Genes "Waiting" for Recruitment by the Adaptive Immune System: The Insights from Amphioxus[1]

Cuiling Yu,[2]* Meiling Dong,[2]* Xiaokun Wu,[2]* Shengguo Li,[§] Shengfeng Huang,* Jing Su,* Jianwen Wei,* Yang Shen,* Chunyan Mou,* Xiaojin Xie,* Jianghai Lin,* Shaochun Yuan,* Xuesong Yu,* Yanhong Yu,* Jingchun Du,* Shicui Zhang,[†] Xuanxian Peng,[‡] Mengqing Xiang,[§] and Anlong Xu[3]*

In seeking evidence of the existence of adaptive immune system (AIS) in ancient chordate, cDNA clones of six libraries from a protochordate, the Chinese amphioxus, were sequenced. Although the key molecules such as TCR, MHC, Ig, and RAG in AIS have not been identified from our database, we demonstrated in this study the extensive molecular evidence for the presence of genes homologous to many genes that are involved in AIS directly or indirectly, including some of which may represent the putative precursors of vertebrate AIS-related genes. The comparative analyses of these genes in different model organisms revealed the different fates of these genes during evolution. Their gene expression pattern suggested that the primitive digestive system is the pivotal place of the origin and evolution of the AIS. Our studies support the general statement that AIS appears after the jawless/jawed vertebrate split. However our study further reveals the fact that AIS is in its twilight in amphioxus and the evolution of the molecules in amphioxus are waiting for recruitment by the emergence of AIS. *The Journal of Immunology,* **2005, 174: 3493–3500.**

The hallmark of the adaptive immune system (AIS)[4] is the presence of cells and molecules participating in the immune recognition of foreign pathogens and the memory of this recognition (1). The cells for AIS primarily are B lymphocytes, T lymphocytes, and APCs. B lymphocytes participate in the humoral immune response by secreting Ig recognizing foreign pathogens. APCs interact with T lymphocytes to form cellular immune response by presenting foreign Ags in the context of MHC to TCR on T lymphocytes. Because no distinct evidence of the existence of either the key molecules such as TCR, Ig, MHC, and RAG or the effector cells was demonstrated in invertebrates, even the jawless vertebrates, it was generally believed that the adaptive immunity emerged suddenly and is only present in jawed vertebrate (2, 3). The studies for the origin of the AIS focus on many aspects: the origin of the Ag receptor, Ag processing and presentation system, and the effector cells (3, 4). Recently, series homologues of vertebrate genes involved in AIS have been isolated and characterized in the lymphocyte-like cells of lampreys, suggesting that the lymphocyte evolution was "waiting" for the emergence of adaptive immunity (5–7). A new type of variable lymphocyte receptors with somatic diversification was just identified in lampreys, suggesting a new strategy that is used to generate the diversified receptor through rearrangement in agnathans (8). Recently, a study of the diversification of Ig superfamily (IgSF) member in snail also indicated that the mechanism providing the diversification to help invertebrate to fight against the varied pathogens has existed in invertebrate independently from the AIS of vertebrate (9). In addition, a multigene family containing Ig-like variable regions, V region-containing chitin-binding protein (VCBP), and an IgSF gene homologous to CD47 have been identified in the intestine of amphioxus (10, 11). The genomic draft of a urochordate, *Ciona intestinalis*, demonstrated the existence of many homologues of the genes involved in immune system especially innate immunity (12). These studies suggested that some possible ancestors of the molecules involved in the adaptive immunity existed in the early chordates. Abi-Rached et al. (13) identified the genes in the proto-MHC regions in amphioxus genome suggesting that genomic duplication occurred after the divergence of cephalochordates and vertebrates. The following studies (14, 15) showed more information about the relationship between this proto-MHC region and its human paralogous regions. The genomic analyses of ascidian also revealed the existence of the preduplicated form of proto-MHC region (12). Flajnik and Kasahara (16) compared the MHC regions of nonmammalian vertebrates and suggested that the ancient genome duplication played an important role in the origin of MHC region.

In this study, we report the identification of genes whose homologues involve the immune system especially the AIS. The large scale expressed sequence tag (EST) analyses in Chinese amphioxus presented in this study provide a link between the genomic

*State Key Laboratory of Biocontrol, The Open Laboratory for Marine Functional Genomics of State High-Tech Development, Department of Biochemistry, Guangzhou Center for Bioinformatics, College of Life Sciences, Sun Yat-Sen (Zhongshan) University, Guangzhou, People's Republic of China; [†]Department of Marine Biotechnology, Ocean University of China, Qingdao, People's Republic of China; [‡]School of Life Sciences, Xiamen University, Xiamen, People's Republic of China; and [§]Center for Advanced Biotechnology and Medicine, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, Piscataway, NJ 08854

Received for publication July 28, 2004. Accepted for publication December 17, 2004.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by the "Outstanding Young Scientist" Award (No. 39725007), Projects 2004AA621030 and 2003AA626010 of State High-Tech Development Project (863) of Ministry of Science and Technology of China, Key Project (0107) of Ministry of Education, Project 30300264 of National Natural Science Foundation, and Key Projects of Commission of Science and Technology of Guangdong Province and Guangzhou City.

[2] C.Y., M.D., and X.W. contributed equally to this work.

[3] Address correspondence and reprint requests to Dr. Anlong Xu, Department of Biochemistry, College of Life Sciences, Sun Yat-Sen University, 135 Xingangxi Road, 510275 Guangzhou, People's Republic of China. E-mail address: ls36@zsu.edu.cn

[4] Abbreviations used in this paper: AIS, adaptive immune system; EST, expressed sequence tag; IgSF, Ig superfamily; VCBP, V region-containing chitin binding protein; VCP, V and C domain-bearing protein; GILT, IFN-γ-induced lysosomal thiol reductase; EBF, early B cell factor.

0022-1767/05/$02.00

study of *Ciona* and the EST analyses in lamprey in further understanding the origin and evolution of AIS. All evidence we demonstrated supports the general statement that the conventional AIS appear after the jawless/jawed vertebrate but not the cephalochordate/vertebrate split. Furthermore, the molecules identified in this study represent the precursors of vertebrate AIS-related genes and the other housekeeping genes that are waiting for recruitment by the emergence of AIS.

Materials and Methods

Animals and tissues

Adults of Chinese amphioxus, *Branchiostoma belcheri tsingtauense* and *B. belcheri* (Gray), were collected from Kioachow Bay near Qingdao and Xiaman, China, respectively, and kept alive with seawater and sea alga. Tissues of intestine, notochord, and ovary were harvested from adult animals. In breeding seasons, gastrula, neurula, and larva were bred up from the collected oosperm in a man-made seawater environment (17, 18).

Library construction and cDNA sequencing

Library construction and cDNA sequencing were performed according to the procedures previously described in the analyses of the neurula cDNA library (19). The adult animal library was constructed from *B. belcheri* (Gray), and the other libraries including ovary, neurula, gastrula, larva, and intestine were all prepared from *B. belcheri tsingtauense*.

