Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey

Zeev Pancer[1,2], **Chris T. Amemiya**[6], **Götz R. A. Ehrhardt**[1,5], **Jill Ceitlin**[7], **G. Larry Gartland**[1,4] **& Max D. Cooper**[1,2,3,4,5]

[1]Division of Developmental and Clinical Immunology, Departments of [2]Medicine, [3]Pediatrics and [4]Microbiology, and the [5]Howard Hughes Medical Institute, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA
[6]Molecular Genetics Program, Benaroya Research Institute at Virginia Mason, Seattle, Washington, 98101, USA
[7]University of Michigan, Ann Arbor, Michigan 48109, USA

....

Although jawless vertebrates are apparently capable of adaptive immune responses, they have not been found to possess the recombinatorial antigen receptors shared by all jawed vertebrates. Our search for the phylogenetic roots of adaptive immunity in the lamprey has instead identified a new type of variable lymphocyte receptors (VLRs) composed of highly diverse leucine-rich repeats (LRR) sandwiched between amino- and carboxy-terminal LRRs. An invariant stalk region tethers the VLRs to the cell surface by means of a glycosyl-phosphatidyl-inositol anchor. To generate rearranged VLR genes of the diversity necessary for an anticipatory immune system, the single lamprey *VLR* locus contains a large bank of diverse LRR cassettes, available for insertion into an incomplete germline VLR gene. Individual lymphocytes express a uniquely rearranged VLR gene in monoallelic fashion. Different evolutionary strategies were thus used to generate highly diverse lymphocyte receptors through rearrangement of LRR modules in agnathans (jawless fish) and of immunoglobulin gene segments in gnathostomes (jawed vertebrates).

Adaptive immune responses in jawed vertebrates are initiated when antigens are recognized by specific lymphocyte receptors. Antigen receptor diversity is generated through recombination of variable, diversity and joining gene segments in the immunoglobulin and T-cell receptor (TCR) gene loci. This combinatorial rearrangement generates vast repertoires of antibodies against unprocessed antigens and of TCRs that recognize antigen fragments presented by major histocompatibility complex (MHC) class I and II molecules. Clonally diverse lymphocytes thus form the cornerstone of vertebrate adaptive immunity in the form of immunoglobulin-bearing B cells and TCR-bearing T cells that differentiate from stem-cell precursors within primary haematopoietic tissues and the thymus. The principal elements of this recombinatorial immune system are conserved in all jawed vertebrates, and the multigene TCR and immunoglobulin loci are remarkably complex even in the most basal gnathostome representatives, such as sharks, skates and rays[1–3].

Adaptive immune responses have also been reported in the jawless vertebrates lamprey and hagfish, the only surviving descendents from the early vertebrate radiation[4]. For example, lampreys produce specific circulating agglutinins in response to immunization[5–10], reject second set skin allografts at an accelerated rate[9,11] and exhibit delayed type hypersensitivity reactions[5]. These responses have been attributed to agnathan cells that morphologically resemble the lymphocytes found in jawed vertebrates[5,9,11–16]. As with their mammalian counterparts, lamprey lymphocytes are more irradiation-sensitive than other blood cell types[9], aggregate and proliferate in response to antigenic stimulation[5,12], and express transcription factors such as PU.1/Spi-B and Ikaros that are involved in mammalian lymphocyte differentiation[17–20]. However, immunoglobulin, TCR and MHC genes have not been identified in jawless vertebrates nor in the draft genome sequence of the invertebrate urochordate *Ciona intestinalis*[21].

With this in mind, we initiated a search for primordial molecular elements of the vertebrate immune system in the sea lamprey *Petromyzon marinus*. In an earlier analysis of transcripts expressed by lymphocyte-like cells from lamprey haematopoietic tissues we identified several homologues of immune system molecules, but

none of the immunoglobulin superfamily receptor elements used by jawed vertebrates for specific adaptive immunity[16,22,23]. Reasoning that immunocompetent cells in the circulation would be more likely to express the genes involved in adaptive responses, we began with a survey of the transcriptome of activated lymphocytes from the blood of immunostimulated lamprey larvae. This search revealed a new type of highly variable lymphocyte receptors.

Immunostimulated blood lymphocyte transcripts

To survey the transcriptome of activated lymphocytes, lamprey larvae were stimulated by intraperitoneal injections of an antigen/mitogen cocktail two–four times at weekly intervals. The fraction of large lymphocytes among peripheral blood leukocytes 3 days after the second booster stimulation was 13-fold greater than in unstimulated individuals, and the fraction of myeloid cells was sixfold greater (Fig. 1a). Compared with the small lymphocyte-like cells, the activated lymphocytes were nearly double in size, had extensive azurophilic cytoplasm and featured prominent nucleoli (Fig. 1b). These cells were sorted on the basis of their light scatter characteristics and used to construct subtracted complementary DNA libraries enriched in messages of activated lymphocytes.

The most abundant group of sequences identified among 1,507 subtracted clones consisted of 319 proteins featuring variable numbers of diverse LRR motifs, which clustered with a set of 52 LRR-containing expressed sequence tags (ESTs) from our database of unstimulated lymphocyte-like transcripts. A set of 239 uniquely diverse LRR proteins was obtained after purging sequences with only the 3′ ends, and 22 ESTs from this set encoded most or all of the open reading frames (ORFs) of 239–304 residues in length (Supplementary Fig. 1). These 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 or exclusively by lymphocytes (Fig. 1c). Unstimulated animals showed highest VLR levels in lymphocytes from haematopoietic tissues, whereas immune stimulation resulted in enhanced VLR transcription by the large blood lymphocytes. The VLR transcripts observed in the myeloid lane of Fig. 1c may reflect

an approximately 5% contamination with activated lymphocytes.

The basic composition of VLRs includes 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/proline-rich stalk, a glycosyl-phosphatidyl-inositol (GPI)-anchor site and a hydrophobic tail (Fig. 1d; see also Supplementary Fig. 2). When a recombinant epitope-tagged VLR was expressed in a mammalian cell line, immunofluorescence analysis confirmed the cell-surface localization of the protein. Treatment with bacterial GPI-specific phospholipase C significantly reduced the level of surface VLR (Fig. 1e) and released the protein into the supernatant. The diversity region from the longest VLR sequence, consisting of 12 LRR modules, was threaded on the crystal structure coordinates of structurally related LRR proteins to generate a three-dimensional model[24]. This model predicts a concave solenoid structure capped on both ends by the LRRNT and LRRCT, in which LRRs 1–9 form β-sheets and the connecting peptide forms an α-helix (Fig. 1f), resembling the model predicted for Toll-like receptor (TLR) ecto-domains[25].

