THE JOURNAL OF IMMUNOLOGY

This information is current as of March 11, 2013.

Ornithorhynchus anatinus (Platypus) Links the Evolution of Immunoglobulin Genes in Eutherian Mammals and Nonmammalian Tetrapods

Yaofeng Zhao, Huiting Cui, Camilla M. Whittington, Zhiguo Wei, Xiaofeng Zhang, Ziding Zhang, Li Yu, Liming Ren, Xiaoxiang Hu, Yaping Zhang, Lars Hellman, Katherine Belov, Ning Li and Lennart Hammarström

Supplementary Material	http://www.jimmunol.org/content/suppl/2009/08/12/jimmunol.0900469.DC1.html
References	This article **cites 38 articles**, 14 of which you can access for free at: http://www.jimmunol.org/content/183/5/3285.full#ref-list-1
Subscriptions	Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscriptions
Permissions	Submit copyright permission requests at: http://www.aai.org/ji/copyright.html
Email Alerts	Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/cgi/alerts/etoc

Ornithorhynchus anatinus (Platypus) Links the Evolution of Immunoglobulin Genes in Eutherian Mammals and Nonmammalian Tetrapods[1,2]

Yaofeng Zhao,[3,4]* Huiting Cui,[3]* Camilla M. Whittington,[†] Zhiguo Wei,* Xiaofeng Zhang,[‡] Ziding Zhang,[§] Li Yu,[¶] Liming Ren,* Xiaoxiang Hu,* Yaping Zhang,[¶] Lars Hellman,[∥] Katherine Belov,[†] Ning Li,[4]* and Lennart Hammarström[4#]

The evolutionary origins of mammalian immunoglobulin H chain isotypes (IgM, IgD, IgG, IgE, and IgA) are still incompletely understood as these isotypes differ considerably in structure and number from their counterparts in nonmammalian tetrapods. We report in this study that the platypus (*Ornithorhynchus anatinus*) Ig H chain constant region gene locus contains eight Ig encoding genes, which are arranged in an μ-δ-o-$\gamma2$-$\gamma1$-$\alpha1$-ϵ-$\alpha2$ order, spanning a total of ∼200 kb DNA, encoding six distinct isotypes. The *o* (*o* for *Ornithorhynchus*) gene encodes a novel Ig H chain isotype that consists of four constant region domains and a hinge, and is structurally different from any of the five known mammalian Ig classes. This gene is phylogenetically related to ν (ϵ) and γ, and thus appears to be a structural intermediate between these two genes. The platypus δ gene encodes ten heavy chain constant region domains, lacks a hinge region and is similar to IgD in amphibians and fish, but strikingly different from that in eutherian mammals. The platypus Ig H chain isotype repertoire thus shows a unique combination of genes that share similarity both to those of nonmammalian tetrapods and eutherian animals and demonstrates how phylogenetically informative species can be used to reconstruct the evolutionary history of functionally important genes.   *The Journal of Immunology,* 2009, 183: 3285–3293.

Immunoglobulins are the key components of the adaptive immune system in jawed vertebrates including mammals, birds, reptiles, amphibians, bony fish, and cartilaginous fish (1, 2). Eutherian (placental) mammals are known to express five classes of Igs including IgM, IgD, IgG, IgE, and IgA, among which only IgM, IgD, and IgA (or its functional equivalent, like amphibian IgX) are found in nonmammalian tetrapods such as birds, reptiles, and amphibians (1–7). Although IgM is structurally conserved in all tetrapods thus far investigated, both mammalian IgD and IgA differ from their respective counterparts in nonmammalian tetrapods by having a structurally flexible hinge and fewer numbers of constant (C) region domains.

IgD is an enigmatic Ig class, primarily because its function remains elusive. Until an ortholog was discovered in the teleost catfish (*Ictalurus punctatus*) a decade ago (8), IgD had been thought to be present only in primates and rodents (9). Recently, the gene encoding IgD was identified in various classes of animals including bony fish, amphibians, reptiles, and additional eutherian mammals such as pigs, cow, sheep, panda, and dogs (3, 4, 6, 10–12), thus showing a distribution across a wide phylogenetic spectrum. Based on analysis of the δ gene in *Xenopus tropicalis* and bony fish (3, 13), investigators uncovered a phylogenetic relationship of IgD with IgW (found in lungfish and cartilaginous fish; Ref. 14, 15), suggesting that the IgD/W lineage is as evolutionarily ancient as IgM. Therefore, the absence of the δ gene in birds (16, 17) and rabbits (18) is probably an outcome of evolutionary loss. Unlike IgM which is structurally conserved across all jawed vertebrates, IgD exhibits extensive variation, particularly in size. The length of IgD can range from two heavy chain constant region domains $(C_H)^5$ to more than ten C_H domains in different species (8–10, 13, 19–21). In addition to the length variations, in some species, such as catfish and pigs, IgD can also be expressed as a chimeric Ig H chain with an inclusion of the IgM C_H1 (8, 22). Compared with their counterparts in nonmammalian vertebrates, IgD molecules

*State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, People's Republic of China; [†]Faculty of Veterinary Science, University of Sydney, Sydney, Australia; [‡]Centre for Structural Biochemistry, Karolinska Institutet at Novum, Huddinge, Sweden; [§]Bioinformatics Center, College of Biological Sciences, China Agricultural University, Beijing, People's Republic of China; [¶]Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming, China; [∥]Department of Cell and Molecular Biology, Biomedical Center, Uppsala, Sweden; and [#]Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska Institutet at Karolinska University Hospital Huddinge, Stockholm, Sweden

[5] Abbreviations used in this paper: C_H, heavy chain constant region domain; IGHC, immunoglobulin heavy chain constant region gene; PFGE, pulsed field gel electrophoresis; TM, transmembrane; BAC, bacterial artificial chromosome.
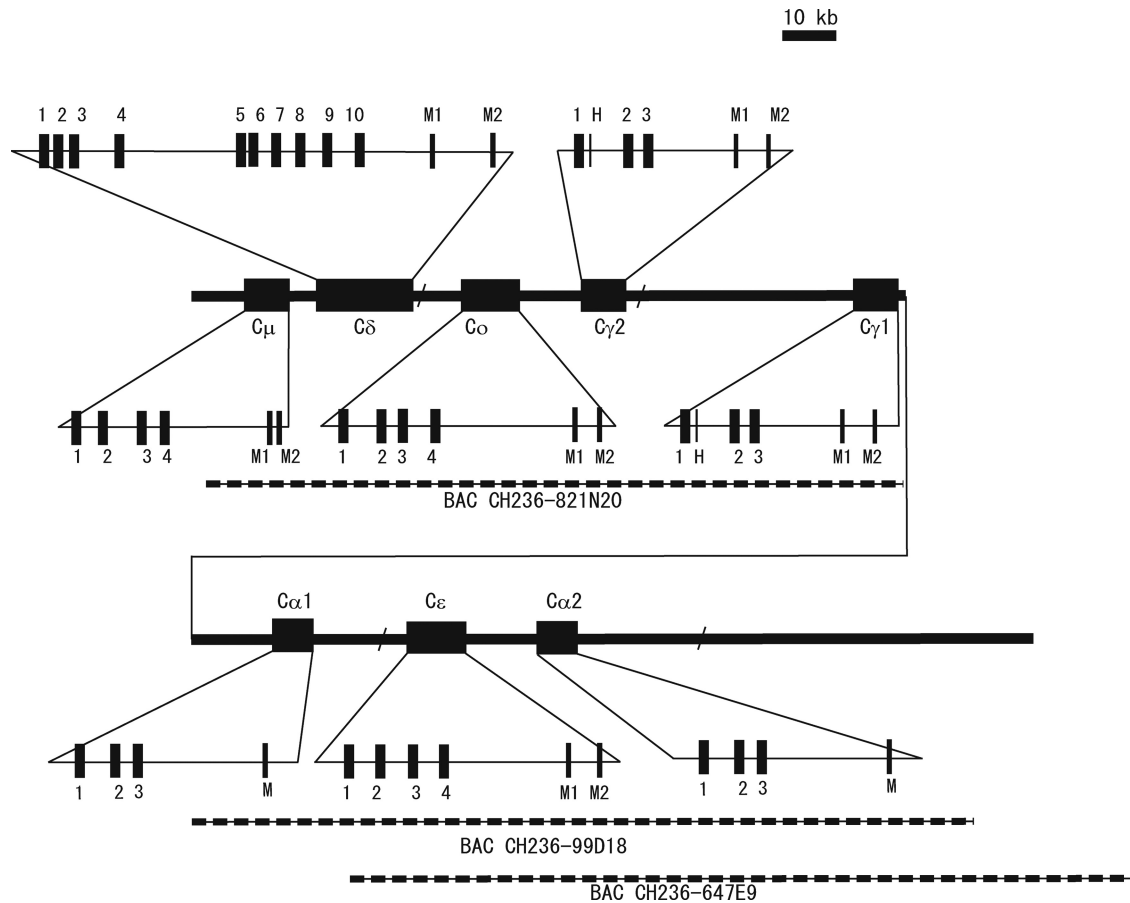
FIGURE 1. Genomic organization of the platypus IgH locus. μ, IgM encoding gene; δ, IgD encoding gene; o, IgO encoding gene; $\gamma1$, IgG1 encoding gene; $\gamma2$, IgG2 encoding gene; ε, IgE encoding gene; $\alpha1$, IgA1 encoding gene; $\alpha2$, IgA2 encoding gene; M, transmembrane region encoding exon. C_H encoding exons for each gene are indicated using Arabic numbers. Slashes indicate sequence gaps longer than 2 kb. Dashed lines indicate the regions covered by the BAC clones used in this study.

from eutherian mammals contain fewer C_H domains and a structurally disordered, flexible hinge region which is not seen in nonmammalian vertebrates. However, the evolutionary process that led to shortening of the IgD molecule and creation of the hinge region remains unclear.

The mammalian IgG and IgE molecules are believed to have evolved through gene duplication and subsequent evolution of IgY (23), an ancient, four-C_H-domain, hinge-lacking Ig class found in birds, reptiles and amphibians. Evolution of IgY to IgE seems to be straightforward, as both Igs have four C_H domains and no hinge, and essentially share sequence homology and similar functions. However, no evidence is available to explain how the hinge of mammalian IgG evolved from IgY, although two hypotheses have been proposed. The first suggests that the IgG hinge region was evolutionarily condensed from the pre-existing IgY C_H2 (23), while the second argues that the hinge originated from the 3′ sequence of an intron (24).

