M13 primer mutagenesis and PCR mutagenesis. Amino acid substitutions are typically of single residues, but can occur at a number of different locations at once; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. Deletions or insertions preferably are made in
10 adjacent pairs, i.e. a deletion of 2 residues or insertion of 2 residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final construct. The mutations must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. Substitutional variants are those in which at least one residue has been
15 removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Tables 1 and 2 and are referred to as conservative substitutions.

TABLE 1:Amino Acid Abbreviations

Amino Acid	Abbreviations
alanine	Ala A
allosoleucine	Alle
arginine	Arg R
asparagine	Asn N
aspartic acid	Asp D
cysteine	Cys C
glutamic acid	Glu E
glutamine	Gln Q
glycine	Gly G
histidine	His H
isoleucine	Ile I
leucine	Leu L
lysine	Lys K
phenylalanine	Phe F
proline	Pro P
pyroglutamic acidp	pGlu
serine	Ser S
threonine	Thr T
tyrosine	Tyr Y
tryptophan	Trp W
valine	Val V

TABLE 2:Amino Acid Substitutions	
Original Residue	Exemplary Conservative Substitutions, others are known in the art.
	Ala; Ser
	Arg;Lys; Gln
	Asn; Gln; His
	Asp; Glu
	Cys; Ser
	Gln; Asn, Lys
	Glu; Asp
	Gly; Pro
	His; Asn; Gln
	Ile; Leu; Val
	Leu; Ile; Val
	Lys; Arg; Gln;
	Met; Leu; Ile
	Phe; Met; Leu; Tyr
	Ser; Thr
	Thr; Ser
	Trp; Tyr
	Tyr; Trp; Phe
	Val; Ile; Leu

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those in Table 2, i.e., selecting residues that

5 differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the protein properties will be those in which (a) a hydrophilic

10 residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is

15 substituted for (or by) one not having a side chain, e.g., glycine, in this case, (e) by increasing the number of sites for sulfation and/or glycosylation.

For example, the replacement of one amino acid residue with another that is biologically and/or chemically similar is known to those skilled in the art as a conservative substitution. For example, a conservative substitution would be replacing one hydrophobic residue for another, or one polar residue for another. The substitutions
5 include combinations such as, for example, Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Such conservatively substituted variations of each explicitly disclosed sequence are included within the mosaic polypeptides provided herein.

Substitutional or deletional mutagenesis can be employed to insert sites for N-glycosylation (Asn-X-Thr/Ser) or O-glycosylation (Ser or Thr). Deletions of cysteine or
10 other labile residues also may be desirable. Deletions or substitutions of potential proteolysis sites, e.g. Arg, is accomplished for example by deleting one of the basic residues or substituting one by glutaminy or histidyl residues.

Certain post-translational derivatizations are the result of the action of recombinant host cells on the expressed polypeptide. Glutaminy and asparaginy
15 residues are frequently post-translationally deamidated to the corresponding glutamyl and asparyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Other post-translational modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of
20 the o-amino groups of lysine, arginine, and histidine side chains (T.E. Creighton, Proteins: Structure and Molecular Properties, W. H. Freeman & Co., San Francisco pp 79-86 [1983]), acetylation of the N-terminal amine and, in some instances, amidation of the C-terminal carboxyl.

As used herein, the term “variable lymphocyte receptor” or “variable
25 lymphocyte receptors” can also refer to polypeptides that have been modified to have reduced immunogenicity when administered to a subject. For example, human amino acid sequences may be inserted within or added to the polypeptide to make a version less immunogenic to a human subject, much like antibodies are humanized. Many non-human variable lymphocyte receptors (e.g., those derived from lampreys, mice, rats, or
30 rabbits) can be naturally antigenic in humans, and thus can give rise to undesirable

immune responses when administered to humans. Therefore, the use of modified polypeptides in the methods of the invention can serve to lessen the chance that a polypeptide administered to a human will evoke an undesirable immune response.

Modification techniques can involve the use of recombinant DNA technology to
5 manipulate the DNA sequence encoding one or more polypeptide regions of the variable lymphocyte receptor molecule. Accordingly, the humanized form of the variable lymphocyte receptor (or a fragment thereof) is a chimeric variable lymphocyte receptor, preferably the antigen (agent)-binding portion of the variable lymphocyte
10 (donor) variable lymphocyte receptor integrated into human (recipient) amino acid sequence.

It is understood that the nucleic acids that can encode those protein sequences, variants and fragments thereof are also disclosed. This would include all degenerate sequences related to a specific protein sequence, i.e. all nucleic acids having a sequence
15 that encodes one particular protein sequence as well as all nucleic acids, including degenerate nucleic acids, encoding the disclosed variants and derivatives of the protein sequences. Thus, while each particular nucleic acid sequence may not be written out herein, it is understood that each and every sequence is in fact disclosed and described herein through the disclosed protein sequence.

20 Humanized variable lymphocyte receptors can also contain amino acid sequences which are found neither in the recipient variable lymphocyte receptor nor in the imported human sequences.

The polypeptides of the invention can also used to make fusion proteins. The polypeptides can serve a targeting function in the fusion protein. Thus the polypeptide of
25 the invention can be conjugated to or otherwise linked by recombinant engineering to a second moiety. The second moiety can comprise a toxin, for example, if cell killing is desired. Thus, for example, the polypeptide that selectively binds a protozoan can target the protozoan and the toxin moiety of the fusion protein can kill the cell. Similarly, the polypeptide of the invention can perform a delivery function. Thus the second moiety can
30 be a therapeutic agent.

The polypeptide of the invention can be linked to a detectable tag. A “detectable tag” is any tag that can be visualized with imaging or detection methods, in vivo or in vitro. The detectable tag can be a radio-opaque substance, radiolabel, a chemoluminescent label, a fluorescent label, or a magnetic label. The detectable tag can be selected from the group consisting of gamma-emitters, beta-emitters, and alpha-emitters, gamma-emitters, positron-emitters, X-ray-emitters and fluorescence-emitters. Suitable fluorescent compounds include fluorescein sodium, fluorescein isothiocyanate, phycoerythrin, and Texas Red sulfonyl chloride, Allophycocyanin (APC), Cy5-PE, CY7-APC, and Cascade yellow.

Suitable radioisotopes for labeling include Iodine-131, Iodine-123, Iodine-125, Iodine-126, Iodine-133, Bromine-77, Indium-111, Indium-113m, Gallium-67, Gallium-68, Ruthenium-95, Ruthenium-97, Ruthenium-103, Ruthenium-105, Mercury-107, Mercury-203, Rhenium-99m, Rhenium-105, Rhenium-101, Tellurium-121m, Tellurium-122m, Tellurium-125m, Thulium-165, Thulium-167, Thulium-168, Technetium-99m and Fluorine-18.

Optionally the detectable tag can be visualized using histochemical techniques, ELISA-like assays, confocal microscopy, fluorescent detection, cell sorting methods, nuclear magnetic resonance, radioimmunosciintigraphy, X-radiography, positron emission tomography, computerized axial tomography, magnetic resonance imaging, and ultrasonography.

Alternatively, the polypeptide can be biotinylated and a subsequent detectable label like a fluorescently labeled strepavidin can be used to indirectly detect the polypeptide. Biotin is detected by any one of several techniques known in the art. For example, the biotin is detectable by binding with a fluorescence-labeled avidin and the avidin is labeled with a phycoerythrin or a catenated fluorescent label to increase the signal associate with each binding event.

Optionally the polypeptide is bound to a solid support such as a slide, a culture dish, a multiwell plate, column, chip, array or stable beads. An “array” includes one or more multiwell arraying means such as microplates or slides.

Optionally the polypeptide is bound to a mobile solid support, e.g., beads, which can be sorted using cell sorting technology. "Mobile solid support" refers to a set of distinguishably labeled microspheres or beads. Preferably, the microspheres are polystyrene-divinylbenzene beads. Sets of microspheres marked with specific fluorescent dyes and having specific fluorescent profiles can be obtained commercially, for example, from Luminex Corporation (Austin, TX).

The invention also provides a plurality of polypeptides of the invention. Optionally the LRRs of the polypeptides are highly variable across polypeptides. Thus, the plurality can include polypeptides with different binding specificities, based on the variability of the internal LRRs.

Also provided are kits that include a container with polypeptides of the invention or a stable or mobile solid support with polypeptides of the invention. Optionally the polypeptides are bound to the solid support or the kit. Optionally the kit contains the polypeptides the solid support, and a linking means for binding the polypeptide to the solid support.

The invention provides isolated nucleic acids that encode the polypeptides of the invention. One example of such a nucleic acid comprises the nucleotide sequence of SEQ ID NO:156, the ORF of a representative *VLR*. Other examples of nucleic acids that encode VLRs or fragments thereof include SEQ ID NO:44, SEQ ID NO:53-55, SEQ ID NO:57-59, SEQ ID NO:66-67, SEQ ID NO:73-74, SEQ ID NO:76, SEQ ID NOs:79-80, and SEQ ID NOs:172-302. There are a variety of sequences related to the VLR gene having Genbank Accession Numbers AY57791-AY578059, AY964719-AY964931, AY965520-AY965612, AY965658-AY965681, and CK988414-CK988652. These sequences are herein incorporated by reference in their entireties as well as for individual subsequences (regions or fragments) contained therein.

Such nucleic acid sequences are provided by way of example of the genus of nucleic acids and are not intended to be limiting. Also provided are expression vectors comprising these nucleic acids, wherein the nucleic acids are operably linked to an expression control sequence. Further provided are cultured cells comprising the

expression vectors. Such expression vectors and cultured cells can be used to make the polypeptides of the invention.

There are a variety of molecules disclosed herein that are nucleic acid based, including for example the nucleic acids that encode, for example VLR or fragments or variants thereof. The disclosed nucleic acids are made up of nucleotides, nucleotide
5 analogs, or nucleotide substitutes.

A nucleotide analog is a nucleotide which contains some type of modification to either the base, sugar, or phosphate moieties. Modifications to the base moiety would include natural and synthetic modifications of A, C, G, and T/U as well as different
10 purine or pyrimidine bases, such as uracil-5-yl (.psi.), hypoxanthin-9-yl (I), and 2-aminoadenin-9-yl. A modified base includes but is not limited to 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine,
15 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine
20 and 3-deazaguanine and 3-deazaadenine. Additional base modifications can be found for example in U.S. Pat. No. 3,687,808, Englisch et al., *Angewandte Chemie*, International Edition, 1991, 30, 613, and Sanghvi, Y. S., Chapter 15, *Antisense Research and Applications*, pages 289-302, Crooke, S. T. and Lebleu, B. ed., CRC Press, 1993. Certain nucleotide analogs, such as 5-substituted pyrimidines,
25 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine can increase the stability of duplex formation. Often time base modifications can be combined with for example a sugar modification, such as 2'-O-methoxyethyl, to achieve unique properties such as increased duplex stability. There are numerous United States
30 patents such as 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272;

5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121, 5,596,091; 5,614,617; and 5,681,941, which detail and describe a range of base modifications. Each of these patents is herein incorporated by reference.

Nucleotide analogs can also include modifications of the sugar moiety.

5 Modifications to the sugar moiety would include natural modifications of the ribose and deoxy ribose as well as synthetic modifications. Sugar modifications include but are not limited to the following modifications at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C₁ to C₁₀, alkyl or C₂ to C₁₀ alkenyl and
 10 alkynyl. 2' sugar modifications also include but are not limited to -O[(CH₂)_n O]_m CH₃, -O(CH₂)_n OCH₃, -O(CH₂)_n NH₂, -O(CH₂)_n CH₃, -O(CH₂)_n -ONH₂, and -O(CH₂)_nON[(CH₂)_n CH₃]₂, where n and m are from 1 to about 10.

Other modifications at the 2' position include but are not limited to: C₁ to C₁₀ lower alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH₃,
 15 OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂ CH₃, ONO₂, NO₂, N₃, NH₂, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having
 20 similar properties. Similar modifications may also be made at other positions on the sugar, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Modified sugars would also include those that contain modifications at the bridging ring oxygen, such as CH₂ and S. Nucleotide sugar analogs may also have sugar mimetics such as cyclobutyl
 25 moieties in place of the pentofuranosyl sugar. There are numerous United States patents that teach the preparation of such modified sugar structures such as 4,981,957; 5,118,800; 5,319,080; 5,359,044; 5,393,878; 5,446,137; 5,466,786; 5,514,785; 5,519,134; 5,567,811; 5,576,427; 5,591,722; 5,597,909; 5,610,300; 5,627,053; 5,639,873; 5,646,265; 5,658,873; 5,670,633; and 5,700,920, each of which is herein
 30 incorporated by reference in its entirety.

Nucleotide analogs can also be modified at the phosphate moiety. Modified phosphate moieties include but are not limited to those that can be modified so that the linkage between two nucleotides contains a phosphorothioate, chiral phosphorothioate, phosphorodithioate, phosphotriester, aminoalkylphosphotriester, methyl and other alkyl phosphonates including 3'-alkylene phosphonate and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates. It is understood that these phosphate or modified phosphate linkage between two nucleotides can be through a 3'-5' linkage or a 2'-5' linkage, and the linkage can contain inverted polarity such as 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms are also included. Numerous United States patents teach how to make and use nucleotides containing modified phosphates and include but are not limited to, 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, each of which is herein incorporated by reference.

It is understood that nucleotide analogs need only contain a single modification, but may also contain multiple modifications within one of the moieties or between different moieties.

Nucleotide substitutes are molecules having similar functional properties to nucleotides, but which do not contain a phosphate moiety, such as peptide nucleic acid (PNA). Nucleotide substitutes are molecules that will recognize nucleic acids in a Watson-Crick or Hoogsteen manner, but which are linked together through a moiety other than a phosphate moiety. Nucleotide substitutes are able to conform to a double helix type structure when interacting with the appropriate target nucleic acid.

Nucleotide substitutes are nucleotides or nucleotide analogs that have had the phosphate moiety and/or sugar moieties replaced. Nucleotide substitutes do not contain a standard phosphorus atom. Substitutes for the phosphate can be for example, short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or

cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene
5 formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Numerous United States patents disclose how to make and use these types of phosphate replacements and include but are not limited to 5,034,506;
10 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439, each of which is herein incorporated by reference.

15 It is also understood in a nucleotide substitute that both the sugar and the phosphate moieties of the nucleotide can be replaced, by for example an amide type linkage (aminoethylglycine) (PNA). United States patents 5,539,082; 5,714,331; and 5,719,262 teach how to make and use PNA molecules, each of which is herein incorporated by reference. (See also Nielsen et al., *Science*, 1991, 254, 1497-1500).

20 It is also possible to link other types of molecules (conjugates) to nucleotides or nucleotide analogs to enhance for example, cellular uptake. Conjugates can be chemically linked to the nucleotide or nucleotide analogs. Such conjugates include but are not limited to lipid moieties such as a cholesterol moiety (Letsinger et al., *Proc. Natl. Acad. Sci. USA*, 1989, 86, 6553-6556), cholic acid (Manoharan et al., *Bioorg. Med. Chem. Lett.*, 1994, 4, 1053-1060), a thioether, e.g., hexyl-S-tritylthiol (Manoharan et al., *Ann. N.Y. Acad. Sci.*, 1992, 660, 306-309; Manoharan et al., *Bioorg. Med. Chem. Lett.*, 1993, 3, 2765-2770), a thiocholesterol (Oberhauser et al., *Nucl. Acids Res.*, 1992, 20, 533-538), an aliphatic chain, e.g., dodecandiol or undecyl residues (Saison-Behmoaras et al., *EMBO J.*, 1991, 10, 1111-1118; Kabanov et al., *FEBS Lett.*,
25 1990, 259, 327-330; Svinarchuk et al., *Biochimie*, 1993, 75, 49-54), a phospholipid,
30

e.g., di-hexadecyl-rac-glycerol or triethylammonium

1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654; Shea et al., *Nucl. Acids Res.*, 1990, 18, 3777-3783), a polyamine or a polyethylene glycol chain (Manoharan et al., *Nucleosides & Nucleotides*, 1995, 14, 969-973), or adamantane acetic acid (Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654), a palmityl moiety (Mishra et al., *Biochem. Biophys. Acta*, 1995, 1264, 229-237), or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety (Crooke et al., *J. Pharmacol. Exp. Ther.*, 1996, 277, 923-937. Numerous United States patents teach the preparation of such conjugates and include, but are not limited to U.S. Pat. Nos. 4,828,979; 4,948,882; 5,218,105; 5,525,465; 5,541,313; 5,545,730; 5,552,538; 5,578,717; 5,580,731; 5,580,731; 5,591,584; 5,109,124; 5,118,802; 5,138,045; 5,414,077; 5,486,603; 5,512,439; 5,578,718; 5,608,046; 4,587,044; 4,605,735; 4,667,025; 4,762,779; 4,789,737; 4,824,941; 4,835,263; 4,876,335; 4,904,582; 4,958,013; 5,082,830; 5,112,963; 5,214,136; 5,082,830; 5,112,963; 5,214,136; 5,245,022; 5,254,469; 5,258,506; 5,262,536; 5,272,250; 5,292,873; 5,317,098; 5,371,241; 5,391,723; 5,416,203; 5,451,463; 5,510,475; 5,512,667; 5,514,785; 5,565,552; 5,567,810; 5,574,142; 5,585,481; 5,587,371; 5,595,726; 5,597,696; 5,599,923; 5,599,928 and 5,688,941, each of which is herein incorporated by reference.

Disclosed are compositions including primers and probes, which are capable of interacting with the VLR gene, or comparable genes. In certain embodiments the primers are used to support DNA amplification reactions. Typically the primers will be capable of being extended in a sequence specific manner. Extension of a primer in a sequence specific manner includes any methods wherein the sequence and/or composition of the nucleic acid molecule to which the primer is hybridized or otherwise associated directs or influences the composition or sequence of the product produced by the extension of the primer. Extension of the primer in a sequence specific manner therefore includes, but is not limited to, PCR, DNA sequencing, DNA extension, DNA polymerization, RNA transcription, or reverse transcription. Techniques and conditions that amplify the primer in a sequence specific manner are preferred. In certain embodiments the primers are used for the DNA amplification reactions, such as PCR or

direct sequencing. It is understood that in certain embodiments the primers can also be extended using non-enzymatic techniques, where for example, the nucleotides or oligonucleotides used to extend the primer are modified such that they will chemically react to extend the primer in a sequence specific manner.

5 The size of the primers or probes for interaction with the VLR gene in certain embodiments can be any size that supports the desired enzymatic manipulation of the primer, such as DNA amplification or the simple hybridization of the probe or primer. A typical VLR primer or probe would be at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 10 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 15 nucleotides long.

 The polypeptides and nucleic acids of the invention can be used in a variety of techniques. For example, the polypeptides can be used to detect a selected agent, to block the activity of a selected agent, to purify an agent, as an imaging tool, and as a therapeutic agent.

20 Provided herein are methods of detecting an agent in a sample, comprising the steps of contacting the sample with the polypeptide, under conditions in which the polypeptide can bind to the agent in the sample, and detecting the polypeptide bound to the agent in the sample. The bound polypeptide indicates the agent in the sample. Detection methods are well known in the art. For example, the polypeptide can be 25 labeled with a detectable tag as described above. The detection method can be used to note the presence or absence of an agent in the sample. The detection method, however, can be further combined with quantification methods. In vitro assay methods include colorimetric assays such as ELISA that allow the quantification of the agent based on a comparison to a control sample or samples of known agent quantity which 30 can be used to establish an amount relative to a standard. The methods can also include

radiometric assays that allow for quantification based on emitted radiation and fluorescent assays or any means of visualization and quantification described above.