Individually diverse VLR repertoires

Blood leukocyte messenger RNA from three immunostimulated and four unstimulated larvae was used for polymerase chain reaction with reverse transcription (RT–PCR) amplification of

VLRs with primers flanking the diversity region (Supplementary Table 1). Sequencing 9–11 clones per animal yielded 69 unique VLRs and only one case of two identical clones from one individual. VLR sequences from two animals are shown to illustrate the protein diversity (Fig. 2; entire set included in Fig. 3 and Supplementary Fig. 3). The size variation of 129–202 residues is primarily due to differences in the number of LRR modules. Each sequence contains an LRRNT of 30–38 residues, an 18-residue LRR1, 1–9 LRRs almost invariably consisting of 24 residues, a 13-residue connecting peptide and a C-terminal LRRCT of 48–58 residues. Regions of pronounced sequence diversity are evident for each LRR motif, almost always consisting of non-similar residues, yet the first seven residues in the LRRNT and the last 20 residues in the LRRCT are nearly invariant.

To assess VLR diversity at the level of individual lymphocytes we used RT–PCR with two sets of nested primers flanking the entire ORF. Analysis of the amplicons obtained from six single lymphocytes sorted from the blood of one unstimulated animal and seven lymphocytes from an immunostimulated larva showed that 12 of the 13 lymphocytes expressed a single VLR (Fig. 3; animals 8, 9). Only one cell isolate yielded two VLR sequences (9.16S, 9.16L), and this isolate might have contained two lymphocytes. Control reactions consisted of four cDNA mixture pairs from the first round of eight single-lymphocyte RT–PCR reactions; subsequent amplification with the nested primers yielded two amplicons for each reaction. In addition, five out of six VLR clones from a pool of
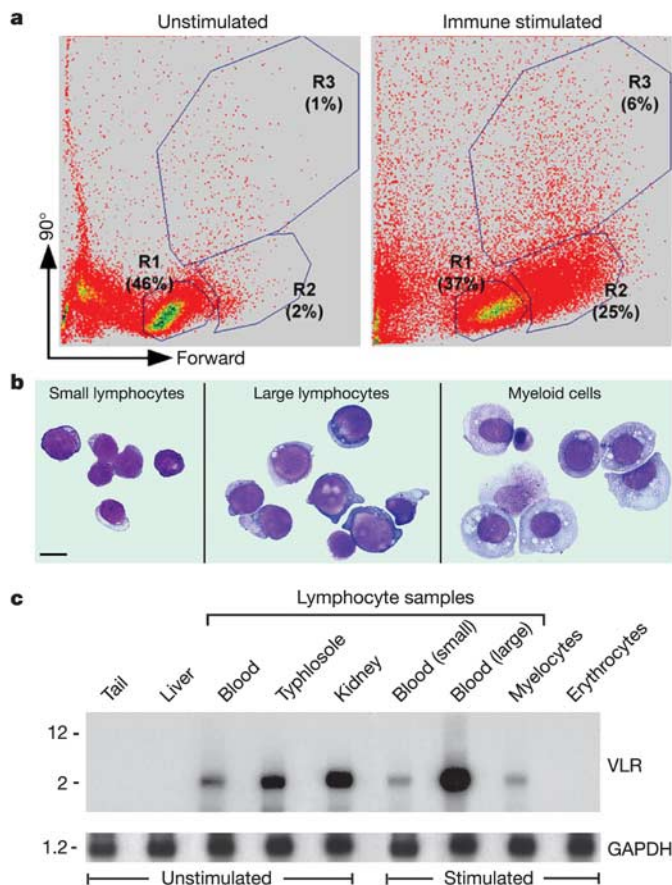
Figure 1 Lamprey leukocytes and VLRs. **a**, FACS analysis of blood leukocytes before and after antigen/mitogen cocktail immunostimulation. **b**, Immunostimulated leukocytes. Small lymphocytes, large lymphocytes and myeloid cells correspond with R1, R2 and R3 respectively, in **a**. Wright–Giemsa stain. Scale bar, 10 μm. **c**, VLR Virtual northern blot. Amplified cDNA from larval tissues, or sorted cells from unstimulated and immunostimulated blood and haematopoietic organs. Molecular size is indicated in
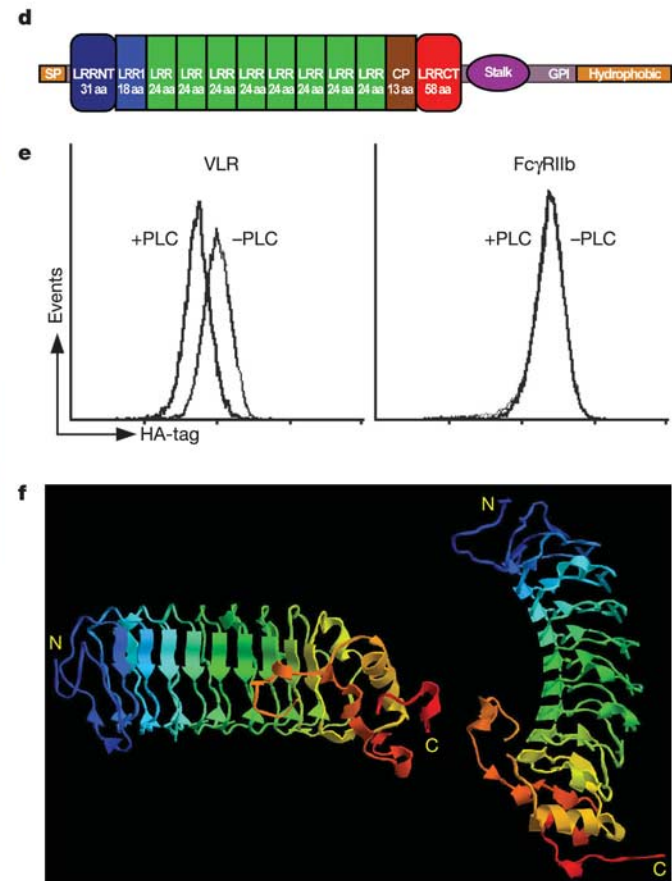
kilobases. **d**, VLR stick model. Sections from left to right are: signal peptide, N-terminal LRR, nine LRRs, connecting peptide, C-terminal LRR, threonine/proline-rich stalk, GPI anchor and hydrophobic tail (clone 12.26, AY577974). **e**, Epitope-tagged VLR and FcγRIIb (control) expressed in mouse thymoma cells, treated with (+PLC) or without (−PLC) GPI-phospholipase C. **f**, Three-dimensional model of VLR diversity region (two rotations).

ten unstimulated cells were unique. These controls indicate that if more than one VLR template were present in a single lymphocyte sample, these would be amplified simultaneously. Notably, combinations of identical VLRs were identified among five lymphocytes from the immunostimulated larva (VLR 9.1 = 9.16S; 9.2 = 9.16L; 9.7 = 9.9; See Fig. 3). These findings are indicative of monoallelic expression of the VLRs, and provide preliminary evidence for clonal expansion of VLR-bearing lymphocytes.

