

ARTICLE

Received 10 Apr 2012 | Accepted 22 Nov 2012 | Published 8 Jan 2013

DOI: 10.1038/ncomms2317

Extensive diversification of IgH subclass-encoding genes and IgM subclass switching in crocodylians

Gang Cheng¹, Yang Gao¹, Tao Wang¹, Yi Sun¹, Zhiguo Wei², Lingxiao Li¹, Liming Ren¹, Ying Guo¹, Xiaoxiang Hu¹, Yan Lu³, Xumin Wang⁴, Guiming Liu⁴, Chenglin Zhang³, Jun Yu⁴, Qiang Pan-Hammarström⁵, Lennart Hammarström⁵, Xiaobing Wu⁶, Ning Li¹ & Yaofeng Zhao^{1,7}

Crocodylians are a group of reptiles that are closely related to birds and are thought to possess a strong immune system. Here we report that the IgH locus in the Siamese crocodile and the Chinese alligator contains multiple μ genes, in contrast to other tetrapods. Both the $\mu 2$ and $\mu 3$ genes are expressed through class-switch recombination involving the switch region and germline transcription. Both IgM1 and IgM2 are present in the serum as polymers, which implies that IgM class switching may have significant roles in humoral immunity. The crocodylian α genes are the first IgA-encoding genes identified in reptiles, and these genes show an inverted transcriptional orientation similar to that of birds. The identification of both α and δ genes in crocodylians suggests that the IgH loci of modern living mammals, reptiles and birds share a common ancestral organization.

¹State Key Laboratory of Agrobiotechnology, College of Biological Sciences, National Engineering Laboratory for Animal Breeding, China Agricultural University, Beijing 100193, China. ²College of Animal Science and Technology, Henan University of Science and Technology, Henan 471003, China. ³Beijing Zoo, Beijing 100044, China. ⁴Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China. ⁵Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska University Hospital Huddinge, Stockholm SE-141 86, Sweden. ⁶College of Life Sciences, Anhui Normal University, Anhui Provincial Key Laboratory of the Conservation and Exploitation of Biological Resources, Wuhu 241000, China. ⁷Key Laboratory of Animal Reproduction and Germplasm Enhancement in Universities of Shandong, College of Animal Science and Technology, Qingdao Agricultural University, Qingdao 266109, China. Correspondence and requests for materials should be addressed to Y.Z. (email: yaofengzhao@cau.edu.cn).

Immunoglobulins (Igs) are essential components of the adaptive immune system and are expressed only in jawed vertebrates, including cartilaginous fish, bony fish, amphibians, reptiles, birds and mammals^{1–3}. These defence molecules emerged ~500 million years after the divergence of jawed vertebrates from jawless vertebrates (cyclostomes such as lampreys and hagfish)². No Igs are found in cyclostomes, which were recently shown to have developed a distinct recombinatorial adaptive system^{4–7}.

Antigen-binding regions (variable regions) and effector regions (constant regions) are pivotal to Ig functions. From an evolutionary perspective, comparative studies have proven to be useful in understanding how the Ig-gene variable regions have diversified and how the effector regions have multiplied (that is, the diversification of the Ig classes and subclasses) in different species¹. A number of molecular mechanisms that are involved in shaping the IgV (V: variable) repertoire, such as V(D)J (D: diversity segments; J: joining segments) recombination, somatic hypermutation (SHM) and gene conversion, have been intensively studied. Both V(D)J recombination and SHM are utilized by all of the species that have been examined, whereas gene conversion has a major role only in birds^{1,8,9}. Additionally, more than 10 genes encoding different Ig classes, such as IgM, IgD (IgW), IgNAR, IgZ (IgT), IgA (IgX), IgY, IgF, IgO, IgG and IgE, have been identified in various species^{3,10–20}, and IgG and IgA are further diversified into a variable number of subclasses in mammalian species. Differential diversification of the Ig classes or subclasses may have arisen from a long evolutionary period of environmental selection pressure, thus conferring survival advantages. IgM and IgD are thought to be the most primitive Ig classes^{17,21}, as IgM has been identified in all species examined to date and IgD is found in most species except for birds and certain mammals²². Logically, other IgH (H: heavy-chain genes) classes should be evolutionarily derived from IgM and IgD, likely through various mechanisms such as gene duplication, gene conversion or recombination. These processes led to the origin of IgNAR in cartilaginous fish and IgZ(T) in bony fish^{13,14,23}. Although both IgY and IgA(X) appeared only after the emergence of tetrapods, the former likely preceded the latter in evolution based on the evolutionary evidence that IgA (X) resulted from a genetic recombination between IgM and IgY^{24,25}. From lower tetrapods to mammals, IgY has been functionally diversified into IgG and IgE²⁶.

In addition to the insightful clues regarding Ig evolution, comparative studies have provided unexpected observations in recent decades, particularly in reptiles and birds. For example, IgD is absent in all birds examined to date²⁷, although it has been identified in all other groups of jawed vertebrates, including reptiles^{22,25,28}. Furthermore, although IgA (or IgX) is expressed in mammals, birds and amphibians¹, this Ig class, which is involved in mucosal immunity, is missing in several reptiles^{25,28}. This absence is surprising considering that even bony fish express a special Ig class, IgT, that is dedicated to mucosal immunity¹⁵. Notably, the α gene in birds shows a reverse transcriptional orientation to the μ and ν genes²⁷. Together, these observations suggest that the IgH locus in birds and reptiles has experienced genetic rearrangements, resulting in either the deletion or inversion of certain genes.

Crocodylians (including Alligatorinae, such as caimans and alligators, Crocodylinae, such as crocodiles and false gavials, and Gavialinae, such as gavials) are thought to be the closest relatives of living birds²⁹. As these species provide a phylogenetic link to other reptiles and birds, analysis of their Ig genes may provide significant clues to understanding Ig evolution. Crocodylians are thought to have a strong immune system, as they are rarely subject to infection despite harsh living conditions^{30,31}. In this

study, we have therefore performed a thorough analysis of the IgH genes in two species of crocodylians: the Siamese crocodile (*Crocodylus siamensis*) and the Chinese alligator (*Alligator sinensis*). We found that the IgH locus in these species contains multiple μ genes, and that IgM subclasses can be expressed through class-switch recombination.

Results

Construction of IgH-specific mini-cDNA libraries. To analyse the IgH isotypes expressed in the Siamese crocodile, we first amplified a fragment of the μ gene using degenerate primers for the conserved sequences of known μ genes in other species. We subsequently performed rapid amplification of cDNA ends (RACE) to clone the 5'-portion of IgM heavy-chain transcripts, which revealed a number of heavy-chain joining (JH) segments. It is known that in a given species of tetrapods, distinct IgH classes can share the same set of JH segments when expressed¹. Using the JH-derived primers, we performed 3'-RACE with RNA isolated from the spleen or small intestine. The amplified 3'-RACE PCR products were cloned to construct both spleen and intestine IgH-specific mini-cDNA libraries. Theoretically, these IgH cDNA libraries should contain all of the IgH isotype transcripts provided that each transcript is expressed at a sufficient level.

Multiple IgH genes are expressed in crocodylians. Using PCR and sequencing, we identified 526 clones in the spleen IgH library and 428 clones in the small intestine library. Despite their different frequencies (Supplementary Fig. S1), we detected heavy-chain transcripts of three distinct μ -encoding genes (termed $\mu 1$, $\mu 2$ and $\mu 3$, sharing 59.2 to 65% sequence identity at the protein level), three ν -encoding genes (termed $\nu 1$, $\nu 2$ and $\nu 3$, sharing 49.3–51.3% sequence identity at the protein level) and two α -encoding genes (termed $\alpha 1$ and $\alpha 2$, sharing 62.4% sequence identity at the protein level) in the two libraries. A single δ heavy-chain transcript containing two CH (heavy-chain constant region) domains was also identified in the spleen library. An additional 3'-RACE reaction with primers based on the identified δ CH revealed two further transcripts: one containing four CH domains and another containing seven CH domains. The identities of these IgH genes were all confirmed by phylogenetic analysis (Fig. 1).

