Analysis of the Immunoglobulin Light Chain Genes in Zebra Finch: Evolutionary Implications

Sabyasachi Das,*[1,2] Uzra Mohamedy,[3] Masayuki Hirano,[1] Masatoshi Nei,[2] and Nikolas Nikolaidis[3]

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

[2]Department of Biology, Institute of Molecular Evolutionary Genetics, Pennsylvania State University

[3]Department of Biological Science, California State University, Fullerton

***Corresponding author:** E-mail: sdas8@emory.edu.

Associate editor: Yoko Satta

Abstract

All jawed vertebrates produce immunoglobulins (IGs) as a defense mechanism against pathogens. Typically, IGs are composed of two identical heavy chains (IGH) and two identical light chains (IGL). Most tetrapod species encode more than one isotype of light chains. Chicken is the only representative of birds for which genomic information is currently available and is an exception to the above rule because it encodes only a single IGL isotype (i.e., lambda). Here, we show that the genome of zebra finch, another bird species, encodes a single IGL isotype, that is, lambda, like the chicken. These results strongly suggest that the second isotype (i.e., kappa) present in both reptiles and mammals was lost in a very early stage of bird evolution. Furthermore, we show that both chicken and zebra finch contain a single set of functional variable, joining, and constant region genes and multiple variable region pseudogenes. The latter finding suggests that this type of genomic organization was already present in the common ancestor of these bird species and remained unchanged over a long evolutionary time. This conservation is in contrast with the high levels of variation observed in the mammalian *IGL* loci. The presence of a single functional variable region gene followed by multiple variable pseudogenes in zebra finch suggest that this species may be generating antibody diversity by a gene conversion-like mechanism like the chicken.

Key words: immunoglobulin genes, cladistic molecular markers, antibody diversity, gene conversion, comparative genomics, repetitive elements.

Research article

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015

Introduction

A typical immunoglobulin (antibody, Ig) in jawed vertebrates is composed of two identical heavy chains (IGH) and two identical light chains (IGL) and provides defense against all extracellular and some intracellular pathogens (Klein and Hořejší 1997). Jawed vertebrate species, with the exception of chickens, ducks, and bats, express more than one immunoglobulin light (IGL) chain isotype (Lundqvist et al. 2006; Criscitiello and Flajnik 2007; Das et al. 2008). In mammals, *IGL* genes generally exist in two distinct isotypes called kappa (κ) and lambda (λ). The genes for the two light chain isotypes are encoded at separate and unlinked loci, and the organization of κ and λ chain locus differs significantly (Wahlstrom et al. 1988; Lai et al. 1989). In general, the κ chain–encoding locus is arranged with multiple *IGVK* genes (variable kappa), a small cluster of *IGJK* (joining kappa) genes, and a single *IGCK* (constant kappa) gene, whereas in the λ chain–encoding locus multiple *IGVL* (variable lambda) genes are followed by *IGJL* (joining lambda) and *IGCL* (constant lambda) genes, which occur as *IGJL*–*IGCL* blocks, usually present in multiple copies (Frippiat et al. 1995; Kirschbaum et al. 1996; Kawasaki et al. 1997, 2001; Das et al. 2008).

Unlike humans and mice, chickens and ducks have been shown to exclusively express λ light chains (Sanders and Travis 1975; Magor et al. 1994; Lundqvist et al. 2006). In the chicken *IGL* locus, there is only one functional *IGVL*, *IGJL*, and *IGCL* gene, whereas there are multiple *IGVL* pseudogenes located upstream of the functional *IGVL* gene (Parvari et al. 1987). In contrast to the humans and mice that depend on gene rearrangements to generate light chain diversity, chickens generate light chain diversity through intrachromosomal gene conversion, a nonreciprocal recombination process that uses the upstream pseudo-*IGVL* genes as donor sequences (McCormack et al. 1991).

Birds are an enormously diverse group of vertebrates comprising around 9,000 species (Shukla and Tyagi 2004). However, with the exception of chicken (galliforms) and ducks (anseriforms), the characterization of avian IGL isotypes is very limited (Reynaud et al. 1983; Magor et al. 1994; Lundqvist et al. 2006). Furthermore, analysis of the complete genomic organization of the *IGL* locus currently exists only for chicken. With the breadth of limited knowledge regarding avian *IGL* genes and their genomic organization, the recently available draft genomic sequence of zebra finch (*Taeniopygia guttata*) provides an opportunity to study the *IGL* genes in another avian model species. Zebra finch is a member of Passeriformes, which diverged from chicken (galliforms) more than 100 Ma (Brown et al. 2008).

In the present study, we analyzed the IGL sequences and their genomic organization in zebra finch to investigate whether the overall organization of the *IGL* locus in this bird species is similar to the chicken *IGL* locus. We also compared the genomic organization of the *IGL* locus

© The Author 2009. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

between avian (chicken–zebra finch) and mammalian (human–horse) species to understand the evolutionary mechanisms that generated the *IGL* repertoire in these species.