Mammals (class Mammalia) are taxonomically classified into Prototheria (monotremes) and Theria (Metatheria: marsupials, and Eutheria: eutherians). Monotremes contain only five living species, the platypus and four species of echidna. The platypus is indigenous to Australia and displays an unusual mixture of mammalian and reptilian features, both at the genome level (25) and anatomically: females lay eggs but also provide colostral milk for their pups, and males are venomous. Monotremes last shared a common ancestor with therian mammals ~166 million years ago (26), placing them in a prime position for reconstructing the evolutionary history of biologically important molecules.

The duck-billed platypus (*Ornithorhynchus anatinus*) has previously been shown to express four Ig classes: IgM, IgG1/IgG2, IgA1/IgA2, and IgE, that all structurally resemble their respective counterparts in eutherian mammals (27, 28). To explore the evolutionary origins of the mammalian Igs, we analyzed the platypus Ig heavy (H) chain constant region gene locus and identified a novel Ig isotype encoding gene (IgO, O, or *o* for *Ornithorhynchus*) and a ten-C_H domain encoding δ gene.

Materials and Methods

Platypus spleen cDNA library, cDNA, and bacterial artificial chromosome (BAC) clones

The platypus spleen cDNA library was previously constructed by Vernersson et al. (28). BAC clones (CH236-821N20, CH236-99D18, and CH236-647E09) were purchased from BACPAC Resource Center (BPRC) at Children's Hospital Oakland Research Institute in Oakland, California. All platypus tissues were collected opportunistically from animals found dead in the wild. Tissues were frozen at −80°C before DNA/RNA extraction. RNA was extracted using TriReagent according to the manufacturer's instructions (Molecular Research Center). RNA samples were subjected to DNase digestion using standard protocols (Promega). The Invitrogen SuperScript III First-Strand Synthesis System for RT-PCR was then used to make cDNA using the Oligo(dT)$_{20}$ primer as per the manufacturer's instructions (Invitrogen).

Database searches

BLAST searches were performed against the platypus genome sequence deposited in the National Center for Biotechnology Information GenBank (http://www.ncbi.nlm.nih.gov/genome/guide/platypus/). Genomic contigs were retrieved for further analysis. To search for putative Ig domains, the

FIGURE 2. Phylogenetic analysis using the terminal C_H domain. Note: C_H8 for the gecko IgD, C_H7 for both the *X. tropicalis* and platypus IgD were used for the phylogenetic analysis. The scale shown as a bar represents the genetic distance. The credibility value for each node is shown. Except the sequences of the platypus IgD and IgO (obtained in this study), and *Xenopus tropicalis* IgF, IgM, and IgX (4), the remaining Ig sequences were obtained from the NCBI GenBank with the following accession numbers: α or χ genes: chicken, S40610; duck, AJ314754; echidna, AF416951; human, J00220; mouse, J00475; opossum: AF108225; platypus, AY055778, AY055779; *X. laevis*, BC072981. δ genes: cow, AF515672; dog, DQ297185; gecko, EU312156; human, BC021276; horse, AY631942; mouse, J00449; pig, AF515674; rat, AY148494; sheep, AF515673; *X. tropicalis*, DQ350886. γ genes: echidna, AF416949; human, J00228; mouse, J00453; opossum, AF035195; platypus, AY055781, AY055782. ε genes: cow, U63640; echidna, AY099258; human, J00222; mouse, X01857; opossum, AF035194; platypus, AY055780. μ gene: chicken, X01613; duck, AJ314754; echidna, AF416952; gecko, EU287911; human, X14940; mouse, V00818; opossum, AF108226; platypus, AY168639; turtle, U53567; *X. laevis*, BC084123. ν genes: chicken, X07175; duck, X78273; gecko, EU827594; *X. laevis*, X15114.
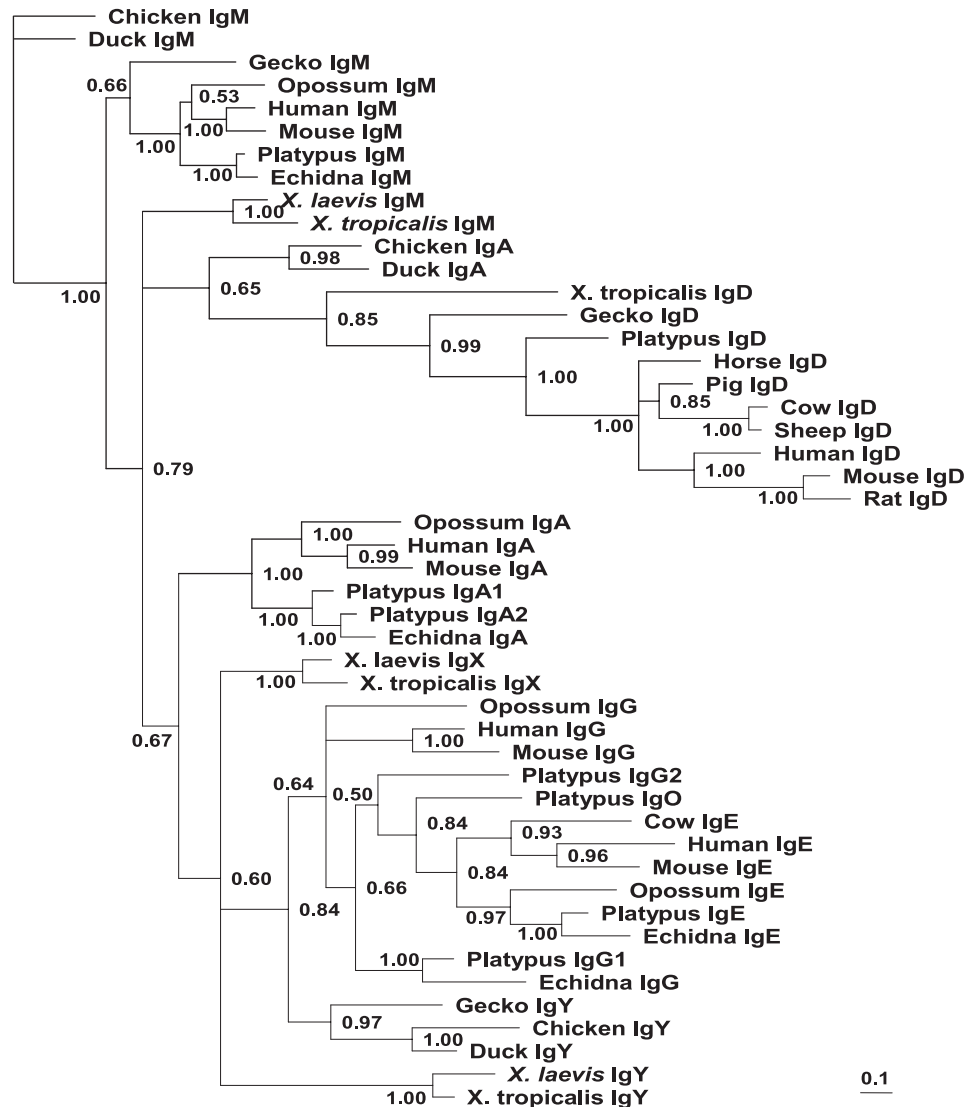
genomic DNA sequences were translated into proteins, which were subsequently subject to protein to protein BLAST searches in the NCBI GenBank. The platypus BAC end sequence database (http://genome.wustl.edu/tools/blast/index.cgi?gsc_link_id = 69) is maintained by the Genome Sequencing Center at Washington University School of Medicine.

Pulsed field gel electrophoresis (PFGE)

Purified BAC DNA was digested using *Not*I and separated on a 1% agarose gel using a CHEF-DR III system (Bio-Rad). The DNA marker used was Midrange I PFG Marker (New England Biolabs). According to PFGE data, the insert size in CH236-821N20, CH236-99D18, and CH236-647E09 were shown to be ~120, 135, and 150-kb, respectively.

Amplification of IgD and IgO constant region cDNA

The platypus IgD constant region cDNA (membrane-bound form) was amplified using a nested PCR amplification with the primers, JHs1 (5′ CAC TGG GGC CAA GGC ACC ATG GT 3′), JHs2 (5′ GGC ACC ATG GTC ACC ATG GT 3′), IgDTMas1 (5′ GGC TTC CTC ACT GGT GGG CAT AG 3′), and IgDTMas2 (5′ AGG GCT ACG AAG GTG GAC ACG GT 3′). The resulting 3.3-kb PCR product was directly sequenced. The IgO C region encoding cDNA was amplified using a nested PCR from a duck-billed platypus cDNA library using the primers JHs1 and IgOas1 (5′ TTT GCC GCA CTG TCT TCT GTA TG 3′), JHs2 and IgOas2 (5′ TGC AGG TGA AGG AAT CTC CGC GT 3′). In both amplifications, the DNA polymerase used was LA *Taq*DNA polymerase (Takara, Dalian), a proof-reading enzyme.

