Evolutionary Genomics of Immunoglobulin-Encoding Loci in Vertebrates

Sabyasachi Das[1], Masayuki Hirano[1], Rea Tako[2], Chelsea McCallister[2] and Nikolas Nikolaidis*,[2]

[1]*Department of Pathology and Laboratory Medicine, Emory Vaccine Center, School of Medicine, Emory University, USA*

[2]*Department of Biological Science, California State University Fullerton, USA*

Abstract: Immunoglobulins (or antibodies) are an essential element of the jawed vertebrate adaptive immune response system. These molecules have evolved over the past 500 million years and generated highly specialized proteins that recognize an extraordinarily large number of diverse substances, collectively known as antigens. During vertebrate evolution the diversification of the immunoglobulin-encoding loci resulted in differences in the genomic organization, gene content, and ratio of functional genes and pseudogenes. The tinkering process in the immunoglobulin-encoding loci often gave rise to lineage-specific characteristics that were formed by selection to increase species adaptation and fitness. Immunoglobulin loci and their encoded antibodies have been shaped repeatedly by contrasting evolutionary forces, either to conserve the prototypic structure and mechanism of action or to generate alternative and diversified structures and modes of function. Moreover, evolution favored the development of multiple mechanisms of primary and secondary antibody diversification, which are used by different species to effectively generate an almost infinite collection of diverse antibody types. This review summarizes our current knowledge on the genomics and evolution of the immunoglobulin-encoding loci and their protein products in jawed vertebrates.

INTRODUCTION

All animal species from sponges to mammals have the ability to defend against infection and protect themselves from re-infection using a variety of mechanisms, which collectively are referred to as immunity. However, only vertebrates, a small fraction of the animal phylogenetic tree, have developed a specific type of immunity, the adaptive immune response. The adaptive immune system (AIS) has evolved progressively over a long period of time and in all vertebrates contains specific organs, cells, molecular cascades, and molecules [1-3].

Random evolutionary events shaped by selective forces resulted in the generation of the vertebrate AIS, which is able to discriminate between self and non-self, protect against pathogens, and allow for memory of subsequent pathogenic encounters [3-5]. The tinkering evolutionary process resulted in two types of AIS in the two major lineages of extant vertebrate species, the jawed and the jawless vertebrates [6-9]. These two radically different systems use entirely different types of molecules that are randomly assembled to generate receptor molecules on the surface of specialized cells called lymphocytes. Jawless fishes, represented by lampreys and hagfish, use the Variable Lymphocyte Receptors, in which the basic rearranging element is a leucine-rich repeat cassette, and the jawed vertebrates use immunoglobulin (Ig) receptors, in which the

basic component is the immunoglobulin domain [6-9]. The use of different types of molecules represents either a case of common ancestry in which the vertebrate ancestor possessed both types of rearranging molecules or is the result of convergent evolution [7].

Immunoglobulins (Igs or antibodies) are key components of the AIS in all jawed vertebrates [3, 5], which include elasmobranches, teleosts, amphibians, reptiles, birds, and mammals. The canonical antibodies are composed of four polypeptide chains: two identical heavy (H) chains and two identical light (L) chains held together by disulfide bonds as well as extremely stable non-covalent interactions. Both the heavy and light chains are composed of a variable (V) and a constant (C) domain. The V domain is responsible for antigen-binding, whereas the C domain is responsible for binding to effector molecules, which by triggering complex signaling pathways, eliminate the antibody-coated foreign material [3, 10].

In the germline configuration, the Ig genes are composed of multiple sets of genetic elements that by recombining with each other produce the mature variable region gene that encodes the V domain. The H-chain V domain is encoded by three genetic elements: the variable-segment (V_H), diversity-segment (D_H), and joining-segment (J_H) genes, whereas the light-chain variable domain is encoded by the variable-segment (V_L) and joining-segment (J_L) genes. The V domain can further be subdivided into two regions based on the amount of sequence divergence and structural delimitations. These regions are: the framework regions (FR) and the hypervariable or complementarity-determining regions (CDR). All functional genes encoding the V domains of both

*Address correspondence to this author at the 800 North State College Blvd, Department of Biological Science, California State University Fullerton, Fullerton, CA 92834, USA; Tel: 657-278-4526, 657-278-5045; Fax: 657-278-3426; E-mail: nnikolaidis@fullerton.edu

H and L chains are flanked by conserved recombination signal sequences (RSS). The consensus RSS consists of a highly conserved heptamer and less conserved AT-rich nonamer sequences separated by either 12±1 or 23±1 base-pair spacer [11-13].

The antibodies and their encoding-loci have been greatly differentiated during vertebrate evolution resulting in a diverse repertoire of Ig-encoding genes and Ig isotypes [2, 3, 10, 12-18]. The comparative and evolutionary genomics studies of the Ig-encoding loci and other immune-related genes have special significance on our understanding of the evolution and complexity of the immune system itself. Understanding how the system and its components evolved will allow us to understand how and why the immune system fails in certain individuals and what can be done to rescue these phenotypes. Such knowledge, which can be achieved only by identifying and using the information provided by the plethora of natural experiments over the course of millions of years of evolution, has the potential to contribute specific tools for vaccination, diagnosis, and treatment of several human conditions. Recently the availability of multiple draft genomic sequences from several vertebrate species has facilitated and permitted the exploration of the genomic organization and molecular evolution of several gene families that have essential roles in immunity. This review will provide a comprehensive overview of recent genomic studies of the vertebrates' Ig-encoding loci with a focus on their mode of evolution and their evolutionary dynamics.