Materials and Methods

Retrieval of Ig Light Chain Genes from Zebra Finch Genome Sequence

An exhaustive gene search was conducted to identify all the light chain genes in the draft genome sequences of zebra finch (*T. guttata*) from the Ensembl genome browser. This is the first release of the zebra finch genome assembly (assembly: Taeniopygia_guttata-3.2.4, Aug 2008). The complete genome sequence of zebra finch was produced by The Genome Center at Washington University School of Medicine in St Louis. To retrieve the light chain variable and constant genes in zebra finch, we performed TBlastN searches (cutoff *E*-value of 10^{-15}) using as queries the encoded amino acid sequences of nine functional light chain variable region genes [three IGVK and three IGVL from human, one IGVL from chicken, and two IGV sigma–encoding sequences (IGVS) from frog] and five functional constant sequences (one IGCK and two IGCL from human, one IGCL from chicken, and one IGCS from frog), respectively (Das et al. 2008). The nine IGL variable sequences and the five IGL constant sequences in the query data set aligned to the same genomic regions because they are similar to one another. For this reason, we extracted only nonoverlapping genomic sequences that produced alignments with the lowest *E*-values. To identify the joining genes, which are very short and cannot be detected by Blast searches, we manually screened 7 kb upstream of the constant gene, taking into account the location of the recombination signal sequence (RSS) at the 5′ end of the joining gene.

To retrieve any expression data, the identified genomic sequences were used as queries in BlastN searches against the Expressed Sequence Tag (EST) database of NCBI and the ESTIMA database (http://titan.biotec.uiuc.edu/cgi-bin/ESTWebsite/estima_start?seqSet=songbird). Multiple sequence alignments between the retrieved cDNA sequences and the genomic *IGVL* genes were performed using MAFFT (Katoh et al. 2009).

Determination of Functional Genes and Pseudogenes

The variable region genes can be divided into two hypervariable or complementarity-determining regions (CDR1 and CDR2) and three framework regions (FRl, FR2, and FR3) (Kabat and Wu 1991). For the variable region genes, any retrieved sequence that aligned with the query sequence without any frameshift mutations and/or premature stop codons in the leader exon and the V-exon, possessed the two conserved Cys residues in FR1 and FR3 regions, respectively, and had a proper RSS was regarded as a potentially functional gene. All other sequences, including truncated ones, were regarded as pseudogenes. For the constant and joining region genes, the retrieved sequences that did not have any frameshift mutations and/or internal stop codons were regarded as potentially functional genes. In addition, for the joining region gene, we have examined the RSS to determine putative functionality. All the retrieved sequences were aligned with the query sequence (MI0003665) using the ClustalW program (Thompson et al. 1994), and the alignments were inspected manually to maximize similarity.

Repetitive Sequence Analysis

The repetitive elements were identified using the CENSOR software tool (Kohany et al. 2006).

Estimation of Sequence Divergence

The number of nucleotide differences and the proportion of differences per site (*P* distance) (Nei and Kumar 2000) were calculated using MEGA4 (Tamura et al. 2007) and SWAAP v1.0.3 (http://asiago.stanford.edu/SWAAP/SwaapPage.htm). In these analyses, all three codon positions were included. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). Standard errors were obtained by a bootstrap procedure of 500 replicates.

Homology Modeling

To map the unique insertion QQQSST (see Results) on the tertiary structure of zebra finch, IGL homology modeling and fold recognition were performed using the SWISS-MODEL (http://swissmodel.expasy.org) (Arnold et al. 2006) and PHYRE (http://www.sbg.bio.ic.ac.uk/~phyre) (Bennett-Lovsey et al. 2008) web servers. Pairwise structural alignments and structural superimposition were performed using the SSAP (http://cathdb.info/cgi-bin/SsapServer.pl) (Taylor and Orengo 1989) and DaliLite (http://www.ebi.ac.uk/Tools/dalilite) (Holm and Park 2000) web servers. Tertiary structure figures were generated using PyMol (DeLano Scientific; http://pymol.org).

