<u>**Proposal for construction of a nurse shark (*Ginglymostoma cirratum*) BAC library**</u>

Martin Flajnik and Yuko Ohta, Dept. of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201

1. The importance of the nurse shark to biomedical or biological research

      The cartilaginous fish, specifically the elasmobranchs (sharks, skates, and rays) are the oldest group of living vertebrates shown to have developed an adaptive immune system with underlying molecules and mechanisms similar to those of mammals. This ancient taxon is the oldest with jaws and diverged from the common ancestor of all vertebrates 460-520 million years ago (the oldest group of jawed vertebrates, the placoderms, are extinct). A second evolutionary wave occurred about 250 million years ago, which gave rise to extant elasmobranchs (1). Studies of this old group, in comparison to representatives in other vertebrate taxa, allow us to theorize about fundamental genetic, developmental, and functional characteristics in the common ancestor of all vertebrates. Our work focuses on evolution of the adaptive immune system (2), and our aim is to uncover those essential elements in all immune systems as well as to perhaps reveal some characteristics that were evolutionary forerunners of the human system (the former task obviously much easier than the latter!) The comparative approach has been instrumental in deciphering the necessary characteristics of the adaptive immune system (3). Among elasmobranch species, we and others use the nurse shark *Ginglymostoma cirratum* to study the genetics, biochemistry, and mechanisms that regulate the immune system. Compared to other shark species, the nurse shark genome size is relatively small, only slightly larger than that of humans. We believe that this vertebrate is a candidate for becoming a model species in the biomedical and general biological field.

2. Uses to which the BAC library would be put, in addition to genomic sequencing

      Our laboratory is most interested in the adaptive immune system. While all living organisms possess innate immunity to defend themselves from invading pathogens, adaptive or acquired immunity is found only in the jawed vertebrates. In search of the origins of adaptive immunity, several groups have focused on the oldest vertebrates shown to have an adaptive immune system, the elasmobranchs (2). Studies of this evolutionarily old group have shown that the basic elements of this system (Immunoglobulin (Ig), T cell receptors (TCR), Major Histocompatibility Complex (MHC), gene rearrangement, somatic hypermutation, and segregation of lymphocytes into secondary lymphoid tissues) were all present in the ancestor of elasmobranchs and mammals (3), suggesting that in a relatively short period of evolutionary time all of these essential elements emerged in order to effect and to regulate this new manner of distinguishing self from non-self. However, elasmobranchs also have unique functional and genetic features of their immune systems that may reveal novel mechanisms of gene regulation. Here we will touch on the Ig and MHC systems.

Unusual Ig genes in the cartilaginous fish
To assemble a functional antibody heavy (H) chain gene, variable (V) genes must recombine with diversity (D) and joining (J) segments (for light (L) chains, only V and J

recombine). This rearrangement at the DNA level results in a complete, functional 'VDJ' gene encoding the antigen recognition segment of the antibody. At the RNA level this VDJ is spliced to constant (C) region exons, and the mature mRNA is translated into a complete Ig chain.

In 1986 Gary Litman's group discovered that IgM heavy chain genes from the horn shark (*Heterodontus francisci*) are arranged in the so-called "cluster-type" organization (i.e. [VDJC]n) rather than the streamlined "translocon" organization (i.e. VnDnJnC) found in all tetrapod species (4). In this species there are perhaps 200 clusters of H chain genes, and this basic finding has been extended to all cartilaginous fish examined and to all of the different immunoglobulin isotypes. One one hand, this cluster organization seems primordial, with duplication of rearranging clusters preceding duplication of individual gene segments. On the other hand, the cluster organization presents problems with gene expression. In the translocon organization one immediately can see how regulation of expression can occur: after rearrangement, all of the D segments are deleted at the Ig locus between the V and J segments and therefore rearrangement cannot proceed; furthermore, since the VDJ rearrangements result in out-of-frame junctions in two out of three cases, and there are only two H chain loci, expression of only one H chain/cell is expected. But how is regulation of the hundreds of elasmobranch clusters accomplished? Preliminary evidence from single cell PCR experiments suggests that indeed only one IgH chain gene is expressed/cell (5), but we know nothing of the mechanism that regulates this expression. The problem is further complicated by the fact that some of the genes have already been rearranged in rare germline events ("germline-joined genes" refs 6,7), so it would appear that mere expression of an H chain from such genes does not extinguish the rearranging machinery. Finally, all elasmobranchs studied have two other isotypes called IgNAR and IgW, which are also encoded in IgM-like clusters. Preliminary evidence suggests that each lymphocyte expresses only one isotype from the earliest stages of its differentiation. Isolation of all of the IgM, IgNAR, and IgW H chain genes from one individual is the obvious first step in understanding the regulation of expression. The nurse shark seems to have fewer IgM H chain genes than the horn shark (~50 rather than ~200, ref 8), which is perhaps reflective of its smaller genome size (see below), and there are only a few IgNAR and IgW genes/haploid genome (9,10).

