UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

BRUNA CRISTINA DIAS

MOLECULAR EVOLUTION OF ADAPTIVE IMMUNE SYSTEM GENES IN AQUATIC MAMMALS

EVOLUÇÃO MOLECULAR DE GENES DO SISTEMA IMUNOLÓGICO ADAPTATIVO EM MAMÍFEROS AQUÁTICOS

CAMPINAS
2022

BRUNA CRISTINA DIAS

MOLECULAR EVOLUTION OF ADAPTIVE IMMUNE SYSTEM GENES IN AQUATIC MAMMALS

EVOLUÇÃO MOLECULAR DE GENES DO SISTEMA IMUNOLÓGICO ADAPTATIVO EM MAMÍFEROS AQUÁTICOS

Dissertação apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Mestra em Genética e Biologia Molecular, na área de Genética Animal e Evolução.

Dissertation presented to the Biology Institute of the University of Campinas in partial fulfillment of the requirements for the degree of Master in the area of Genetics and Molecular Biology, in the field of Animal Genetics and Evolution.

Orientadora: Profa. Dra. Mariana Freitas Nery

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO/TESE DEFENDIDA PELA ALUNA BRUNA CRISTINA DIAS, E ORIENTADA PELA PROFA. DRA. MARIANA FREITAS NERY

CAMPINAS

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

Dias, Bruna Cristina, 1993-
D543m Molecular evolution of adaptive immune system genes in aquatic mammals / Bruna Cristina Dias. – Campinas, SP: [s.n.], 2021.

Orientadora: Mariana Freitas Nery.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Evolução molecular. 2. Imunidade adaptativa. 3. Mamífero aquático. 4. Imunogenética. 5. Evolução. I. Nery, Mariana Freitas, 1983-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Evolução molecular de genes do sistema imunológico adaptativo em mamíferos aquáticos
Palavras-chave em inglês:
Molecular evolution
Adaptive immunity
Aquatic mammals
Immunogenetics
Evolution
Área de concentração: Genética Animal e Evolução
Titulação: Mestra em Genética e Biologia Molecular
Banca examinadora:
Mariana Freitas Nery [Orientador]
Juliana de Abreu Vianna
Marcelo Mendes Brandão
Data de defesa: 09-07-2021
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-9817-1005
- Currículo Lattes do autor: http://lattes.cnpq.br/7525609810215272

Campinas, 09 de Julho de 2021

COMISSÃO EXAMINADORA

Profa. Dra. Mariana Freitas Nery

Prof. Dr. Marcelo Mendes Brandão

Dra. Juliana de Abreu Vianna

A Ata da defesa com as respectivas assinaturas dos membros da Comissão Examinadora encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

A mind that is stretched by a new experience
can never go back to its old dimensions.

Oliver Wendell Holmes Jr.

AGRADECIMENTOS

À minha orientadora, Profa. Dra. Mariana Nery, pela ótima orientação, por me direcionar nos momentos que me perdi, pela parceria, paciência e compreensão, por acreditar em mim e me acalmar nos momentos de ansiedade, pela oportunidade de crescimento pessoal e profissional que me deu, permitindo com que eu experimentasse o início da vida acadêmica;

Aos amigos que fiz no Laboratório de Genômica Evolutiva e apoiadores oficiais nos momentos mais difíceis, Érica Souza, Flávia Akemi e Lucas Canesin, pelas conversas e conselhos, pelo conhecimento compartilhado, pelas risadas e puxões de orelha, pelo companheirismo. Sem vocês, este trabalho jamais teria acontecido; Aos meus companheiros de laboratório, Elisa Ramos e Lucas Freitas, por tornarem o ambiente de trabalho um local agradável e por me ajudarem sempre que precisei;

Ao Prof. Dr. Michael Criscitiello por me permitir trabalhar em seu laboratório e pelo acolhimento em um momento tão intenso, pela incrível orientação e pelos ensinamentos. Assim como a todos os integrantes do Laboratório de Imunogenética Comparada da Universidade Texas A&M, Jenna Rocket, Kaitlyn Romoser, Jeannine Ott e Christian Mitchell, pela receptividade, auxílio e parceria.

À minha psicóloga Rita Ferramola, pelo suporte, autoconhecimento e direcionamento. Seu papel foi essencial para a finalização deste trabalho;

Aos membros da banca de qualificação, a Profa. A Dra. Vera Solferini e o Prof. Dr. Daniel Lahr pelos ótimos comentários e observações;

Aos membros da banca de defesa da dissertação, a Profa. Dra. Juliana Vianna e ao Prof. Dr. Marcelo Brandão, pela devolutiva da dissertação com valiosas sugestões e correções, tornando o trabalho mais adequado e superior, além da presença e sensibilidade de ambos em um dia tão importante em minha vida profissional e pessoal.

A toda a minha família, pelo amor, pelo apoio e por me permitirem chegar até aqui. Por todos os esforços prestados e pelo carinho, pela ajuda a encontrar meios para solucionar problemas e pelos momentos de alegria e comemoração vividos juntos;

Aos meus amigos, pela parceria eterna e pelo companheirismo, pelo suporte e presença em todos os momentos, bons e ruins. Pelas risadas e por tornar a caminhada mais leve, sempre.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento deste trabalho (processo número 2015/18269-1), pela bolsa concedida a mim (2017/14831-2) e pela incrível oportunidade de crescimento proporcionada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

Apesar dos constantes desafios microbianos do meio ambiente, o corpo humano evita infecções por um sistema específico e incrivelmente diverso que pode combater uma miríade de patógenos. O sistema imune adaptativo (AIS) é um mecanismo de defesa muito complexo. O AIS possui dois grupos diferentes de células e moléculas, denominados linfócitos B e T, que se defendem contra micróbios extracelulares e intracelulares. Os linfócitos B sintetizam exclusivamente anticorpos - chamados coletivamente de imunoglobulinas (IG) - produzidos em grande variedade, cada um com diferentes sequências de DNA e, consequentemente, diferentes sítios de ligação ao antígeno. O complexo de genes ativadores de recombinação 1 (RAG1) e 2 (RAG2) catalisa a montagem aleatória de segmentos gênicos variáveis (V), diversidade (D) e junção (J) que estão presentes no genoma em inúmeras cópias e geram a enorme variedade dos anticorpos montados e receptores de antígeno. Acredita-se que a origem dos RAGs deve ter sido crucial para a imunidade adaptativa, já que todos os genes BCR e TCR usam o mesmo mecanismo de rearranjo V(D)J. A ocupação de novos ambientes é um cenário favorável ao surgimento e diversificação de novas espécies. A transição dos tetrápodes da terra para a água é um exemplo clássico da ocupação de novos ambientes, caracterizados por grandes transformações morfológicas e fisiológicas. Cetáceos e sirênios constituem modelos interessantes de espécie a serem estudados considerando suas mudanças ambientais da terra para a água, bem como sua segunda radiação de ambientes marinhos para fluviais. A fim de aprofundar nosso conhecimento sobre o sistema imune adaptativo de mamíferos aquáticos por meio de uma perspectiva evolutiva considerando diferentes transições ambientais, o Capítulo 1 traz análises evolutivas dos genes RAG1 e RAG2 em Cetáceos, demonstrando que os genes RAG1 e RAG2 permaneceram bastante conservados entre os tetrápodes, e evolução diferencial acontecendo nesses genes intimamente ligados na linhagem dos cetáceos, com RAG1 sendo menos conservado quando comparado a outros mamíferos da filogenia. O capítulo 2 é a primeira caracterização do *locus* da cadeia leve da imunoglobulina em *Trichechus manatus* e *Trichechus inunguis*, mostrando que o número de segmentos codificados genomicamente e, consequentemente, a diversidade segmentar é limitada nos peixes-boi. Os peixes-boi parecem ter apenas alguns pseudogenes V dentro da cadeia lambda organizados de maneira diferente como visto na maioria dos tetrápodes, permitindo que os peixes-boi criem muito pouca diversidade de anticorpos através da região da cadeia leve da imunoglobulina.

ABSTRACT

Despite the constant microbial challenges from the environment, the body prevents infections by a specific and incredibly diverse system that can fight myriad pathogens. The adaptive immune system (AIS) is a very complex defense mechanism. The AIS has two different groups of cells and molecules, called B and T lymphocytes, that defend against extracellular microbes and intracellular microbes. B lymphocytes exclusively synthesize antibodies - collectively called immunoglobulins (IG) - produced in great variety, each with different DNA sequences and, consequently, different antigen-binding sites. Recombination-activating genes 1 (RAG1) and 2 (RAG2) genes complex catalyzes random assembly of variable (V), diversity (D), and joining (J) gene segments that are present in the genome in numerous copies and generate the enormous variety of the assembled antibodies and antigen receptors. It is believed that the RAGs origin must have been crucial for adaptive immunity, as all of the BCR and TCR genes use the same mechanism of V(D)J rearrangement. The occupation of new environments is a favorable scenario for the emergence and diversification of new species, and the molecular changes taking place in the genome of organisms during this occupation have ceased to be a mystery. Tetrapods' transition from land to water is a classic example of the occupation of new environments, characterized by great morphological and physiological transformations. Cetaceans and sirenians make an interesting model species to study considering their environmental changes from land to water, as well as their second radiation from marine to fluvial environments. In order to deepen our knowledge about the adaptive immune system of aquatic mammals through an evolutionary perspective considering different environmental transitions, Chapter 1 brings evolutionary analyses of the RAG1 and RAG2 genes in Cetaceans, demonstrating that RAG1 and RAG2 genes remained fairly conserved among tetrapods, and differential evolution happening in these closely linked genes in the Cetacea lineage, with RAG1 being less conserved when compared to other mammals of the phylogeny. Chapter 2 is the first characterization of both *Trichechus manatus* and *Trichechus inunguis* immunoglobulin light chain locus, showing that the number of genomically encoded segments and, consequently, the segmental diversity is limited in manatees. Manatees seem to have only a few V pseudogenes within the lambda chain organized in a different manner as seen in most tetrapods, enabling manatees to create very little antibody diversity through the immunoglobulin light chain region.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIAÇÕES E SIGLAS

- **MGK+F3X4+G4 model of substitution:**
  - **MG**: Nonsynonymous/synonymous (dn/ds) rate ratio with additional transition/transversion (ts/tv) rate ratio;
  - **+F3X4**: unequal nucleotide frequencies and unequal nucleotide frequencies over three codon positions;
  - **+G4**: discrete Gamma model with default 4 rate categories.
- **GY+F1X4+G4 model of substitution:**
  - **GY**: Nonsynonymous/synonymous and transition/transversion rate ratios;
  - **+F1X4**: Unequal nucleotide frequencies but equal nucleotide frequencies over three codon positions;
  - **+G4**: discrete Gamma model with default four rate categories.
  - **Markov-modulated Markov models of codon evolution**: this means that the future of the process (the various states possibly reached and their probabilities of occurrence) depends only on the present state, not on the past. Markov processes can be in discrete time, when states are assigned to successive "steps," or "generations," or in continuous time when the time to next event is an exponential random variable;
- **LRTs**: likelihood ratio tests LRTs;
- **TE**: transposable element;
- **LRTs**: likelihood ratio tests LRTs;
- **TE**: transposable element;
- **RAG1**: recombination activation gene 1;
- **RAG2**: recombination activation gene 2;
- **IG**: immunoglobulin;
- **IGH**: immunoglobulin heavy chain;
- **IGL**: immunoglobulin light chain;
- **VL**: variable lambda;
- **VK**: variable kappa;
- **CL**: constant lambda;

- **CK**: constant kappa;
- **JL**: joining lambda;
- **JK**: joining kappa;
- **IGHV**: immunoglobulin heavy chain variable gene;
- **IGHJ**: immunoglobulin heavy chain joining gene;
- **IGHC**: immunoglobulin heavy chain constant region;
- **IGLV**: immunoglobulin lambda chain variable gene;
- **IGLJ**: immunoglobulin lambda chain joining gene;
- **IGLC**: immunoglobulin lambda chain constant region;
- **IGKV**: immunoglobulin kappa chain variable gene;
- **IGKJ**: immunoglobulin kappa chain joining gene;
- **IGKC**: immunoglobulin kappa chain constant region.

SUMÁRIO

CHAPTER 1

1. INTRODUCTION

1.1. The Adaptive Immune System

The Adaptive Immune System (AIS) is a very complex defense mechanism. It consists of lymphocytes and their products, such as the antibodies providing specialized defense against infections (Abbas et al., 2015). Lymphocytes express receptors that specifically recognize a wide variety of molecules produced by pathogens. This recognition creates a 'memory' that enables a large and effective response to repeated exposures to previously encountered pathogens (Flajnik and Kasahara, 2010; Abbas et al., 2015). The AIS has two different groups of cells and molecules, called B and T lymphocytes, that defend against extracellular microbes and intracellular microbes. Plasma membrane antigen receptors and signaling molecules in T lymphocytes are called the T cell receptor (TCR) complex, whereas in lymphocytes B, it is called the B cell receptor (BCR) complex (Abbas et al., 2015).

B lymphocytes exclusively synthesize antibodies - collectively called immunoglobulins (IG) - produced in great variety, each with different DNA sequences and, consequently, different antigen-binding sites (Alberts et al., 2002). When secreted in the blood, immunoglobulins can recognize microbial antigens and toxins by their variable domains (Abbas et al., 2015). Immunoglobulins have a characteristic three-dimensional Y shape composed of four polypeptide chains: two identical heavy (H) chains and two identical light (L) chains. Each chain contains a variable (V) and a constant (C) region, composing together with the immunoglobulin (IG) domain (Abbas et al., 2015). The discovery of antibodies (immunoglobulins) over 100 years ago and the detection of somatic hypermutation in 1970 (Weigert et al., 1970), together with VDJ rearrangement (variable–diversity–joining) of immunoglobulin (IG) genes findings (Tonegawa, 1976) a few years later, solved important questions relating to the generation of diversity of the AIS. Like most molecules of the AIS, immunoglobulins were first found in cartilaginous fish generated by a somatic recombination mechanism (Flajnik, 2002). In general, all vertebrates have multiple isotypes that provide distinct effector functions. B cell receptor heavy chain isotypes IgG, IgD, IgA, IgM, and IgE

each had different origins along with jawed vertebrates' evolutionary history, with the IgM isotype being the most ancient antibody class and has the same function in all gnathostomes (Flajnik, 2002). The B cell receptor light chains kappa and lambda were previously believed to have originated late in evolution, probably in the mammalian lineage, being later shown that both isotypes emerged in a common ancestor of all living vertebrates (Greenberg et al., 1993; Rast et al., 1994). Moreover, studies identified another light chain (σ-chain) in frogs that arose early in evolution and were retained in all cold-blooded gnathostomes (Schwager et al., 1991).

In the 1980s, studies showed that TCR and BCR genes have similar domain organization and rely on the same rearrangement mechanism to generate diversity, being hypothesized that they shared a common ancestor (Davis et al., 1984). Later, the discovery of two enzymes called RAG1 and RAG2 that were involved in the rearrangement of BCR and TCR genes (Oettinger et al., 1990) and hypermutation of BCR genes (Muramatsu et al., 2000) raised other important questions about how a system capable of generating such a great diversity evolved.

The evolution of immune genes - changing from expressing generalized reactions to highly specific responses along with the history of life - is intriguing. Consequently, many efforts are being made to conduct comparative studies of AIS genes on extant vertebrates, taking advantage of the growing availability of new whole-genome information and the development of new approaches to analyze gene expression and protein interactions. In this study, we present two newly sequenced whole-genome immune genes information about two different manatee species and two different dolphin species, adding new valuable data to collaborate on our understanding of the evolutionary history of the Adaptive Immune System genes.

1.1.1. RAGs

One of the suggested origins of the RAG1 and RAG2 genes was a genetic transposition event that might have happened more than 500 million years ago and could have involved a recombination-activating gene (RAG)-like element. This event, which gave rise to the predecessors of the rearranging antigen-binding receptors of the jawed vertebrates, was defined as the determining point in the evolution of RAG-mediated adaptive immunity

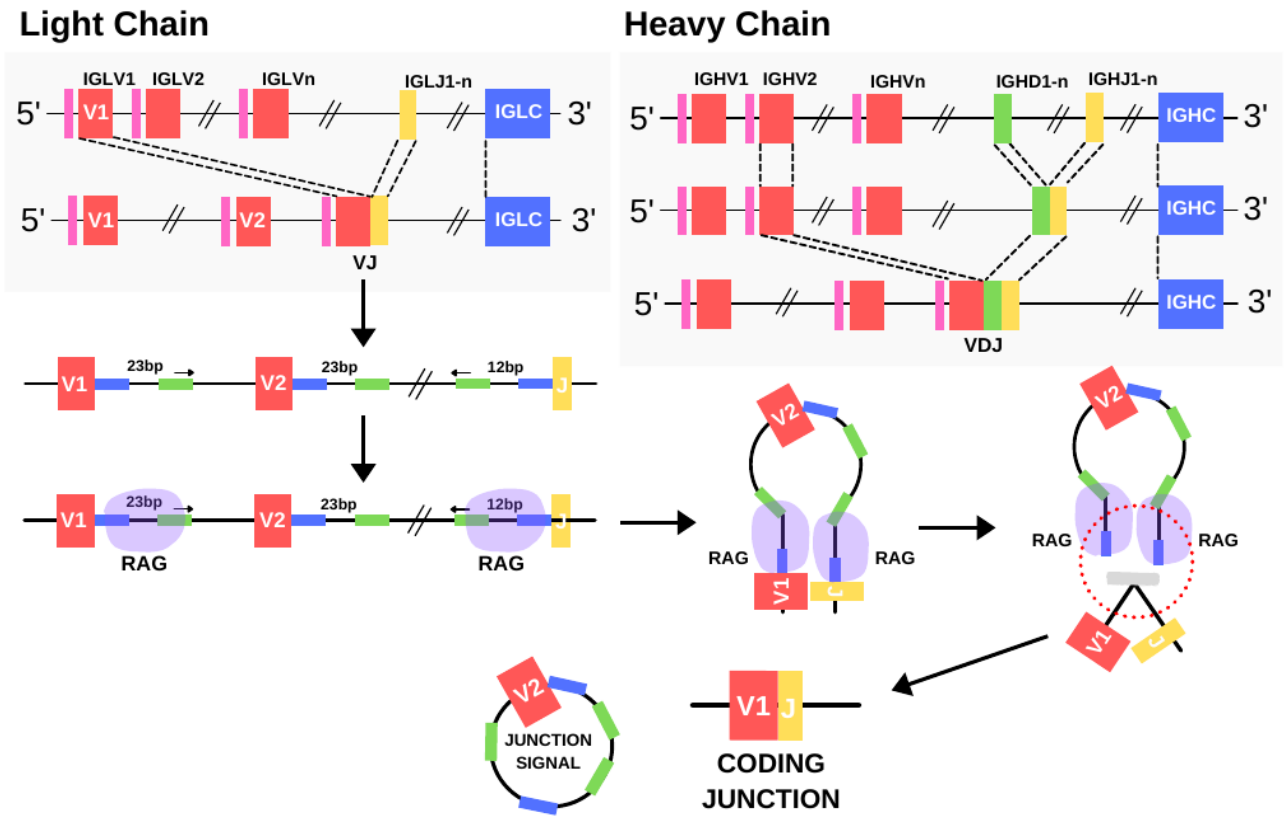
(Agrawal et al., 1998; Hiom et al., 1998). The RAG-mediated adaptive immunity would evolve throughout building a mechanism for acquiring almost unlimited variation from very few genes. Yet, the origins of RAG1 and RAG2 genes are still unclear. Clear identification of two linked RAG1 and RAG2 genes were found in the sea-urchin genome, suggesting that a RAG1-like gene could have invaded the sea-urchin's genome and inserted near a RAG2-like gene. This RAG2-like gene would then be included in the original transposon (Cannon et al., 2004; Kapitonov and Jurka, 2005; Fugmann et al., 2006). Similarly, RAG1 and RAG2 genes are closely linked in the human genome, located just eight kilobases apart from one another (Oettinger et al., 1990; Greenhalgh et al., 1993). Other hypotheses suggest that a RAG1-like gene may have originated from the *Transib* superfamily, one of ten superfamilies of DNA transposons detected in eukaryotes (Kapitonov and Jurka, 2003; Kapitonov and Jurka, 2005), or even from a viral genetic element (Dreyfus, 2009). In addition, studies have not been successful in isolating complete RAG genes from jawless fish, reinforcing the idea that the acquisition of the RAG transposon helped trigger BCR- and TCR- based immunity (Litman et al., 2005).

The RAG1-RAG2 complex catalyzes random assembly of variable (V), diversity (D), and joining (J) gene segments that are present in the genome in numerous copies and generate the enormous variety of the assembled antibodies and antigen receptors (Gellert, 2002; Kapitonov and Koonin, 2015). In the germline, the V region is encoded by noncontiguous gene portions split into V, J, and, in some cases, D segments. Each of these gene segments is joined in a site-specific recombination reaction, known as the V(D)J recombination (Figure 1), to form the exon that encodes the antigen-binding portion of a polypeptide. The RAG transposon had a considerable role in the emergence of the jawed vertebrate AIS, as all BCR and TCR genes use the same VDJ rearrangement mechanism. Major features of the system arose early and have been relatively conserved over evolutionary time (Flajnik and Kasahara, 2010). Jawed vertebrate AIS seems to display conserved features that allow for the general functioning of the system, like RAG1 and RAG2 genes; and plastic features that would enable rapid changes over evolutionary time to respond to different pathogens (Klein et al., 1993; Flajnik and Kasahara, 2010), like the immunoglobulins.

The figure 1 simplified scheme of the V(D)J system and the role of RAG1 to RAG2 enzymes within that system. The light chain V region genes (on the left) are made up of two

segments: a variable (V) and a junction (J) segments (red and yellow blocks, respectively), which together form a complete exon of the light chain V region (VJ). The heavy chain regions (on the right) are made up of three segments: a variable (V), a diversity (D), and a junction (J) segments (red, green, and yellow blocks, respectively), which together form a complete exon of the V region of the heavy chain (VDJ). The RAG complex binds to the 23bp and 12bp Recombination Signal Sequences (RSS). The DNA molecules are broken at the blue ends highlighted by the red dotted circle (heptamer sequences) and are joined together to form a circle of non-functional DNA that is discarded. Segments V and J join to form the coding junction region. Scheme built based on "The Immune System" book by Parham (2001).

