Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Molecular Immunology 99 (2018) 75–81

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Genomic organization and adaptive evolution of IGHC genes in marine mammals

Lili Li, Xinghua Rong, Guiting Li, Yingying Wang, Bingyao Chen, Wenhua Ren, Guang Yang*, Shixia Xu*

Jiangsu Key Lab for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu Province, 210023, China

ARTICLE INFO

Keywords:
Marine mammals
IGHC
Genomic organization
Positive selection
Aquatic adaptation

ABSTRACT

Immunoglobulins are important elements of the adaptive immune system that bind to an immense variety of microbial antigens to neutralize infectivity and specify effector functions. In the present study, the immunoglobulin heavy chain constant region (IGHC) genes from marine mammals were identified and compared with those of their terrestrial relatives to explore their genomic organization and evolutionary characteristics. The genomic organization of marine mammal IGHC genes was shown to be conservative with other eutherian mammals. Stronger signals of positive selection on IGHC were revealed in terrestrial mammals than that in marine mammals with the branch-site model, displaying different selective pressure, which might suggest their divergent adaptations to contrasted environments.

1. Introduction

Immunoglobulins (Igs) are glycoprotein molecules that function as antibodies and are only expressed in jawed vertebrates (Flajnik and Kasahara, 2010). Typically, Igs compose two identical heavy (IgH) and two identical light chains (IgL). Both heavy and light chains can be divided into two regions, i.e., the constant region and the variable region, based on the variability of amino acid sequences. In general, the variable region recognizes antigens, whereas the constant region specifies critical effector or signaling functions, such as phagocytosis, binding Fc (fragment crystallizable) receptors and activation of complement (Schroeder and Cavacini et al., 2010). The classification of Igs is based on their heavy chain constant (IGHC) genes (Bengtén et al., 2000). To date, several IgH isotypes, including IgM, IgD, IgG, IgE, and IgA, have been identified in multiple vertebrates, and they are encoded by the IGHM, IGHD, IGHG, IGHE and IGHA genes respectively (Schroeder and Cavacini et al., 2010). Both IgM and IgD have been reported to be the most ancient IgH isotypes among jawed vertebrates, with IgM very conserved but IgD showing a high degree of structural and (presumably) functional diversity and even being lost in some species over evolutionary time (Sun et al., 2011; Ohta and Flajnik, 2006). IgG and IgE are both exclusively present in mammals (Bengtén et al., 2000; Flajnik and Kasahara, 2010), the former being the predominant serum antibody and binding receptors on phagocytic cells and the latter predominantly functioning in immunity against parasites (Maizels, 2005). Additionally, IgA was found in reptiles first, with similar functions in all animals (Flajnik and Kasahara, 2010). Participating in mucosal immunity, IgA coats pathogens to prevent proliferation and defend against local infection (Macpherson et al., 2008).

Extant marine mammals descended from terrestrial ancestors that re-entered in the aquatic environment at different times (Uhen, 2010). There are five main groups of extant marine mammals: Cetacea (dolphins, porpoises and whales), Sirenia (manatees, dugongs), Pinnipedia (walruses, sea lions and seals), polar bear (*Ursus maritimus*) and sea otter (*Enhydra lutris*) (Uhen, 2007). Cetacea and Sirenia are the only known completely aquatic mammals, and they appeared to return to the aquatic environment at approximately the same time, i.e. ~50 Mya (million years ago) (Uhen, 2010; Thewissen et al., 2007). In contrast, semi-aquatic pinnipeds, with a close relationship to Musteloidea, later transferred to water environment at ~28 Mya (Uhen, 2010; Flynn et al., 2005). Despite the independent evolutionary origins of these marine mammals groups, they have developed a series of specializations for aquatic lifestyles, such as loss of hair, thickened blubber, derivation of echolocation, and so on (Uhen, 2007).

During their transition from terrestrial environment to the sea, marine mammals must have faced dramatic changes in pathogens. Recent studies have shown that cetacean TLR4 and MHC were subjected to adaptive evolution, particularly in the lineages with dramatic habitat transition (e.g. the lineage leading to hippo + whale representing the early stage of aquatic adaptation) or rapid radiation

* Corresponding authors.
E-mail addresses: MollyLi2008@outlook.com (L. Li), 821393856@qq.com (X. Rong), 936877504@qq.com (G. Li), jane9375@163.com (Y. Wang), chby2632@163.com (B. Chen), 08162@njnu.edu.cn (W. Ren), gyang@njnu.edu.cn (G. Yang), xushixia78@163.com (S. Xu).

https://doi.org/10.1016/j.molimm.2018.04.011
Received 30 January 2018; Received in revised form 19 April 2018; Accepted 19 April 2018
Available online 26 May 2018

lineages such as extant dolphins (Shen et al., 2012; Zhang et al., 2016). However, the evolution of Ig in the aquatic adaptation of marine mammals remains poorly explored, except for a few reports about the cloning and sequencing of IGHC genes in bottlenose dolphin (*Tursiops truncatus*) (Mancia et al., 2006, 2007; Lundqvist et al., 2002) and the Florida manatee (*Trichechus manatus latirostris*) (Breaux et al., 2017) and the identification of IgM, IgG and IgA and lineage-specific IGHG gene duplication in some cetaceans (Mancia et al., 2006), and so on. Therefore, in the present study, we first retrieved IGHC genes from the genomes of ten representative marine mammals. The genomic organization of IGHC cluster was identified and positive selection on the marine mammals was detected to address whether they experienced adaptive evolution in response to aquatic environments.