IMMUNOGLOBULIN HEAVY CHAINS

Evolution of Immunoglobulin Heavy Chain Isotypes

Cartilaginous fishes express Igs comprised of conventional heavy-light chain isotypes that are called IgM and IgD/W. Apart from these two isotypes, a third isotype called IgNAR (Ig-new antigen receptor) was found in sharks. The latter is an unusual disulphide bonded homodimeric isotype that consists of two heavy chains [19, 20]. The IgNAR isotype in contrast to the conventional antibodies, lacks Ig light chains. Such antibodies that lack light chains have also been found in camelids, but their occurrence is considered to be a lineage specific event not related to the emergence of IgNAR in sharks [21]. It has been shown that homodimeric antibody in camelids evolved from a specific lineage of V$_H$ genes and have gone through a number of adaptive changes [21]. To increase the antigen-binding repertoire, the variable gene encoding homodimeric antibody has experienced higher proportions of sequence variability compared to that of variable genes of the conventional

tetrameric antibodies. However, the differences in the types of antigens being recognized by tetrameric and dimeric antibodies are not clear.

The IgM and IgD isotypes are regarded as ancient isotypes that were likely present in the vertebrate ancestor predating the emergence of cartilaginous fishes. Unlike IgM, however, IgD has undergone important structural changes over the course of evolution and shows a discontinuous distribution, i.e., lost in certain birds and mammals [20, 22-24]. In addition to IgM and IgD, a third heavy-chain isotype, IgZ/IgT, has been found in bony fishes [25, 26]. The IgZ/IgT isotype, which has certain unique features, is thought to be restricted to bony fishes. The presence of both IgM and IgD isotypes in the earliest jawed vertebrates, suggests that the divergence of these two isotypes predates the separation of the bony and cartilaginous fish lineages.

One important time interval of the evolution of vertebrates is the emergence of tetrapods, which include amphibians, reptiles, birds, and mammals. Tetrapods exhibit the most diversified antibodies compared to non-tetrapod species. Placental mammals (eutheria) including humans, express five types of IgHC chains, which are denoted by the Greek letters: α (IgA), δ (IgD), ε (IgE), γ (IgG), and μ (IgM) [3] (Table **1**). In addition to these five main types, all mammals studied thus far also express variable numbers of subspecialized isotypes [27-31]. A distinct IgHC isotype has been described in platypus, termed IgO (o) [32]. The *o* gene in platypus encodes a unique IgHC isotype that consists of four constant region domains and a hinge, and is structurally different from any of the five known mammalian Ig classes. Phylogenetically, this gene is related to ε and γ, and appears to be a structural intermediate between these two genes [32]. Since the IgO isotype has thus far not been observed in any other tetrapod species, it is plausible that this isotype is restricted in prototherian mammals.

Birds possess three different H chain isotypes, IgM, IgA, and IgY (Table **1**). Among the three different classes, the IgM and IgA are homologous to the corresponding mammalian chains. The third class of antibody, the IgY, although has some similarities with both IgG and IgE of mammals does not have a homolog in the latter species. However, the IgY class has been described in both reptiles and amphibians [33-37].

Reptiles have four H-chain isotypes, μ, υ, δ, α, and four corresponding classes IgM, IgY, IgD, and IgA (Table **1**). More recently, genomic organization studies showed that the IgH locus of the Asian lizard (*Eublepharis macularius*) contains genes that code for IgM, IgD, IgY, and IgA [38-40]. In contrast to the Asian lizard, the IgH locus of the American

Table 1. Distribution of the Immunoglobulin Heavy and Light Chain Isotypes in Vertebrates

	Cartilaginous fishes	Bony fishes	Amphibians	Reptiles	Birds	Mammals
Ig heavy chain	μ, δ, IgNAR	μ, δ, ζ/τ	μ, δ, ν, χ, φ, π$^\$$	μ, δ, α, ν	μ, α, ν	μ, δ, α, γ, ε, o[*]
Ig light chain	σ-cart, σ, κ, λ	σ, κ, λ	σ, κ, λ	κ, λ	λ	κ, λ

$^\$$IgP (φ) isotype is found in newts (*Pleurodeles waltl*) and is preferentially expressed in the larval stage.
[*]IgO (o) isotype is found only in platypus (*Ornithorhynchus anatinus*).

lizard (*Anolis carolinensis*) codes only for IgM, IgD, and IgY [38, 41]. Taking into account the fact that both mammals and birds encode the IgA isotype, the absence of this gene in certain reptiles indicates that the IgA gene was independently lost in some reptilian lineages.

Amphibians, which represent the transitional group in the dawn of tetrapod evolution, and in particular, the frog, express five different IgH isotypes, IgM, IgD, IgX, IgY, and IgF [42] (Table **1**). Apart from these five isotypes an additional isotype, designated as IgP (π), was identified in the Iberian ribbed newt (*Pleurodeles waltl*) (Class: Amphibia; Order: Urodela) [43]. The IgF (φ) isotype contains only two constant domains and a hinge region, which covalently links the two heavy chains [42]. IgY is found in a variety of amphibians, reptiles, and birds, usually consists of four constant region domains, and is regarded as a progenitor of both mammalian IgG and IgE [44, 45]. Common biological properties and sequence homology between the IgY and IgE heavy chains suggest that IgY is the immediate ancestor of mammalian IgE, whereas the transition from IgY to IgG involved structural changes that led to the formation of a hinge region in the IgG molecules. The amphibian IgX (χ) isotype is considered an analogue of the mammalian IgA because of similarities in expression patterns and tissue distribution [44, 45]. However, somewhat different from mammalian IgA, IgX is structurally similar to avian IgA and the reptilian IgA-like molecule [40]. The mammalian IgAs therefore, may be the descendants of the amphibian IgX.