Figure 1: Scheme of the V(D)J system and the role of RAG1 to RAG2 enzymes within that system



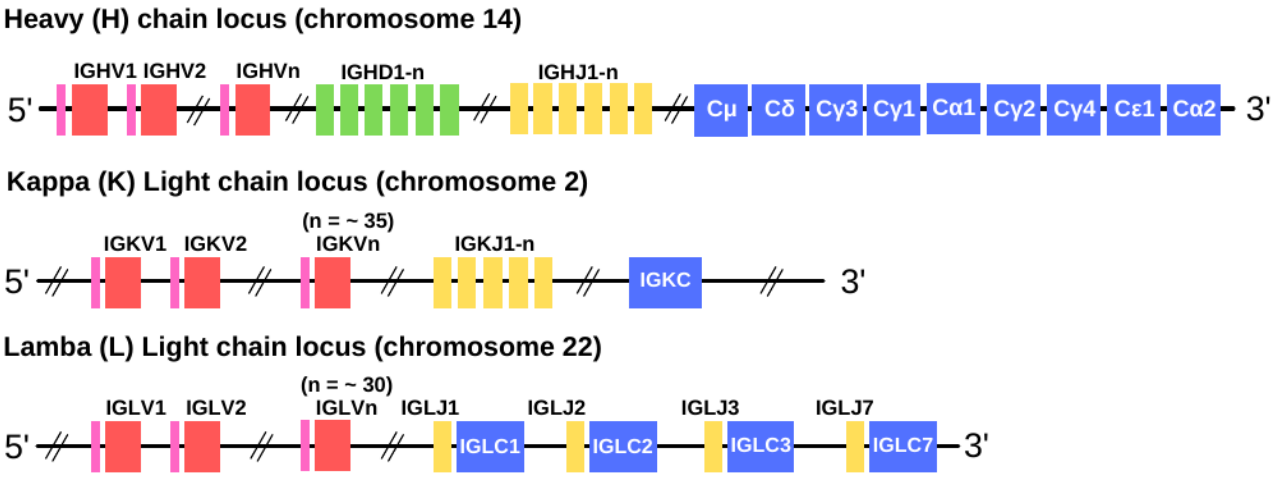
Source: Own design.

1.1.2. Immunoglobulins

Immunoglobulins (IG), also called antibodies, have been studied in many vertebrates (Du Pasquier, 1989), and they were first found in cartilaginous fish, such as sharks, skates, rays, and chimeras (Flajnik and Kasahara, 2010). Immunoglobulins carry four polypeptide chains that form a Y-shaped molecule when assembled. These chains are divided into two identical heavy chains (IGH) and two identical light chains (IGL). An IGH is composed of one variable (V) and three or four constant (C) domains, and an IGL is composed of one V and one C domain (Figure 2). Antibodies can recognize microbial antigens and toxins through their V domains (Abbas et al., 2015). The IGL, which will be our focus in Chapter 2, is divided into the kappa and lambda chains. Each chain defines the immunoglobulin types kappa (K) and lambda (L), respectively (Milsten and Pink, 1970). Both chains can be associated with heavy chains (Hobbs et al., 1966; Hansson et al., 1966). The genomic organization of tetrapods' IGH and IGL loci are in tandem with duplicated V genes located in a chromatin domain, followed by similarly arranged diversity (D) genes and joining (J) genes in the IGH loci, and only by J genes in the IGL loci (Sun et al., 2013) (Figures 2 and 3).

The Figure 2 germline organization of antigen receptor gene loci. Coding segments (exons) are shown as colored blocks of different sizes. These blocks are separated by segments that are not expressed (introns) shown as black lines. Leader sequences are shown as pink rectangles. The diagrams illustrate the antigen receptor gene loci in humans. The sizes of the segments and the distances between them are not drawn to scale. D, Diversity; J, joining; V, variable. Scheme built based on the "Basic Immunology" book by Abbas et al. (2015).

Figure 2: Germline organization of antigen receptor gene loci



Source: Own design.

1.1.2.1. Light Chains

Genome projects and model vertebrate species studies have shown that the IGL isotypes - kappa (K) and lambda (L) - emerged in the common ancestor of all living vertebrates (Greenberg et al., 1993; Rast et al., 1994). Three IGL genes were identified in the cartilaginous fish, called types I, II, and III (Greenberg et al., 1993; Sun et al., 2013). The shark V region of the type III light chain resembles the mammalian kappa light chain. Yet, types I and II could not be classified as either kappa or lambda chains (Greenberg et al., 1993). The IGL organization in bony fishes is an enigma as different patterns of IGL genes appear in different classes of bony fish. Teleosts IGL organization resembles more cartilaginous fish than ray-finned fish (Edholm et al., 2011; Guselnikov et al., 2018; Smith et al., 2019). In contrast, the IGL organization in a more primitive ray-finned fish, the sturgeon, has an IGL chain locus that looks like a mammalian kappa locus (Lundqvist et al., 1996). Amphibians have an additional IGL chain called the ômega (σ) chain and two chains homologous to the mammalian kappa and lambda chains. Thus, the three types of light chain identified in frogs are the lambda chain, homologous to the mammalian lambda chain; the rho chain (ρ), homologous to the mammalian kappa chain; and the σ chain, which has no apparent mammalian counterpart (Schwager et al., 1991). Little is known about the IGL chain in reptiles. Although, two distinct light chain isotypes have been identified in alligators (Wang et al., 2016). As for birds, a single Ig light chain locus has been identified containing

one or two functional VL gene segments with a single JL segment closely linked to it, followed by a single CL exon L-like (Thompson and Neiman, 1987; Das et al., 2010).

Light chain genes are usually found in two separate loci found in almost all mammalian species (Sun et al., 2013). Mammals' IGLV genes can be grouped into six clans, further divided into ten subgroups (Sitnikova and Nei, 1998; Sitnikova and Su, 1998). In most species, preferential use of one kind of IGL isotype can be noticed over another. In mice, for example, there are many VL gene segments and only three functional VK segments, resulting in a primary use of the VL segments (L:K, 95%:5%) (Almagro et al., 1998). Most of the knowledge about mammalian IG was based on placental mammals' studies (Sun et al., 2013). Studies have demonstrated that the placental mammals that express relatively restricted VH repertoires also seem to express limited VL repertoires (Saini et al., 2003; Sun et al., 2010). Mammals' IGL chain genes are organized in a cluster of IGKV genes found 5′ of a small number of IGKJ genes, with the IGKJ gene cluster located 5′ of a single IGKC gene (Ermert et al., 1995; Kirschbaum et al., 1998; Martinez-Jean et al., 2001; Schwartz et al., 2012; Schwartz et al., 2018).

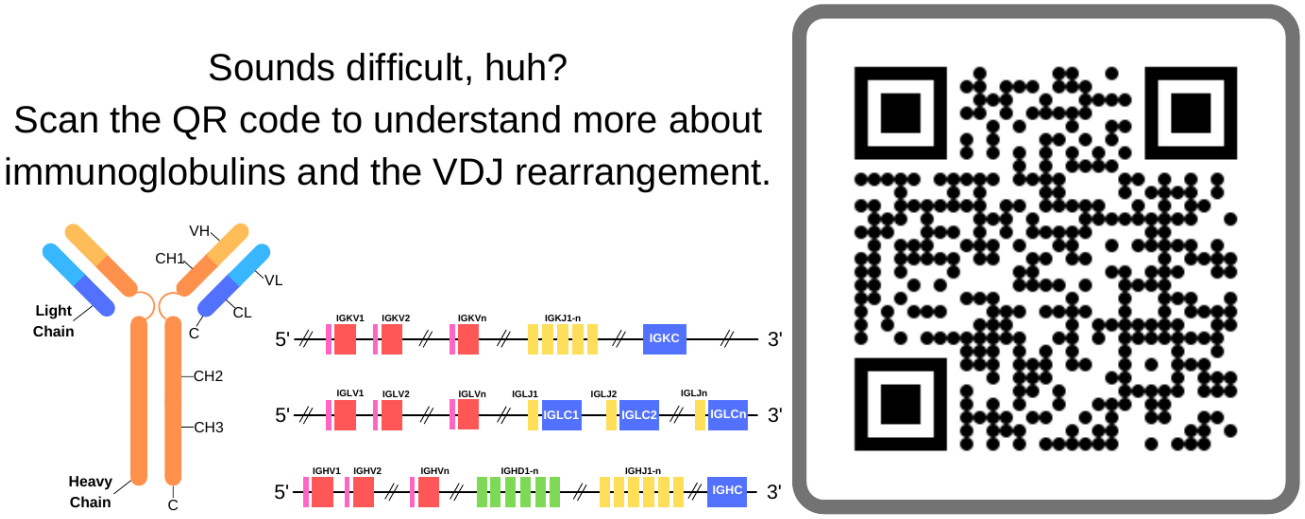
1.1.2.2. Heavy Chains

The presence of IGH chains dates back to cartilaginous fish with a single IgM-like antibody and two additional IGH chain genes called IGW (Greenberg et al., 1996). In bony fishes, the IGH gene's organization is much more dispersed, starting to resemble more mammals' organization, and RAG-mediated rearrangement becomes more important (Rast et al., 1998). Amphibians have an IGH locus organization similar to mammals (Wilson et al., 1992). Whereas in reptiles, studies are far behind compared to amphibians or mammals. Yet, transcripts were found in turtles that could produce IGM or IGY antibodies (Turchin and Hsu, 1996).

Mammalian IGH chain repertoire involves a single V gene rearranging to the downstream D and J segments (Tonegawa, 1983). Tens to hundreds of V genes in the mammalian repertoire can be separated into three clans based on sequence similarity (Kirkham et al., 1992). Most mammals studied to date have germline V segment representatives for each clan, although a few species lack V segments from one or two clans (Zhao et al., 2009). As examples, we have cows, sheep, horses, and manatees that lack Clan

III V segments, and they appear to be coupled with low numbers of V segments (Dufour et al., 1996; Niku et al., 2012; Walther et al., 2015; Breaux et al., 2017).

Figure 3: QR code linking to an educational video about how the Adaptive Immune System creates antibody diversity[1].



Source: Own design.

1.2. The Study Models

Tetrapods' transition from land to water was characterized by great morphological and physiological transformations and is a classic example of the occupation of new environments (Hoelzel, 2009). Colonization of the aquatic environment repeatedly occurred in phylogenetically distant tetrapod lineages. In mammals, three extant lineages have returned independently to the sea: cetaceans (whales and dolphins), pinnipeds (sea lions, seals, and walruses), and sirenians (manatee) (Hoelzel, 2009). However, gaps on the genetic basis of evolutionary transition to the aquatic environment still exist, despite many efforts that have been made in this direction in the last years (McGowen et al. 2012; Nery et al. 2013a,b,c; Zhou et al. 2013; Yim et al. 2014; Nery et al. 2014).

Another interesting transition that occurred more recently was the secondary radiation of marine mammals to the fluvial environment or aquatic river mammals to the marine environment. This recolonization happened several times and independently within different genera of cetaceans and sirens (Berta et al. 2006). As examples of these marine mammals, we

[1] You can enable the video caption for English subtitles.

have dolphins from the *Sotalia* genera that can be found both in the marine environment, specifically along the Central Atlantic and South America coasts (the Guiana dolphin, *Sotalia guianensis*), and in the freshwater environment along the Amazon River and its tributaries (the tucuxi dolphin, *Sotalia fluviatilis*) (Carr & Bonde 2000; Cunha et al. 2005; Caballero et al. 2007). Other examples are the manatees from the *Trichechus* genera.

The Order Sirenia belongs to the Afrotheria clade and includes three manatee species, the West Indian (*Trichechus manatus*), the Amazonian (*T. inunguis*), and the African (*T. senegalensis*) manatee that inhabit the Caribbean and Atlantic coasts, the Amazon River drainage, and the West African coasts, respectively (Domning, 1982). In Brazil, there are two manatee species: *T. manatus*, the marine manatee that inhabits the Atlantic coasts, and *T. inunguis*, the freshwater manatee, endemic to the Amazon Basin (Catanhede et al. 2005). Some earlier studies have quantified the robustness of the manatee immune system (Bossart, 1995; McGee, 2012; Sweat et al., 2005; Walsh et al., 2003, 2005), but the specific antigen-binding diversity of the adaptive immune receptors has not been described.

There is little data on the lymphocyte antigen receptors of the Afrotherian branch of mammalian evolution, making manatees an interesting non-model species to study (Guo et al., 2011). Extant marine mammal species descended from terrestrial ancestors that recolonized the marine environment (Uhen, 2007; Meredith et al., 2011). Marine mammals present unique adaptations to the aquatic environment (Bik et al., 2016). Studies have shown that the most basal split within eutherian mammals might have been between Afrotheria and all other placental mammals, which makes the clade Afrotheria, a possible representative of important radiation in mammalian evolution (Foley et al., 2016; Murphy et al., 2001, 2004; Springer et al., 2007).

This work aims to deepen our knowledge about the adaptive immune system of aquatic mammals through an evolutionary perspective, considering different environmental transitions that occurred during their evolutionary history. In Chapter 2, considering the generally conserved nature of RAG1 and RAG2 genes, we carry out an exploratory molecular study of these genes focusing on the primary transition of cetaceans from the terrestrial to the marine environment under an evolutionary perspective. Finally, in Chapter 3, considering that little is known about the manatee light chain immunoglobulins' organization and repertoire, we describe the organization of these loci in two newly sequenced genomes of *Trichechus*

manatus and *Trichechus inunguis* and analyze the IGL genes under an evolutionary perspective.

CHAPTER 2

2. ANALYSES OF RAG1 AND RAG2 GENES SUGGEST DIFFERENT EVOLUTIONARY RATES IN THE CETACEA LINEAGE

2.1. Introduction

The adaptive immune system has only become possible after the acquisition of a retroposon that, millions of years ago, invaded the genome of an early vertebrate, as only vertebrates have both elements of the retroposon: two sites of recognition signal sequences (RSS) and the presence of recombination-activation genes (Janeway, 2001). These genes, known as RAG1 and RAG2, encode a site-specific recombinase that acts on germline gene segments to produce all antibody molecules and T cell receptors of the adaptive immune system (Janeway, 2001).

The jawed vertebrate adaptive immune system relies on a diverse array of immunoglobulins (Ig) and T-cell antigen receptors (TCRs) for specific recognition of antigens (Agrawal et al., 1998; Teng and Schatz, 2015). This system consists of cells that provide pathogen-specific immunity to the host through somatic rearrangement of antigen receptor genes (Thompson, 1995). Each Ig or TCR polypeptide consists of a constant region and a variable region, and it is in the variable region that the antigen recognition and specificity are determined (Schatz, 2004). In the germline, the variable region is encoded by noncontiguous gene portions split into V (variable), J (joining), and, in some cases, D (diversity) segments. Each of these gene segments is joined in a site-specific recombination reaction, known as the V(D)J recombination, to form the exon that encodes the antigen-binding portion of the polypeptide (Agrawal et al., 1998; Schatz, 2004).

The discovery of RAG genes is considered a hallmark of adaptive immunity, as their proteins constitute the enzymatic core of V(D)J recombination machinery (Fugmann et al., 2006; Kapitonov and Koonin, 2015; Poole et al., 2017). The RAG1-RAG2 complex catalyzes random assembly of V, D, and J gene segments present in the genome in numerous copies and generates an enormous variety of the assembled antibodies and antigen receptors (Gellert, 2002; Kapitonov and Koonin, 2015). RAG1 and RAG2 genes are closely linked in

the human genome, located just eight kilobases apart from one another, and they are convergently transcribed (Oettinger et al., 1990; Greenhalgh et al., 1993).

Among jawed vertebrates, mammals possess great ecological diversity, and their complex evolutionary history associated with their radiation to different environments might have depicted many distinctly pathogenic challenges from different habitats (Tian et al., 2018). The aquatic environment was one of these habitats occupied by different mammalian lineages. Cetaceans comprise a mammalian order of fully aquatic mammals that originated about 50 million years ago in the Eocene epoch from a complete terrestrial ancestor (Thewissen et al., 2009). In addition to anatomical and physiological innovations required for life in water, cetaceans must have been confronted with challenges from changing environmental pathogens while they transitioned from land to sea (Shen et al., 2012; Ishengoma and Agaba, 2017). These challenges exerted intensified selection pressure on the genomes of colonizing species, especially on those genes and gene families related to the immune system, as already reported (e.g., Haldane, 1949; Wlasiuk and Nachman, 2010; Areal et al., 2011; Ishengoma and Agaba, 2017).

In this context, cetaceans comprise the ideal candidate taxa to study the molecular evolutionary mechanisms behind the vertebrate immune system since this lineage has experienced a great habitat change during its evolutionary history. Accordingly, this study aimed to investigate the molecular evolution of RAG1 and RAG2 genes in a phylogenetic framework, exploring the possible role of natural selection acting on these genes focusing on cetaceans compared to their terrestrial counterparts.

2.2. Material and Methods

2.2.1. Phylogenetic Reconstruction

Vertebrate RAG1 and RAG2 coding sequences were retrieved from GenBank (NCBI) and Ensembl public databases. All accession numbers are depicted in Supplementary Table 1. A total of 49 terrestrial vertebrates, seven other marine mammals, and eight cetacean sequences were selected to perform the analyses. Nucleotide and amino acid sequences were aligned using MUSCLE software (Edgar, 2004). We used the PAL2NAL program to generate a codon alignment, and this alignment was used to estimate the type and rate of nucleotide

substitutions in coding DNA (Suyama et al., 2006). The phylogenetic tree was reconstructed using IQ-TREE (Trifinopoulos et al., 2015) software using a maximum likelihood approach and implemented the fast method to substitution model selection with the ModelFinder program (Kalyaanamoorthy et al., 2017). The ultrafast bootstrap approximation (UFBoot) (Minh et al., 2013) with 1,000 replicates also was implemented in the IQ-TREE software package. The models MGK+F3X4+G4 and GY+F1X4+G4 were used for RAG1 and RAG2, respectively, as selected by the Bayesian Information Criterion (BIC).

2.2.2. Selection analyses

To test for signatures of positive selection and to infer sites under positive selection, we used the codeml program in PAML 4.7 package (Yang, 2007) and FITMODEL program (Guindon et al., 2004), which implements Markov-modulated Markov models of codon evolution or switching site evolution models on RAG1 and RAG2 genes separately.

On PAML, we applied branch models, which allow the ω ratio to vary among branches in the phylogeny and are useful for detecting positive selection acting on particular lineages (Yang, 1998; Yang and Nielsen, 1998). Within branch models strategy, we estimated the likelihood of the free-ratio model, which assumes an independent ω (=dN/dS) ratio for each branch; the one-ratio model, which estimates a unique ω value for all branches along the tree; and the two-ratio model, which estimates one ω for the Cetacea lineage and another for the rest of the phylogeny. Also, we tested the branch-site model that attempts to detect positive selection that affects only a few sites along a specified branch (Zhang et al., 2005). The branch-site analysis divides the tree into foreground branches (Cetacea), where sites may be under positive selection, and background branches where positive selection is absent (the rest of the phylogeny) (Yang and Nielsen, 2002; Yang et al., 2005; Zhang et al., 2005). Under this model, sites are categorized into four classes 0, 1, 2a, and 2b. Site class 0 includes codons that evolve under purifying selection on the foreground and background branches, with $0 < \omega 0 < 1$. In site class 1, codons evolve neutrally in all lineages, with $\omega 1 = 1$, whereas in classes 2a and 2b, positive selection is allowed on the foreground branches with $\omega 2 > 1$, but not on the background branches. This model is compared with the corresponding null hypothesis of neutral evolution, where ω2 is fixed to 1. If the null hypothesis is rejected by the likelihood ratio test (LRT), a Bayes empirical Bayes approach is used to calculate the posterior

probabilities that each site has evolved under positive selection on the foreground lineage (Yang et al., 2005).

To further investigate the molecular evolution of RAG1 and RAG2 genes, we performed likelihood analyses under a nested set of codon-substitution models with FITMODEL version 0.5.3 (Guindon et al., 2004). We used models M0 and M3 and the switching models M3+S1 and M3+S2. Model M0 assumes that all sites in a sequence alignment are subject to the same selection process (homogeneous). As implemented in FITMODEL, under the M3 model, variation in selective constraint across sites is modeled as three rate ratio classes with ω1, ω2, and ω3. Switching was modeled as a time-reversible Markov process with three additional parameters: the overall rate of interchange among rate ratio classes (δ), a coefficient for shifts between ω1 and ω3 (α), and a coefficient for changes between ω2 and ω3 (β). The S1 model implemented in FITMODEL imposes equal switching rates among ω1, ω2 and ω3 rate ratio classes ($\alpha = \beta = 1$), and the S2 model allows α and β to vary freely, accounting for unequal rates of switches between selection classes (Guindon et al., 2004).

For both PAML and FITMODEL, nested likelihood ratio tests (LRTs) were performed for model comparisons. For PAML models, the LRTs were performed between the free-ratio model vs. the one-ratio model and the two-ratio model vs. the one-ratio model (Table 2). For FITMODEL, the comparisons were performed between the following models: no rate heterogeneity vs. variation across sites (M0 vs. M3), variation across sites without vs. with switching among substitution rate ratio classes (M3 vs. M3+S1), and equal switching rates vs. class-dependent switching rates across branches (M3+S1 vs. M3+S2) (Table 2). The chi-square test was employed to estimate the statistical difference ($P < 0.05$). Degrees of freedom for each test equal the difference in the number of parameters estimated for the models under comparison.

Table 1: Likelihood analyses of the branch models and branch-site models in the PAML program for RAG1 and RAG2[2]

PAML							
BRANCH MODEL							
	Model	**lnL**	**ω**	**np**	**LRT**	**df**	**P-value**
RAG1	One-ratio model	-57620.68	0.09074	126	--	--	--
	Free-ratio model	-57221.58	various	249	798.2	123	**0**
	Two-ratio model	-57599.56	0.08868; 0.22258	127	42.24	1	**0**
RAG2	One-ratio model	-26095.59	0.13293	126	--	--	--
	Free-ratio model	-25925.59	various	249	340	123	**0**
	Two-ratio model	-26095.29	0.13250; 0.16241	127	0.6	1	0.43
BRANCH-SITE MODEL							
	Model	**lnL**	**np**	**LRT**	**df**	**P-value**	
RAG1	Null Model	-56293.55	128	--	--	--	
	Model A	-56293.55	129	0	1	1	
RAG2	Null Model	-25656.34	128	--	--	--	
	Model A	-25655.46	129	1.76	1	0.18	

Source: Own data.