EST analysis

All 5′ EST reads were treated with software PHRED (20, 21) to remove vector sequences and low-quality regions, and then assembled into consensus sequences with software STACKPACK (version 2.1 patch 1) (22, 23). The consensus sequences were used as ESTs to search against GenBank with the BLASTX program (24). EST clustering was performed using STACKPACK2.1 with default setting.

Phylogenetics analyses

Consensus cDNA sequences of genes were searched using the BLASTX algorithm against NR Database from the National Center for Biotechnology Information (NCBI). The homologous sequences were selected from the BLASTX search results as the input for phylogenetic analysis. Consensus phylogenetic tree was constructed with the Phylogeny Inference Package (PHYLIP) (25). The input data for phylogenetic methods is the most consistent alignment that was obtained with the CLUSTALW program and reduced by GBLOCKS (26). Two methods were used to construct the phylogenetic tree using the NEIGHBOUR and PROTPARS programs (PHYLIP). The statistical significance of phylogenies was estimated by using the SEQBOOT program (PHYLIP) by bootstrap analysis with 100 pseudoreplicated data sets, grouping the bootstrap values higher than 50 as significant. The consensus trees were generated by CONSENSE program (PHYLIP) and were plotted by TREEVIEW (version 1.3).

Bioinformatics analyses

Some immune-related ESTs were obtained from the BLASTX results of our database (E value $<1e^{-5}$ and the length of amino acid sequence >50). All domains that each molecule contained are identified through RPS-BLAST in NCBI first. Then Pfam (27) and Prosite (28) databases are queried to find whether the model organisms have the molecules that contain those domains. For the IgSF members, the whole sequences from amphioxus are used to search against the proteomes of the model organisms. Considering the proteome of ascidian is unavailable, TBLASTX program was used to search against the ascidian cDNA database, which is from ⟨http://ghost.zool.lyoto-u.ac.jp/indexrl.html⟩, and the ENSEMBL server was used to predict the homologous genes to refine our analyses (29).

In situ hybridization

Adults of the Chinese amphioxus, *B. belcheri tsingtauense*, were collected and kept in filtered seawater for 2 days. Then the whole body was fixed with 4% paraformaldehyde in PBS and later was made into 14-μm transverse sections in slices by frozen sectioning. The digoxigenin-labeled probes were prepared by using the plasmids that contain the sequences of the interesting ESTs with the SP6 promoter sequence at the 3′ end of the sequences as template and got the antisense probes with the SP6 RNA polymerase according to the protocol of the digoxigenin DIG RNA labeling kit (Roche). In situ hybridization was according to Li et al. (30). The negative control probes were prepared in the opposite direction. The ESTs used as probes of in situ hybridization were chosen from the database and cover the IgSF, MHC region genes, and the CD molecules.

Results

Global profile of EST database

A total of 23,095 sequences were generated by randomly sequencing the clones of six cDNA libraries of Chinese amphioxus, including ovary, neurula, gastrula, larva, intestine, and the whole body of an adult animal. In total, 9009 derived consensus sequences were fed to BLASTX search against GenBank, and 3795 were identified under the expect value of 0.01.

Previously, Panopoulou et al. (31) released 14,189 ESTs from two normalized cDNA libraries of another amphioxus species, *Branchiostoma floridae*, and this data set was downloaded and subjected to an overall clustering with our own data collection. Only 2.94% EST clusters were shared by two data sets, these clusters accounted for 9508 ESTs of our own data set and 5547 ESTs of Florida amphioxus, respectively, and these ESTs mainly represented abundant mRNAs. Nonetheless, this implied a considerable amount of novel genes residing in our data set exclusively, although most sequences from both data sets were not supposed to be species-specific.

As major contributors of our whole EST collection (64.7%), libraries of intestine, neurula, and ovary provided 80.1% of immune-related ESTs in this study. The percentages of unique sequences from these three libraries were all above 71%, indicating that these libraries had a good coverage of the low abundant mRNAs. Conversely, as expected from their different sources, the profiles of gene expression, delineated according to BLASTX search results, exhibited apparent discrepancies among the three libraries.

In addition to the global profile, up to 291 EST clusters (accounting for 937 ESTs) were sorted out as immune-related EST clusters from six libraries. These clusters of EST were broken down into 10 catalogs (Table I), among which the most interesting entries in our dataset were listed in supplemental Table 1.[5] In this immune-related data set, 183 ESTs were singletons, and the other 754 ESTs were grouped into 108 clusters, of which 18 clusters contained over 10 copies.

ESTs for these immune-related genes had an uneven distribution in six different cDNA libraries. A pairwise *U* test was introduced to examine the immune-related EST ratios (data not shown) and indicated that the intestine library was the most important source of immune-related ESTs, in accordance with the fact that intestine was a major battlefield for fighting against foreign pathogenic invasion in amphioxus.

The genes encoding IgSF members

Through the recombination of multiple copies of various interspersed V, (D)J, and C segments in somatic cells, Igs and TCRs have developed the diversity to recognize various Ags. These two pivotal effectors of adaptive immunity are generally believed to be restricted in jawed vertebrates (32). Igs and TCRs share an essential "Ig-fold" that comprises a sandwich-like domain of two β-pleated sheets and they belong to the IgSF. According to the length between two cysteines and the number of the β-sheets, the IgSF domain can be classified into four types: V, C1, C2, and I (33). C1 type is restricted in some molecules only in jawed vertebrates such as TCR, Ig, MHC class I and class II, tapasin, signal regulatory proteins, and butyrophilins. However, the other three types are widespread in several kinds of molecules through the

[5] The online version of this article contains supplemental material.

Table I. *cDNA analysis of the immune-related gene*

Functional Group (total cDNA amount)	Total Amount		Gastrula	Intestine	Larva	Neurula	Ovary	Adult
	Cluster	EST	(2248)	(2114)	(1840)	(8265)	(4556)	(4072)
Genes may be involved in rearrangement	3	7	0	0	2	3	2	0
Genes of IgSF	2	4	0	4	0	0	0	0
Genes controlling lymphocyte ontogeny	18	22	1	3	2	8	4	4
Genes involved in lymphocyte signaling	67	180	8	28	20	92	20	12
Genes required for lymphocyte proliferation and migration	5	40	4	7	4	18	7	0
Genes involved in Ag presentation	60	238	17	19	14	124	56	8
Other immune-related genes	88	346	12	43	43	172	53	23
Adhesion molecules	7	10	0	1	3	6	0	0
Apoptosis-related genes	20	51	0	27	2	12	8	2
Other CD molecules	21	39	4	12	0	17	5	1
Total	291	937	46	144	90	452	155	50

metazoa. Using the short-primer PCR, largely based on the conserved short motifs in the Ig-fold along with other methods (34), many genes bearing the typical V-C structures have been isolated (10, 35, 36) and most of them function as the receptors of innate immunity, suggesting the evolutionary relationship between innate immunity and adaptive immunity (37). Recently, some IgSF members bearing the V or V-C structures have been found in ascidian genome (12) and in amphioxus (10, 11). The architecture of these genes showed more or less similarities to the Ag receptors and provided insights to the primitive forms of Ag binding receptors.