Complexity of the *VLR* locus

Genome blot hybridization with a C-terminal VLR probe revealed a single band for all three digests used (Fig. 4a). The N-terminal probe, consisting of the 5′ untranslated region (UTR) and signal peptide (Fig. 5a), hybridized with 2–3 bands depending on the restriction enzyme used, except for one individual whose blot showed two additional *Bam*HI bands. For high-resolution genome spread we probed a pulse-field blot of erythrocyte DNA (Fig. 4b). In all six digests a single hybridizing band was detected with the C-terminal probe, and the N-terminal probe produced a matching pattern with one additional 350-kilobase *Not*I band. These findings indicate a single *VLR* locus, with the N terminus and C terminus of the VLR gene (germline VLR; *gVLR*) being contained within 100–150 kb of the genome (Fig. 4b; *Pac*I digest). To characterize further 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 adult. In an analysis of the five PACs that hybridized with both probes, a single 14-kb *gVLR* amplicon was identified by long-range PCR (LR–PCR). Restriction-enzyme analysis of these amplicons revealed identical *Eco*RI bands and two allelic *Bam*HI patterns. PAC clones representing the two *gVLR* alleles were sequenced (PAC3 and PAC16) with inserts of 33 and 44 kb, respectively (Fig. 5a). Their sequences overlapped a 20-kb region containing *gVLR*, and furthermore, PAC16 extended 25 kb upstream from *gVLR* whereas PAC3 extended 18 kb downstream. The overlap region between PAC3 and PAC16 was identical, except for short deletions in *gVLR* from PAC16 (24, 43 and 78 bp). These

sequences were therefore melded into a *gVLR* contig preserving the slightly longer sequence of PAC3.

gVLR in the PAC3/16 contig consists of four 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, whereas other exon–intron boundaries in *gVLR* were determined by alignment to cDNA sequences. Notably, the *gVLR* sequence did not contain a 3′ LRRNT, LRR1 or any of the 24-residue LRRs. Upstream from *gVLR*, six cassettes of singlet or doublet variable LRR modules were identified, including LRRNTs, LRR1 and 24-residue LRRs positioned either in forward or reverse orientation. These LRR cassettes spanned the first 6 kb of the contig, whereas two diverse 5′ LRRCT cassettes were located 7 kb downstream from *gVLR*.

A survey of PAC clones that hybridized only with the N-terminal probe—by PCR with LRRNT and LRR1 consensus primers—indicated that PAC4 encoded multiple LRRs. The entire insert was 58 kb long (Fig. 5b) and its sequence aligned to 11.7 kb of the *gVLR* contig with minor gaps (four gaps of 210–738 bp in PAC4; 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. Thirty diverse LRR modules, arranged in 17 cassettes of 1–3 modules, were identified in a 31-kb region located 15-kb upstream from the partial 5′ *gVLR*. Sequence comparison between PAC4 and the PAC3/16 contig revealed additional 1–5-kb regions with >90% identity, but these were disrupted by unrelated sequences. Hence, the 3′ terminal ~12 kb of PAC4 might represent either a duplication of about half of the *gVLR* and the 5′ flank, or a highly divergent VLR allele. To estimate the level of polymorphism in the *VLR* locus we compared the pattern of hybridization in the N-terminal Southern blot to the map of restriction sites in the *gVLR* genes from all three PAC inserts. The pattern of bands and restriction maps was compatible except for a 5.7-kb PAC4 *Hin*dIII fragment instead of the 2-kb band in the
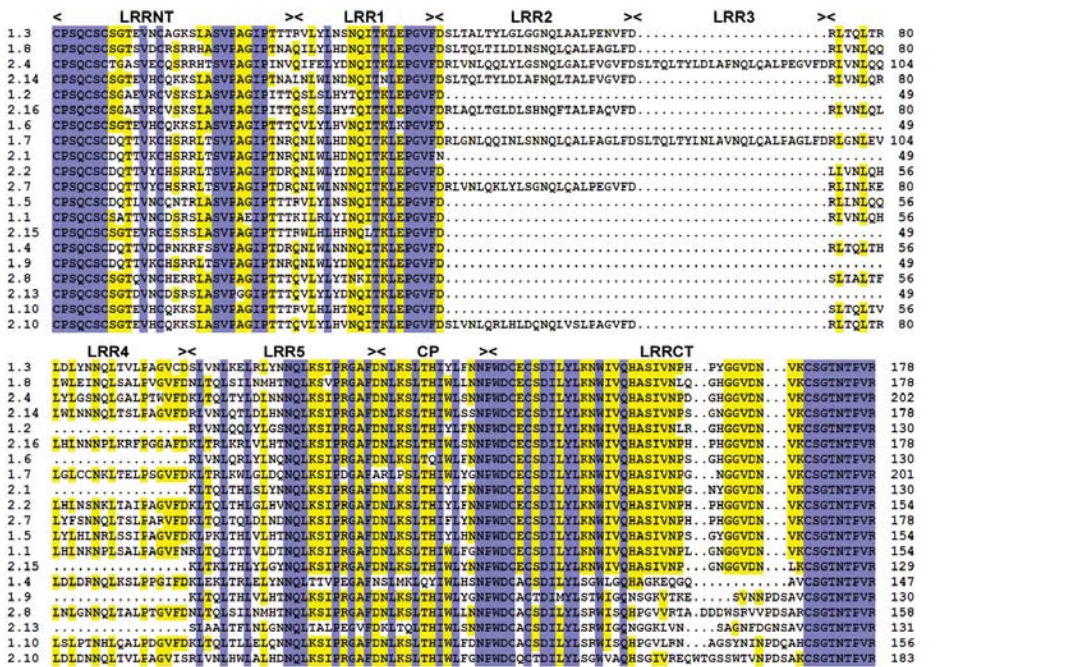