Additionally, we used the HyPhy package (Pond et al., 2005) in the DataMonkey Server (Weaver et al., 2018) to implement the RELAX model (Wertheim et al., 2014), which tests whether the strength of natural selection has been relaxed or intensified along with a specified set of test branches.

Also, it is important to consider that GC-biased gene conversion (gBGC) is a recombination-associated process that causes variation in GC content and has an important influence on substitution patterns, leading to sequence accelerated evolution (Gaultier et al., 2009). To test for gBGC on our alignment, we implemented the program phastBias (Capras et al., 2013) available as part of the PHAST software package (Hubisz et al., 2011) that uses a

[2] Abbreviations of table's cells are as follows: likelihood value (lnL), omega value (ω), number of parameters (np), likelihood ratio test (LRT), degrees of freedom (df), and the P values.

hidden Markov model and statistical phylogenetic models that consider the influence of both natural selection and gBGC on substitution rates and patterns.

2.3. Results

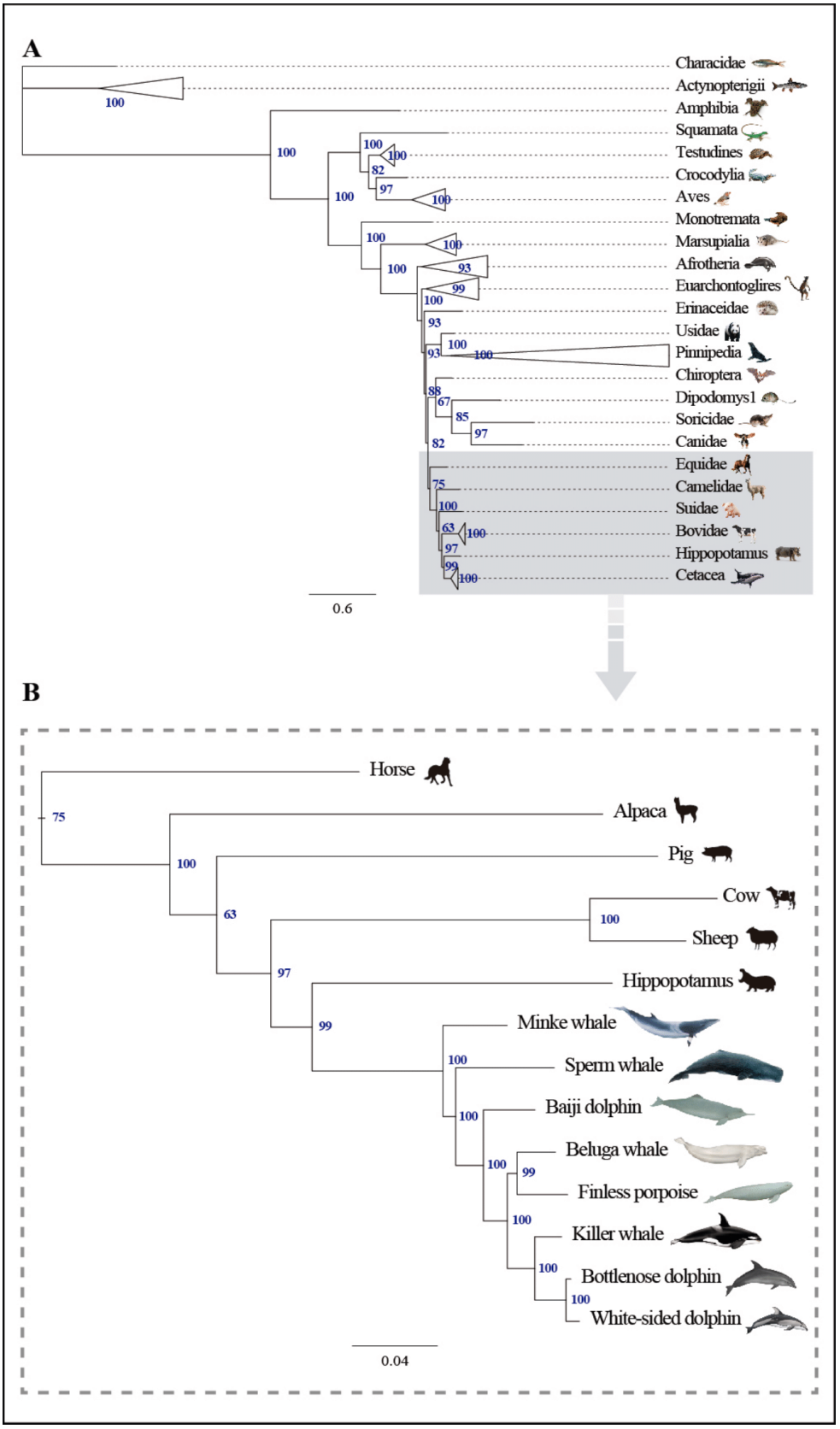
2.3.1. Phylogenetic Analyses of RAG1 and RAG2 Genes

RAG1 and RAG2 phylogenetic trees resulted in similar topologies, with occasional differences in some phylogenetic relationships. In both trees (Figures 4 and 5), mammals form a monophyletic clade, with the lineage of cetaceans being part of Cetartiodactyla and the sister clade of Hippopotamidae. In RAG1, both cetaceans and bovids clade are closer to Camelidae, while in RAG2, they are closer to Suidae. Tree lengths indicate that RAG1 and RAG2 experienced accelerated evolutionary rates in the ancestral lineage of all tetrapods after their separation from fishes. The lengths of tree branches also show that both genes had great accelerated evolutionary rates in the Pinnipedia lineage, with most evolutionary modifications happening in *Arctocephalus* for both genes (Figures 4 and 5).

The figure 4 is about Phylogenetic relationships were recovered through maximum likelihood analyses from the molecular data of the RAG1 gene. (A) Phylogenetic relationships among all the species in the study. The gray chart represents the Euungulata clade. (B) Phylogenetic relationships at the species level of the Euungulata clade, with emphasis on cetaceans. The numbers in blue represent the bootstrap values that support each node of the phylogenies.

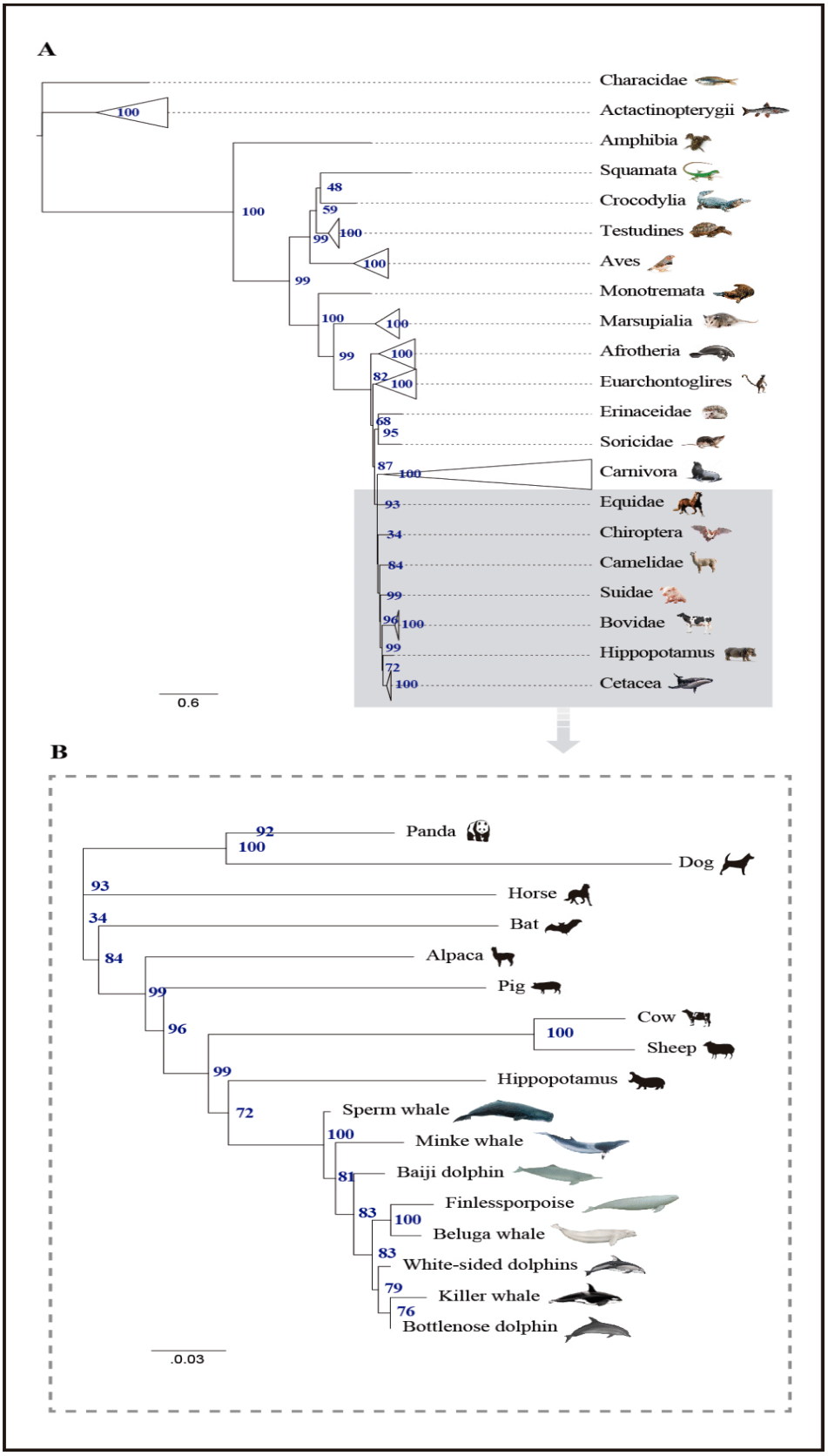
The figure 5 is about Phylogenetic relationships were recovered through maximum likelihood analyses from the molecular data of the RAG2 gene. (A) Phylogenetic relationships among all the species in the study. (B) Phylogenetic relationships at the species level of the Euungulata clade, with emphasis on cetaceans, and the insertion of microbats closely related to the Artiodactyls. The numbers in blue represent the bootstrap values that support each node of the phylogenies.

Figure 4: Phylogenetic relationships were recovered through maximum likelihood analyses from the molecular data of the RAG1 gene



Source: Own design.

Figure 5: Phylogenetic relationships were recovered through maximum likelihood analyses from the molecular data of the RAG2 gene



Source: Own design.

2.3.2. Identifying Sites Under Positive Selection

For RAG1, the model that best fitted our data on branch model was the two-ratio model, where Cetacea lineage has a greater ω value (0.22) when compared to the rest of the tree (0.08), suggesting that cetaceans accumulated more modifications on the RAG1 sequence throughout their evolutionary history (Table 2). The RELAX algorithm implemented on DataMonkey identified a significant relaxation in the selection pressure in the cetacean lineage for this gene, which could explain this evolutionary acceleration. For RAG2, the two-ratio model did not fit the data better. This result matches the result of RELAX, which did not identify a significant relaxation in the cetacean lineage compared to other lineages. The results from the branch-site model were not significant for both genes (Table 2), not being able to identify sites under positive selection in cetacean lineage.

2.3.3. Shift in the Site-Specific Selection Process

We carried out a maximum-likelihood analysis on FITMODEL software under a set of branch-site codon substitution models to investigate the substitution process on RAG1 and RAG2 genes. We performed likelihood analyses under a nested set of codon-substitution models (M3, M0, M3+S1, M3+s2) (Guindon et al., 2004). Table 2 shows that log-likelihoods improved significantly as parameters were added to the nested substitution models ($P < 0.001$). These results suggest that M3+S2 (unequal switching rates among three ω rate ratio classes) is the best codon substitution model for RAG1 and RAG2 genes. Under this model, the substitution rate ratio estimated for the three classes were $\omega1 = 0.0$, $\omega2 = 0.3$ and $\omega3 = 1.3$ (Table 2). The switching rate (represented by the R letters in table 2) between ω2 and ω3 (R23 = 1.18 for RAG1 and R23 = 0.70 for RAG2) was significantly higher than the switching rates between ω1 and ω2 and between ω1 and ω3 (Table 2). These results imply that site-specific shifts between moderate purifying selection (ω2) and relaxed selection (ω3) occurred more frequently than shifts that involve the most highly constrained rate ratio classes for both RAG1 and RAG2 genes. With an ω considerably lower than one for almost all the rate ratio classes, excluding ω3 for both genes, the sum of the parameter estimates (p1,

p2, and p3) values of the M3+S2 model suggests that most sites are under purifying selection for both genes (RAG1 = 80% and RAG2 = 90%; Table 2).

Table 2: Likelihood analyses and P values of the models in the FITMODEL program for RAG1 and RAG2 sequence data[3]

FITMODEL				
	M0	**M3**	**M3+S1**	**M3+S2**
RAG1				
lnL	-65937.20	-62608.04	-62441.29	-62381.36
ω1 ω2 ω3	0.3	0.0 0.3 0.9	0.0 0.3 1.1	0.0 0.3 1.3
p1 p2 p3	1.0	0.3 0.3 0.2	0.4 0.3 0.2	0.4 0.3 0.1
R12 R13 R23	--	--	0.15 0.15 0.15	0.19 0.00 1.18
P value	--	0	P <0.001	P <0.001
RAG2				
lnL	-31291.80	-30688.42	-30635.45	-30629.07
ω1 ω2 ω3	0.3	0.0 0.3 0.9	0.0 0.3 1.1	0.0 0.7 2.4
p1 p2 p3	1.0	0.3 0.3 0.2	0.4 0.3 0.2	0.1 0.5 0.3
R12 R13 R23	--	--	0.15 0.15 0.15	0.39 0.00 0.70
P value	--	0	P <0.001	P <0.001

Source: Own data.

No position in our alignment was identified with a probability of being in one of the gBGC states, thus rejecting the action of GC-biased gene conversion in RAG1/RAG2 genes in cetaceans.

2.4. Discussion

This study represents the first molecular evolution analyses of RAG1 and RAG2 genes focusing on cetaceans. The phylogenetic trees generated for both genes depicted short

[3] Abbreviations of table's cells are as follows: likelihood value (lnL), omega value (ω), parameter estimated values (p), switching rates values ®, and P values.

branches among all cetacean species (Figures 4 and 5). On both trees, the phylogeny is fully resolved for Cetacea. It forms a definite clade for mammals, with the lineage of cetaceans being part of Cetartiodactyla and the sister clade of Hippopotamidae. Considering RAG1, the Cetartiodactyla clade is better supported in relation to RAG2 phylogeny. For the RAG1 gene tree, cetaceans and bovids are closer to camelids. This result does not support previous phylogenetic studies using the RAG1 gene, which places cetaceans and bovids closer to suids (Waddell and Shelley, 2003). On the other hand, our RAG2 phylogenetic analyses show cetaceans and bovids closer to suids than camelids, which is the most accepted position of this phylogeny considering previous phylogenetic studies (e.g., Waddell and Shelley, 2003; Gatesy et al., 2013). In a previous study of extant cetaceans, including the RAG1 gene, the topology derived from Bayesian analyses of nucleotide resembled the Cetacea topology generated by our analyses, with the same relationship found between whales and dolphins (McGowen et al., 2009).

The ω ratio is a measure of natural selection acting on the protein. In short, values of $\omega < 1$, $= 1$, and > 1 indicate negative purifying selection, neutral evolution, and positive selection respectively. However, the ratio averaged over all sites, and all lineages is rarely > 1 since positive selection is unlikely to affect all sites over a prolonged time (Yang, 1997). Our PAML results for RAG1 indicated a statistically higher ω value for cetaceans (0.22) when compared to the rest of the phylogeny (0.08), which means that this lineage accumulated more modifications throughout its evolutionary history, i.e., the gene on this branch experienced an acceleration in the rate of evolution. This result corroborates with the result from the RELAX model that identified a relaxation in the purifying selection pressure for RAG1 in the cetacean lineage, suggesting that such relaxation may have accelerated the evolutionary rate in cetaceans, as seen with their higher ω value. Taken together, the results corroborate a relaxation on the selective pressure of the RAG1 gene rather than a stronger and direct action of positive selection. For RAG2, the two-ratio model did not fit the data better. This result agrees with the RELAX result, which did not identify a significant relaxation along the cetacean lineage compared to other lineages. Other lineages of our set of species may have experienced an acceleration in the evolution rate of this gene since the free-ratio model better fits the data than the one-ratio model.

On FITMODEL, a posterior probability greater than 90% for selection class ω3 is considered to have evolved under relaxed selection. In contrast, a posterior probability lower than 20% for selection class ω3 is considered to be subject to purifying selection. For RAG1, under the M3+S2 model, only one site (270) presented ω3 posterior probability of 95% and thus, was identified as evolving under relaxed selection through the phylogeny. The core RAG1 domain contains the nonamer binding domain (NBD) and two domains, namely the central with zinc finger B (ZFB) region and C-terminal domains. Site 270 is located in the CDN region, which possesses sequence signals critical for nuclear localization, zinc coordination, and interactions with nucleic acid, and is conserved from the sea urchin to human (Arbuckle et al., 2011; Kumar et al., 2015). Mutations in the first 90 amino acids of RAG2 severely inhibit basic recombination reaction, formation of signal joints by deletion, and formation of both signal and coding joints by inversional recombination (Cuomo and Oettinger, 1994). The first 90 amino acids of RAG2 sequences on our FITMODEL analyses were conserved and had no sites evolving under relaxed selection. A total of 24 sites were under purifying selection (ω3 < 20%), and the rest had posterior probabilities ranging from 21% to 89%, whereas the first 63 amino acids had the same ω3 values (33%), confirming their conservation. A total of four sites had a posterior probability greater than 90% for ω3 in RAG2 analyses (113, 124, 397, 428). Site 428 is found in the C-terminus of the protein (Cuomo and Oettinger, 1994). None of the estimated values for both genes were near 1, which would suggest neutrality.

Although RAG1 and RAG2 genes are highly conserved, previous comparisons of RAG2 amino acids in frogs, mammals, and chickens indicated that RAG2 is less conserved than RAG1 (Greenhalgh et al., 1993). Our FITMODEL analyses for RAG2 from all tetrapods confirm this statement, considering the four sites found to be under relaxed selection compared to one site of RAG1.

RAG1 appears to have originated from a TE of the Transib family. It can mediate low levels of recombination in the absence of RAG2 since this gene contain all the essential domains and activities needed to bind and cleave DNA, placing RAG2 in the role of an accessory or regulatory factor (Ji et al., 2010; Teng et al., 2015; Carmona et al., 2016). Despite this RAG1 ability, it is still proposed that RAG2 plays a critical role in the establishment and evolution of V(D)J recombination (Carmona et al., 2016). Furthermore, it

is believed that, when acquiring a RAG2-like element, both genes were able to provide functional advantages that allowed for the evolution of the adaptive immune system of early jawed vertebrates (Carmona et al., 2016). Thus, despite being closely linked in the genome and working together to generate diversity in the adaptive immune system, RAG1 and RAG2 genes have different evolutionary origins, function separately, and thus, may have different evolutionary rates.

Drastic environmental changes such as the transition from a terrestrial to marine habitat should select for numerous evolutionary adaptations (Chikina et al., 2016). The oceans harbor an enormous diversity and number of viruses and prokaryotes, which could frequently become a threat to marine mammals (Whitman et al., 1998; Suttle, 2007). The abundance of viruses exceeds bacteria and archaea by approximately 15-fold (Bettarel et al., 2000). Estimates of cell density, volume, and carbon indicate that prokaryotes are ubiquitous in marine environments (Ducklow and Carlson, 1992; Simon, 1994). The RAG1 gene structure is not conserved in fishes. Still, it is highly conserved among tetrapods (Kumar, 2015). One could hypothesize that this gene behaves differently in different environments, being less evolutionarily conserved in the aquatic environment, considering the greater number and diversity of pathogens in this habitat.

2.5. Conclusion

In summary, our findings demonstrate that RAG1 and RAG2 genes remained relatively conserved among tetrapods, with purifying selection happening on both RAG1 (80% of sites under purifying selection) and RAG2 (90% of sites under purifying selection) genes, and evidence for a few punctuated shifts in nucleotide substitution rates of both genes along with tetrapod evolution. When considering only cetacean lineage, the RAG1 gene shows an accelerated rate of evolution with a relaxation of the selective pressure. In contrast, for RAG2, this relaxation was not observed, and no specific acceleration on the evolutionary rate. These results demonstrate differential evolution happening in the closely linked genes RAG1 and RAG2 in cetaceans, with RAG1 being less conserved when compared to other mammals of the phylogeny. It is important to note that this is only a brief part of the whole story. Since RAG genes act on DNA substrates and a complex panorama of recombination signals, future work focusing on locus subject to the action of RAG genes, such as Ig and

TCR genes, will be necessary to clarify further how molecular evolution acts on immunological genes during the occupation of new environments.