In our intestine cDNA library of Chinese amphioxus, two kinds of IgSF members are found: one is the VCBP and the other is a novel member of IgSF. The VCBP molecules in our Chinese amphioxus library have just ~70% identities in amino acid sequence with the VCBP4 of the Florida amphioxus. In addition, the different splicing forms of the VCBP have been detected by RT-PCR (data not shown). The alternate splicing that operates in this multigene family suggests its potential diversity and flexibility to different ligands.

The novel member of IgSF encodes 338 amino acid residues and comprises a signal peptide, one V-type domain, one C-type domain, a transmembrane region, and a cytoplasmic region, and named as V and C domain-bearing protein (VCP). In terms of the entire protein organization, this gene is dramatically different from the VCBPs and the CD47-like (V domain-bearing) genes found previously in amphioxus. Despite the variety of the Ag binding receptors, the amino acid sequences of IgSF domains still have some canonical positions at residues of G16, C23, W41, L89, D98, Y102, and C104 (10, 33, 34). Alignments of the first Ig domain of VCP with the other V-type domain of rearranging and nonrearranging IgSF molecules indicate that some canonical residues are highly conserved in VCP, especially the residues that are crucial for the formation of the Ig-fold structure (Fig. 1). The secondary structure prediction programs (Jpred and PSIpred) (38, 39) indicate the presence of eight β-strands in the V domain and seven in the C domain (Fig. 2). Through IgBLAST server at NCBI (⟨www.ncbi.nlm.nih.gov/igblast/⟩), two loops corresponding to the CDR1 and CDR2 of Ag receptor molecules can be inferred from the V domain of VCP, whereas a loop equivalent to the CDR3 loop of Ag receptor molecules also can be identified between F and G β-strands of the V domain (Fig. 2). However, the typical diglycine bulge (GxG or FGxG) of J region in the Ag receptor molecules is missing. The consensus pattern, [FY]-x-C-x-[VA]-x-H, is found to exist in the F β-strand of the second IgSF domain of VCP within the C1 type IgSF domain for the formation of conserved disulfide bond according to the analysis using the PROSITE program (29) (Fig. 2). An extra cysteine exists between the C domain and transmembrane region of VCP, suggesting the possibility of the dimerization of two VCP molecules or other partners. Along with CDRs, VCP also contains some predicted phosphorylation sites, such as

FIGURE 1. The alignment of amino acid sequences of V type IgSF domain of VCP with the other V-type domain of rearranging and nonrearranging IgSF molecules. Periods indicate identity residues with the sequence of VCP and dashes indicate gaps introduced to improve the alignment. Four conserved residues in all sequence are indicated with shadow. BTIP, brain and testis-specific IgSF protein; JAM, junctional adhesion molecule; NITR, novel immune-type receptor; SIRP, signal regulatory protein. Accession numbers: BTIP, NP_689751; CD83, AAO6299; TCR-δ, AAF26858; butyrophilin-like, AAC05289; Ig-λ, AAM76525; ChT1, AAD17523; NITR, CAD12507; VCBP4, AAN62910; IgG-κ, AAB97649; TCR-α, CAA26435; SIRP-α, CAA71403; CD86, NP_062261; VDB, AAQ57589; CD80, NP_033985; Nectin, AAG22808; JAM, AAH65309.

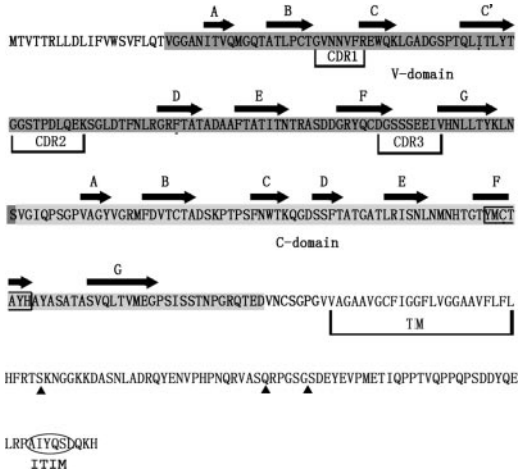

FIGURE 2. The secondary structure and motif analyses of two IgSF domains of VCP. The V and C domains are shaded. Predicted β-strands A–G are indicated with arrows. A small consensus pattern in C domain, [FY]-x-C-x-[VA]-x-H, is boxed. Phosphorylation sites (▲) are shown. TM, transmembrane region.

protein kinase C phosphorylation site and casein kinase II phosphorylation site in the cytoplasmic region (Fig. 2), suggesting the ability of VCP to mediate the signal transduction through these phosphorylation sites. It is intriguing that an ITIM-like motif is found in the cytoplasmic region of VCP, indicating the possibility of the existence of ITIM-dependent signal transduction (Fig. 2).

In terms of the whole molecule, VCP has similar architecture with the cortical thymocyte marker of *Xenopus* (CTX) family and junctional adhesion molecule (JAM) family, as well as CD80 and CD86. When searched against the NCBI database through the position-specific iterated (PSI)-BLAST program, ignoring the big IgSF members containing many IgSF domains, the V and C domain of VCP shows some similarity with the CTX family members. However, VCP just contains two cysteines in its C domain but not four cysteines as in the CTX family (40).

In this study, in situ hybridization to transverse section through the upper or lower pharyngeal region of sexual-matured amphioxus indicates a selective expression of VCP in the wall cells of pharynx and hepatic cecum, a diverticulum of the intestine that may be the precursor of the liver (or pancreas) of vertebrates (Fig. 3, *A–C*), similar to the observation on the other IgSF molecules such as VCBPs in previous studies (10).

In the somatic rearrangement of the bona fide Ag binding receptor, RAG enzymes play the pivotal role. The origin of these transposon-like genes also interested researchers for a long period (41). As the key participants for Ag binding receptor, RAG enzymes were restricted in the jawed vertebrates. In our EST database, no homologues of RAG1/2 genes were found. Nonetheless, one cDNA clone from amphioxus encoding 210 amino acids was found to be homologous to a mouse gene that facilitates RAG1 gene activation (59). In the human lymphoid progenitor cell line, two separate signals from stromal cells and cytokines can induce the expression of RAG1/2 genes (42). Using the mRNA differential display method, a stromal cell-derived protein was found to be associated with RAG1 gene activation in the lymphoid progenitor cell line (43). Whether the homologue of that activator gene of RAG1 has the activation function in amphioxus will need further studies.