Results

Analysis of the Light Chain Variable Region Genes in Zebra Finch

Using the human (kappa and lambda), chicken (lambda), and frog (sigma) IGL variable sequences as queries, we identified 21 *IGL* variable region genes located in a cluster in chromosome 15 of the zebra finch genome. The genomic location of these variable region genes in zebra finch is given in supplementary table S1 (Supplementary Material online). The sequence comparison with functional sequences of human, chicken, and frog indicated that among the 21 *IGL* variable region genes, only one sequence is functional because it contains a complete coding sequence without frameshift mutations and/or internal stop codons, two conserved Cys residues in FR1 and FR3 regions and a proper RSS (fig. 1). The genomic structure of the single functional *IGL* variable region gene in zebra finch is shown in supplementary figure S1 (Supplementary Material online). The remaining *IGVL* genes either lacked the proper leader and/or RSS or were truncated in their 5′ or 3′ ends

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015

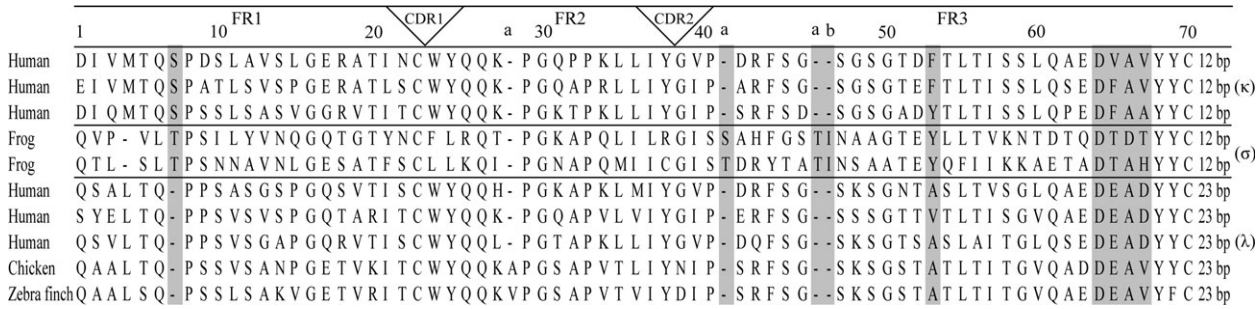

Fig. 1. Alignment of IGL variable sequences. The randomly chosen reference sequences of known kappa (three IGVK sequences from human), lambda (three IGVL sequences from human and one IGVL from chicken), and sigma (the IGVS from frog) isotypes were taken from Das et al. (2008). The cladistic molecular markers that distinguish the three isotypes (κ, λ, and σ) are highlighted. The lengths of the RSS spacer sequences are given. The numbering of the amino acid positions in the V-segment is based on human IGVK and IGVL sequences, and the gaps relative to the frog IGVS sequences are indicated by "a" and "b" (Das et al. 2008).

(supplementary fig. S2, Supplementary Material online), like the chicken *IGVL* pseudogenes (Reynaud et al. 1987). Only four sequences contained internal stop codons (supplementary fig. S2 and table S1, Supplementary Material online).

Analysis of *IGL* Joining and Constant Region Genes in Zebra Finch

From the similarity search using as queries, five functional IGL constant sequences (one IGCK and two IGCL from human, one IGCL from chicken, and one IGCS from frog), we identified only a single functional *IGL* constant–encoding gene in the zebra finch genome. This gene is located 4.5 kb downstream of the single functional variable region gene. ESTs and cDNA sequences confirm the presence of a single functional *IGL* constant–encoding gene in zebra finch because the identified ESTs and cDNA sequences align—with almost 100% identity and no gaps—to a single genomic position in the zebra finch genome that corresponds to the position of the single *IGCL* functional gene (supplementary table S2 and fig. S3, Supplementary Material online). To identify the *IGL* joining region gene in the zebra finch genome, we scanned for the conserved RSS in the 4.5 kb region between the functional variable and constant region genes because the joining region gene is too short (usually 12 amino acids in length) to be identified by Blast searches. Once the potential RSS was identified in this 4.5 kb region, we translated the nucleotide sequences at the 3′ end of the RSS into amino acids and compared the translated sequence with the human, chicken, and frog IGL joining region sequences, which were identified in our previous study (Das et al. 2008). Using this method, we identified a single functional *IGL* joining region gene in zebra finch.

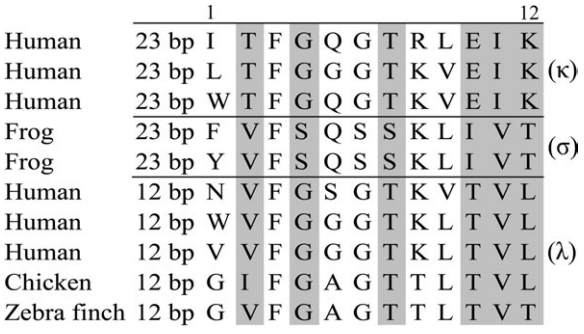

Fig. 2. Alignment of IGL joining region sequences. The randomly chosen reference sequences of known kappa (three IGJK sequences from human), lambda (three IGJL sequences from human and one IGJL from chicken), and sigma (two IGJS from frog) isotypes were taken from Das et al. (2008). The cladistic molecular markers that distinguish the three isotypes (κ, λ, and σ) are highlighted (Das et al. 2008). The lengths of the RSS spacer sequences are shown.