Assembly of the MHC
While most immunological studies in elasmobranchs have focused on the Ig system, genetics of the MHC has been analyzed only in nurse sharks. The MHC is a large genetic region (400 Kb in humans) encoding the so-called class I and class II molecules that bind to foreign antigens (in the form of peptides) and then display these antigens to T cells. In addition, genes with no structural similarity to class I/II but that are important for the "processing" of proteins for display to T cells are also encoded in the MHC. Furthermore, other genes of immunological interest having no obvious role in the processing or presentation of foreign antigens to T cells are also found in the MHC in the so-called class III region (e.g. the complement components C3 and C4). Thus, determining the evolutionary timescale for the accumulation of these genes in the MHC is obviously of great interest (11).

Exogenous foreign antigens are taken up by cells and transported into intracellular vesicles. Here the proteins are degraded into peptides by lysosomal enzymes, which are then "loaded" onto MHC class II molecules for presentation to CD4-positive "helper" T cells. Such T cells stimulate B cells to produce antibodies or activate macrophages to destroy intracellular pathogens. Upon virus infection, misfolded viral are processed in the cytosol by the multicatalytic proteasome, transported by a specialized transporter (TAP) into the endoplasmic reticulum, loaded onto class I molecules and then presented to CD8-positive "killer" T cells that then destroy the infected cells. Nurse shark class I and class II genes have been isolated and show unambiguous structural similarity to the mammalian counterparts (2). Additionally, in mammals the catalytic components of the proteasome are substituted upon viral infection, which changes the specificity of cleavage sites on proteins to make the generated peptides more suited for binding to class I molecules. These specialized proteasome genes (*LMP2* and *LMP7*), as well as the genes encoding the aforementioned TAP transporter proteins, have also been isolated from nurse sharks and shown to map to MHC (12).

In the human and mouse the MHC is organized into the order class II—class III—class I. Surprisingly, the proteasome and TAP genes involved in the class I pathway are embedded in the middle of the class II region. By contrast, in the MHC of all non-mammalian species studied to date, there are usually a low number of class I genes linked tightly to the proteasome and TAP genes, forming a true "class I region," which is likely to be primordial. In all teleost fish examined, the "class I region" is further pronounced because the class II and class III genes are found on other chromosomes (see below).

Our analyses of nurse shark MHC with family studies and cosmid cloning (see below) strongly suggest that a "class I region" exists in nurse sharks as well, but clearly a complete physical map is needed. "Class I region" genes reside within a 200-300Kb region in teleosts such as the medakafish, zebrafish, and pufferfish. In the human class I region, the TAP and proteasome genes exist within 100Kb. Our preliminary analysis of MHC genes isolated from cosmid (avg. insert size 40Kb) and lambda phage (avg. insert size 8-17Kb) genomic libraries revealed unexpectedly large gene sizes; all genes studied to date have large introns (3-5X that of human), and intergenic distances are also disappointingly huge, as it is rare to find two genes within a single cosmid clone. Thus, acquisition of a large-insert library (~100Kb) is obviously required.