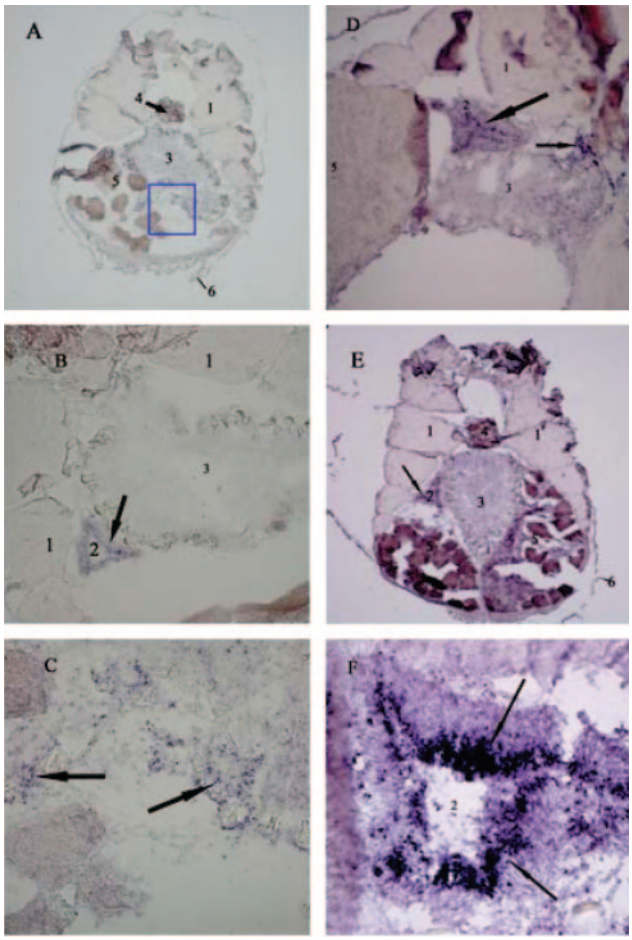

FIGURE 3. The selective expressions of VCP, KE2-like, and Amphi-CD53 in adult amphioxus were revealed by in situ hybridization in transverse sections. *A*, Expression of VCP specifically in pharynx at magnification ×40. *B*, Expression of VCP specifically in pharynx and hepatic cecum at magnification ×100. *C*, Highlighted region of pharynx wall in *A* at magnification ×200. *D*, Expression of KE2-like specifically in pharynx and hepatic cecum, magnification ×100. *E*, Expression of Amphi-CD53 specifically in pharynx and hepatic cecum, magnification ×40. *F*, Expression of Amphi-CD53 specifically in hepatic cecum, magnification ×400. In all sections, tissues and organs are indicated as flowed: 1, myomere; 2, hepatic cecum; 3, pharynx (intestines); 4, notochord; 5, gonad; 6, metapleural fold. Positive signals are shown by arrow.

Homologues of Ag peptide processing and presentation genes and MHC anchor genes

The Ag presentation system in AIS involved many proteases and chaperones, and Danchin et al. (44) demonstrated that many of them gained their specific function in Ag processing and presentation through various levels of co-option. Many homologues of proteasomes and chaperones were found in our database. Besides the genes encoding proteasome (proteasome, macropain) subunit β-type (PSMB)7/10 and PSMB5/8 previously reported (45, 46), other homologues of the housekeeping genes PSMB6/9 and PA28γ were also found in our library. Like the genes found in lamprey (6) and ascidian (12), amphioxus seems to have no inducible forms that change their expression pattern or even their biochemical behavior for Ag processing after genomic duplication. Homologues of the chaperones involved in the process of MHC I molecules assembly were identified in our library such as calreticulin, chaperone protein Dnak (Hsp70), and the glucose-regulated

protein of 78 kDa (GRP78). As for the ATP-binding cassette transporters, just homologues of ATP-binding cassette BI and C sequences were found in our database. A homologue of stress-induced ribosome-associated protein 4, RAMP4, which interacts with nascent MHC class II-associated invariant chains in mammalian lymphocytes (47), was also identified.

An EST similar to IFN-γ-induced lysosomal thiol (GILT) reductase was found in our database. In vertebrate this enzyme is expressed constitutively in APCs and induced by IFN-γ in other cell types. By reducing disulfide bonds under acid condition, GILT can facilitate the processing and presentation of Ag peptides containing disulfide bonds by MHC II molecule to Ag-specific T cells during the proteolysis within lysosomes. In the absence of GILT, presentation of two major lysozyme antigenic epitopes to T cell is partially or completely abrogated (48). The Amphi-GILT found in our database contains the conserved active site CxxC, and nine cysteines downstream of the active site just like the counterparts in human and mouse (Fig. 4). The phylogenetic tree (Fig. 5) shows the evolutionary relationship between Amphi-GILT and other proteins of the GILT family that contain the conserved active domain. According to its sequence similarity and phylogenetic analysis, Amphi-GILT gene is more similar to vertebrate ones than invertebrate, suggesting a closer function to its counterparts in vertebrates. The similar result has been reported in the phylogenetic analysis of immunoproteasomes (46), in which the rate of amino acid substitution has been substantially accelerated in all the IFN-γ-inducible forms. Their accelerated evolving rate beginning with amphioxus might be functionally required for themselves: evolve fast to deal with multifarious foreign antigenic peptides during evolution.

Although there were some genes with Ig domains found in our cDNA libraries, the classical MHC molecules or even homologues of their Ag binding domains had not been discovered as previously studied (13). Recent studies of amphioxus and ascidian genome revealed that just a proto-MHC region existed in these protochordate (12, 44). In our amphioxus libraries, we also found the homologues of BAT5, allograft inflammatory factor 1 (AIF1), RING2 (KE6), KE2 (Fig. 3*D*) and IκBL (NF of κ light polypeptide gene enhancer in B cell inhibitor-like 1), which are linked anchor genes in vertebrates MHC region. The study on the genomic organization of these new anchor genes will help to ultimately reveal the evolution of MHC genomic organization from nonvertebrates to vertebrates.

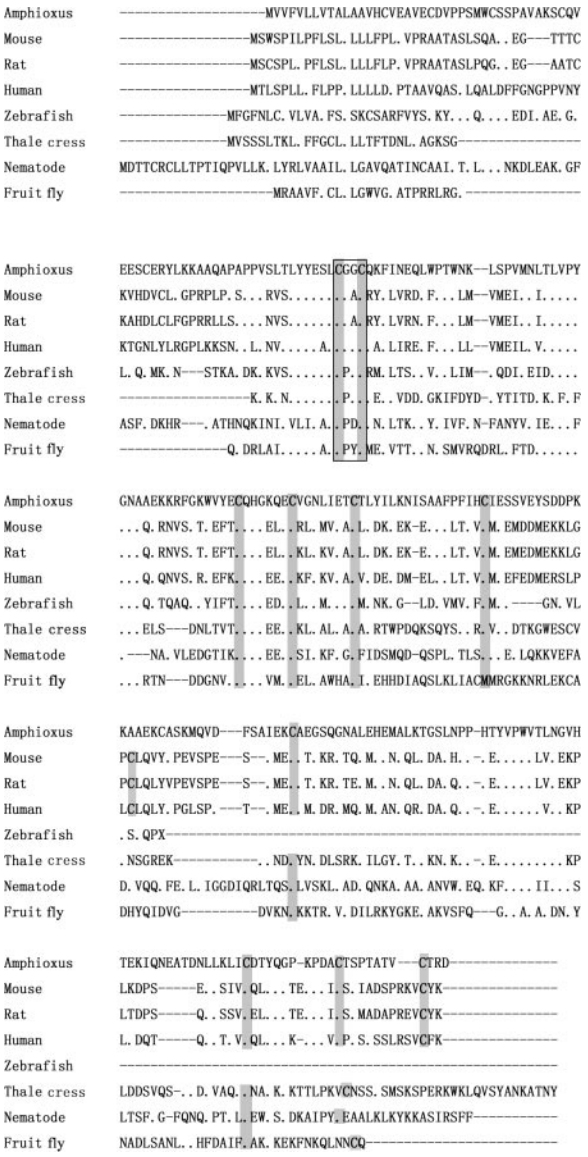

FIGURE 4. The alignment of amino acid sequences of Amphi-GILT with proteins in human, mouse, and putative proteins in zebrafish, nematode, and cress. Periods indicate identity residues with the sequence of Amphi-GILT and dashes indicate gaps introduced to improve the alignment. The motif of active site is boxed and the conserved cysteines are shaded.