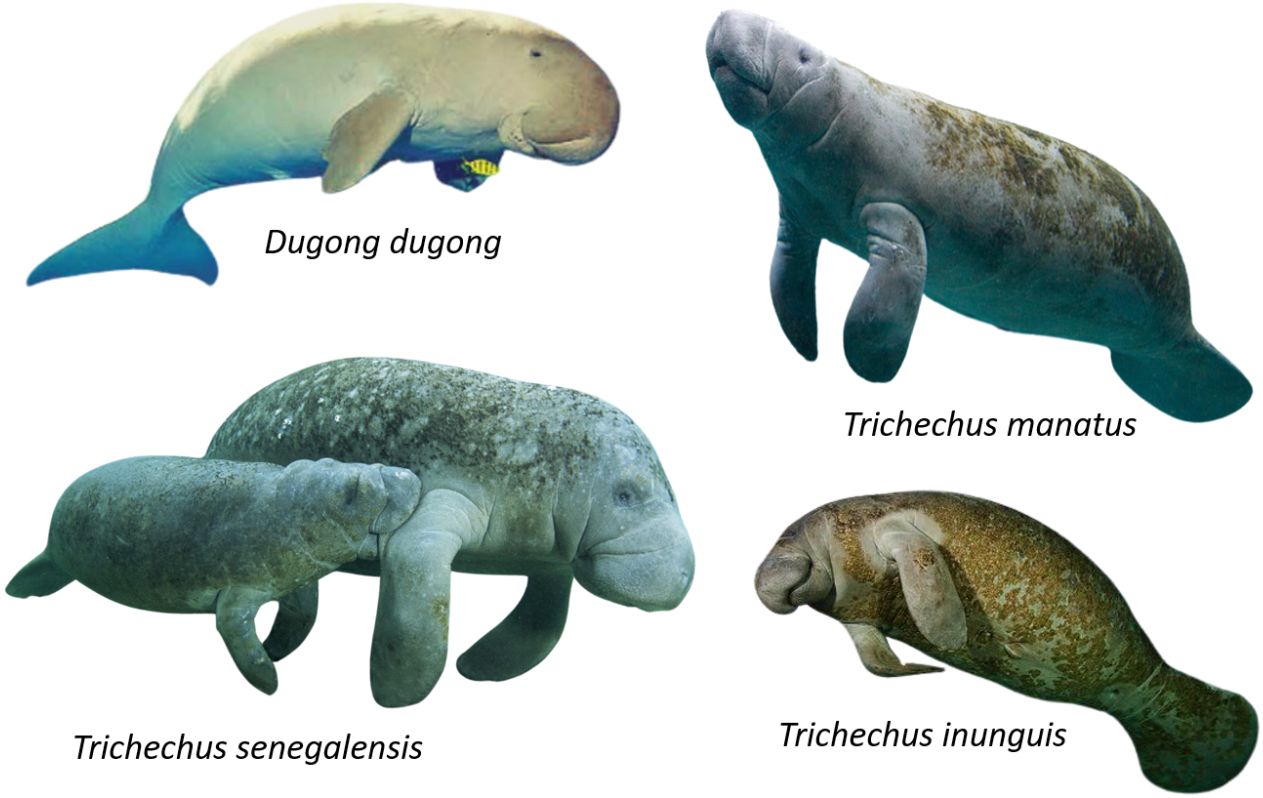
CHAPTER 3

3. GENOMIC ORGANIZATION AND NATURAL SELECTION ANALYSES OF THE IMMUNOGLOBULIN LIGHT CHAIN REGION IN THE AMAZONIAN (TRICHECHUS INUNGUIS) AND WEST INDIAN MANATEE (TRICHECHUS MANATUS)

3.1. Introduction

Sirenians comprise modern sea cows (manatees and dugongs) and are the only living herbivorous aquatic mammals. Their fossil record extends over 50 million years, and only two genera live today, the genus *Trichechus* and the genus *Dugong* (Figure 6) (Domning, 1982; Reeves et al., 2002; Domning, 2002).

Figure 6. Extant manatees from Trichechus and Dugong genera[4].



Source: Own design.

[4] Pictures were provided by the marine life photographer David Schrichte and the underwater photographer David Fleetham.

The genus *Trichechus* is primarily tropical in distribution. Three living species of manatees are distributed across shallow tropical and subtropical waters of America (*Trichechus manatus* and *Trichechus inunguis*) and the western coast of Africa (*Trichechus senegalensis*) (Berta et al., 2006). More specifically, the West Indian manatee (*Trichechus manatus*) is distributed from Florida (USA) to the northeast coast of Brazil (Berta et al., 2006). In contrast, the Amazonian manatee (*Trichechus inunguis*) occurs only in freshwater, being the smallest sirenian with its distribution limited to the rivers, lakes, and tributaries of the Amazon Basin (Reeves et al., 2002; Vianna et al., 2006) (Figure 7).

Figure 7. QR code linked to an educational video about manatees[5].

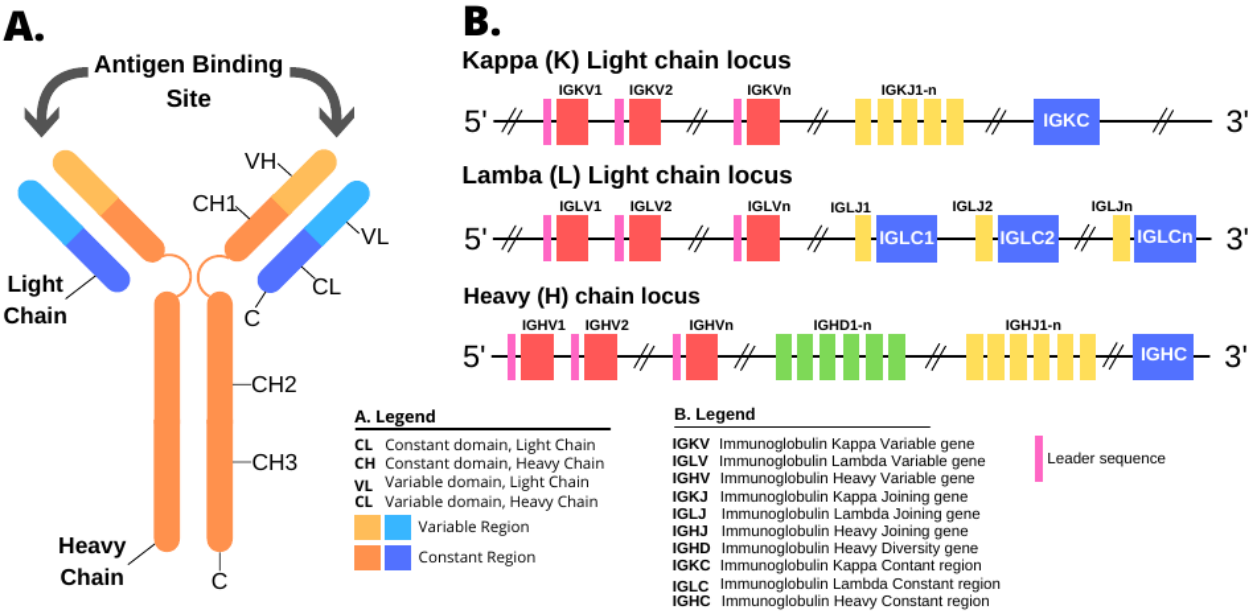
Source: Own design.

Immunoglobulins (IG), also called antibodies, have been studied in several vertebrates (Du Pasquier, 1989), and they are first found in cartilaginous fish, such as sharks, skates, rays, and chimeras (Flajnik and Kasahara, 2010). When secreted in the blood, immunoglobulins can recognize microbial antigens and toxins by their variable domains, just like the membrane-bound antigen receptors on B cells that recognize antigens to initiate the immune defense responses (Abbas et al., 2015). Immunoglobulins have a characteristic three-dimensional Y shape composed of four polypeptide chains: two identical heavy (H) chains and two identical light (L) chains. Each chain contains a variable (V) and a constant (C) region, composing together with the immunoglobulin (IG) domain (Abbas et al., 2015) (Figure 8A). Considering light chains (IGL), they comprise two types of chains that differ in

[5] You can enable the video caption for English subtitles.

their C regions: the kappa (K - IGKC) and the lambda (L - IGLC) chains. Unlike the heavy chains types, light chains have no functional differences, and each type can build complexes to any kind of heavy chain antibody molecule (Abbas et al., 2015) (Figure 8B).

Figure 8. Diagram of the immunoglobulin protein (A) and the genome structure of the kappa (K) and lambda (L) Light chains locus, as well as the Heavy (H) chain locus.



Source: Own design.

In mammals, germline variable light chain (VL) genes segments can be grouped into clans based on their nucleotide sequence similarity (Schroeder et al., 1990). A clan is a set of subgroups from different species that appear related to phylogenetic trees. For example, human and mouse variable genes of the heavy (IGHV) and kappa (IGKV) chains can be divided into three clans (I, II, III), while variable genes from the lambda (IGLV) chain can be divided into three and seven clans (I, II, III, IV, V, VI, VII) in mouse and humans, respectively. The 'clan' concept is part of the 'classification' concept of IMGT-ONTOLOGY (Giudicelli and Lefranc, 1999). The IMGT-ONTOLOGY describes the classification and specification of terms needed for immunogenetics and bioinformatics, which allows an extensive and standardized description and characterization of immunoglobulin data (Giudicelli and Lefranc, 1999).

Immune studies in manatees have thus far been limited, and immunological data for these species has focused mainly on correlations with cold stress syndrome and red tide exposure (Bossart et al., 1998; Walsh et al., 2004). However, regarding the manatee

immunoglobulin (IG) repertoire, data on baseline levels of circulating IgG in *T. manatus* is available (McGee, 2012), as well as a characterization of the *T. manatus* IG heavy chain locus (Breaux et al., 2017). Further, the Florida manatee is the first ungulate species to show evidence of the loss of all Clan III V segments (Breaux et al., 2017).

Before seeking further immune-related research and diagnostic studies, it is necessary to establish a basic definition and characterization of the immune system to correlate immune-related variables and evolutionary aspects. Furthermore, incorporating new immunological data allows more accurate assessments of health in endangered species, like manatees. Accordingly, for purposes of research, diagnostics, and health monitoring in manatees, here we report the genomic organization of the immunoglobulin light chain (IGL) locus of *Trichechus manatus* and *Trichechus inunguis*. Our results expand the findings by previous studies characterizing the manatee immunoglobulin heavy chain (IGH) (Breaux et al., 2017), and we report the identification of the V segments clans, selection analyses of the kappa and lambda chains constant regions, and the comparison of the IGL locus with other vertebrates, through an evolutionary perspective.

3.2. Materials and Methods

3.2.1. Sample Collection

Skin tissue was collected from a female *Trichechus inunguis* from the Mamirauá Institute in the Manacabi community from the Amazonian Japurá river, Brazil, in November of 2014. For *Trichechus manatus*, cookie tissue was collected from a male individual during wild capture in the Retiro Grande beach, Ceará, Brazil, in November of 2016. All samples had their respective licenses and followed the necessary ethics advice.

3.2.2. DNA Extraction and Whole Genome Sequencing

Genomic DNA was isolated from the manatees' samples through the Wizard® Genomic DNA Purification Kit protocol, with Phenol-Chloroform according to manufacture. Both genomes were sequenced through the Next Generation Sequencing (NGS) strategies from the Illumina platform in Macrogen (Korea). The genome from *Trichechus inunguis* was

sequenced using TruSeq DNA 350bp library, HiseqX 150bp Paired End, with three lanes of ~300Gb total, and Nextera Mate Pair (3kb, 8kb) libraries, Hiseq4000 100bp Paired End, eight lanes, and a total of ~550Gb. As a result, the genome has coverage of ~111.47x. *Trichechus manatus* re-sequencing used the TruSeq DNA 350bp library, HiseqX 150bp Paired End, with ~110Gb per lane. As a result, the genome has coverage of ~37.21x. The genomic libraries were assembled at the Central Laboratory of High-Performance Life Sciences Technologies (LaCTAD).

The quality control was made using FASTQC (Andrews, 2011) for each library. Next, mate-pair reads were trimmed using TRIMMOMATIC (Bolger et al., 2014). The quality of trimmed reads was processed using NXTRIM (O'Connell et al., 2015) to remove libraries' junction adaptors and classify the read pairs.

3.2.3. Identification of IGL genes in the manatee's genome and other mammals

The manatee immunoglobulin light chain (IGL) was identified through BLASTn within our server by using the African elephant (*Loxodonta africana*) immunoglobulin light chain variable (V) and joining (J) genes and constant (C) regions (Guo et al., 2011) as queries. We also used some previously annotated IG lambda chain V genes of the publicly available data from *T. manatus* to BLAST it against *T. inunguis*. Putative V segments and C regions were compared to other vertebrates' V and C domain alignment (Criscitiello and Flajnik, 2002). If the DNA segment was a potential V gene or C region, it was then BLASTed through Geneious 9.0.5 software (Biomatters, Ltd.) (Kearse et al., 2012) to check their reliability against the NCBI available data. We only considered DNA segments with 80 to 100% identity to annotation.

The *Recombination Signal Sequence Site* (RSS) (http://www.itb.cnr.it/rss/index.html) was used to predict recombination signal sequences to identify J segments, which were predicted based on a reverse orientation 23 RSS that was followed by either an FGXG or a WGXG motif. Joining (J) segments are short coding sequences between the variable (V) and constant (C) gene segments in all IG loci. Regarding the Heavy chain, J segments together with a D short coding sequence are somatically recombined with V segments during antibody

development. Only V and J segments are recombined in the Light chain to form the VDJ and VJ recombination DNA codes for the ends of the antigen receptor V regions. It is also believed that the J segments contribute to the diversity of the antigen receptor repertoire (Abbas et al., 2016).

To identify IGL genes in other mammals, we first retrieved immunoglobulin's V and C genes from nine terrestrial mammals from the international ImMunoGeneTics (IMGT) information system (http://www.imgt.org), including rabbit (*Oryctolagus cuniculus*), mouse (*Mus musculus*), human (*Homo sapiens*), monkey (*Macaca mulatta*), cat (*Felis catus*), dog (*Canis familiaris*), goat (*Capra hircus*), cattle (*Bos taurus*) and pig (*Sus scrofa*). We then used the IGLV (immunoglobulin light chain variable gene) and IGLC (immunoglobulin light chain constant region) sequences of these terrestrial relatives as a query to search for the orthologous genes in the genomes of 10 cetaceans, two rodents, two odd-toed ungulates, four carnivores, two artiodactyls, one xenarthran, six afrotherians, one bat, one scandentian and one Monotremata, with BLASTN approaches through NCBI, Ensembl, and UCSC Genome Browser. In addition, V and C genes from *T. inunguis* and *T. manatus* identified in this study were also included. The accession numbers of the sequences and the detailed information about the genome assemblies used are compiled in Supplemental Table 2.

3.2.4. Definition of the Variable genes clans

To identify IGKV and IGLV clans in *T. manatus* and *T. inunguis*, we used human and mouse IGKV and IGLV data from the IMGT® database, compiling genes representative of all the clans present on each genome. However, only three and four IGLV genes were identified in *T.. manatus*' and *T. inunguis*' genomes, respectively. Furthermore, none of those genes appears to be functional genes, as they all have a stop codon. Due to this reason, we only analyzed the clan's division happening on IGKV genes on both species.

We retrieved mouse and human IGKV sequences from the IMGT® database. Sequences representative of each clan were chosen at random among the functional genes. We also included IGKV genes from the elephant (*Loxodonta africana*) in the analyses (Guo et al., 2011).

3.2.5. Phylogenetic reconstruction

The number of different lambda chain constant regions among mammals is highly variable. For example, in humans, seven IGLC genes were identified (Kawasaki et al., 1997) in contrast with two IGLC genes found in goats (Schwartz J.C. et al., 2018). Nine IGLC genes were identified in cattle (Chen L. et al., 2008), two in sheep (Giudicelli V. et al., 2005), eleven in cats (Giudicelli V. et al., 2005), and four in rats (Steen, M.L. et al., 1987). For that reason, we chose the common constant region (IGLC1) of all of the species used in the study to reconstruct these animals' phylogeny. Amino acid alignments with a total of 35 and 38 species for immunoglobulins kappa (IGKC) and lambda (IGLC1) constant regions genes, respectively, were aligned using the Geneious (version 9.1.5) software and the MUSCLE alignment tool (Edgar, 2004). Phylogenetic trees for IGKC and IGLC1 genes were reconstructed through a Bayesian inference using the MrBayes3.1 program (Huelsenbeck and Ronquist, 2001) and a maximum likelihood inference through IQ-TREE (Trifinopoulos et al., 2016) and viewed with FigTree v1.4.4 software (Rambaut, 2018). Regarding MrBayes' analyses, the Generalised Time Reversible (GTR) evolutionary model was used with gamma-distributed rate variation across sites and a proportion of invariable sites. The analyses ran for 10,000,000 generations, sampling every 1,000, with four independent analyses running simultaneously, each starting from different random trees and a burn-in of 25% of the trees. Platypus was used as the outgroup for IGLC1 data, and armadillo as the IGKC data outgroup. The choice to use only IGLC1 was based on the fact that the number of constant regions present in the lambda locus amongst mammals has great variability, and IGLC1 is present in almost all lineages.

A maximum-likelihood tree of the mouse, human, and elephant IGKV genes, together with potentially functional IGKV gene segments from *T. manatus* and *T. inunguis,* were reconstructed through IQ-TREE to study the grouping of variable genes clans (Trifinopoulos et al., 2016) and viewed with FigTree v1.4.4 software (Rambaut, 2018).

3.2.6. Selection analyses

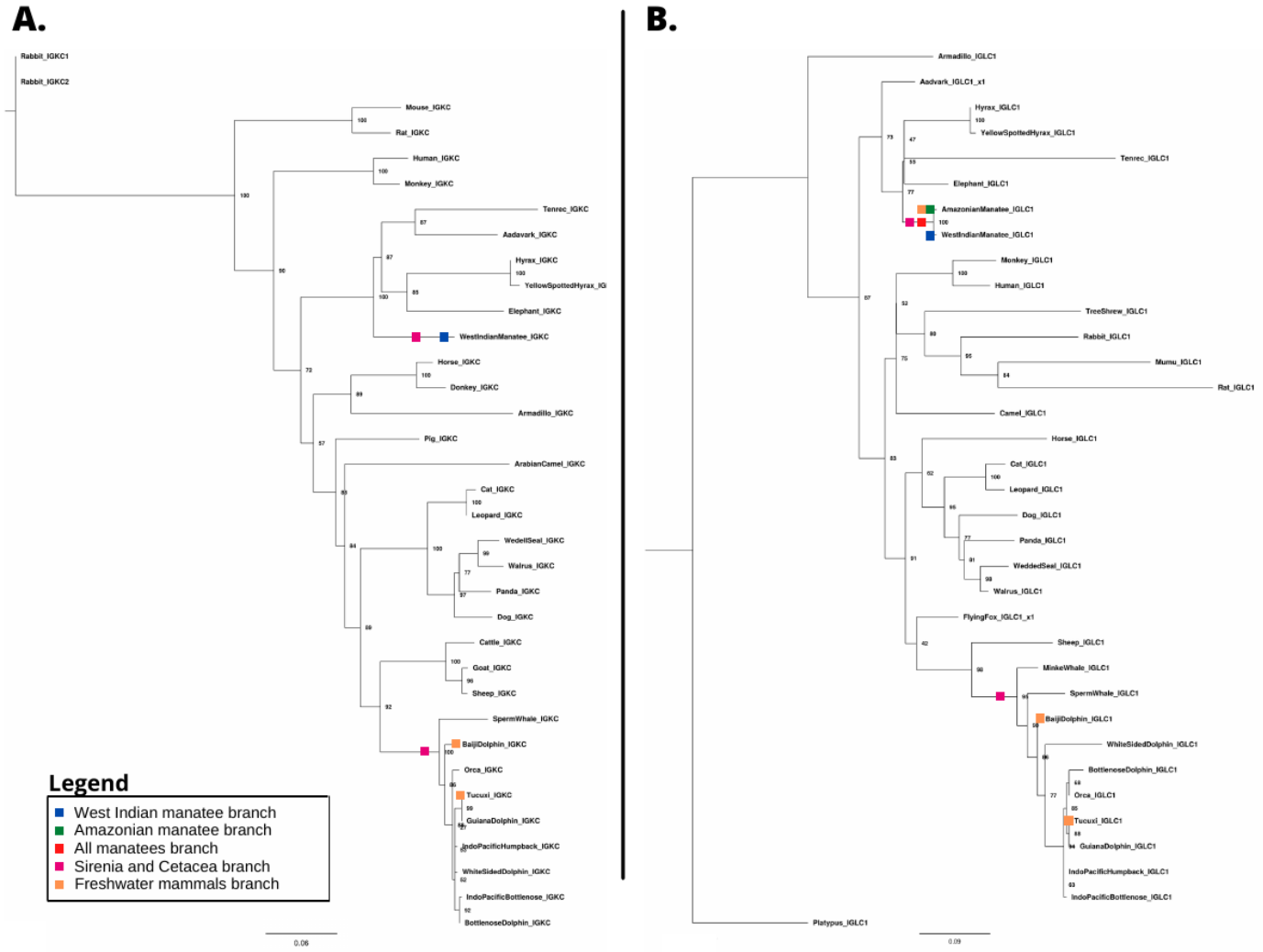
In the natural selection analyses, the same alignments used for phylogenetic reconstruction were used as datasets. As the variable regions of the immunoglobulins are in constant change to create diverse antibodies to recognize different and many antigens, selection analyses on these genes could be superfluous and lead to frail results. Thus, we conducted natural selection analyses on both kappa and lambda constant regions in the IGL locus. We used the PAL2NAL program (Suyama et al., 2006) to generate a codon alignment, and this alignment was used to estimate the type and rate of nucleotide substitutions in coding DNA. To explore patterns of natural selection and identify sites targeted by positive selection on IGLC1 and IGKC genes, we investigated variation in the omega (ω) ratio (*dN*/*dS*, where *dN* is the non-synonymous substitutions rate and *dS* is the synonymous substitutions). The ω ratio measures the direction and intensity of natural selection, with $\omega=1$, $\omega>1$ and $\omega<1$ indicating neutral evolution, positive and purifying selection, respectively. We performed statistical tests based on the ω values using the Godon software (Davydov et al., 2019). Godon is a classical codon model software that allows detecting a robust signal of positive selection affecting most positions over a long period (Davydov et al. 2019). In addition, the software has a heuristic algorithm to avoid LRT statistics overestimation - which often causes false positives in the PAML software (Yang, 2007) –, also correcting for LRT underestimation. Within Godon, we ran the M0 model (Goldman and Yang, 1994), which assumes that all sequence alignment sites are subject to the same selection process (homogeneous). Under this model, five different foreground branch selection analyses were conducted on IGKC and IGLC1 genes: (1) the West Indian manatee branch; (2) the Amazonian manatee branch; (3) all manatees branch; (4) the Sirenia and Cetacea branch; and the (5) freshwater mammals branch.