Genomic Organization of the IgH-Encoding Loci

In jawed vertebrates, the Ig heavy chain (IgH) encoding locus contains variable (V_H), diversity (D_H), joining (J_H), and constant (C_H) region genes. The IgH loci of the cartilaginous fishes, the earliest extant jawed vertebrates, show two exceptional characteristics as compared to all other vertebrates. Specifically, the IgH locus of cartilaginous fishes is organized as a cluster (Fig. **1**), with the most common type being a cluster composed of closely linked (V_H-D_H1-D_H2-J_H-C_H)$_n$ genes [46-50], which are dispersed throughout the genome of cartilaginous fishes [46-50]. The presence of germline-joined IgH genes is another characteristic unique to the elasmobranches. Approximately 50% of the germline IgH loci are partially ($V_H D_H$-J_H) or fully ($V_H D_H J_H$) joined and lack intervening sequences between the

V-domain encoding genes [46-48]. The identification of germline joined Ig genes raises questions concerning their origin, evolution, and function. Accumulating evidence suggests that these sequences are the products of recombination between split genes in germ cells and most probably do not represent the primordial form of rearranging Ig genes [51, 52]. Additionally, the presence of germline joined genes may be advantageous, because these genes are no longer tied to somatic Recombination activating gene (RAG) expression and subsequent RAG-mediated rearrangement, and may not be related to the restriction of antibody diversity.

In contrast to the cartilaginous fishes, the IgH loci of bony fishes and tetrapods are organized in a locus containing multiple variable region (V_H), joining region (J_H) /diversity region (D_H), genes followed by constant region (C_H) genes (Fig. **1**). This type of organization, known as the translocon type, is conserved to all higher vertebrates. However, in some primitive bony fishes like the channel catfish (*Ictalurus punctatus*), the IgL gene organization is that of the cluster type [49, 53, 54]. The conversion from one type of organization to another, cluster versus translocon, may have represented an evolutionary advantage towards enhanced antibody diversity.

The length of the IgH loci organized as translocons varies significantly from species to species, spanning from hundreds to thousands of kilobases. The number of genes also varies extensively between different vertebrate lineages. For example, in the frog, *Xenopus tropicalis*, the IgH locus contains nearly eighty V_H genes followed by the five D_H and seven J_H [17, 42], whereas the IgH locus in anole lizards (*A. carolinensis*) contains more than 70 V_H, 22 D_H, and nine J_H genes [38]. The IgH locus of birds has an unusual organization as compared to other tetrapods [17, 33, 34, 37]. Specifically, the chicken IgH locus consists of a single functional V_H gene and 58 V_H-pseudogenes [17, 55]. Approximately fifteen D_H genes and a single J_H are found in the chicken IgH locus [56]. In the chicken IgH-chain encoding loci (and the IgL-chain encoding loci) after rearrangement, the V(D)J region is modified extensively by gene conversion using the upstream pseudogenes as conversion donors. In humans, depending on the haplotypes, the IgH locus contains nearly 100 V_H, 20 D_H, and nine J_H genes [17, 57, 58]. More than 50% of V_H genes in humans are pseudogenes. Whether these V_H pseudogenes in humans
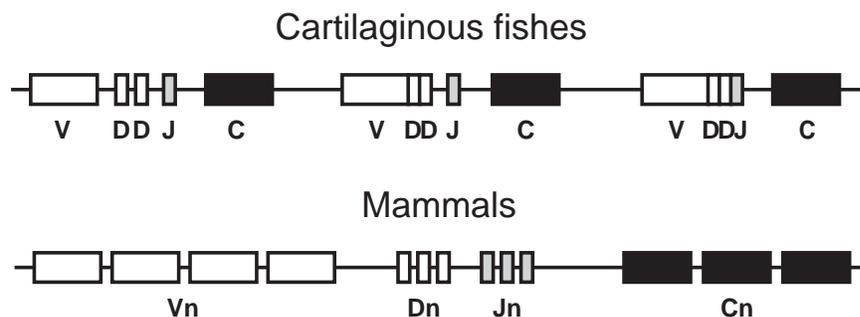

Fig. (1). Diagrammatic representation of cluster and translocon type of immunoglobulin genomic organization with reference to the IgH loci. Cartilaginous fishes have cluster type of organization including germline-joined genes, whereas higher vertebrates have translocon type of Ig organization. C = constant gene; J = joining gene; D = diversity gene; V = variable gene; and n=variable number of genes.

have any role in the antibody diversity through gene conversion events is not yet precisely known. Similar variation in the gene content is also observed in the translocon-type of the IgL loci. These variations in the number of component genes suggest that in spite of the similar overall organization of the Ig-encoding loci in tetrapods, the number of component genes varies extensively and is consistent with a model of birth-and-death evolution [17, 58-61].

Evolution of IgH Class Switching and Somatic Hypermutation

To combat a large number of antigens, a wide range of antibodies is produced by several diversifying mechanisms. During Ig gene recombination, a specialized DNA polymerase enzyme known as terminal deoxynucleotidyl transferase (TDT) introduces junctional diversity by adding N-nucleotides at the junction of V(D)J recombination; whereas B-lymphocytes upon encountering antigens in the peripheral lymphoid compartments can change the class of expressed antibody from IgM to IgG, IgA, or IgE through a recombination/deletion process termed immunoglobulin heavy chain (IgH) class switch recombination (CSR). CSR is a deletional recombination event that occurs *via* the introduction of DNA double-stranded breaks into two participating switch (S) regions, rejoining of the broken S regions to each other accompanied by deletion of all of the intervening sequences including various CH genes. A B-cell-specific protein known as activation-induced cytidine deaminase (AID) is required for the CSR activity [62]. AID is also involved in the initiation of somatic hypermutation, which introduces point mutations, and sometimes small insertions and deletions, into the variable region gene segments at a very high rate during B-cell clonal expansion.