The expression of multiple Ig subclass-encoding genes, including the μ genes, in the Siamese crocodile was unexpected because few of the tetrapod species examined thus far have been reported to express more than one μ gene. The only known exception is cattle, in which two μ genes have been detected (located in BTA21 and BTA11)³². To determine whether multiple Ig subclasses are expressed by other crocodylians, we analysed the IgH genes expressed in the Chinese alligator using the same approach and also detected the expression of three μ , three ν and two α heavy chains (Supplementary Fig. S1). Interestingly, each of these heavy-chain transcripts corresponded with a high degree of sequence identity (> 85%) to a specific class or subclass identified in the Siamese crocodile (Supplementary Fig. S1), whereas the different subclasses in a single species share < 70% sequence identity. These data suggest that the divergence of the Ig subclass-encoding genes likely occurred before the divergence of these two species.

Physical mapping of the IgH gene locus in the Siamese crocodile. We subsequently determined how these μ genes are organized in the genome, that is, whether they are organized in a single IgH locus with a 'translocon' structure (Vn-Dn-Jn-Cn) as in other tetrapods or in (V-D-J-Cn) clusters located on either the same or different chromosomes as found in cartilaginous fish.

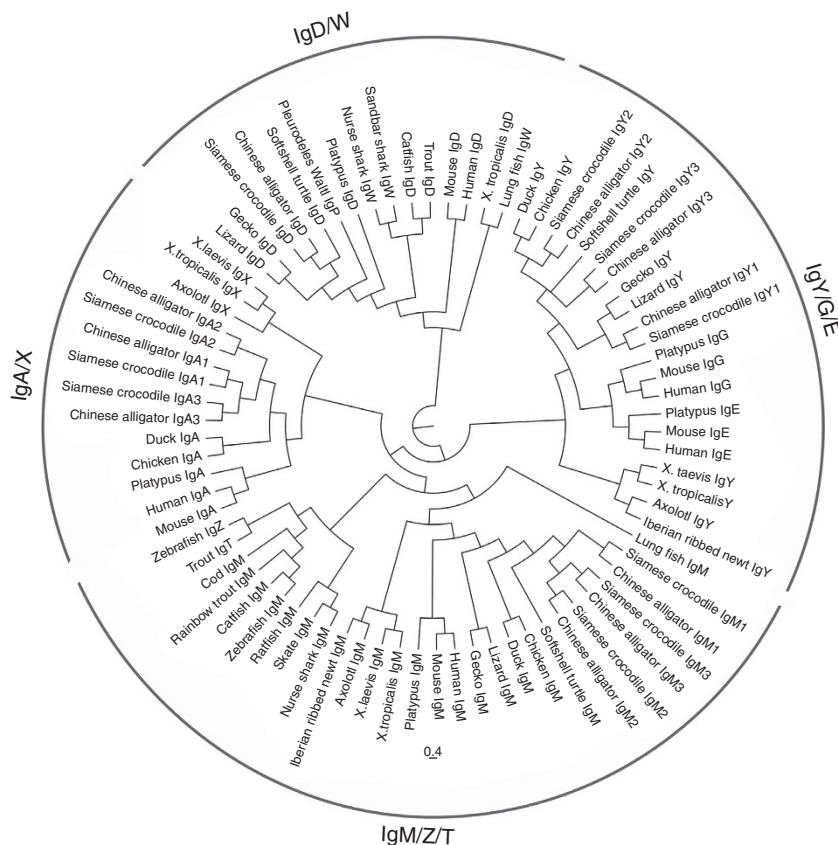


Figure 1 | Phylogenetic analysis of IGHC genes in jawed vertebrates. The scale bar shows the genetic distance.

To address this issue, we employed genome walking to map the IgH locus of the Siamese crocodile. The corresponding genomic sequences of each IgH cDNA (except IgD) were first amplified by long-range PCR and then sequenced. These sequences served as the starting points for genome walking in both directions until overlapping sequences could be found for an upstream or downstream gene. Using this approach, we identified a genomic region of ~260 kb in which all of the identified IgH class- and subclass-encoding genes were contained. Surprisingly, an additional pseudo μ gene and two additional α genes (one structurally functional (termed $\alpha 3$) and one pseudo α) were identified in this region (Fig. 2a). The pseudo μ gene was found to have a mutation at the CH1 3'-splice site (or the 5'-splice site of the following intron), changing **GTAAG** to **ATAAG**. Although the expression of the $\alpha 3$ gene could be confirmed at the cDNA level, the pseudo α was found to be mutated. Notably, all α genes, including the pseudo α gene, showed an opposite transcriptional orientation to all other genes within the locus.

The δ gene was shown to have seven CH exons corresponding to the seven cloned CH-containing cDNA sequences. Additionally, four mutated CH exons were identified between δ CH4 and δ CH5 (Fig. 2a). It was also shown that the two cloned short IgD transcripts contained the first two and first four CH domains, respectively.

The $\mu 1$ gene was located at the most 5'-end of the IGHC (Ig heavy-chain constant region gene) locus, and a JH gene locus containing nine JH gene segments was found 10 kb upstream of the $\mu 1$ gene (Fig. 2a). The IgH gene locus in the Siamese crocodile was arranged as JH- $\mu 1$ - δ - $\alpha 1$ - $\mu 2$ - $\alpha 3$ - $\mu 3$ - $\psi \alpha$ - $\psi \mu$ - $\alpha 2$ - $v 3$ - $v 2$ - $v 1$. The expression of these IgH genes in different tissues was examined by quantitative RT-PCR as shown in Fig. 3. Although all of the three μ genes were expressed at the highest levels in the spleen,

both the $\alpha 1$ and $\alpha 2$ genes were expressed at the highest levels in the intestine. All three v genes were expressed at relatively high levels in the liver and spleen.

Physical mapping of the IgH gene locus in the Chinese alligator. To analyse whether the IgH genes are similarly organized in the Chinese alligator, we generated a BAC (bacterial artificial chromosome) genomic library using peripheral blood leucocytes isolated from a Chinese alligator. The library consisted of 2.1×10^5 clones with an average insert size of ~100 kb (Supplementary Fig. S2), representing $\sim 9 \times$ genomic coverage (~2.5 Gb). Five IgH gene-positive BAC clones (Y368I17, Y236C22, Y234H2, Y29J4 and Y88L23) were identified using a PCR-based approach and sequenced. Upon covering a gap between Y236C22 and Y234H2 by genome walking, we obtained a ~432 kb genomic sequence covering the alligator IgH genes. Furthermore, in addition to the genes identified at the cDNA level, a pseudo μ ($\psi \mu$), a δ , a third functional α and a pseudo α ($\psi \alpha$) gene were arranged in the same order as in the Siamese crocodile (Fig. 2b). Within nearly 200 kb upstream of the first μ , we identified two functional VH (heavy-chain variable gene), six pseudo VH, 68 DH (heavy-chain diversity segment) and 10 JH segments (Supplementary Figs S3 and S4). The identification of these regions allowed us to deduce the physical map of the alligator IgH gene locus as VH_n -DH₍₁₋₆₈₎-JH₍₁₋₁₀₎- $\mu 1$ - δ - $\alpha 1$ - $\mu 2$ - $\alpha 3$ - $\mu 3$ - $\psi \alpha$ - $\psi \mu$ - $\alpha 2$ - $v 3$ - $v 2$ - $v 1$ (Fig. 2b).