PCR detection of the platypus Ig gene expression in different tissues

cDNA samples derived from seven tissues including gastrointestinal tract, kidney, liver, lung, reproductive tract, spleen, and testis, were used in PCR detection using normal *Taq*DNA polymerase (Tiangen Biotech). PCR primers were: IgM-detections: 5′ ACA AGC CTA TTC CAC GAC CTC 3′; IgM-detectionAs: 5′ GTT GAA GTG CTT GGC CTG ACA 3′; IgE-detections: 5′ GCG CTG GAA CTC TGG AGA CCT GG 3′; IgE-detectionAs: 5′ TCC TGG CAC TGT CTT GGT AGG TT 3′; IgD-detections: 5′ CTT AGC CAG TGT CCA GAT CAG 3′; IgD-detectionAs: 5′ GCA GAG CAG AGT CGG ATG ATC 3′; IgO-detections: 5′ TTC TAC AAC AGC AAA GCC CAG AC 3′; IgO-detectionAs: 5′ GGG CAA TGA AGA GGT CAG TAG GG 3′; IgG1-detections: 5′ GTA CCA AGG CGA CAA GGA ACT 3′; IgG1-detectionAs: 5′ TGT TCA TCT CAT CTC GGT GAG 3′; IgG2-detections: 5′ CGC ACA GAC GTC ATC AGA TAC 3′; IgG2-detectionAs: 5′ TGC TCA CAG ATT CCT TGC TCA 3′; IgA1-detections: 5′ GAC AGT GAC CAG GCT TCT TCC 3′; IgA1-detectionAs: 5′ GGA GAG CAC TTG CAA CAC TGT 3′; IgA2-detections: 5′ CTG ACG ACA AGA CCT ACC AAT 3′; IgA2-detectionAs: 5′ GTT AGT TCA CAG GTC AGG TTG 3′; platypus-GAPDHs: 5′ ACC GGC CAA ATA CGA TGA TAT 3′; platypus-GAPDHAs: 5′ TAC CAG GAG ATG AGC TTG ACA 3′.

DNA computations, structural prediction, and construction of phylogenetic trees

DNA and protein sequence editing, alignments and comparisons were performed using the DNAstar program (DNAstar). Phylogenetic trees were
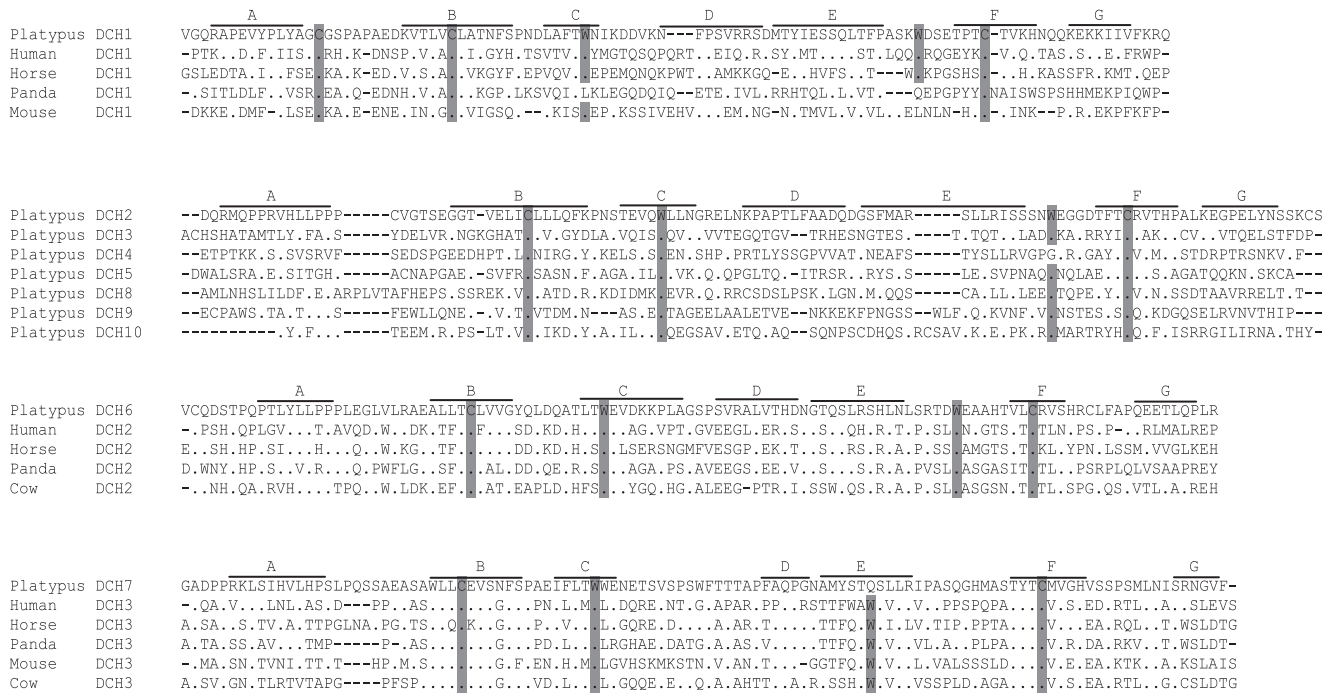
```
                    A                     B                  C                   D             E                         F              G
Platypus DCH1   VGQRAPEVYPLYAGCGSPAPAEDKVTLVCLATNFSPNDLAFTWNIKDDVKN---FPSVRRSDMTYIESSQLTFPASKWDSETPTCVTVKHNQQKEKKIIVFKRQ
Human    DCH1   -PTK..D.F.IIS..RH.K-DNSP.V.A..I.GYH.TSVTV..YMGTQSQPQRT..EIQ.R.SY.MT....ST.LQQ.RQGEYK..V.Q.TAS.S..E.FRWP-
Horse    DCH1   GSLEDTA.I..FSE..KA.K-ED.V.S.A..VKGYF.EPVQV..EPEMQNQKPWT..AMKKGQ-E..HVFS..T---W.KPGSHS.-..H.KASSFR.KMT.QEP
Panda    DCH1   -.SITLDLF..VSR..EA.Q-EDNH.V.A...KGP.LKSVQI.LKLEGQDQIQ--ETE.IVL.RRHTQL.L.VT.---QEPGPYY.NAISWSPSHHMEKPIQWP-
Mouse    DCH1   -DKKE.DMF-.LSE..KA.E-ENE.IN.G..VIGSQ.--.KIS..EP.KSSIVEHV...EM.NG-N.TMVL.V.VL..ELNLN-H...INK--P.R.EKPFKFP-
```

```
                    A                        B                 C                  D                E                    F               G
Platypus DCH2   --DQRMQPPRVHLLPPP-----CVGTSEGGT-VELICLLLQFKPNSTEVQWLLNGRELNKPAPTLFAADQDGSFMAR-----SLLRISSSNWEGGDTFTCRVTHPALKEGPELYNSSKCS
Platypus DCH3   ACHSHATAMTLY.FA.S-----YDELVR.NGKGHAT..V.GYDLA.VQIS..QV..VVTEGQTGV--TRHESNGTES.----T.TQT..LAD.KA.RRYI..AK..CV..VTQELSTFDP-
Platypus DCH4   --ETPTKK.S.SVSRVF-----SEDSPGEEDHPT.L.NIRG.Y.KELS.S..EN.SHP.PRTLYSSGPVVAT.NEAFS-----TYSLLRVGPG.R.GAY..V.M..STDRPTRSNKV.F--
Platypus DCH5   -DWALSRA.E.SITGH.-----ACNAPGAE.-SVFR.SASN.F.AGA.IL..VK.Q.QPGLTQ.-ITRSR..RYS.S-----LE.SVPNAQ.NQLAE....S.AGATQQKN.SKCA---
Platypus DCH8   --AMLNHSLILDF.E.ARPLVTAFHEPS.SSREK.V..ATD.R.KDIDMK.EVR.Q.RRCSDSLPSK.LGN.M.QQS-----CA.LL.LEE.TQPE.Y..V.N.SSDTAAVRRELT.T--
Platypus DCH9   --ECPAWS.TA.T...S-----FEWLLQNE.-.V.T.VTDM.N---AS.E.TAGEELAALETVE--NKKEKFPNGSS--WLF.Q.KVNF.V.NSTES.S.Q.KDGQSELRVNVTHIP---
Platypus DCH10  -----------.Y.F...----TEEM.R.PS-LT.V..IKD.Y.A.IL..QEGSAV.ETQ.AQ--SQNPSCDHQS.RCSAV.K.E.PK.R.MARTRYH.Q.F.ISRRGILIRNA.THY-
```

```
                    A                      B                 C                 D              E                         F              G
Platypus DCH6   VCQDSTPQPTLYLLPPPLEGLVLRAEALLTCLVVGVYQLDQATLTWEVDKKPLAGSPSVRALVTHDNGTQSLRSHLNLSRTDWEAAHTVLCRVSHRCLFAPQEETLQPLR
Human    DCH2   -..PSH.QPLGV...T.AVQD.W..DK.TF..F...SD.KD.H....AG.VPT.GVEEGL.ER.S..S..QH.R.T.P.SL.N.GTS.T.TLN.PS.P-..RLMALREP
Horse    DCH2   E..SH.HP.SI...H...Q..W.KG..TF......DD.KD.H.S..LSERSNGMFVESGP.EK.T..S..RS.R.A.P.SS.AMGTS.T.KL.YPN.LSSM.VVGLKEH
Panda    DCH2   D.WNY.HP.S..V.R...Q.PWFLG..SF...AL.DD.QE.R.S...AGA.PS.AVEEGS.EE.V..S...S.R.A.PVSL.ASGASIT.TL..PSRPLQLVSAAPREY
Cow      DCH2   -..NH.QA.RVH....TPQ..W.LDK.EF..I.AT.EAPLD.HFS...YGQ.HG.ALEEG-PTR.I.SSW.QS.R.A.P.SL.ASGSN.T.TL.SPG.QS.VTL.A.REH
```

```
                    A                      B                  C                 D                E                     F              G
Platypus DCH7   GADPPRKLSIHVLHPSLPQSSAEASAWLLCEVSNFSPAEIFLTWWENETSVSPSWFTTTAPFAQPGNAMYSTQSLLRIPASQGHMASTYTCMVGHVSSSPSMLNISRNGVF-
Human    DCH3   -.QA.V...LNL.AS.D---PP..AS...,...G...PN.L.M.L.DQRE.NT.G.APAR.PP..RSTTFWAW.V..V..PPSPQPA....V.S.ED.RTL..A..SLEVS
Horse    DCH3   A.SA...S.TV.A.TTPGLNA.PG.TS...Q.K..G...P..V..L.GQRE.D....A.AR.T.....TTFQ.W.I.LV.TIP.PPTA...V..EA.RQL.T.WSLDTG
Panda    DCH3   A.TA.SS.AV...TMP-----P-.AS...,...G...PD.L..L.RGHAE.DATG.A.AS.V.....TTFQ.W.V..VL.A..PLPA....V.R.DA.RKV..T.WSLDT-
Mouse    DCH3   -.MA.SN.TVNI.TT.T---HP.M.S.......G.F.EN.H.M.LGVHSKMKSTN.V.AN.T...-GGTFQ.W.V..L.VALSSSLD....V.E.EA.RTK..A.KSLAIS
Cow      DCH3   A.SV.GN.TLRTVTAPG----PFSP....,...L.GQQE.E..Q.A.AHTT..A.R.SSH.W.V..VSSPLD.AGA....V.S.EA.RTL..G.CSLDTG
```

FIGURE 3. Sequence alignment of the platypus IgD with its counterparts in placental mammals. Putative IgSF strands are indicated as A-G with lines over the sequences. Dots are used to denote identical amino acids, whereas dashes are used to adjust the sequence alignment. Canonical cysteines and tryptophans are shaded. The alignment was performed using *ClustalW* with some manual adjustments.

made using MrBayes 3.1.2 (29) and viewed in TREEVIEW (30). Multiple sequence alignments were performed using *Clustalw*. Protein disordered regions were predicted using a series of online programs (http://www.disprot.org/predictors.php).