From an evolutionary point of view, somatic hypermutation is observed in the cartilagenious fishes, whereas amphibians are the most primitive vertebrates known to use DNA recombination to switch antibody classes. Both mechanisms for antibody diversity have been maintained in every vertebrate group that has subsequently evolved [63]. Despite the expression of AID, the CSR process has not been demonstrated in bony fish and elasmobranches. However, AID from zebrafish can restore normal CSR in AID-deficient mouse B-cells, indicating that the AID functional domains required for CSR existed before the emergence of land vertebrates [64]. It is possible that components required for CSR other than AID are absent in cartilaginous and bony fishes. For example, the appropriate DNA switch regions are missing in the fish heavy chain loci. Therefore, it seems possible that the limiting step for class switching was not the evolution of a switch-capable AID protein, but rather the evolution of appropriate DNA switch regions in the translocon type configuration.

IMMUNOGLOBULIN LIGHT CHAINS

Evolution of Immunoglobulin Light Chain Isotypes

In humans there are two IgL isotypes, kappa (κ) and lambda (λ), which were initially recognized serologically [65]. The κ and λ denomination has been extended from humans to all other vertebrate species, most commonly by comparisons of nucleotide or amino acid sequences. These two isotypes are conserved in all jawed vertebrates (Table **1**) [49, 50]. Like humans, most mammals studied contain both κ and λ IgL chains, while analyses of IgL genes in different avian species indicated that only the λ IgL isotype is present in birds [15, 23, 59]. Use of molecular cladistic markers showed that amphibians in addition to κ and λ chains also contain genes that code for a third type, named sigma (σ) [49, 59]. The IgL isotype discrimination in lower vertebrates is problematic because the cladistic molecular markers which define κ, λ, and σ isotypes in tetrapods are not well preserved in bony and cartilaginous fishes. Nevertheless, application of these markers, as well as phylogenetic and sequence analysis suggest the presence of three isotypes in bony fishes, which are homologous to the amphibian κ, λ, and σ [50, 54]. Moreover, four IgL isotypes (σ, σ-cart, κ, and λ) have been reported in cartilaginous fishes based on phylogenetic reconstruction and the length of the CDRs [66]. The distribution of the IgL isotypes in jawed vertebrates (Table **1**) indicates that IgL isotype divergence occurred early in jawed vertebrate evolution.

Genomic Organization of Ig Light Chain Loci

The Ig light (IgL) locus possesses variable (V_L), joining (J_L), and constant (C_L) region genes. Like the IgH loci, the IgL chain encoding genes in cartilaginous fishes are present as multiple V_L-J_L-C_L clusters in the genome, with some loci being capable of rearranging while others contain fused $V_L J_L$ genes [51, 52]. Similar organization has been also identified in most bony fishes, while chondrostean fishes, like the sturgeon, have their IgL genes organized as translocons like all tetrapod species [50].

In most species, including tetrapods, the genes encoding the different IgL isotypes are located in different genomic regions. In the κ-encoding locus of tetrapods, multiple $J_κ$ genes are present in a cluster, followed by a single $C_κ$ gene. A notable exception is the κ-encoding locus of rabbits, which contains two $C_κ$ genes as a result of a lineage-specific duplication [67]. In contrast, in the λ-encoding locus $J_λ$ and $C_λ$ genes occur as $J_λ$–$C_λ$ blocks and usually are present in multiple copies. In the λ locus, the $J_λ$ gene is located within a 3-kb 5′ region upstream of the $C_λ$ gene, whereas in the κ locus the $J_κ$ cluster is located around 6-kb upstream of the $C_κ$ gene. Birds also have unique characteristics in the organization of their IgL locus. In contrast to the organization observed in amphibians, reptiles, and mammals, birds contain only one $J_λ$–$C_λ$ block [15, 59]. Interestingly, the organization of the IgL-encoding loci of reptiles resembles more the organization observed in amphibians and mammals than the organization found in birds. Finally, the genomic organization of the amphibian σ-encoding locus resembles that of the κ-encoding locus, suggesting an evolutionary link between the two regions (Fig. **2**).

The numbers of the IgL chain variable genes in tetrapods vary significantly even between closely related species. For example, in rodents $V_κ$ genes are more abundant than $V_λ$ genes [59], while the $V_κ$ and $V_λ$ genes in humans are present in similar numbers. The numbers of $V_κ$ and $V_λ$ genes in platypus and opossum (non-placental mammals) also differ considerably. On the other hand, birds contain only one
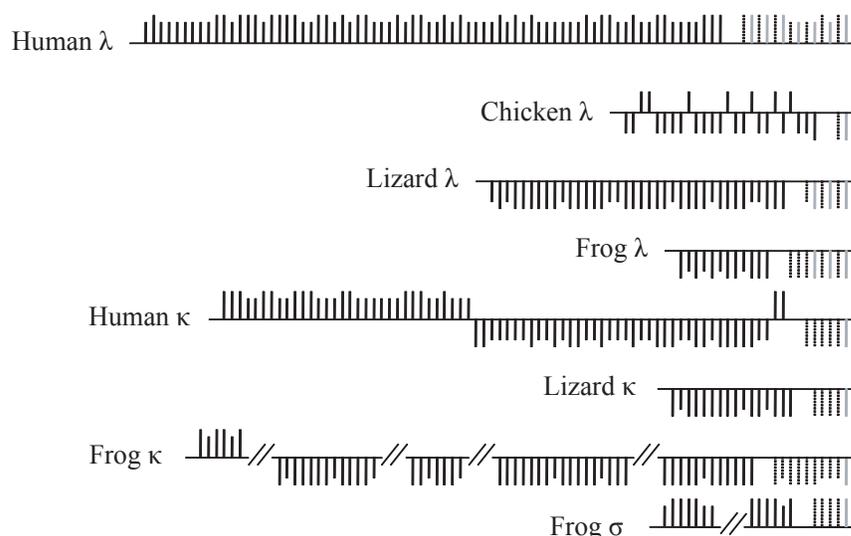
Fig. (2). Genomic organization of immunoglobulin light chain loci in representative tetrapod species. Constant, joining and variable genes are represented by gray, dotted black, and solid black vertical lines, respectively. Long vertical lines indicate functional genes whereas short vertical lines indicate pseudogenes. Vertical lines above and below the horizontal lines indicate that the Ig genes are located on opposite strands. In the frog κ- and σ-encoding loci the variable genes are located in different scaffolds. For this reason, these two loci are represented by separate horizontal lines. The figure is not drawn to scale. See details in Das *et al.* 2008.

functional V_λ gene and multiple V_λ pseudogenes [15, 59]. In the anole lizard there are 16 V_κ and 38 V_λ genes [59, 68, 69]. In addition to the number of variable genes, the copy numbers of *IGCL*, *IGJL*, and *IGJK* genes also vary from species to species.