Both the $\mu 2$ and $\mu 3$ genes are expressed through CSR. In tetrapods, the expression of IgH classes other than IgM and IgD is mediated by class-switch recombination (CSR), which involves somatic DNA rearrangements that delete the μ and δ genes as

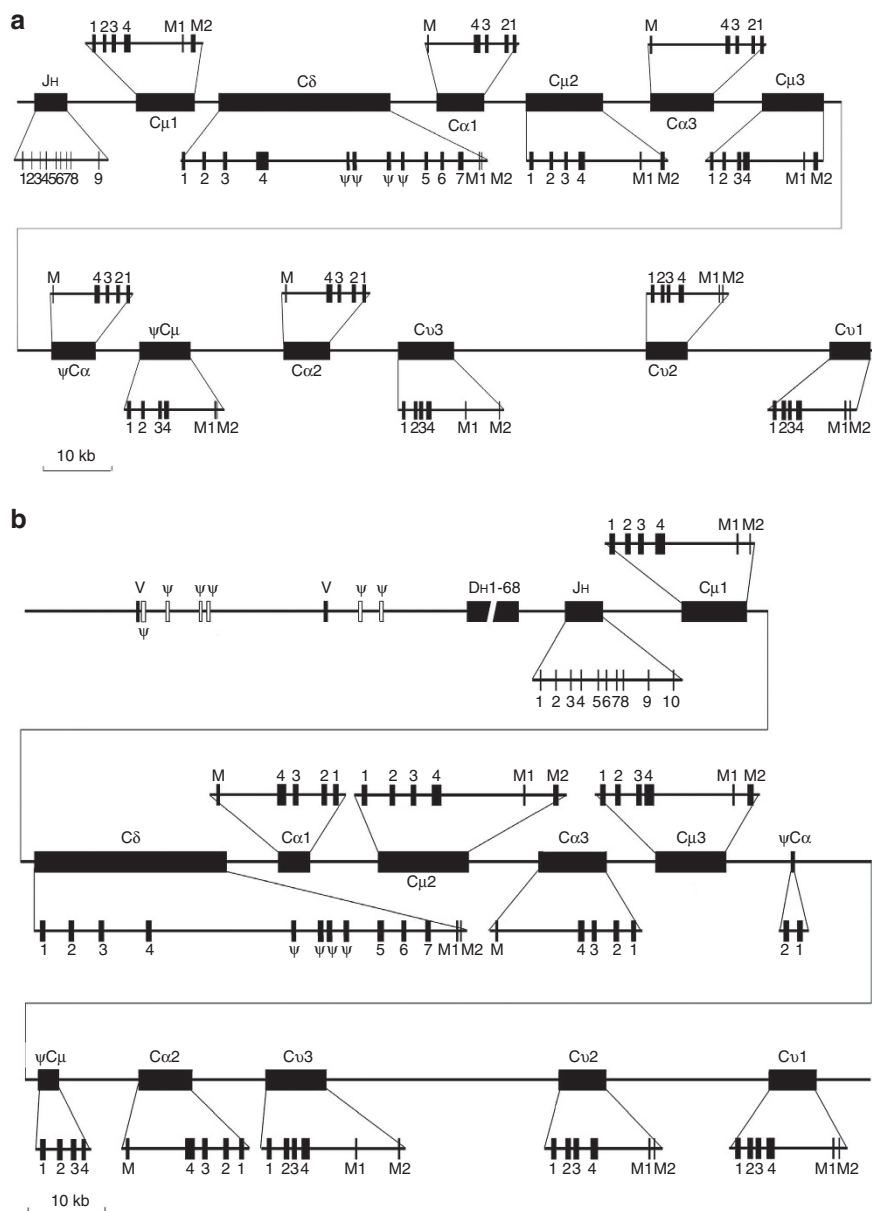


Figure 2 | Physical map of the IgH gene locus in two crocodilian species. **(a)** Physical map of the IgH gene locus in the Siamese crocodile. **(b)** Physical map of the IgH gene locus in the Chinese alligator. A detailed structure for each gene is also shown under or above the backbone map with the exons indicated by Arabic numerals. All of the α genes in the locus were found to be inverted. ‘ Ψ ’ is used to denote pseudo genes or pseudo exons. M: membrane region-encoding exon.

well as other genes upstream of the expressed genes. As IgM is the first Ig class expressed during B-cell development, CSR switches an IgM-expressing B cell to express a non-IgM/IgD class. As two additional functional μ genes were found downstream of the most 5' μ within the crocodilian IgH locus, we questioned whether these genes are expressed through CSR. CSR is mediated by recombination between switch (S) regions, which are located upstream of the μ gene and other IgH constant genes (except for δ). The switch regions typically contain short, repetitive sequences that are rich in AGCT motifs. Using these criteria, we identified S regions for all of the three μ genes and ν genes in the IgH locus of the Siamese crocodile (Supplementary Fig. S5).

To determine whether CSR is involved in $\mu 2$ and $\mu 3$ expression, sense primers derived from the S $\mu 1$ 5'-flanking region and anti-sense primers from the S $\mu 2$, S $\mu 3$ and S $\nu 1$ 3'-flanking regions were designed and used in two-round nested PCR amplifications of the recombined S $\mu 1$ -S $\mu 2$, S $\mu 1$ -S $\mu 3$ and S $\mu 1$ -S $\nu 1$ fragments

(Fig. 4a). Using spleen genomic DNA as templates, these primers generated the desired bands only if somatic recombination had occurred between S $\mu 1$ and the other S regions (Fig. 4b). After cloning and sequencing the amplified PCR products, we obtained 54 unique S $\mu 1$ -S $\mu 2$ recombined fragments, 53 S $\mu 1$ -S $\mu 3$ fragments and 45 S $\mu 1$ -S $\nu 1$ fragments (Fig. 4c, Supplementary Fig. S6), suggesting that CSR occurred between $\mu 1$ and the two downstream μ genes.

It is well known that CSR is achieved through the non-homologous end-joining pathway in humans and mice³³. An analysis of the junctions in the recombined fragments described above revealed a similar pattern to that observed in humans and mice³⁴. Approximately one quarter (28%) of the junctions indicated direct joining of S $\mu 1$ and the other S regions. However, more than half (57%) of the junction sites appeared to be derived from microhomology-mediated ligation, where 1–6 nucleotides were found to be shared by the two S regions involved

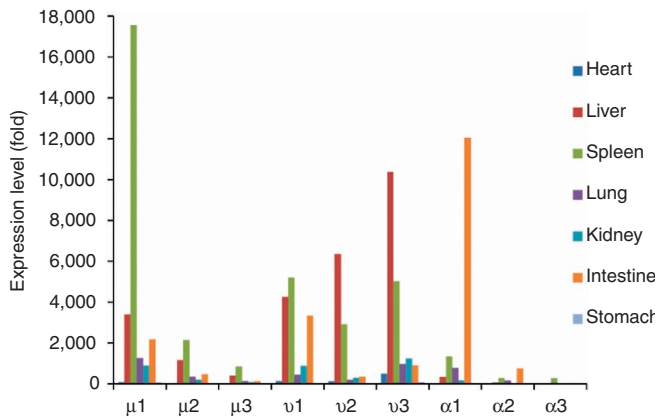


Figure 3 | Quantitative RT-PCR analysis of Siamese crocodile IgH gene expression in different tissues. The *EEF1A1* gene was used as an internal control, and the values shown in the figure were calculated using the $\Delta\Delta C_t$ method.

(Table 1). A small portion (16.6%) of the junctions was also shown to have nucleotide insertions between the two recombined S regions.