We also used maximum likelihood frameworks running analyses on FEL, RELAX, and BUSTED from the HyPhy package v. 2.1 (Sergei et al., 2015), implemented on the Datamonkey web server (Weaver et al. 2018). The Fixed Effects Likelihood (FEL) model (Pond and Frost, 2005) infer *dN*/*dS* substitution rates on a per-site basis for a given alignment and its corresponding phylogeny applying the MG94xREV model to each codon site, assuming that the selection pressure for each site is constant along the entire phylogeny. We

also used the RELAX algorithm (Wertheim et al., 2015) to test whether the strength of natural selection has been relaxed (several neutral changes; K<1) or intensified (purifying or positive changes; K>1) in a given branch. Finally, we used the Branch-Site Unrestricted Statistical Test for Episodic Diversification (BUSTED) (Murrell et al., 2015) to investigate whether the genes under analyses have diversifying selection signs for a specific branch, testing for positive selection by asking whether a gene has experienced positive selection at least one site on at least one branch. On Godon and HyPhy analyses, five different branches were tested on both IGKC and IGLC1 genes, respectively. However, we could not identify IGKC on the *T. inunguis*' genome, keeping this branch out of the analyses for this gene.

Under FEL, RELAX, and BUSTED models, the same five foreground branch selection analyses were conducted on IGLC1 and IGKC genes: (1) the West Indian manatee branch; (2) the Amazonian manatee branch; (3) all manatees branch; (4) the Sirenia and Cetacea branch; and (5) the freshwater mammals branch (Figure 9). In addition, for all selection analyses based on ML, nested models were compared using the likelihood ratio test (LRT), and the results were evaluated against χ^2 distributions with different degrees of freedom according to each test.

Figure 9. Maximum likelihood trees generated by the IGKC (A) and IGLC1 (B) alignments used on Godon, FEL, BUSTED and RELAX analyses[6].



Source: Own design.

3.3. Results

We identified *T.manatus*' IGL locus within six scaffolds. The kappa chain locus spans approximately 1,296,911 kb from the most 5' VK segment to the most 3' CK gene. We identified 23 VK gene segments and 1 CK gene. In this class of antibody light chains, there is usually a single constant region at the end of the locus, as we found in our analysis. The lambda chain spans approximately 366,776 bp. *Trichechus inunguis*' IGL locus is located in eight scaffolds of the genome assembly. The lambda locus spans about 987,359bp. We

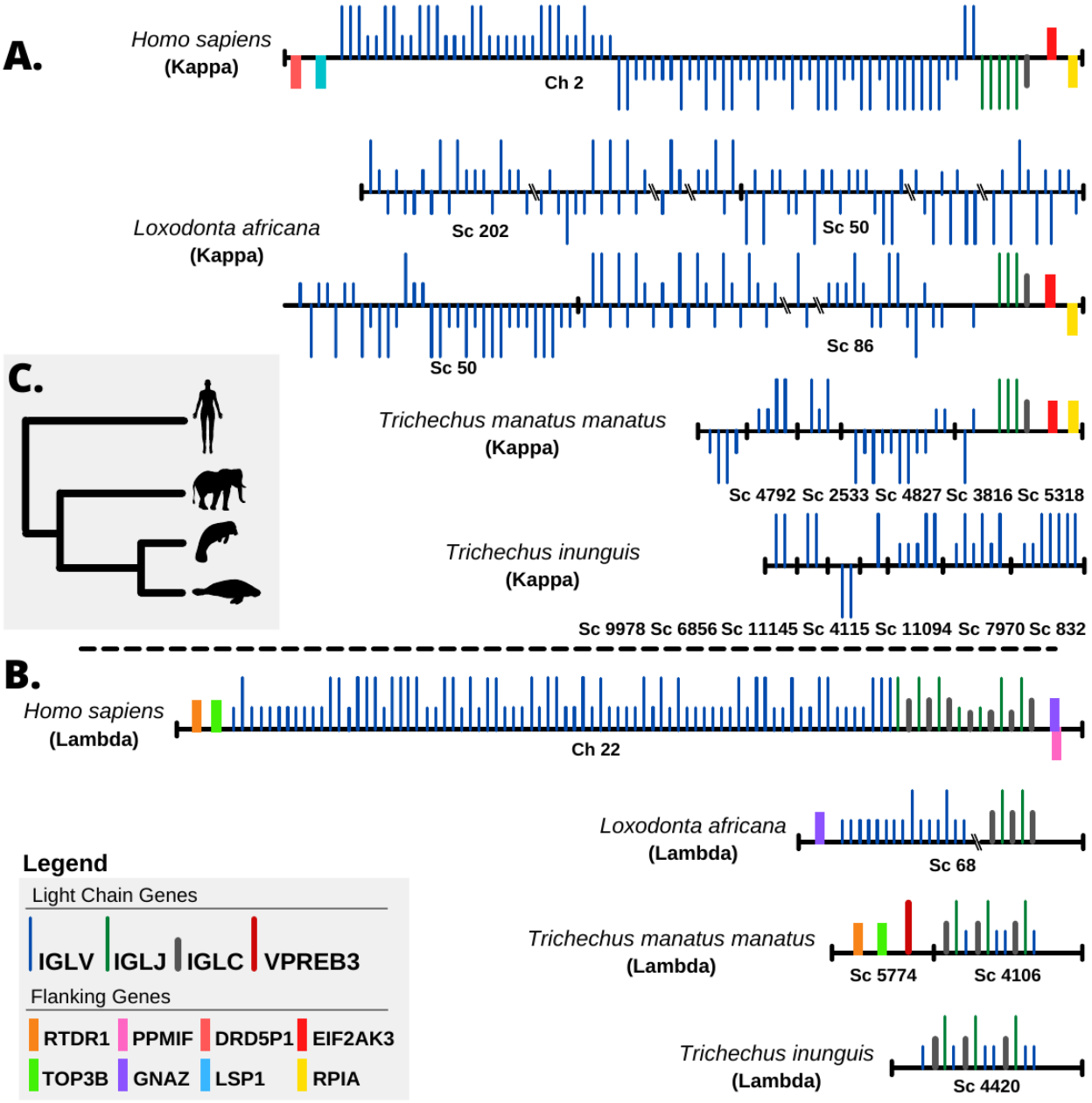
[6] Foreground branches used on each analysis are depicted as colored squares.

identified 24 variable kappa chain (VK) gene segments and one VPREB3 gene. The VPREB3 gene (V-set pre-B cell surrogate light chain 3) encodes for the Pre-B lymphocyte protein 3, which is thought to be involved in B-cell maturation and may play a role in the assembly of the pre-B cell receptor biosynthesis (pre-BCR) (Rosnet et al., 2004; Felizola et al., 2015). Unfortunately, we could not find the kappa constant region gene (IGKC) on the *T. inunguis* genome, which helps to determine the end of the locus. The lambda chain spans approximately 167,859bp. We found four variable lambda chain (VL) gene segments and three constant region genes (CL) in the same scaffold. This assembly's high genome coverage (~111,47x) confers high reliability on the gene repertoire characterization. However, to confirm if the absence of the kappa constant region gene on *T. inunguis* genome is a real evolutionary pattern or an artifact caused by genome fragmentation, an improvement in the contiguity of this genome is suggested, as well as the assembling of a chromosome level genome using long reads, or a more specific laboratory test for this purpose.

3.3.1. Constant region genes and Joining genes segments

As found in humans, mouse and elephants, *Trichechus manatus* IGKJ and IGKC genes are organized after the cluster of IGKV genes found 5′ of a small number (3) of IGKJ genes, with the IGKJ gene cluster located 5′ of a single IGKC gene (Figure 10). Due to the preliminary nature of *T. inunguis*' genome assembly, we could not identify the kappa joining genes and constant region genes (IGKJ and IGKC). In the lambda chain, both *T. inunguis*' and *T. manatus*' IGLJ and IGLC genes are organized in pairs, and their IGL gene locus consists of three IGLJ–IGLC gene recombination units in a tandem array (Figure 10). Although most mammalian species include a set of IGLV genes that are located 5′ from a variable number of cassettes, each made up of an IGLJ gene and an IGLC gene, the manatees' IGLV genes are found to be amongst the IGLJ-IGLC pairs, rather than in 5′ of the IGLJ-IGLC pairs (Figure 10).

Figure 10. Schematic diagrams of the genomic organizations of IGL loci in human, elephant, and the West Indian and Amazonian manatees (not to scale).



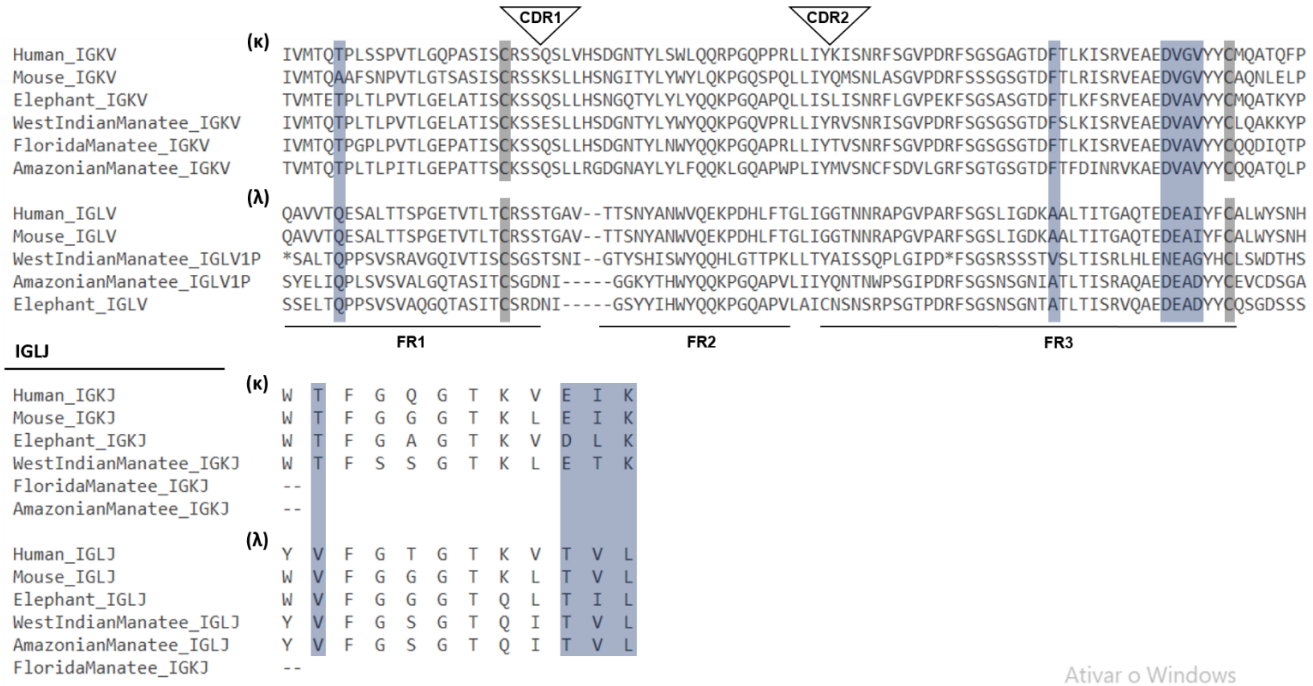
Source: Own design.

The figure 10 is about Schematic diagrams of the genomic organizations of IGL loci in human, elephant, and the West Indian and Amazonian manatees (not to scale). Black thick and vertical lines represent the species scaffold, and crossed strokes indicate long spaces with no immunoglobulin genes along the scaffold. The blue and green thin rods represent IGLV and IGLJ, respectively. The long thick marron rod represents the VPREB3 gene, and the gray

thick rods lines represent IGLC genes. Long rods show functional genes, and short rods indicate pseudogenes. Rods above and below the lines indicate genes located on opposite strands. Colored rectangles represent the flanking genes, as shown in the legend. Here, the gene symbols used are as follows: RPIA, ribose-5-phosphate isomerase; EIF2AK3, eukaryotic translation initiation factor 2 alpha kinase 3; LSP1, lymphocyte-specific protein 1; DRD5P1, dopamine receptor D5 pseudogene 1; GNAZ, guanine nucleotide-binding protein; RTDR1, rhabdoid tumor deletion region gene 1; TOP3B, topoisomerase III beta; and PPM1F, protein phosphatase 1F. Ch and Sc correspond to chromosomes and scaffold, respectively.

Six probably functional J segments were identified for *Trichechus manatus* with 23-spaced RSS and an FGXG or FGXR motif. Three of them belong to the kappa chain, and three to the lambda chain. The predicted sequence length of the functional J segments is 36 amino acids each (Figure 11).

Figure 11. Alignments of IGLV and IGLJ sequences. One IGLV and one IGLJ sequence (randomly chosen) from each isotype of each species were taken from the large-scale alignment.



Source: Own design.

The figure 11 is about Alignments of IGLV and IGLJ sequences. One IGLV and one IGLJ sequence (randomly chosen) from each isotype of each species were taken from the large-scale alignment. The cladistic molecular markers that distinguish between the two isotypes (Kand L) are highlighted in dark blue. Black triangles indicate CDR regions, while

the lines below the sequences indicate FR regions and their delimitations. Both regions were determined using the IMGT numbering system. Conserved cysteines in V segments are highlighted in gray.

3.3.2. Variable gene segments

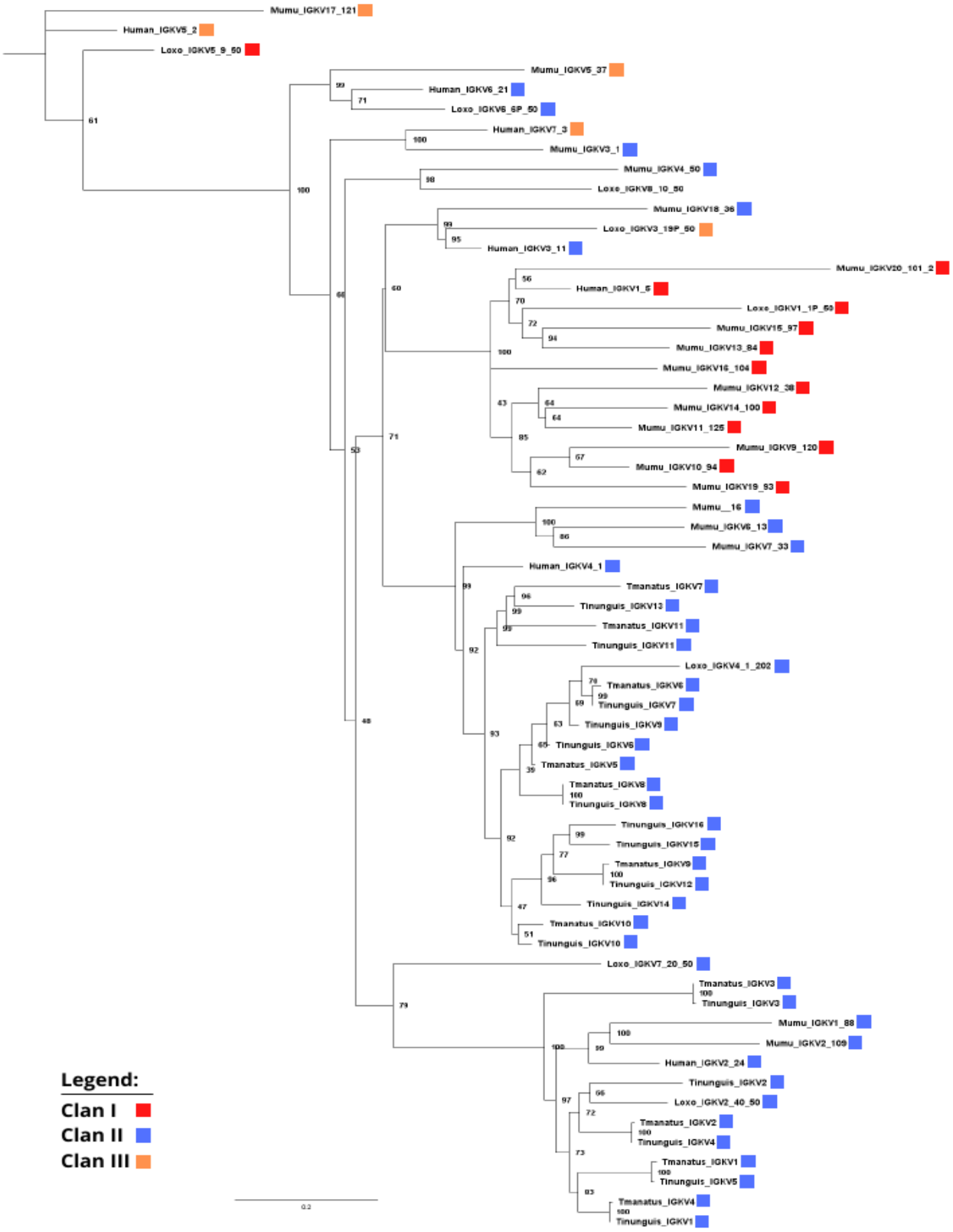
A total of 23 V kappa chain (VK) segments and four V lambda chain (VL) segments were identified on the T. manatus genome assembly. Of these V segments, 13 VK segments and all of the four VL segments were designated as pseudogenes due to the presence of stop codons or frameshifts, leaving ten potentially functional VK segments and no potentially functional VL segments.

In T. inunguis genome assembly, a total of 24 VK segments and three VL segments were identified. Of these V segments, eight VK segments and all of the three VL segments were designated as pseudogenes, leaving 16 potentially functional VK segments and no potentially functional VL segments. More functional V segments and pseudogenes may exist, but no other V segments have been identified on different scaffolds.

Since all VL gene segments from both species are potential pseudogenes, we analyzed clans clustering only through VK gene segments. Compared to human, mouse, and elephant sequences representative of the three VK clans (Clan I, II, III) defined by the IMGT database, all of the 11 functional VK segments from T. manatus cluster with V segments representatives of Clan II, whereas 15 functional VK segments from T. inunguis were all also clustered with V segments representatives of Clan II (Figure 12). No functional VK segments belong to Clan I and III on both species.

The figure 12 is about Maximum likelihood tree of Human (Human), mouse (Mumu), and elephant's (Loxo) potentially functional IGKV sequences representative of the 3 VK clans (Clans I, II, III) defined by IMGT database, and the representation of the manatees' IGKV sequences clustering altogether within Clan II. Sequences were randomly chosen, and the tips' numbers indicate the names of the sequences in the IMGT database.

Figure 12: Maximum likelihood tree of Human (Human), mouse (Mumu), and elephant's (Loxo).



Source: Own design.

3.3.3. Cladistic markers of IGLV, IGLJ, and IGLC genes sequences

Our analysis found, at position six of the IGLV alignment, mostly Thr (T) in the kappa chain, whereas in the lambda chain, the same position contains only Gln (G) (Figure 11). In addition, the two groups can be distinguished by the amino acid position 75, which includes a bulky aromatic side chain (Phe – F) in kappa chain and mostly a residue with a smaller side chain (Ala – A) in lambda chain sequences, except in *T. manatus* which contains a Val (V) in the same position. Furthermore, there is a conserved DEAD (Asp-Glu-Ala-Asp) motif in lambda chain sequences for *T. inunguis* and elephant and a DEAI motif for humans and mice.. *manatus* kept only Glu, and Ala conserved from this motif, having a NEAG motif instead (Asn-Glu-Ala-Gly). This motif has degenerated in kappa chain sequences, and all the Afrotherians have a conserved DVAV (Asp-Val-Ala-Val) motif (Figure 11).

Kappa and lambda groups defined by the markers of IGLV sequences are also supported by a cladistics analysis of the IGLJ sequences (Figure 11). In particular, the Thr at position 2 discriminates kappa sequences from the lambda sequences, which contains the Val at the same position. Kappa sequences have the EIK motif for humans and mice, mutated to DLK in elephant and ETK in *T.. manatus* at positions 10–12. In contrast, lambda sequences contain the TVL (or TIL) motif (Figure 11). IGKC and IGLC1 sequences residues discriminate K sequences from L sequences, specifically at positions 13, 56, 71, and 108 (Figure 13). These markers distinguish IGL isotypes in all species included in the study.

Figure 13: Alignment of representative IGKC and IGLC1 sequences.

Source: Own design.

The conserved markers that can be used for identifying the two isotypes (K and L) are highlighted in dark blue. Conserved cysteines in C segments are highlighted in gray. Sites possibly under positive selection are highlighted in dark red.