Comparisons to teleosts
Teleost (e.g. pufferfish) genome sequencing and mapping projects have been underway in the past few years. The small genome sizes of some teleosts, and the ability to produce mutants and manipulate embryos have made these species attractive models. However, analysis of teleost MHC has demonstrated that despite the tight linkage of LMP/TAP/class I genes in the "class I region," class II and class II region genes are not found in a supergene cluster but instead are scattered on many different chromosomes. By contrast, we have shown in nurse sharks that class I and class II genes are linked (13), and recently we found that, like in the tetrapods, the complement (class III) genes C4 and factor B are also MHC-linked. Such data from elasmobranchs prove that the teleost MHC has been rent apart, either through translocations or because of differential

3

silencing of MHC genes after a teleost-specific genome-wide duplication, suggested by studies of homeobox (and other) genes (14,15). If this disruption of ancient syntenies is a common feature in teleosts, then these species are clearly not the genetic models of choice, at least when we attempt to understand whether a particular linkage group is primordial or derived. It is logical (imperative?) that we should also initiate definitive genetic studies in elasmobranchs, which predate teleosts in the phylogenetic tree, and at least in the preliminary studies seem to be more "mainstream" than the teleosts, i.e. more likely to provide the primordial state.

3. The size of the research community that could potentially use the BAC library and the community's interest in and support for having a BAC library

As described above, elasmobranch genome analyses of any sort has been limited. Genetic information obtained from the nurse shark BAC library can be immediately applied to most other cartilaginous fish, since we expect the types of genes to be similar in all animals despite differences in genome sizes. For example, thus far all elasmobranchs have been shown to have the same types of Ig H and L chain genes (3 types of H chains and 3 types of L chains), but the numbers of clusters of each type varies according to the species. Thus, we believe that almost any gene isolated from the nurse shark BAC library can be used to identify orthologous genes from other elasmobranchs (from our very limited analysis of MHC genes in nurse shark and banded houndshark we also predict that the genetic organization will be similar in different species, but this limb is long and thin and we are at the distal end). We think that this will hold true not only for those labs examining the immune system, but also for those scientists interested in the evolutionary genetics of any physiological system in elasmobranchs.

In phylogenetic studies, there is usually a large gap when it comes to the elasmobranchs; it is a shame that more information is not available for this taxon at the base of the jawed vertebrate tree. Finally, there are over 350 members of the American Elasmobranch Society, including biologists in many different fields ranging from population genetics to hard-core gene regulation (Flajnik is an AES member). We plan to make the BAC library (and the other genomic and cDNA libraries) available to this broad community.

4. Whether the nurse shark has been proposed to NHGRI or another publicly funded agency for BAC-based genomic sequencing and the status of that request 5. Other genomic resources that will complement this resource; 6. The strain of nurse shark proposed and the rationale for its selection; 7. The size of the genome; 8. The availability of a source of DNA for construction of the BAC library; 9. Specifications for the library (e.g. library depth, BAC insert size) and supporting scientific rationale for the specs.

Among elasmobranch species, the adaptive immune system has been studied in the horn shark, clearnose skate (*Raja eglanteria*), sandbar shark (*Carcharhinus plumbeus*), banded houndshark (*Triakis scyllium*), and nurse shark. While horn shark (16) and skate (17) PAC libraries were made previously, a large-insert nurse shark library has never been proposed or attempted. The horn shark and nurse shark diverged about 120 million years ago (long before the emergence of all extant classes of placental mammals), and the horn shark has approximately double the amount of DNA/haploid genome as the

nurse shark (horn shark, 7-8 pg; nurse shark, 3.8 pg; human, 3.5 pg, ref 18). The number of IgM H chain gene clusters (~200 for horn shark and ~50 for nurse shark) reflects the genome sizes. Skates and sharks diverged from a common ancestor 220-250 million years ago (about the time birds split from reptiles). We believe that a large-insert library from nurse shark, with its genome of rather low complexity and long divergence time from horn shark and skate, complements very well resources that are currently available.

We would like to use DNA from one nurse shark (called "Yellow"), from which we have prepared our previous genomic and cDNA libraries. We have obtained and partially mapped several MHC and MHC-related genes by screening these libraries, but we realize that only with a large-insert library (~100 Kb) can we make headway into isolation of overlapping clones and elucidation of the entire MHC. Our present data will complement the cosmid/phage studies, and these other libraries will also perhaps allow us to fill in gaps missing from the BACs. High quality (high molecular weight) DNA has been obtained from erythrocytes (such cells are nucleated in all cold-blooded vertebrates); Dr. Ohta has had much experience in preparing high quality DNA for such purposes (17).