Homologues of genes involved in immune signaling

In our database, homologues to many kinds of conserved enzymes used not only by signaling of AIS but also by general signal transduction were identified, including protein tyrosine kinase, protein tyrosine phosphatase, MAPK, MEK, and small GTPase (Ras, Rho, and Ran). Some homologues to transcriptional factors, the final effectors of the signal transduction, were found in our libraries, such as Collier/Olf1/early B cell factor (COE/EBF) gene family member, Ets, and lymphoid transcription factor (LyF). EBF belongs to a conserved helix-loop-helix gene family. The EST homologous to EBF in our database was found to have 80% identity over the entire molecule and >90% identity over a stretch of 370 residues, suggesting its essential feature of dimerization and DNA binding. In situ hybridization also indicated the expression of Amphi-EBF in pharynx and hepatic cecum (data not shown). In addition, homologues to B cell adaptor for PI3K and CD3ϵ-associated signal transducer found recently in lymphocyte-like cells library in lamprey (7) were detected in our database. Our database also contains sequences encoding several other mammalian homologues involved in the signaling pathways shared by many different cell types including lymphocytes, such as TNFR and pre-B cell colony-enhancing factor. Though the existence of these signal-related molecules was not regarded as the direct evidence for the existence of AIS signaling in amphioxus, they provided a sufficient preparation for the appearance of the AIS.

In contrast, many homologues of CD molecules were identified in amphioxus, such as CD9, CD20, CD53, CD63, CD81, and CD82. Some of these CD molecules, such as CD9, CD53, CD63, and CD81, belong to the tetraspanins family that feature four transmembrane domains, a small extracellular loop 1 (EC1), and a mushroom-like large EC2 (49). Functionally, tetraspanins act as molecule organizers of membrane microdomains and multimolecular complexes, such as signaling complexes on the lymphocyte surface (50). By their nature, tetraspanins widely participate in cell migration, motility, adhesion, activation, and proliferation. One member of the tetraspanins family, CD53, is specifically expressed

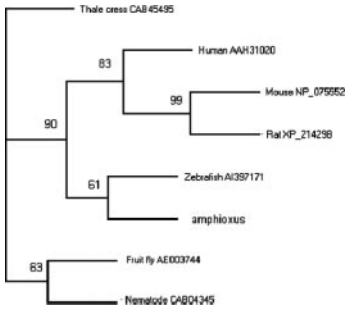

FIGURE 5. The phylogenetic analysis of Amphi-GILT. The topological structure of the trees for Amphi-GILT constructed by NEIGHBOUR and PROTPARS programs is the same. The number at the nodes represents the percentage of 100 bootstrap pseudo-replicated that contained the cluster distal to the node. The sequences used for constructing the phylogenetic tree are the same as the sequences used in the alignment described in Fig. 4.

in B cells, T cells, dendritic cells, monocytes, NK cells, and granulocytes, and involved in T cell activation, positive selection, and B cell integrin adhesion (51). In our database we found a CD53 homologue with 37% identity and 55% similarity to the mouse CD53 and has a conserved site of potential *N*-glycosylation in EC2. However, unlike the canonical pattern of mammalian CD53 (CCG, DW, PxSC, GC), this amphioxus homologue has two extra cysteines in the EC2 (Fig. 6). The expression of the Amphi-CD53 gene was detected by in situ hybridization of adult amphioxus and the signal was specifically located in the pharynx and hepatic cecum particularly in the pharynx wall (Fig. 3, *E* and *F*). The tetraspanin family has many members involved in signal transduction including immune-related signaling between cells. Although some homologues of the family members were found in our database, whether they have the immune function needs further study. Just like the other families whose members are involved in immune function, a certain member in the tetraspanin family may be waiting for obtaining the new immune-related function in amphioxus.

Comparative analyses between different organisms

Because many homologues of housekeeping genes involved in not only AIS but also other biologic systems have been detected in our database, the comparative analyses of these genes are performed in this study. The results of domain analyses are indicated in Table II, in which the two members of IgSF show intriguing results. For VCP, when using the whole sequence to search against the database of model organisms, there is no distinct homologue, whereas VCBP just has homologue in ascidian. In contrary, the domain analyses reveal the fact that Ig domains in VCP and VCBP are present in many kinds of molecules in metazoa, but the Ig domains coming from different molecules show lower similarities with each other. For VCBP, the chitin-binding domain is also widespread in metazoa. These comparative analyses indicate some scenarios of the evolution of these two molecules. Although the functions of these two IgSF members are unclear now, but for the molecular characteristics of these genes, they might have their function in the defense system in amphioxus. To deal with the increasing diversity of ligands they might evolve rapidly through changes of sequence or expression pattern to have novel function, followed by the recruitment into new interaction pathways. Another scenario is that with the appearance of V(D)J rearrangement machinery, these two IgSF members become nonessential for the defense system, thus lost after evolutionary split of cephalochordate/vertebrate. However, their domains became the "candidates" of some shuffling events and were involved with other new molecules existing in vertebrates.

For other molecules homologous to the genes involved in signal transduction, Ag peptide processing and presentation, the results of domain analyses, indicate that the domains in some of these genes are widespread in metazoa. These molecules might be the members of some families sharing the conserved domain and involve some important biologic system common in metazoa. Although the function of these molecules in amphioxus is unknown, it is possible that when the key molecules in AIS come into being, these ordinary families might gain a chance to obtain new members with new functions (or still retain their original functions) just for AIS.

Discussion

The evolutionary origin of acquired immune mechanism has been one of focal interests in immunology for a long period (2–4, 27). Owing to their important roles in the adaptive immunity and special somatic rearrangement that produces the diversity, the two Ag binding receptors, TCRs and Igs, have been the essential focus of recent studies (52). The mechanism used by vertebrates to diversify their Ag binding receptors might not be the one used by invertebrates, but at least some invertebrates and pregnathostome vertebrates could use other mechanisms to generate enough diversity to fight against the varied pathogens as indicated in recent reports (8, 9). It was assumed that the rearranging Ag receptor genes derive from a single primitive nonrearranging form in a series of transpositions and duplications (53, 54). Even though none of the bona fide rearranging Ag receptor molecules has been found in urochordate and cephalochordate, many membrane-bound molecules bearing the Ag receptor architecture including the novel V-C structure were found in our database as well as in