Similar to SHM, which occurs in the variable regions of Ig genes, CSR is also an AID (activation-induced cytidine deaminase)-dependent process. In humans and mice, it has been recognized that AID can introduce mutations into S regions involved in CSR³⁵. We therefore analysed the mutation pattern in the cloned switch fragments. The analysis was limited to ± 25 -bp sequences flanking the junction site of each recombined fragment. The analysed region accounted for 7,600 nucleotides, and a total of 120 nucleotide mutations were identified by comparison to the germline S region sequences, indicating a mutation frequency of 15.7 per 1,000 bp (Table 1). Although nucleotide transitions occurred at the same rate as transversions, significantly more mutations were found at G/C sites than at A/T sites (90% versus 10%) (Table 1). Approximately 73% of the mutations were located at the WRC or GYW motifs, suggesting that AID was involved in the mutational process (Table 1).

Identification of germline transcripts and intronic exons of μ genes. In humans and mice, germline transcription through the S regions, which allows the S regions to be accessible to AID, is required to initiate CSR. Along with germline transcription, a short sequence (termed the I exon) located in the 5' segment of each S region is spliced to the CH exons to generate germline transcripts. As $\mu 2$ and $\mu 3$ are expressed through CSR, both the I exon and the corresponding germline transcript are expected to be located upstream of these genes. Upon performing 5'-RACE to amplify the 5'- $\mu 3$ heavy-chain portion in the Siamese crocodile, we obtained three cDNAs with no rearranged VDJ sequence but with a sequence located ~ 0.7 kb from the 5'- $S\mu 3$ region, which was spliced to the $\mu 3$ CH1 exon (Fig. 5a). This finding indicated the presence of an $I\mu 3$ exon and germline $\mu 3$ transcription. As neither $I\mu 1$ nor $I\mu 2$ could be identified using the same approach, we designed a series of sense primers for the 5'- $S\mu 1$ and - $S\mu 2$ flanking regions. These primers were used for RT-PCR amplification together with anti-sense primers derived from $\mu 1$ and $\mu 2$. This approach allowed us to locate $I\mu 1$ at ~ 1.3 kb 5' of $S\mu 1$ and $I\mu 2$ at ~ 0.4 kb 5' of $S\mu 2$ (Fig. 5a). Germline $\mu 1$ and $\mu 2$ gene transcripts were mainly detected in the spleen and intestine, whereas germline $\mu 3$ transcripts were mainly detected in the spleen (Fig. 5b).

Promoter and enhancer activity for germline transcription of μ genes. Germline transcription of each Ig heavy-chain constant region gene is driven by an intronic promoter located at the 5'-region of the mammalian I exon. To determine whether there are I promoters for the Siamese crocodile μ genes, we cloned the sequences (two fragments, M1L1 and M1L2 for $I\mu 1$, M2L for $I\mu 2$ and M3L for $I\mu 3$) upstream of each $I\mu$ exon into pGL3 luciferase reporter vectors, which were then transfected into the chicken B-cell line DT40. As shown in Fig. 5c, we identified I promoters in the 5'-regions of all three $I\mu$ exons, with M1L1, M2L and M3L fragments demonstrating promoter activities when cloned into the pGL3-enhancer vector. However, only a weak enhancer activity was detected for the $I\mu 1$ promoter, and enhancer activity was not observed for the $I\mu 2$ and $I\mu 3$ promoters; only M1L1 showed enhancer activity when cloned into the pGL3-promoter vector (Fig. 5d), whereas M2L and M3L decreased the promoter activity to a certain extent (Fig. 5d).

Both IgM1 and IgM2 can form polymers. To examine the secreted forms of IgM in serum, we developed mouse monoclonal antibodies specific for Chinese alligator IgM1 and IgM2. Western blotting showed that under reducing conditions, both IgM1 and IgM2 heavy chains were detected as an ~ 80 -kDa band, which is 15–20 kDa greater than the predicted size and roughly equal to the size of the mouse IgM heavy chain (Fig. 6a,b). The sizes of both IgM1 and IgM2 were reduced to approximately the predicted size when the serum sample was treated with PNGase F to remove N-linked saccharides, suggesting substantial N-linked glycosylation of both subclasses (Fig. 6b). The size was only slightly reduced upon O-glycosidase treatment (Fig. 6b). In contrast to mouse IgM (which is present as pentamers in serum), IgM1 was present as both pentamers and hexamers, whereas IgM2 was detected as tetramers under non-reducing conditions (Fig. 6c).

Discussion

Compared with birds and mammals, few studies have been conducted on reptilian Ig genes^{25,28,36–41}. Crocodylians are a group of reptiles that phylogenetically link reptiles and birds, and they are thought to have a strong immune system^{29–31}. It is thus of great interest to investigate the immunogenetic components of these crocodylians.

The family Crocodylidae comprises three subfamilies (Alligatorinae, Crocodylinae and Gavialinae). The Siamese crocodile and the Chinese alligator belong to Crocodylinae and Alligatorinae, respectively. In both of these crocodylians, the IgH gene was organized similarly, as Vn-Dn-Jn- $\mu 1$ - δ - $\alpha 1$ - $\mu 2$ - $\alpha 3$ - $\mu 3$ - $\psi\alpha$ - $\psi\mu$ - $\alpha 2$ - $\nu 3$ - $\nu 2$ - $\nu 1$ in a 'translocon' configuration (Fig. 2), and each IGHC gene in one species clearly had a homologous counterpart in the other species. This finding strongly suggests that Alligatorinae and Crocodylinae share a common IgH gene locus that developed before the divergence of these subfamilies ~ 140 million years ago⁴². This IgH gene locus is interesting in several ways. First, this locus contains orthologs of all IGHC genes found in mammals (given that the mammalian γ and ϵ evolved from ν) even though the δ or α genes have not been observed in other reptiles and birds studied to date. Second, this is the first IgH gene locus in tetrapods that has been found to possess multiple μ genes. Third, it is the first IgH locus in non-mammalian tetrapods that has been shown to have subclass divergence of both the ν and α genes, although the subclass divergence of the counterparts of these genes is commonly observed in mammals. Fourth, similar to birds²⁷, all of the α genes in this IgH locus showed the opposite transcriptional orientation to that of the remaining genes, indicating that α gene inversion may have occurred before the