2. Methods

2.1. Genomic identification of IGHC genes

We first downloaded the IGHC genes of 13 terrestrial mammals, including cow (*Bos taurus*), pig (*Sus scrofa*), horse (*Equus caballus*), dog (*Canis lupus familiaris*), giant panda (*Ailuropoda melanoleuca*), black flying fox (*Pteropus alecto*), little brown bat (*Myotis lucifugus*), human (*Homo sapiens*), rhesus monkey (*Macaca mulatta*), rat (*Rattus norvegicus*), rabbit (*Oryctolagus cuniculus*), opossum (*Monodelphis domestica*) and platypus (*Ornithorhynchus anatinus*) from the National Center for Biotechnology Information (NCBI), Ensemble databases and the International ImMunoGeneTics information system (IMGT). The accession numbers of these sequences are compiled in Supplemental Table 1. Then we used IGHC sequences of these terrestrial relatives as a query to search for the orthologous genes from 10 marine mammal genomes with TBLASTN and BLASTN approaches, including bottlenose dolphin (*Tursiops truncatus*), killer whale (*Orcinus orca*), Yangtze finless porpoise (*Neophocuena asiaeorientalis*), baiji (*Lipotes vexillifer*), sperm whale (*Physeter catodon*), minke whale (*Balaena acutorostrata*), bowhead whale (*Balaena mysticetus*), Weddell seal (*Leptonychotes weddellii*), Pacific walrus (*Odobenus rosmarus*), and Florida manatee (*Trichechus manatus latirostris*). In addition, IGHC genes were also scanned from the genomes of three additional terrestrial mammals, i.e., hedgehog (*Erinaceus europaeus*), European shrew (*Sorex araneus*), and elephant (*Loxodonta africana*). The detailed information about the genome assemblies used in this study is given in Table 1. Finally, we checked whether the newly identified putative IGHC genes were their best hit by a blast search against the non-redundant database from GenBank. The newly identified IGHC genes were further separated into intact genes, partial genes and pseudogenes based on the amino acid alignment and blast results. The genomic coordinates of the newly identified IGHC genes are listed in Supplemental Table 2.

2.2. Sequence alignment and phylogenetic analysis

To confirm that the regions of alignment had sufficient confidence to test for positive selection and determine phylogenetic relationships, the unaligned multiple sequence alignment (MSA) of IGHC genes translated to amino-acid sequences and further back-translated into the corresponding codon were submitted to GUIDANCE (Penn et al., 2010a, b) to obtain alignment confidence scores. GUIDANCE used a PRANK alignment program (version 140,110) and returned a colored MSA that allowed to delimit aligned regions ambiguously. Then the columns with confidence scores lower than the threshold were excluded from further analyses.

Phylogenetic trees were reconstructed using Bayesian inference (BI) in MrBayes 3.2.2 (Huelsenbeck and Ronquist, 2001) and Maximum likelihood (ML) in RAxML 8.0.26 (Stamatakis, 2006). The best-fit evolutionary model for each dataset was chosen by MrModeltest 2.3 (Nylander, 2004) with the Akaike Information Criterion (AIC). Bayesian inference analysis was run for 10^6 generations with one cold and three heated Markov chains to ensure the analysis remaining in global rather than local optima and trees were sampled every 100 generations. The first 1000 trees were discarded as the "burn-in". Because the samples from the beginning of the chain are unreliable. For ML analysis, the maximum likelihood tree was reconstructed with 1000 bootstrap replications.

2.3. Detection of positive selection

The rate of nonsynonymous (d_N) and synonymous (d_S) substitution (d_N/d_S or omega, ω) measures the direction and intensity of natural selection, with $\omega = 1$, $\omega > 1$ and $\omega < 1$ indicating neutral evolution, positive and purifying selection, respectively. To detect the impact of natural selection on IGHC genes, we performed a number of statistical tests based on calculating the ω values using the CODEML program implemented in PAML 4.7 package (Yang, 2007). The selection analyses using the gene tree reconstructed by BI and ML methods generated nearly the same results as those from the well-supported species phylogeny (Perelman et al., 2011; Zhou et al., 2011a); thus, only the latter analysis is shown here.

To detect positively selected sites on IGHC genes, we first performed the site specific models, i.e., M8a (nearly neutral, beta distribution: $0 < \omega_0 < 1$ and $\omega_1 = 1$) versus M8 (positive selection, beta distribution: $0 < \omega_0 < 1$ and $\omega_1 > 1$), as implemented in CODEML program. In addition, positively selected sites were identified using Datamonkey, which incorporated the rate of synonymous substitution, whereas the d_S was fixed in PAML. Three ML methods, i.e. the single likelihood ancestor counting (SLAC), fixed-effect likelihood (FEL), and random-effect likelihood (REL) implemented in the Datamonkey

Table 1
List of mammals genome included in this study.

Classification	Species name	Assembly	Date	Coverage and sequence technology	Contig N50 ($\times 10^3$)[9])
Cetartiodactyla	Tursiops truncatus	turTru2	Jan.2012	2.5x Sanger; 3.5 × 454 FLX; 30x Illumina HiSeq	11.8 (2.5)
	Orcinus orca	Oorc_1.1	Jan.2013	200x Illumina HiSeq	70.3 (2.4)
	Neophocuena asiaeorientalis	Unreleased			
	Lipotes vexillifer	Lipotes_vexillifer_v1	Jul.2013	115x Illumina Hiseq 2000	31.9 (2.4)
	Physeter catodon	Physeter_macrocephalus-2.0.2	Sep.2013	75x Illumina	35.3 (2.3)
	Balaena acutorostrata	BalAcu1.0	Oct.2013	92x Illumina HiSeq 2000	22.7 (2.4)
	Balaena mysticetus	http://www.bowheadwhale.org/	Jan.2015	150x Illumina HiSeq	34.8 (2.3)
Carnivora	Leptonychotes weddellii	LepWed1.0	Mar.2013	82x Illumina HiSeq	23.7 (3.2)
	Odobenus rosmarus	Oros_1.0	Jan.2013	200x Illumina	90.0 (2.4)
Chiroptera	Pteropus alecto	ASM32557v1	Dec.2013	110x Illumina HighSeq 2000	31.8 (2.0)
	Myotis lucifugus	Myoluc2.0	Sep.2010	7x Sanger	64.3 (2.0)
Eulipotyphla	Erinaceus europaeus	EriEur2.0	Sep.2012	79x Illumina Hi-Seq	21.4 (2.7)
	Sorex araneus	SorAra2.0	Aug.2012	120x Illumina Hi-Seq	22.6 (2.4)
Superorder Afrotheria	Trichechus manatus latirostris	TriManLat1.0	Jan.2012	150x Illumina HiSeq	37.8 (3.1)
	Loxodonta africana	Loxafr3.0	Jul.2009	7x Sanger; ABI	69.0 (3.2)

website (Pond and Frost, 2005a), were used to infer the positively selected sites with default settings and significance levels of 0.1, 0.1, and 50, respectively. Although SLAC, a conservative method, is suitable for large data sets, it usually underestimates the substitution rate. FEL method seems to capture the pattern of rate variation better than SLAC and REL, but it tends to be less conservative than SLAC to detect selection in data sets of intermediate size. REL approach is more powerful than SLAC and FEL, but it has the higher rates of false positives for small data sets (Pond and Frost, 2005a,b). These three predictions were conducted to estimating ω value of each site.