The sample can be any sample to be tested including any biologic sample. Samples can include fluid samples (like water, blood, urine, etc.), tissue samples,
5 culture samples, cellular samples, etc.

The polypeptides of the invention may also be used to block the activity of any agent to which it binds, comparable to a blocking antibody. Thus also disclosed are methods of blocking the activity of an agent, comprising contacting the agent with the polypeptide of the invention under conditions for the polypeptide to bind the agent.
10 The binding of the polypeptide to the agent blocks the activity of the agent. The contacting step can be in vivo or in vitro. Thus, for example, to reduce contamination of a sample, a polypeptide that binds a toxin can be added to the sample and block the toxin activity.

The polypeptides of the invention may also be used to promote the activity of an agent to which it binds, comparable to an agonistic antibody. Thus also disclosed are
15 methods of promoting the activity of an agent, comprising contacting the agent with the polypeptide of the invention under conditions for the polypeptide to bind the agent. The binding of the polypeptide to the agent promotes the activity of the agent.

The polypeptides disclosed herein can be used to determine the function of a gene with unknown function. Thus, disclosed herein are methods of using the disclosed
20 polypeptides in protein knock-down assays. For example, the disclosed polypeptides can be expressed in the cytoplasm of a cell which comprises a gene of unknown function. When the RNA transcript is being translated in the cytoplasm of the cell, the disclosed polypeptides can bind the protein product of the gene question. By
25 monitoring the effect the loss of protein expression has on the cell, the proteins function can be determined. Thus, specifically disclosed are polypeptides specific for a gene product of unknown function. Also are methods of determining the function of a gene comprising introducing a polypeptide specific for the protein product of the gene into the cytoplasm of a cell expressing the gene and monitoring the effect due to the loss of
30 protein product of the gene with unknown function.

The polypeptides of the invention can also be used in imaging methods. For example, the invention provides an imaging method comprising administering to a subject an effective amount of the polypeptide and detecting the localization of the bound polypeptide in the subject. Examples of imaging methods are described above.

5 The invention also provides methods of purification. Disclosed herein are methods of purifying an agent from a sample comprising contacting the sample with a polypeptide under conditions for the polypeptide to bind the agent and form a polypeptide/agent complex; and isolating the agent from the polypeptide/agent complex. For example, the polypeptide can be bound to a column and the sample can
10 be passed through the column under conditions that allow the agent in the sample to bind to the bound polypeptide. The agent can subsequently be eluted from the column in a desired eluant. The purification methods would be useful as research methods and as commercial methods. For example, such a method would be useful in removing contaminants from pharmacological compounds.

15 The polypeptides can also be used in therapeutic methods. For example, provided herein is a method of reducing or preventing a pathogenic effect in a subject comprising administering to the subject an effective amount of a polypeptide that binds a pathogen. Also provided is a method of blocking or promoting the activity of an agent so as to reduce deleterious effects or promote positive effects.

20 Provided herein are composition comprising the polypeptides or nucleic acids of the invention and a pharmaceutically acceptable carrier. The compositions of the invention can also be administered *in vivo*. The compositions may be administered orally, parenterally (e.g., intravenously), by intramuscular injection, by intraperitoneal injection, transdermally, extracorporeally, topically or the like, although topical
25 intranasal administration or administration by inhalant is typically preferred. As used herein, "topical intranasal administration" means delivery of the compositions into the nose and nasal passages through one or both of the nares and can comprise delivery by a spraying mechanism or droplet mechanism, or through aerosolization of the nucleic acid or vector. The latter may be effective when a large number of animals is to be
30 treated simultaneously. Administration of the compositions by inhalant can be through

the nose or mouth via delivery by a spraying or droplet mechanism. Delivery can also be directly to any area of the respiratory system (e.g., lungs) via intubation. The exact amount of the compositions required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the allergic disorder being treated, the particular nucleic acid or vector used, its mode of
5 administration and the like. Thus, it is not possible to specify an exact amount for every composition. However, an appropriate amount can be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein.

Parenteral administration of the composition, if used, is generally characterized
10 by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant dosage is maintained. See, e.g., U.S. Patent No. 3,610,795, which is
15 incorporated by reference herein.

The materials may be in solution, suspension (for example, incorporated into microparticles, liposomes, or cells). These may be targeted to a particular cell type via antibodies, receptors, or receptor ligands. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Senter, et al.,
20 *Bioconjugate Chem.*, 2:447-451, (1991); Bagshawe, K.D., *Br. J. Cancer*, 60:275-281, (1989); Bagshawe, et al., *Br. J. Cancer*, 58:700-703, (1988); Senter, et al., *Bioconjugate Chem.*, 4:3-9, (1993); Battelli, et al., *Cancer Immunol. Immunother.*, 35:421-425, (1992); Pietersz and McKenzie, *Immunolog. Reviews*, 129:57-80, (1992); and Roffler, et al., *Biochem. Pharmacol.*, 42:2062-2065, (1991)). Vehicles such as "stealth" and
25 other antibody conjugated liposomes (including lipid mediated drug targeting to colonic carcinoma), receptor mediated targeting of DNA through cell specific ligands, lymphocyte directed tumor targeting, and highly specific therapeutic retroviral targeting of murine glioma cells *in vivo*. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Hughes et al., *Cancer Research*,
30 49:6214-6220, (1989); and Litzinger and Huang, *Biochimica et Biophysica Acta*,

1104:179-187, (1992)). In general, receptors are involved in pathways of endocytosis, either constitutive or ligand induced. These receptors cluster in clathrin-coated pits, enter the cell via clathrin-coated vesicles, pass through an acidified endosome in which the receptors are sorted, and then either recycle to the cell surface, become stored
5 intracellularly, or are degraded in lysosomes. The internalization pathways serve a variety of functions, such as nutrient uptake, removal of activated proteins, clearance of macromolecules, opportunistic entry of viruses and toxins, dissociation and degradation of ligand, and receptor-level regulation. Many receptors follow more than one intracellular pathway, depending on the cell type, receptor concentration, type of ligand,
10 ligand valency, and ligand concentration. Molecular and cellular mechanisms of receptor-mediated endocytosis has been reviewed (Brown and Greene, *DNA and Cell Biology* 10:6, 399-409 (1991)).

By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to a subject, along with
15 the polypeptide of the invention, without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained. The carrier would naturally be selected to minimize any degradation of the active ingredient and to minimize any adverse side effects in the subject, as would be well known to one of skill in the art.
20 Suitable carriers and their formulations are described in *Remington: The Science and Practice of Pharmacy* (19th ed.) ed. A.R. Gennaro, Mack Publishing Company, Easton, PA 1995. Typically, an appropriate amount of a pharmaceutically-acceptable salt is used in the formulation to render the formulation isotonic. Examples of the pharmaceutically-acceptable carrier include, but are not limited to, saline, Ringer's
25 solution and dextrose solution. The pH of the solution is preferably from about 5 to about 8, and more preferably from about 7 to about 7.5. Further carriers include sustained release preparations such as semipermeable matrices of solid hydrophobic polymers containing the variable lymphocyte receptor, which matrices are in the form of shaped articles, e.g., films, liposomes or microparticles. It will be apparent to those
30 persons skilled in the art that certain carriers may be more preferable depending upon,

for instance, the route of administration and concentration of variable lymphocyte receptor being administered.

Pharmaceutical carriers are known to those skilled in the art. These most typically would be standard carriers for administration of drugs to humans, including
5 solutions such as sterile water, saline, and buffered solutions at physiological pH. The compositions can be administered intramuscularly or subcutaneously, for example. Other compounds will be administered according to standard procedures used by those skilled in the art.

Pharmaceutical compositions may include carriers, thickeners, diluents, buffers,
10 preservatives, surface active agents and the like in addition to the molecule of choice. Pharmaceutical compositions may also include one or more active ingredients such as antimicrobial agents, anti-inflammatory agents, anesthetics, and the like.

Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are
15 propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient
20 replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

Formulations for topical administration may include ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical
25 carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets, or tablets. Thickeners, flavorings, diluents, emulsifiers, dispersing aids or binders may be desirable.

Some of the compositions may potentially be administered as a pharmaceutically acceptable acid- or base- addition salt, formed by reaction with inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, and phosphoric acid, and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, malonic acid, succinic acid, maleic acid, and fumaric acid, or by reaction with an inorganic base such as sodium hydroxide, ammonium hydroxide, potassium hydroxide, and organic bases such as mono-, di-, trialkyl and aryl amines and substituted ethanolamines.

The dosage ranges for the administration of the compositions are those large enough to produce the desired effect in which the symptoms disorder are effected. The dosage should not be so large as to cause adverse side effects, such as unwanted cross-reactions, anaphylactic reactions, and the like. Generally, the dosage will vary with the age, condition, sex and extent of the disease in the patient and can be determined by one of skill in the art. The dosage can be adjusted by the individual physician in the event of any contraindications. Dosage can vary, and can be administered in one or more dose administrations daily, for one or several days.

The variable lymphocyte receptors and variable lymphocyte receptor fragments and variants of the invention can also be administered to patients or subjects as a nucleic acid preparation (e.g., DNA or RNA) that encodes the variable lymphocyte receptor or variable lymphocyte receptor fragment or variant, such that the patient's or subject's own cells take up the nucleic acid and produce and secrete the encoded variable lymphocyte receptor or variable lymphocyte receptor fragment.

There are a number of compositions and methods which can be used to deliver nucleic acids to cells, either in vitro or in vivo. These methods and compositions can largely be broken down into two classes: viral based delivery systems and non-viral based delivery systems. For example, the nucleic acids can be delivered through a number of direct delivery systems such as, electroporation, lipofection, calcium phosphate precipitation, plasmids, viral vectors, viral nucleic acids, phage nucleic acids, phages, cosmids, or via transfer of genetic material in cells or carriers such as cationic

liposomes. Appropriate means for transfection, including viral vectors, chemical transfectants, or physico-mechanical methods such as electroporation and direct diffusion of DNA, are described by, for example, Wolff, J. A., et al., *Science*, 247, 1465-1468, (1990); and Wolff, J. A. *Nature*, 352, 815-818, (1991). Such methods are well known in the art and readily adaptable for use with the compositions and methods described herein. In certain cases, the methods will be modified to specifically function with large DNA molecules. Further, these methods can be used to target certain diseases and cell populations by using the targeting characteristics of the carrier.

Transfer vectors can be any nucleotide construction used to deliver nucleic acids into cells (e.g., a plasmid), or as part of a general strategy to deliver genes, e.g., as part of recombinant retrovirus or adenovirus (Ram et al. *Cancer Res.* 53:83-88, (1993)). As used herein, plasmid or viral vectors are agents that transport the disclosed nucleic acids, such as VLR into the cell without degradation and include a promoter yielding expression of the gene in the cells into which it is delivered. Viral vectors are, for example, Adenovirus, Adeno-associated virus, Herpes virus, Vaccinia virus, Polio virus, AIDS virus, neuronal trophic virus, Sindbis and other RNA viruses, including these viruses with the HIV backbone. Also preferred are any viral families which share the properties of these viruses which make them suitable for use as vectors.

Retroviruses include Murine Maloney Leukemia virus, MMLV, and retroviruses that express the desirable properties of MMLV as a vector. Retroviral vectors are able to carry a larger genetic payload, i.e., a transgene or marker gene, than other viral vectors, and for this reason are a commonly used vector. However, they are not as useful in non-proliferating cells. Adenovirus vectors are relatively stable and easy to work with, have high titers, and can be delivered in aerosol formulation, and can transfect non-dividing cells. Pox viral vectors are large and have several sites for inserting genes, they are thermostable and can be stored at room temperature. A preferred embodiment is a viral vector which has been engineered so as to suppress the immune response of the host organism, elicited by the viral antigens. Preferred vectors of this type will carry coding regions for Interleukin 8 or 10.

Viral vectors can have higher transaction (ability to introduce genes) abilities than chemical or physical methods to introduce genes into cells. Typically, viral vectors contain, nonstructural early genes, structural late genes, an RNA polymerase III transcript, inverted terminal repeats necessary for replication and encapsidation, and promoters to control the transcription and replication of the viral genome. When engineered as vectors, viruses typically have one or more of the early genes removed and a gene or gene/promotor cassette is inserted into the viral genome in place of the removed viral DNA. Constructs of this type can carry up to about 8 kb of foreign genetic material. The necessary functions of the removed early genes are typically supplied by cell lines which have been engineered to express the gene products of the early genes in trans.

A retrovirus is an animal virus belonging to the virus family of Retroviridae, including any types, subfamilies, genus, or tropisms. Retroviral vectors, in general, are described by Verma, I.M., Retroviral vectors for gene transfer. *In Microbiology*-1985, American Society for Microbiology, pp. 229-232, Washington, (1985), which is incorporated by reference herein. Examples of methods for using retroviral vectors for gene therapy are described in U.S. Patent Nos. 4,868,116 and 4,980,286; PCT applications WO 90/02806 and WO 89/07136; and Mulligan, (Science 260:926-932 (1993)); the teachings of which are incorporated herein by reference.

A retrovirus is essentially a package which has packed into it nucleic acid cargo. The nucleic acid cargo carries with it a packaging signal, which ensures that the replicated daughter molecules will be efficiently packaged within the package coat. In addition to the package signal, there are a number of molecules which are needed in cis, for the replication, and packaging of the replicated virus. Typically a retroviral genome, contains the gag, pol, and env genes which are involved in the making of the protein coat. It is the gag, pol, and env genes which are typically replaced by the foreign DNA that it is to be transferred to the target cell. Retrovirus vectors typically contain a packaging signal for incorporation into the package coat, a sequence which signals the start of the gag transcription unit, elements necessary for reverse transcription, including a primer binding site to bind the tRNA primer of reverse transcription,

terminal repeat sequences that guide the switch of RNA strands during DNA synthesis, a purine rich sequence 5' to the 3' LTR that serve as the priming site for the synthesis of the second strand of DNA synthesis, and specific sequences near the ends of the LTRs that enable the insertion of the DNA state of the retrovirus to insert into the host
5 genome. The removal of the gag, pol, and env genes allows for about 8 kb of foreign sequence to be inserted into the viral genome, become reverse transcribed, and upon replication be packaged into a new retroviral particle. This amount of nucleic acid is sufficient for the delivery of a one to many genes depending on the size of each transcript. It is preferable to include either positive or negative selectable markers
10 along with other genes in the insert.

Since the replication machinery and packaging proteins in most retroviral vectors have been removed (gag, pol, and env), the vectors are typically generated by placing them into a packaging cell line. A packaging cell line is a cell line which has been transfected or transformed with a retrovirus that contains the replication and
15 packaging machinery, but lacks any packaging signal. When the vector carrying the DNA of choice is transfected into these cell lines, the vector containing the gene of interest is replicated and packaged into new retroviral particles, by the machinery provided in cis by the helper cell. The genomes for the machinery are not packaged because they lack the necessary signals.

20 The construction of replication-defective adenoviruses has been described (Berkner et al., *J. Virology* 61:1213-1220 (1987); Massie et al., *Mol. Cell. Biol.* 6:2872-2883 (1986); Haj-Ahmad et al., *J. Virology* 57:267-274 (1986); Davidson et al., *J. Virology* 61:1226-1239 (1987); Zhang "Generation and identification of recombinant adenovirus by liposome-mediated transfection and PCR analysis" *BioTechniques*
25 15:868-872 (1993)). The benefit of the use of these viruses as vectors is that they are limited in the extent to which they can spread to other cell types, since they can replicate within an initial infected cell, but are unable to form new infectious viral particles. Recombinant adenoviruses have been shown to achieve high efficiency gene transfer after direct, in vivo delivery to airway epithelium, hepatocytes, vascular
30 endothelium, CNS parenchyma and a number of other tissue sites (Morsy, *J. Clin.*

Invest. 92:1580-1586 (1993); Kirshenbaum, *J. Clin. Invest.* 92:381-387 (1993); Roessler, *J. Clin. Invest.* 92:1085-1092 (1993); Moullier, *Nature Genetics* 4:154-159 (1993); La Salle, *Science* 259:988-990 (1993); Gomez-Foix, *J. Biol. Chem.* 267:25129-25134 (1992); Rich, *Human Gene Therapy* 4:461-476 (1993); Zabner, *Nature Genetics* 6:75-83 (1994); Guzman, *Circulation Research* 73:1201-1207 (1993); Bout, *Human Gene Therapy* 5:3-10 (1994); Zabner, *Cell* 75:207-216 (1993); Caillaud, *Eur. J. Neuroscience* 5:1287-1291 (1993); and Ragot, *J. Gen. Virology* 74:501-507 (1993)). Recombinant adenoviruses achieve gene transduction by binding to specific cell surface receptors, after which the virus is internalized by receptor-mediated endocytosis, in the same manner as wild type or replication-defective adenovirus (Chardonnet and Dales, *Virology* 40:462-477 (1970); Brown and Burlingham, *J. Virology* 12:386-396 (1973); Svensson and Persson, *J. Virology* 55:442-449 (1985); Seth, et al., *J. Virol.* 51:650-655 (1984); Seth, et al., *Mol. Cell. Biol.* 4:1528-1533 (1984); Varga et al., *J. Virology* 65:6061-6070 (1991); Wickham et al., *Cell* 73:309-319 (1993)).

A viral vector can be one based on an adenovirus which has had the E1 gene removed and these virions are generated in a cell line such as the human 293 cell line. In another preferred embodiment both the E1 and E3 genes are removed from the adenovirus genome.

Another type of viral vector is based on an adeno-associated virus (AAV). This defective parvovirus is a preferred vector because it can infect many cell types and is nonpathogenic to humans. AAV type vectors can transport about 4 to 5 kb and wild type AAV is known to stably insert into chromosome 19. Vectors which contain this site specific integration property are preferred. An especially preferred embodiment of this type of vector is the P4.1 C vector produced by Avigen, San Francisco, CA, which can contain the herpes simplex virus thymidine kinase gene, HSV-tk, and/or a marker gene, such as the gene encoding the green fluorescent protein, GFP.

In another type of AAV virus, the AAV contains a pair of inverted terminal repeats (ITRs) which flank at least one cassette containing a promoter which directs cell-specific expression operably linked to a heterologous gene. Heterologous in this

context refers to any nucleotide sequence or gene which is not native to the AAV or B19 parvovirus.

Typically the AAV and B19 coding regions have been deleted, resulting in a safe, noncytotoxic vector. The AAV ITRs, or modifications thereof, confer infectivity
5 and site-specific integration, but not cytotoxicity, and the promoter directs cell-specific expression. United states Patent No. 6,261,834 is herein incorporated by reference for material related to the AAV vector.

The vectors of the present invention thus provide DNA molecules which are capable of integration into a mammalian chromosome without substantial toxicity.

10 The inserted genes in viral and retroviral usually contain promoters, and/or enhancers to help control the expression of the desired gene product. A promoter is generally a sequence or sequences of DNA that function when in a relatively fixed location in regard to the transcription start site. A promoter contains core elements required for basic interaction of RNA polymerase and transcription factors, and may
15 contain upstream elements and response elements.