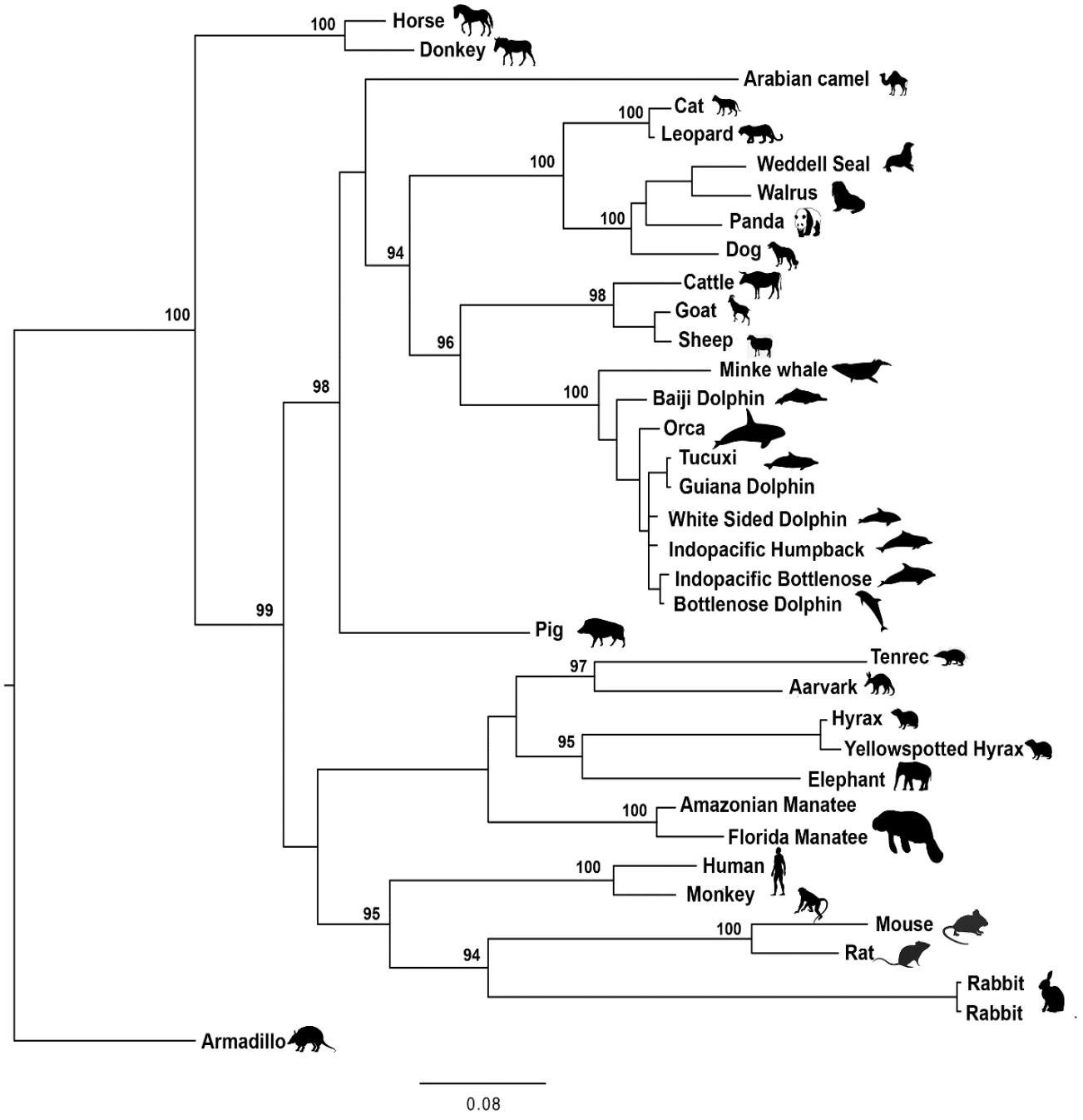
3.3.4. Genes Flanking the IGL Loci

The 3’ end of the *Trichechus manatus* K locus is flanked by the non-IGL genes RPIA and EIF2AK3. This pattern is also seen in humans, which have the 3' end of the same locus flanked by the genes RPIA and EIF2AK3 (see Figure 10 for these abbreviated genes). Similarly, the 5’ end of the *T. manatus* lambda locus seems to be flanked by the TOP3B and PPM1F genes because of a possible L-like VPREB gene found close to the flanking genes. In humans, the L-encoding locus is also flanked by TOP3B and PPM1F genes, yet, in other species, these genes don't seem to be in conserved synteny. On the other hand, the 3’ end of the human lambda locus contains the GNAZ and the RTDR1 genes. In manatees and in other species, GNAZ and the RTDR1 genes are not in conserved synteny with the L-encoding locus, and their distances from the L-encoding locus are not conserved. However, GNAZ and the RTDR1 genes were identified in a different scaffold, and no sign of V segments was found within this scaffold to consider a sequel of the locus. In *T. inunguis*’ genome assembly, the gene pairs GNAZ and RTDR1 and TOP3B and PPM1F were identified together in different scaffolds, each pair. The L-encoding locus of the elephant genome assembly, the GNAZ gene, was identified at 5’ of the locus, rather than the 3’ end seen in humans. The RTDR1 gene could not be found. Both the TOP3B and PPM1F genes were identified together in a different scaffold than the one containing the lambda locus (Figure 10).

3.3.5. Phylogenetic reconstruction

Maximum Likelihood (ML) and Bayesian inferences recovered similar topologies for both IGKC and IGLC1 datasets with bootstrap values higher than 60% and posterior probabilities (PP > 90%) support for most branches. Since ML and Bayesian approaches yielded similar topologies for both datasets, only ML inferences are shown (Figures 14 and 15).

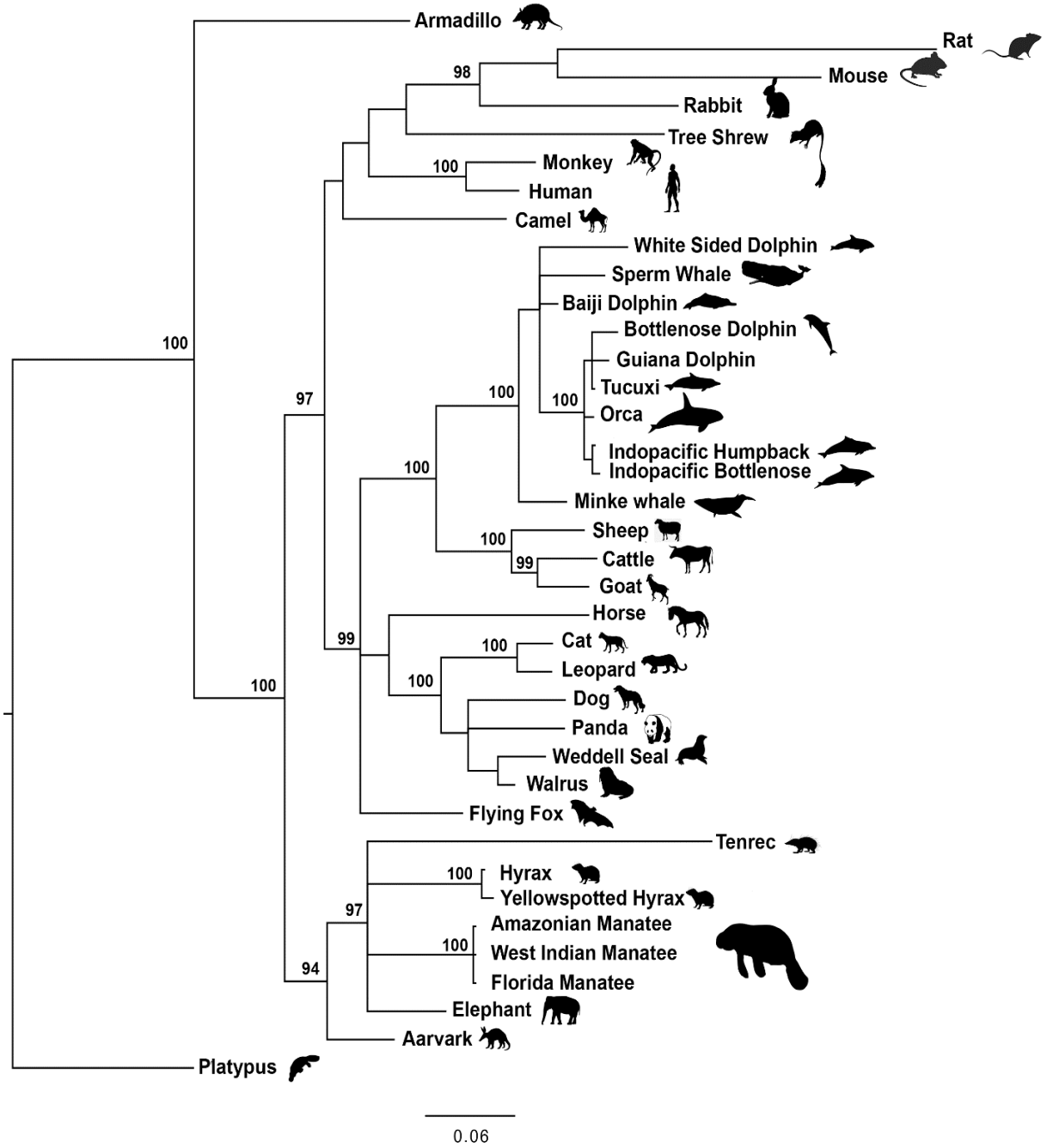
Figure 14: Maximum likelihood of functional mammalian IGKC gene sequences.



Source: Own design.

Armadillo was used as an outgroup. Numbers in branches indicate values supported above 90% bootstrap value. The scale shown as a bar represents the genetic distance. Phylogeny branch lengths represent genetic changes i.e. the longer the branch, the more genetic change has occurred.

Figure 15: Maximum likelihood tree of functional mammalian IGLC1 gene sequences.



Source: Own design.

Platypus was used as an outgroup. Numbers in branches indicate values supported above 90% bootstrap value. The scale shown as a bar represents the genetic distance.

Phylogeny branch lengths represent genetic changes i.e the longer the branch, the more genetic change has occurred.

3.3.6. Selection analyses

We investigated five different foreground branches on Godon analyses and FEL, RELAX, and BUSTED maximum likelihood analyses. Considering Godon analyses, we detected evidence for positive selection only at the 104th position in the constant region of the kappa chain (IGKC) of the West Indian manatee branch (LRT = 7.0917, p = 0.0077, Bayes Empirical Bayes posterior probability = 0.987). We also found eight sites (11, 17, 41, 48, 52, 81, 98, 104) for the branch comprising all the manatees and three sites (11, 81, 104) for the Sirenia and Cetacea branch. Site number 104 is found as positively selected on all of the analyses (Table 3). FEL identified one site (104) as positively selected in the West Indian branch, one site (104) for the branch with all the manatees, four sites (8, 18, 81, 104) for Sirenia and Cetacea branches, and two sites (7, 17) on freshwater mammals. Two sites (90, 94) were under negative selection on the Sirenia and Cetacea branches. These findings are consistent with RELAX results that tested significant intensified selection (K >1) acting on IGKC in each branch analyzed, except for the freshwater branch identified under relaxation. Thus, the intensified selection identified in the West Indian branch, all manatees branch, and Sirenia and Cetacea branches can either lean to a purifying selection or a positive selection. Furthermore, BUSTED detected significant diversifying selection on all branches investigated, except for the freshwater mammals branch (Figure 9, Table 3), consistent with the results found under the RELAX analysis.

As to the IGLC1 gene, the same four different branches were investigated in the analyses, with the Amazonian manatee branch. Godon analyses found three sites (**22**, 31, 92) positively selected along the Sirenia and Cetacea branches, while no sites were identified as positively selected in the manatees' branches. Five sites (**22**, 80, 91, 95, 96) were identified under positive selection by FEL in the Sirenia and Cetacea branch, and 14 sites (9, 12, 13, 30, 32, 47, 57, 61, 62, 71, 72, 74, 76, 77) were under negative selection in the same foreground branches. RELAX identified significant intensified selection (K >1) acting on IGLC1 in the Sirenia and Cetacea branches, while BUSTED detected evidence of diversifying selection happening in the same branches (Figure 9, Table 3).

Table 3. The results from selection analyses using different programs for IGKC and IGLC1 genes and the foreground branches used on each analysis.

FOREGROUND BRANCHES	SITES UNDER POSITIVE SELECTION		BUSTED	RELAX
IGKC	**Godon**	**FEL**	**PS Evidence**	**Strength of NS**
West Indian manatee	**103**	**103**	Evidence	Intensified
Amazonian manatee	-	-	-	-
All manatees	11,17,41,48,52,**80**,98,**103**	**103**	Evidence	Intensified
Sirenia and Cetacea	11,**80**,**103**	8, 18, **80**, **103**	Evidence	Intensified
Freshwater mammals	-	7, 17	No Evidence	Relaxed
IGLC1	**Godon**	**FEL**	**BUSTED**	**RELAX**
West Indian manatee	-	31	No Evidence	Relaxed
Amazonian manatee	-	**22**	No Evidence	Relaxed
All manatees	-	**22**,48	No Evidence	Relaxed
Sirenia and Cetacea	**22**, 31, 92, 95	**22**, 80, 91, 95, 96	Evidence	Intensified
Freshwater mammals	-	**22**, 92	No Evidence	Relaxed

Source: Own data.

Here, PS stands for Positive Selection and NS for Natural Selection. Dashes in the Amazonian manatee line represent no data regarding that gene for that branch. Dashes in Godon analyses represent no sites found under positive selection on that foreground branch.

3.4. Discussion

This study is the first to characterize the immunoglobulin light chain (IGL) genomic organization in manatees using the recently sequenced genome of *Trichechus inunguis* (Amazonian manatee) and the re-sequenced genome of *Trichechus manatus* (West Indian manatee). The organization of the IGL genes is particularly complex and the number of genes per species within the kappa (K) chain and lambda (L) chain is highly variable as a result of dynamic evolutionary processes in these gene families (Collins and Watson, 2018). Sequencing the human kappa locus identified 44 functional IGKV genes (Kawasaki et al., 1997; Watson et al., 2015), and 54 IGKV genes were characterized in genomic sequences from the rhesus macaque (Ramesh et al., 2017). In the guinea pig and rat reference sequences, 111 and 135 potentially functional IGKV genes have been found, respectively (Guo et al., 2012; Dammers et al., 2014). Studies of the elephant genome reference sequence identified 53 functional IGKV genes (Guo et al., 2011) and only 19 potentially functional IGKV genes in the horse genome (Sun et al., 2010). Compared to most mammals, *T. manatus*

and *T. inunguis* have a limited repertoire based on their low segmental diversity of V and J segments. Manatee's IGKV genes are more abundant than IGLV genes, considering that all of the IGLV genes identified are pseudogenes and that we found only four (*T. manatus*) and three (*T. inunguis*) of them, compared to 23 and 24 IGKV genes identified within the *T. manatus* and *T. inunguis* genomes, respectively. Of these 23 and 24 IGKV genes, only 13 and eight of them were considered potentially functional in *T. manatus* and *T. inunguis*, respectively. The manatee's IGL locus has one of the fewest numbers of functional V segments observed in mammals, and they are all concentrated in the kappa locus.

Within the kappa locus of humans, other primates, mice, pigs, goats, and elephants, genes are organized in a similar way to the genes of the heavy chain locus: a cluster of IGKV genes are found 5′ of a small number of IGKJ genes, with the IGKJ gene cluster located 5′ of a single IGKC gene (Ermert et al., 1995; Kirschbaum et al., 1998; Guo et al., 2011; Schwartz et al., 2012a; Schwartz et al., 2018). On the other hand, the lambda locus of most species investigated to date, such as humans (Kawasaki et al., 1997), the rhesus macaque (Ramesh et al., 2017), pigs (Schwartz et al., 2012b), and elephants (Guo et al., 2011), includes a set of IGLV genes that are located 5′ from a variable number of tandem cassettes, each made up of an IGLJ gene and an IGLC gene. On both the *T. manatus* and *T. inunguis*, the organization of the kappa locus is similar to the species mentioned above. However, the lambda locus of the manatees is differently organized. It is common to identify IGLV genes located in 5′ of functional J-C pairs, but this organization is not invariant. For example, we identified 4 and 3 IGLV genes within the lambda locus for *T. manatus* and *T. inunguis,* respectively. All of them were designated as pseudogenes due to the presence of stop codons or frameshifts. Furthermore, the IGLV genes are located amongst the J-C pairs rather than in 5′ of the J-C pairs.

It is also possible to identify different organizations of the L-encoding locus in horses and mice. In horses, a total of 144 IGLV genes have been identified, with 110 genes being located upstream and 34 downstream of the IGLJ/IGLC cluster (Sun et al., 2010; Walther et al., 2015). As seen in manatees, low diversity in the lambda locus is also a characteristic in mice of all three *Mus musculus* subspecies (Scott and Potter, 1984). The C57BL/6 mouse genome includes three IGLV genes. Two of them are associated with one functional J-C pair, while the third is associated with a second J-C pair (Scott and Potter, 1984). In each of the

two mice, VJC units are where the lambda rearrangement occurs, showing little or no recombination between the units (Sanchez et al., 1991). This might also be the case of manatees, causing them to manage their antibodies diversity mainly through the kappa chain.

Different ratios of IGKV and IGLV genes are found in different mammalian species. For example, we have mice with 95% of their light chain repertoire based on IGKV genes and humans with about 60% of their IGL repertoire generated from IGKV genes (Almagro et al., 1998). Considering these findings, it has been proposed that the preference to use a certain light chain isotype at the protein level may be correlated with the overall quantity of V gene segments (Almagro et al., 1998). For the light chain genes, manatee IGKV genes are more abundant than IGLV genes, being possible that the kappa chain predominates over the lambda chain at the protein level in these species. Compared to potentially functional IGKV genes, segments from both manatee species group into three families: one family of one segment from each species, one family of three segments of *T. manatus* and four segments of *T. inunguis*, and another family of seven and eleven segments of *T. manatus* and *T. inunguis*, respectively. All of the IGKV segments together (27) group to Clan II, and none group to Clan I or III (Figure 9).

The birth-and-death evolutionary model has been proposed to describe the evolution of immunoglobulin multigene families. Under this model, the birth of genes occurs by gene duplication with subsequent divergent evolution, and the death of genes is caused by either gene deletion or loss of functionality (pseudogenes) (Nei et al., 1997). Besides lacking Clan I and Clan III IGKV segments in the IGL locus, manatees also lack Clan III V segments from the immunoglobulin heavy chain (IGH) locus (Breaux et al., 2017). Also, it is possible to notice a great number of pseudogenes present in IGL and IGH from most tetrapods studied to date. In humans, 27 of the 76 IGKV genes are pseudogenes (Tomlinson et al., 1996), and 100 of the 153 IGKV genes found in elephants were considered non-functional (Guo et al., 2011). In manatees' IGH locus, a total of 139 IGHV segments were identified, and 118 of them were designated as pseudogenes (Breaux et al., 2017). In the IGL lambda locus, manatees do not appear to have any functional IGLV segments, only pseudogenes. In the kappa locus, 13 of the 23 IGKV genes in *T. manatus* were designated as pseudogenes. The riverine *T. inunguis* manatee has fewer IGKV genes but a higher number of potentially functional genes. Of the 24 IGKV segments, eight were designated pseudogenes. This evidence reflects an

evolutionary loss of IGHV, IGKV, and IGLV genes, according to the “birth-and-death” model, in which the number of genes in a family (or “library”) is allowed to expand and contract.

Another possible reason for the loss of V segments among mammals, especially manatees, is the drastic environmental change through their evolutionary history. The disparity of the V segments quantity present in terrestrial mammals compared to marine mammals is extensive, suggesting that the aquatic environment may have influenced the V segments diversity in manatees. The birth-and-death model of evolution (Ota and Nei 1994; Nei et al. 1997; Sitnikova and Su 1998; Das et al. 2008) resembles the accordion model of MHC evolution (Klein et al., 1993). The accordion model theorizes that depending on the need to protect the host from different pathogens, V segments libraries would be duplicated and diverged functionally. In contrast, others could become pseudogenes and/or be deleted from the genome. Accordingly, the ever-changing pathogens found in the aquatic environment may have influenced the loss of IGKV and IGLV diversity in manatees.

Trichechus inunguis has almost the same number of IGKV genes (24) as *T. manatus* has (23). However, the number of potentially functional IGKV genes compared to pseudogenes in both species is different. *Trichechus inunguis* has 16 potentially functional IGKV genes compared to ten from *T. manatus*. This indicates a higher level of IGKV gene sequence variation in *T. inunguis* that might be correlated to its secondary radiation from the marine environment to an exclusively freshwater sirenian.

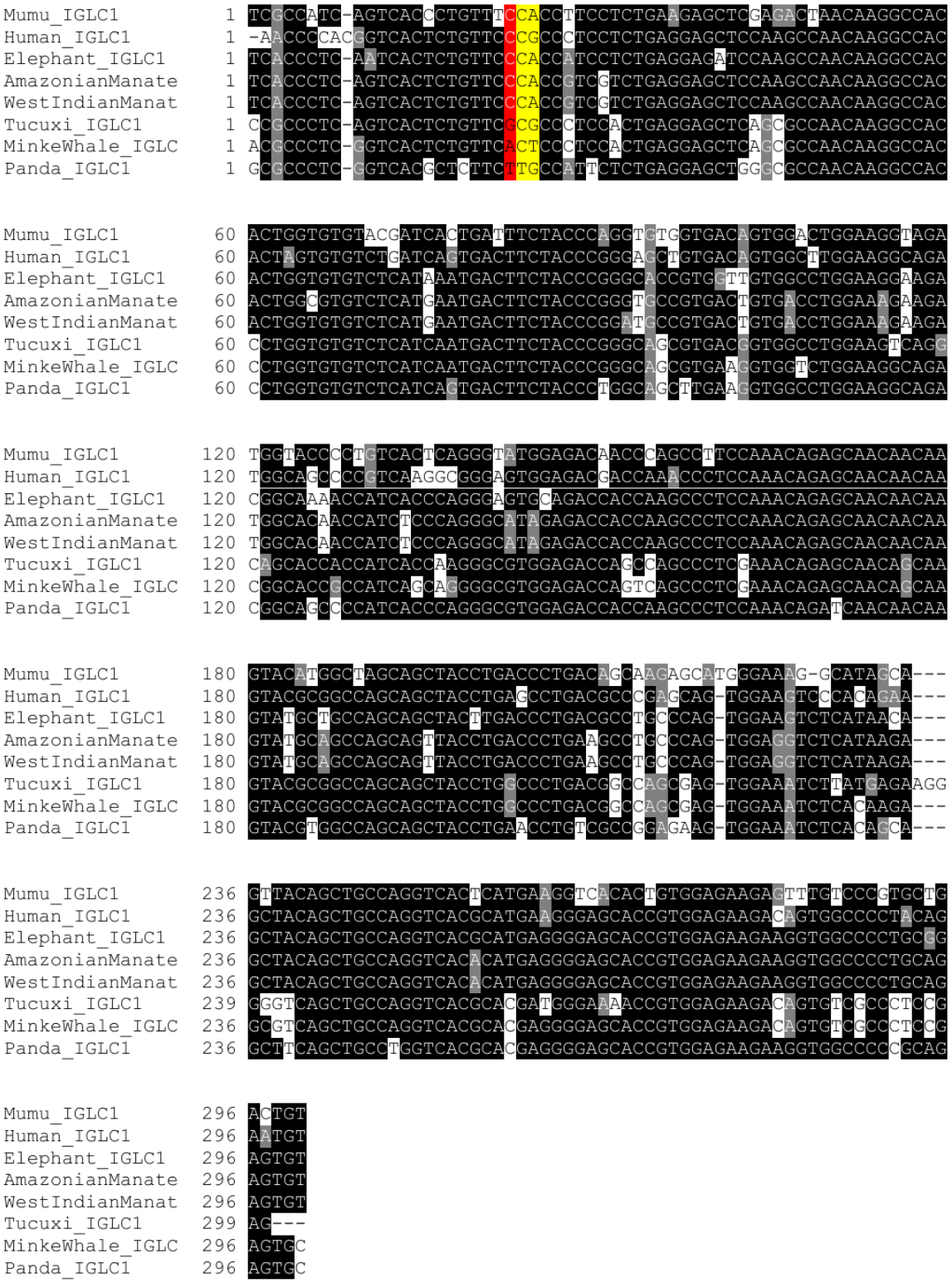
A different approach based on using cladistic molecular markers to characterize IGLV, IGLJ, and IGLC protein sequences has been previously suggested. This alternative can determine different IGL isotypes in tetrapods by aligning framework regions (FR1 – FR3) of functional sequences to reveal the presence of conserved amino acid residues or alignment gaps, useful to distinguish IGLV sequences and assigning them into groups (Das et al., 2008). *Trichechus manatus* kappa genomic region is flanked by the RPIA and EIF2AK3 genes as seen in humans, elephants (Figure 7), and other tetrapods (Das et al., 2008). The human lambda genomic region is flanked by the GNAZ and RTDR1 genes in the 3' end and by the TOP3B and PPM1F in the 5' end of the locus. These genes are not in conserved synteny with the lambda genomic region in other species like chicken, lizard, frog (Das et al., 2008), and

manatee (this study). However, they are mostly placed on the same scaffold as that of the lambda genomic region, and their distances from the lambda locus are not conserved.