Identification of Zebra Finch IGL Isotype

To characterize the isotype of the zebra finch IGL sequences, we used the cladistic molecular markers, which we previously described (Das et al. 2008). Like the IGVL sequences of other tetrapods, the only functional variable region sequence of zebra finch lacks Ser or Thr at position 7 and does not possess a bulky aromatic residue (Phe or Tyr) at position 53 (fig. 1). The tetrapods IGVL sequences generally have a fairly conserved DEAD (Asp–Glu–Ala–Asp) motif in the FR3 region (Das et al. 2008). However, like chicken IGVL sequence (i.e., DEAV at position 64–67), the Asp residue is substituted for Val at position 67 (fig. 1).

Consistent with the variable region sequence, the molecular markers in the joining and constant region sequences in zebra finch also categorize them as lambda light chain sequences (figs. 2 and 3). In addition, like mammalian lambda light chain genes, the RSS sequences flanking the single functional variable region gene and the joining region gene are interrupted by a 23-bp and a 12-bp spacer, respectively, in the zebra finch *IGL* locus. Hence, the molecular markers in the variable, joining, and constant region sequences in zebra finch indicate that like chicken (Sanders and Travis 1975) and duck (Magor et al. 1994), the zebra finch genome encodes only the lambda isotype of Ig.

Strikingly, the zebra finch IGL constant region sequence can be distinguished from the tetrapod IGL constant region sequences because it contains a unique insertion of a six amino acid stretch (QQQSST) (fig. 3). This insertion is confirmed by the fact that all ESTs/cDNAs sequences, when

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015

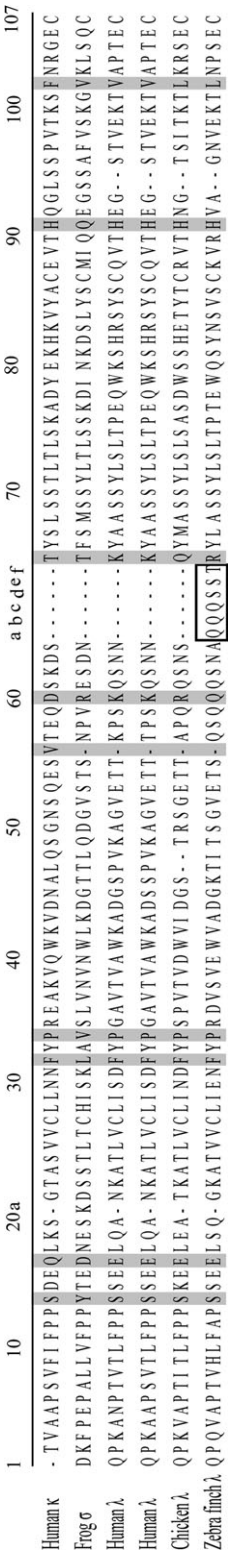

FIG. 3. Alignment of IGL constant sequences. The reference sequences of known kappa (one IGCK sequence from human), lambda (two IGCL from human and one IGCL from chicken), and sigma (one IGCS from frog) isotypes were taken from Das et al. (2008). The numbering of the amino acid positions in the C-segment is based on Das et al. (2008). The gap relative to IGCS sequence of frog and the gaps relative to zebra finch IGCL sequence are characterized by alphabets. The cladistic molecular markers that distinguish the three isotypes (κ, λ, and σ) are highlighted.

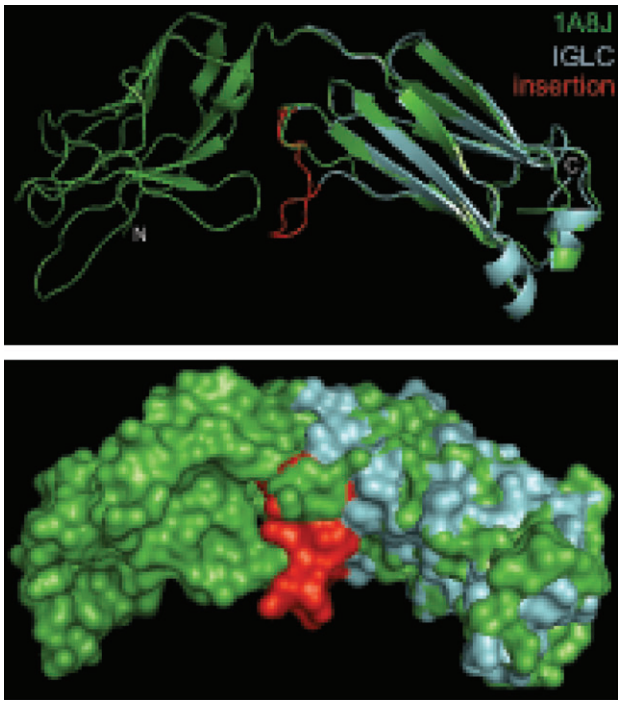

FIG. 4. The amino acid insertion of the zebra finch IGCL protein sequence is mapped on a 3D model. The theoretical 3D model for the zebra finch IGCL (shown in light cyan) was constructed using as template the experimentally resolved structure of a human IGL lambda chain (shown in green; PDB accession code: 1A8J). The predicted and the template structures were structurally aligned to highlight the insertion region (shown in red). The upper panel shows a cartoon representation and the lower shows a surface representation of the two structures.