Molecular genetic experiments with large human herpesviruses have provided a means whereby large heterologous DNA fragments can be cloned, propagated and established in cells permissive for infection with herpesviruses (Sun et al., *Nature genetics* 8: 33-41, 1994; Cotter and Robertson, *Curr Opin Mol Ther* 5: 633-644, 1999).

20 These large DNA viruses (herpes simplex virus (HSV) and Epstein-Barr virus (EBV), have the potential to deliver fragments of human heterologous DNA > 150 kb to specific cells. EBV recombinants can maintain large pieces of DNA in the infected B-cells as episomal DNA. Individual clones carried human genomic inserts up to 330 kb appeared genetically stable The maintenance of these episomes requires a specific EBV
25 nuclear protein, EBNA1, constitutively expressed during infection with EBV.

Additionally, these vectors can be used for transfection, where large amounts of protein can be generated transiently in vitro. Herpesvirus amplicon systems are also being used to package pieces of DNA > 220 kb and to infect cells that can stably maintain DNA as episomes.

Other useful systems include, for example, replicating and host-restricted non-replicating vaccinia virus vectors.

The disclosed compositions can be delivered to the target cells in a variety of ways. For example, the compositions can be delivered through electroporation, or
5 through lipofection, or through calcium phosphate precipitation. The delivery mechanism chosen will depend in part on the type of cell targeted and whether the delivery is occurring for example *in vivo* or *in vitro*.

Thus, the compositions can comprise, in addition to the disclosed vectors for example, lipids such as liposomes, such as cationic liposomes (e.g., DOTMA, DOPE,
10 DC-cholesterol) or anionic liposomes. Liposomes can further comprise proteins to facilitate targeting a particular cell, if desired. Administration of a composition comprising a compound and a cationic liposome can be administered to the blood afferent to a target organ or inhaled into the respiratory tract to target cells of the respiratory tract. Regarding liposomes, see, e.g., Bringham et al. *Am. J. Resp. Cell. Mol.*
15 *Biol.* 1:95-100 (1989); Felgner et al. *Proc. Natl. Acad. Sci USA* 84:7413-7417 (1987); U.S. Pat. No.4,897,355. Furthermore, the compound can be administered as a component of a microcapsule that can be targeted to specific cell types, such as macrophages, or where the diffusion of the compound or delivery of the compound from the microcapsule is designed for a specific rate or dosage.

20 In the methods described above which include the administration and uptake of exogenous DNA into the cells of a subject (i.e., gene transduction or transfection), delivery of the compositions to cells can be via a variety of mechanisms. As one example, delivery can be via a liposome, using commercially available liposome preparations such as LIPOFECTIN, LIPOFECTAMINE (GIBCO-BRL, Inc.,
25 Gaithersburg, MD), SUPERFECT (Qiagen, Inc. Hilden, Germany) and TRANSFECTAM (Promega Biotec, Inc., Madison, WI), as well as other liposomes developed according to procedures standard in the art. In addition, the nucleic acid or vector of this invention can be delivered *in vivo* by electroporation, the technology for which is available from Genetronics, Inc. (San Diego, CA) as well as by means of a
30 SONOPORATION machine (ImaRx Pharmaceutical Corp., Tucson, AZ).

The materials may be in solution, suspension (for example, incorporated into microparticles, liposomes, or cells). These may be targeted to a particular cell type via VLRs, antibodies, receptors, or receptor ligands. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Senter, et al.,
5 *Bioconjugate Chem.*, 2:447-451, (1991); Bagshawe, K.D., *Br. J. Cancer*, 60:275-281, (1989); Bagshawe, et al., *Br. J. Cancer*, 58:700-703, (1988); Senter, et al., *Bioconjugate Chem.*, 4:3-9, (1993); Battelli, et al., *Cancer Immunol. Immunother.*, 35:421-425, (1992); Pietersz and McKenzie, *Immunolog. Reviews*, 129:57-80, (1992); and Roffler, et al., *Biochem. Pharmacol.*, 42:2062-2065, (1991)). These techniques can be used for a
10 variety of other specific cell types. Vehicles such as "stealth" and other antibody or VLR conjugated liposomes (including lipid mediated drug targeting to colonic carcinoma), receptor mediated targeting of DNA through cell specific ligands, lymphocyte directed tumor targeting, and highly specific therapeutic retroviral targeting of murine glioma cells *in vivo*. The following references are examples of the use of this
15 technology to target specific proteins to tumor tissue (Hughes et al., *Cancer Research*, 49:6214-6220, (1989); and Litzinger and Huang, *Biochimica et Biophysica Acta*, 1104:179-187, (1992)). In general, receptors are involved in pathways of endocytosis, either constitutive or ligand induced. These receptors cluster in clathrin-coated pits, enter the cell via clathrin-coated vesicles, pass through an acidified endosome in which
20 the receptors are sorted, and then either recycle to the cell surface, become stored intracellularly, or are degraded in lysosomes. The internalization pathways serve a variety of functions, such as nutrient uptake, removal of activated proteins, clearance of macromolecules, opportunistic entry of viruses and toxins, dissociation and degradation of ligand, and receptor-level regulation. Many receptors follow more than one
25 intracellular pathway, depending on the cell type, receptor concentration, type of ligand, ligand valency, and ligand concentration. Molecular and cellular mechanisms of receptor-mediated endocytosis has been reviewed (Brown and Greene, *DNA and Cell Biology* 10:6, 399-409 (1991)).

Nucleic acids that are delivered to cells which are to be integrated into the host
30 cell genome, typically contain integration sequences. These sequences are often viral

related sequences, particularly when viral based systems are used. These viral intergration systems can also be incorporated into nucleic acids which are to be delivered using a non-nucleic acid based system of deliver, such as a liposome, so that the nucleic acid contained in the delivery system can be come integrated into the host genome.

Other general techniques for integration into the host genome include, for example, systems designed to promote homologous recombination with the host genome. These systems typically rely on sequence flanking the nucleic acid to be expressed that has enough homology with a target sequence within the host cell genome that recombination between the vector nucleic acid and the target nucleic acid takes place, causing the delivered nucleic acid to be integrated into the host genome. These systems and the methods necessary to promote homologous recombination are known to those of skill in the art.

As described above, the compositions can be administered in a pharmaceutically acceptable carrier and can be delivered to the subject's cells *in vivo* and/or *ex vivo* by a variety of mechanisms well known in the art (e.g., uptake of naked DNA, liposome fusion, intramuscular injection of DNA via a gene gun, endocytosis and the like).

If *ex vivo* methods are employed, cells or tissues can be removed and maintained outside the body according to standard protocols well known in the art. The compositions can be introduced into the cells via any gene transfer mechanism, such as, for example, calcium phosphate mediated gene delivery, electroporation, microinjection or proteoliposomes. The transduced cells can then be infused (e.g., in a pharmaceutically acceptable carrier) or homotopically transplanted back into the subject per standard methods for the cell or tissue type. Standard methods are known for transplantation or infusion of various cells into a subject.

The nucleic acids that are delivered to cells typically contain expression controlling systems. For example, the inserted genes in viral and retroviral systems usually contain promoters, and/or enhancers to help control the expression of the desired gene product. A promoter is generally a sequence or sequences of DNA that function when in a relatively fixed location in regard to the transcription start site. A

promoter contains core elements required for basic interaction of RNA polymerase and transcription factors, and may contain upstream elements and response elements.

Preferred promoters controlling transcription from vectors in mammalian host cells may be obtained from various sources, for example, the genomes of viruses such as: polyoma, Simian Virus 40 (SV40), adenovirus, retroviruses, hepatitis-B virus and most preferably cytomegalovirus, or from heterologous mammalian promoters, e.g. beta actin promoter. The early and late promoters of the SV40 virus are conveniently obtained as an SV40 restriction fragment which also contains the SV40 viral origin of replication (Fiers et al., *Nature*, 273: 113 (1978)). The immediate early promoter of the human cytomegalovirus is conveniently obtained as a *Hind*III E restriction fragment (Greenway, P.J. et al., *Gene* 18: 355-360 (1982)). Of course, promoters from the host cell or related species also are useful herein.

Enhancer generally refers to a sequence of DNA that functions at no fixed distance from the transcription start site and can be either 5' (Laimins, L. et al., *Proc. Natl. Acad. Sci.* 78: 993 (1981)) or 3' (Lusky, M.L., et al., *Mol. Cell Bio.* 3: 1108 (1983)) to the transcription unit. Furthermore, enhancers can be within an intron (Banerji, J.L. et al., *Cell* 33: 729 (1983)) as well as within the coding sequence itself (Osborne, T.F., et al., *Mol. Cell Bio.* 4: 1293 (1984)). They are usually between 10 and 300 bp in length, and they function in cis. Enhancers function to increase transcription from nearby promoters. Enhancers also often contain response elements that mediate the regulation of transcription. Promoters can also contain response elements that mediate the regulation of transcription. Enhancers often determine the regulation of expression of a gene. While many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, -fetoprotein and insulin), typically one will use an enhancer from a eukaryotic cell virus for general expression. Preferred examples are the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

The promoter and/or enhancer may be specifically activated either by light or specific chemical events which trigger their function. Systems can be regulated by

reagents such as tetracycline and dexamethasone. There are also ways to enhance viral vector gene expression by exposure to irradiation, such as gamma irradiation, or alkylating chemotherapy drugs.

In certain embodiments the promoter and/or enhancer region can act as a
5 constitutive promoter and/or enhancer to maximize expression of the region of the transcription unit to be transcribed. In certain constructs the promoter and/or enhancer region be active in all eukaryotic cell types, even if it is only expressed in a particular type of cell at a particular time. A preferred promoter of this type is the CMV promoter (650 bases). Other preferred promoters are SV40 promoters, cytomegalovirus (full
10 length promoter), and retroviral vector LTR.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human or nucleated cells) may also contain sequences necessary for the termination of transcription which may affect mRNA expression. These regions are transcribed as polyadenylated segments in the untranslated portion of the mRNA
15 encoding tissue factor protein. The 3' untranslated regions also include transcription termination sites. It is preferred that the transcription unit also contain a polyadenylation region. One benefit of this region is that it increases the likelihood that the transcribed unit will be processed and transported like mRNA. The identification and use of polyadenylation signals in expression constructs is well established. It is
20 preferred that homologous polyadenylation signals be used in the transgene constructs. In certain transcription units, the polyadenylation region is derived from the SV40 early polyadenylation signal and consists of about 400 bases. It is also preferred that the transcribed units contain other standard sequences alone or in combination with the above sequences improve expression from, or stability of, the construct.

25 The viral vectors can include nucleic acid sequence encoding a marker product. This marker product is used to determine if the gene has been delivered to the cell and once delivered is being expressed. Preferred marker genes are the *E. Coli* lacZ gene, which encodes β -galactosidase, and green fluorescent protein.

In some embodiments the marker may be a selectable marker. Examples of
30 suitable selectable markers for mammalian cells are dihydrofolate reductase (DHFR),

thymidine kinase, neomycin, neomycin analog G418, hydromycin, and puromycin.

When such selectable markers are successfully transferred into a mammalian host cell, the transformed mammalian host cell can survive if placed under selective pressure.

There are two widely used distinct categories of selective regimes. The first category is
5 based on a cell's metabolism and the use of a mutant cell line which lacks the ability to grow independent of a supplemented media. Two examples are CHO DHFR- cells and mouse LTK- cells. These cells lack the ability to grow without the addition of such nutrients as thymidine or hypoxanthine. Because these cells lack certain genes necessary for a complete nucleotide synthesis pathway, they cannot survive unless the
10 missing nucleotides are provided in a supplemented media. An alternative to supplementing the media is to introduce an intact DHFR or TK gene into cells lacking the respective genes, thus altering their growth requirements. Individual cells which were not transformed with the DHFR or TK gene will not be capable of survival in non-supplemented media.

15 The second category is dominant selection which refers to a selection scheme used in any cell type and does not require the use of a mutant cell line. These schemes typically use a drug to arrest growth of a host cell. Those cells which have a novel gene would express a protein conveying drug resistance and would survive the selection. Examples of such dominant selection use the drugs neomycin, (Southern P. and Berg,
20 P., *J. Molec. Appl. Genet.* 1: 327 (1982)), mycophenolic acid, (Mulligan, R.C. and Berg, P. *Science* 209: 1422 (1980)) or hygromycin, (Sugden, B. et al., *Mol. Cell. Biol.* 5: 410-413 (1985)). The three examples employ bacterial genes under eukaryotic control to convey resistance to the appropriate drug G418 or neomycin (geneticin), xgpt (mycophenolic acid) or hygromycin, respectively. Others include the neomycin analog
25 G418 and puramycin.

In the methods described above which include the administration and uptake of exogenous DNA into the cells of a subject (i.e., gene transduction or transfection), the nucleic acids of the present invention can be in the form of naked DNA or RNA, or the nucleic acids can be in a vector for delivering the nucleic acids to the cells, whereby the
30 antibody-encoding DNA fragment is under the transcriptional regulation of a promoter,

as would be well understood by one of ordinary skill in the art. The vector can be a commercially available preparation, such as an adenovirus vector (Quantum Biotechnologies, Inc. (Laval, Quebec, Canada). Delivery of the nucleic acid or vector to cells can be via a variety of mechanisms. As one example, delivery can be via a liposome, using commercially available liposome preparations such as LIPOFECTIN,
5 LIPOFECTAMINE (GIBCO-BRL, Inc., Gaithersburg, MD), SUPERFECT (Qiagen, Inc. Hilden, Germany) and TRANSFECTAM (Promega Biotec, Inc., Madison, WI), as well as other liposomes developed according to procedures standard in the art. In addition, the nucleic acid or vector of this invention can be delivered *in vivo* by
10 electroporation, the technology for which is available from Genetronics, Inc. (San Diego, CA) as well as by means of a SONOPORATION machine (ImaRx Pharmaceutical Corp., Tucson, AZ).

As one example, vector delivery can be via a viral system, such as a retroviral vector system which can package a recombinant retroviral genome (see e.g., Pastan et al., *Proc. Natl. Acad. Sci. U.S.A.* 85:4486, 1988; Miller et al., *Mol. Cell. Biol.* 6:2895,
15 1986). The recombinant retrovirus can then be used to infect and thereby deliver to the infected cells nucleic acid encoding a broadly neutralizing antibody (or active fragment thereof) of the invention. The exact method of introducing the altered nucleic acid into mammalian cells is, of course, not limited to the use of retroviral vectors. Other
20 techniques are widely available for this procedure including the use of adenoviral vectors (Mitani et al., *Hum. Gene Ther.* 5:941-948, 1994), adeno-associated viral (AAV) vectors (Goodman et al., *Blood* 84:1492-1500, 1994), lentiviral vectors (Naidini et al., *Science* 272:263-267, 1996), pseudotyped retroviral vectors (Agrawal et al., *Exper. Hematol.* 24:738-747, 1996). Physical transduction techniques can also be used,
25 such as liposome delivery and receptor-mediated and other endocytosis mechanisms (see, for example, Schwartzenberger et al., *Blood* 87:472-478, 1996). This invention can be used in conjunction with any of these or other commonly used gene transfer methods.

As one example, if the antibody-encoding nucleic acid of the invention is
30 delivered to the cells of a subject in an adenovirus vector, the dosage for administration

of adenovirus to humans can range from about 10^7 to 10^9 plaque forming units (pfu) per injection but can be as high as 10^{12} pfu per injection (Crystal, *Hum. Gene Ther.* 8:985-1001, 1997; Alvarez and Curiel, *Hum. Gene Ther.* 8:597-613, 1997). A subject can receive a single injection, or, if additional injections are necessary, they can be repeated
5 at six month intervals (or other appropriate time intervals, as determined by the skilled practitioner) for an indefinite period and/or until the efficacy of the treatment has been established.

Parenteral administration of the nucleic acid or vector of the present invention, if used, is generally characterized by injection. Injectables can be prepared in
10 conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant dosage is maintained. See, e.g., U.S. Patent No. 3,610,795, which is incorporated by reference herein. For additional
15 discussion of suitable formulations and various routes of administration of therapeutic compounds, see, e.g., *Remington: The Science and Practice of Pharmacy* (19th ed.) ed. A.R. Gennaro, Mack Publishing Company, Easton, PA 1995.

The invention further provides a method of making a polypeptide of the invention comprising culturing a cell comprising a vector comprising a nucleic acid that
20 encodes the polypeptide and purifying the polypeptide from the cell or from the medium. Further provided are methods of making a polypeptide of the invention using protein synthesis techniques.

Also disclosed are methods of screening for one or more variable lymphocyte receptors in a subject comprising identifying in the subject one or more polypeptides
25 comprising an N-terminal leucine rich repeat (LRRNT), one or more leucine rich repeats (LRRs), a C-terminal leucine rich repeat (LRRCT), and a connecting peptide, wherein the connecting peptide comprises an alpha helix.

Examples

The following examples are put forth so as to provide those of ordinary skill in
30 the art with a complete disclosure and description of how the compounds,

compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary of the invention and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.), but some
5 errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric.

Example 1: Variable Lymphocyte Receptors in Sea Lamprey

Analysis of Transcripts from Immunostimulated Blood Lymphocytes

10 In order to survey the transcriptome of activated lymphocytes, lamprey larvae were stimulated by intraperitoneal injections of an antigen/mitogen cocktail, including live *E. coli* bacteria, sheep erythrocytes, phytohemagglutinin and pokeweed mitogen, two to four times at weekly intervals. The fraction of large lymphocytes among peripheral blood leukocytes three days after the second booster stimulation was 13-fold
15 greater than in unstimulated individuals, and the fraction of myeloid cells was also 6-fold greater (Fig. 1a). Compared to the small blood lymphocytes, the large lymphocytes were nearly double in size, had extensive azurophilic cytoplasm and featured prominent nucleoli (Fig. 1b). These cells were sorted and used to construct cDNA libraries enriched in messages of activated lymphocytes by subtraction against cDNA from
20 lamprey activated myeloid cells or erythrocytes.

The most abundant group of sequences identified among 1,507 clones from the subtracted libraries predicted 319 proteins with variable numbers of diverse leucine-rich repeat (LRR) motifs, that clustered with a set of 52 LRR-containing expressed sequence tags (EST) from a survey of unstimulated lymphocyte transcripts. After purging the 3'
25 end sequences, a set of 239 uniquely diverse LRR proteins were identified, 22 of which encoded most or all of the open reading frames (ORF) of 239-304 aa (Fig. 6). These lamprey proteins were provisionally named variable lymphocyte receptors (VLR) because each of these 239 sequences was unique and their transcripts were found to be expressed predominantly by lymphocytes (Fig. 1c). Lymphocytes from hematopoietic
30 tissues showed highest VLR levels in unstimulated animals, and immune stimulation

resulted in enhanced VLR transcription by the large blood lymphocytes. The basic composition of these VLRs included a conserved signal peptide, N-terminal LRR (LRRNT), a variable number of diverse LRRs, a connecting peptide followed by a C-terminal LRR (LRRCT) and a conserved C-terminus composed of a threonine- and
5 praline rich stalk, a generic glycosyl-phosphatidyl-inositol (GPI)-anchor site and a hydrophobic tail (Fig. 1d and Fig. 7). When a retroviral construct encoding an epitope tagged VLR was transfected into a mammalian cell line, immunofluorescence analysis confirmed the cell surface localization of the protein, and treatment with bacterial GPI-specific phospholipase C significantly reduced the level of cell surface expression (Fig.
10 1e) and released VLR protein into the supernatant. The longest VLR sequence consisting of 11 LRRs was threaded on the crystal structure coordinates of related LRR proteins to generate a 3-dimensional structural model (Schwede et al., 2000). The model provides a concave solenoid structure in which nine β -sheets are capped on both ends by the LRRNT and LRRCT (Fig. 1f), similar to the model predicted for Toll-like
15 receptor (TLR) ectodomains (Bell et al., 2003).