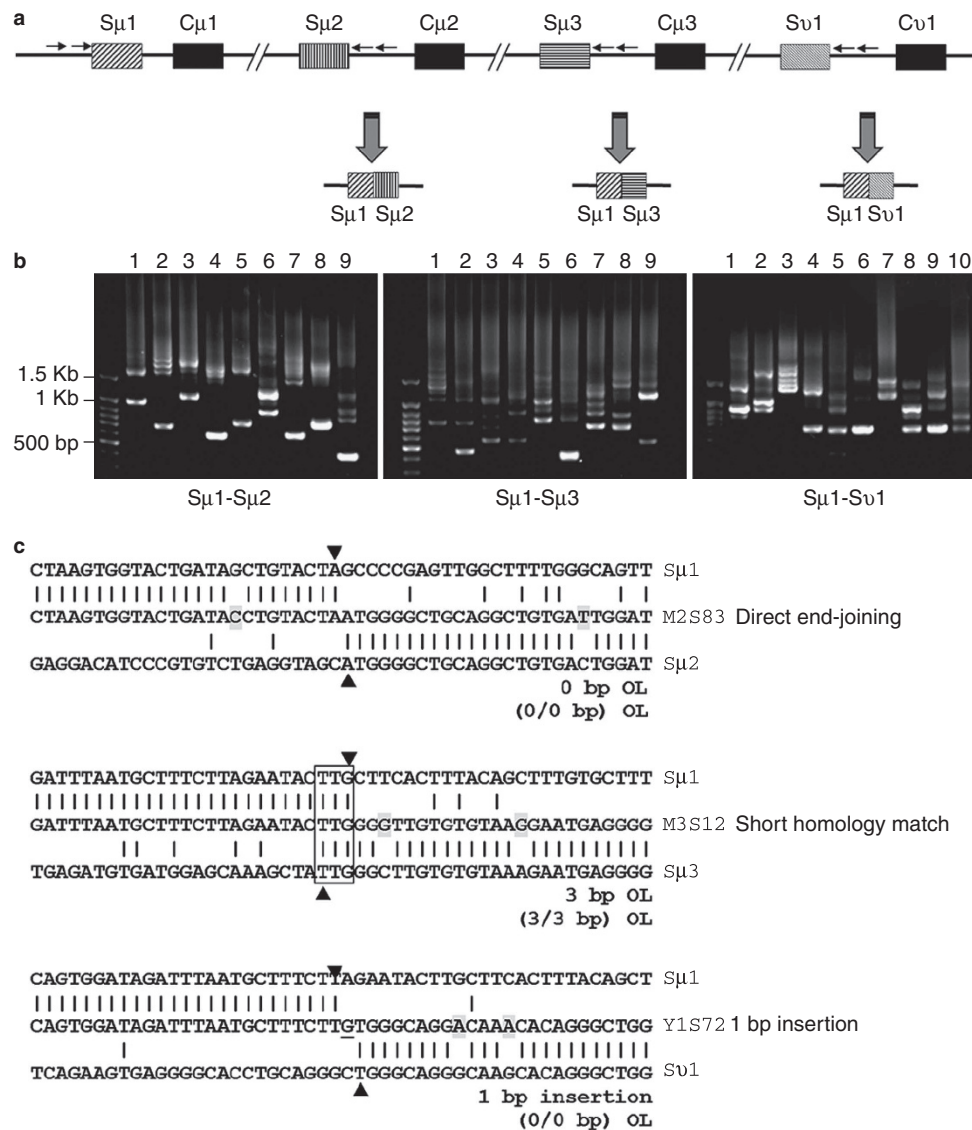


Figure 4 | Amplification and analysis of the recombined switch fragments $S\mu 1$ - $S\mu 2$, $S\mu 1$ - $S\mu 3$ and $S\mu 1$ - $Sv 1$. **(a)** Schematic demonstration of the PCR strategy used for the amplification of recombined switch fragments. **(b)** PCR amplification of recombined switch fragments. **(c)** Three representative junctions of the cloned switch fragments. For each alignment, the sequence of the cloned fragment is shown in the middle, and the germline sequences of the two S regions that are involved are shown above and below. The arrows indicate the break points, whereas boxed nucleotides denote microhomology shared by two S regions. The inserted nucleotide in the junction site is underlined. OL: overlap.

divergence of crocodylians and birds. Last, a large number (68) of *DH* gene segments was found in the IgH locus of the Chinese alligator, suggesting that *DH* segments may contribute significantly to antibody diversity in crocodylians.

The present work allows us to more clearly track the evolutionary history of IgH genes in jawed vertebrates, particularly tetrapods (Fig. 7). As IgM and IgD(W) are now commonly accepted to be the most primordial IgH classes in evolution^{17,21,22}, it is reasonable to conclude that all other IgH classes evolved from them through various molecular mechanisms such as gene duplication, conversion or recombination. These processes gave rise to IgNAR and IgZ(T) after the divergence of cartilaginous fish and bony fish, respectively^{12–14}. In comparison with fish, two additional IgH classes, IgY and IgA(X), have evolved in tetrapods, including amphibians, reptiles and birds. Although the origin of these classes is unknown, the present study suggests that both classes emerged before the divergence of tetrapods and that the common

tetrapod ancestor expressed IgM, IgD, IgY and IgA. During tetrapod evolution, lineage-specific IgH class addition or deletion was adopted to shape the IgH locus of the modern living tetrapods. In *Xenopus*, the new isotype IgF, apparently derived from IgY gene duplication, was identified in addition to IgM, IgD, IgY and IgX¹⁷ (Fig. 7). Interestingly, IgA was found to be lost in a number of reptilian species, including lizards and turtles^{25,41,43}, whereas IgD is absent in all birds investigated thus far^{22,27,44–46} (Fig. 7). Another important point regarding Ig evolution in tetrapods is that mammalian species inherited all IgH classes from the common tetrapod ancestor, with IgY subsequently functionally diversifying into IgG and IgE²⁶.

The finding of three functional μ genes was unexpected. Both $\mu 2$ and $\mu 3$ genes were expressed via the CSR mechanism, which was previously shown to induce a non-IgM class switch in IgM-expressing B cells³³. CSR is not utilized in bony or cartilaginous fish as a mechanism for expressing non-IgM/IgD antibody classes, even though the CSR initiator AID has been identified in

Table 1 | Summary of the junctional patterns and nucleotide mutations flanking the junction sites in the recombined switch fragments.

Junction	Direct end-joining	Perfectly matched short homology			Insertions			No. of fragments
		1-3 bp	4-6 bp	7-9 bp	1 bp	2-3 bp	≥4 bp	
S μ 1-S μ 2	13 (24%)	28 (52%)	3 (6%)	0 (0%)	7 (13%)	1 (2%)	2 (4%)	54
S μ 1-S μ 3	16 (30%)	29 (55%)	1 (2%)	0 (0%)	7 (13%)	0 (0%)	0 (0%)	53
S μ 1-S ν 1	13 (29%)	19 (42%)	7 (16%)	0 (0%)	5 (11%)	1 (2%)	0 (0%)	45
Total junctions	42 (28%)	76 (50%)	11 (7%)	0 (0%)	19 (13%)	2 (1%)	2 (1%)	152

	Mutations in recombined fragments					
	AT versus GC mutations	Transitions	Mutations in WRC or GYW	Frequency (per 1,000 bp)	No. of bp sequenced	No. of mutations
S μ 1-S μ 2	5 versus 34 (13% versus 89%)	13 (33%)	27 (69%)	14.4	2,700	39
S μ 1-S μ 3	5 versus 39 (11% versus 89%)	23 (52%)	34 (77%)	16.6	2,650	44
S μ 1-S ν 1	2 versus 35 (5% versus 95%)	18 (49%)	26 (70%)	16.4	2,250	37
Total junctions	12 versus 108 (10% versus 90%)	54 (50%)	87 (73%)	15.7	7,600	120

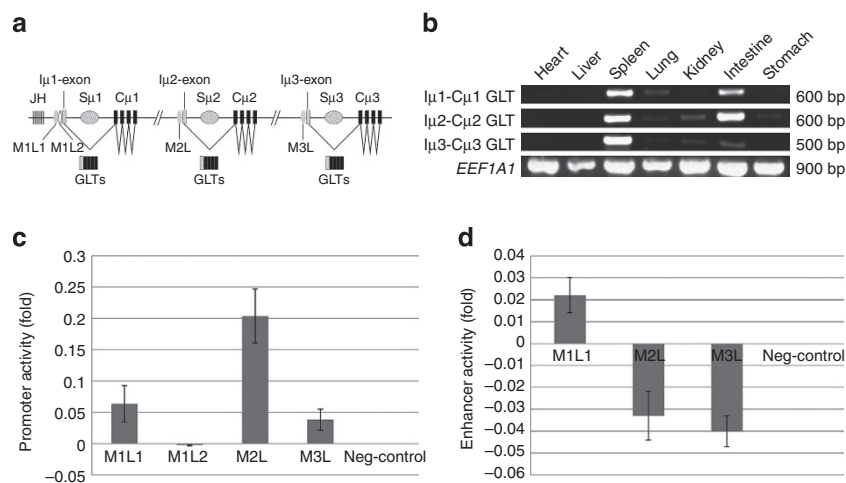


Figure 5 | Analysis of GL transcription, including the I promoter and enhancer for the three μ genes. (a) Schematic map of the location of the I μ exons and splicing of the GL transcripts. **(b)** GL transcription of the μ genes in various tissues. **(c)** Analysis of I promoter activity in the three μ genes. **(d)** Analysis of enhancer activity in the three I promoters. In both **(c)** and **(d)**, the pGL3-promoter or enhancer vector was used as a negative control. The Y axis indicates the relative response ratio. Data in **(c)** and **(d)** were based on three independent experiments. Error bars \pm s.e.m.