translated, include the additional six amino acids (see supplementary table S2 and fig. S3, Supplementary Material online). To further analyze this insertion, we generated a 3D model of the zebra finch IGCL protein and mapped the insertion on the 3D model (fig. 4). Our analysis predicts that the QQQSST insertion extends a loop located in the region between the IGVL and IGCL by approximately 9 Å (fig. 4; supplementary fig. S4, Supplementary Material online). Although this extended loop is predicted to reduce the distance between the Ig light and the heavy chains by almost 6 Å (supplementary fig. S4, Supplementary Material online), it does not seem to affect the contact between the different Ig chains. This supposition deserves to be experimentally tested.

Genomic Organization of the IGL Locus in Zebra Finch

In the kappa- and sigma-encoding locus of most tetrapods, multiple joining region genes are present in a cluster, followed by a single constant gene, whereas in the lambda-encoding locus, joining and constant genes occur as *IGJL–IGCL* blocks, which usually have multiple copies. Only chicken has one *IGJL–IGCL* block (Das et al. 2008). The *IGL* locus in zebra finch contains one functional *IGVL* gene, multiple pseudo-*IGVL* genes, and only one *IGJL–IGCL* block (fig. 5), like the chicken locus (Reynaud et al. 1985, 1987).

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015

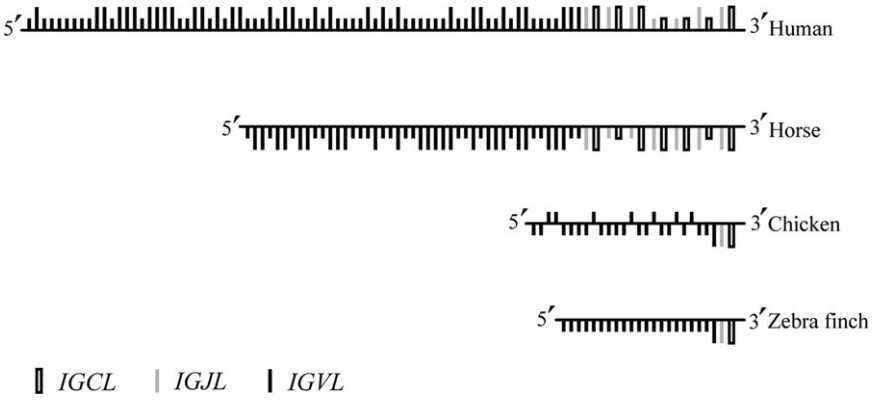

Fig. 5. Schematic diagrams of the genomic organizations of lambda light chain loci in human, horse, chicken, and zebra finch (not to scale). Rods above and below the lines indicate genes located on opposite strands based on the particular genome sequence. Long rods show functional genes, and short rods indicate pseudogenes. The positional information of human, horse, and chicken lambda chain genes are taken from Das et al. (2008).

The comparison of the IGL locus between the chicken and the zebra finch shows that in these species both the number and the position of the *IGL* genes are very similar (fig. 5), despite the fact that these species have diverged more than 100 Ma (Brown et al. 2008). The main difference between the two loci is the presence of a few pseudogenes with reverse orientation in chicken. In contrast to the general conservation of IGL locus observed between the two bird species, comparison of the *IGL* locus between different mammalian species (i.e., human–horse) that have diverged approximately 100 Ma showed many differences both in the constant and variable regions (fig. 5). In the constant region, both human and horse contain seven *IGJL–IGCL* blocks, but the distribution of functional genes and pseudogenes are different between the two species (fig. 5). In the variable region, both the number and the distribution of functional genes and pseudogenes also vary between the human (32 functional and 42 variable region pseudogenes) and the horse (25 variable region functional and 20 pseudogenes) lambda loci.