Figure 2 Diversity of VLRs from two lampreys. Alignment of 20 diversity regions, from LRRNT to near C terminus of LRRCT, PCR amplified from blood leukocyte mRNA. Locations of LRR motifs, clone number (for example, 1.3 indicates animal 1, clone 3) and length are indicated. Blue, 100% identity; yellow, 60–99%; white, <60%. All substitutions are non-similar residues.
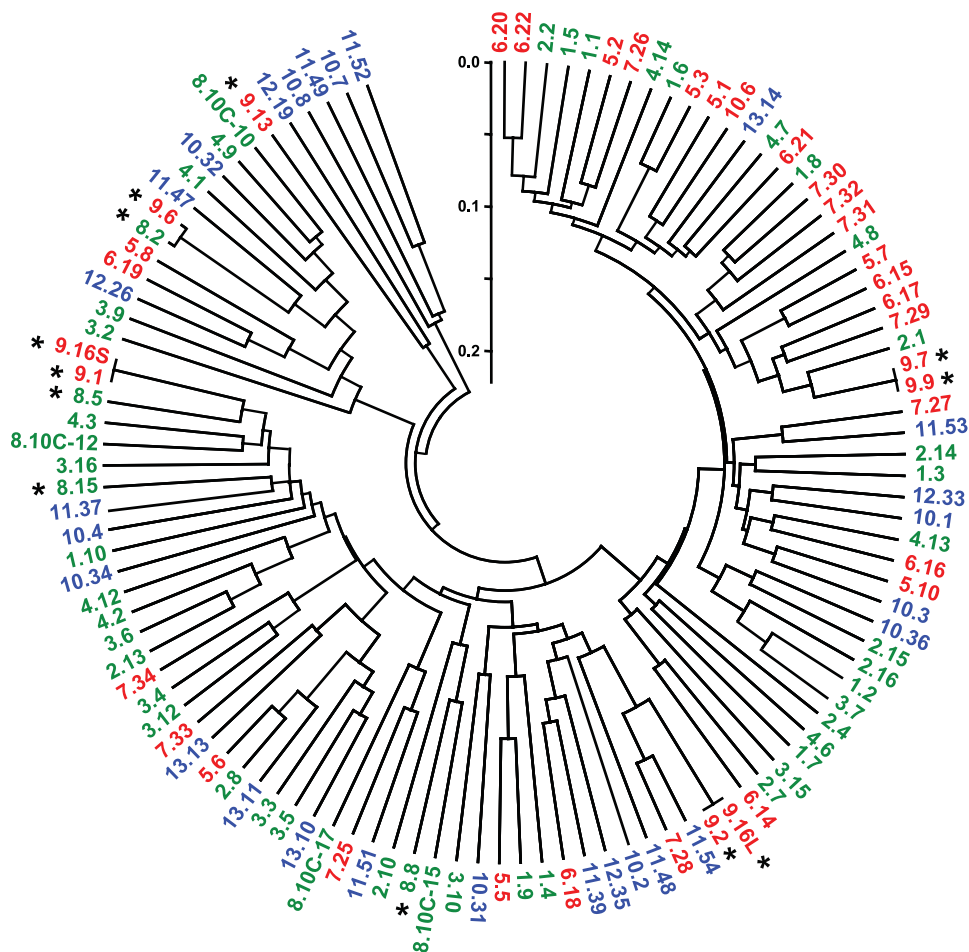
Figure 3 Graphic representation of protein divergence among 112 VLR diversity regions PCR amplified from 13 individual lampreys. Animal and clone numbers are indicated in red for immunostimulated ($N = 27$) and green for unstimulated ($N = 41$) larvae. Asterisks indicate VLRs derived from single lymphocytes ($N = 12$), including two VLRs from one isolate (9.16S, 9.16L); clones derived from a control ten-cell pool are denoted 10C ($N = 4$). Mature VLR sequences derived from genomic DNA are in blue ($N = 28$; blood, animal 10, 12; carcass, animal 11, 13). The mean diversity is 0.36 ± 0.03 ($\pm$s.d.), ranging from 0.28 to 0.54 within the groups of sequences from 13 individuals. Scale bar indicates per cent amino acid divergence.
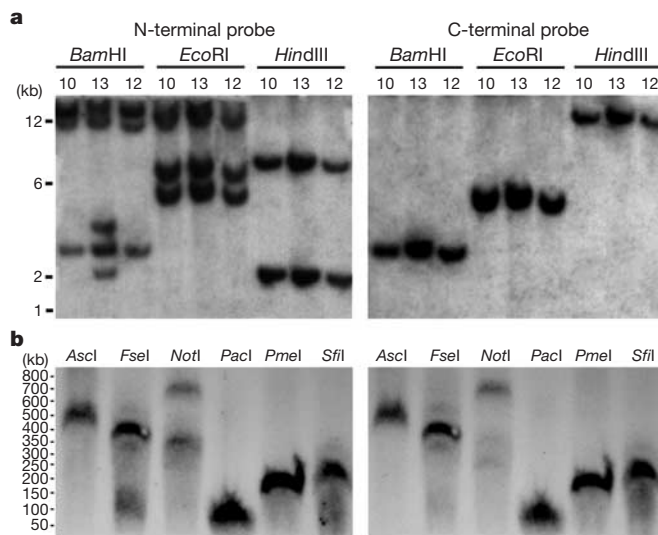
Figure 4 Sea lamprey genome harbours a single *VLR* locus. Genome blots of restriction-enzyme-digested DNA were hybridized with VLR N-terminal or C-terminal probes. **a**, Blots of three lampreys (blood DNA, animal 10, 12; carcass, animal 13). Only animal 13 shows a polymorphic *Bam*HI pattern. **b**, Genome spread of erythrocytes pooled from ten 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.

Southern blot (Fig. 4a). As only limited polymorphism was observed in an analysis of *gVLR* from three larvae and the PAC library donor, it is unlikely that the 5′ half of *gVLR* from PAC4 represents a divergent allele. The limited genetic diversity indicated here for the *VLR* locus is consistent with that reported for microsatellite loci[26], thus indicating a high level of inbreeding among the recently landlocked population of sea lamprey in the North American Great Lakes. We conclude from this analysis that a 5′ half *gVLR* duplication has occurred within ~100 kb of *gVLR*.