both groups of fish⁴⁷⁻⁴⁹. Although many *cis* elements, such as the switch region, the I promoter and germline transcription, are required for CSR in tetrapods, none of these elements have been detected in fish³³. Although the evolution of CSR must have been accompanied by the acquisition of these *cis* elements, it remains puzzling how these genetic elements were acquired by non-IgM/IgD-encoding genes. The discovery of μ to μ CSR in this study strongly suggests that these elements developed by gene duplication of the most 5'- μ gene, which is similar to the finding that multiple crocodilian μ genes likely resulted from gene duplication. Furthermore, it has long been hypothesized that non-IgM/IgD-encoding genes in tetrapods were initially derived by duplication of the μ gene²⁶. In addition, as there is an E μ enhancer with promoter activity between the JH region and the first μ in mammals and birds^{33,50}, the I promoters of non-IgM/IgD-encoding genes are likely to be derived from the E μ enhancer accompanying the duplication process.

Identification of the crocodilian α genes in this study also provides a number of significant clues to the evolution of this gene. First, with the addition of the crocodilian α genes (representing the first *bone fide* reptilian α gene cloned thus

far), the phylogenetic analysis clearly revealed that the α genes in reptiles, birds and mammals, as well as IgX in amphibians, share a common ancestor⁵¹. This finding suggests that the amphibian IgX is orthologous to IgA⁵¹ and that the α gene emerged before the divergence of tetrapods (including amphibians, reptiles, birds and mammals). Considering the early emergence of the tetrapod α gene, it is remarkable that this antibody class, with a major importance in mucosal immunity, is missing in some reptiles²⁵. Second, similar to birds, the crocodilian α genes were all inverted, suggesting that the α gene inversion also occurred before the divergence of crocodilians and birds. Third, this study provides the first evidence in non-mammalian tetrapods for the evolution of IgA subclass-encoding genes, which may have led to functional divergence and a benefit to mucosal immunity.

In conclusion, this study reports the identification of a distinct IgH gene locus in crocodilians that demonstrates extensive subclass divergence of the μ , α and ν genes. Our data provide significant insight into Ig-gene evolution and improve the understanding of how a flexible Ig-gene system can develop in various species in response to long-term environmental selection pressure.

Methods

Sample collection. The Siamese crocodile (*Crocodylus siamensis*) was purchased from a crocodile breeding farm in Tianjin, and the Chinese alligator (*Alligator sinensis*) tissue samples were collected from the Anhui Research Centre for the Reproduction of the Chinese Alligator. Blood samples for the BAC library construction were collected from the Beijing Zoo. These studies were approved by the Animal Care and Use Committee of the China Agricultural University.

BAC library construction. The BAC library was constructed using a service provided by Bioestablish Biotechnology Co., Ltd (Beijing, China). Briefly, high-

molecular-weight DNA was isolated from the blood cells of a Chinese alligator and embedded in agarose plugs. The DNA-containing agarose plugs were partially digested by *Hind* III according to a standard protocol. The partially digested DNA was subjected to two rounds of size selection by pulsed-field gel electrophoresis in a 1% agarose gel. DNA fragments ranging from 100 to 300 kb were excised from the gel and recovered. The size-selected, digested DNA was ligated into a *Hind* III-digested and dephosphorylated BAC vector. The ligation reaction was transformed into electrocompetent EPI300 cells to obtain $\sim 2 \times 10^5$ recombinant clones, which were loaded onto 560 384-well plates. Super and secondary pools were also established for PCR screening.

Genome walking. The IgH locus of the Siamese crocodile was mapped by genome walking using a genome walking kit (Clontech, CA, USA) according to the manufacturer's instructions. All amplifications were conducted using LA Taq DNA polymerase (Takara, Dalian, China) with proof-reading activity.

Construction and screening of an IgH-specific mini-cDNA library. JH-derived primers were used to perform 3'-RACE (rapid amplification of cDNA 3'-ends) using RNA isolated from the spleen and small intestine. The resultant PCR products were cloned into a pMD19-T vector. Recombinant clones were screened by sequencing and PCR.

5'-RACE amplification of the IgH heavy-chain variable regions. Primers derived from the IgM heavy chain constant regions were used for 5'-RACE amplification using spleen RNA as a template. The 5'-RACE kit used was 5'-RACE System Version 2.0 (Invitrogen, NY, USA). The resultant PCR products were cloned into a pMD19-T vector and sequenced.

Quantitative RT-PCR. Total RNA was isolated from various tissues using an RNeasy Mini Kit (Qiagen, Valencia, CA, USA). cDNA was prepared using the QuantiTect Reverse Transcription Kit (Qiagen). The TaqMan gene expression assays were performed using the TaqMan Matrix Mix and TaqMan probes.

Analysis of recombined switching fragments. Two sense primers were designed using the 5'-flanking sequence of S μ 1, while six anti-sense primers (two for each S region) were designed based on the 3'-flanking sequences of S μ 2, S μ 3 and S μ 1. These primers were used for a nested PCR to amplify the recombined fragments of S μ 1-S μ 2, S μ 1-S μ 3 and S μ 1-S μ 1. The PCR was performed using the high-fidelity enzyme FastPfu DNA Polymerase (TansGen, Beijing, China). The resulting PCR products were cloned into the pMD19-T vector and sequenced. The primers sequences were as follows: M1U1 (5'-TGTGCTTGAAGTGCATAGGA-3') and M1U2 (5'-AACCTTGGAGCGATTTCTGAT-3'); M2L1 (5'-AACCTCTTTGAAAGCCAGCTC-3') and M2L2 (5'-GCCACAAACCACAGAGCTAAG-3'); M3L1 (5'-ATCATCTCATTACAGCCACCTG-3') and M3L2 (5'-TAATGGGAGAAGCCGGAAGTA-3'); and Y1L1 (5'-CACAGTGCCCAACTGGTTTAT-3') and Y1L2 (5'-TAACCAGCCTAGCCAGTCTCA-3'). The junction sites of these recombined fragments and the mutations in the flanking regions were determined by sequence alignment with germline S regions.

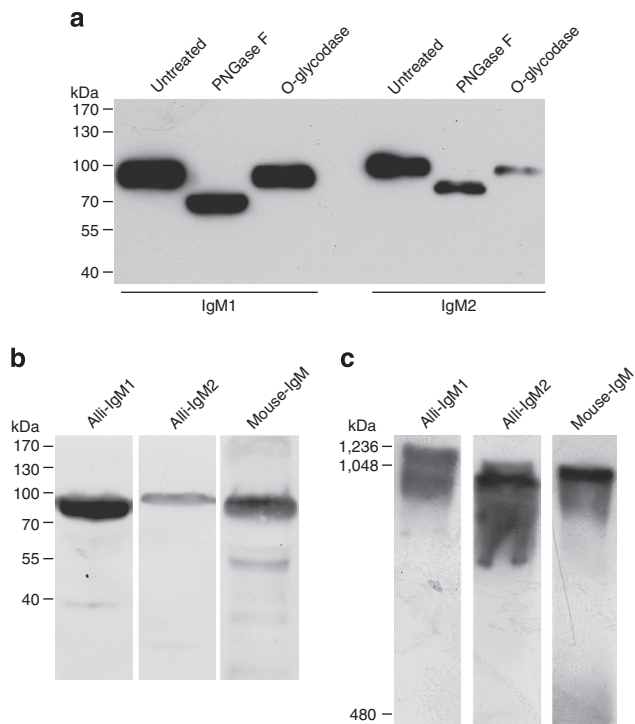


Figure 6 | Western blotting for IgM1 and IgM2 in the serum of a Chinese alligator. (a) Analysis of IgM1 and IgM2 treated by glycosidase under reducing conditions. (b) Analysis of IgM1 and IgM2 under reducing conditions. (c) Analysis of IgM1 and IgM2 under non-reducing conditions. In both (b) and (c), mouse serum IgM was used for comparison.