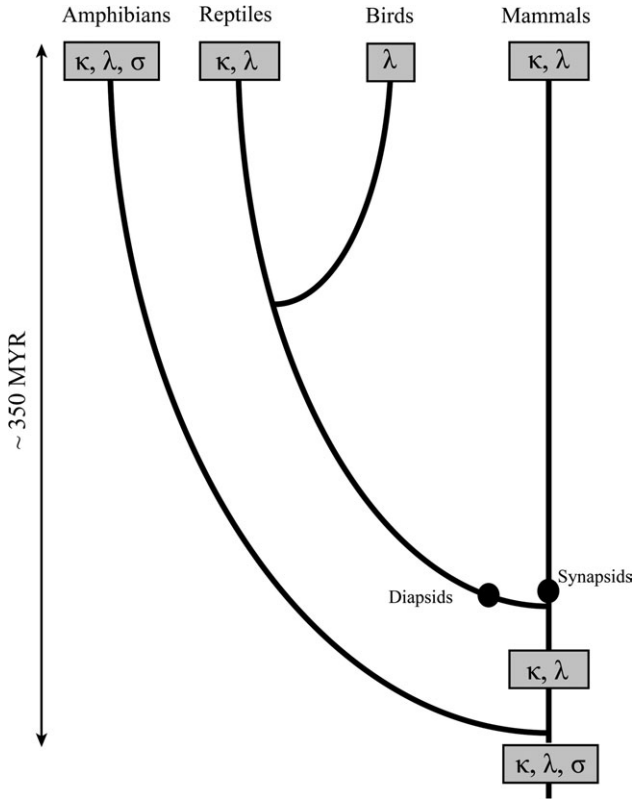

Fig. 6. A hypothetical evolutionary scheme of the evolution of the different IGL isotypes in tetrapods. The scale indicates the approximate divergence times for different classes of tetrapods.

Discussion

Our analysis of the *IGL* locus in zebra finch (Passeriformes) shows that the genome of this species encodes a single IGL isotype, that is, lambda (figs. 1–3), like the chicken (Galliformes). Taking into account that ducks (Arseniformes) contain only lambda IGL chains (Magor et al. 1994), we can safely suggest that birds encode a single IGL chain. If this idea proves to be true for all birds, then birds are exceptional to most studied tetrapod species, which encode at least two different types of IGL chains (Pilstrom 2002; Das et al. 2008). Indeed, three IGL isotypes (i.e., kappa, lambda, and sigma) are present in frogs, whereas reptiles and most mammals contain two IGL isotype (i.e., kappa and lambda) (Criscitiello and Flajnik 2007; Das et al. 2008; Qin et al. 2008). Assuming that the genomes of birds encode only lambda IGL chains, then the IGL kappa–encoding genes were lost before the divergence of Passeriformes and Galliformes (ca. 100 Ma) and after the divergence of the bird and the reptilian lineages (fig. 6).

Our data show that the genomic organization of the *IGL* locus is very similar between zebra finch and chicken, and both species contain a single set of functional *IGVL*, *IGJL*, and *IGCL* genes and multiple *IGVL* pseudogenes (fig. 5). The presence of a single *IGVL* gene and multiple pseudogenes has been reported for other bird species with the sole exception of muscovy ducks (*Cairina moschata*), which contain an additional potentially functional *IGVL* gene

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015

Table 1. Repetitive Elements in Ig Lambda Locus.

Repeat Class	Zebra Finch (25.7 kb)	Chicken (31.1 kb)	Human (884.3 Kb)	Horse (892.4 Kb)
DNA transposon	5 (0.19/kb)	2 (0.06/kb)	239 (0.27/kb)	208 (0.23/kb)
Endogenous retrovirus	0 (0/kb)	3 (0.09/kb)	307 (0.35/kb)	314 (0.35/kb)
LTR retrotransposon	3 (0.12/kb)	2 (0.06/kb)	64 (0.07/kb)	51 (0.06/kb)
Non-LTR retrotransposon	3 (0.12/kb)	10 (0.32/kb)	752 (0.85/kb)	550 (0.61/kb)
Total	11 (0.43/kb)	17 (0.54/kb)	1362 (1.54/kb)	1123 (1.26/kb)

NOTE.—The length of the lambda locus is given in parenthesis. LTR, long terminal repeat.

(McCormack et al. 1989). Given that the mallard duck (*Anas platyrhynchos*) contains only one functional *IGVL* gene (Lundqvist et al. 2006) and that ducks and chickens are more closely related than chickens and zebra finch (Hackett et al. 2008), we speculate that the presence of an additional functional *IGVL* gene in muscovy ducks is the result of a recent lineage-specific gene duplication. Therefore, the similar genomic organization between the IGL loci in zebra finch and chicken suggest that this type of organization was already present in the common ancestor of these two species and remained largely unchanged for a long evolutionary time.