10. Time frame in which the library is needed; 11. Other support that is available or has been requested for the construction of the desired library; 12. The need for an additional BAC library if one or more already exists; and any other relevant information.

For the study of the MHC, we have already analyzed over 300Kb of cosmid sequence so far (sadly) with no overlapping clones; thus, the sooner we can obtain the library the better! We have an ongoing collaboration with Drs. Shiina and Inoko at Toaki University in Japan in which we have isolated the clones and do partial mapping, and they perform shotgun sequencing. This collaboration has worked quite well (we share in the analysis), and we hope it will continue with the BAC clones. Since our initial work suggests that shark genes are 2-3 times the size of the human genes, we expect that the entire shark MHC to be between 8-10,000Kb. While we have a major interest in comparative MHC genomics, we also think it is quite possible that we will uncover other immune- or antigen presentation-related genes that have been lost from the human MHC.

For the Ig genes, BACs are important, not only for the isolation of template genes to study expression and somatic hypermutation, but also to uncover the regulatory regions, which from studies in other vertebrates do not appear to be evolutionarily conserved (19), i.e. i.e. it is important to have a large amount of sequence surrounding the coding segments when functional studies are initiated.

Lastly, we would certainly be will to spearhead elasmobranch genomic EST projects, or at least help to organize a website that lists all available resources.

References

1. Carroll RL (1988) Vertebrate Paleontology and Evolution. Freeman, New York.
2. Flajnik MF and Rumfelt LL (2000) The immune system of cartilaginous fish. Curr. Topics Microbiol. Immunol. 248:249-270.

3. Flajnik MF (2002) Comparative analysis of immunoglobulin genes: Surprises and portents. Nature Rev Immunol 2:688-698.
4. Hinds KR and Litman GW (1986) Major reorganization of vertebrate VH segmental elements during vertebrate evolution. Nature 320:546-549.
5. Litman GW (2002) Seminars Immunol, in press.
6. Kokubu F, Litman R, Shamblott MJ, Hinds K, and Litman GW (1988) Diverse organization of immunoglobulin VH loci in a primitive vertebrate. EMBO J 7:3413-3422.
7. Lee SS, Fitch D, Flajnik MF, and Hsu E (2000) Rearrangement of immunoglobulin genes in germ cells. J. Exp. Med. 191:1637-1648.
8. Rumfelt LL, Avila D, Diaz M, Bartl S, McKinney EC, and Flajnik MF (2001) A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG. Proc. Natl. Acad. Sci. 98:1775-1780.
9. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, and Flajnik MF (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374:168-173.
10. Greenberg AS, Hughes AL, Guo J, Avila D, McKinney EC, and Flajnik MF (1996) A novel "chimeric" antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin. Eur J Immunol 26:1123-1129.
11. Flajnik MF and Kasahara M (2001) Comparative genomics of the MHC: Glimpses into the evolution of the adaptive immune system. Immunity 15:351-362.
12. Ohta Y, McKinney EC, Criscitiello MF, and Flajnik MF (2002) Proteasome, TAP, and class I genes in the nurse shark *Ginglymostoma cirratum*: Evidence for a stable class I region and MHC haplotype lineages. J Immunol 168:771-781.
13. Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, and Flajnik MF (2000) Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc Natl Acad Sci 97:4712-4717.
14. Amores A, Force A et al. (1998) Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711-1714.
15. Chiu CH, Amemiya C, Dewar K, Kim CB, Ruddle FH, and Wagner GP (2002) Molecular evolution of the HoxA cluster in the three major gnathostomes lineages. Proc Natl Acad Sci 99:5492-5497.
16. Kim CB, Amemiya C et al. (2000) Hox clusters in the horn shark *Heterodontus francisci*. Proc Natl Acad Sci 97:1655-1660.
17. Strong SJ, Ohta Y, Litman GW, Amemiya CT (1997) Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. Nucleic Acids Res 25:3959-3961.
18. Schwartz FJ and Maddock MM (2002) Cytogenetics of the Elasmobranchs: genome evolution and phylogenetic implications. Marine Freshwater Res 53:491-502.
19. Magor BG, Ross DA, Pilström L, and Warr GW (1999) Transcriptional enhancers and the evolution of the IgH locus. Immunol Today 20:13-17.