The VLR repertoire is highly diverse in individual lampreys

The VLR diversity was surveyed in individual lampreys by RT-PCR. Blood leukocytes mRNA from three immunostimulated and four unstimulated larvae was amplified with primers flanking the VLRs diversity region. Sequencing of ~10 clones
20 per animal yielded 69 unique VLRs and only two identical clones from one individual. Variable sequences of 20 VLRs from two animals illustrate the protein diversity (Fig. 2; entire set included in Fig. 3 and Fig. 8). The size variation, 134-214 aa, is primarily due to differences in number of LRR modules. Each sequence contains an LRRNT, an 18 aa LRR1, 1-9 LRRs almost invariably 24 aa long, a 13 aa connecting peptide and C-
25 terminal LRRCT; the LRRNTs have 30-38 aa and the LRRCTs 48-58 aa. While regions of pronounced sequence diversity are evident for each LRR motif, the first seven residues in LRRNT and the last 20 residues in LRRCT are nearly invariant.

To assess VLRs diversity at the level of individual lymphocytes RT-PCR with primers flanking the whole ORF was used. Single cell isolates were sorted from the
30 blood of an immunostimulated and an unstimulated larvae. Analysis of the PCR

products obtained from six single cell reactions from the unstimulated animal and seven reactions from the immunostimulated larva showed that 12 of the 13 lymphocytes expressed a single VLR (Fig. 3), and five of six VLR clones from a control pool of 10 unstimulated cells were unique. One cell isolate yielded two VLRS (9.16S, 9.16L), but
5 the possibility that this isolate contained two lymphocytes cannot be excluded. Three of the VLRS had in-frame stop codons predicting truncated proteins. Interestingly, combinations of identical VLRS were identified among five lymphocytes from the immunostimulated larva (9.1=9.16S; 9.2=9.16L; 9.7=9.9). The analysis of blood samples from three additional immunostimulated larvae (#5-7) revealed only unique
10 VLRS (N=27). These findings are indicative of monoallelic expression of the diverse VLRS, and provide preliminary evidence for clonal expansion of VLR-bearing lymphocytes.

Complexity of the *VLR* locus

Genome blot hybridization with a conserved C-terminal probe revealed a single
15 band (Fig. 4a). The N-terminal probe, consisting of the conserved 5' UTR and signal peptide, reacted with 2-3 bands depending upon the restriction enzyme employed, except for an individual whose blot showed 2 additional BamHI bands. In addition, a genomic pulse-field CHEF blot revealed a single hybridization band with the C-terminal probe in all six digests, whereas the N-terminal probe produced a matching
20 pattern with one additional 350 kb *NotI* band (Fig. 4b). These findings indicate a single *VLR* locus, with the *N-terminus* and *C-terminus* of the *germline VLR* gene (*gVLR*) contained within 100-150 kb of the genome (Fig. 4b; *PacI* digest). To further characterize the locus, these probes were used to screen a large insert sea lamprey P1 bacterial artificial chromosome (PAC) library constructed from erythrocyte DNA of one
25 adult. In an analysis of five PACs that hybridized with both probes, a single 14 kb VLR gene (*gVLR*) amplicon was identified by long range PCR (LR-PCR) using the ORF-flanking primers. Restriction-enzyme analysis of the PCR products revealed identical *EcoRI* bands and two allelic *BamHI* patterns. PAC clones representing the two *gVLR* alleles were sequenced, PAC3 and PAC16 with 33 and 44 kb inserts respectively. Their
30 sequences overlapped a 20 kb region containing the *gVLR*; PAC16 extended 25 kb

upstream from the gVLR and PAC3 extended 18 kb downstream. The overlap region between PACs 3 and 16 was nearly identical, except for short deletions in the gVLR of PAC16 (24, 43 and 78 bp). These sequences were therefore melded into a gVLR contig preserving the slightly longer sequence of PAC3 (Fig. 5a).

5 The gVLR in the PAC3/16 contig consist of 4 exons. The first contains part of the 5' UTR; exon 2 contains the rest of the 5' UTR, a signal peptide and the 5' half of LRRNT; exon 3 encodes the 5' half of LRRCT, and exon 4 encodes the 3' half of LRRCT, the C-terminus and 3' UTR. Canonical eukaryotic splice sites were identified only in the 5' UTR intron, while other exon/intron boundaries in the gVLR were
10 determined by alignment to cDNA sequences. Notably, the gVLR sequence did not contain a 3' LRRNT, LRR1 or any of the 24 aa LRRs. Upstream from this gVLR, six cassettes of variable LRR modules were identified, singlet or doublet, including LRRNT, LRR1 and LRR positioned either in forward or reverse orientation. These LRR cassettes spanned the first 6 kb of the contig, while two diverse 5' LRRCT
15 cassettes were located 7 kb downstream from the gVLR.

Another clone, PAC4, hybridized only with the N-terminal probe but it was found to encode multiple LRRs that were identified by PCR with LRRNT and LRR1 consensus primers. The entire insert was 58 kb long (Fig. 5b), and the sequence overlapped 11.7 kb of the gVLR contig with minor gaps (four gaps of 210-738 bp in
20 PAC4 and eight gaps of 25-55 bp in the PAC3/16 contig). The overlap extended into the intervening sequence between gVLR exons 2 and 3, but the 553 bp terminal sequence of PAC4 was unique. Seventeen cassettes of 1-3 diverse LRR modules, 30 in total, were encoded in a 31 kb region in PAC4 located 15 kb upstream from the partial gVLR. Comparison of the PAC3/16 gVLR contig and PAC4 sequences revealed
25 additional 1-5 kb regions with >90% identity, but these were disrupted by unrelated sequences. PAC4 could represent either a duplication of ~12 kb, encompassing the 5' flank and about half of the gVLR, or a highly divergent VLR allele. To distinguish between these possibilities the pattern of genomic hybridization was compared with the N-terminal probe (Fig. 4a) to the map of restriction sites in the gVLRs from these PAC
30 inserts. The blot pattern and restriction map were compatible for all fragments except

for a 5.7 kb HindIII fragment from PAC4 that was different than the 2 kb band in the blot (Fig. 4a). In view of such limited variability amongst three blotted genomes and the genome from the PAC library, PAC4 seems unlikely to represent a polymorphic gVLR allele. Limited VLR allelic variation would be consistent with other evidence of low allelic diversity even in microsatellite loci (Bryan et al., 2003), indicating the sea lamprey populations in the North American Great Lakes and other landlocked populations are highly inbred. The analysis thus indicates the single lamprey gVLR locus harbors an additional copy of the N-terminal half of the gVLR.

Somatic *gVLR* rearrangement generates diverse *mature VLRs*

When larval DNA samples were analyzed by PCR amplification with primers flanking the VLR diversity region, six unique intron-less VLR ORFs were obtained (Fig. 3, animals #10, 12). In accordance with this intriguing finding, PCR amplification of larval DNA samples with the ORFflanking primers produced VLR clones of 1.5-2 kb including the 5' UTR intron, revealing unique sequence in 13 of 14 clones (#10, 11). Because these genomic PCR clones contained uninterrupted VLR ORFs, they were provisionally named mature VLRs to distinguish them from the 'incomplete' germline VLR. Sequence analysis indicated that these mature VLRs should generate 1-1.3 kb polymorphic EcoRI bands hybridizing with the N-terminal probe, but these bands were observed only in a lymphocyte DNA blot (Fig. 4c to be included). These observations indicate that lamprey DNA samples extracted from pelleted blood erythrocytes or whole larval bodies contain mature VLRs, but only copies of the germline VLR are sufficiently abundant to be detected in DNA blots from these samples.

To address this enigma it was theorized that somatic gene rearrangement in lamprey lymphocytes generated the small mature VLRs, replacing non-coding DNA from the germline gVLR with diverse LRRs from the upstream and downstream cassettes. To test this hypothesis primers were designed for PCR amplification across the germline gVLR, including ~3 kb of upstream and ~3 kb of downstream flanks (Fig. 5a). LR-PCR amplification from larval DNA samples yielded a minor band of ~20 kb, similar to the gVLR amplicon from PAC16 plasmid, plus an additional prominent band of ~8 kb (Fig. 5c). Sequence analysis of the 8 kb amplicons from two larval samples

revealed 9 of 10 clones encoding unique mature VLRs (Fig. 3), the flanks of which were identical to those of the gVLR (Fig. 5d). Altogether 28 unique mature VLRs were identified among the PCR products from four larval DNA samples. Lymphocyte DNA was most likely the template for these mature VLRs, as a small fraction of the pelleted erythrocytes or whole larval bodies used to extract these DNA samples. Apparently, the shorter templates of lymphocyte mature VLRs were preferentially amplified during the LR-PCR. A similar PCR bias was observed when amplifying with primers that flanked the gVLR ORF, resulting in two amplicons, the 1.5-2 kb of mature VLRs and the 14 kb gVLRs.

The search for lymphocyte receptors that could trigger adaptive immune responses in lampreys thus identifies a system of variable lymphocyte receptors that is entirely different from the Ig and TCR of jawed vertebrates. The VLRs consist of multiple LRR modules and an invariant stalk region that is attached to the lymphocyte plasma membrane via a GPI-anchor. The flanking tips of the N-terminal and C-terminal LRRs are invariant and the remarkable VLR diversity is contributed by variation in number and sequences of the intervening LRRs. The potential VLR diversity is vast, with 345 out of 354 unique sequences, and only three pairs of identical VLRs from immunostimulated lymphocytes and three other nearly identical VLRs. The VLRs thus endow this agnathan representative with a diverse repertoire of lymphocyte receptors.

These highly diverse VLRs serve a role in recognition of pathogens. Proteins featuring diverse LRR modules are cardinal innate immune receptors of animals and plants due to their propensity to interact with an extraordinary vast array of ligands. Animal TLRs are implicated in recognition of conserved epitopes on viruses, bacteria, fungi and protozoa, activating signal transduction cascades that culminate in inflammatory responses (Beutler, 2004). CD14, a GPI-anchored LRR protein that is also found in a soluble form, binds bacterial lipopolysaccharide and phospholipids to form a signaling complex with the TLR4 receptor (Landmann, 2000). Yet another mammalian family of cytosolic LRR proteins, the NBS-LRRs, recognize intracellular pathogens (Chamaillard et al., 2003). Plant disease resistance genes are members of large multigene families including hundreds of NBS-LRR proteins, LRR-receptorlike

kinases and LRR-receptor-like proteins, many of which have been shown to be involved in specific activation of anti-pathogen responses (Jones et al., 2004). Antigen-binding VLRs with their remarkable diversity mediate the adaptive immune responses observed in lampreys. The GPI-anchorage of VLRs to the surface of lymphocytes allow
5 GPI-specific phospholipase release of these receptors (Ikezawa 2002), endowing VLRs with dual functionality both as surface receptors and humoral agglutinins in an anticipatory immune system.

Sequencing genomic PAC clones a germline gVLR consisting of 4 exons that encoded only the signal peptide, 5' LRRNT, 5' LRRCT, 3' LRRCT and the C-terminus
10 was identified. The gVLR lacked diversity LRR modules except for a 5' LRRCT, indicating that without modification it could not encode the highly diverse VLR messages. However, multiple diverse LRR cassettes were found upstream and downstream from the gVLR, and these could be available for insertion into the gVLR to assemble mature VLR genes. To test the hypothesis that mature VLRs are generated
15 through somatic replacement of non-coding DNA in the germline gVLR with upstream and downstream LRR cassettes, LR-PCR was used to detect the presence of both germline and mature VLR genes. The expected product of ~20 kb from the gVLR was obtained from genomic DNA of two lampreys and in addition, the predicted 8 kb amplicon from mature VLRs, that was found to encode a diverse set of mature VLRs.
20 Moreover, in a few cases candidate LRR donors could be identified among the gVLR neighboring cassettes based on identity to VLR sequences, and the highly conserved sequences in the gVLR 5' LRRNT and 3' LRRCT could potentially serve as anchoring regions for a gene conversion process. VLRs are generated by a mechanism of somatic DNA rearrangement.

25 Non-meiotic DNA rearrangements are known from other systems. For example, rearrangement of genes encoding surface components is a strategy used by several pathogens to evade immune recognition during chronic infection. Antigenic variation in the pilin of *Neisseria gonorrhoeae* involves non-reciprocal recombination between the pilE locus and multiple silent pilS copies (Hamrick, 2001), and antigenic variation in
30 Lyme disease *Borrelia spirochaetes* is generated by gene conversion between an array of

15 silent cassettes and the vlsE expression site (Wang et al., 2003). Also the protozoan *Trypanosoma brucei* alternate expression of their variant surface coat glycoprotein by repeated DNA rearrangements (Donelson, 2003), as well as the malaria parasite *Plasmodium falciparum* and the intestinal dweller *Giardia lamblia* that frequently
5 switch among multiple surface antigen genes. In the evolutionary arms race between hosts and parasites, vertebrates adopted a similar strategy to combat infectious disease by somatic rearrangement of germline receptors. Diverse lymphocyte antigen receptors are assembled via the cut-and-paste activity of the paired transposase-like RAG1 and RAG2 in gnathostomes (Schluter et al., 1999) and via an as yet uncharacterized
10 mechanism in agnatha.

Features of the lamprey VLR system bear analogy to the Ig and TCR of jawed vertebrate lymphocytes, with two notable differences. First, lamprey VLRs consist of LRR modules whereas gnathostome antigen receptors consist of Ig domains. Lampreys immunity underwent a gradual evolutionary process, replacing the ancestral germline
15 encoded diversity of LRR receptors with a system of variable lymphocyte LRR receptors that are somatically diversified versions of their germline VLR gene. In contrast, Ig domains as core components of jawed vertebrates recombinatorial lymphocyte receptors is an intriguing untraceable evolutionary drift from their predecessors, since no Ig superfamily member has yet been shown to play a role in any
20 type of immune recognition of pathogens or allografts in animals other than the jawed vertebrates (Kaufman, 2002). Second, no evidence for the existence of MHC molecules in the lamprey has been found. In jawed vertebrates polymorphic MHC molecules are essential for efficient presentation of antigen peptides to T-cells, whereas inbred MHC homozygotes appear to suffer from impaired disease resistance (Penn et al., 2002;
25 Grimholt et al., 2003). Since lampreys thrive as an inbred population in the Great Lakes, this indicates their VLR system may have evolved to function independent of polymorphic components.

Animals

Larvae (8–13 cm long) of the sea lamprey were from tributaries to Lake
30 Michigan (Lamprey Services, Ludington, MI), or tributaries to Lake Huron (Hammond

Bay Biological Station, Millersburg, MI). Larvae for immunostimulation were sedated (100 mg/l MS222; Sigma) and injected intraperitoneally with 75 μ l 0.67X PBS containing: 10⁷ E. coli BL21(DE3), 10⁷ sheep erythrocytes, 50 μ g phytohemagglutinin and 25 μ g pokeweed mitogen (Sigma). Immunostimulation was repeated 2 or 4 times at
5 weekly intervals and cells were collected 3-4 days after last immunization. Blood was drained from tail-severed larvae, diluted 1:1 with 0.57X PBS and 30 mM EDTA. Buffy coat leukocytes were collected after 5 min centrifugation at 50 g. Cells were sorted using MoFlo cytometer as described (Mayer et al., 2002a).

Subtracted immunostimulated lymphocyte cDNA libraries

10 Super SMART PCR cDNA Synthesis (BD Biosciences) was used with mRNA from large blood lymphocytes, myeloid cells and erythrocytes sorted from larvae immunostimulated 4 times at weekly intervals. Activated lymphocyte cDNA was subtracted in 2 reactions against cDNA of myeloid cells or erythrocytes (PCR-Select, BD Biosciences). Subtracted products were cloned in pGEM-T Easy (Promega) and
15 1,507 sequences were analyzed.

Table 3. PCR primers

Primer (10 pmole/ μ l)	Sequence (5'-3')	Position (cDNA clone)	Position (gVLR contig)
Slit.F	CTCGGCTCTGCAGCTCTCA (SEQ ID NO: 159)	2-20 (LRR-2913)	24872-24890
LRR.F1	TGGCGCCCTGGTGCAAAGT (SEQ ID NO: 160)	153-171 "	25643-25661
Slit.R	GAACACTGCGAGGGACATG (SEQ ID NO: 161)	179-197 "	25669- 25687
Dis_LRR.F	AAAAGATCTTGTCCCTCGCAGTGTTTC (SEQ ID NO: 162)	181-197 "	
LRR.R1	ACGGACGGGGGTATTGGTA (SEQ ID NO: 163)	633-651 "	37969- 37987
LRR_C.F1	ATCCCTGAGACCACCACCT (SEQ ID NO: 164)	739-757 "	38075- 38093
LRR_C.R1	CACGCCGATCAACGTTTCCT (SEQ ID NO: 165)	928-947 "	38264- 38283
Dis_LRR.R1	AAAGTCGACACGCCGATCAACGTTTC (SEQ ID NO: 166)	930-946 "	
LRR_C.R2	CCGCCATCCCCGACCTTTG (SEQ ID NO: 167)	948-966 "	38302- 38284
gVLR.F1	CCGGTTGGACACTAGTGTTG (SEQ ID NO: 168)		22285-22304
gVLR.R1	GTGCCATTGGGATCAGTGGT (SEQ ID NO: 169)		42099-42118
GAPDH.F	GAACATCGGCATCAATGGGT (SEQ ID NO: 170)	71-90 (PmGAPDH)	
GAPDH.R	GAGGCCTTATCGATGGTGGT (SEQ ID NO: 171)	366-385 "	

5

VLR RT-PCR

Buffy coat leukocytes from unstimulated larvae (#1-4), or immunostimulated twice at one week intervals (#5-7), were pelleted 5 min at 300 g. First strand cDNA was primed with 50 ng random hexamers (SuperScript III; Invitrogen). VLR diversity regions were amplified with Expand High Fidelity (Roche) using LRR.F1+LRR.R1 (Table 3). Thermal cycling was as follows: 94°C 1 min, then 35 cycles of 94°C 30 sec, 59°C 30 sec, 72°C 1 min. Per animal 10-12 clones were sequenced.

VLR single cell RT-PCR

Single lymphocytes, or a 10-cell pool, from buffy coats of unstimulated larva (#8), and one immunostimulated twice at one week interval (#9), were sorted into 0.2 ml TRIzol (Invitrogen). First strand cDNA was primed with LRR_C.R2. VLRs were

amplified by 2 rounds of nested PCR, first Slit.F+LRR_C.R2 using Advantage II (BD Biosciences) then LRR_N.F1+LRR_C.R1 using Expand High Fidelity. Cycling parameters were: 94°C 1 min, then 40 cycles of 94°C 30 sec, 60°C 30 sec, 72°C 1 min. Colony PCR with vector primers revealed a single size insert in 6 colonies from each of the 12 cells, 3 of which were sequenced. Colonies from cell 9.16 revealed 2 sizes and 3 short and 3 long inserts were sequenced. From the pool of 10 unstimulated cells 6 clones were sequenced.