Somatic rearrangement generates mature VLRs

PCR amplification from blood or larval carcass DNA with primers flanking the VLR diversity region produced intronless VLR ORFs of 0.5–0.8 kb. PCR with two other sets of ORF-flanking primers produced similar amplicons of ~1 kb, or ~1.7 kb including the 5′ UTR intron. Analysis of these PCR products revealed unique VLR sequences in all 19 clones (Fig. 3; numbers 10–13). Because these genomic amplicons encoded uninterrupted VLR ORFs, which contained all the diversity LRR modules identified in VLR transcripts, they were provisionally named mature VLRs to distinguish them from the 'incomplete' *gVLR*. Sequence analysis indicated that the ~1.7-kb mature VLRs should be represented by polymorphic *Eco*RI bands of 1–1.5 kb in the N-terminal Southern blot, but none of these genomic DNA samples yielded visible *Eco*RI bands of <5 kb (Fig. 4a). These observations indicate that mature VLR templates are included in DNA samples extracted from larval carcasses or blood, yet paradoxically only *gVLR* was detected in DNA blots from these samples.

To address this issue we proposed that somatic gene rearrangement[1] in lamprey lymphocytes might generate the mature VLRs by insertion of diverse LRR modules from upstream and downstream cassettes into *gVLR*. Primers were therefore designed for PCR amplification across *gVLR*, including flanking regions ~3 kb upstream and ~3 kb of *gVLR* (Fig. 5a). LR–PCR amplification from genomic DNA samples yielded a minor band of ~20 kb, similar to the *gVLR* amplicon from PAC16, plus a prominent band of ~8 kb (Fig. 5c). Sequence analysis of the 8-kb amplicons from two larval DNA samples revealed unique mature VLRs in nine out of

ten clones (Fig. 3; animal 10, blood; animal 13, carcass—two identical clones), the flanks of which were identical to those of *gVLR* (Fig. 5e). In contrast, LR–PCR amplification with primers flanking the *gVLR* ORF from 100-cell pools of erythrocytes or lymphocytes revealed the 14-kb *gVLR* amplicon in both, whereas the ~1-kb mature VLRs were observed only in lymphocytes (Fig. 5d). We conclude that mature VLR amplicons from blood or carcass DNA are derived from the lymphocytes in these samples, with preferential amplification favouring the shorter mature VLR templates over the more abundant but much longer *gVLR* templates. Altogether 28 unique mature VLRs were identified among random samples from four lamprey genomic amplicons, thus indicating a level of diversity equivalent to that found in VLR transcripts.

Discussion

Our search for receptors that could mediate adaptive immune responses in the lamprey has identified a novel system of variable lymphocyte receptors. The VLRs consist of multiple LRR modules and an invariant stalk region that is attached to the lymphocyte plasma membrane by a GPI anchor. The flanking ends of the N-terminal and C-terminal LRRs are invariant and the remarkable VLR diversity derives from the variation in sequence and number of the LRR modules. Our initial assessment identified unique transcripts in 345 out of 354 sequences, three pairs of identical VLRs from immunostimulated lymphocytes and only three nearly identical sequences. Random samples of VLR clones from individual lampreys also revealed a prevalence of unique sequences; for example, all 12 VLR clones from animal 10 were unique. This agnathan representative is thus endowed with a vast repertoire of lymphocyte receptors that is expressed predominantly in a monoallelic fashion.

A single germline VLR gene (*gVLR*) was identified in the lamprey genome. It comprises four exons that encode only the signal peptide, 5′ LRRNT, 5′ LRRCT, 3′ LRRCT and the C terminus. Lacking in diversity LRR modules, this incomplete *gVLR* could not encode the highly diverse VLR transcripts. However, clusters of diverse LRR cassettes were found upstream and downstream from
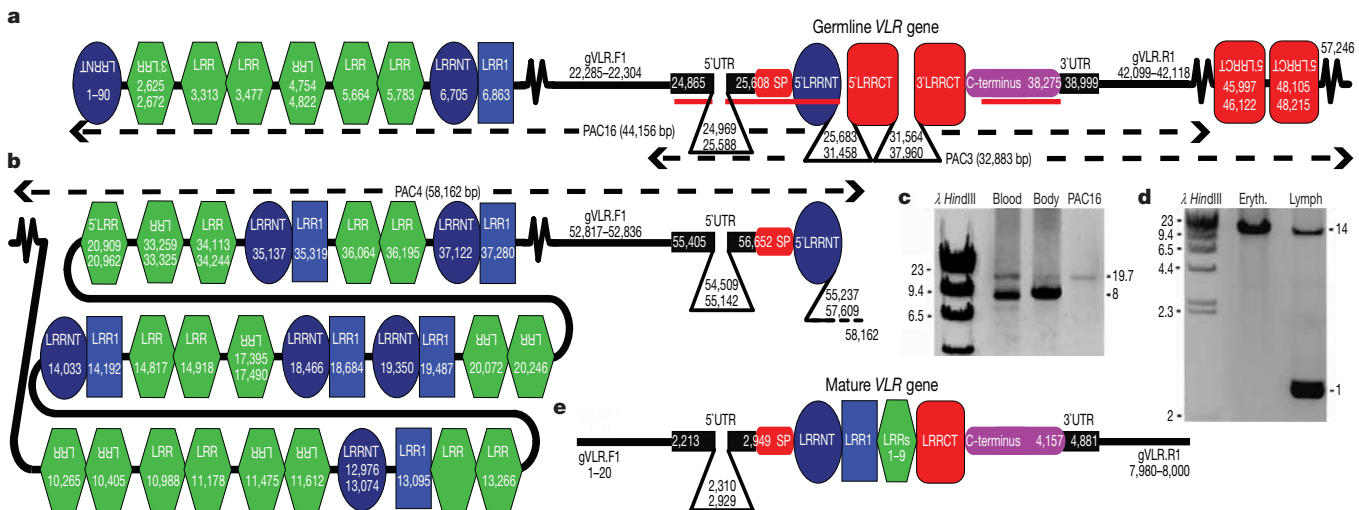
Figure 5 The *VLR* locus. **a**, Motifs in the 57-kb *gVLR* contig (PAC16, 44 kb; PAC3, 33 kb; AY577941). Cassettes of 1–2 LRRs are positioned in forward or reverse orientations. Dashed lines represent PAC inserts; red bars indicate N-terminal and C-terminal probes; gVLR.F1 and gVLR.R1 indicate LR–PCR primers. **b**, Motifs in PAC4 (58 kb, nucleotides 45,882–57,609, align with the *gVLR* contig; AY577942). **c**, LR–PCR analysis of *gVLR*, amplified from blood DNA (animal 10), carcass (animal 13) or PAC16 (control). The

germline VLR is ~20 kb whereas the mature VLRs are ~8 kb. **d**, Lymphocyte-specific rearrangement of mature VLRs. LR–PCR with ORF-flanking primers from pools of 100 lymphocytes or erythrocytes: germline VLR ~14 kb; lymphocyte mature VLRs ~1 kb. Numbers along left and right sides of panels in **c** and **d** indicate molecular size in kilobases. **e**, An 8-kb mature VLR amplicon.