Primate Specific Innovation: Novel Association between V_λ and microRNA Genes

MicroRNAs (miRNAs) are single stranded small non-coding RNA molecules that regulate gene expression at the post-transcriptional level [70]. Recently, it was found that specific V_λ genes contain a particular miRNA gene, known as miR-650 [71]. The *miR-650* gene (hairpin structure is 96 nt long) overlaps in the same transcription orientation with the leader exon (89 nt long) of V_λ genes (Fig. **3**). The Untranslated Region (UTR) of the V_λ leader exon contains the mature miRNA sequence, whereas its complementary sequence is located in protein-coding region of the leader exon (CDS). Nine miR-650-bearing V_λ genes were found in humans, present in both functional and pseudogenes, and all of them belong to a specific phylogenetic group, which indicates that those V_λ genes have common origin [71]. Sequence comparison and structural prediction suggested that this novel association between *miR-650* gene and the leader exon of V_λ gene is a primate-specific innovation [71]. Computational analysis of the promoter region of *miR-650*-bearing V_λ genes indicated that V_λ and the *miR-650* genes use the same promoter region for their transcription; however these two genes are apparently transcribed independently as they are expressed in different cell types [71].

CONCLUSIONS

The immunoglobulin, one of the major components of the jawed vertebrate's adaptive immune system, has evolved to recognize and respond to an exceptionally diverse range of antigens. The recent availability of several vertebrate genome sequences shows a diverse and complex picture of their evolutionary dynamics. The high degree of sequence identity between specific regions of Ig genes and proteins indicates that within this multigene family, negative selection may be operating to conserve the overall folding pattern and regions responsible for recombination and cell-signaling. However, other regions of the antibodies may have been targets of diversifying selection aiming to increase diversity. Furthermore, gene copy numbers and the ratio between functional genes and pseudogenes vary greatly among species suggesting the action of a birth-and-death mode of evolution. The Ig-loci exhibit a high degree of evolutionary turnover and might be less constrained than other genomic regions for recombination, duplication, deletion, gene conversion, point mutation, and translocation. These processes facilitate the birth-and-death of individual genes in the Ig-loci. Additionally, certain IgH isotypes have independently appeared and disappeared in different vertebrate lineages, which may affect the effector functions of the antibodies and induce specialized responses in specific cell types and different species.

Like the evolution of the IgH isotypes, the evolution of Ig light chains in vertebrates is also puzzling. Four isotypes (κ, λ, σ, and σ-cart) are found in sharks, three in bony fishes and amphibians, only two (κ, and λ) in reptiles and mammals, while birds contain only on isotype (λ). This phylogenetic distribution suggests that more recently diversified vertebrate species possess lower number of different IgL chain types. This interpretation contrasts the generally accepted and well established view of increased complexity in vertebrates like mammals. The evolutionary pattern and multiplicity of the IgL types is also enigmatic, since there is no known advantage in having multiple and different IgL isotypes. Moreover, the independent evolution of antibodies composed exclusively of heavy chains, like the IgNAR in cartilaginous fishes and the heavy chain antibodies in
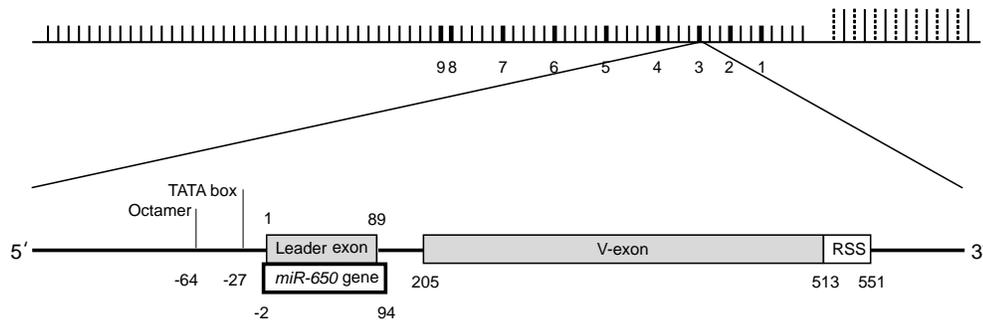
Fig. (3). Top Panel: Genomic architecture of *miRNA-650*-bearing V_λ genes in the human Ig-λ locus. The numbers the diagram indicate the positions of *miR-650*-bearing V_λ genes. Short vertical lines indicate V_λ genes, whereas long vertical lines indicate either Ig-λ constant genes (solid line) or joining genes (dotted line). Lower Panel: Schematic diagram of the overlap of *miR-650* gene and the leader exon of V_λ gene is shown in a representative *miR-650*-bearing V_λ gene. Numbers correspond to the position of *miR-650* gene and the positions of octamer, TATA box, leader exon, V-exon, and recombination signal sequence (RSS).

camelids, further challenges the existing paradigms as to the immunoglobulin structure and antibody repertoire generation. Perhaps, the main advantage of multiple IgL isotypes is redundancy; if the rearrangement of genetic elements encoding one light chain isotype becomes nonfunctional, the other ones can take over.