Genomic DNA and genomic PCR

Genomic DNA was isolated from 1/3 whole larval body, erythrocytes from 0.25 ml blood pelleted for 5 min at 50 g, or 10⁷ sorted lymphocytes. PCR was from 400 ng gDNA using Expand Long Template (Roche). VLR diversity regions were amplified from larvae #10 and 12, using LRR.F1+LRR.R1. Mature VLRs were amplified from animals #10 and 11, using Slit.F+LRR_C.R2, or LRR_N.F1+LRR_C.R1. Amplification across the gVLR was from animals #10 and 13, with gVLR.F1+gVLR.R1. The 8 kb band was cloned in pCR-XL (Invitrogen) and sequenced with: M13.Forward, M13.Reverse, Slit.F and LRR_C.R2.

Virtual Northern and DNA blots

Virtual Northern was prepared as recommended (Super SMART manual). Twenty cycleamplified cDNA was from larval tail, liver and sorted lymphocytes from blood, typhlosole and kidneys of unstimulated animals, or small and large blood lymphocytes, myeloid cells and erythrocytes sorted from blood of larvae immunostimulated 4 times at weekly intervals.

Genomic DNA from larvae #10, 12 and 13, 10 µg per lane, was digested with BamHI, EcoRI or HindIII (Roche); 5 µg lymphocyte DNA was digested with EcoRI. For the pulse-field CHEF blot, erythrocytes from 10 larvae were embedded in agarose, and 20 µg DNA per lane were digested with AscI, FseI, NotI, PacI, PmeI, or SfiI.

The following 32P-labeled probes were used: VLR N-terminal probe, 196 bp, PCR amplified from clone LRR-2913 using Slit.F+Slit.R, and C-terminal probe, 208 bp, amplified with LRR_C.F1+LRR_C.R1; GAPDH probe, 314 bp, amplified from clone PmGAPDH using GAPDH.F+GAPDH.R.

PAC library and clones

Arrayed sea lamprey PAC library in pCYPAC6 (AF133437) was constructed from erythrocyte DNA of one Lake Michigan adult using partial MboI digests. The 6x10⁴ clones had 65 kb average inserts with 1-2 fold genome coverage. Library was
 5 screened using both N-terminal and Cterminal probes. Plasmids of positive clones were EcoRI digested, blotted and hybridized either with the N-terminal or C-terminal probes. Five PACs hybridized with both probes (2, 3, 15, 16, 17) and 5 PACs hybridized only with the N-terminal probe (4, 9, 14, 35, 42, 43).

The gVLR was amplified with Expand Long Template from plasmids of PACs
 10 2, 3, 15, 16 and 17 using Slit.F+LRR_C.R2. All PCR products were of 14 kb, with 2 sets of BamHI patterns (PACs 2, 3 and 15-17). PACs 3, 4 and 16 were sequenced at McGill University (Quebec, Canada).

VLR GPI-anchor

A VLR insert, LRRNT to stop codon, was amplified from clone LRR-2913 with
 15 Expand High Fidelity using Dis_LRR.F+Dis_LRR.R1 and fused to Ig κ signal peptide and Hemagglutinin epitope in pDisplay (Invitrogen). Surface localization and VLR GPI-anchor were analyzed in BW1547 cells, or controls expressing mFc γ RIIb. Cells were treated with 1 unit/ml bacterial GPIspecific phospholipase C (Sigma) 45 min at 30°C. Surface staining of epitope tagged proteins was with anti-HA-tag mAb 12CA5.

20 Sequence analysis

Sequence variability was estimated using MEGA 2.1 UPGMA (Kumar et al., 2001). GPI-anchor site was identified via: <http://129.194.185.165/dgpi/>. SWISS-MODEL VLR 3D structure was via:
http://cubic.bioc.columbia.edu/predictprotein/submit_meta.html. Residues 22-319 from
 25 clone 12.26 were threaded on crystal coordinates of CD42a (1m10.pdb) and NOGO-66 receptor (1p8t.pdb).

Example 2: Variable Lymphocyte Receptors in Hagfish

Cyclostome VLR homologs

Two distinct types of VLR, VLR-A and VLR-B, were identified among
 30 expressed sequence tags from 12,000 leukocyte cDNA clones of the Inshore hagfish,

Eptatretus burgeri (Suzuki et al., 2004B). Matching VLR were then cloned by RT-PCR from transcripts of lymphocyte-like cells of the Pacific hagfish, *E. stoutii*. Figure 9 depicts an alignment of the amino acid sequences of hagfish VLR-A and VLR-B, the Sea lamprey VLR (*Petromyzon marinus*) and VLRs of two non-parasitic lampreys, American brook lamprey (*Lampetra appendix*) and Northern brook lamprey (*Ichthyomyzon fossor*). These VLR share similar structural domains: a signal peptide (SP), N-terminal LRR (LRRNT), 18-residue LRR1 followed by a variable number of 24-residue LRRs, a 13-residue connecting peptide (CP) and C-terminal LRR (LRRCT). At the beginning of the C-terminus the lamprey VLR and hagfish VLR-B have a threonine/proline-rich region, but this region is not well conserved in the hagfish VLR-A. All VLR proteins end with a hydrophobic tail region that is required for modification of the protein to add a glycosyl-phosphatidyl-inositol (GPI) cell surface membrane anchor. Like the sea lamprey VLR, hagfish VLR-A was predicted to be a GPI-anchored protein although no cleavage site was identified (DGPI <http://129.194.185.165/dgpi/>); the C-terminal hydrophobicity profile for VLR-B is also predictive of GPI modification.

Transcripts of hagfish VLR are abundant in lymphocyte-like cells, but not in myeloid cells or erythrocytes sorted by their light scatter characteristics. VLR-A transcript levels were ~3-fold higher than VLR-B levels in blood leukocyte samples. Both VLR types of the Pacific hagfish are highly heterogeneous (Fig. 10A and B), exhibiting variable numbers of the 24-residue LRR modules and pronounced LRR sequence diversity. Comparable diversity was observed for VLR-A (N=66) and VLR-B (N=18) sequences from Inshore hagfish (Fig. 13). Interestingly, five clusters of 2-4 VLR-A clones that were identical or differed by only 1-2 residues were found among the 40 transcripts from hagfish #5 (marked by asterisks in Fig. 10A), that was given four weekly injections of an antigen and mitogen cocktail. The finding that 30% of the VLR-A transcripts from this hagfish consisted of clusters of related sequences indicates clonal expansion of VLR-A bearing lymphocytes. The clones with 1-2 amino acid substitutions reflect additional VLR diversification through somatic hypermutation.

The dataset of unique sequence Pacific hagfish VLR-A (N=130) reveals 2-6 copies per transcript of the 24-residue LRRs (N=527; average 4). In the VLR-B dataset

(N=69) there are 1-6 copies of the 24-residue LRRs (N=195; average 2.8), while in the set of 129 Sea lamprey VLR (19; GenBank accessions AY577943-AY578059) there were 1-9 copies of 24-residue LRRs (N=325; average 2.5). The individual components of these VLR, except for LRRNT and LRRCT that were too diverse among the species for reliable alignment (Table 4; 328 LRR1 domains, 328 CP domains, and 1,047 single domains of the 24-residue LRRs) were then analyzed separately in a Neighbor Joining phylogenetic tree.

Table 4. Components of unique hagfish and Sea lamprey VLR

	Unique LRR motifs			Diversity LRR consensus*
	LRR1 (18 aa)	CP (13 aa)	Diversity LRR (24 aa)	
Es_VLR-A	77/130 (59%)	71/130 (55%)	477/527 (90%)	-L--L--L-L--NqL--IP-G-FD (SEQ ID NO: 304)
Es_VLR-B	68/69 (98%)	46/69 (67%)	190/195 (97%)	KLT-Lt-L-L--NqL-S-P-GvFD (SEQ ID NO: 305)
Pm_VLR	68/129 (53%)	36/129 (28%)	269/325 (83%)	-L--L--L-L--NQL---P-G-FD (SEQ ID NO: 306)

*Consensus - capital letters: 80-100% identity; small letters: 60-79%

The clusters were nearly exclusively of the same type and species origin, i.e., Pacific hagfish VLR-A, VLR-B or Sea lamprey VLR clustering. There were no instances of identical LRR domains between the different VLR types. However, a large portion of the LRR1 and CP domains within hagfish VLR-A and lamprey VLR clusters were identical (Table 4). In contrast, the LRR1 domains in hagfish VLR-B were 98% unique; the sets of 24-residue LRRs also consisted predominantly of unique sequences: 97% were unique in hagfish VLR-B, 90% in VLR-A and 83% in the Sea lamprey VLR. This remarkably high degree of diversity is especially remarkable given that consensus sequences derived for each of the 24-residue LRR types share at least 10 framework residues.

20 Hagfish VLR genes

Genomic organization of the Pacific and Inshore hagfish VLR loci was determined from sequences of large insert genomic clones isolated from bacterial artificial chromosome (BAC) libraries, one BAC for each VLR type (Fig. 11). Only one copy of each of the gVLRs was identified in hagfish genomes. The sequences and

organization of the loci are nearly identical in both species and fairly conserved between gVLR-A and gVLR-B. Hagfish gVLR begin with a 5' untranslated region (UTR) that is followed by two coding regions (Fig. 12A). As in the Sea lamprey gVLR, the 5' UTR is split by an intron, 6.4 kb long in the Pacific hagfish gVLR-A and 220 bp in gVLR-B. The first coding region in the hagfish gVLR encodes the signal peptide and an LRRNT domain in gVLR-A and only residues 1-13 of the 23-residue signal peptide in gVLR-B. Next, there are short intervening sequences of 171 and 211 bp for gVLR-A and gVLR-B, respectively. The second coding region consists of the 3' end of LRRCT and the C-terminus, as in the Sea lamprey gVLR, except that the lamprey region coding for the 5' end of LRRCT is missing. The hagfish gVLR are compact, 671 bp from start-to-stop codons in gVLR-A and 410 bp in gVLR-B.

The hagfish gVLR loci harbor cassettes encoding diverse LRR motifs located ~20-40 kb downstream from the germline genes (Fig. 11). In the VLR-A locus there is a cassette encoding 6-8 terminal residues of a diverse CP domain and a 5' LRRCT that includes a 4-residue identical overlap with the gVLR-A 3' LRRCT. Farther downstream there is a cassette of two diverse LRRs positioned in reverse orientation relative to the gVLR-A and then an inverted incomplete 5' LRRCT. In the gVLR-B locus, there is a cassette encoding residues 7-23 of the signal peptide and a 5' LRRNT, then a diverse CP domain and 5' LRRCT, one inverted LRR and, farther downstream, another inverted LRR cassette consisting of the 12-terminal residues and 8-proximal residues of LRRs. No other diverse LRR modules were identified in flanking DNA spanning ~50 kb upstream and ~70 kb downstream from the gVLRs. However, diverse LRR elements likely exist elsewhere in the genome to provide missing components of the mature VLR genes identified in samples of genomic PCR amplicons from lymphocyte-like cells: 35 unique mature VLR-A and 38 VLR-B sequences from two animals (Fig. 10). Thus, the hagfish mature VLR genes must be assembled through somatic recombination, as is the case for lamprey.

Germline VLR genes in hagfish lymphocyte-like cells are actively transcribed prior to gene rearrangement. PCR amplicons of VLR-A germline transcripts are ~0.7 kb long and ~0.5 kb for VLR-B (Fig. 12B, RT-PCR; position of PCR primers indicated in

Fig. 12A) while the larger amplicons correspond to transcripts from the rearranged mature VLR genes, ~1.1 and ~0.8 kb for VLR-A and VLR-B respectively. The corresponding PCR amplicons from blood genomic DNA are ~0.7 kb for the germline genes and ~1.1 kb for the mature VLR-A and VLR-B genes (Fig. 12B, genomic PCR).
5 In transcripts from germline and mature VLR genes, the 5' intron is spliced out to yield RT-PCR products shorter than the corresponding genomic PCR amplicons (see VLR-B in Fig. 12B; gVLR-A amplicons do not include the 6.4 kb intron). However, the intervening sequences between the coding exons are retained in the germline transcripts because they lack consensus eukaryotic splice sites. The germline transcription may be
10 required for gVLR rearrangement, as is the case in mammalian antibody class switch recombination for which germline switch region transcription is obligatory (Bottaro et al., 1994; Hein et al., 1998).

VLR Phylogeny

A phylogenetic analysis of the agnathan VLR proteins reveals three distinct
15 clusters respectively composed by lamprey VLR, hagfish VLR-A and VLR-B sequences (Fig. 12C). The hagfish VLR-B and lamprey VLR cluster in a separate branch from that with the hagfish VLR-A. The same tree topology was seen when only the VLR diversity regions, LRRNT to LRRCT or LRR1 to CP, were aligned. Hence, either the hagfish VLR-A arose by duplication of the ancestral gene (Fig. 12D) or the
20 lamprey lost their VLR-A ortholog after the split between the hagfish and lamprey lineages, dating 499 ± 38 Myr ago in the Cambrian period (Hedges et al., 2001). It is also possible that a lamprey VLR-A ortholog exists, but was not detected in >18,000 cDNA sequences derived from lamprey lymphocyte-like cells (Pancer et al., 2004) because it is expressed at very low levels or in non-lymphoid cells.

25 The presence of VLRs in both of the extant cyclostome orders is indicative of strong evolutionary pressure for vertebrates to develop an anticipatory molecular recognition system. The analysis indicates that, within less than 40 million years in the Cambrian, two radically different systems evolved in agnathans and gnathostomes in which either LRR or Ig gene fragments undergo recombinatorial assembly to generate
30 diverse repertoires of lymphocyte receptors. This evolutionary scenario raises many

intriguing questions, one of which concerns the issue of whether the two adaptive immune strategies represent convergent evolution or if one was ancestral to the other. Whether VLRs were forerunner vertebrate immune receptors or the rearranging VLRs and Igs evolved independently will become certain only with an unambiguous resolution of the phylogenetic relationships among the groups of living and extinct jawless and jawed vertebrates (Mallatt et al., 2003; Meyer et al., 2003). In this regard, however, the presence of VLRs in both orders of contemporary agnathans lends additional molecular evidence favoring a monophyletic origin of cyclostomes.

Animals.

Live specimens of Pacific hagfish *Eptatretus stoutii* (30–60 cm long) were purchased from Marinus Scientific (Long Beach, CA) and maintained for two months at 12°C in artificial sea water (Oceanic System, Dallas, TX). Larvae (15–20 cm long) of the American brook lamprey (*Lampetra appendix*) and Northern brook lamprey (*Ichthyomyzon fossor*), were from tributaries to the Great Lakes (Lamprey Services, Ludington, MI).

Hagfish were sedated by immersion for 15 min in 0.5 gr/liter MS222 (Sigma, St. Louis, MO) buffered to pH=7 before intraperitoneal injection with an antigen/mitogen cocktail in 0.5 ml hagfish PBS (per liter: 28 gr NaCl, 0.2 gr KCL, 1.44 gr Na₂HPO₄, 0.24 gr KH₂PO₄, pH=7.4, 1 osmole). The cocktail contained 10⁹ live *E. coli* TG1 bacteria, 10⁹ sheep erythrocytes (Colorado Serum Company, Denver, CO) and 100 µg each phytohemagglutinin and pokeweed mitogen (Sigma). Immune stimulation was repeated at weekly intervals and four days after the fourth stimulation blood was collected with a syringe from the tail blood sinus and diluted 1:1 with hagfish PBS containing 30 mM EDTA. Buffy coat leukocytes collected after 5 min centrifugation at 50x g were sorted by their light scatter characteristics as described (Newton et al., 1994; Raison et al., 1994) using a MoFlo cytometer (Cytomation, Fort Collins, CO).

Hagfish VLR.

Inshore hagfish *Eptatretus burgeri* VLR homologs were identified using lamprey VLR as BLAST queries against the database of expressed sequence tags from leukocyte RNA of unstimulated animals #7,8 (Suzuki et al., 2004B). Clones with significant

matches were sequenced on both strands: 64 VLR-A and 15 VLR-B cDNA clones. For the Pacific hagfish, unseparated blood cells and buffy coat leukocytes from three unstimulated individuals (#1-3,6), and buffy coat leukocytes from two immunostimulated animals (#4,5) were used for extraction of blood genomic DNA and leukocyte RNA. Extraction of RNA was with TRIzol Reagent (Invitrogen, Carlsbad, CA) and PolyA RNA was selected with Dynabeads mRNA purification Kit (Dyna-
 5 Biotech, Lake Success, NY). First strand cDNA synthesis was primed with 20 pmoles of the HgVLRA.F1 (Table 5) for VLR-A, or HgVLRB.F1 for VLR-B, using SuperScript III First Strand cDNA Synthesis kit (Invitrogen), and the products were
 10 column purified (QIAquick PCR purification; QIAGEN, Valencia, CA).

Table 5. VLR PCR primers

Name	Sequence 5'-3'	Position in cDNA clone	Position in Eb_gVLR Contig
HgVLRA.F1	TGGTGATAACCTCAAGGTGCT (SEQ ID NO: 322)	35-55 (Eb7VLRA.21)	9597-9614
HgVLRA.F2	CAGAGATGATGGGTCCGGT (SEQ ID NO: 323)	60-78 (Eb7VLRA.21)	15509-15527
HgVLRA.R1	GGCAAGTGAGACACTGGTTC (SEQ ID NO: 324)	1023-1042 (Eb7VLRA.21)	16166-16185
HgVLRA.R2	TCTTGAGAAAGTGGAAGACGTA (SEQ ID NO: 325)	995-1016 (Eb7VLRA.21)	16138-16159
HgVLRB.F1	CACGAGGATTGCACGTGAAGA (SEQ ID NO: 326)	49-69 (Eb7VLRB.15)	59421-59441
HgVLRB.F2	TTCCACCTCGAGGAAGATGA (SEQ ID NO: 327)	93-112 (Eb7VLRB.15)	59677-59696
HgVLRB.R1	GGCAAAATGTTGGACGGTGT (SEQ ID NO: 328)	866-885 (Eb7VLRB.15)	60116-60135
HgVLRB.R2	GGCGTGACATATGAGGTAAAC (SEQ ID NO: 329)	826-846 (Eb7VLRB.15)	60076-60096
Slit.F	CTCGGCTCTGCAGCTCTCA (SEQ ID NO: 330)	1-19 (LaVLR.2)	
LRR_N.F1	CTCCGCTACTCGGCCTGCA (SEQ ID NO: 331)	1-19 (IfVLR.15)	
VLR_3UT.R	GATGAAGCGAAGACAGACGTG (SEQ ID NO: 332)	1607-1627 (LaVLR.2)	
VLR_3UT.R	GATGAAGCGAAGACAGACGTG (SEQ ID NO: 333)	1405-1425 (IfVLR.15)	

VLRs were then PCR amplified using Expand High Fidelity PCR (Roche Applied
 15 Science, Indianapolis, IN), from the cDNA or from genomic DNA, in 50 µl reactions containing: 1 µl each of the sets of forward and reverse primers (F1 or F2 and R1 or

R2) at 10 pmole/ μ l, 5 μ l 10x buffer, 36.25 μ l DDW, 5 μ l cDNA or genomic DNA (250 ng) and 0.75 μ l Expand enzyme. Reactions were amplified using one cycle of 94°C 1 min, then 35 cycles of 94°C 30 sec, 58°C 30 sec and 72°C 1 min, and a final 7 min elongation at 72°C. Products were column purified, cloned in pCRII-TOPO (Invitrogen) and the inserts were sequenced. For the Pacific hagfish, 109 VLR-A RT-PCR clones were sequenced (four contained in-frame stop codons), and 36 genomic mature VLR-A amplicons (two contained in-frame stop codons). For VLR-B, 37 RT-PCR clones were sequenced (one contained an in-frame stop codon), and 38 genomic mature VLR-B amplicons (four contained in-frame stop codons). Liver genomic DNA from Inshore hagfish #9 (Suzuki et al., 2004B) was used for PCR cloning and sequencing mature VLRs: 4 mature VLR-A amplicons (two contained in-frame stop codons) and 3 mature VLR-B amplicons.