Results

Identification of the genomic sequences encoding IgM (μ), IgG1/G2 ($\gamma 1/\gamma 2$), IgA1/A2 ($\alpha 1/\alpha 2$), and IgE (ε) in the platypus

The platypus has previously been shown to express four Ig classes, encoded by six genes: IgM (μ), IgG1/G2 ($\gamma 1/\gamma 2$), IgA1/A2 ($\alpha 1/\alpha 2$), and IgE (ε) at the cDNA level (27, 28). Using these cDNA sequences as templates, we performed BLAST searches against the platypus genomic sequences deposited in the NCBI GenBank. This allowed us to identify a 680-kb genomic contig (NW_001794226) containing those previously known μ, $\gamma 1$, ε, $\alpha 2$. The $\gamma 2$ and $\alpha 1$ genes were found on two other short contigs (NW_001705893.1, NW_001777603.1, the latter contig contains only a part of the $\alpha 1$ genomic sequence), which are not assembled into the 680-kb contig in the current version of the platypus genome assembly. Alignments of these identified genomic sequences with their respective cDNA revealed that all genes had a similar genomic organization as their corresponding genes in eutherian mammals (Fig. 1).

Identification of the platypus IgD encoding gene

There is an ~90-kb of DNA sequence (including sequence gaps with unknown sizes) between the μ and $\gamma 1$ genes in the 680-kb contig, which is of sufficient size to accommodate additional Ig isotype encoding genes. Approximately 4.3-kb downstream of the μTM2 exon, an Ig C_H domain encoding exon was identified. Further downstream, nine additional Ig C_H domain-encoding exons and two typical IgM/D transmembrane exons were also observed, indicating the presence of a 10 C_H domain-encoding Ig gene (Fig. 1). Blast searches using the deduced amino acid sequence revealed that it shared a relatively high degree of homology with IgD from fish, *Xenopus*, leopard gecko, and some mammals. Together with

phylogenetic analyses (Fig. 2, supplementary Fig. 1),[6] these data strongly suggest that the identified gene is the platypus δ.

We failed to amplify the δ-encoding cDNA from a platypus cDNA library (28), probably due to its low rate of expression. We thus resorted to a nested RT-PCR directly using total RNA as a template and primers derived from J_H and the predicted TM1. An ~3.3 kb PCR product was successfully amplified and directly sequenced, and was shown to contain all the ten predicted C_H exons (supplementary Fig. 2). This clearly suggests that the platypus IgD can be expressed as a membrane-bound form consisting of multiple C_H domains. However, we did not succeed in cloning a secreted form encoding cDNA. Like in most other mammals, the platypus δ gene is most likely expressed through alternative splicing of a long primary transcript which consists of both the δ and μ genes, as no putative switch region sequence was identified in the μ-δ intron (data not shown) (31).

Despite its apparent absence in birds (chickens and ducks), IgD has recently been identified in some reptiles (Ref. 6, and our unpublished data) showing a roughly similar size (11 C_H domains) as the platypus IgD. We subsequently performed a domain to domain sequence comparison of IgD between the platypus and two reptiles (leopard gecko and the green anole lizard), that revealed that nearly all the 10 platypus IgD C_H domains have a corresponding homologous C_H domain in both reptile IgD molecules (supplementary Table 1), supporting the view that these molecules are related.

A domain-to-domain sequence comparison revealed that the $C_H 1$, $C_H 6$, and $C_H 7$ domains of the platypus IgD are homologous to the $C_H 1$, $C_H 2$, and $C_H 3$ of IgD of eutherian mammals, respectively (supplementary Table 2 and Fig. 3), indicating that this molecule was shortened by a selective loss of C_H domains during

---

no hinge region (Fig. 3), indicating that the IgD hinge developed after divergence of therian mammals from monotremes.

Identification of a novel Ig gene in the platypus

In addition to the above mentioned five genes including the μ, δ, γ, ε, and α, a sixth H chain isotype encoding gene was unexpectedly found downstream of the δ. This gene consists of four C_H and two TM encoding exons and is structurally similar to μ and ε. However, it has only 29.4 and 45.1% amino acid identity to the platypus IgM and IgE, respectively. We termed this gene o (o for *Ornithorhynchus anatinus,* encoding IgO) (Fig. 1). RT-PCR shows that IgO is exclusively expressed in the spleen (Fig. 4). We cloned the IgO C region encoding cDNA using a nested PCR-amplification of a duck-billed platypus spleen cDNA library (Fig. 5). Surprisingly, alignment of the IgO C region with that of the platypus IgE revealed an extended C_H2 domain at the N-terminal (Fig. 6). This extended sequence is abundant in prolines and very similar, in amino acid composition, to hinges of mammalian IgA and IgG (supplementary Fig. 3). Proline rich regions usually play a structural role in proteins as spacers and are normally devoid of a secondary structure, i.e., being disordered or forming a random coil. Definite Ig hinges usually display an extended (at least partially), disordered protein structure, conferring structural flexibility (32). A series of prediction programs suggested that the extended peptide in the IgO C_H2 region is structurally disordered, as exemplified in supplementary Fig. 4. Therefore, it is likely to form a hinge unit together with the first cysteine residue in the C_H2 domain. An Ig class with four C_H domains plus a hinge has thus far not been
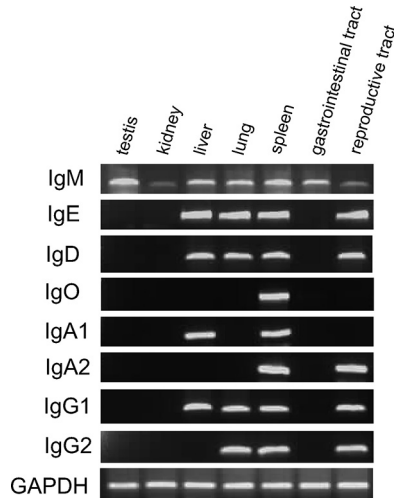

FIGURE 4. RT-PCR detection of the platypus Ig gene expression in different tissues. We were not able to conduct Northern blotting to detect tissue expression of the platypus Ig genes due to unavailability of high quality RNA, as all platypus tissues were collected opportunistically from animals found dead in the wild. Tissues of gastrointestinal tract, liver, spleen, and reproductive tract were collected from one male animal, and testis from another male, whereas the lung and kidney were from a female animal.
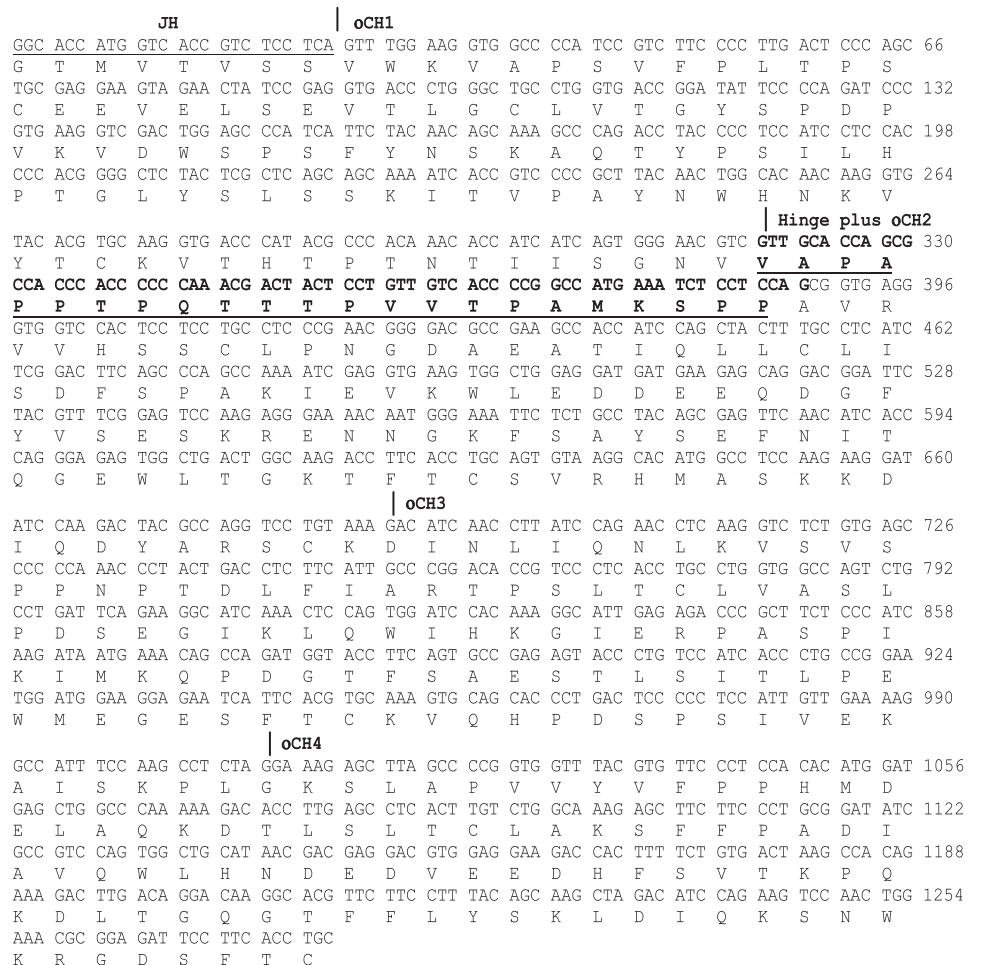
evolution. IgD in eutherian mammals usually contains a structurally flexible hinge in the H chain constant region (9, 12, 22). However, a detailed examination of the platypus IgD sequence reveals