gVLR, and these would be available for insertion into *gVLR* to assemble mature VLR genes. To test the hypothesis that mature VLRs are generated through insertion of upstream and downstream LRR cassettes, which replace non-coding DNA in the lymphocyte *gVLR*, we used LR–PCR to detect both germline and mature VLR genes in lamprey genomic DNA from blood and carcasses. Two amplicons resulted: the expected *gVLR* product of ~20 kb and the predicated ~8-kb amplicon that consisted of a diverse set of mature VLRs. In contrast, when purified populations of erythrocytes and lymphocytes were examined, mature VLR amplicons were found only in the lymphocytes. The mature VLRs are thus generated through a process of somatic DNA rearrangement in lymphocytes.

The highly diverse VLRs may serve a role in recognition of pathogens. Proteins featuring heterogeneous LRR modules are principal innate immune receptors of animals and plants owing to their ability to interact with a wide range of pathogen-associated ligands. For instance, the LRR ectodomains of vertebrate TLRs are implicated in recognition of conserved epitopes of viruses, bacteria, fungi and protozoa, thereby activating signal transduction cascades that culminate in inflammatory responses[27]. CD14, a GPI-anchored LRR protein that is also found in a soluble form, binds bacterial lipopolysaccharide and phospholipids to form a signalling complex with the TLR4 receptor[28]. Yet another mammalian family of cytosolic LRR proteins, the NBS-LRRs, recognize intracellular pathogens[29]. Plant disease resistance genes are members of large multigene families including hundreds of NBS-LRR proteins, LRR-receptor-like kinases and LRR-receptor-like proteins, many of which are involved in specific activation of anti-pathogen responses[30]. In the likely event that VLRs prove to be capable of binding antigens, their remarkable diversity makes them excellent candidates to mediate the humoral immune responses observed in lampreys. The GPI anchorage of VLRs to the surface of lymphocytes might allow GPI-specific phospholipase release of these receptors[31], thereby endowing VLRs with dual functionality, both as cell-surface receptors and humoral agglutinins in an anticipatory immune system.

Many features of the lamprey VLR system bear analogy to the variable B and T lymphocyte receptors of jawed vertebrates, with the notable difference that lamprey VLRs consist of LRR modules whereas gnathostome antigen receptors consist of immunoglobulin domains. Lamprey immunity may therefore represent a traceable evolutionary process in which the ancestral strategy of multigene families of heterogeneous LRR receptors gave rise to a novel system of variable lymphocyte LRR receptors that are somatically diversified versions of the single germline VLR gene. In contrast, the deployment of immunoglobulin domains as core components of jawed vertebrate recombinatorial lymphocyte receptors represents an intriguing although as yet untraceable evolutionary innovation, as immune recognition of pathogens and allografts by means of immunoglobulin superfamily members have been shown only in the jawed vertebrates[32].

Elucidation of the precise mechanism underlying the somatic generation of mature VLRs will require further study, but we speculate that the conserved flanking ends in the *gVLR* 5′ LRRNT and 3′ LRRCT might serve as anchoring regions for a gene conversion process similar to that elucidated for avian antibody diversification[33,34]. Indeed, candidate LRR-module donors could be identified among the *gVLR* neighbouring cassettes on the basis of partial identity to VLR sequences. Furthermore, rearrangement of genes encoding surface components is a strategy often used by pathogens to evade immune recognition during chronic infection. Antigenic variation in the pilin of *Neisseria gonorrhoeae* involves non-reciprocal recombination between multiple genes[35], and surface antigen variation in *Borrelia* spirochaetes, the causative agent of Lyme disease, is generated by gene conversion between an array of 15 silent cassettes and one expression site[36]. Recurring DNA rearrangements are also used by the protozoan *Trypanosoma brucei*

for alternative expression of its surface coat glycoprotein genes[37], and a similar mechanism may drive the frequent switching among multiple surface antigens in the malaria parasite *Plasmodium falciparum*[38] and the intestinal dweller *Giardia lamblia*[39]. In the evolutionary arms race between hosts and parasites, vertebrates have adopted a similar strategy to combat infectious disease by somatic rearrangement of germline receptor units. The diverse lymphocyte antigen receptors are assembled by means of the cut-and-paste activity of the paired transposase-like RAG1/RAG2 in gnathostomes[40], and via a mechanism that we are just beginning to unravel in Agnatha. □

Methods

Animals, immune stimulation and blood leukocytes

Sea lamprey larvae, 8–13 cm long, from tributaries to Lake Michigan (Lamprey Services) or Lake Huron (Hammond Bay Biological Station), were sedated (100 mg l^{-1} MS222; Sigma) for intraperitoneal injection with 75 μl 0.67 × PBS containing: 10^7 live *Escherichia coli* bacteria, 10^7 sheep erythrocytes, 50 μg phytohaemagglutinin and 25 μg pokeweed mitogen (Sigma). Larvae were injected two or four times at weekly intervals and cells were collected 3–4 days later. Blood was drained from tail-severed larvae and diluted 1:1 with 0.57 × PBS, 30 mM EDTA. Buffy coat leukocytes were collected after 5 min centrifugation at 50 g and sorted using a MoFlo cytometer (Cytomation).

Subtracted lymphocyte cDNA libraries

The Super SMART PCR cDNA Synthesis kit (BD Biosciences) was used with mRNA from activated blood lymphocytes, myeloid cells and erythrocytes sorted from 21 larvae immunostimulated four times. Activated lymphocyte cDNA was subtracted in two reactions, against myeloid or erythrocyte cDNA (PCR-Select, BD Biosciences). Subtracted products were cloned in pGEM-T (Promega) and 1,507 sequences were analysed.