Non-parasitic lamprey VLR

First strand cDNA was synthesized as above using the reverse primer VLR_3UT.R (Sea lamprey 3' UTR primer, Table 5). For the American brook lamprey the forward primer was Slit.F (Sea lamprey 5' UTR primer), and for the Northern brook lamprey LRR_N.F1 (another Sea lamprey 5' UTR primer). In total 13 unique VLR clones of the American brook lamprey and seven of the Northern brook lamprey were sequenced.

BAC libraries and clones.

An Inshore hagfish BAC library (Suzuki et al., 2004A) was screened by PCR using VLR primers as above (F1 or F2 and R1 or R2). The Pacific hagfish BAC library (VMRC23) was constructed from EcoRI partial digests of erythrocyte DNA from a single specimen in the vector pCCBACE1 (Epicentre Technologies, Madison WI). This library consists of ~184,000 recombinants and encompasses ~5x coverage of the hagfish genome. The entire library was screened by hybridization with 5' and 3' VLR-A and VLR-B probes and positive clones were authenticated by PCR. One BAC for each VLR type from the Pacific and Inshore hagfish were sequenced at ~10x coverage and assembled into contigs (Macrogen, Seoul, Korea). In case of incomplete sequence of the inserts only portions containing the gVLR and LRR cassettes were included with

uncaptured gaps in the contigs: Eb_gVLR-A, 43,362 bp; Eb_gVLR-B, 92,072 bp; Es_gVLR-A, 81,648 bp; Es_gVLR-B, 76,730 bp.

Sequence analysis

Neighbor Joining and UPGMA trees were constructed with the pairwise
5 deletion option using the programs from MEGA 3 Molecular Evolutionary Genetics Analysis (Kumar et al., 2004). Prediction of genes in the BAC inserts was accomplished by using local BLAST downloaded from <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/> and the GenScan server: genes.mit.edu/GENSCAN.html.

Throughout this application, various publications are referenced. The
10 disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains. The references disclosed are also individually and specifically incorporated by reference herein for the material contained in them that is discussed in the sentence in which the reference is relied upon.

15 It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of the invention. Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as
20 exemplary only, with a true scope and spirit of the invention being indicated by the following claims.

References

- 25 Anderson MK, Sun X, Miracle AL, Litman GW and Rothenberg EV (2001) Evolution of hematopoiesis: Three members of the PU.1 transcription factor family in a cartilaginous fish, *Raja eglanteria*. *Proc. Natl. Acad. Sci. USA* 98:553-8
- Ardavin CF and Zapata A (1987) Ultrastructure and changes during metamorphosis of the lympho-hemopoietic tissue of the larval anadromous sea lamprey *Petromyzon marinus*. *Dev. Comp. Immunol.*, 11:79-93
30
- Azumi K et al.,. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics* 55: 570-81, 2003

- Bell, JK., Mullen, GED., Leifer, CA. Mazzoni, A., Davies, DR. and Segal, DM. Leucine-rich repeats and pathogen recognition in Toll-like receptors. *Trends in Immunology* 2003, 24: 528-533.
- 5 Beutler, B. Innate immunity: an overview. *Molecular Immunology* 40 (2004) 845–859.
- Bryan, M.B., Libants, S.V., Warrillow, J.A., Li, W. and Scribner, K.T. Polymorphic microsatellite markers for the landlocked sea lamprey, *Petromyzon marinus*. *Conservation Genetics* 4: 113–116, 2003
- 10 Bottaro, A., Lansford, R., Xu, L., Zhang, J., Rothman, P. & Alt, F. W. (1994) *EMBO J.* 13, 665-674.
- Chamaillard, M., Girardin, SE., Viala, J. and Philpott, DJ. Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation. *Cellular Microbiology* 15 (2003) 5: 581–592.
- Cooper AJ (1971) Ammocoete lymphoid cell populations in vitro. In: *4th Leukocyte Culture Conference*. O.R. McIntyre (Ed). New York Appleton Century-Crofts. pp. 137-47
- 20 Donelson JE. Antigenic variation and the African trypanosome genome. *Acta Trop.* 2003, 85: 391-404.
- 25 Finstad J and Good RA (1964) The evolution of the immune response. III. Immunologic responses in the lamprey. *J. Exp. Med.*, 120: 1151-67
- Finstad J, Papermaster BW and Good RA (1964) Evolution of the immune response. II. Morphologic studies of the thymus and organized lymphoid tissue. *Lab Invest.*, 13:490-512
- 30 Flajnik MF and Kasahara M (2001) Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* 15:351-62
- 35 Flajnik MF (2002) Comparative analyses of immunoglobulin genes: surprises and portents. *Nat. Rev. Immunol.*, 2:688-98
- Forey PL and Janvier P (1993) Agnathans and the origin of jawed vertebrates. *Nature* 361:129–134
- 40 Fujii T (1982) Electron microscopy of the leukocytes of the typhlosole in ammocoetes, with special attention to the antibody-producing cells. *J. Morphol.*, 173:87-100
- Fujii T and Hayakawa I (1983) A histological and electron-microscopic study of the cell types involved in rejection of skin allografts in ammocoetes. *Cell Tissue Res.*, 231:301-
- 45

12

Good, R. A., Finstad, J. & Litman, G. W. in *The biology of lampreys II: Immunology* (Eds Hardisty, M.V. & Potter, I.C.) 405-432 (Academic Press, London 1972).

5

Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, Storset A, Saebo S, Stet RJ. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics*. 55:210-9, 2003

10

Hagen M, Filosa MF and Youson JH (1985) The immune response in adult sea lamprey (*Petromyzon marinus* L.): the effect of temperature. *Comp. Biochem. Physiol.*, 82:207-10

15

Haire RN, Miracle AL, Rast JP and Litman GW (2000) Members of the Ikaros gene family are present in early representative vertebrates. *J. Immunol.*, 165:306-12

20

Hamrick TS, Dempsey JA, Cohen MS, Cannon JG. Antigenic variation of gonococcal pilin expression in vivo: analysis of the strain FA1090 pilin repertoire and identification of the pilS gene copies recombining with pilE during experimental human infection. *Microbiology* 2001, 147: 839-49.

25

Hedges, S. B. (2001) in *Major events in early vertebrate evolution, Systematics Association special vol. 61: Molecular evidence for the early history of living vertebrates*, ed Ahlberg, P. E.(Taylor & Francis, London), pp. 119-134.

Hein, K., Lorenz, M. G., Siebenkotten, G., Petry, K., Christine, R. & Radbruch, A (1998) *J. Exp. Med.* 188, 2369-2374.

30

Ikezawa, H. Glycosylphosphatidylinositol (GPI)-Anchored Proteins. *Biol. Pharm. Bull.* 25: 409- 417 (2002)

35

Jones, D.A. and Takemoto, D. Plant innate immunity – direct and indirect recognition of general and specific pathogen-associated molecules. *Current Opinion in Immunology* 2004, 16:48–62

Kaufman J (2002) The origins of the adaptive immune system: whatever next? *Nat. Immunol.*, 3:1124-5

40

Kilarski W and Plytycz B (1981) The presence of plasma cells in the lamprey (Agnatha). *Dev. Comp. Immunol.*, 5:361-6

45

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, AR Laird DJ, De Tomaso AW, Cooper MD and Weissman IL (2000) 50 million years of chordate

- evolution: seeking the origins of adaptive immunity. *Proc. Natl. Acad. Sci., USA* 97:6924-6
- 5 Kumar, S., Tamura, K. & Nei, M. (2004) *Brief. Bioinform.* 5, 150-163.
- Landmann, R., Müller, B. and Zimmerli, W. CD14, new aspects of ligand and signal diversity. *Microbes and Infection*, 2, 2000, 295-304.
- 10 Litman GW, Frommel D, Finstad FJ, Howell J, Pollara BW and Good RA (1970) The evolution of the immune response. VIII. Structural studies of the lamprey immunoglobulin. *J. Immunol.*, 105:1278-85
- Mallatt, J. & Chen, J. Y. (2003) *J. Morphol.* 258, 1-31.
- 15 Marchalonis JJ and Edelman GM (1968) Phylogenetic origins of antibody structure. 3. Antibodies in the primary immune response of the sea lamprey, *Petromyzon marinus*. *J. Exp. Med.*, 127:891-914
- 20 Mayer WE, Uinuk-Ool T, Tichy H, Gartland LA, Klein J and Cooper MD (2002 a) Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl. Acad. Sci., USA* 99:14350-5
- Mayer WE, O'Huigin C, Tichy H, Terzic J and Saraga-Babic M (2002 b) Identification of two Ikaros-like transcription factors in lamprey. *Scand. J. Immunol.*, 55:162-70
- 25 Meyer, A. & Zardoya, R. (2003) *Annu. Rev. Ecol. Evol. Syst.* 34, 311-338.
- Newton, R. A., Raftos, D. A., Raison, R. L. & Geczy, C. L. (1994) *Dev. Comp. Immunol.* 18, 295-303.
- 30 Pancer, Z., Mayer, W. E., Klein, J. & Cooper, M. D. (2004) *Proc. Natl. Acad. Sci. USA* 101, 13273-13278.
- Penn DJ, Damjanovich K, Potts WK. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci* 99:11260-42002, 2002
- 35 Perey DY, Finstad J, Pollara B and Good RA (1968) Evolution of the immune response. VI. First and second set skin homograft rejections in primitive fishes. *Lab. Invest.*, 19:591-7
- 40 Piavis GW and Hiatt JL (1971) Blood cell lineage in the sea lamprey *Petromyzon marinus* (Pisces: Petromyzontidae). *Copeia* 4:722-8
- Pollara B, Litman GW, Finstad J, Howell J and Good RA (1970) The evolution of the immune response. VII. Antibody to human "O" cells and properties of the
- 45

- immunoglobulin in lamprey. *J. Immunol.*, 105:738-45
- Raison, R. L., Coverley, J., Hook, J. W., Towns, P., Weston, K. M. & Raftos, D. A (1994) *Immunol. Cell Biol.* 72, 326-332.
- 5 Rast, J.P., Michele K. Anderson, M.K., Strong, S.J., Luer, C., Litman, R.T., and Litman, G.W. α , β , γ , and δ T Cell Antigen Receptor Genes Arose Early in Vertebrate Phylogeny. *Immunity*, 6:1-11, 1997.
- 10 Schluter SF, Bernstein RM, Bernstein H and Marchalonis JJ (1999) 'Big Bang' emergence of the combinatorial immune system. *Dev. Comp. Immunol.*, 23:107-11
- Schwede, T., Diemand, A. Guex, N. and Peitsch, M.V. Protein structure computing in the
 15 genomic era. *Research in Microbiology* 151: 107-112 (2000)
- Shintani S, Terzic J, Sato A, Saraga-Babic M, O'hUigin C, Tichy H and Klein J (2000) Do lampreys have lymphocytes? The Spi evidence. *Proc. Natl. Acad. Sci., USA* 97:7417-22
- 20 Suzuki, T., Ota, T., Fujiyama, A. & Kasahara, M. (2004A) *Genes Genet. Syst.* 79, 251-253.
- Suzuki, T., Shin-I, T., Kohara, Y. & Kasahara, M. (2004B) *Dev. Comp. Immunol.* 28,
 25 993-1003.
- Uinuk-Ool T, Mayer WE, Sato A, Dongak R, Cooper MD and Klein J (2002) Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc. Natl. Acad. Sci., USA* 99:14356-61
- 30 Uinuk-Ool TS, Mayer WE, Sato A, Takezaki N, Benyon L, Cooper MD and Klein J (2003) Identification and characterization of a TAP-family gene in the lamprey. *Immunogenetics* 55:38-48
- 35 Wang D, Botkin DJ, Norris SJ. Characterization of the vls antigenic variation loci of the Lyme disease spirochaetes *Borrelia garinii* Ip90 and *Borrelia afzelii* ACAI. *Mol. Microbiol.* 2003, 47: 1407-17.
- Zapata A, Ardavin CF, Gomariz RP and Leceta J (1981) Plasma cells in the ammocoete of *Petromyzon marinus*. *Cell Tissue Res.*, 221:203-8.
- 40

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 73

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 73

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

What is claimed is:

1. An isolated polypeptide comprising an N-terminal leucine rich repeat (LLRNT), one or more leucine rich repeats (LRRs), a C-terminal leucine rich repeat (LLRCT), and a connecting peptide, wherein the connecting peptide comprises an alpha helix.
2. The polypeptide of claim 1, wherein the connecting peptide is linked to the LLRCT.
3. The polypeptide of claim 1, further comprising a stalk region and a glycosyl-phosphatidyl-inositol anchor.
4. The polypeptide of claim 3, further comprising a hydrophobic tail.
5. The polypeptide of claim 3, wherein the stalk region comprises a threonin-proline rich region.
6. The polypeptide of claim 1, further comprising a signal peptide.
7. The polypeptide of claim 1, wherein there are 1-9 LRRs, with LRR1 adjacent to the LLRNT.
8. The polypeptide of claim 7, wherein LRR1 comprises less than about 20 amino acids.
9. The polypeptide of claim 7, wherein LRR1 comprises about 18 amino acids.
10. The polypeptide of claim 7, wherein each of LRR 2-9 comprises less than about 25 amino acids.

ises less than about 40

amino acids.

12. The polypeptide of claim 11, wherein the LLRNT comprises the amino acid sequence of SEQ ID NO:157.
13. The polypeptide of claim 12, wherein the LLRNT comprises the amino acid sequence of SEQ ID NO:157 with one or more conservative amino acid substitutions.
14. The polypeptide of claim 1, wherein the LLRCT comprises less than about 60 amino acids.
15. The polypeptide of claim 14, wherein the LLRCT comprises the amino acid sequence of SEQ ID NO:158.
16. The polypeptide of claim 14, wherein the LLRCT comprises the amino acid sequence of SEQ ID NO:158 with one or more conservative amino acid substitutions.
17. The polypeptide of claim 1, wherein the connecting peptide comprises less than about 15 amino acids.
18. The polypeptide of claim 1, wherein the LRRs differ in amino acid sequence from each other and from the LLRNT and the LLRCT.
19. The polypeptide of claim 1, wherein the polypeptide is about 130 to about 225 amino acids in length.
20. The polypeptide of claim 1, wherein the polypeptide selectively binds an agent.
21. The polypeptide of claim 20, wherein the agent is a pathogen.

22. The polypeptide of claim 21, wherein the pathogen is a bacterium.
23. The polypeptide of claim 20, wherein the agent is a toxin.
24. An isolated nucleic acid that encodes the polypeptide of claim 1.
25. An expression vector comprising the nucleic acid of claim 24 operably linked to an expression control sequence.
26. A cultured cell comprising the vector of claim 25.
27. The polypeptide of claim 1, wherein the polypeptide is bound to a solid support.
28. The polypeptide of claim 27, wherein the solid support is a mobile solid support.
29. The polypeptide of claim 27, wherein the solid support is a column.
30. The polypeptide of claim 27, wherein the solid support is a chip.
31. The polypeptide of claim 27, wherein the solid support is a multiwell plate.
32. A plurality of the polypeptides of claim 1, wherein the LRRs are highly variable across polypeptides.
33. The plurality of polypeptides of claim 32, wherein the plurality is bound to a solid support.
34. The plurality of polypeptide of claim 32, wherein the solid support is a mobile solid support.

lid support is a column.

36. The plurality of polypeptide of claim 32, wherein the solid support is a chip.

37. The plurality of polypeptide of claim 32, wherein the solid support is a multiwell plate.

38. A method of detecting an agent in a sample, comprising the steps of

- a. Contacting the sample with the polypeptide of claim 20 under conditions for the polypeptide to bind to the agent in the sample;
- b. Detecting the polypeptide bound to the agent in the sample, bound polypeptide indicating agent in the sample.

39. The method of claim 38, wherein the polypeptide is labeled with a detectable moiety.

40. The method of claim 38, further comprising quantification of the agent in the sample.

41. A method of blocking the activity of an agent, comprising contacting the agent with the polypeptide of claim 20 under conditions for the polypeptide to bind the agent, binding of the polypeptide to the agent blocking the activity of the agent.

42. The method of claim 38, wherein contacting step is *in vitro*.

43. The method of claim 38, wherein contacting step is *in vivo*.

44. An imaging method comprising administering to a subject an effective amount of the polypeptide of claim 1 and detecting the localization of the bound polypeptide in the subject.

45. The method of claim 44, wherein the polypeptide is labeled with a detectable moiety.
46. A method of purifying an agent from a sample comprising
- a. contacting the sample with the polypeptide of claim 20 under conditions for the polypeptide to bind the agent and form a polypeptide/agent complex;
 - b. isolating the polypeptide/agent complex from the sample;
 - c. isolating the agent from the polypeptide/agent complex.
47. A method of reducing or preventing a pathogenic effect in a subject comprising administering to the subject an effective amount of the polypeptide of claim 21.
48. A method of making a polypeptide comprising culturing the cell of claim 26 under condition permitting expression of the polypeptide and purifying the polypeptide from the cell or medium of the cell.
49. A method of screening for one or more variable lymphocyte receptors in a subject comprising identifying in the subject one or more polypeptides comprising an N-terminal leucine rich repeat (LLRNT), one or more leucine rich repeats (LRRs), a C-terminal leucine rich repeat (LLRCT), and a connecting peptide, wherein the connecting peptide comprises an alpha helix.
50. A method of determining the function of a gene comprising introducing a polypeptide specific for the protein product of the gene into the cytoplasm of a cell expressing the gene and monitoring the effect due to the loss of protein product of the gene with unknown function.