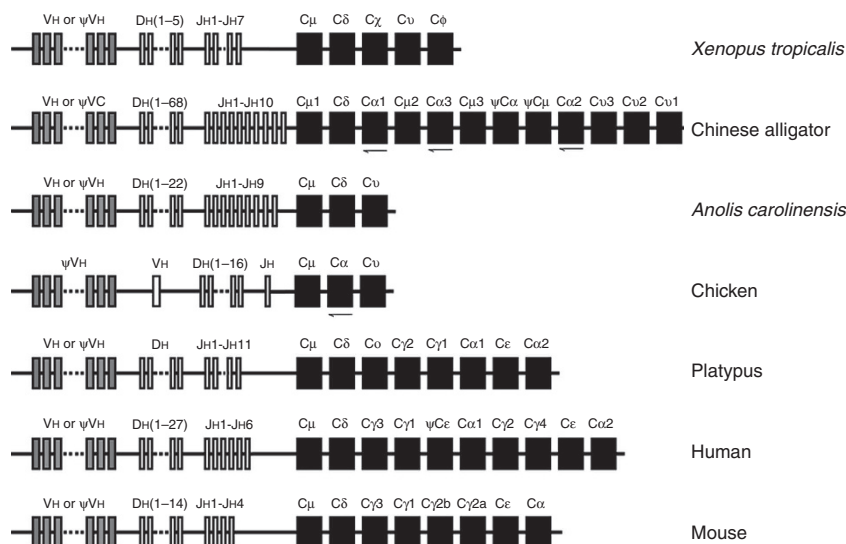


Figure 7 | Genomic organization of the IGHC gene locus in different species. 'ψ' is used to denote pseudo genes. Arrows indicate the transcriptional orientation of the chicken and Chinese alligator genes.

Promoter and enhancer activity analysis. The sequences upstream of the $I\mu 1$, $I\mu 2$ and $I\mu 3$ regions were cloned into both the pGL3 enhancer and pGL3-promoter vectors (Promega, WI, USA). These plasmid vectors were transfected into chicken DT40 cells using Lipofectamine 2000 (Invitrogen, NY, USA) together with the PRL-TK vector at a ratio of 1:9. The pGL3 enhancer and pGL3 control vectors were used as negative and positive controls, respectively. The activity of firefly luciferase and renilla luciferase was detected using the Dual-Glo Luciferase Assay System E2920 (Promega, WI, USA).

Preparation of mAbs against the Chinese alligator IgM1 and IgM2. The mouse mAbs were prepared using a service provided by Abmart (Shanghai, China). The specificity of the obtained mAbs was confirmed by western blotting for the constant regions of IgM1 and IgM2 expressed in 293T cells. 1B2 and 3B1 mAbs were selected for the detection of IgM1 and IgM2, respectively, in the serum.

Phylogenetic analysis. The phylogenetic trees were constructed using MrBayes3.1.2⁵² and were viewed using FigTree or Tree View⁵³. The amino-acid sequences of the entire heavy-chain constant region of most isotypes were used in the analysis, whereas the first four CH regions were used for the IgD analysis of non-mammalian vertebrates. The NCBI GenBank accession numbers of the sequences used were as follows: μ , nurse shark, M92851; skate, M29679; trout, X65261; cod, CAA41680.1; rainbow trout, AAB27359.2; zebrafish, AF281480; catfish, CAB38072.1; lung fish, AAO52808.1; Chinese soft-shelled turtle, ACU45376.1; chicken, X01613; duck, AJ314754; *X. laevis*, X15114; lizard, ABV66128; gecko, ABY74509; platypus, AY168639; human, X14940; mouse, V00818; *X. tropicalis*, AAH89670; axolotl, A46532; and Iberian ribbed newt, CAE02685; δ , catfish, U67437; *X. tropicalis*, DQ350886; human, BC021276; platypus, ACD31540; mouse, J00449; lizard, ABV66130; gecko, ABY67439; Chinese soft-shelled turtle, ACU45375; and trout, AAW66977; α or γ , chicken, S40610; duck, AJ314754; human, J00220; mouse, J00475; platypus, AY055778; *X. laevis*, BC072981; *X. tropicalis*, AAI57651; and axolotl, CAO82107; γ , human, J00228; mouse, J00453; and platypus, AY055781; ϵ , human, J00222; mouse, X01857; and platypus, AY055780; ν : chicken, X07175; duck, X78273; *X. laevis*, X15114; axolotl, CAA49247; lizard, ABV66132; gecko ACF60236; *X. tropicalis* BC089679; Chinese soft-shelled turtle, ACU45374; and Iberian ribbed newt, CAE02686; $\zeta/\tau/\pi$, zebrafish, AY643752; trout, AY872256; and Iberian ribbed newt, CAL25718; and ω , sandbar shark, U40560; lungfish, AF437727; and nurse shark, U51450.

Statistical analysis. The statistical analysis was performed using a one-way analysis of variance in SPSS1.5.