VLR RT–PCR

All PCR primers are listed in Supplementary Table 1. Buffy coat leukocytes from unstimulated larvae (numbers 1–4), or larvae that had been immunostimulated twice (numbers 5–7), were pelleted for 5 min at 300 g. VLR diversity regions were amplified from random primed cDNA (SuperScript III; Invitrogen) with Expand High Fidelity (Roche) using LRR.F1 + LRR.R1. Cycling parameters were: 94 °C for 1 min, then 35 cycles of 94 °C for 30 s, 59 °C for 30 s, 72 °C for 1 min.

Single lymphocytes ($N = 32$), or ten-cell pools of buffy coat leukocytes from an unstimulated larva (animal 8) and one twice-immunostimulated larva (animal 9), were sorted into 0.2 ml TRIzol (Invitrogen) and cDNA was primed with LRR_C.R2. VLRs were amplified by two rounds of nested PCR—first Slit.F + LRR_C.R2 (Advantage II; BD Biosciences) then LRR_N.F1 + LRR_C.R1—using Expand High Fidelity. Colony PCR with vector primers revealed single-sized inserts in six colonies from each of 12 lymphocytes, three of which were sequenced. Colonies from isolate 9.16 revealed two insert sizes, and six short and long inserts were sequenced. Six clones were sequenced from a control pool of ten unstimulated cells.

Genomic DNA and PCR

Genomic LR–PCR (Expand Long Template, Roche) was carried out from 400 ng DNA, isolated from larval carcass or 0.25 ml blood pelleted for 5 min at 50 g. VLR diversity regions were amplified using LRR.F1 + LRR.R1 (0.5–0.8 kb; animals 10, 12). Mature VLRs were amplified using Slit.F + LRR_C.R2 or LRR_N.F2 + LRR_C.R1 (~1 or ~1.7 kb respectively; animals 10, 11). Amplification across the *gVLR* was with gVLR.F1 + gVLR.R1 (animals 10, 13), and the 8-kb amplicon was cloned in pCR-XL (Invitrogen). Pools of 100 erythrocytes or blood lymphocytes were sorted into 0.2 ml TE, 0.1% SDS, 5 μl β-mercaptoethanol and 5 μg proteinase K (Invitrogen). After 1 h at 37 °C and precipitation *gVLR* was amplified by two rounds of nested PCR, first Slit.F + LRR_C.R2 then LRR_N.F2 + LRR_C.R1.

Virtual northern and genome blots

Virtual northern blot was prepared according to the SMART manual. Twenty-cycle amplified cDNA was from unstimulated larvae (tail, liver and sorted lymphocytes from blood, typhlosole and kidneys) and blood cells sorted from four times immunostimulated larvae (small and large lymphocytes, myeloid cells and erythrocytes).

For genome blots, 10 μg larval DNA per lane (animal 10, 12, 13) was digested with *Bam*HI, *Eco*RI and *Hind*III (Roche). For the pulse-field blot (CHEF DRIII, Bio-Rad), erythrocytes from ten larvae were embedded in agarose[41] and 20 μg DNA per lane was digested with *Asc*I, *Fse*I, *Not*I, *Pac*I, *Pme*I and *Sfi*I (New England Biolabs). ^{32}P-labelled probes were: VLR N terminus (196 bp) amplified with Slit.F + Slit.R, and C terminus (208 bp) amplified with LRR_C.F1 + LRR_C.R1; GAPDH (314 bp) was amplified with GAPDH.F + GAPDH.R.

PAC clones

The arrayed sea lamprey PAC library in pCYPAC6 (AF133437) was constructed from erythrocytes of one Lake Michigan adult using partial *Mbo*I digests[41]. The $6 × 10^4$ clones had ~65-kb inserts with 1–2-fold genome coverage. The library was screened with N-terminal and C-terminal probes; five PACs hybridized with both probes (2, 3, 15–17) and six others hybridized only with the N terminus (4, 9, 14, 35, 42, 43). *gVLR* was amplified from PACs 2, 3, 15–17 using Slit.F + LRR_C.R2. The 14-kb amplicons revealed

two sets of *Bam*HI patterns: PACs 2, 3 and 15–17. PACs 3, 4, 16 were sequenced at McGill University.

VLR GPI anchor

A VLR insert, LRRNT to stop codon, was amplified using Dis_LRR.F + Dis_LRR.R1 (Expand High Fidelity) and fused to Igκ signal peptide and haemagglutinin (HA) epitope in pDisplay (Invitrogen). Transfectants in BW1547 cells, the GPI-anchored VLR, or mFcγRIIb as control, were treated with 1 unit ml^{-1} bacterial GPI-specific phospholipase C (Sigma) for 45 min at 30 °C. Epitope-tagged proteins were stained with anti-HA-tag monoclonal antibody 12CA5 (Roche) for FACS analysis.

Sequence analysis

A genetic distance dendrogram was generated with MEGA 2.1 UPGMA[42]. GPI-anchor site was predicted by means of http://129.194.185.165/dgpi/. SWISS-MODEL[24] VLR three-dimensional prediction was made using http://cubic.bioc.columbia.edu/predictprotein/submit_meta.html. Residues 22–319 from clone 12.26 were threaded on crystal coordinates of CD42α (1m10.pdb) and NOGO-66 receptor (1p8t.pdb).