FIGURE 5. Sequence of the platypus IgO constant region cDNA. The putative hinge region is in bold and underlined.

```
                        JH                   |  oCH1
GGC ACC ATG GTC ACC GTC TCC TCA GTT TGG AAG GTG GCC CCA TCC GTC TTC CCC TTG ACT CCC AGC  66
 G   T   M   V   T   V   S   S   V   W   K   V   A   P   S   V   F   P   L   T   P   S
TGC GAG GAA GTA GAA CTA TCC GAG GTG ACC CTG GGC TGC CTG GTG ACC GGA TAT TCC CCA GAT CCC 132
 C   E   E   V   E   L   S   E   V   T   L   G   C   L   V   T   G   Y   S   P   D   P
GTG AAG GTC GAC TGG AGC CCA TCA TTC TAC AAC AGC AAA GCC CAG ACC TAC CCC TCC ATC CTC CAC 198
 V   K   V   D   W   S   P   S   F   Y   N   S   K   A   Q   T   Y   P   S   I   L   H
CCC ACG GGG CTC TAC TCG CTC AGC AGC AAA ATC ACC GTC CCC GCT TAC AAC TGG CAC AAC AAG GTG 264
 P   T   G   L   Y   S   L   S   S   K   I   T   V   P   A   Y   N   W   H   N   K   V

                                                                   |   Hinge plus oCH2
TAC ACG TGC AAG GTG ACC CAT ACG CCC ACA AAC ACC ATC ATC AGT GGG AAC GTC GTT GCA CCA GCG 330
 Y   T   C   K   V   T   H   T   P   T   N   T   I   I   S   G   N   V   V   A   P   A
CCA CCC ACC CCC CAA ACG ACT ACT CCT GTT GTC ACC CCG GCC ATG AAA TCT CCT CCA GCG GTG AGG 396
 P   P   T   P   Q   T   T   T   P   V   V   T   P   A   M   K   S   P   P   A   V   R
GTG GTC CAC TCC TCC TGC CTC CCG AAC GGG GAC GCC GAA GCC ACC ATC CAG CTA CTT TGC CTC ATC 462
 V   V   H   S   S   C   L   P   N   G   D   A   E   A   T   I   Q   L   L   C   L   I
TCG GAC TTC AGC CCA GCC AAA ATC GAG GTG AAG TGG CTG GAG GAT GAT GAA GAG CAG GAC GGA TTC 528
 S   D   F   S   P   A   K   I   E   V   K   W   L   E   D   D   E   E   Q   D   G   F
TAC GTT TCG GAG TCC AAG AGG GAA AAC AAT GGG AAA TTC TCT GCC TAC AGC GAG TTC AAC ATC ACC 594
 Y   V   S   E   S   K   R   E   N   N   G   K   F   S   A   Y   S   E   F   N   I   T
CAG GGA GAG TGG CTG ACT GGC AAG ACC TTC ACC TGC AGT GTA AGG CAC ATG GCC TCC AAG AAG GAT 660
 Q   G   E   W   L   T   G   K   T   F   T   C   S   V   R   H   M   A   S   K   K   D
                          |  oCH3
ATC CAA GAC TAC GCC AGG TCC TGT AAA GAC ATC AAC CTT ATC CAG AAC CTC AAG GTC TCT GTG AGC 726
 I   Q   D   Y   A   R   S   C   K   D   I   N   L   I   Q   N   L   K   V   S   V   S
CCC CCA AAC CCT ACT GAC CTC TTC ATT GCC CGG ACA CCG TCC CTC ACC TGC CTG GTG GCC AGT CTG 792
 P   P   N   P   T   D   L   F   I   A   R   T   P   S   L   T   C   L   V   A   S   L
CCT GAT TCA GAA GGC ATC AAA CTC CAG TGG ATC CAC AAA GGC ATT GAG AGA CCC GCT TCT CCC ATC 858
 P   D   S   E   G   I   K   L   Q   W   I   H   K   G   I   E   R   P   A   S   P   I
AAG ATA ATG AAA CAG CCA GAT GGT ACC TTC AGT GCG GAG AGT ACC CTG TCC ATC ACC CTG CCG GAA 924
 K   I   M   K   Q   P   D   G   T   F   S   A   E   S   T   L   S   I   T   L   P   E
TGG ATG GAA GGA GAA TCA TTC ACG TGC AAA GTG CAG CAC CCT GAC TCC CCC TCC ATT GTT GAA AAG 990
 W   M   E   G   E   S   F   T   C   K   V   Q   H   P   D   S   P   S   I   V   E   K
               |  oCH4
GCC ATT TCC AAG CCT CTA GGA AAG AGC TTA GCC CCG GTG GTT TAC GTG TTC CCT CCA CAC ATG GAT 1056
 A   I   S   K   P   L   G   K   S   L   A   P   V   V   Y   V   F   P   P   H   M   D
GAG CTG GCC CAA AAA GAC ACC TTG AGC CTC ACT TGT CTG GCA AAG AGC TTC TTC CCT GCG GAT ATC 1122
 E   L   A   Q   K   D   T   L   S   L   T   C   L   A   K   S   F   F   P   A   D   I
GCC GTC CAG TGG CTG CAT AAC GAC GAG GTG GAG GAA GAC CAC TTT TCT GTG ACT AAG CCA CAG CAG 1188
 A   V   Q   W   L   H   N   D   E   D   V   E   E   D   H   F   S   V   T   K   P   Q
AAA GAC TTG ACA GGA CAA GGC ACG TTC TTC CTT TAC AGC AAG CTA GAC ATC CAG AAG TCC AAC TGG 1254
 K   D   L   T   G   Q   G   T   F   F   L   Y   S   K   L   D   I   Q   K   S   N   W
AAA CGC GGA GAT TCC TTC ACC TGC
 K   R   G   D   S   F   T   C
```
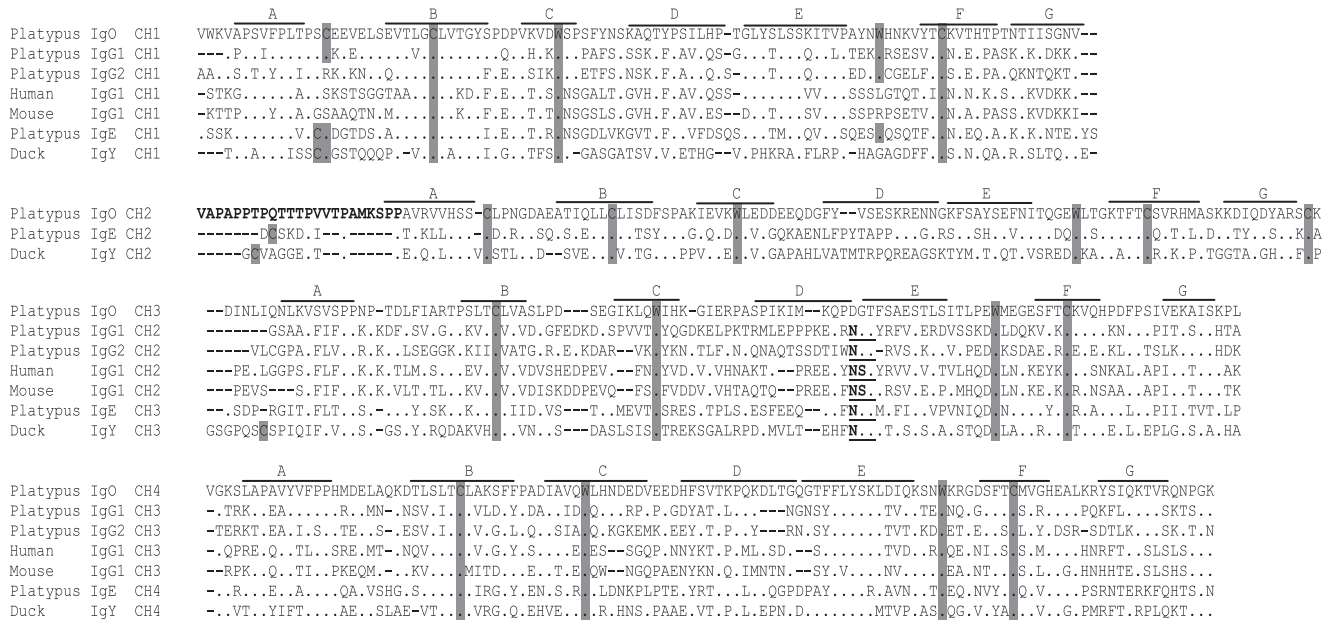
```
                            A              B              C              D              E              F              G
Platypus IgO  CH1  VWKVAPSVFPLTPSCEEVELSEVTLGCLVTGYSPDPVKVDWSPSFYNSKAQTYPSILHP-TGLYSLSSKITVPAYNWHNKVYTCKVTHTPTNTIISGNV--
Platypus IgG1 CH1  ---.P..I......K.E.......V........Q..H.K...PAFS.SSK.F.AV.QS-G...T...Q...L.TEK.RSESV..N.E.PASK.K.DKK.--
Platypus IgG2 CH1  AA..S.T.Y..I..RK.KN..Q........F.E..SIK...ETFS.NSK.F.A..Q.S--..T...Q...ED..CGELF..S.E.PA.QKNTQKT.--
Human     IgG1 CH1  -STKG......A..SKSTSGGTAA.......KD.F.E..T.SLNSGALT.GVH.F.AV.QSS-......VV...SSSLGTQT.I.N.K.S..KVDKK.--
Mouse     IgG1 CH1  -KTTP...Y..A.GSAAQTN.M.......K..F.E..T.TSNSGSLS.GVH.F.AV.ES--D..T...SV...SSPRPSETV..N.A.PASS.KVDKKI.--
Platypus IgE  CH1  .SSK.......V.CDGTDS.A.........I.E..T.R.NSGDLVKGVF..VFDSQS...TM..QV..SQES.QSQTF..N.EQ.A.K.K.NTE.YS
Duck      IgY  CH1  ---T..A...ISSC.GSTQQQP.-V..A...I.G..TFS.-GASGATSV.V.ETHG--V.PHKRA.FLRP.-HAGAGDFF..S.N.QA.R.SLTQ..E-

                            A              B              C              D              E              F              G
Platypus IgO  CH2  VAPAPPTPQTTTPVVTPAMKSPPAVRVVHSS-CLPNGDAEATIQLLCLISDFSPAKIEVKWLEDDEEQDGFY--VSESKRENNGKFSAYSEFNITQGEWLTGKTFTCSVRHMASKKDIQDYARSCK
Platypus IgE  CH2  -------DCSKD.I--.-----.T.KLL....C.D.R..SQ.S.E....TSY...G.Q.DC.V.GQKAENLFPYTAPP...G.RS..SH..V...DQ..S......Q.T.L.D..TY..S..K.A
Duck      IgY  CH2  -----GCVAGGE.T--.-----.E.Q.L...VCSTL..D--SVE...V.TG...PPV..E..V.GAPAHLVATMTRPQREAGSKTYM.T.QT.VSRED.KA..A...R.K.P.TGGTA.GH..F.P

                            A              B              C              D              E              F              G
Platypus IgO  CH3  --DINLIQNLKVSVSPPNP-TDLFIARTPSLTCLVASLPD---SEGIKLQWIHK-GIERPASPIKIM--KQPDGTFSAESTLSITLPEWMEGESFTCKVQHPDFPSIVEKAISKPL
Platypus IgG1 CH2  -------GSAA.FIF..K.KDF.SV.G..KV..V.VD.GFEDKD.SPVVT.YQGDKELPKTRMLEPPPKE.RN..YRFV.ERDVSSKD.LDQKV.K.....KN...PIT.S..HTA
Platypus IgG2 CH2  -----VLCGPA.FLV..R.K..LSEGGK.KII.VATG.R.E.KDAR--VK.YKN.TLF.N.QNAQTSSDTIWN..-RVS.K..V.PED.KSDAE.R.E.E.KL..TSLK....HDK
Human     IgG1 CH2  ---PE.LGGPS.FLF..K.K.TLM.S...EV..V.VDVSHEDPEV.--FN.YVD.V.VHNAKT.--PREE.YNS.YRVV.V.TVLHQD.LN.KEYK...SNKAL.API..T...AK
Mouse     IgG1 CH2  ---PEVS---S.FIF..K.K.VLT.TL..KV..V.VDISKDDPEVQ--FS.FVDDV.VHTAQTQ--PREE.FNS.RSV.E.P.MHQD.LN.KE.K.R.NSAA..API..T...TK
Platypus IgE  CH3  -..SDP-RGIT.FLT..S--...Y.SK..K....IID.VS---T..MEVT.SRES.TPLS.ESFEEQ--..FN..M.FI..VPVNIQD.N....Y..R.A...L..PII.TVT.LP
Duck      IgY  CH3  GSGPQS.CSPIQIF.V..S.-GS.Y.RQDAKVH...VN..S---DASLSIS.TREKSGALRPD.MVLT--EHF.N..T.S.S.A.STQD.LA..R...T...E.L.EPLG.S.A.HA

                            A              B              C              D              E              F              G
Platypus IgO  CH4  VGKSLAPAVYVFPPHMDELAQKDTLSLTCLAKSFFPADIAVQWLHNDEDVEEDHFSVTKPQKDLTGQGTFFLYSKLDIQKSNWKRGDSFTCMVGHEALKRYSIQKTVRQNPGK
Platypus IgG1 CH3  -..TRK..EA......R..MN--.NSV.I...VLD.Y.DA..ID..Q...RP.P.GDYAT.L...----NGNSY......TV..TE.NQ.G....S.R....PQKFL....SKTS..
Platypus IgG2 CH3  -TERKT.EA.I.S..TE..S-.ESV.I...V.G.L.Q..SIA..Q.KGKEMK.EEY.T.P..Y----RN.SY......TVT.KD.ET.E...S.L.Y.DSR-SDTLK...SK.T.N
Human     IgG1 CH3  -.QPRE.Q..TL..SRE.MT-.NQV....V.G.Y.S....E.ES--SGQP.NNYKT.P.ML.SD.--S.......TVD..R.QE.NI.S.S.M...HNRFT..SLSLS...
Mouse     IgG1 CH3  --RPK..Q..TI..PKEQM.-..KV...MITD...E..T.E.QW--NGQPAENYKN.Q.IMNTN.---SY.V....NV....EA.NT...S.L..G.HNHHTE.SLSHS...
Platypus IgE  CH4  -..R...E..A.....QA.VSHG.S....IRG.Y.EN.S.R.LDNKPLPTE.YRT....L..QGPDPAY....R.AVN..T.EQ.NVY..Q.V...PSRNTERKFQHTS.N
Duck      IgY  CH4  -..VT..YIFT....AE..SLAE-VT...VRG.Q.EHVE...R.HNS.PAAE.VT.P.L.EPN.D........MTVP.AS.QG.V.YA...V..G.PMRFT.RPLQKT...
```