The results of our study have three major implications for the evolution of the IGL lambda–encoding locus in tetrapods. First, the presence of a single functional *IGJL–IGCL* block is unique in birds because multiple copies of paralogous *IGJL–IGCL* blocks are present in the Ig lambda–encoding locus of all other tetrapods (Das et al. 2008). Second, the presence of multiple *IGVL* pseudogenes upstream of a single set of functional *IGVL*, *IGJL*, and *IGCL* genes and the nearly similar number of *IGVL* pseudogenes in both birds contradict the evolution of IGL lambda–encoding loci observed in other species. For example, in mammals, the ratio between functional genes and pseudogenes varies significantly, even between closely related species, like mice and rats or humans and macaques (Das et al. 2008; Das 2009). A prevalent hypothesis concerning large scale changes in genomic sequences, including the Ig heavy chain–encoding loci, suggests that such alterations may be partially explained by the content of different repetitive elements (Straubinger et al. 1987; Mazzarella and Schlessinger 1997; Matsuda et al. 1998). To test this hypothesis, we compared the content and distribution of repetitive elements in the lambda-encoding loci of zebra finch, chicken, human, and horse. Our analysis revealed that the content of repetitive elements is much higher in mammals than that of avian species (table 1). We speculate that the differences in the repetitive element content in the IGL locus can be one of the reasons to explain the higher gene content heterogeneity of the IGL locus in mammals as compared with birds. Third, our analysis together with previous results on the IGL lambda–encoding loci of multiple tetrapods (Das et al. 2008) suggest that the general organization of the lambda loci in amphibians and reptiles is more similar to that of mammals than that of birds. Whether this type of organization occurred by random genomic drift (Nei 2007) or whether this genomic organization was selected due to functional constraints in the common ancestor of birds remains an open question.

The conservation of the genomic organization of the *IGL* loci between the two bird species raises an additional implication that concerns the generation of antibody diversity in birds. It has been shown that the chicken generates light chain diversity through intrachromosomal gene conversion, which uses the upstream pseudo-*IGVL* genes as donor sequences (Carlson et al. 1990; McCormack and Thompson 1990; McCormack et al. 1991). The presence of a single functional *IGVL* followed by multiple *IGVL* pseudogenes in zebra finch suggests that this species may also use a similar mechanism to generate light chain diversity. The limited number (seven) of ESTs/cDNAs that we identified seems to support this notion (see supplementary table S2 and fig. S5, Supplementary Material online). This analysis suggests that there are at least three stretches of nucleotides in the IGVL-encoding mRNA that could be the result of a gene conversion-like mechanism. The estimated sequence divergence between the genomic *IGVL* sequence and the ESTs/cDNAs *IGVL* fragments (table 2; supplementary table S3 and fig. S6, Supplementary Material online) suggests that the extent of diversity in zebra finch is similar to the diversity in other bird species (Magor et al. 1994; Lundqvist et al. 2006).

Table 2. The Number of Nucleotide Differences between the IGVL Genes of Zebra Finch is Shown. Standard Error Estimate(s) are Shown above the Diagonal and Were Obtained by a Bootstrap Procedure (500 Replicates). All Positions Containing Alignment Gaps and Missing Data Were Eliminated Only in Pairwise Sequence Comparisons.

	IGVL	CK304555	FE719512	DQ214165	DQ214164	CK307295	CK307761	CK307451
IGVL		2.16	1.859	4.683	4.507	4.522	5.416	3.619
CK304555	5		1.002	4.743	4.664	4.082	5.266	3.979
FE719512	4	1		4.66	4.583	4.229	5.389	3.829
DQ214165	25	26	25		4.177	4.674	6.176	4.848
DQ214164	22	25	24	20		4.541	5.957	4.837
CK307295	22	19	20	25	23		5.968	4.906
CK307761	33	31	32	42	44	41		5.214
CK307451	16	18	17	24	28	27	28	

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015

In conclusion, our analysis of the *IGL* locus in zebra finch and the comparison of the evolution of the *IGL* loci between birds and mammals indicate that the mammals and the birds have used different evolutionary processes to achieve the same physiological result, which is the generation of antibody diversity and ultimately the defense against pathogens.

Supplementary Material

Supplementary tables S1–S3 and figures S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Max Cooper, Jan Klein, Parimal Majumder, Jianxu Li, Masafumi Nozawa, and Sayaka Miura for their valuable comments and suggestions. U.M. was supported by the Associate Student Incorporated Grant from California State University, Fullerton (CSUF). This work was supported by the National Institutes of Health (grant GM020293-35 to M.N.), by the CSUF (start-up money to N.N.), and by a CSUF Junior Faculty Research Grant (to N.N.).

References

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 22:195–201.

Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA. 2008. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*. 70: 611–625.

Brown JW, Rest JS, Garcia-Moreno J, Sorenson MD, Mindell DP. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol*. 6:6.

Carlson LM, McCormack WT, Postema CE, Humphries EH, Thompson CB. 1990. Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion. *Genes Dev*. 4:536–547.

Criscitiello MF, Flajnik MF. 2007. Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. *Eur J Immunol*. 37: 2683–2694.

Das S. 2009. Evolutionary origin and genomic organization of microRNA genes in immunoglobulin lambda variable region gene family. *Mol Biol Evol*. 26:1179–1189.

Das S, Nikolaidis N, Klein J, Nei M. 2008. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proc Natl Acad Sci USA*. 105:16647–16652.