To further detect whether the possibility of positive selection was limited on specific lineages, we compared the free-ratio model, which estimated a separate ω ratio for each lineage, with the one-ratio model that estimated a same ω ratio for all lineages in the tree (Yang, 1998). In addition, the branch-site model were conducted to detect evidence of positive selection on individual codon along a specific lineage (Zhang et al., 2005). In branch-site model, we compared Ma (positive selection, $0 < \omega_0 < 1$, $\omega_1 = 1$ and $\omega_2 \geq 1$) with Ma0 (neutral selection, $0 < \omega_0 < 1$, $\omega_1 = 1$ and $\omega_2 = 1$) to detect positive selection. The significance of the differences between two nested models was tested using likelihood ratio tests (LRT) statistic (-2[LogLikelihood1-LogLikelihood2]) with a chi-square distribution. Bayes Empirical Bayes (BEB) approach (Yang et al., 2005) implemented in the CODEML was used to test the posterior probabilities (PPs) of positively selected sites, where sites with PPs > 0.90 were considered as candidates undergoing positive selection. Finally, to support the PAML results, we employed a complementary protein-level approach implemented in TreeSAAP (Woolley et al., 2003) that detected the significant changes in amino acid physicochemical properties.

3. Results and discussion

3.1. Genomic organization of IGHC genes in marine mammals

In the present study, we retrieved IGHC genes from the genomes of ten marine mammals (including seven cetaceans, one manatee and two pinnipeds) and compared them with those of terrestrial relatives to provide insights into the evolution of IGHC genes in marine mammals. We found that the genomic regions containing an IGHC cluster are conservative with special flanking sequences, i.e., the TMEM121 and IGH diversity gene cluster (Eguchi-Ogawa et al., 2012), which helped us obtain the complete IGHC genes in mammalian genomes. A total of 182 IGHC genes were identified in the 26 mammals in our study (Fig. 1), including 79 sequences that were newly identified in 13 species. For the 10 marine mammals, a total of 59 sequences were identified in this study.

To define the orthologous relationships among IGHC genes, we reconstructed phylogenetic trees for each gene, as well as a phylogenetic tree incorporating all members of IGHC genes from 26 mammals using two different methods (BI and ML). Similar tree topologies were inferred (Fig. 2, Supplemental Figs. 1–11). The phylogenetic relationship showed that the IGHC genes of marine mammals could be clustered well with those of their terrestrial relatives, suggesting that marine mammals possess the typical isotype repertoire as other eutherian mammals. Similar to other mammals, marine mammals were found to contain a single IGHM, IGHD, IGHE, and IGHA genes, and several IGHG genes based on the phylogenetic reconstruction. The order of IGHC genes in the IgH locus of marine mammals was determined to be 5′-IGHM-IGHD-IGHG-IGHE-IGHA-3′, which is in agreement with their terrestrial relatives (Supplemental Fig. 12). Furthermore, we found that the first domain (CH1) of IGHM and the CH1 of IGHD of cetaceans shared nearly identical sequences, as shown in artiodactylas (Zhao et al., 2002) (Supplemental Fig. 13). The genomic organization of marine mammal IGHC genes is similar to that of other species examined, which suggested a constraint acting on their genomic arrangement and the conserved genomic organization of the IGHC genes.

The expansion of the IgG subclass is primarily a feature of mammals (Hsu et al., 2006). The IgG-like Igs of chicken and lizard showed no subclass diversification (Butler et al., 2011). The gene duplication of IGHG genes is believed to have occurred most recently, and it belongs to a lineage-specific event (Wagner et al., 2002). In the present study, we found that there are usually two or four IGHG genes in marine mammals (Fig. 1). Like other mammals, the IGHG gene in marine mammals comprises three CH domains and a hinge segment. Furthermore, they have close sequence similarities to the IGHG genes in their terrestrial relatives. Based on the phylogenetic tree, it was found that IGHG genes within one species clustered together rather than with those of other species (Fig. 2). However, the IGHG genes of five species grouped together with their true orthologues in closely related species, including bottlenose dolphin, killer whale, Yangtze finless porpoise, Weddell seal and Pacific walrus, likely due to the species of Delphinidae (e.g., dolphin, killer whale) rapidly diverging during a very narrow time frame of ~3 million years (Zhou et al., 2011b) from the last common ancestral IGHG genes. These observations provided evidence to support the view that species-specific duplication of IGHG genes occurred during the evolution of mammalian species (Wagner et al., 2002). Overall, the presence of IgG subclasses and species-specific duplication of IGHG genes in marine mammals further confirmed the finding that IGHC genes of marine mammals fit the canonical pattern of their terrestrial relatives.

The number of IGHC genes has been noted to vary among different mammals. For example, 2 IGHA genes were present in humans and platypuses (Mage et al., 2006), whereas 13 IGHA genes were found in rabbits (Mccallum et al., 2004). For IGHG genes, seven copies were found in horses (Wagner et al., 2004), five in humans (Takahashi et al., 1982), and only one in rabbits (Mage et al., 2006) and opossum (Wang et al., 2009). No IGHG gene was detected in the bowhead whale and little brown bat (Fig. 1), which may due to the bad genomic assembly of both species considering that IgG subclasses have been identified in other whales (Mancia et al., 2006) and partial sequence encoding of IgG in little brown bat (*Myotis lucifugus*) was recovered by PCR (Butler et al., 2011). We also failed to retrieve IGHD genes from the genomes of sperm whale, hedgehog, European shrew, little brown bat, black flying fox and elephant despite considerable efforts (Fig. 1). One possible explanation is the poor genome quality, and the an alternative explanation is that the IGHD gene might have really been lost in evolution, similar to the situation in opossums (Wang et al., 2009) and rabbits (Mage et al., 2006). Further investigations to recover IGHD transcript by PCR will be needed in the future. Despite necessary further verification of IGHC genes in some species(e.g. with transcript evidence), changes in the copy number of IGHC genes among different mammalian lineages might reflect the rapid evolution of Ig genes (Hsu et al., 2006), corresponding to the marked variability found in the pathogens in different environments.