FIGURE 6. Sequence alignment of the platypus IgO with other Ig classes. Putative IgSF strands are indicated as A–G with lines over the sequences. Dots are used to denote identical amino acids, whereas dashes are used to adjust the sequence alignment. Canonical cysteines and tryptophans are shaded, while noncanonical cysteines are also shaded. The alignment was performed using *ClustalW* with some manual adjustments. The putative IgO hinge is in bold, while the conserved N-linked glycosylation site is both in bold and underlined.

observed in any species, suggesting that IgO represents a distinct Ig class in mammals.

IgO, a structural intermediate between IgY(E) and IgG?

To determine the phylogenetic position of IgO, we performed thorough phylogenetic analyses using its separate C_H domains or the entire constant region sequence. All analyses showed that IgO is phylogenetically related to IgG, IgE, and IgY (relatively closer to IgG) (Fig. 2 and 7, supplementary Fig. 1). Despite the close phylogenetic relationship to IgG, two other lines of evidence, in addition to the obvious structural difference ($4C_H$ vs $3C_H$), support the notion that the IgO is not a subclass of IgG, but a distinct Ig H chain isotype. All mammalian IgG molecules contain a functionally important N-linked glycosylation site (Asn 297) in the C_H2 domain. This site is conserved in homologous regions of IgM, IgE, and IgD (12, 33, 34). This site is absent in the corresponding position of the IgO constant region (Fig. 6). Moreover, the IgO hinge region lacks cysteines that are often seen in IgG hinges and used to covalently hold heavy chains. Taken together, these analyses support the notion that likely, IgO represents a structural intermediate between IgY (E) and IgG. Phylogenetic analysis indicates that the C_H1, C_H2, and C_H3 domains of IgG correspond to the C_H1, C_H3, and C_H4 of IgO or IgY(E), respectively (supplementary Fig. 5). The finding of a hinge region in the C_H2 of IgO suggests that, in the evolutionary route from a four C_H domain Ig to a three C_H domain IgG, a hinge region was first developed at the N-terminal of the C_H2 (of an IgY like Ig) domain and afterward, the C_H2 (of an IgY like Ig) was lost. This is consistent with the hypothesis proposed by Blattner et al. (24) that the Ig hinge region may have evolved by duplication, leading to incorporation of a 3' RNA splice site into the coding portion of a C_H domain, and later, mutations creating a 5' RNA splice site at the junction of the hinge region and the C_H domain, caused detachment of the hinge exon and nonfunctionality of the C_H domain involved (24). Examination of the putative IgO hinge encoding sequence did indeed reveal features of splice sites, where the 3' end of the hinge encoding sequence is pyrimidine rich and terminated by a conserved AG

di-nucleotide, exhibiting features of a 3' splice site (supplementary Fig. 6). Immediately downstream of the hinge sequence, two cryptic 5' splice sites (GTGAG, GTGGT) can be identified (supplementary Fig. 6). To test whether these two 5' splice sites are functional, we designed two sense primers by fusing nucleotides upstream of the two splice sites respectively, with the first five nucleotides of the C_H3 exon. These two primers were used in nested RT-PCR amplifications (data not shown), and showed that the second cryptic 5' splice site (GTGGT) was functional (albeit less efficient) in RNA splicing based on sequencing of the PCR products. Thus, the hinge sequence can be directly spliced onto the C_H3 exon, generating a transcript (VDJ-C_H1-hinge-C_H3-C_H4) that structurally resembles an IgG H chain (supplementary Fig. 7). However, this alternatively spliced variant appears to be nonfunctional due to a shift of the open reading frame (supplementary Fig. 8). Nevertheless, it provides data to support the above-mentioned model, explaining how a genetic hinge (encoded by a separate exon) could be developed.