References

- Flajnik, M. F. Comparative analyses of immunoglobulin genes: surprises and portents. *Nat. Rev. Immunol.* **2**, 688–698 (2002).
- Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59 (2010).
- Das, S., Hirano, M., McCallister, C., Tako, R. & Nikolaidis, N. Comparative genomics and evolution of immunoglobulin-encoding loci in tetrapods. *Adv. Immunol.* **111**, 143–178 (2011).
- Hirano, M., Das, S., Guo, P. & Cooper, M. D. The evolution of adaptive immunity in vertebrates. *Adv. Immunol.* **109**, 125–157 (2011).
- Cooper, M. D. & Herrin, B. R. How did our complex immune system evolve? *Nat. Rev. Immunol.* **10**, 2–3 (2010).
- Guo, P. *et al.* Dual nature of the adaptive immune system in lampreys. *Nature* **459**, 796–801 (2009).
- Alder, M. N. *et al.* Antibody responses of variable lymphocyte receptors in the lamprey. *Nat. Immunol.* **9**, 319–327 (2008).
- Reynaud, C. A., Anquez, V., Grimal, H. & Weill, J. C. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48**, 379–388 (1987).
- Reynaud, C. A., Dahan, A., Anquez, V. & Weill, J. C. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* **59**, 171–183 (1989).
- Berstein, R. M., Schluter, S. F., Shen, S. & Marchalonis, J. J. A new high molecular weight immunoglobulin class from the carcharhine shark: implications for the properties of the primordial immunoglobulin. *Proc. Natl Acad. Sci. USA* **93**, 3289–3293 (1996).
- Ota, T., Rast, J. P., Litman, G. W. & Amemiya, C. T. Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. *Proc. Natl Acad. Sci. USA* **100**, 2501–2506 (2003).
- Greenberg, A. S. *et al.* A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* **374**, 168–173 (1995).
- 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.* **6**, 295–302 (2005).
- 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. USA* **102**, 6919–6924 (2005).
- Zhang, Y. A. *et al.* IgT, a primitive immunoglobulin class specialized in mucosal immunity. *Nat. Immunol.* **11**, 827–835 (2010).
- Wilson, M. *et al.* A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. *Proc. Natl Acad. Sci. USA* **94**, 4593–4597 (1997).
- Zhao, Y. *et al.* Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*. *Proc. Natl Acad. Sci. USA* **103**, 12087–12092 (2006).
- Zhao, Y. *et al.* Ornithorhynchus anatinus (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods. *J. Immunol.* **183**, 3285–3293 (2009).
- Hsu, E., Flajnik, M. F. & Du Pasquier, L. A third immunoglobulin class in amphibians. *J. Immunol.* **135**, 1998–2004 (1985).
- Gambon-Deza, F., Sanchez-Espinel, C. & Magadan-Mompo, S. The immunoglobulin heavy chain locus in the platypus (*Ornithorhynchus anatinus*). *Mol. Immunol.* **46**, 2515–2523 (2009).
- Ohta, Y. & Flajnik, M. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proc. Natl Acad. Sci. USA* **103**, 10723–10728 (2006).
- Sun, Y., Wei, Z., Hammarstrom, L. & Zhao, Y. The immunoglobulin delta gene in jawed vertebrates: a comparative overview. *Dev. Comp. Immunol.* **35**, 975–981 (2011).
- Litman, G. W., Anderson, M. K. & Rast, J. P. Evolution of antigen binding receptors. *Annu. Rev. Immunol.* **17**, 109–147 (1999).
- Deza, F. G., Espinel, C. S. & Beneitez, J. V. A novel IgA-like immunoglobulin in the reptile *Eublepharis macularius*. *Dev. Comp. Immunol.* **31**, 596–605 (2007).
- Wei, Z. *et al.* Expression of IgM, IgD, and IgY in a reptile, *Anolis carolinensis*. *J. Immunol.* **183**, 3858–3864 (2009).
- Warr, G. W., Magor, K. E. & Higgins, D. A. IgY: clues to the origins of modern antibodies. *Immunol. Today* **16**, 392–398 (1995).
- 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.* **276**, 46729–46736 (2001).
- Gambon Deza, F., Sanchez Espinel, C. & Magadan Mompo, S. The immunoglobulin heavy chain locus in the reptile *Anolis carolinensis*. *Mol. Immunol.* **46**, 1679–1687 (2009).
- Janke, A. & Arnason, U. The complete mitochondrial genome of Alligator mississippiensis and the separation between recent archosauria (birds and crocodiles). *Mol. Biol. Evol.* **14**, 1266–1272 (1997).
- Darville, L. N., Merchant, M. E., Hasan, A. & Murray, K. K. Proteome analysis of the leukocytes from the American alligator (*Alligator mississippiensis*) using mass spectrometry. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **5**, 308–316 (2010).
- Merchant, M. E. *et al.* Broad spectrum antimicrobial activity of leukocyte extracts from the American alligator (*Alligator mississippiensis*). *Vet. Immunol. Immunopathol.* **110**, 221–228 (2006).
- Zhao, Y., Kacsokovics, I., Rabbani, H. & Hammarstrom, L. Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus. *J. Biol. Chem.* **278**, 35024–35032 (2003).
- Stavnezer, J., Guikema, J. E. & Schrader, C. E. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* **26**, 261–292 (2008).
- Pan-Hammarstrom, Q., Zhao, Y. & Hammarstrom, L. Class switch recombination: a comparison between mouse and human. *Adv. Immunol.* **93**, 1–61 (2007).
- Pan-Hammarstrom, Q. *et al.* ATM is not required in somatic hypermutation of VH, but is involved in the introduction of mutations in the switch mu region. *J. Immunol.* **170**, 3707–3716 (2003).
- Litman, G. W. *et al.* Complete nucleotide sequences of three VH genes in Caiman, a phylogenetically ancient reptile: evolutionary diversification in coding segments and variation in the structure and organization of recombination elements. *Proc. Natl Acad. Sci. USA* **82**, 844–848 (1985).
- Turchin, A. & Hsu, E. The generation of antibody diversity in the turtle. *J. Immunol.* **156**, 3797–3805 (1996).
- Wu, Q. *et al.* Phylogeny, genomic organization and expression of lambda and kappa immunoglobulin light chain genes in a reptile, *Anolis carolinensis*. *Dev. Comp. Immunol.* **34**, 579–589.
- Das, S., Nikolaidis, N., Klein, J. & Nei, M. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proc. Natl Acad. Sci. USA* **105**, 16647–16652 (2008).
- Gambon-Deza, F. & Espinel, C. S. IgD in the reptile leopard gecko. *Mol. Immunol.* **45**, 3470–3476 (2008).

41. Xu, Z., Wang, G. L. & Nie, P. IgM, IgD and IgY and their expression pattern in the Chinese soft-shelled turtle *Pelodiscus sinensis*. *Mol. Immunol.* **46**, 2124–2132 (2009).
42. Janke, A., Gullberg, A., Hughes, S., Aggarwal, R. K. & Arnason, U. Mitogenomic analyses place the gharial (*Gavialis gangeticus*) on the crocodile tree and provide pre-K/T divergence times for most crocodylians. *J. Mol. Evol.* **61**, 620–626 (2005).
43. Gambon-Deza, F., Sanchez-Espinel, C., Mirete-Bachiller, S. & Magadan-Mompo, S. Snakes antibodies. *Dev. Comp. Immunol.* **38**, 1–9 (2012).
44. Huang, T. *et al.* Analysis of immunoglobulin transcripts in the ostrich *Struthio camelus*, a primitive avian species. *PLoS One* **7**, e34346.
45. Zhao, Y., Rabbani, H., Shimizu, A. & Hammarstrom, L. Mapping of the chicken immunoglobulin heavy-chain constant region gene locus reveals an inverted alpha gene upstream of a condensed epsilon gene. *Immunology* **101**, 348–353 (2000).
46. 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.* **30**, 93–100 (2006).
47. Barreto, V. M. *et al.* AID from bony fish catalyzes class switch recombination. *J. Exp. Med.* **202**, 733–738 (2005).
48. Zhao, Y., Pan-Hammarstrom, Q., Zhao, Z. & Hammarstrom, L. Identification of the activation-induced cytidine deaminase gene from zebrafish: an evolutionary analysis. *Dev. Comp. Immunol.* **29**, 61–71 (2005).
49. Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K. & Neuberger, M. S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* **22**, 367–377 (2005).
50. Lundqvist, M. L., McElveen, B. R., Middleton, D. L. & Warr, G. W. Enhancer and promoter activity in the JH to IGHM intron of the Pekin duck, *Anas platyrhynchos*. *Dev. Comp. Immunol.* **31**, 286–295 (2007).
51. Musmann, R., Du Pasquier, L. & Hsu, E. Is *Xenopus* IgX an analog of IgA? *Eur. J. Immunol.* **26**, 2823–2830 (1996).
52. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
53. Page, R. D. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357–358 (1996).

Acknowledgements

This work was supported by the National Science Fund for Distinguished Young Scholars (30725029), the National Basic Research Programme of China (973 Programme-2010CB945300) and the Taishan Scholar Foundation of Shandong Province. The authors wish to thank Dr Martin F Flajnik for his advice.

Author contributions

Y.Z., G.C., N.L., Q.P.H. and L.H. designed the experiments and wrote the manuscript. G.C., Y.G., T.W., Y.S., Z.W., L.L., L.R., Y.G. and X.H. performed the experiments. X.W., G.L. and J.Y. helped with the sequencing work. Y.L., C.Z. and X.W. provided materials and technical help.

Additional information

Accession codes: Sequence data has been deposited in DDBJ/EMBL/GenBank under accession codes JQ417410 to JQ417429 and JQ479335 to JQ479336.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Cheng, G. *et al.* Extensive diversification of IgH subclass-encoding genes and IgM subclass switching in crocodylians. *Nat. Commun.* **4**:1337 doi: 10.1038/ncomms2317 (2013).