For the generation of Ig repertoires different species use different molecular and cellular strategies but overall these strategies have evolved towards the same goal, the production of a large number of diverse Ig molecules carrying different antigen-recognizing domains. For example, unlike teleosts, the rearranged *VDJ* gene segments in the heavy chain loci of tetrapods can join to different *C* gene segments through class-switch recombination. Avian species diversify their Ig repertoires mainly by gene conversion, while B-cells in some mammalian species like sheep and cow diversify their Ig repertoire mostly by hypermutation. In spite of such diversities in the development of the Ig repertoire, the recombination mechanisms and genetic elements that promote the formation of multiple specific Igs are fairly conserved between the different vertebrate lineages. The differential use of these mechanisms suggests that evolution using the same raw materials followed different routes in the individual vertebrate lineages to generate a large number of Igs containing diverse antigen-binding regions and effector functions. The new efforts to sequence the genomes of a large number of vertebrate species using next generation sequencing technologies will enhance our understanding on the diversification processes of the immunoglobulin-encoding loci and ultimately elucidate the evolution of the adaptive immune system.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Danilova, N.; Amemiya, C. T. Going adaptive: the saga of antibodies. *Ann. N. Y. Acad. Sci.,* **2009,** *1168*, 130-55.

[2]   Flajnik, M. F.; Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.,* **2010,** *11* (1), 47-59.

[3]   Klein, J.; Horejsi, V. *Immunology*. second ed.; Blackwell Science Ltd: Malden, 1997; p 722.

[4]   Litman, G. W.; Cannon, J. P.; Rast, J. P. New insights into alternative mechanisms of immune receptor diversification. *Adv. Immunol.,* **2005,** *87*, 209-36.

[5]   Klein, J.; Nikolaidis, N. The descent of the antibody-based immune system by gradual evolution. *Proc. Natl. Acad. Sci. U S A,* **2005,** *102* (1), 169-74.

[6]   Herrin, B. R.; Cooper, M. D. Alternative adaptive immunity in jawless vertebrates. *J. Immunol.,* **2010,** *185* (3), 1367-74.

[7]   Hirano, M.; Das, S.; Guo, P.; Cooper, M. D. The evolution of adaptive immunity in vertebrates. *Adv. Immunol.,* **2011,** (109), 125-157.

[8]   Mariuzza, R. A.; Velikovsky, C. A.; Deng, L.; Xu, G.; Pancer, Z. Structural insights into the evolution of the adaptive immune system: the variable lymphocyte receptors of jawless vertebrates. *Biol. Chem.,* **2010,** *391* (7), 753-60.

[9]   Saha, N. R.; Smith, J.; Amemiya, C. T. Evolution of adaptive immune recognition in jawless vertebrates. *Semin. Immunol.,* **2010,** *22* (1), 25-33.

[10]   Burton, D. R. Structure and function of antibodies. In *Molecular Genetics of Immunoglobulins*, Calabi, F.; Neuberger, M. S., Eds. Elsevier Science Publishers B. V.: New York, 1987; Vol. 17, pp 1-50.

[11]   Early, P.; Huang, H.; Davis, M.; Calame, K.; Hood, L. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell,* **1980,** *19* (4), 981-92.

[12]   Tonegawa, S. Somatic generation of antibody diversity. *Nature,* **1983,** *302* (5909), 575-81.

[13]   Wood, C.; Tonegawa, S. Diversity and joining segments of mouse immunoglobulin heavy chain genes are closely linked and in the same orientation: implications for the joining mechanism. *Proc. Natl. Acad. Sci. U S A,* **1983,** *80* (10), 3030-4.

[14]   Baltimore, D. Gene conversion: some implications for immunoglobulin genes. *Cell,* **1981,** *24* (3), 592-4.

[15]   Das, S.; Mohamedy, U.; Hirano, M.; Nei, M.; Nikolaidis, N. Analysis of the immunoglobulin light chain genes in zebra finch: evolutionary implications. *Mol. Biol. Evol.,* **2010,** *27* (1), 113-20.

[16]   Das, S.; Nikolaidis, N.; Nei, M. Genomic organization and evolution of immunoglobulin kappa gene enhancers and kappa

deleting element in mammals. *Mol. Immunol.,* **2009,** *46* (15), 3171-7.

[17]   Das, S.; Nozawa, M.; Klein, J.; Nei, M. Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. *Immunogenetics,* **2008,** *60* (1), 47-55.

[18]   Davis, M. M. The evolutionary and structural 'logic' of antigen receptor diversity. *Semin. Immunol.,* **2004,** *16* (4), 239-43.

[19]   Greenberg, A. S.; Avila, D.; Hughes, M.; Hughes, A.; McKinney, E. C.; Flajnik, M. F. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature,* **1995,** *374* (6518), 168-73.

[20]   Ohta, Y.; Flajnik, M. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proc. Natl. Acad. Sci. U S A,* **2006,** *103* (28), 10723-8.

[21]   Nguyen, V. K.; Su, C.; Muyldermans, S.; van der Loo, W. Heavy-chain antibodies in Camelidae; a case of evolutionary innovation. *Immunogenetics,* **2002,** *54* (1), 39-47.

[22]   Lundqvist, M. L.; Middleton, D. L.; Hazard, S.; Warr, G. W. The immunoglobulin heavy chain locus of the duck. Genomic organization and expression of D, J, and C region genes. *J. Biol. Chem.,* **2001,** *276* (50), 46729-36.

[23]   Lundqvist, M. L.; Middleton, D. L.; Radford, C.; Warr, G. W.; Magor, K. E. Immunoglobulins of the non-galliform birds: antibody expression and repertoire in the duck. *Dev. Comp. Immunol.,* **2006,** *30* (1-2), 93-100.

[24]   Ros, F.; Puels, J.; Reichenberger, N.; van Schooten, W.; Buelow, R.; Platzer, J. Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene,* **2004,** *330*, 49-59.

[25]   Danilova, N.; Bussmann, J.; Jekosch, K.; Steiner, L. A. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat. Immunol.,* **2005,** *6* (3), 295-302.