FIGURE 6. The alignment of amino acid sequences of Amphi-CD53 with other homologues. The alignment produced by T-Coffee and rendered by GeneDoc software (⟨www.psc.edu/biomed/genedoc⟩). Periods indicate identity residues with the sequence of Amphi-CD53 and dashes indicate gaps introduced to improve the alignment. The four transmembrane domains (*) are shown: EC1 locates between the 1st and 2nd transmembrane domains, EC2 locates between the 3rd and 4th transmembrane domains. The canonical motifs of mammalian CD53 are displayed on each *bottom row*. The two extra cysteines are shaded. Accession numbers: human, AAH40693; rat, NP_036655; mouse, NP_031677; African clawed frog, BC077949; zebrafish, NM_001002186; sea squirt, AK112403.

```
Amphioxus           ----------MEGCGLQCIKYLLFAFNLLFWLAGAGHVGVGIYVMVA-YGVSNVPQLLNLSELLLRGGGIALIVVGVITHIVG
Human               ----------.GHSS.KLL..V..F.....IC.CCIL.F...LLI-HDNFGVL.H----NLPS.TL.NV-FVI..S.I.V.A
Rat                 ----------.GHSS.KLL..V..F..F...VC.CCIL.F..HLL.-QNTYGIL.R----NLPF.TL.NV-.VI..S.I.V.A
Mouse               ----------.GHSS.KLL..V..I.....VC.CCIL.F...FL.-QNTYGVL.R----NLPF.TL.N.-.VI..S.I.V.A
African_clawed_frog HRKYKAIITEFKKCN.ARGC.C.L..MM.L...I...C.C.LL....WLS.SQGNFATFSP-----.FPS.SAANL-V..I.TVV.VT.
Zebrafish           ---------EFEKLK.ARGC.C.V..MM.L.......S.C.LL....WLS.SQGSFATFSP-----.FPS.SAANL-V.TL.SVV.VT.
Sea_squirt          --------FWTKTRK..SK..N.C...M.L....IF.C.CVLL.LSVWAY.NADSFKKIIS-----.DPVIFNSL.V..A..AVLIVA.
                                        1*************************1                2*******************

Amphioxus           FLGCCGAIKENRCHLGAFFTLLLILLLVEVGVAIYAFVQRGQFFSLIGTVVDTTSTVPYTSLDQAA--RTLVDFTQTYLKCCGHNDTN
Human               ....M.S....K.L.MS..I....I..A..TL..LL..YEQKLNEYVAKGLT--DSIHRYHS.NST--KAAW.SI.SF.Q...I.G.S
Rat                 ....M.S....K.L.MS..V...LI..A..TL..LL..YEKKINT.VAEGLN--DSIQHYHS.NST--.MAW..I.SQ.Q...V.GSS
Mouse               ....M.S....K.L.MS..V....I..A..TI..LL..YEQKLNT.VAEGLN--DSIQHYHS.NST--HKAW..I..Q.Q...V.GSS
African_clawed_frog ....L......K.L.LS..II...I..T.LILL.LF..YMDKVNENAKQDLK--DGLLLYNSENNVGLKNAWNII.AEMH...VT.YT
Zebrafish           ....L......K.L.LS..IV...I..A.LILL.LF..YTEKVSENAKQDLK--DGLRLYNT.NNVGL.NAWNII.AEWQ...VTGLS
Sea_squirt          .F..L...Q.SK.L..T...IV.VIFIA.IVG..LIY.YYPKAKD.ALQSHQ------NYNT--TT--KQPW.IL..TF....FTNYS
                    *2              3*********************3                                        CCG

Amphioxus           NYNP-------TPASCSAASVSTTSITGGN-YNQPCKTAIISFLAEKSLIVCGIGIGIGLTQLIGHIFACCLFQAKGKESDVV--
Human               DW------TSGP....PSDR--K----------VEG.YAKARLWFHSNF.YI.I.T.CVCVIEVL..S..LT.NCQID.T.QTIGL
Rat                 DW------ISGP.S..PSG.--D----------V.G.YKKGQAWFHSNF.YI.IVT.CVCVI.VL..S..LT.NCQID.T.QALGL
Mouse               DW------TSGP.S..PSG.--D----------V.G.YNKAK.WFHSNF.YI.I.T.CVCVI.VL..S..LT.NCQID.T.QALGL
African_clawed_frog DWY.VLGENTV-.DR.QEN.-QDGHNSTSLVWKTG.YEKVKNWFDDNKHVL.T..MC.LII.IL..A.SMT...QIHRTGKKYDA
Zebrafish           DWHEALQEKSV-.DR.QEH-YTEGRN.TNVFWS.G.YEKVEEW.NDNKHLL.T.AHCVLVL..L..A.SHT.Y.QIHRAGKKYDA
Sea_squirt          DW------GSTI.ST.QSD.-A.PP.P-.LITF.PLG.EA..RKYF----W.I.......IVFEILA.........NI.DYEMA---
                    DW          PxSC                  GC          4*********************4
```

Table II. *Comparative analyses of the domains of some ESTs in our database in different organisms*

ID	Similarity to	Nematode	Fruit Fly	Ascidian	Amphioxus	Zebrafish/Fugu	Mouse	Human
Members of IgSF[a]								
2234	VCP (Ig-like PF00047)	−	−	−	+	−	−	−
2372	VCBP (Ig-like PF00047, CBD, PF01607)	−	−	+	+	−	−	−
Genes may be involved in rearrangement								
2806	RGA (MtN3_slv PF03083)	+	+	+	+	?	+	+
Transcription factors								
4140	LyF (Znf_C2H2 PF00096)	+	+	+	+	+	+	+
1467	IκBL (ANK PF00023)	+	+	+	+	+	+	+
898	ETS (ETS PF000178)	+	+	+	+	+	+	+
4966	COE (TF_COE PS01345)	+	+	+	+	+	+	+
Genes involved in immune signaling								
4530	CAST (Calpain_inhib PF00748)	−	−	+	+	+	+	+
6024	PBEF (NAPRTase PF04095)	+	+	+	+	+	+	+
1596	CD20 (CD20 PF04103)	−	−	+	+	+	+	+
4626	CD75 (Glyco_trans_29 PF00777)	−	+	+	+	+	+	+
2195	CD53 (Transmem_4 PF00335)	+	+	+	+	+	+	+
Genes involved in Ag processing and presentation								
1100	PSMB5/8 (protsme_protease PF00227)	+	+	+	+	+	+	+
676	PSMB6/9 (protsme_protease PF00227)	+	+	+	+	+	+	+
1628	PSMB7/10 (protsme_protease PF00227)	+	+	+	+	+	+	+
972	GILT (GILT PF03227)	+	+	+	+	+	+	+
3311	KE2 (KE2 PF01920)	+	+	+	+	+	+	+
1259	GRP78 (Hsp70 PF00012)	+	+	+	+	+	+	+

[a] For IgSF members, the sequences from amphioxus are used to do the whole molecule analyses and results are indicated here. +, Positive result; −, negative result; ?, information unavailable; the domains in the molecules and the Pfam or PROSITE accession number are showed in the parentheses.

other early chordates (12, 55). Their vertebrate counterparts function as the nonimmune or innate immune effector molecules. Notwithstanding that the functions of those putative ancestors in early chordates needs further study, it is possible that AIS co-opts one of these V-C structural membrane-bound molecules for the function of recognizing Ags with the specific diversification mechanism through some events such as genomic duplication, exon shuffling, and transpositions during evolution.