3.2. Adaptive evolution of IGHC genes in mammals

To detect the possible roles of positive selection during the evolution of IGHC genes in mammalian lineages, evolutionary analyses were conducted using a series of selection models implemented in the CODEML program and DataMonkey website. Because of the absence of intact IGHD genes in many mammals and the species-specific duplication of IGHG genes, only datasets of the IGHM, IGHE and IGHA genes were used to test the selection. Site model analysis of all mammals revealed that the models incorporating positive selection (M8) fit the data significantly better than the neutral model (M8a) for both the IGHM and IGHA genes, whereas no evidence of positive selection was identified at the IGHE genes (Table 2). Specially, M8 model detected nine positively selective sites at the IGHM and IGHA genes, whereas a total of 39 codons from the IGHM, IGHE and IGHA genes were also found to be under positive selection using the FEL, REL, and SLAC methods implemented in Datamonkey. Eleven codons from three genes

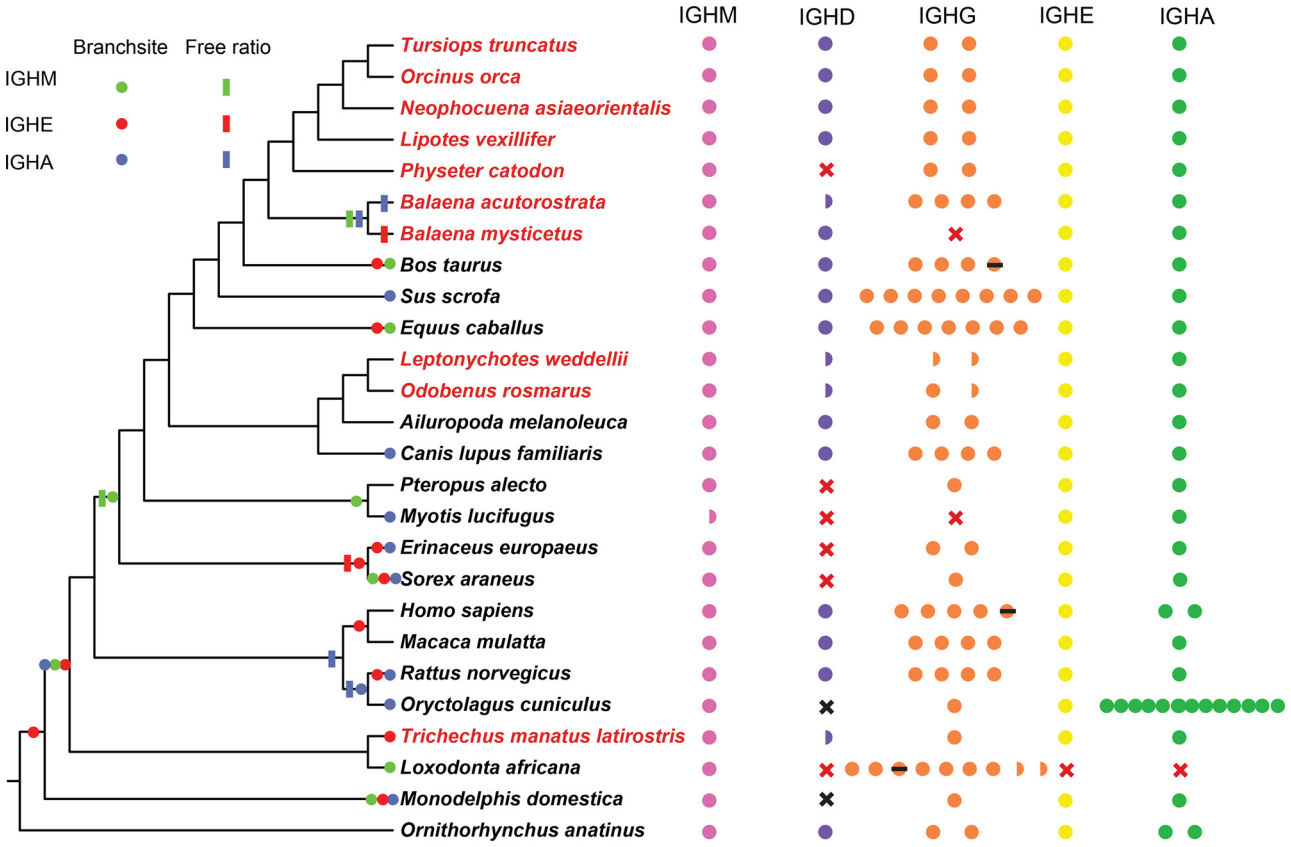

Fig. 1. Positive selection at IGHC genes across the mammal phylogeny and the presence/absence patterns of mammal IGHC genes.
A widely accepted phylogeny of mammals was used for selective pressure analysis in the free ratio model and the branch-site model. The branches indicated by vertical lines or circles show evidence of undergoing positive selection. The red font indicates marine mammals; black font indicates terrestrial mammals. The IGHC genes of corresponding species were listed in the right part of Fig. 1. Each circle represents a gene member; circles represent a complete gene sequence; semicircles indicate a partial gene sequence; red cross means no sequences found by TBLASTN and BLASTN approaches; black cross means the gene really lost; the species highlighted with a black line mean a pseudogene.