[26]   Hansen, J. D.; Landis, E. D.; Phillips, R. B. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish. *Proc. Natl. Acad. Sci. U S A,* **2005,** *102* (19), 6919-24.

[27]   Bao, Y.; Guo, Y.; Xiao, S.; Zhao, Z. Molecular characterization of the VH repertoire in Canis familiaris. *Vet. Immunol. Immunopathol.,* **2010,** *137* (1-2), 64-75.

[28]   Butler, J. E.; Sun, J.; Wertz, N.; Sinkora, M. Antibody repertoire development in swine. *Dev. Comp. Immunol.,* **2006,** *30* (1-2), 199-221.

[29]   Butler, J. E.; Wertz, N.; Zhao, Y.; Zhang, S.; Bao, Y.; Bratsch, S.; Kunz, T. H.; Whitaker, J. O., Jr.; Schountz, T. The two suborders of chiropterans have the canonical heavy-chain immunoglobulin (Ig) gene repertoire of eutherian mammals. *Dev. Comp. Immunol.,* **2011,** *35* (3), 273-84.

[30]   Wagner, B. Immunoglobulins and immunoglobulin genes of the horse. *Dev. Comp. Immunol.,* **2006,** *30* (1-2), 155-64.

[31]   Zhao, Z.; Zhao, Y.; Pan-Hammarstrom, Q.; Liu, W.; Liu, Z.; Zhao, Y.; Li, N.; Hammarstrom, L. Physical mapping of the giant panda immunoglobulin heavy chain constant region genes. *Dev. Comp. Immunol.,* **2007,** *31* (10), 1034-49.

[32]   Zhao, Y.; Cui, H.; Whittington, C. M.; Wei, Z.; Zhang, X.; Zhang, Z.; Yu, L.; Ren, L.; Hu, X.; Zhang, Y.; Hellman, L.; Belov, K.; Li, N.; Hammarstrom, L. Ornithorhynchus anatinus (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods. *J. Immunol.,* **2009,** *183* (5), 3285-93.

[33]   Magor, K. E.; Higgins, D. A.; Middleton, D. L.; Warr, G. W. One gene encodes the heavy chains for three different forms of IgY in the duck. *J. Immunol.,* **1994,** *153* (12), 5549-55.

[34]   Magor, K. E.; Higgins, D. A.; Middleton, D. L.; Warr, G. W. Opposite orientation of the alpha- and upsilon-chain constant region genes in the immunoglobulin heavy chain locus of the duck. *Immunogenetics,* **1999,** *49* (7-8), 692-5.

[35]   Magor, K. E.; Warr, G. W.; Middleton, D.; Wilson, M. R.; Higgins, D. A. Structural relationship between the two IgY of the duck, Anas platyrhynchos: molecular genetic evidence. *J. Immunol.,* **1992,** *149* (8), 2627-33.

[36]   Mansikka, A. Chicken IgA H chains. Implications concerning the evolution of H chain genes. *J. Immunol.,* **1992,** *149* (3), 855-61.

[37]   Parvari, R.; Avivi, A.; Lentner, F.; Ziv, E.; Tel-Or, S.; Burstein, Y.; Schechter, I. Chicken immunoglobulin gamma-heavy chains: limited VH gene repertoire, combinatorial diversification by D gene segments and evolution of the heavy chain locus. *EMBO J.,* **1988,** *7* (3), 739-44.

[38]   Gambon Deza, F.; Sanchez Espinel, C.; Magadan Mompo, S. The immunoglobulin heavy chain locus in the reptile Anolis carolinensis. *Mol. Immunol.,* **2009,** *46* (8-9), 1679-87.

[39]   Gambon-Deza, F.; Espinel, C. S. IgD in the reptile leopard gecko. *Mol. Immunol.,* **2008,** *45* (12), 3470-6.

[40]   Gambon-Deza, F.; Espinel, C. S.; Beneitez, J. V. A novel IgA-like immunoglobulin in the reptile Eublepharis macularius. *Dev. Comp. Immunol.,* **2007,** *31* (6), 596-605.

[41]   Wei, Z.; Wu, Q.; Ren, L.; Hu, X.; Guo, Y.; Warr, G. W.; Hammarstrom, L.; Li, N.; Zhao, Y. Expression of IgM, IgD, and IgY in a reptile, Anolis carolinensis. *J. Immunol.,* **2009,** *183* (6), 3858-64.

[42]   Zhao, Y.; Pan-Hammarstrom, Q.; Yu, S.; Wertz, N.; Zhang, X.; Li, N.; Butler, J. E.; Hammarstrom, L. Identification of IgF, a hinge-region-containing Ig class, and IgD in Xenopus tropicalis. *Proc. Natl. Acad. Sci. U S A,* **2006,** *103* (32), 12087-92.

[43]   Schaerlinger, B.; Bascove, M.; Frippiat, J. P. A new isotype of immunoglobulin heavy chain in the urodele amphibian Pleurodeles waltl predominantly expressed in larvae. *Mol. Immunol.,* **2008,** *45* (3), 776-86.

[44]   Mussmann, R.; Du Pasquier, L.; Hsu, E. Is Xenopus IgX an analog of IgA? *Eur. J. Immunol.,* **1996,** *26* (12), 2823-30.

[45]   Warr, G. W.; Magor, K. E.; Higgins, D. A. IgY: clues to the origins of modern antibodies. *Immunol. Today,* **1995,** *16* (8), 392-8.

[46]   Harding, F. A.; Amemiya, C. T.; Litman, R. T.; Cohen, N.; Litman, G. W. Two distinct immunoglobulin heavy chain isotypes in a primitive, cartilaginous fish, Raja erinacea. *Nucleic Acids Res.,* **1990,** *18* (21), 6369-76.

[47]   Harding, F. A.; Cohen, N.; Litman, G. W. Immunoglobulin heavy chain gene organization and complexity in the skate, Raja erinacea. *Nucleic Acids Res.,* **1990,** *18* (4), 1015-20.