Although classical MHC molecules are not yet to be identified in amphioxus, several genes with housekeeping functions in Ag processing and presentation have been found, as similarly found in ascidian (12, 13). Our data in this study are consistent with the previous research: the inducible forms of those housekeeping genes that are directly involved in the Ag processing and presentation are not present in amphioxus. Those housekeeping genes, among which the human counterparts map to the different MHC paralogous regions, tended to be in a contiguous genomic fragment in amphioxus and ascidian genome. The absence of the functional forms and the existence of this proto-MHC region revealed the possibility that those housekeeping genes are ready to be co-opted by the bona fide MHC molecules during evolution.

The origin of lymphocytes is another interesting issue (56). Thus far, no distinct lymphoid tissues have been found in agnathan and protochordate by histologic studies but some lymphocyte-like cells were isolated from the intestine and the associated typhlosole of lamprey (5). In our studies, many genes homologous to the counterparts for the mammalian lymphocytes signaling and development were identified, including some transcriptional factors and many enzymes being used in different cell lines in immune system. In contrast, the gene expression pattern analyzed in this study showed amazing similarity among different genes. These genes were expressed mostly in pharynx and intestine (hepatic cecum), the first defense against the foreign pathogens in amphioxus, providing the possibility that the ancestor of lymphocytes might derive from the cells located in intestine through the evolution.

In summary, we identified extensive ESTs homologous to genes participating in immune response in vertebrate. Although none of the key genes specific to the AIS was found, the existence of members belonging to multigene families (IgSF, proteasomes, enzymes, transcription factors, and other) involved in immune or nonimmune responses suggested that precursors of the genes involved in vertebrate AIS may have existed in protochordate. Many of these precursors gained new roles in adaptive immune response when the key genes evolved into the jawed vertebrates. Amphioxus represents the pivotal position during evolution, and it is generally known that some important steps such as the genomic duplication took place at the cephalochordate/vertebrate split (57, 58). This large-scale genomic duplication provided the opportunities for the emergence of the new forms of those multifamily members that may evolve specific roles in the AIS. In short, the adaptive immune system is in its twilight in amphioxus. The Ag binding molecules, Ag processing and presentation molecules, and signal transduction-related molecules were all in their evolution and waiting for immunologic "Big Bang" to be recruited for AIS. As such, the functional analyses on those genes identified and interactions among them in comparison with other vertebrates like human and mouse will be of particular interest for further studies. However, the sequencing of the whole genome of amphioxus along with the analyses of the other protochordate genome sequence might be pivotal to the study of the original status of the immune-related gene organization, especially the genes related to AIS before vertebrate.

Acknowledgments

We thank Dr. Pierre Pontarotti (Université de Provence, Marseille, France) for the helpful advices on the comparative analyses and the critical reading of the manuscript. We thank Dr. Louis Du Pasquier (University of Basel, Basel, Switzerland) for the helpful suggestions on the analyses of the VCP molecules.

Disclosures

The authors have no financial conflict of interest.