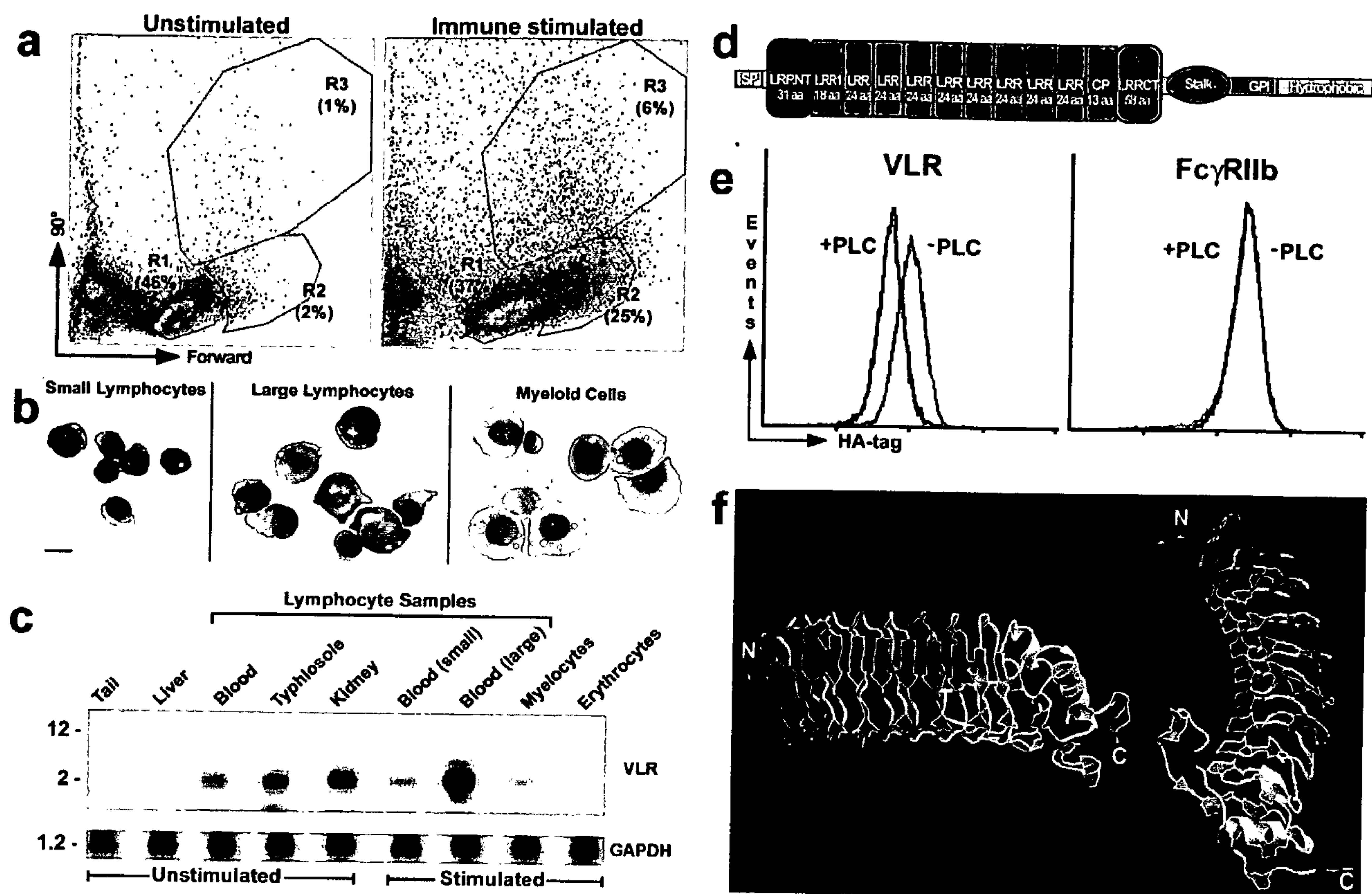
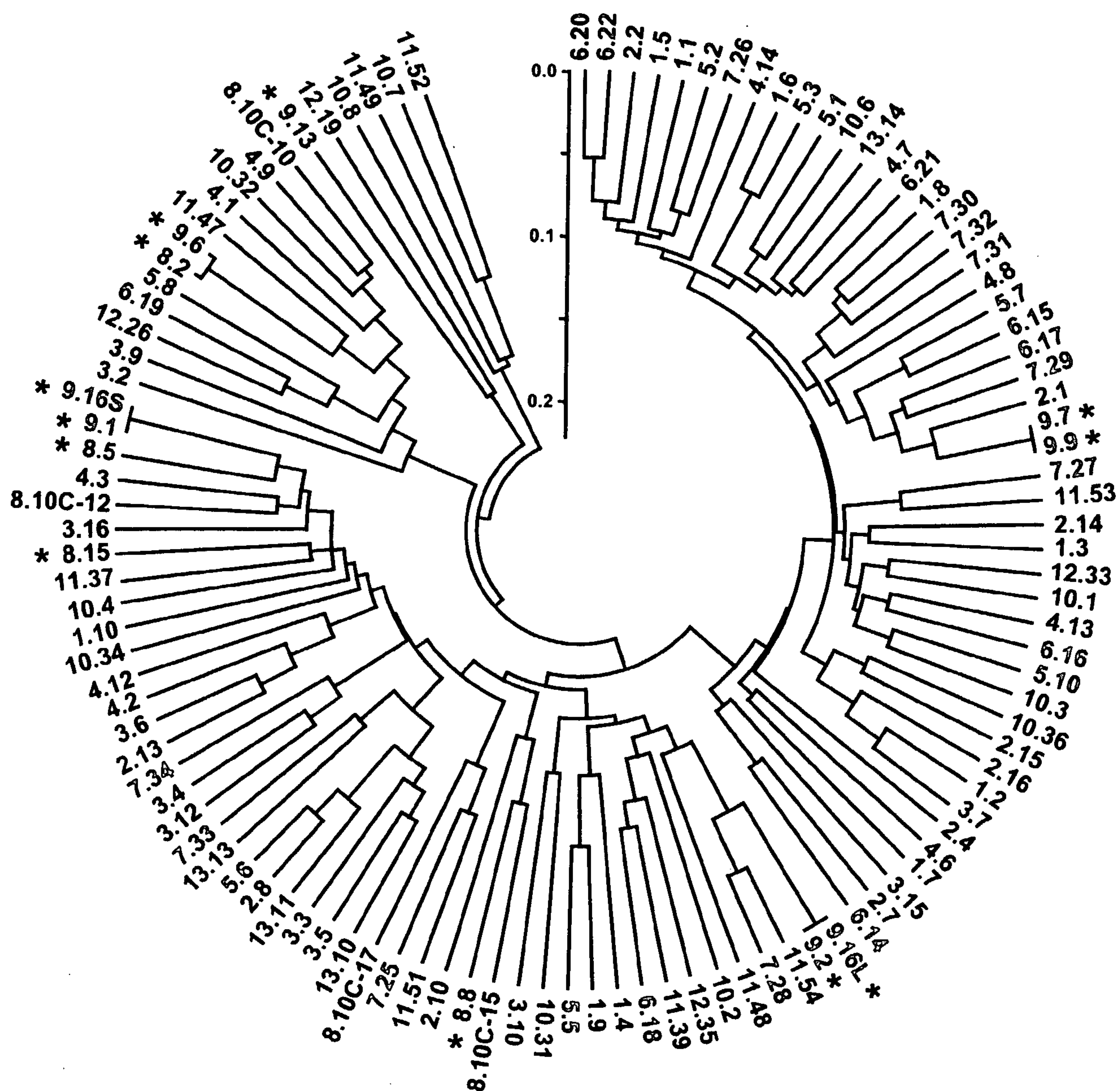


Figure 1

	<	LRRNT	><	LRR1	><	LRR2	><	LRR3	><	
1.3		GALVQSAVACPSQCSCSGTEVNCAGKSLASVPAGIPTTTRVLYLNSNQITKLEPGVFD		SLTALTYLGLGGHQLAALPENVFD						RL 84
1.8		GALVQSAVACPSQCSCSGTSDCRSRRHASVPAGIPTNAQITLYLHDNQITKLEPGVFD		SLTQLTILDLSHQQLALPAGLFD						RL 84
2.4		GALVQSAVACPSQCSCSGTSGASVECSRRTSVPAGIPINVQIFELYDNQITKLEPGVFD		DLVNLQQLYLGSHQLGALPVGVD		SLTQLTLYDLAPNQLALPEGVFD				RL 108
2.14		GALVQSAVACPSQCSCSGTEVNCQKKSLSASVPAGIPTNALNITWLNQITKLEPGVFD		SLTQLTLYDLAPNQLALPVGVD						RL 84
1.2		GALVQSAVACPSQCSCSGAEVRCVSKSLASVPAGIPTTQSSLSHYTQITKLEPGVFD								58
2.16		GALVQSAVACPSQCSCSGAEVRCVSKSLASVPAGIPTTQSSLSHYTQITKLEPGVFD		DLAQLTGLDLSHQFTALPAQVFD						RL 84
1.6		GALVQSAVACPSQCSCSGTEVNCQKKSLSASVPAGIPTTQVLYLHVNQITKLEPGVFD								58
1.7		GALVQSAVACPSQCSCDQITVYCHSRRLTSVPAGIPTNRQNTWLNQITKLEPGVFD		DLGNLQQLNLSNQLALPAGLFD		SLTQLTLYNLAVNQLALPAGLFD				RL 108
2.1		GALVQSAVACPSQCSCDQITVYCHSRRLTSVPAGIPTNRQNTWLNQITKLEPGVFD								58
2.2		GALVQSAVACPSQCSCDQITVYCHSRRLTSVPAGIPTDRQNTWLYDNQITKLEPGVFD								60
2.7		GALVQSAVACPSQCSCDQITVYCHSRRLTSVPAGIPTDRQNTWLNQITKLEPGVFD		DLVNLQQLYLSGHQLALPEGVFD						RL 84
1.5		GALVQSAVACPSQCSCDQILVNCQTRLSASVPAGIPTTTRVLYLNSNQITKLEPGVFD								RL 60
1.1		GALVQSAVACPSQCSCSATVNCDSRLASVPAEIPITTKILRLVYNQITKLEPGVFD								RL 60
2.15		GALVQSAVACPSQCSCSGTEVRCESRLASVPAGIPTTTRWDLHLRNQITKLEPGVFD								58
1.4		GALVQSAVACPSQCSCDQITVDNRNKRFSVPAGIPTDRQNTWLNQITKLEPGVFD								RL 60
1.9		GALVQSAVACPSQCSCDQITVYCHSRRLTSVPAGIPTNRQNTWLYDNQITKLEPGVFD								58
2.8		GALVQSAVACPSQCSCSGTQVNCERRLSASVPAGIPTTQVLYLYTNQITKLEPGVFD								SL 60
2.13		GALVQSAVACPSQCSCSGTQVNCDSRLASVPGGIPITTQVLYLYDNQITKLEPGVFD								58
1.10		GALVQSAVACPSQCSCSGTEVNCQKKSLSASVPAGIPTTTRVLYLHTNQITKLEPGVFD								SL 60
2.10		GALVQSAVACPSQCSCSGTEVNCQKKSLSASVPAGIPTTQVLYLHVNQITKLEPGVFD		SLVNLQRLHLDQNLVSLPAGVFD						RL 84
		LRR4	><	LRR5	><	CP	><	LRRCT		
1.3		TQITRILQDYNTQITVLEAGVCSLVNKEIRIYNNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNEH		PHYGGVDN		VKCSGTNTPVR		187
1.8		VNIQOILWPEIQISALEVGVEDNLTQISINMHTNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNLQ		GHGGVDN		VKCSGTNTPVR		187
2.4		VNIQOILYIGSTQIGALPTWEDKILQITQIDNNNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNEP		GHGGVDN		VKCSGTNTPVR		211
2.14		VNIQRIWQWNTQITSLPAGVEDRLVNLQITQIDHNNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNPS		GNGGVDN		VKCSGTNTPVR		187
1.2		RLVNLQQLYLGSHQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNLR		GHGGVDN		VKCSGTNTPVR		139
2.16		VNIQILHILNIPDKRFEGGAEKILRLKRVQHTNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNEH		PHGGVDN		VKCSGTNTPVR		187
1.6		RLVNLQRIYNNQNLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNPS		GHGGVDN		VKCSGTNTPVR		139
1.7		GNIEVIGITCCNITELSCVEDKILRLKRWGQDNQLKSI		PDGAFRLSLTHIYLYGNPNWDCECSDIILYLNWIVOHASIVNPS		NGGVDN		VKCSGTNTPVR		210
2.1		KILQITHTISTYNNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNPS		HYGGVDN		VKCSGTNTPVR		139
2.2		VNIQHHTHNSNKTATPAGVEDKILQITQIDHVNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNEH		PHGGVDN		VKCSGTNTPVR		163
2.7		INIKEMYFSNQTSLPAREVDKILQITQIDHNDQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNEH		PHGGVDN		VKCSGTNTPVR		187
1.5		LNQOILYCHLIRSSIPAGVEDKILPKLTHIVHTNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNPS		GYGGVDN		VKCSGTNTPVR		163
1.1		VNIQHHTHNSNKTATPAGVEDKILQITQIDHVNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNPL		GNGGVDN		VKCSGTNTPVR		163
2.15		KILKLTHTIYLYGYNQLKSI		RGAFDNLKSLTHIYLSNPNWDCECSDIILYLNWIVOHASIVNEH		GNGGVDN		VKCSGTNTPVR		138
1.4		TQITHTIDDRNQLKSLPSTEDKLEKLTREIYNNQLTTPV		EGAFNSLMKLOYLHLSNPNWDCA		CSDIILYLSGWLQAHAGKEQQ		AVCSGTNTPVR		156
1.9		KILQITHTIYVHTNQLKSI		RGAFDNLKSLTHIYLYGNPNWDCA		CDIILYLSGWLQAHAGKEQQ		AVCSGTNTPVR		139
2.8		TATITFNTGNTQITALETVEDNLTQISINMHTNQLKSI		RGAFDNLKSLTHIYLSNPNWDCA		CSDIILYLSRTHSQHPGVRTA		DDWSRVVPSARCSGTNTPVR		167
2.13		SLAATITFNTGNTQITALETVEDNLTQISINMHTNQLKSI		RGAFDNLKSLTHIYLSNPNWDCA		CSDIILYLSRTHSQHPGVRTA		DDWSRVVPSARCSGTNTPVR		140
1.10		TQITVIESIPTTHIQALPQVEDKILQITQIDHNNQLKSI		RGAFDNLKSLTHIYLSNPNWDCA		CSDIILYLSRTHSQHPGVRLN		AGSYNIPDQAHCSGTNTPVR		165
2.10		TQITRILQDYNTQITVLEAGVCSLVNKEIRIYNNQLKSI		RGAFDNLKSLTHIYLSNPNWDCA		CDIILYLSGWLQAHAGKEQQ		AVCSGTNTPVR		192