1. Litman, G. W., Anderson, M. K. & Rast, J. P. Evolution of antigen binding receptors. *Annu. Rev. Immunol.* **17,** 109–147 (1999).
2. Flajnik, M. F. & Kasahara, M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* **15,** 351–362 (2001).
3. Flajnik, M. F. Comparative analyses of immunoglobulin genes: surprises and portents. *Nature Rev. Immunol.* **2,** 688–698 (2002).
4. Forey, P. L. & Janvier, P. Agnathans and the origin of jawed vertebrates. *Nature* **361,** 129–134 (1993).
5. Finstad, J. & Good, R. A. The evolution of the immune response. III. Immunologic responses in the lamprey. *J. Exp. Med.* **120,** 1151–1167 (1964).
6. Marchalonis, J. J. & Edelman, G. M. Phylogenetic origins of antibody structure. 3. Antibodies in the primary immune response of the sea lamprey, *Petromyzon marinus. J. Exp. Med.* **127,** 891–914 (1968).
7. Litman, G. W. *et al.* The evolution of the immune response. VIII. Structural studies of the lamprey immunoglobulin. *J. Immunol.* **105,** 1278–1285 (1970).
8. Pollara, B., Litman, G. W., Finstad, J., Howell, J. & Good, R. A. The evolution of the immune response. VII. Antibody to human "O" cells and properties of the immunoglobulin in lamprey. *J. Immunol.* **105,** 738–745 (1970).
9. 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, London, 1972).
10. Hagen, M., Filosa, M. F. & Youson, J. H. The immune response in adult sea lamprey (*Petromyzon marinus* L.): the effect of temperature. *Comp. Biochem. Physiol.* **82,** 207–210 (1985).
11. Perey, D. Y., Finstad, J., Pollara, B. & Good, R. A. Evolution of the immune response. VI. First and second set skin homograft rejections in primitive fishes. *Lab. Invest.* **19,** 591–597 (1968).
12. Cooper, A. J. in *Ammocoete Lymphoid Cell Populations In vitro. 4th Leukocyte Culture Conference* (ed. McIntyre, O. R.) 137–147 (Appleton Century-Crofts, New York, 1971).
13. Piavis, G. W. & Hiatt, J. L. Blood cell lineage in the sea lamprey *Petromyzon marinus* (Pisces: Petromyzontidae). *Copeia* **4,** 722–728 (1971).
14. Fujii, T. Electron microscopy of the leukocytes of the typhlosole in ammocoetes, with special attention to the antibody-producing cells. *J. Morphol.* **173,** 87–100 (1982).
15. Ardavin, C. F. & Zapata, A. Ultrastructure and changes during metamorphosis of the lympho-hemopoietic tissue of the larval anadromous sea lamprey *Petromyzon marinus. Dev. Comp. Immunol.* **11,** 79–93 (1987).
16. Mayer, W. E. *et al.* Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl Acad. Sci. USA* **99,** 14350–14355 (2002).
17. Mayer, W. E., O'Huigin, C., Tichy, H., Terzic, J. & Saraga-Babic, M. Identification of two Ikaros-like transcription factors in lamprey. *Scand. J. Immunol.* **55,** 162–170 (2002).
18. Haire, R. N., Miracle, A. L., Rast, J. P. & Litman, G. W. Members of the Ikaros gene family are present in early representative vertebrates. *J. Immunol.* **165,** 306–312 (2000).
19. Shintani, S. *et al.* Do lampreys have lymphocytes? The Spi evidence. *Proc. Natl Acad. Sci. USA* **97,** 7417–7422 (2000).
20. Anderson, M. K., Sun, X., Miracle, A. L., Litman, G. W. & Rothenberg, E. V. Evolution of hematopoiesis: three members of the PU.1 transcription factor family in a cartilaginous fish, *Raja eglanteria. Proc. Natl Acad. Sci. USA* **98,** 553–558 (2001).
21. Azumi, K. *et al.* Genomic analysis of immunity in a urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics* **55,** 570–581 (2003).
22. Uinuk-Ool, T. S. *et al.* Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc. Natl Acad. Sci. USA* **99,** 14356–14361 (2002).
23. Uinuk-Ool, T. S. *et al.* Identification and characterization of a TAP-family gene in the lamprey. *Immunogenetics* **55,** 38–48 (2003).
24. Schwede, T., Diemand, A., Guex, N. & Peitsch, M. V. Protein structure computing in the genomic era. *Res. Microbiol.* **151,** 107–112 (2000).
25. Bell, J. K. *et al.* Leucine-rich repeats and pathogen recognition in Toll-like receptors. *Trends Immunol.* **24,** 528–533 (2003).
26. Bryan, M. B., Libants, S. V., Warrillow, J. A., Li, W. & Scribner, K. T. Polymorphic microsatellite markers for the landlocked sea lamprey, *Petromyzon marinus. Conserv. Genet.* **4,** 113–116 (2003).
27. Beutler, B. Innate immunity: an overview. *Mol. Immunol.* **40,** 845–859 (2004).
28. Landmann, R., Müller, B. & Zimmerli, W. CD14, new aspects of ligand and signal diversity. *Microbes Infect.* **2,** 295–304 (2000).
29. Chamaillard, M., Girardin, S. E., Viala, J. & Philpott, D. J. Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation. *Cell. Microbiol.* **5,** 581–592 (2003).
30. Jones, D. A. & Takemoto, D. Plant innate immunity—direct and indirect recognition of general and specific pathogen-associated molecules. *Curr. Opin. Immunol.* **16,** 48–62 (2004).
31. Ikezawa, H. Glycosylphosphatidylinositol (GPI)-anchored proteins. *Biol. Pharm. Bull.* **25,** 409–417 (2002).
32. Kaufman, J. The origins of the adaptive immune system: whatever next? *Nature Immunol.* **3,** 1124–1125 (2002).
33. Thompson, C. B. & Nieman, P. E. Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. *Cell* **48,** 369–378 (1987).
34. Reynaud, C. A., Anquez, V., Grimal, H. & Weill, J. C. A hyper conversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48,** 379–388 (1987).
35. Hamrick, T. S., Dempsey, J. A., Cohen, M. S. & Cannon, J. G. 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. *Microbiol.* **147,** 839–849 (2001).
36. Wang, D., Botkin, D. J. & Norris, S. J. Characterization of the *vls* antigenic variation loci of the Lyme disease spirochaetes *Borrelia garinii* Ip90 and *Borrelia afzelii* ACAI. *Mol. Microbiol.* **47,** 1407–1417 (2003).
37. Berriman, M. *et al.* The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei. Mol. Biochem. Parasitol.* **122,** 131–140 (2002).
38. Duffy, M. F., Reeder, J. C. & Brown, G. V. Regulation of antigenic variation in *Plasmodium falciparum*: censoring freedom of expression? *Trends Parasitol.* **19,** 121–124 (2003).
39. Nash, T. E. Surface antigenic variation in *Giardia lamblia. Mol. Microbiol.* **45,** 585–590 (2002).
40. Schluter, S. F., Bernstein, R. M., Bernstein, H. & Marchalonis, J. J. 'Big Bang' emergence of the combinatorial immune system. *Dev. Comp. Immunol.* **23,** 107–111 (1999).
41. Amemiya, C. T., Ota, T. & Litman, G. W. in *Nonmammalian Genomic Analysis: A Practical Guide. Construction of P1 Artificial Chromosome (PAC) Libraries from Lower Vertebrates* (eds Birren, B. & Lai, E.) 223–256 (Academic, New York, 1996).
42. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17,** 1244–1245 (2001).