[48]   Litman, G. W.; Amemiya, C. T.; Harding, F. A.; Haire, R. N.; Hinds, K. R.; Litman, R. T.; Ohta, Y.; Shamblott, M. J.; Varner, J. A. Evolutionary development of immunoglobulin gene diversity. *Adv. Exp. Med. Biol.,* **1991,** *292*, 11-7.

[49]   Edholm, E. S.; Wilson, M.; Bengten, E. Immunoglobulin light (IgL) chains in ectothermic vertebrates. *Dev. Comp. Immunol.,* **2011,** *35* (9), 906-15.

[50]   Hikima, J.; Jung, T. S.; Aoki, T. Immunoglobulin genes and their transcriptional control in teleosts. *Dev. Comp. Immunol.,* **2011,** *35* (9), 924-36.

[51]   Anderson, M. K.; Shamblott, M. J.; Litman, R. T.; Litman, G. W. Generation of immunoglobulin light chain gene diversity in Raja erinacea is not associated with somatic rearrangement, an exception to a central paradigm of B cell immunity. *J. Exp. Med.,* **1995,** *182* (1), 109-19.

[52]   Lee, S. S.; Fitch, D.; Flajnik, M. F.; Hsu, E. Rearrangement of immunoglobulin genes in shark germ cells. *J. Exp. Med.,* **2000,** *191* (10), 1637-48.

[53]   Bengten, E.; Clem, L. W.; Miller, N. W.; Warr, G. W.; Wilson, M. Channel catfish immunoglobulins: repertoire and expression. *Dev. Comp. Immunol.,* **2006,** *30* (1-2), 77-92.

[54]   Edholm, E. S.; Wilson, M.; Sahoo, M.; Miller, N. W.; Pilstrom, L.; Wermenstam, N. E.; Bengten, E. Identification of Igsigma and Iglambda in channel catfish, Ictalurus punctatus, and Iglambda in Atlantic cod, Gadus morhua. *Immunogenetics,* **2009,** *61* (5), 353-70.

[55]   Reynaud, C. A.; Bertocci, B.; Dahan, A.; Weill, J. C. Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. *Adv. Immunol.,* **1994,** *57*, 353-78.

[56]   McCormack, W. T.; Tjoelker, L. W.; Thompson, C. B. Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion. *Annu. Rev. Immunol.,* **1991,** *9*, 219-41.

[57]   Corbett, S. J.; Tomlinson, I. M.; Sonnhammer, E. L.; Buck, D.; Winter, G. Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. *J. Mol. Biol.,* **1997,** *270* (4), 587-97.

[58]   Ravetch, J. V.; Siebenlist, U.; Korsmeyer, S.; Waldmann, T.; Leder, P. Structure of the human immunoglobulin mu locus: characterization of embryonic and rearranged J and D genes. *Cell,* **1981,** *27* (3 Pt 2), 583-91.

[59]   Das, S.; Nikolaidis, N.; Klein, J.; Nei, M. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using

molecular markers. *Proc. Natl. Acad. Sci. U S A,* **2008,** *105* (43), 16647-52.

[60] Nei, M.; Gu, X.; Sitnikova, T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. U S A,* **1997,** *94* (15), 7799-806.

[61] Nei, M.; Rooney, A. P. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.,* **2005,** *39*, 121-52.

[62] Chaudhuri, J.; Basu, U.; Zarrin, A.; Yan, C.; Franco, S.; Perlot, T.; Vuong, B.; Wang, J.; Phan, R. T.; Datta, A.; Manis, J.; Alt, F. W. Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv. Immunol.,* **2007,** *94*, 157-214.

[63] Stavnezer, J.; Amemiya, C. T. Evolution of isotype switching. *Semin. Immunol.,* **2004,** *16* (4), 257-75.

[64] Barreto, V. M.; Pan-Hammarstrom, Q.; Zhao, Y.; Hammarstrom, L.; Misulovin, Z.; Nussenzweig, M. C. AID from bony fish catalyzes class switch recombination. *J. Exp. Med.,* **2005,** *202* (6), 733-8.

[65] Korngold, L.; Lipari, R. Multiple-myeloma proteins. III. The antigenic relationship of Bence Jones proteins to normal gammaglobulin and multiple-myeloma serum proteins. *Cancer,* **1956,** *9* (2), 262-72.

[66] Criscitiello, M. F.; Flajnik, M. F. Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. *Eur. J. Immunol.,* **2007,** *37* (10), 2683-94.

[67] Akimenko, M. A.; Mariame, B.; Rougeon, F. Evolution of the immunoglobulin kappa light chain locus in the rabbit: evidence for differential gene conversion events. *Proc. Natl. Acad. Sci. U S A,* **1986,** *83* (14), 5180-3.

[68] Das, S.; Hirano, M.; Nikolaidis, N. Immunoglobulins in Reptiles. In *Reptiles: Biology, Behavior and Conservation*, Baker, K. J., Ed. Nova Publishers: New York, 2011; pp 101-114.

[69] Wu, Q.; Wei, Z.; Yang, Z.; Wang, T.; Ren, L.; Hu, X.; Meng, Q.; Guo, Y.; Zhu, Q.; Robert, J.; Hammarstrom, L.; Li, N.; Zhao, Y. Phylogeny, genomic organization and expression of lambda and kappa immunoglobulin light chain genes in a reptile, Anolis carolinensis. *Dev. Comp. Immunol.,* **2010,** *34* (5), 579-89.

[70] Lee, R. C.; Ambros, V. An extensive class of small RNAs in Caenorhabditis elegans. *Science,* **2001,** *294* (5543), 862-4.

[71] Das, S. Evolutionary origin and genomic organization of micro-RNA genes in immunoglobulin lambda variable region gene family. *Mol. Biol. Evol.,* **2009,** *26* (5), 1179-89.