(5 in IGHM, 2 in IGHE, 4 in IGHA) were predicted to be robust sites under positive selection identified by at least two ML methods. Furthermore, TreeSAAP analysis showed that these 11 positively selected sites have undergone radical changes in their physicochemical properties (Table 3), and this may be regarded as another evidence to support significant positive selection. Then, we marked the putatively selected sites in the alignment and found that 36.4% (4/11) of these sites were located in, or close to, the functionally important regions of IGHC genes (Supplemental Fig. 14). This result suggested that positive selection on mammals may be the major driving force for the evolution of IGHC genes to adapt to environmental pathogens. Besides, for expanded subclasses, such as the 13 IGHA genes of rabbit and multiple copies of IGHG gene in mammals, we tested whether this species-specific expansion has been influenced by positive selection. The site model analysis (M8 vs. M8a, M1a vs. M2a implemented in PAML) showed that the IGHA genes of rabbits were under positive selection with similar positively selected sites (Supplemental Table 3). The branch-site model was used to examine whether some branches leading to gene duplication of IGHG genes (branch *a*-*o*) have been positively selected. The results showed that signals of positive selection were detected in branches *a* (Delphinoidea), *c* (*Physeter catodon*), *e* (*Bos taurus*), *f* (*Sus scrofa*), *g* (*Equus caballus*), and *h* (*Macaca mulatta*) (Supplemental Table 4, Supplemental Fig. 15). Overall, pervasive positive selection was detected on IGHC genes of mammals.

As is well known, many physicochemical properties, such as temperature, pressure, illumination and osmotic pressure, are different between water and land environments, possibly causing the difference in the diversity and abundance of pathogens. To date, there has been no deterministic conclusion on whether the pathogenic microbes in marine environment or in terrestrial environment are more diverse. Some early studies have supported weaker pathogenic pressure in the water environment (Zhang et al., 2016; Trowsdale et al., 1989; Slade, 1992), while several recent studies have examined the microbiome of marine mammals, suggesting that although the taxonomic composition of microbial communities in marine mammals is distinct from those of terrestrial mammals, its diversity is not lower than that of terrestrial mammals (Bik et al., 2016; Nelson et al., 2015).

To further test whether marine and terrestrial mammals were subjected to different pathogenic pressures during their evolution, we used the branch model in our study. The LRTs showed that the free-ratio model fits better than one-ratio model for IGHM, IGHE and IGHA genes, indicating heterogeneous selective pressures on different lineages. The ω values greater than 1 are limited to some branches in marine and terrestrial mammals, including the terminal branch of minke whale, the last common ancestral (LCA) branch of baleen whales, the LCA branch of Glires, and the LCA branch of Euarchontoglires at IGHA gene; the terminal branch of bowhead whale, and the LCA branch of Eulipotyphla at IGHE genes; the LCA branch of Laurasiatheria at IGHM gene (Fig. 1), suggesting similar selective pressures acting on IGHC genes in marine and terrestrial mammals. This result could provide evidence to support that the diversity and abundance of pathogens in water was comparable to that in land.

We also used more stringent branch-site model to predict positive selection acting on specific sites in each lineage. Almost all signs of positive selection were identified in terrestrial lineages (e.g., artiodactyla, perissodactyla, chiroptera, eulipotyphla, primates, Glires,