Frippiat JP, Williams SC, Tomlinson IM, Cook GP, Cherif D, Le Paslier D, Collins JE, Dunham I, Winter G, Lefranc MP. 1995. Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. *Hum Mol Genet*. 4:983–991.

Hackett SJ, Kimball RT, Reddy S, et al. (18 co-authors). 2008. A phylogenomic study of birds reveals their evolutionary history. *Science*. 320:1763–1768.

Holm L, Park J. 2000. DaliLite workbench for protein structure comparison. *Bioinformatics*. 16:566–567.

Kabat EA, Wu TT. 1991. Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *J Immunol*. 147:1709–1719.

Katoh K, Asimenos G, Toh H. 2009. Multiple Alignment of DNA Sequences with MAFFT. *Methods Mol Biol*. 537:39–64.

Kawasaki K, Minoshima S, Nakato E, et al. (11 co-authors). 2001. Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes. *Eur J Immunol*. 31:1017–1028.

Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Schmeits JL, Wang J, Shimizu N. 1997. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res*. 7:250–261.

Kirschbaum T, Jaenichen R, Zachau HG. 1996. The mouse immunoglobulin kappa locus contains about 140 variable gene segments. *Eur J Immunol*. 26:1613–1620.

Klein J, Hořejší V. 1997. Immunology. Oxford: Blackwell Science Ltd.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*. 7:474.

Lai E, Wilson RK, Hood LE. 1989. Physical maps of the mouse and human immunoglobulin-like loci. *Adv Immunol*. 46:1–59.

Lundqvist ML, Middleton DL, Radford C, Warr GW, Magor KE. 2006. Immunoglobulins of the non-galliform birds: antibody expression and repertoire in the duck. *Dev Comp Immunol*. 30: 93–100.

Magor KE, Higgins DA, Middleton DL, Warr GW. 1994. cDNA sequence and organization of the immunoglobulin light chain gene of the duck, Anas platyrhynchos. *Dev Comp Immunol*. 18: 523–531.

Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, Honjo T. 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med*. 188:2151–2162.

Mazzarella R, Schlessinger D. 1997. Duplication and distribution of repetitive elements and non-unique regions in the human genome. *Gene*. 205:29–38.

McCormack WT, Carlson LM, Tjoelker LW, Thompson CB. 1989. Evolutionary comparison of the avian IgL locus: combinatorial diversity plays a role in the generation of the antibody repertoire in some avian species. *Int Immunol*. 1:332–341.

McCormack WT, Thompson CB. 1990. Chicken IgL variable region gene conversions display pseudogene donor preference and 5′ to 3′ polarity. *Genes Dev*. 4:548–558.

McCormack WT, Tjoelker LW, Thompson CB. 1991. Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion. *Annu Rev Immunol*. 9:219–241.

Nei M. 2007. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci USA*. 104:12235–12242.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford: Oxford University Press.

Parvari R, Ziv E, Lentner F, Tel-Or S, Burstein Y, Schechter I. 1987. Analyses of chicken immunoglobulin light chain cDNA clones indicate a few germline V lambda genes and allotypes of the C lambda locus. *EMBO J*. 6:97–102.

Pilstrom L. 2002. The mysterious immunoglobulin light chain. *Dev Comp Immunol*. 26:207–215.

Qin T, Ren L, Hu X, et al. (11 co-authors). 2008. Genomic organization of the immunoglobulin light chain gene loci in Xenopus tropicalis: evolutionary implications. *Dev Comp Immunol*. 32:156–165.

Reynaud CA, Anquez V, Dahan A, Weill JC. 1985. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell*. 40:283–291.

Reynaud CA, Anquez V, Grimal H, Weill JC. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell*. 48:379–388.

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015

Reynaud CA, Dahan A, Weill JC. 1983. Complete sequence of a chicken lambda light chain immunoglobulin derived from the nucleotide sequence of its mRNA. *Proc Natl Acad Sci USA*. 80: 4099–4103.

Sanders BG, Travis JC. 1975. Evidence for one immunoglobulin light-chain type in chickens: absence of a blocked N-terminal light-chain type. *Biochem Genet*. 13:779–782.

Shukla AN, Tyagi R. 2004. Encyclopaedia of birds. New Delhi: Anmol Publications Pvt. Ltd.

Straubinger B, Osterholzer E, Zachau HG. 1987. Three transposed elements in the intron of a human VK immunoglobulin gene. *Nucleic Acids Res*. 15:9567–9675.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596–1599.

Taylor WR, Orengo CA. 1989. Protein structure alignment. *J Mol Biol*. 208:1–22.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.

Wahlstrom G, Pear WS, Steen ML, Szpirer J, Levan G, Klein G, Sumegi J. 1988. Localization of the rat immunoglobulin lambda light chain locus to chromosome 11. *Immunogenetics*. 28:182–183.

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 28, 2015