References

1. Klein, J., and V. Horejsi. 1997. *Immunology*. Blackwell Science, Oxford.

2. Hughes, A. L. 2002. Natural selection and the diversification of vertebrate immune effectors. *Immunol. Rev. 190:161*.
3. Kasahara, M., T. Suzuki, and L. Du Pasquier. 2004. On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. *Trends Immunol. 25:105*.
4. Du Pasquier, L. 2004. Speculations on the origin of the vertebrate immune system. *Immunol. Lett. 92:3*.
5. Shintani, S., J. Terzic, A. Sato, M. Saraga-Babic, C. O'hUigin, H. Tichy, and J. Klein. 2000. Do lampreys have lymphocytes? The Spi evidence. *Proc. Natl. Acad. Sci. USA 97:7417*.
6. Mayer, W. E., T. Uinuk-Ool, H. Tichy, L. A. Gartland, J. Klein, and M. D. Cooper. 2002. Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl. Acad. Sci. USA 99:14350*.
7. Uinuk-Ool, T., W. E. Mayer, A. Sato, R. Dongak, M. D. Cooper, and J. Klein. 2002. Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc. Natl. Acad. Sci. USA 99:14356*.
8. Pancer, Z., C. T. Amemiya, G. R. Ehrhardt, J. Ceitlin, G. Larry Gartland, and M. D. Cooper. 2004. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature 430:174*.
9. Zhang, S. M., C. M. Adema, T. B. Kepler, and E. S. Loker. 2004. Diversification of Ig superfamily genes in an invertebrate. *Science 305:251*.
10. Cannon, J. P., R. N. Haire, and G. W. Litman. 2002. Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. *Nat. Immunol. 3:1200*.
11. Sato, A., W. E. Mayer, and J. Klein. 2003. A molecule bearing an immunoglobulin-like V region of the CTX subfamily in amphioxus. *Immunogenetics 55:423*.
12. Azumi, K., R. De Santis, A. De Tomaso, I. Rigoutsos, F. Yoshizaki, M. R. Pinto, R. Marino, K. Shida, M. Ikeda, M. Ikeda, et al. 2003. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics 55:570*.
13. Abi-Rached, L., A. Gilles, T. Shiina, P. Pontarotti, and H. Inoko. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet. 31:100*.
14. Castro, L. F., R. F. Furlong, and P. W. Holland. 2004. An antecedent of the MHC-linked genomic region in amphioxus. *Immunogenetics 55:782*.
15. Vienne, A., T. Shiina, L. Abi-Rached, E. Danchin, V. Vitiello, F. Cartault, H. Inoko, and P. Pontarotti. 2003. Evolution of the proto-MHC ancestral region: more evidence for the plesiomorphic organisation of human chromosome 9q34 region. *Immunogenetics 55:429*.
16. Flajnik, M. F., and M. Kasahara. 2001. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity 15:351*.
17. Hirakow, R., and N. Kajita. 1994. An electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: the neurula and larva. *Kaibogaku Zasshi 69:1*.
18. Yasui, K., S. C. Zhang, M. Uemura, S. Aizawa, and T. Ueki. 1998. Expression of a twist-related, Bbtwist, during the development of a lancelet species and its relation to cephalochordate anterior structures. *Dev. Biol. 195:49*.
19. Mou, C.-Y., S.-C. Zhang, J.-H. Lin, W.-L. Yang, W.-Y. Wu, J.-W. Wei, X.-K. Wu, J.-C. Du, Z.-Y. Fu, L.-T. Ye, et al. 2002. EST analysis of mRNAs expressed in neurula of Chinese amphioxus. *Biochem. Biophys. Res. Commun. 299:74*.
20. Ewing, B., L. D. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Res. 8:175*.
21. Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using PHRED. II. Error probabilities. *Genome Res. 8:186*.
22. Miller, R. T., A. G. Christoffels, C. Gopalakrishnan, J. Burke, A. A. Ptitsyn, T. R. Broveak, and W. A. Hide. 1999. A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledgebase. *Genome Res. 9:1143*.
23. Burke, J., D. Davison, and W. Hide. 1999. D2_cluster: a validated method for clustering EST and full-length cDNA sequences. *Genome Res. 9:1135*.
24. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25:3389*.
25. Felsenstein, J. 2000. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genetics, University of Washington, Seattle.
26. Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol. 17:540*.
27. Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. L. Sonnhammer, et al. 2004. The Pfam protein families database. *Nucleic Acids Res. 32:D138*.
28. Hofmann, K., P. Bucher, L. Falquet, and A. Bairoch. 1999. The PROSITE database: its status in 1999. *Nucleic Acids Res. 27:215*.
29. Clamp, M., D. Andrews, D. Barker, P. Bevan, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, et al. 2003. Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Res. 31:38*.
30. Li, S., F. Qiu, A. Xu, S. M. Price, and M. Xiang. 2004. Barhl1 regulates migration and survival of cerebellar granule cells by controlling expression of the neurotrophin-3 gene. *J. Neurosci. 24:3104*.
31. Panopoulou, G., S. Hennig, D. Groth, A. Krause, A. J. Poustka, R. Herwig, M. Vingron, and H. Lehrach. 2003. New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res. 13:1056*.
32. Laird, D. J., A. W. De Tomaso, M. D. Cooper, and I. L. Weissman. 2000. 50 million years of chordate evolution: seeking the origins of adaptive immunity. *Proc. Natl. Acad. Sci. USA 97:6924*.
33. Harpaz, Y., and C. Chothia. 1994. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol. 238:528*.
34. Hawke, N. A., J. A. Yoder, and G. W. Litman. 1999. Expanding our understanding of immunoglobulin, T-cell antigen receptor, and novel immune-type receptor genes: a subset of the immunoglobulin gene superfamily. *Immunogenetics 50:124*.
35. Strong, S. J., M. G. Mueller, R. T. Litman, N. A. Hawke, R. N. Haire, A. L. Miracle, J. P. Rast, C. T. Amemiya, and G. W. Litman. 1999. A novel multigene family encodes diversified variable regions. *Proc. Natl. Acad. Sci. USA 96:15080*.
36. Yoder, J. A., M. G. Mueller, S. Wei, B. C. Corliss, D. M. Prather, T. Willis, R. T. Litman, J. Y. Djeu, and G. W. Litman. 2001. Immune-type receptor genes in zebrafish share genetic and functional properties with genes encoded by the mammalian leukocyte receptor cluster. *Proc. Natl. Acad. Sci. USA 98:6771*.
37. Van den Berg, T. K., J. A. Yoder, and G. W. Litman. 2004. On the origins of adaptive immunity: innate immune receptors join the tale. *Trends Immunol. 25:11*.
38. Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay, and G. J. Barton. 1998. Jpred: a consensus secondary structure prediction server. *Bioinformatics 14:892*.
39. McGuffin, L. J., K. Bryson, and D. T. Jones. 2000. The PSIPRED protein structure prediction server. *Bioinformatics 16:404*.
40. Chretien, I., A. Marcuz, M. Courtet, K. Katevuo, O. Vainio, J. K. Heath, S. J. White, and L. Du Pasquier. 1998. CTX, a *Xenopus* thymocyte receptor, defines a molecular family conserved throughout vertebrates. *Eur. J. Immunol. 28:4094*.
41. Agrawal, A., Q. M. Eastman, and D. G. Schatz. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature 394:744*.
42. Tagoh, H., H. Kishi, A. Okumura, T. Kitagawa, T. Nagata, K. Mori, and A. Muraguchi. 1996. Induction of recombination activating gene expression in a human lymphoid progenitor cell line: requirement of two separate signals from stromal cells and cytokines. *Blood 88:4463*.
43. Tagoh, H., H. Kishi, and A. Muraguchi. 1996. Molecular cloning and characterization of a novel stromal cell-Derived cDNA encoding a protein that facilitates gene activation of recombination activating gene (RAG)-1 in human lymphoid progenitors. *Biochem. Biophys. Res. Commun. 221:744*.
44. Danchin, E., V. Vitiello, A. Vienne, O. Richard, P. Gouret, M. F. McDermott, and P. Pontarotti. 2004. The major histocompatibility complex origin. *Immunol. Rev. 198:216*.
45. Hayashi, M., T. Ishibashi, K. Tanaka, and M. Kasahara. 1997. The mouse genes encoding the third pair of β-type proteasome subunits regulated reciprocally by IFN-γ: structural comparison, chromosomal localization, and analysis of the promoter. *J. Immunol. 159:2760*.
46. Takezaki, N., Z. Zaleska-Rutczynska, and F. Figueroa. 2002. Sequencing of amphioxus PSMB5/8 gene and phylogenetic position of agnathan sequences. *Gene 282:179*.
47. Schroder, K., B. Martoglio, M. Hofmann, C. Holscher, E. Hartmann, S. Prehn, T. A. Rapoport, and B. Dobberstein. 1999. Control of glycosylation of MHC class II-associated invariant chain by translocon-associated RAMP4. *EMBO J. 18:4804*.
48. Maric, M., B. Arunachalam, U. T. Phan, C. Dong, W. S. Garrett, K. S. Cannon, C. Alfonso, L. Karlsson, R. A. Flavell, and P. Cresswell. 2001. Defective antigen processing in GILT-free mice. *Science 294:1361*.
49. Boucheix, C., and E. Rubinstein. 2001. Tetraspanins. *Cell Mol. Life Sci. 58:1189*.
50. Hemler, M. E. 2003. Tetraspanin proteins mediate cellular penetration, invasion, and fusion events and define a novel type of membrane microdomain. *Annu. Rev. Cell. Dev. Biol. 19:397*.
51. Tarrant, J. M., L. Robb, A. B. van Spriel, and M. D. Wright. 2003. Tetraspanins: molecular organisers of the leukocyte surface. *Trends Immunol. 24:610*.
52. Marchalonis, J. J., I. Jensen, and S. F. Schluter. 2002. Structural, antigenic and evolutionary analyses of immunoglobulins and T cell receptors. *J. Mol. Recognit. 15:260*.
53. Marchalonis, J. J., and S. F. Schluter. 1998. A stochastic model for the rapid emergence of specific vertebrate immunity incorporating horizontal transfer of systems enabling duplication and combinational diversification. *J. Theor. Biol. 193:429*.
54. DuPasquier, L., and I. Chretien. 1996. CTX, a new lymphocyte receptor in *Xenopus*, and the early evolution of Ig domains. *Res. Immunol. 147:218*.
55. Du Pasquier, L., I. Zucchetti, and R. De Santis. 2004. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev. 198:233*.
56. Anderson, M. K., R. Pant, A. L. Miracle, X. Sun, C. A. Luer, C. J. Walsh, J. C. Telfer, G. W. Litman, and E. V. Rothenberg. 2004. Evolutionary origins of lymphocytes: ensembles of T cell and B cell transcriptional regulators in a cartilaginous fish. *J. Immunol. 172:5851*.
57. Holland, P. W. 1999. Introduction: gene duplication in development and evolution. *Semin. Cell. Dev. Biol. 10:515*.
58. Holland, P. W. 1999. Gene duplication: past, present and future. *Semin. Cell. Dev. Biol. 10:541*.
59. Dong, M., Y. Fu, C. Yu, J. Su, S. Huang, X. Wu, J. Wei, S. Yuan, Y. Shen, and A. Xu. Identification and characterization of an activation gene for the recombination activating gene 1 (RAG1) in amphioxus. *Fish Shellfish Immunol. In press*.