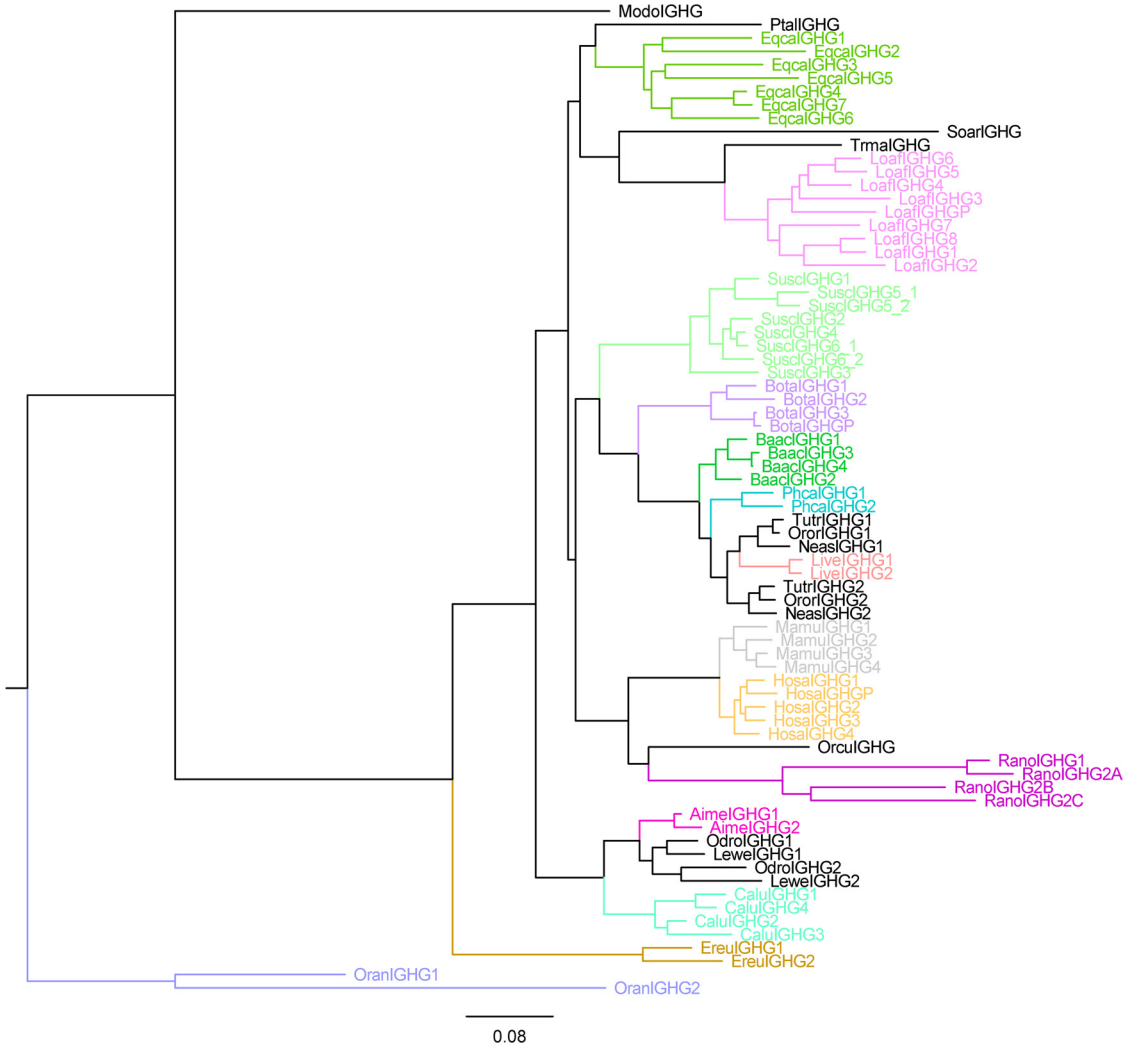

Fig. 2. Phylogenetic tree of mammal IGHG genes produced using MrBayes.
The CH domains of IGHG genes were used in the tree reconstruction. Tutr, Tursiops truncatus; Oror, Orcinus orca; Neas, Neophocuena asiaeorientalis; Live, Lipotes vexillifer; Phca, Physeter catodon; Baac, Balaena acutorostrata; Bamy, Balaena mysticetus; Lewe, Leptonychotes weddellii; Odro, Odobenus rosmarus; Trma, Trichechus manatus latirostris; Mylu, Myotis lucifugus; Ptal, Pteropus alecto; Ereu, Erinaceus europaeus; Soar, Sorex araneus; Loaf, Loxodonta africana; Bota, Bos taurus; Susc, Sus scrofa; Eqca, Equus caballus; Calu, Canis lupus familiaris; Aime, Ailuropoda melanoleuca; Hosa, Homo sapiens; Mamu, Macaca mulatta; Rano, Rattus norvegicus; Orcu, Oryctolagus cuniculus; Modo, Monodelphis domestica; Oran, Ornithorhynchus anatinus.

Table 2
Selective pressure analyses of mammal IGHC genes by site model.

Gene	Models	-lnL[a]	*P* value	M8 ω value	Positively Select Sites[b]($P > 0.90$)
IGHA	M8	10704.474		2.045	7 T 0.997; 203 A 0.937; 213 A 1.000; 237 R 1.000; 254 T 0.961; 321 N 0.977
	M8a	10716.250	< 0.001		
IGHE	M8	13389.996	0.065	1.371	
	M8a	13391.701			
IGHM	M8	14298.945	0.006	1.888	3 P 0.946; 101 Q 0.974; 294 L 0.990
	M8a	14302.713			

[a] lnL is the log-likelihood score.
[b] PPs of Bayes Empirical Bayes (BEB) analysis with $P > 0.9$ was regarded as candidates for selection.

Fig. 1, Supplemental Table 4). By contrast, for marine mammal lineages, only the terminal branch of Florida manatee was subjected to positive selection (Fig. 1, Supplemental Table 4). This result supported that pathogenic pressure was weaker in aquatic environment than in terrestrial environment. Moreover, recent studies showed a very low number of variable genes that directly binds to antigens in marine mammals, which also supported this conclusion (Olivieri et al., 2014). But further study on the variable region will be needed, providing insights to understand whether the difference in selective pressure between marine and terrestrial mammals existed and whether such a difference could respond to the different pathogenic pressure between aquatic and land environments.

In the present study, IGHC genes were determined to have undergone positive selection in Chiroptera, which is now known as notorious reservoir hosts for a number of pathogenic viruses, including Nipah, Hendra, Ebola, and severe acute respiratory syndrome (SARS) responsible for severe human and livestock disease outbreaks. However, they rarely display clinical symptoms (Zhang et al., 2013; Wang, 2009). The ability of bats to coexist with viruses presents an interesting immunological problem. Previous studies have founded that some immunoglobulin gene superfamilies underwent expansion in Brandt's bat (Seim et al., 2013) and that some positively selected genes were involved in the immune response (Zhang et al., 2013). Thus, IGHC genes subject to positive selection in chiropterans may be related to the long-term coexistence of bats and viruses.

Table 3
Positively selected sites of mammal IGHC genes detected by PAML, Datamonkey and TreeSAAP.

Genes	AA positions	PAML		Datamonkey			TreeSAAP	
		Branch-site model ($P > 0.90$)	Site model ($P > 0.90$)	SLAC ($P < 0.1$)	FEL ($P < 0.1$)	REL (BF > 50)	Radical Changes in Amino Acid Properties[a]	Total
IGHA	200			0.0634811	0.0304991	77.8394	*pHi, α_c, E_t,*	3
	213		1.000		0.0406899		*P_α, Ns, B_l, B_r, R_F, P_c, K°, h, F, P_r, α_c, α_m, α_n, R_a, H_p, H_t, E_t, P*	18
	328			0.0351638	0.0935704		*P_r, α_c, E_t*	3
	364			0.0610423		52.2109	*P_α, N_s, P_β, B_r, R_F, P_c, K°, h, F, P_r, p, α_c, α_n, R_a, H_p, H_t, E_t, P*	18
	365			0.034845	0.0453226	52.0384	*P_α, P_c, K°, P_r, p, α_c, E_t, P*	8
IGHE	201			0.0950233	0.098536		*N_s, P_β, B_r, R_F, P_c, pK', C_α, h, pH_i, E_t, F, H_{nc}, V°, P_r, p, E_{sm}, R_a, H_p, H_t, E_t, P*	21
	224			0.099105	0.00221453	312.686	*P_α, K°, H_t*	3
IGHM	48			0.0880392	0.0671209		*P_α, N_s, B_l, R_F, P_c, pH_i, F, α_c, α_n, P*	10
	254			0.0585277	0.0527633		*P_α, K°, H_t*	3
	294		0.990		0.0526114		*P_α, P_β, P_c, K°, pK', pH_i, P_r, α_c, α_n, R_a, P*	11
	345			0.0287169	0.0361345	57.5523	*P_α, N_s, P_r, p, α_c, R_a, H_b P*	8