Localization of the γ2 and α1 genes in the platypus immunoglobulin heavy chain constant region (IGHC) gene locus

Although the genomic sequences of the platypus γ2 and α1 could be found in the genome database, they were not assembled in the main Ig gene locus (680-kb contig). This could be due to the presence of sequence gaps within the genomic contig, as in nearly all tetrapods so far investigated, all the IGHC genes are arranged in a single locus in tandem. To determine the localizations of the γ2 and α1 genes, we used several BAC clones that covered the entire IGHC gene locus. Based on BLAST searches against the platypus BAC end sequence database, BAC clone CH236-821N20 was shown to contain the μ, δ, ο, and γ1 genes. Clone CH236-99D18 was physically linked to the 3' end of CH821N20, while CH236-647E9 was located ~12-kb downstream (Note: there is a long gap with an unknown size in between) (Fig. 1). PCR experiments showed that CH236-99D18, but not other two clones, contained the α1 gene. As there is no α1 gene in the long determined sequence downstream of the α2 in both CH236-99D18 and
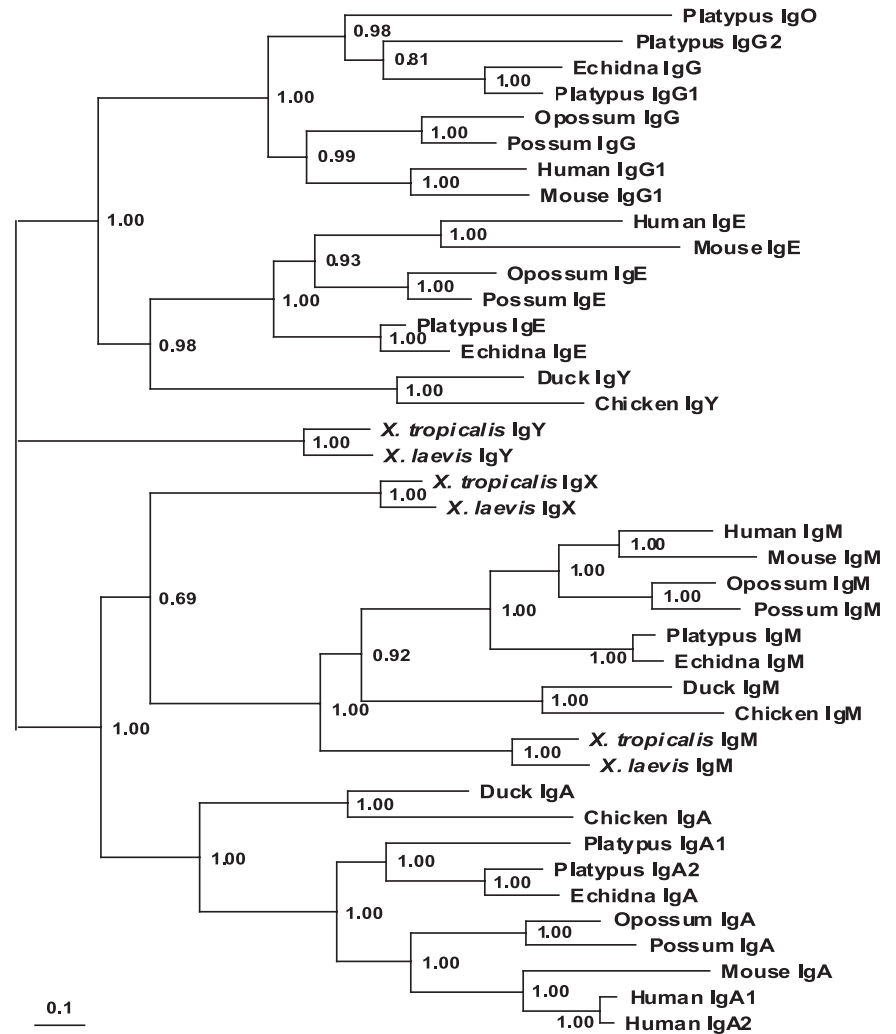
FIGURE 7. Phylogenetic analysis of the IgO using the entire constant region sequence. The scale shown as a bar represents the genetic distance. The credibility value for each node is shown. For accession numbers of the sequences used in the tree see Fig. 2 legend.

CH236-647E9, the most plausible location for α1 is in the gap between CH236-821N20 and CH236-647E9 (Fig. 1). This was confirmed by sequencing of PCR products amplified using primers derived from the α1 and the sequences flanking the gap. The genomic organization of the α1 was deduced by PCR and sequencing.

According to our PCR results, CH236-821N20 was shown to contain the γ2 gene. To determine the location of the γ2 gene, a couple of primers were designed covering all sequence gaps from the δ to the γ1 gene, and used together with the primers derived from γ2 in PCR amplifications using CH236-821N20 BAC DNA as a template. By sequencing of the obtained PCR products, it was deduced that the γ2 gene was located in the first gap downstream of the o gene (Fig. 1). These data allowed us to conclude that the platypus IGHC genes are arranged in an μ-δ-o-γ2-γ1-α1-ε-α2 order (Fig. 1).

Supporting functionality of the identified platypus IGHC locus, a J$_H$ (Ig H chain joining gene segment) locus was identified ~10-kb upstream of the μ gene. The J$_H$ locus spans ~4.2-kb DNA and contains 10 structurally functional J$_H$ gene segments and 1 pseudo J$_H$ (Fig. 8). The D$_H$ and V$_H$ gene loci could not be analyzed as there are too many sequence gaps upstream of the J$_H$ locus.

Discussion

In the present study, we identified two Ig H chain isotypes in the duckbilled platypus, IgD and IgO, in addition to the previously known IgM, IgG1/G2, IgE, and IgA1/A2 classes, showing that the platypus expresses eight Ig H chain constant region genes, arranged in an μ-δ-o-γ2-γ1-α1-ε-α2 order (Fig. 1). This also demonstrates that all Ig H chain isotypes expressed by eutherian mammals are present in monotremes even though these two mammalian lineages diverged ~166 million years ago (26).

Although it has been known for some years that monotremes express typical mammalian IgM, IgG, IgA, and IgE (27, 28, 35–38), the presence of an ortholog of IgD has remained elusive. We have previously tried, but failed, to clone the δ gene in the platypus using a PCR-based approach and degenerate primers covering the conserved δ gene sequence of eutherian mammals, suggesting that if it was present, it should be distinct from that of its counterparts in eutherian mammals. In this study, we show that the platypus contains a δ gene strikingly different from the δ gene in eutherian mammals. It does, however, share great similarity to the δ gene in nonmammalian vertebrates such as reptiles, amphibians, and fish. In eutherian mammals, IgD typically contains three C$_H$ domains and a hinge segment (9), which is much shorter than its counterparts in nonmammalian vertebrates, which usually have eight or more CH domains but no hinge (3, 4, 8). The platypus δ gene encodes ten C$_H$ domains, representing the first long form of IgD molecule identified in mammals. The discovery of such an IgD gene has evolutionary implications as: 1) it establishes a clear relationship between the short IgD molecules in eutherian mammals and the long IgD molecules in nonmammalian vertebrates and 2)

```
        nonamer            23 bp spacer       heptamer
JH1   GGATTTTTGcaggagaggagtggagagttgatCTCTGTGgTACTATTACTTGGAGCACTGGGGCCAAGGCACCATGGTCACCGTCTCCTCAGgtac
                                              Y  Y  Y  L  E  H  W  G  Q  G  T  M  V  T  V  S  S

JHp   GGTTTTTGTgccttggtgggggtctat        CACTGTGTACTGGGGCCAAGGCACCATGGTCACCGTCTCCTCAGgtaa
                                         Y  W  G  Q  G  T  M  V  T  V  S  S

JH2   GGTATTTGTgcctcagaaacctgggggtctgtCACTGTGacTATAACGGCTTCAAGAACTGGGGCCAAGGCACCATGGTCACCGTCTCTTCAGgtaa
                                             Y  N  G  F  K  N  W  G  Q  G  T  M  V  T  V  S  S

JH3   AGTTTTTGTgcctcggaaacaaggggctctgtCATTGTGacTGCTATGCTTTCGATAACTGGGGCCAAGGCACCATGGTCACCGTCTCCTCAGgtaa
                                            C  Y  A  F  D  N  W  G  Q  G  T  M  V  T  V  S  S

JH4   AATTTTTGTccctcggaaacaagggggctctgtCAATGTGacTACTGTGCTCTCGAGAACTGGGGCCAAGGCACCATGGTCACCGTCTCCTCAGgtaa
                                             Y  C  A  L  E  N  W  G  Q  G  T  M  V  T  V  S  S

JH5   GCTTTTTGTgtcccagaaacattggggtttgaCAATGTGacTATGGTATCTTAGAGAACTGGGGCCAAGGCACCATGGTCACCGTTTCCTCAGgtaa
                                            Y  G  I  L  E  N  W  G  Q  G  T  M  V  T  V  S  S

JH6   GGTTTTTGTagctcagaaacaagggggctctgtCAGTGTGacTATGGATACTTAAATCACTGGGGCCAAGGCACCATGGTCACCATCTCCTCAGgtaa
                                             Y  G  Y  L  N  H  W  G  Q  G  T  M  V  T  I  S  S

JH7   GGTTTGTGTgtagtcttcagtgggaactgtgcCCCAGTGGcAATGCTTTTGGGTATTGGGGCCAAGGGACTATGGTGACAGTCTCCTCAGgtaa
                                            N  A  F  G  Y  W  G  Q  G  T  M  V  T  V  S  S

JH8   GGTTTTTGTatgccgggtgaaggcatctgagcCAACGTGatTACTTTGCTAACTGGGGCCAAGGGACCTTCGTGACCGTCTCCTCAGgtaa
                                            Y  F  A  N  W  G  Q  G  T  F  V  T  V  S  S

JH9   AGTTACTGCcaaggtcccattgtgactcgcccCATTGTGacAACCGGTTAGAGCACTGGGGCCAAGGCACCATGGTGACTGTCTCCTCAGgtaa
                                            N  R  L  E  H  W  G  Q  G  T  M  V  T  V  S  S

JH10  TGGTTTTTGtaagacagaatgatgaatttatcCACTGTGacTATGGTATGGATGCATGGGGCCAAGGAACCATGGTGACCGTCTCCTCAGgtaa
                                             Y  G  M  D  A  W  G  Q  G  T  M  V  T  V  S  S
```

FIGURE 8. Sequences of the platypus J_H genes. Spacer sequences and 3′ splice sites are shown in small letters, whereas the heptamer and nonamer sequences are underlined. JHp: pseudo JH.

it shows that the platypus has a unique combination of Ig genes which share similarity both to those of nonmammalian vertebrates and eutherian animals.

Sequence comparisons suggest that the platypus δ gene shares an immediate ancestor with the recently identified δ gene in a reptile, leopard gecko (6), as all its 10 C_H exons are homologous to exons found in the reptile (gecko) δ gene (containing 11 C_H encoding exons). We also recently identified the δ gene in another reptile, the green anole lizard (our unpublished data), displaying a similar structure to that of the gecko gene by having 11 C_H encoding exons. Presence of these closely related δ genes in phylogeny in both mammals and reptiles, strongly suggests that the absence of the δ gene in birds (7, 16) is due to a recent genetic modification that occurred after their divergence from reptiles.