Figure 2



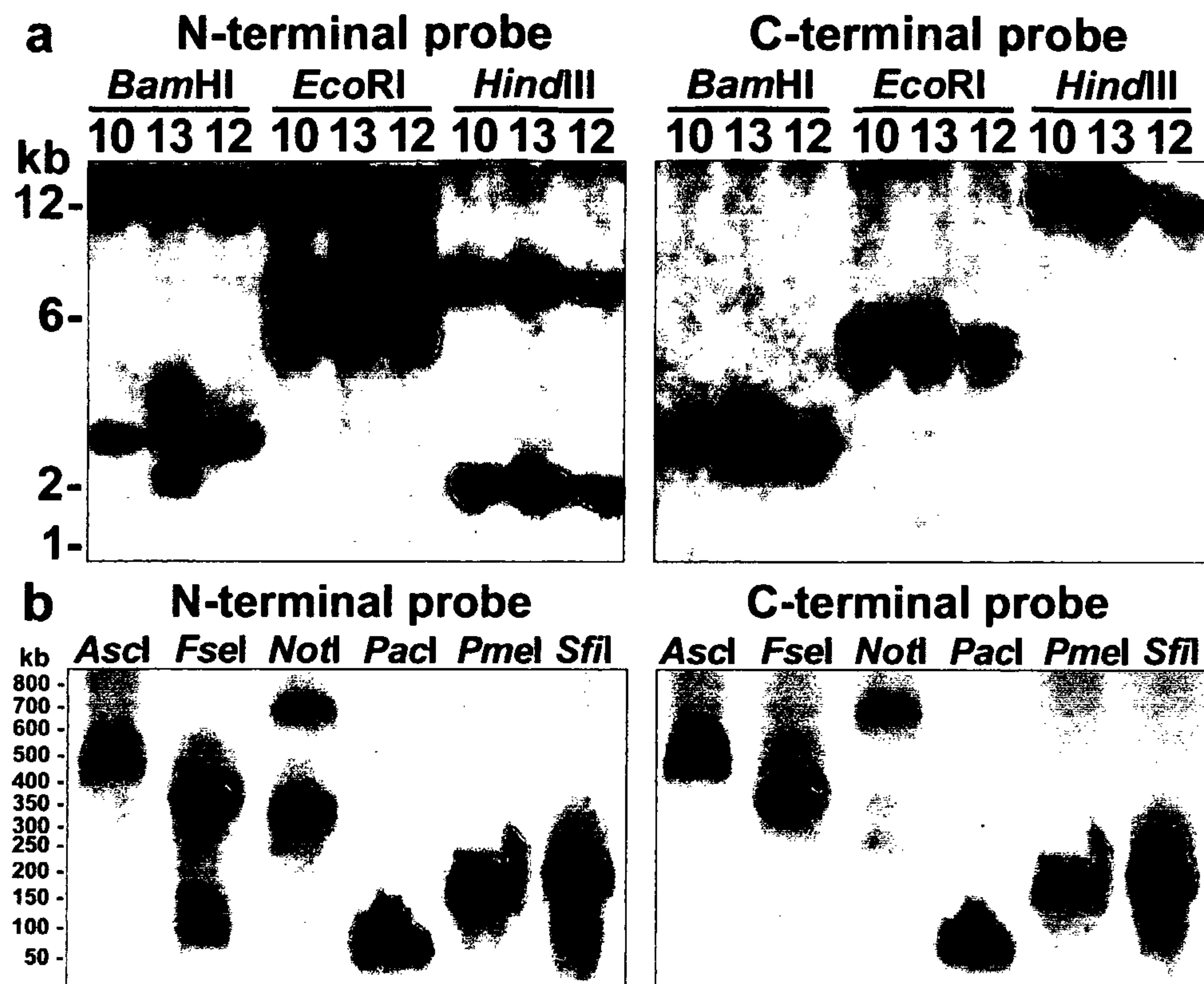


Figure 4

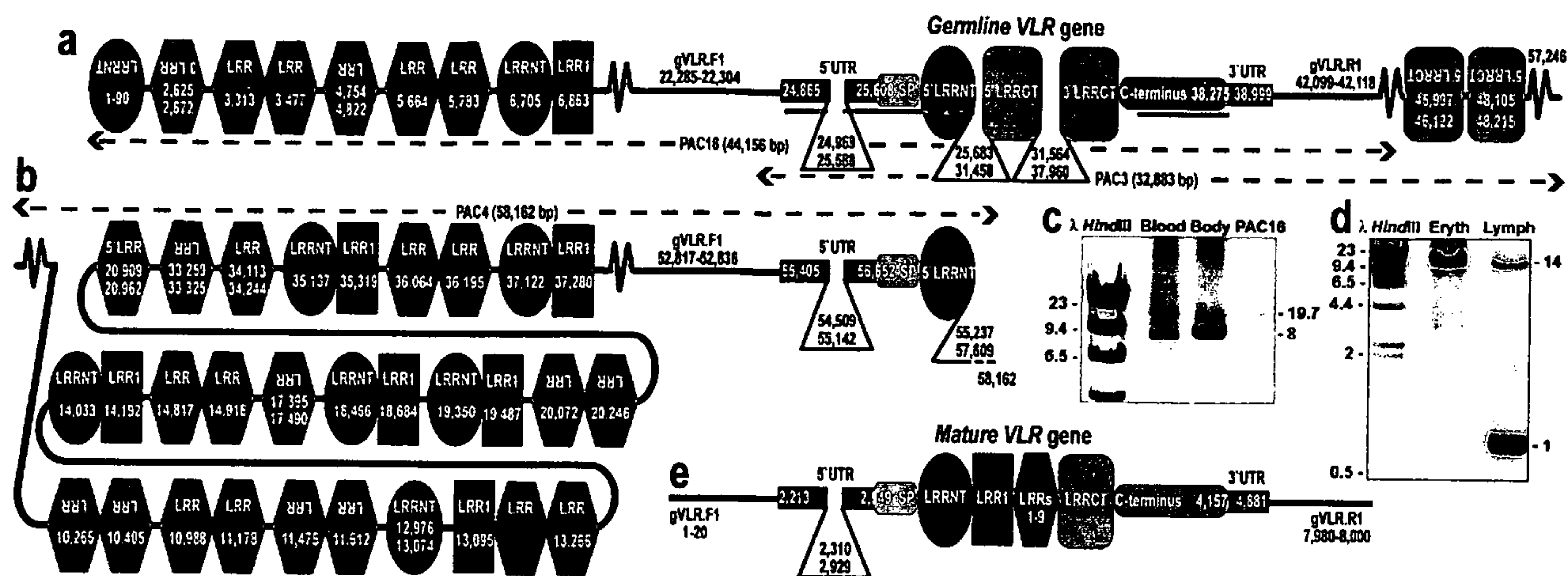


Figure 5

Ly	EST	Accession	Protein	Length
LyEST12061	12061	AF012061	Protein	70
LyEST13098	13098	AF013098	Protein	70
LyEST1703	1703	AF011703	Protein	94
LyEST4217	4217	AF014217	Protein	94
LyEST8432	8432	AF018432	Protein	94
LyEST2123	2123	AF012123	Protein	94
LyEST10428	10428	AF0110428	Protein	94
LyEST481	481	AF010481	Protein	94
LyEST3372	3372	AF013372	Protein	94
LyEST5220	5220	AF015220	Protein	94
LyEST8756	8756	AF018756	Protein	94
LyEST8564	8564	AF018564	Protein	94
LyEST2913	2913	AF012913	Protein	94
LyEST11913	11913	AF0111913	Protein	98
LyEST11865	11865	AF0111865	Protein	111
LyEST10645	10645	AF0110645	Protein	111
LyEST8780	8780	AF018780	Protein	111
LyEST12976	12976	AF0112976	Protein	118
LyEST256	256	AF010256	Protein	118
LyEST8138	8138	AF018138	Protein	111
LyEST5266	5266	AF0105266	Protein	111
LyEST3090	3090	AF0103090	Protein	130
LyEST12061	12061	AF012061	Protein	131
LyEST13098	13098	AF013098	Protein	133
LyEST1703	1703	AF011703	Protein	157
LyEST4217	4217	AF014217	Protein	157
LyEST8432	8432	AF018432	Protein	155
LyEST2123	2123	AF012123	Protein	155
LyEST10428	10428	AF0110428	Protein	155
LyEST481	481	AF010481	Protein	155
LyEST3372	3372	AF013372	Protein	157
LyEST5220	5220	AF015220	Protein	157
LyEST8756	8756	AF018756	Protein	155
LyEST8564	8564	AF018564	Protein	156
LyEST2913	2913	AF012913	Protein	156
LyEST11913	11913	AF0111913	Protein	159
LyEST11865	11865	AF0111865	Protein	179
LyEST10645	10645	AF0110645	Protein	179
LyEST8780	8780	AF018780	Protein	179
LyEST12976	12976	AF0112976	Protein	188
LyEST256	256	AF010256	Protein	187
LyEST8138	8138	AF018138	Protein	204
LyEST5266	5266	AF0105266	Protein	228
LyEST3090	3090	AF0103090	Protein	239
LyEST12061	12061	AF012061	Protein	241
LyEST13098	13098	AF013098	Protein	250
LyEST1703	1703	AF011703	Protein	235
LyEST4217	4217	AF014217	Protein	274
LyEST8432	8432	AF018432	Protein	233
LyEST2123	2123	AF012123	Protein	263
LyEST10428	10428	AF0110428	Protein	270
LyEST481	481	AF010481	Protein	270
LyEST3372	3372	AF013372	Protein	274
LyEST5220	5220	AF015220	Protein	273
LyEST8756	8756	AF018756	Protein	225
LyEST8564	8564	AF018564	Protein	227
LyEST2913	2913	AF012913	Protein	273
LyEST11913	11913	AF0111913	Protein	275
LyEST11865	11865	AF0111865	Protein	269
LyEST10645	10645	AF0110645	Protein	287
LyEST8780	8780	AF018780	Protein	277
LyEST12976	12976	AF0112976	Protein	234
LyEST256	256	AF010256	Protein	304
LyEST8138	8138	AF018138	Protein	287
LyEST5266	5266	AF0105266	Protein	272

Figure 6

1 M W I K W I A T L V A F G A L V Q S A V A C P S Q C S C D Q LRRNT
118 ATGTGGATCAAGTGGATCGCCACGCTGGTGGCTTTGGCGCCCTGGTGGTCAAGTGGTAGCATGTCCCTCGCAGTGTTCGTCGATCAG
Exon 2
31 T L V N C Q N I R L A S V P A G I P T D K Q R L W L N N Q LRR1
208 ACACTTGTGAAGTGGCCAGAAATATACGCCCTCGCATCTGTGCTGCGGGAATCCCAAGCAGAGGCTGTGTTGAACAACAATCAG
LRRNT
61 I T K L E P G V F D H L V N L Q Q L Y F N S N K L T A I P T LRR2
298 ATCAGGAAGCTTGAGCCCGGGGTGTTGACCATCTGTGATCTGTCAGCAGCTCTATTTTAACAGCAACAAGCTAACAGCTATACCCACT
LRR1
91 G V F D K L T Q L T Q L D L N D N H L K S I P R G A F D N L LRR3
388 GGGTGTGTAACAAGTCAACCCAGCTCACTCACTGATTTGAATGACAACCATCTGAAGAGCATTCACAGGGCGCCTTTGACAACCTC
Connecting peptide
121 K S L T H I Y L Y N P W D C E C R D I M Y L R N W V A D H LRRCT
478 AAGAGCCTAACTCACATCTATCTGTACAAACCCATGCGGATTCGGAGTGCAGGACATTAATGTACCTCAGGAAGTGGTGGCAGACAC
Exon 3
151 T S I V M R W D G K A V N D P D S A K C A G T N T P V R A V LRRCT
568 ACTTCTATTGTAATGCGCTGGGATGGGAAGCGCTTAACGACCCCGACTCTGCCAAGTGGCTGGTACCAATACCCCGTCCGTGCGGTC
Exon 4
181 T E A S T S P S K C P G Y V A T T T T P T T T P E F I P E
658 ACCGAGGCCAGCACTAGCCCTCGAATGCCAGGCTACGTGTGCTACGACCAACGCGCCGACGACCAACGCGCCGAAATTCATCCCTGAG

211 T T T S P Q P V I T T Q K P K P L W N F N C T S I Q E R K N
748 ACCACCACTCGCCGCGAGCCCGGTGATCACAAACCCAGAAACCCCAAGCCCTCTGTGGAATTTCAACTGCACCTCAATTCAGGAGGAGAAAC
GPI cleavage
241 D G G D C G K P A C T T L L N C A N F L S C L C S T C A L C
838 GACGGTGGCGACTGCGGAAAGCCCGCTGCACAACTCTCTGAACTGCGGGAATTTCTCAGCTGCTCTGCTGACCTGCGCCCTCTGTC

271 R K R *
928 AGGAACGTTGA

Figure 7

[illegible]

Figure 8

Figure 8 (cont.)

[illegible]

Figure 9

Figure 10

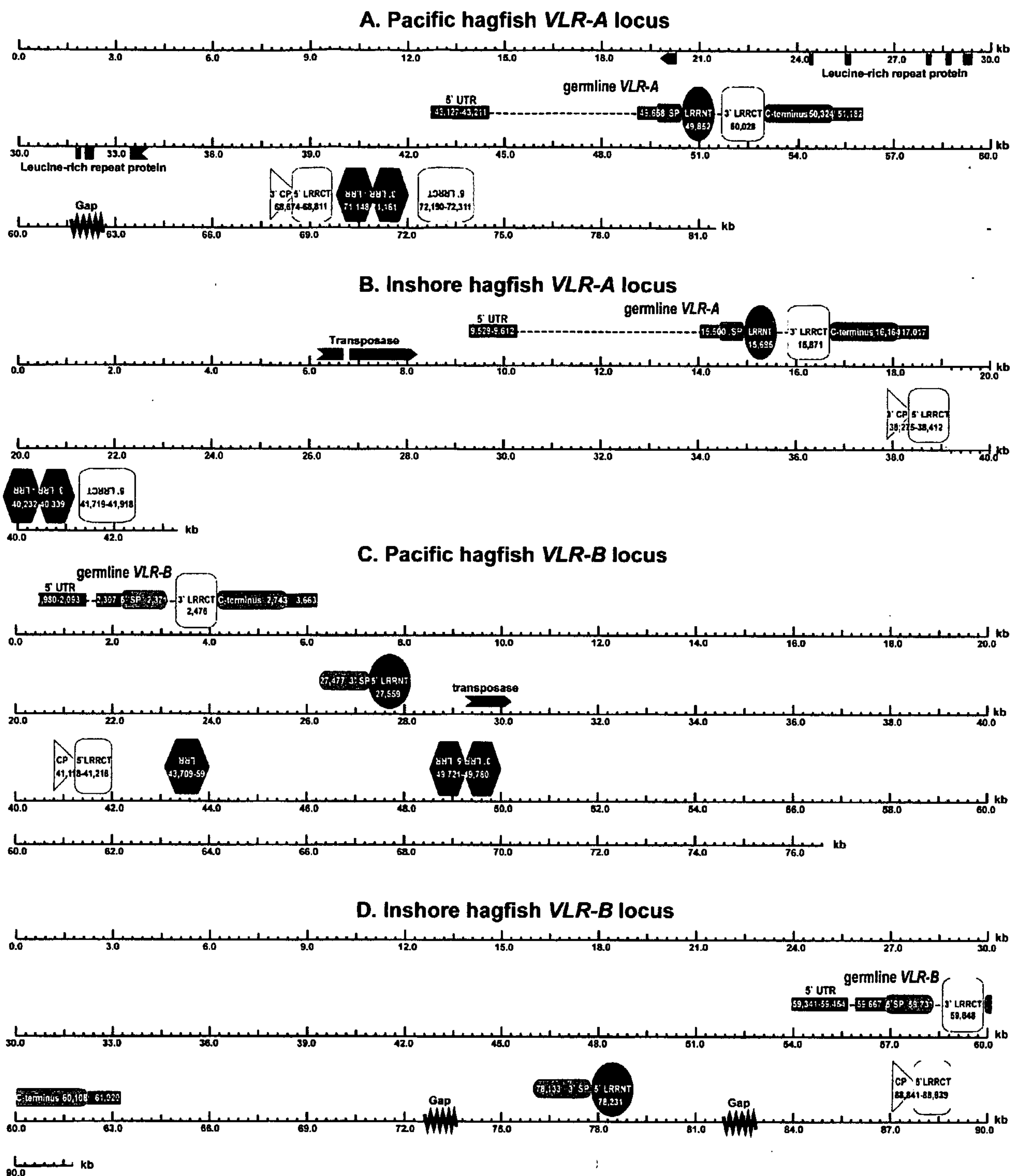


Figure 11

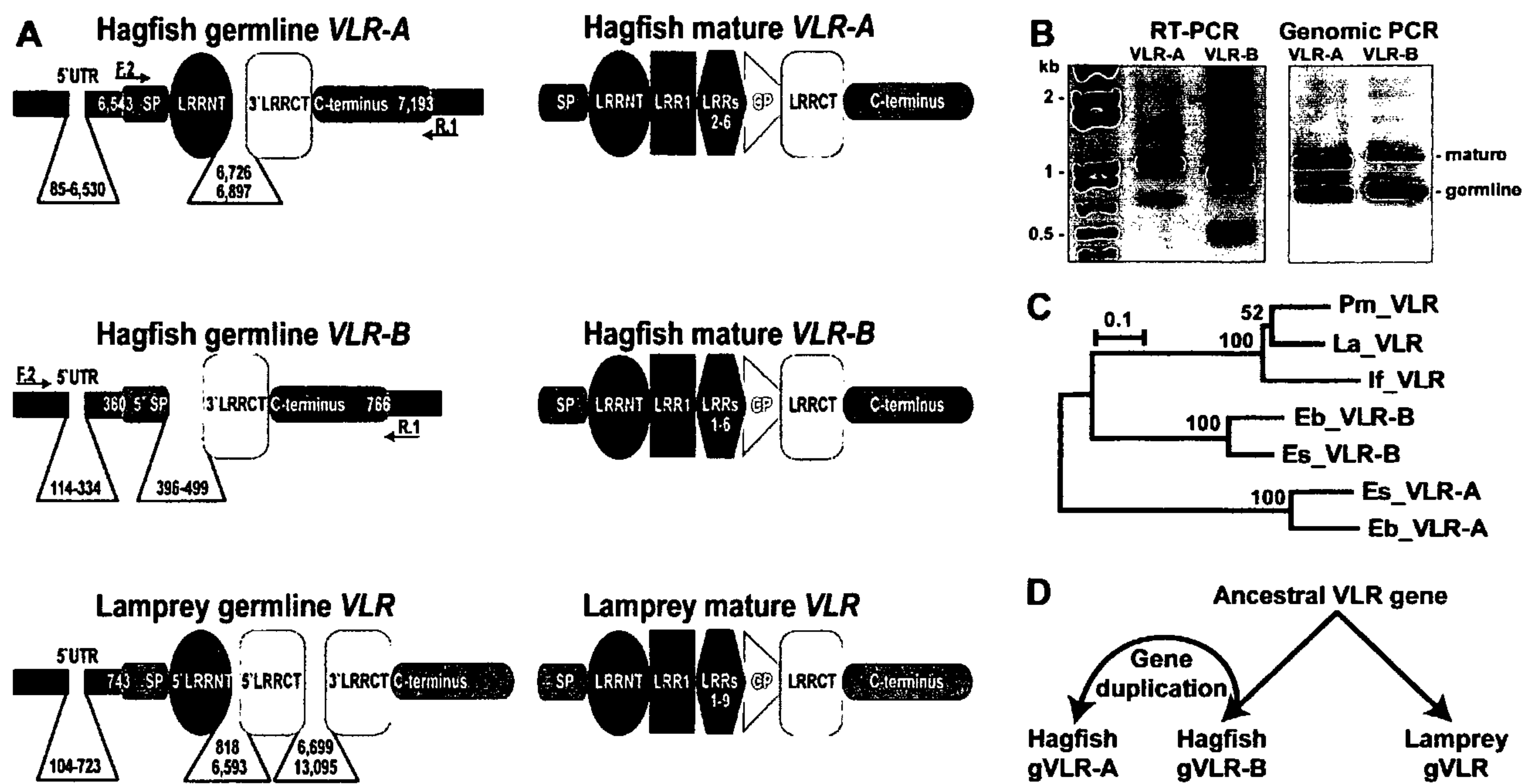


Figure 12

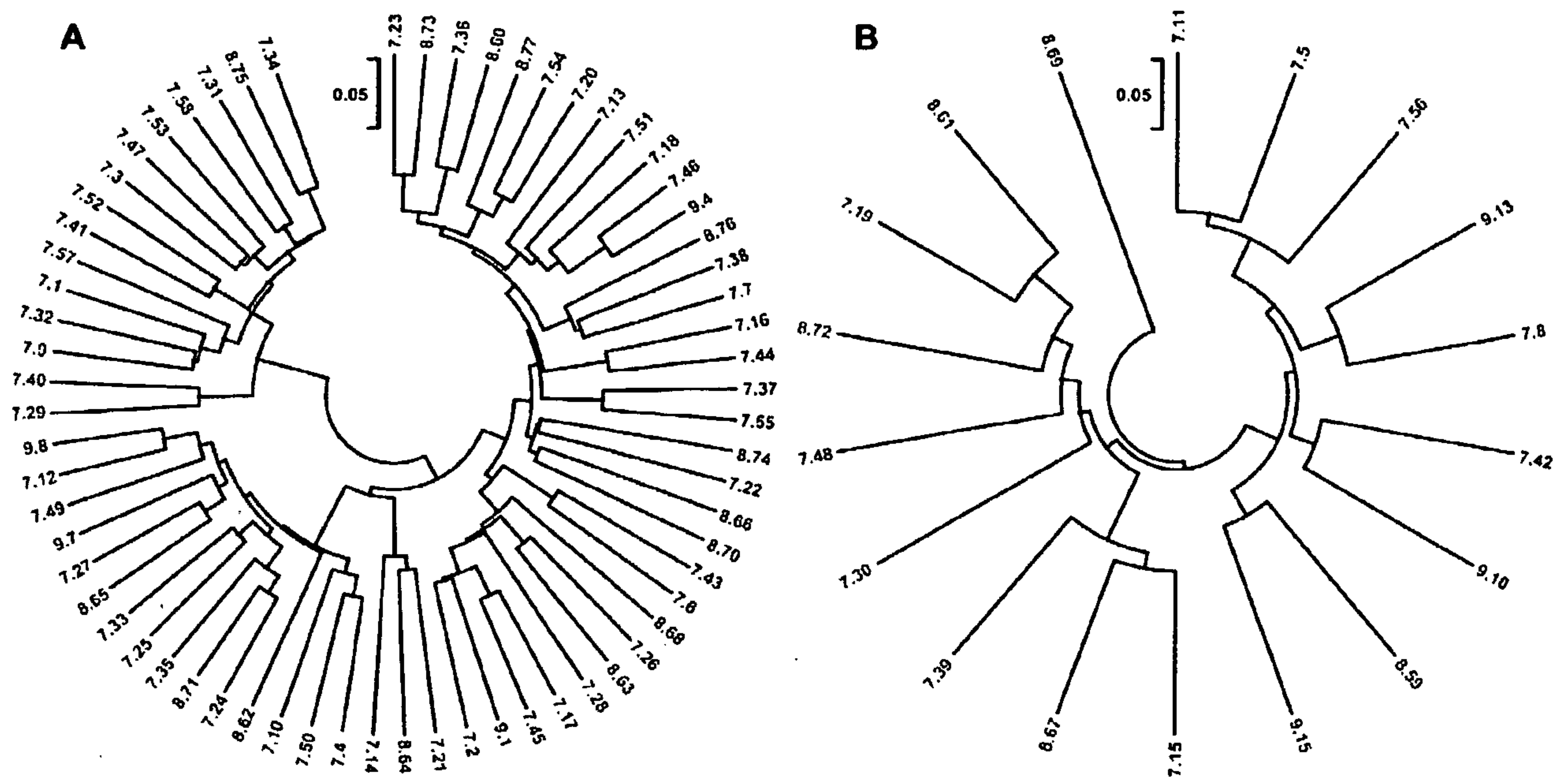


Figure 13