[a] Radical changes in amino acid properties under categories 6–8 were detected in TreeSAAP. P_α = α-helical tendencies, N_s = Average number of surrounding residues, P_β = β-Structure tendencies, B_l = Bulkiness, B_r = Buriedness, R_F = Chromatographic index, P_c = Coil tendencies, K° = Compressibility, pK' = Equilibrium constant of ionization for COOH, C_α = Helical contact energy, h = Hydropathy, pHi = Isoelectric point, E_l = Long-range non-bonded energy, F = Mean r.m.s. fluctuation displacement, H_{nc} = Normal consensus hydrophobicity, V° = Partial specific volume, P_r = Polar requirement, p = Polarity, α_c = Power to be C-terminal, α-helix, α_m = Power to be middle, α-helix, α_n = Power to be N-terminal of an α-helix, E_{sm} = Short- and medium-range non-bonded energy, R_a = Solvent accessible reduction ratio, H_p = Surrounding hydrophobicity, H_t = Thermodynamic transfer hydrophobicity, E_t = Total non-bonded energy, P = Turn tendencies.

Evidences for positive selection has also been detected in some domesticated mammals such as pig, horse and cattle. Approximately 10,000 years ago, the domestication of farm animals by selection for biological, agricultural and biomedical importance began, leading to animals that slowly adapted to local condition. Approximately 200 years ago, selection pressures on livestock increased with the rise of the concept of breeds (Taberlet et al., 2008). It has been previously proposed that artificial selection indirectly drove for rapid adaptation of domestic animals and pathogen genes (Smith et al., 2012). Therefore, we postulated that artificial selection resulted in the adaptive evolution observed in IGHC genes of domestic animals.

4. Conclusions

This study has provided the first characterization of IGHC genes in marine mammals and explored their potential molecular genetic basis in the secondary aquatic adaptation. Our analyses revealed that the IGHC genes in marine mammals fit the canonical pattern of their terrestrial relatives. In addition, different levels of selection were detected between marine and terrestrial mammalian lineages with the branch-site model, suggesting a divergent adaptation to land and aquatic environments probably due to contrasting pathogenic pressures. Further investigation on the variable region of IgH genes is necessary to improve our understanding of the immune adaptation of marine mammals.

Funding

This work was supported by the Key Project of National Natural Science of China (grant number 31630071 to GY); the National Natural Science Foundation of China (NSFC) (grant numbers 31570379; 31772448 to SX); and the National Key Research and Development Programme, Ministry of Science and Technology (Grant number 2016YFC0503200 to GY and SX; the Priority Academic Program Development of Jiangsu Higher Education Institutions to GY and SX).

Declarations of interests

None.

Acknowledgements

We thank Zepeng Zhang, Xiaohui Sun, Yulin Ge, Yuan Mu and Jing Li for support and discussion.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.molimm.2018.04.011.

References

Bengtén, E., Wilson, M., Miller, N., Clem, L.W., Pilström, L., Warr, G.W., 2000. Immunoglobulin isotypes: structure, function, and genetics. In: Pasquier, L.D., Litman, G.W. (Eds.), Origin and evolution of the vertebrate immune system. Springer-Verlag, Berlin Heidelberg, pp. 189–219.

Bik, E.M., Costello, E.K., Switzer, A.D., Callahan, B.J., Holmes, S.P., Wells, R.S., et al., 2016. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nat. Commun. 7, 10516.

Breaux, B., Deiss, T.C., Chen, P.L., Cruz-Schneider, M.P., Sena, L., Hunter, M.E., et al., 2017. The Florida manatee (*Trichechus manatus latirostris*) immunoglobulin heavy chain suggests the importance of clan III variable segments in repertoire diversity. Dev. Comp. Immunol. 72, 57–68.

Butler, J.E., Wertz, N., Zhao, Y., Zhang, S., Bao, Y., Bratsch, S., et al., 2011. The two suborders of chiropterans have the canonical heavy-chain immunoglobulin (Ig) gene repertoire of eutherian mammals. Dev. Comp. Immunol. 35 (3), 273–284.

Eguchi-Ogawa, T., Toki, D., Wertz, N., Butler, J.E., Uenishi, H., 2012. Structure of the genomic sequence comprising the immunoglobulin heavy constant (IGHC) genes from sus scrofa. Mol. Immunol. 52 (3-4), 97–107.

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

Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J., Nedbal, M.A., 2005. Molecular phylogeny of the carnivora (mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. Syst. Biol. 54 (2), 317–337.

Hsu, E., Pulham, N., Rumfelt, L.L., Flajnik, M.F., 2006. The plasticity of immunoglobulin gene systems in evolution. Immunol. Rev. 210 (1), 8–26.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: bayesian inference of phylogenetic trees. Bioinformatics 17 (8), 754–755.

Lundqvist, M.L., Kohlberg, K.E., Gefroh, H.A., Arnaud, P., Middleton, D.L., Romano, T.A., et al., 2002. Cloning of the IgM heavy chain of the bottlenose dolphin (*Tursiops truncatus*), and initial analysis of VH gene usage. Dev. Comp. Immunol. 26 (6), 551–562.

Macpherson, A.J., Mccoy, K.D., Johansen, F.E., Brandtzaeg, P., 2008. The immune geography of IgA induction and function. Mucosal Immunol. 1 (1), 11–22.

Mage, R.G., Lanning, D., Knight, K.L., 2006. B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues. Dev. Comp. Immunol. 30 (1-2), 137–153.

Maizels, N., 2005. Immunoglobulin gene diversification. Annu. Rev. Genet. 39 (1), 23–46.