The platypus IgD does not contain a hinge region. Sequence comparisons suggest that the δCH1, δCH2, and δCH3 of eutherian mammals share homology with the δCH1, δCH6, and δCH7 of the platypus, suggesting a selective loss of other C_H exons during the evolution. To provide insights into the timing of the loss of these domains, we searched for a δ gene in a marsupial, the gray short-tailed opossum, by using its genome sequence. Typical mammalian μ, γ, ε, and α genes are found in tandem in a single locus (data not shown). In the 200-kb of DNA between the opossum μ and γ genes (where the δ usually located), we did not find any evidence of a δ gene (or even any putative Ig encoding sequences). A δ gene was also not found anywhere in the entire opossum genome, suggesting an absence of the gene in this species. A future search for the IgD encoding gene in additional marsupials might be expected to provide more clues to this evolutionary issue.

The o gene represents another interesting finding in platypus. It encodes four Ig CH domains and a hinge attached to the N-terminal of C_H2, giving rise to an Ig H chain constant region that has thus far not been observed in vertebrates. Our data suggest that it might be an immediate evolutionary precursor of IgG. This hypothesis is based on several facts: 1) the IgO encoding gene, o, is physically located in the position where a γ gene is usually found in the mammalian IgH locus; 2) the o gene is phylogenetically related to the mammalian γ gene; and 3) the o gene encodes a

hinge segment. IgG is thought to have originated from IgY, an ancient, hinge-lacking, four CH domain Ig isotype. Compared with IgY, mammalian IgG usually has a hinge segment and one C_H domain less (a loss of CH2 of IgY). The structure of IgO suggests that the hinge was formed before the loss of a C_H domain. This hypothesis agrees with a model that explains the origin of the murine IgA hinges (24). The pivotal point in that model is that the Ig hinge region was originally derived from the noncoding 3′ end sequence of an intron that was rich in pyrimidine (especially cytidine), potentially encoding the prolines that are often seen in Ig hinge regions. This originally noncoding stretch of nucleotides was subsequently attached to the 5′ end of a C_H exon and inserted in frame with the reading frame of the C_H exons. A final loss of the C_H exon (or a gain of a separate genetic hinge exon) involved could be achieved through mutations of hinge-C_H junctional sequence generating a 5′ RNA splice site, which would result in a removal of the C_H exon in mature message RNA. Nonfunctionality of the C_H exon would finally cause it to be gradually lost, forming the genomic structure of the mammalian γ genes.

In summary, the discovery of IgO and IgD in the platypus highlights the important position this species holds in mammalian phylogeny and provides unique insights into the evolution of the mammalian Ig repertoire. It also provides support for the continuous evolution of immunoglobulins in vertebrates and the importance of comparative studies for reconstructing the evolutionary history of the immune system.

Disclosures

References

1. Flajnik, M. F. 2002. Comparative analyses of immunoglobulin genes: surprises and portents. *Nat. Rev. Immunol.* 2: 688–698.
2. Litman, G. W., M. K. Anderson, and J. P. Rast. 1999. Evolution of antigen binding receptors. *Annu. Rev. Immunol.* 17: 109–147.
3. Ohta, Y., and M. Flajnik. 2006. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proc. Natl. Acad. Sci. USA* 103: 10723–10728.
4. Zhao, Y., Q. Pan-Hammarstrom, S. Yu, N. Wertz, X. Zhang, N. Li, J. E. Butler, and L. Hammarstrom. 2006. Identification of IgF, a hinge-region-containing Ig

class, and IgD in *Xenopus tropicalis*. *Proc. Natl. Acad. Sci. USA* 103: 12087–12092.

5. Hsu, E., M. F. Flajnik, and L. Du Pasquier. 1985. A third immunoglobulin class in amphibians. *J. Immunol.* 135: 1998–2004.

6. Gambon-Deza, F., and C. S. Espinel. 2008. IgD in the reptile leopard gecko. *Mol. Immunol.* 45: 3470–3476.

7. Lundqvist, M. L., D. L. Middleton, C. Radford, G. W. Warr, and K. E. Magor. 2006. Immunoglobulins of the non-galliform birds: antibody expression and repertoire in the duck. *Dev. Comp. Immunol.* 30: 93–100.

8. Wilson, M., E. Bengten, N. W. Miller, L. W. Clem, L. Du Pasquier, and G. W. Warr. 1997. A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. *Proc. Natl. Acad. Sci. USA* 94: 4593–4597.

9. Zhao, Y., I. Kacskovics, Q. Pan, D. A. Liberles, J. Geli, S. K. Davis, H. Rabbani, and L. Hammarstrom. 2002. Artiodactyl IgD: the missing link. *J. Immunol.* 169: 4408–4416.

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

11. Wagner, B., D. C. Miller, T. L. Lear, and D. F. Antczak. 2004. The complete map of the Ig heavy chain constant gene region reveals evidence for seven IgG isotypes and for IgD in the horse. *J. Immunol.* 173: 3230–3242.

12. Zhao, Z., Y. Zhao, Q. Pan-Hammarstrom, W. Liu, Z. Liu, Y. Zhao, N. Li, and L. Hammarstrom. 2007. Physical mapping of the giant panda immunoglobulin heavy chain constant region genes. *Dev. Comp. Immunol.* 31: 1034–1049.

13. Hordvik, I., J. Thevarajan, I. Samdal, N. Bastani, and B. Krossoy. 1999. Molecular cloning and phylogenetic analysis of the Atlantic salmon immunoglobulin D gene. *Scand. J. Immunol.* 50: 202–210.

14. Ota, T., J. P. Rast, G. W. Litman, and C. T. Amemiya. 2003. 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.

15. Berstein, R. M., S. F. Schluter, S. Shen, and J. J. Marchalonis. 1996. 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.

16. Lundqvist, M. L., D. L. Middleton, S. Hazard, and G. W. Warr. 2001. 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.

17. Zhao, Y., H. Rabbani, A. Shimizu, and L. Hammarstrom. 2000. Mapping of the chicken immunoglobulin heavy-chain constant region gene locus reveals an inverted α gene upstream of a condensed upsilon gene. *Immunology* 101: 348–353.

18. Ros, F., J. Puels, N. Reichenberger, W. van Schooten, R. Buelow, and J. Platzer. 2004. Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene* 330: 49–59.

19. Stenvik, J., and T. O. Jorgensen. 2000. Immunoglobulin D (IgD) of Atlantic cod has a unique structure. *Immunogenetics* 51: 452–461.

20. Hirono, I., B. H. Nam, J. Enomoto, K. Uchino, and T. Aoki. 2003. Cloning and characterisation of a cDNA encoding Japanese flounder *Paralichthys olivaceus* IgD. *Fish Shellfish Immunol.* 15: 63–70.

21. Saha, N. R., H. Suetake, K. Kikuchi, and Y. Suzuki. 2004. Fugu immunoglobulin D: a highly unusual gene with unprecedented duplications in its constant region. *Immunogenetics* 56: 438–447.

22. Zhao, Y., Q. Pan-Hammarstrom, I. Kacskovics, and L. Hammarstrom. 2003. The porcine Ig δ gene: unique chimeric splicing of the first constant region domain in its heavy chain transcripts. *J. Immunol.* 171: 1312–1318.

23. Warr, G. W., K. E. Magor, and D. A. Higgins. 1995. IgY: clues to the origins of modern antibodies. *Immunol. Today* 16: 392–398.

24. Tucker, P. W., J. L. Slightom, and F. R. Blattner. 1981. Mouse IgA heavy chain gene sequence: implications for evolution of immunoglobulin hinge axons. *Proc. Natl. Acad. Sci. USA* 78: 7684–7688.

25. Warren, W. C., L. W. Hillier, J. A. Marshall Graves, E. Birney, C. P. Ponting, F. Grutzner, K. Belov, W. Miller, L. Clarke, A. T. Chinwalla, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453: 175–183.

26. Bininda-Emonds, O. R., M. Cardillo, K. E. Jones, R. D. Macphee, R. M. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2008. The delayed rise of present-day mammals. *Nature* 456: 274.

27. Belov, K., and L. Hellman. 2003. Immunoglobulin genetics of *Ornithorhynchus anatinus* (platypus) and *Tachyglossus aculeatus* (short-beaked echidna). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 136: 811–819.

28. Vernersson, M., M. Aveskogh, B. Munday, and L. Hellman. 2002. Evidence for an early appearance of modern post-switch immunoglobulin isotypes in mammalian evolution (II); cloning of IgE, IgG1 and IgG2 from a monotreme, the duck-billed platypus, *Ornithorhynchus anatinus*. *Eur. J. Immunol.* 32: 2145–2155.

29. Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

30. Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357–358.

31. Pan-Hammarstrom, Q., Y. Zhao, and L. Hammarstrom. 2007. Class switch recombination: a comparison between mouse and human. *Adv. Immunol.* 93: 1–61.

32. Harris, L. J., E. Skaletsky, and A. McPherson. 1998. Crystallographic structure of an intact IgG1 monoclonal antibody. *J. Mol. Biol.* 275: 861–872.

33. Wright, A., and S. L. Morrison. 1994. Effect of altered CH2-associated carbohydrate structure on the functional properties and in vivo fate of chimeric mouse-human immunoglobulin G1. *J. Exp. Med.* 180: 1087–1096.

34. Prabakaran, P., B. K. Vu, J. Gan, Y. Feng, D. S. Dimitrov, and X. Ji. 2008. Structure of an isolated unglycosylated antibody C(H)2 domain. *Acta Crystallogr. D Biol. Crystallogr.* 64: 1062–1067.

35. Vernersson, M., M. Aveskogh, and L. Hellman. 2004. Cloning of IgE from the echidna (*Tachyglossus aculeatus*) and a comparative analysis of epsilon chains from all three extant mammalian lineages. *Dev. Comp. Immunol.* 28: 61–75.

36. Belov, K., K. R. Zenger, L. Hellman, and D. W. Cooper. 2002. Echidna IgA supports mammalian unity and traditional Therian relationship. *Mamm. Genome* 13: 656–663.

37. Belov, K., L. Hellman, and D. W. Cooper. 2002. Characterisation of echidna IgM provides insights into the time of divergence of extant mammals. *Dev. Comp. Immunol.* 26: 831–839.

38. Belov, K., L. Hellman, and D. W. Cooper. 2002. Characterization of immunoglobulin gamma 1 from a monotreme, *Tachyglossus aculeatus*. *Immunogenetics* 53: 1065–1071.