To detect the possible roles of positive selection during the evolution of kappa (IGKC) and lambda (IGLC1) constant region genes present in the immunoglobulin light chain locus, we conducted evolutionary analyses based on calculating the rate of non-synonymous and synonymous substitutions in their codons. Selection analyses using BUSTED and Godon methods retrieved signs of positive selection in IGKC referring to almost all the branches, except the freshwater branch, mainly indicated by the variation in sites number 80 and 103. Site 80 is found in the 27th codon. Mouse, human and monkey has an Asparagine (AAC) in this position; a Glycine (GGC) is found in the horse, panda, and both Tucuxi and Guiana dolphins alignments, whereas in elephant and both Amazonian and West Indian manatees a Glutamic acid (GAG) is found in this position (Figure 16). In IGLC1, site 22, with signs of positive selection under Godon and FEL analyses corroborated by the BUSTED and RELAX analyses, is found in the 8th codon of the constant domain 1. The Proline amino acid (CCA, CCG) found in almost all alignment is changed to Alanine (GCG, GCT) in cetaceans and platypus, except for the Minke whale and panda, which has Threonine (ACT) and Leucine (TTG), respectively, in the same position (Figure 15). Signs of positive selection in immune system genes have already been described for marine mammals, as in the study of Xu et al. (2019). This study found a total of 148 robust candidates of positive selection sites identified in Toll-like receptors (TLRs) genes in the Cetacea lineage. Amino acid changes in sites that were being positively selected were fortified during the period of an early evolutionary transition from land to semi-aquatic and from semi-aquatic to full-aquatic habitat. These changes might enable cetaceans to adapt quickly to new pathogens as they transition to new environments (Xu et al., 2019). Another interesting finding related to the transition from land to water environment in marine mammals was the great number of gene losses (85 genes) in the cetacean stem lineage (Huelsmann et al., 2019). These gene losses were likely neutral, and their loss happened because of relaxed selection to maintain their function. Moreover, pleiotropic genes with epidermal and hippocampal functions were convergently lost in cetaceans and manatees (Hecker et al., 2017; Huelsmann et al., 2019). Considering these findings, it can be suggested that gene losses in marine mammals may be associated with

aquatic specializations, especially when adding to these results the loss of Clan III in IGHV genes (Breaux et al., 2017) and Clan I and III in IGKV genes in manatees (this study).

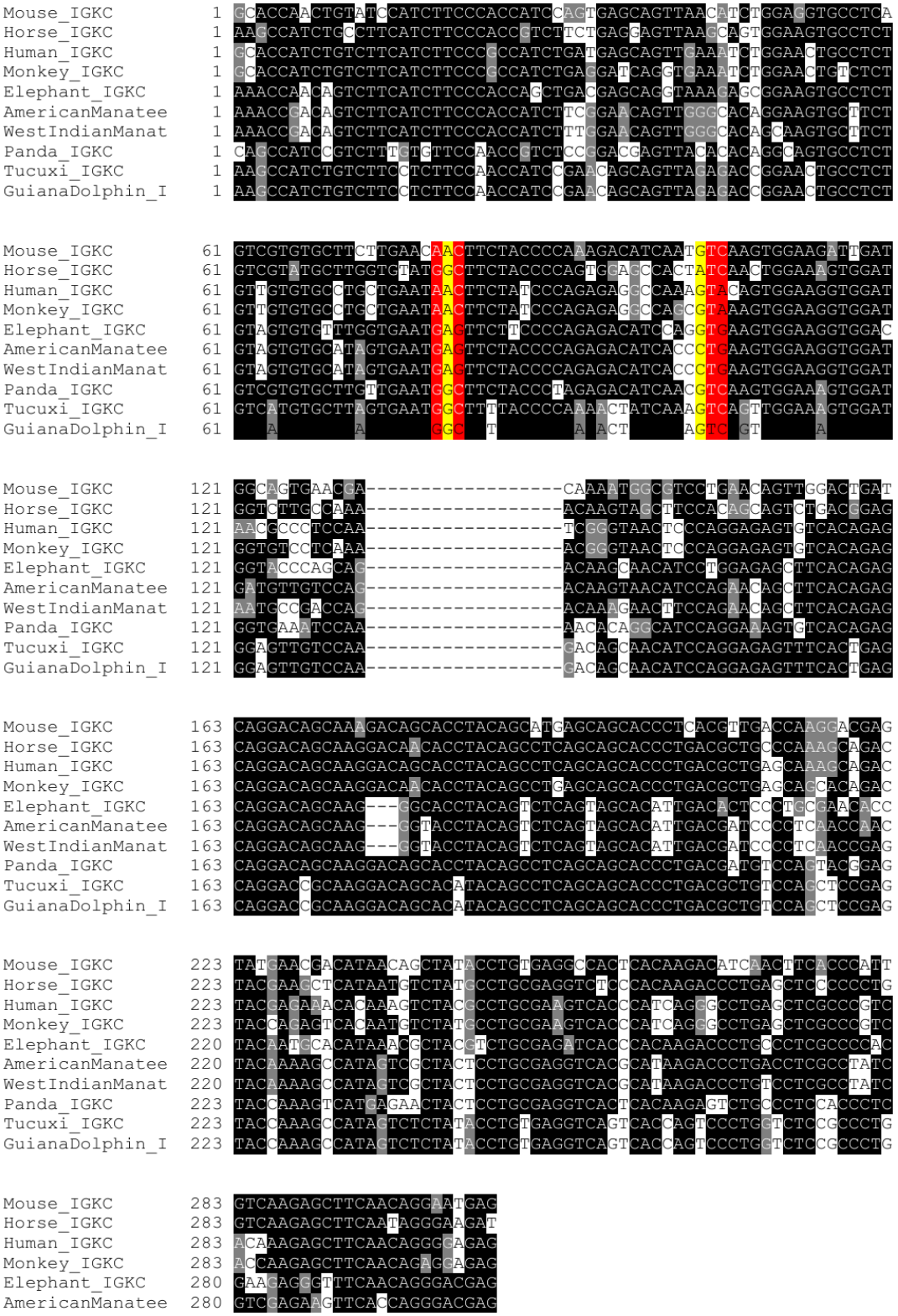
Figure 16: IGLC1 codon alignment of the mouse, human, elephant, Amazonian and West Indian manatees, Tucuxi dolphin, Minke whale, and panda[7].



Source: Own design.

[7] Site number 22 is highlighted in yellow. Sites 23 and 24 (highlighted in red), together with site 22, correspond to the codons mentioned in the text.

Figure 17: IGKC codon alignment of the mouse, horse, human, monkey, elephant, Amazonian and West Indian manatees, Tucuxi and Guiana dolphins, and panda[8].



Source: Own design.

[8] Sites number 80 and 103 are highlighted in yellow. Sites 79 and 81, 104 and 105 (highlighted in red), together with sites 80 and 103, respectively, correspond to the codons mentioned in the text.

Functional studies should be done to indicate whether the variation in sites 80 and 103 for IGKC and 22 for IGLC1 is contained in a functionally relevant portion of the IGKC and IGLC1 genes in manatees, which could indicate that the change occurring at these sites may also be related to the adaptive success of the mammalian immune system in the marine environment. In addition, only a few sites with selection signs were found for the branch comprising all the manatees on IGKC and IGLC1 using Godon. These results, together with the results of intensification of the strength of selection on both genes, retrieved by RELAX, may indicate that these genes are conserved in the branch of manatees.

3.5. Conclusion

Studying the structure and organization of the immunoglobulin gene loci is important to understand the nature of antibody molecules. This study is the first characterization of both the *Trichechus manatus* and *Trichechus inunguis* IGL locus. Our results showed that the number of genomically encoded segments and, consequently, the segmental diversity is limited in manatees. Besides lacking two V segments of clans (Clan I and Clan III) within the kappa chain, manatees seem to have only a few V pseudogenes within the lambda chain organized differently, as seen in most tetrapods. Altogether, these conditions enable manatees to create very little antibody diversity through the immunoglobulin light chain region. Therefore, they probably have to manage their antibody diversity mainly through the kappa chain. Furthermore, the environmental change manatees faced through their evolutionary history might also significantly impact the difference these species have related to their antibody diversity. The “birth-and-death” model of evolution may have influenced the diminishing of the light chain region in manatees upon the need to protect them from ever-changing groups of parasites in the aquatic environment, with some V segments libraries becoming pseudogenes or being deleted from the genome.

This study helps to fill the gap of knowledge in manatee immunity as well as sheds light on the evolutionary mechanisms that maintain diversity at this complex locus. Comparing manatee segmental diversity with other species enables us to understand which segments are critical within and between species. Our results make one wonder about how well-armed manatees are for defense against infectious disease in their environment and the possible impact it might have on manatee conservation. Considering immunoglobulins,

manatees have low segmental diversity available to create antibodies, making their antigens recognition capacity limited, possibly making these animals more vulnerable to new and different pathogens. Thus, the Sirenian lineage may be a probable candidate for close extinction, especially when considering their current vulnerable status and history of overhunting (O'Shea, 1988) and the significant population fragmentation due to habitat destruction (Vianna et al., 2006) of the American manatees. Also, West Indian manatees suffered a historical bottleneck that may have led to a low genetic diversity (Vianna et al., 2006). However, these animals may rely on their defense mechanism through cell-mediated immunity, counting on their T cell receptors (TCR) to compensate for their low BCR diversity. New investigations are necessary to understand better whether manatees rely on other ways to create diverse antigen recognition and how manatees are supplied to defend themselves in their environment. This study provides information for comparative studies of mammalian IG genes, as well as data for further studies of the manatee immunoglobulin genes.

CHAPTER 4

FINAL CONSIDERATIONS

The adaptive immune system (AIS) is a very complex yet fascinating defense mechanism. Despite the constant microbial challenges from the environment, the body prevents infections by a specific and incredibly diverse system that can fight various pathogens (Flajnik and Kasahara, 2010).

Immunoglobulin genes, antibodies, and the enzymes RAG1 and RAG2 work together to generate diversity through VJ rearrangement, in the case of immunoglobulin light chain genes. The RAG transposon had a major role in the emergence of the jawed vertebrate AIS, which has evolved to create a mechanism for deriving almost endless variation from very few genes. The RAG enzymes are conserved features in jawed vertebrate AIS, allowing a general functioning of the system, whereas immunoglobulin genes are plastic features, permitting rapid changes over evolutionary time to respond to different pathogens (Klein et al., 1993; Flajnik and Kasahara, 2010). Considering major environmental changes that happened throughout mammalian evolutionary history, one could wonder whether the conserved feature of the RAG enzymes endured radical transitions from land to water facing different pathogens. Molecular evolutionary analyses suggest that RAG1 and RAG2 genes remained fairly conserved among tetrapods with purifying selection acting on both genes, demonstrating different evolutionary rates, specifically in cetaceans (Dias and Nery, 2019).

The occupation of new environments is a favorable scenario for the emergence and diversification of new species and molecular changes in the genome of organisms during these ecological transitions. Among aquatic mammals, the manatees also had major environmental changes during their evolutionary history. Great effort has been made in quantifying the strength of the manatee immune system. Previous studies characterized blood cells (Bossart, 1995), mucosal-associated immune tissues (McGee, 2012), and the expression of surface interleukin-2 receptor (Sweat et al., 2005), as well as analyzed lymphocyte proliferation through different environmental stressors (Walsh et al., 2005). However, the specific antigen-binding diversity of the AIS receptors has not been described until recently. In 2017, studies showed that *Trichechus manatus* has a limited repertoire of immunoglobulins heavy chain (IGH) genes based on their low segmental diversity from 13 functional V

segments, 48 D segments, and six J segments (Breaux et al., 2017). Compared to humans, mice, and elephants, the manatee's possible segmental combinations are low. Moreover, the Florida manatee lacks IGH Clan III V segments, which may have reduced the levels of recombination (Breaux et al., 2017). To continue studies related to the adaptive immune system of manatees, we characterized the immunoglobulin light chain (IGL) genomic organization of the recently sequenced genome of *Trichechus inunguis* (Amazonian manatee) and *Trichechus manatus* (West Indian manatee). Results showed that the number of genomically encoded segments and, consequently, the segmental diversity is limited in manatees. These animals lack two IGLV segments clans (Clan I and Clan III) within the kappa chain, and they seem to have only a few V pseudogenes within the lambda chain organized in a different manner as seen in most tetrapods (Dias and Nery, unpublished). Altogether, these conditions enable manatees to create very little antibody diversity through the immunoglobulin light chain region. The environmental change manatees had to face through their evolutionary history might also have a big impact in the difference these species have related to their antibody diversity. Different from what we expected from manatee's immune system, considering earlier hypotheses about the robustness of their immune system, manatees have low IGH and IGL repertoire diversity, which may have been influenced by the "birth-and-death" model of evolution upon the need to protect them from ever-changing groups of parasites in the aquatic environment, with some V segments libraries becoming pseudogenes or being deleted from the genome.

Extant jawed vertebrates have integrated multiple systems in adaptive immune responses to achieve a collective goal of host defense. As such, aquatic mammals possess numerous and diverse systems under constant evolutionary pressure from the evolution of pathogens. Thus, it is expected that analyzing and understanding the lineage of a specific immune effector can provide us with only a small part of a complex figure.

Even so, as much as we could imagine that systems related to the adaptive immune responses of model and non-models aquatic mammals would present great variations in evolutionary pressure, cetaceans and sirenians retain their RAG1, RAG2, and IGL genes fairly conserved through their evolutionary history. It is probable that natural selection may have continuously assembled and excluded innovative genes related to immunity over evolutionary time, like immunoglobulin gene types as we can see in manatees IGL and IGH

genes, which makes their antigens recognition capacity limited, possibly making these animals more vulnerable to new and different pathogens. Thus, the Sirenian lineage may be a probable candidate for close extinction, especially when added to the anthropological impact they have on their habitat and their history of overhunting. As additional non-traditional species are explored through comparative studies, novel or more adapted immune system genes may be uncovered holding crucial clues about the evolution of adaptive immune responses.

This work is a part of the big picture of the immune system evolution and brings important clues towards our understanding of how adaptive immune responses work under different evolutionary pressures. We provide valuable information that can help the unraveling of the complex networks extant in aquatic mammals and jawed vertebrates that guide their host defenses. These findings, together with other and novel data from contrasting lineages will continue to direct us in approaching the mechanisms that affect immune function in jawed vertebrates.

REFERENCES

Abbas, A., Lichtman, A., & Pillai, S. (2015). Imunologia Celular e Molecular. 8a Ed..

Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature, 394(6695), 744-751.

Almagro, J. C., Hernández, I., Ramírez, M. C. and Vargas-Madrazo, E. (1998). Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. Immunogenetics, 47(5), 355-363.

Andrews, S. (2011). FastQC - a quality control tool for high throughput sequence data [Online]. Babraham Bioinformatics.

Berta, A., Sumich, J.L., Kovacs, K.M., Folkens, P.A. and Adam, P.J. (2006). Sirenian and other marine mammals: Evolution and systematic. Marine mammals (Second edition), 89-110.

Bik, E.M., Costello, E.K., Switzer, A.D., Callahan, B.J., Holmes, S.P., Wells, R.S., Carlin, K.P., Jensen, E.D., Venn-Watson, S. and Relman, D.A. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nature communications, 7(1), 1-13.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.

Bossart, G.D. (1995). Immunocytes of the Atlantic Bottlenose Dolphin (Tursiops truncatus) and West Indian Manatee (Trichechus manatus latirostris): Morphologic Characterizations and Correlations between Healthy and Disease States Under Free-ranging and Captive Conditions (Doctoral Dissertation). FIU Electronic Theses and Dissertations.

Bossart, G.D., Baden, D.G., Ewing, R.Y., Roberts, B. and Wright, S.D. (1998). Brevetoxicosis in manatees (Trichechus manatus latirostris) from the 1996 epizootic: gross, histologic, and immunohistochemical features. Toxicologic Pathology, 26(2), 276-282.

Breaux, B., Deiss, T. C., Chen, P. L., Cruz-Schneider, M. P., Sena, L., Hunter, M. E. and Criscitiello, M. F. (2017). The Florida manatee (Trichechus manatus latirostris) immunoglobulin heavy chain suggests the importance of clan III variable segments in repertoire diversity. Developmental & Comparative Immunology, 72, 57-68.

Butler, J. E. (2003). Disparate mechanisms drive antibody diversity among mammals: a useful addition to immunology textbooks. Current Trends Immunol, 5, 1-18.

Butler, J.E., Wertz, N., Baker, M.L. (2014). The immunoglobulin genes of bats. In: Kaushil AK, Pasman Y, editors. Comparative Immunoglobulin Genetics. Toronto, ON: Apple Academic Press, 53–84.

Cannon, J. P., Haire, R. N., Rast, J. P. and Litman, G. W. (2004). The phylogenetic origins of the antigen‐binding receptors and somatic diversification mechanisms. Immunological Reviews, 200(1), 12-22.

Chen, L., Li, M., Li, Q., Yang, X., An, X., Chen, Y. (2008) Characterization of the bovine immunoglobulin lambda light chain constant IGLC genes. Veterinary Immunology and Immunopathology, 124(3-4), 284-94.

Collins, A. M. and Watson, C. T. (2018). Immunoglobulin light chain gene rearrangements, receptor editing and the development of a Self-Tolerant antibody repertoire. Frontiers in immunology, 9, 2249.

Criscitiello, M. F. and Flajnik, M. F. (2007). Four primordial immunoglobulin light chain isotypes, including λ and κ, identified in the most primitive living jawed vertebrates. European journal of Immunology, 37(10), 2683-2694.

Criscitiello, M. F. and Flajnik, M. F. (2007). Four primordial immunoglobulin light chain isotypes, including λ and κ, identified in the most primitive living jawed vertebrates. European journal of immunology, 37(10), 2683-2694.

Daggfeldt, A., Bengtén, E. and Pilström, L. (1993). A cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (Gadus morhua L.) and rainbow trout (Oncorhynchus mykiss Walbaum) indicated by nucleotide sequences of cDNAs and hybridization analysis. Immunogenetics, 38(3), 199-209.

Dammers, P.M., Hendricks, J., Terpstra, P., Bos, N.A. and Kroese, F.G. (2014). The organization of the immunoglobulin heavy- and light-chain loci in the rat. Comparative Immunoglobulin Genetics. 107.

Das, S., Mohamedy, U., Hirano, M., Nei, M., and Nikolaidis, N. (2010). Analysis of the immunoglobulin light chain genes in zebra finch: evolutionary implications. Molecular Biology and Evolution, 27(1), 113-120.

Das, S., Nikolaidis, N., Klein, J. and Nei, M. (2008). Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. Proceedings of the National Academy of Sciences, 105(43), 16647-16652.

Davis, M. M., Chien, Y. H., Gascoigne, N. R. and Hedrick, S. M. (1984). A murine T cell receptor gene complex: isolation, structure and rearrangement. Immunological Reviews, 81, 235-258.

Davydov, I. I., Salamin, N. and Robinson-Rechavi, M. (2019). Large-scale comparative analysis of codon models accounting for protein and nucleotide selection. Molecular biology and evolution, 36(6), 1316-1332.

Domning, D. P. (1982). Evolution of manatees: a speculative history. Journal of Paleontology, 599-619.

Domning, D.P. (2002). Sirenian evolution, in Perrin, W.F., et al., eds., The encyclopedia of marine mammals: San Diego, California, Academic Press, 1083–1086.

Dreyfus, D. H. (2009). Paleo-immunology: evidence consistent with insertion of a primordial herpes virus-like element in the origins of acquired immunity. PLoS One, 4(6), e5778.

Du Pasquier, L. (1989). Evolution of the Immune System. Fundamental Immunology, ed. Paul, W.E. New York: Raven Press Ltd, 139-165.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792-1797.

Edholm, E. S., Wilson, M., and Bengten, E. (2011). Immunoglobulin light (IgL) chains in ectothermic vertebrates. Developmental & Comparative Immunology, 35(9), 906-915.

Ermert, K., Mitlohner, H., Schempp, W. and Zachau, H.G. (1995). The immunoglobulin kappa locus of primates. Genomics, 25, 623-629.

Felizola, S.J., Katsu, K., Ise, K., Nakamura, Y., Arai, Y., Satoh, F., Sasano, H. (2015). Pre-B lymphocyte protein 3 (VPREB3) expression in the adrenal cortex: precedent for non-immunological roles in normal and neoplastic human tissues. Endocrine Pathology. [in press].

Flajnik, M. F. (2002). Comparative analyses of immunoglobulin genes: surprises and portents. Nature Reviews Immunology, 2, 688–698.

Flajnik, M. F., Miller, K. and Du Pasquier, L. (2003). Fundamental Immunology, ed. Paul, W. E., Lippincott Williams & Wilkins, Philadelphia, 5, 519–570.

Flajnik, M.F. and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nature Reviews Genetics 11, 47-59.

Foley, N.M., Springer, M.S. and Teeling, E.C. (2016). Mammal madness: is the mammal tree of life not yet resolved?. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1699), 20150140.