Mancia, A., Romano, T.A., Gefroh, H.A., Chapman, R.W., Middleton, D.L., Warr, G.W., et al., 2006. The immunoglobulin G heavy chain (IGHG) genes of the Atlantic bottlenose dolphin, Tursiops truncatus. Comp. Biochem. Phys. B 144 (1), 38–46.
Mancia, A., Romano, T.A., Gefroh, H.A., Chapman, R.W., Middleton, D.L., Warr, G.W., et al., 2007. Characterization of the immunoglobulin A heavy chain gene of the Atlantic bottlenose dolphin (*Tursiops truncatus*) ☆. Vet. Immunol. Immunopathol. 118 (3-4), 304–309.
Mccallum, H.I., Kuris, A., Harvell, C.D., Lafferty, K.D., Smith, G.W., Porter, J., 2004. Does terrestrial epidemiology apply to marine systems? Trends Ecol. Evol. 19 (11), 585–591.
Nelson, T.M., Apprill, A., Mann, J., Rogers, T.L., Brown, M.V., 2015. The marine mammal microbiome: current knowledge and future directions. Microbiol. Aust. 36, 8–13.
Nylander, J.A.A., 2004. MrModeltest 2.3. Program Distributed by the author. Evolutionary Biology Centre, Uppsala University.
Ohta, Y., Flajnik, M., 2006. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proc. Natl. Acad. Sci. U. S. A. 103 (28), 10723–10728.
Olivieri, D., Haeften, B.V., Sánchez-espinel, C., Gambón-Deza, F., 2014. The immunologic V-gene repertoire in mammals. bioRxiv. http://dx.doi.org/10.1101/002667.
Penn, O., Privman, E., Ashkenazy, H., Landan, G., Dan, G., Pupko, T., 2010a. GUIDANCE: a web server for assessing alignment confidence scores. Nucleic Acids Res. 38, W23–W28 Web Server issue.
Penn, O., Privman, E., Landan, G., Dan, G., Pupko, T., 2010b. An alignment confidence score capturing robustness to guide tree uncertainty. Mol. Biol. Evol. 27 (8), 1759–1767.
Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A., et al., 2011. A molecular phylogeny of living primates. Plos Genet. 7 (3), e1001342.
Pond, S.L.K., Frost, S.D., 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21 (10), 2531–2533.
Pond, S.L.K., Frost, S.D., 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22 (5), 1208–1222.
Schroeder Jr., H.W., Cavacini, L., 2010. Structure and function of immunoglobulins. J. Allergy Clin. Immunol. 125 (202), 41–52.
Seim, I., Fang, X., Xiong, Z., Lobanov, A.V., Huang, Z., Ma, S., et al., 2013. Genome analysis reveals insights into physiology and longevity of the brandt' bat myotis brandtii. Nat. Commun. 4 2212-2212.
Shen, T., Xu, S., Wang, X., Yu, W., Zhou, K., Yang, G., 2012. Adaptive evolution and functional constraint at TLR4 during the secondary aquatic adaptation and diversification of cetaceans. BMC Evol. Biol. 12 (1), 39.
Slade, R.W., 1992. Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. Proc. R. Soc. B Biol. Sci. 249 (1325), 163–171.
Smith, S.A., Jann, O.C., Haig, D., Russell, G.C., Werling, D., Glass, E.J., et al., 2012. Adaptive evolution of Toll-like receptor 5 in domesticated mammals. BMC Evol. Biol. 12 (1), 122.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21), 2688–2690.
Sun, Y., Wei, Z., Hammarstrom, L., Zhao, Y., 2011. The immunoglobulin δ gene in jawed vertebrates: a comparative overview. Dev. Com. Immunol. 35, 975–981.
Taberlet, P., Valentini, A., Rezaei, H.R., Naderi, S., Pompanon, F., Negrini, R., et al., 2008. Are cattle, sheep, and goats endangered species? Mol. Ecol. 17 (1), 275–284.
Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S., Honjo, T., 1982. Structure of human immunoglobulin gamma genes: implications for evolution of a gene family. Cell 29 (2), 671–679.
Thewissen, J.G., Cooper, L.N., Clementz, M.T., Bajpai, S., Tiwari, B.N., 2007. Whales originated from aquatic artiodactyls in the eocene epoch of india. Nature 450 (7173), 1190–1194.
Trowsdale, J., Groves, V., Arnason, A., 1989. Limited MHC polymorphism in whales. Immunogenetics 29 (1), 19–24.
Uhen, M.D., 2007. Evolution of marine mammals: back to the sea after 300 million years. Anat. Rec. 290 (6), 514–522.
Uhen, M.D., 2010. The origin(s) of whales. Annu. Rev. Earth Pl. Sci. 38 (1), 189–219.
Wagner, B., Greiser-Wilke, I., Wege, A.K., Radbruch, A., Leibold, W., 2002. Evolution of the six horse IGHG genes and corresponding immunoglobulin gamma heavy chains. Immunogenetics 54 (5), 353–364.
Wagner, B., Miller, D.C., Lear, T.L., Antczak, D.F., 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 (5), 3230–3242.
Wang, L.F., 2009. Bats and viruses: a brief review. Virol. Sin. 24 (2), 93–99.
Wang, X., Olp, J.J., Miller, R.D., 2009. On the genomics of immunoglobulins in the gray, short-tailed opossum monodelphis domestica. Immunogenetics 61 (8), 581–596.
Woolley, S., Johnson, J., Smith, M.J., Crandall, K.A., Mcclellan, D.A., 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19 (5), 671–672.
Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15 (5), 568–573.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24 (8), 1586–1591.
Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22 (4), 1107–1118.
Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22 (12), 2472–2479.
Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K.A., Fang, X., et al., 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science 339 (6118), 456–460.
Zhang, X., Lin, W., Zhou, R., Gui, D., Yu, X., Wu, Y., 2016. Low major histocompatibility complex class II variation in the endangered indo-pacific humpback dolphin (*Sousa chinensis*): inferences about the role of balancing selection. J. Hered. 107 (2), 143–152.
Zhao, Y., Kacskovics, I., Pan, Q., Liberles, D.A., Geli, J., Davis, S.K., et al., 2002. Artiodactyl IgD: the missing link. J. Immunol. 169 (8), 4408–4416.
Zhou, X., Xu, S., Xu, J., Chen, B., Zhou, K., Yang, G., 2011a. Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the laurasiatherian mammals. Syst. Biol. 61 (1), 150–164.
Zhou, X., Xu, S., Yang, Y., Zhou, K., Yang, G., 2011b. Phylogenomic analyses and improved resolution of cetartiodactyla. Mol. Phylogenet. Evol. 61 (2), 255–264.