Fugmann, S. D., Messier, C., Novack, L. A., Cameron, R. A. and Rast, J. P. (2006). An ancient evolutionary origin of the Rag1/2 gene locus. Proceedings of the National Academy of Sciences, 103(10), 3728-3733.

Giudicelli, V. and Lefranc, M. P. (1999). Ontology for immunogenetics: the IMGT-ONTOLOGY. Bioinformatics, 15(12), 1047-1054.

Giudicelli, V., Chaume, D., Lefranc, M.P. (2005). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Research, 33(Database issue), D256-61.

Goldman, N. and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular biology and evolution, 11(5), 725-736.

Good, R.A., Finstad, J., Litman, G.W. (1972) The Biology of Lampreys, eds. Hardisty, M.V. and Potter, I.C. London: Academic, 2, 405–432.

Greenberg, A. S., Hughes, A. L., Guo, J., Avila, D., McKinney, E. C., and Flajnik, M. F. (1996). A novel "chimeric" antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin. European Journal of Immunology, 26(5), 1123-1129.

Greenberg, A. S., Steiner, L., Kasahara, M. and Flajnik, M. F. (1993). Isolation of a shark immunoglobulin light chain cDNA clone encoding a protein resembling mammalian kappa light chains: implications for the evolution of light chains. Proceedings of the National Academy of Sciences, 90(22), 10603-10607.

Guo Y, Bao Y, Meng Q, Hu X, Meng Q, Ren L, et al. (2012). Immunoglobulin genomics in the guinea pig (Cavia porcellus). PLoS one, 7(6), e39298.

Guo, Y., Bao, Y., Wang, H., Hu, X., Zhao, Z., Li, N. and Zhao, Y. (2011). A preliminary analysis of the immunoglobulin genes in the African elephant (Loxodonta africana). PloS one, 6(2), e16889.

Guselnikov, S. V., Baranov, K. O., Najakshin, A. M., Mechetina, L. V., Chikaev, N. A., Makunin, A. I., ... & Taranin, A. V. (2018). Diversity of immunoglobulin light chain genes in non-teleost ray-finned fish uncovers IgL subdivision into five ancient isotypes. Frontiers in Immunology, 9, 1079.

Hecker, N., Sharma, V. and Hiller, M. (2017). Transition to an aquatic habitat permitted the repeated loss of the pleiotropic KLK8 gene in mammals. Genome biology and evolution, 9(11), 3179-3188.

Hinds, K. R., and Litman, G. W. (1986). Major reorganization of immunoglobulin VH segmental elements during vertebrate evolution. Nature, 320(6062), 546-549.

Hiom, K., Melek, M. and Gellert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell, 94(4), 463-470.

Hoelzel, A. R. (2009). Marine mammal biology: an evolutionary approach. John Wiley & Sons.

Huelsenbeck, J. P. and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Huelsmann, M., Hecker, N., Springer, M. S., Gatesy, J., Sharma, V. and Hiller, M. (2019). Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. Science advances, 5(9), 6671.

Janeway Jr., C.A., Travers, P., Walport, M. and Shlomchik, M.J. (2001). Evolution of the innate immune system. Immunobiology: The Immune System in Health and Disease, 5th ed., Garland Science.

Kapitonov, V. V. and Jurka, J. (2005). RAG1 core and V (D) J recombination signal sequences were derived from Transib transposons. PLoS Biology, 3(6), e181.

Kapitonov, V.V., Jurka, J. (2003). Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proceedings of the National Academy of Sciences, USA, 100, 6569–6574.

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J.L., et al. (1997). One-megabase sequence analysis of the human immunoglobulin lambda gene locus. Genome Research. 7, 250-261.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. and Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647-1649.

Kirschbaum, T., Pourrajabi, S., Zocher, I., Schwendinger, J., Heim, V., Roschenthaler, F., et al. (1998). The 3’ part of the immunoglobulin kappa locus of the mouse. European Journal of Immunology, 28, 1458-466.

Klein, J., Ono, H., Klein, D. and O’hUigin, C. (1993). The accordion model of MHC evolution. Progress in Immunology, 8, 137-143.

Klein, J., Satta, Y., O'hUigin, C. and Takahata, N. (1993). The molecular descent of the major histocompatibility complex. Annual Review of Immunology, 11(1), 269-295.

Knight, K. L. (1992). Restricted VH gene usage and generation of antibody diversity in rabbit. Annual review of immunology, 10(1), 593-616.

Kokubu, F., Litman, R., Shamblott, M. J., Hinds, K., and Litman, G. W. (1988). Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. The EMBO journal, 7(11), 3413-3422.

Kosakovsky Pond, S. L. and Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution, 22(5), 1208-1222.

Kosakovsky Pond, S. L. and Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular biology and evolution, 22(5), 1208-1222.

Lefranc, M. P. and Lefranc, G. (2001). The Immunoglobulin Facts Book, Academic Press, London, UK.

Lefranc, M.P. and Lefranc, G. (2004). Immunoglobulin lambda (IGL) genes of human and mouse, In: Molecular Biology of B cells (Honjo T., Alt, F. W. and Neuberger, M. S. eds), Academic Press, Elsevier Science, 37-59.

Lundqvist, M., Bengtén, E., Strömberg, S., and Pilström, L. (1996). Ig light chain gene in the Siberian sturgeon (Acipenser baeri). The Journal of Immunology, 157(5), 2031-2038.

Martinez-Jean, C., Folch, G., and Lefranc, M.P. (2001). Nomenclature and overview of the mouse (Mus musculus and Mus sp.) immunoglobulin kappa (IGK) genes. Experimental and Clinical Immunogenetics, 18, 255–279.

Mashoof, S., Criscitiello, M.F. (2016). Fish immunoglobulins. Biology, 5, E45.

McGowen, M.R., Grossman, L.I. and Wildman, D.E. (2012). Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. Proceedings of the Royal Society B: Biological Sciences.

Mcgee, J. L. (2012). Immunological investigations in the West Indian manatee (Trichechus manatus) and Asian elephant (Elephas maximus) (Doctoral dissertation, University of Florida).

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell, 102(5), 553-563.

Murphy, W. J., Eizirik, E., O'Brien, S. J., Madsen, O., Scally, M., Douady, C. J. and Springer, M. S. (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science, 294(5550), 2348-2351.

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A. and Scheffler, K. (2015). Gene-wide identification of episodic selection. Molecular biology and evolution, 32(5), 1365-1371.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K. and Pond, S. L. K. (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genetics, 8(7), e1002764.

Mussmann, R., Du Pasquier, L. and Hsu, E. (1996). Is Xenopus IgX an analog of IgA? European Journal of Immunology, 26, 2823–2830.

Nei, M., Gu, X. and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proceedings of the National Academy of Sciences, 94(15), 7799-7806.

Nery MF, Arroyo JI and Opazo JC. Increased rate of hair keratin gene loss in the cetacean lineage. BMC Genomics. 2014; 15:869.

Nery, M.F., Arroyo, J.I. and Opazo, J.C. (2013a) Genomic organization and differential signature of positive selection in the alpha and beta globin gene clusters in two cetacean species. Genome Biology and Evolution, 5(12), 2359-2367.

Nery, M.F., Arroyo, J.I. and Opazo, J.C. (2013b) Accelerated evolutionary rate of the myoglobin gene in longdiving whales. Journal of Molecular Evolution, 76(6), 380-387.

Nery, M.F., González, D.J. and Opazo, J.C. (2013c) How to make a dolphin: molecular signature of positive selection in cetacean genome. PloS One, 8(6), e65491.

Nguyen, V. K., Desmyter, A. and Muyldermans, S. (2001). Functional heavy-chain antibodies in Camelidae, 261-296.

O'shea, T. J. (1988). The past, present, and future of manatees in the southeastern United States: realities, misunderstandings and enigmas. In Third Southeastern Nongame and Endangered Wildlife Symposium.

Oettinger, M. A., Schatz, D. G., Gorka, C. and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V (D) J recombination. Science, 248(4962), 1517-1523.

Ohta, Y. and Flajnik, M. (2006). IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proceedings of the National Academy of Sciences, USA, 103, 10723–10728.

Ota, T. and Nei, M. (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. Molecular biology and evolution, 11(3), 469-482.

O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M. M., Gormley, N. A. and Cox, A. J. (2015). NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics, 31(12), 2035-2037.

Pond, S. L. K. and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. Statistical methods in molecular evolution, 125-181.

Rambaut, A. (2018). FigTree v1. 4.2, a graphical viewer of phylogenetic trees.

Ramesh, A., Darko, S., Hua, A., Overman, G., Ransier, A., Francica, J.R., et al. (2017). Structure and diversity of the Rhesus macaque immunoglobulin loci through multiple de novo genome assemblies. Frontiers Immunology, 8, 1407.

Rast, J. P., Amemiya, C. T., Litman, R. T., Strong, S. J., and Litman, G. W. (1998). Distinct patterns of IgH structure and organization in a divergent lineage of chondrichthyan fishes. Immunogenetics, 47(3), 234-245.

Rast, J. P., Anderson, M. K., Litman, R. T., Margittai, M., Litman, G. W., Ota, T. and Shamblott, M. J. (1994). Immunoglobulin light chain class multiplicity and alternative organizational forms in early vertebrate phylogeny. Immunogenetics, 40(2), 83-99.

Reeves, R. R., Stewart, B. S., Clapham, P. J. and Powell, J. A. (2002) Guide to marine mammals of the world, 527.

Reynaud, C. A., Dahan, A., Anquez, V., and Weill, J. C. (1989). Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell, 59(1), 171-183.

Rommel, S., Haubold, E., Costidis, A., Bossart, G. and Meisner, R. (2002). Comparative Distribution of Lymph Nodes in Marine Mammals. Proceedings for the Florida Marine Mammal Health Conference, Gainesville, Florida.

Rosnet, O., Blanco-Betancourt, C., Grivel, K., Richter, K., Schiff, C. (2004). Binding of free immunoglobulin light chains to VpreB3 inhibits their maturation and secretion in chicken B cells. Journal of Biological Chemistry, 279 (11), 10228–36.

Rubinstein, N. D., Doron-Faigenboim, A., Mayrose, I. and Pupko, T. (2011). Evolutionary models accounting for layers of selection in protein-coding genes and their impact on the inference of positive selection. Molecular biology and evolution, 28(12), 3297-3308.

Sanchez, P., Nadel, B. and Cazenave, P.A. (1991). V lambda-J lambda rearrangements are restricted within a V-J-C recombination unit in the mouse. European Journal of Immunology, 21, 907-911.

Scheffler, K., Martin, D. P. and Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. Bioinformatics, 22(20), 2493-2499.

Schroeder, H.W., Jr., Hillson, J.L. and Perlmutter, R.M. (1990). Structure and evolution of mammalian VH families. International Immunology, 2, 41–50.

Schwager, J., Burckert, N., Schwager, M. and Wilson, M. (1991). Evolution of immunoglobulin light chain genes: analysis of Xenopus IgL isotypes and their contribution to antibody diversity. EMBO Journal, 10, 505–511.

Schwartz, J. C., Lefranc, M. P. and Murtaugh, M. P. (2012). Organization, complexity and allelic diversity of the porcine (Sus scrofa domestica) immunoglobulin lambda locus. Immunogenetics, 64(5), 399-407.

Schwartz, J.C., Lefranc, M.P., Murtaugh, M.P. (2012a) Evolution of the porcine (Sus scrofa domestica) immunoglobulin kappa locus through germline gene conversion. Immunogenetics, 64, 303-311.

Schwartz, J.C., Philp, R.L., Bickhart, D.M. et al. (2018). The antibody loci of the domestic goat (Capra hircus). Immunogenetics, 70, 317–326.

Schwartz, J.C., Philp, R.L., Bickhart, D.M., Smith, T.P.L. and Hammond, J.A. (2018). The antibody loci of the domestic goat (Capra hircus). Immunogenetics, 70, 317-326.

Scott, C.L. and Potter, M. (1984). Variation in V lambda genes in the genus Mus. Journal of Immunology, 132, 2638-2643.

Sergei, L., Kosakovsky, P., Simon, D.W.F., Spencer V.M. (2005). HyPhy: hypothesis testing using phylogenies, Bioinformatics, 21(5), 676–679.

Shamblott, M. J., and Litman, G. W. (1989). Genomic organization and sequences of immunoglobulin light chain genes in a primitive vertebrate suggest coevolution of immunoglobulin gene organization. The EMBO journal, 8(12), 3733-3739.

Sitnikova, T. and Su, C. (1998). Coevolution of immunoglobulin heavy-and light-chain variable-region gene families. Molecular biology and evolution, 15(6), 617-625.

Smith, N. C., Rise, M. L., and Christian, S. L. (2019). A comparison of the innate and adaptive immune systems in cartilaginous fish, ray-finned fish, and lobe-finned fish. Frontiers in Immunology, 10, 2292.

Springer, M. S., Burk-Herrick, A., Meredith, R., Eizirik, E., Teeling, E., O'Brien, S. J. and Murphy, W. J. (2007). The adequacy of morphology for reconstructing the early history of placental mammals. Systematic Biology, 56(4), 673-684.

Steen, M. L., Hellman, L., and Pettersson, U. (1987). The immunoglobulin lambda locus in rat consists of two Cλ genes and a single Vλ gene. Gene, 55(1), 75-84.

Sun, Y., Wang, C., Wang, Y., Zhang, T., Ren, L., Hu, X., et al. (2010). A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. Developmental and Comparative Immunology, 34, 1009-1020.

Sun, Y., Wei, Z., Li, N. and Zhao, Y. (2013) A comparative overview of immunoglobulin genes and the generation of their diversity in tetrapods. Developmental and Comparative Immunology, 39, 103–109.

Suyama, M., Torrents, D. and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research, 34, W609-W612.

Sweat, J.M., Johnson, C.M., Marikar,Y. And Gibbs, E.P. (2005). Characterization of surface interleukin-2 receptor expression on gated populations of peripheral blood mononuclear cells from manatees, Trichechus manatus latirostris. Veterinary Immunology and Immunopathology, 108, 269-281.

Thompson, C. B. (1995). New insights into V (D) J recombination and its role in the evolution of the immune system. Immunity, 3(5), 531-539.

Thompson, C. B., and Neiman, P. E. (1987). Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. Cell, 48(3), 369-378.

Tomlinson, I. M., Walter, G., Jones, P. T., Dear, P. H., Sonnhammer, E. L. and Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. 813-817.

Tonegawa, S. (1976). Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. Proceedings of the National Academy of Sciences, 73(1), 203-207.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature, 302(5909), 575-581.

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic acids research, 44(W1), W232-W235.

Turchin, A., and Hsu, E. (1996). The generation of antibody diversity in the turtle. The Journal of Immunology, 156(10), 3797-3805.

Vianna, J. A., Bonde, R. K., Caballero, S., Giraldo, J. P., Lima, R. P., Clark, A. and Rodríguez‐Lopez, M. A. (2006). Phylogeography, phylogeny and hybridization in trichechid sirenians: implications for manatee conservation. Molecular Ecology, 15(2), 433-447.

Walsh, C.J., Luer, C.A. and Noyes, D.R. (2004). Effects of environmental stressors on lymphocyte proliferation in Florida manatees, Trichechus manatus latirostris. Veterinary Immunology and Immunopathology, 103, 247-256.

Walther S, Rusitzka TV, Diesterbeck US and Czerny CP. (2015). Equine immunoglobulins and organization of immunoglobulin genes. Developmental and Comparative Immunology, 53, 303-319.

Wang, X., Cheng, G., Lu, Y., Zhang, C., Wu, X., Han, H., ... and Ren, L. (2016). A comprehensive analysis of the phylogeny, genomic organization and expression of immunoglobulin light chain genes in Alligator sinensis, an endangered reptile species. PloS one, 11(2), e0147704.

Watson, C.T., Steinberg, K.M., Graves, T.A., Warren, R.L., Malig, M., Schein, J., et al. (2015). Sequencing of the human IG light chain loci from a hydatidiform mole BAC library reveals locus-specific signatures of genetic diversity. Genes & Immunity, 16, 24-34.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V. and Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Molecular biology and evolution, 35(3), 773-777.

Weigert, M. G., Cesari, I. M., Yonkovich, S. J. and Cohn, M. (1970). Variability in the lambda light chain sequences of mouse antibody. Nature, 228(5276), 1045-1047.

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. and Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. Molecular biology and evolution, 32(3), 820-832.

Wilson DE, Reeder DM. 2005 Mammal species of the world: a taxonomic and geographic reference. Baltimore, MD: Johns Hopkins University Press.

Wilson, M., Hsu, E., Marcuz, A., Courtet, M., Du Pasquier, L., and Steinberg, C. (1992). What limits affinity maturation of antibodies in Xenopus--the rate of somatic mutation or the ability to select mutants?. The EMBO Journal, 11(12), 4337-4347.

Xu, S., Tian, R., Lin, Y., Yu, Z., Zhang, Z., Niu, X., Wang, X., and Yang, G. (2019). Widespread positive selection on cetacean TLR extracellular domain. Molecular Immunology, 106, 135-142.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution, 24(8), 1586-1591.

Yim, H.S., Cho, Y.S., Guang, X., Kang, S.G., Jeong, J.Y., Cha, S.S., Oh, H.M., Lee, J.H., Yang, E.C., Kwon, K.K. et al. (2014). Minke whale genome and aquatic adaptation in cetaceans. Nature Genetics, 46(1), 88-92.

Zhang, J., Nielsen, R. and Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular biology and evolution, 22(12), 2472-2479.

Zhao, Y., Pan-Hammarström, Q., Yu, S., Wertz, N., Zhang, X., Li, N., ... and Hammarström, L. (2006). Identification of IgF, a hinge-region-containing Ig class, and IgD in Xenopus tropicalis. Proceedings of the National Academy of Sciences, 103(32), 12087-12092.

Zhou, X., Sun, F., Xu, S., Fan, G., Zhu, K., Liu, X., Chen, Y., Shi, C., Yang, Y., Huang, Z. et al. (2013). Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptation. Nature Communications, 4, 2708.

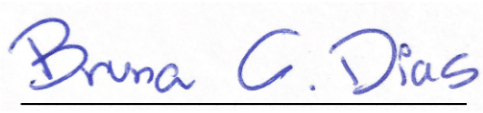
Zimmerman, A. M., Yeo, G., Howe, K., Maddox, B. J. and Steiner, L. A. (2008). Immunoglobulin light chain (IgL) genes in zebrafish: Genomic configurations and inversional rearrangements between (VL–JL–CL) gene clusters. Developmental & Comparative Immunology, 32(4), 421-434.

ATTACHMENTS

DECLARAÇÃO

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Evolução molecular de genes do sistema imunológico adaptativo em mamíferos aquáticos**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 25 de Novembro de 2021



Bruna Cristina Dias
RG: 40.024.331-3

Assinatura : _____
Nome do(a) orientador(a):
RG n.° 34.544.823-6

Ministério do Meio Ambiente
CONSELHO DE GESTÃO DO PATRIMÔNIO GENÉTICO

SISTEMA NACIONAL DE GESTÃO DO PATRIMÔNIO GENÉTICO E DO CONHECIMENTO TRADICIONAL ASSOCIADO

Comprovante de Cadastro de Acesso

Cadastro nº AB375AB

A atividade de acesso ao Patrimônio Genético, nos termos abaixo resumida, foi cadastrada no SisGen, em atendimento ao previsto na Lei n° 13.123/2015 e seus regulamentos.

Número do cadastro: **AB375AB**
Usuário: **UNICAMP**
CPF/CNPJ: **46.068.425/0001-33**
Objeto do Acesso: **Patrimônio Genético**
Finalidade do Acesso: **Pesquisa**

Espécie

Inia geoffrensis

Sotalia fluviatilis

Título da Atividade: **USANDO GENÔMICA COMPARATIVA PARA ENTENDER A EVOLUÇÃO CONVERGENTE EM MAMÍFEROS: EM BUSCA DAS PEGADAS MOLECULARES DA OCUPAÇÃO DO AMBIENTE MARINHO E FLUVIAL**

Equipe

Mariana Freitas Nery	**UNICAMP**
Bruna Cristina Dias	**UNICAMP**
Érica Martinha Silva de Souza	**UNICAMP**
Lucas Eduardo Costa Canesin	**UNICAMP**
Elisa Karen da Silva Ramos	**UNICAMP**
Pedro de Gusmão Ribeiro	**UNICAMP**
Maria Luiza Andreani	**UNICAMP**
Letícia Magpali Moura Estevão	**UNICAMP**

João Pedro Vieira Mariz	**UNICAMP**
Juliana Mitie Pereira	**UNICAMP**
Michelle Carneiro Rachid Ribeiro	**UNICAMP**
Flávia Akemi Nitta Fernandes	**UNICAMP**

Data do Cadastro: **21/02/2018 20:48:11**

Situação do Cadastro: **Concluído**

Ministério do Meio Ambiente
CONSELHO DE GESTÃO DO PATRIMÔNIO GENÉTICO

SISTEMA NACIONAL DE GESTÃO DO PATRIMÔNIO GENÉTICO E DO CONHECIMENTO TRADICIONAL ASSOCIADO

Comprovante de Cadastro de Acesso

Cadastro nº A94D205

A atividade de acesso ao Patrimônio Genético, nos termos abaixo resumida, foi cadastrada no SisGen, em atendimento ao previsto na Lei n° 13.123/2015 e seus regulamentos.

Número do cadastro: **A94D205**
Usuário: **UNICAMP**
CPF/CNPJ: **46.068.425/0001-33**
Objeto do Acesso: **Patrimônio Genético**
Finalidade do Acesso: **Pesquisa**

Espécie

Trichechus Inunguis

Título da Atividade: **Usando genômica comparativa para entender a evolução convergente em mamíferos: em busca das pegadas moleculares da ocupação do ambiente marinho e fluvial**

Equipe

Mariana Freitas Nery	**UNICAMP**
Bruna Cristina Dias	**UNICAMP**
Érica Martinha Silva de Souza	**UNICAMP**
Lucas Eduardo Costa Canesin	**UNICAMP**
Elisa Karen da Silva Ramos	**UNICAMP**
Pedro de Gusmão Ribeiro	**UNICAMP**
Maria Luiza Andreani	**UNICAMP**
Letícia Magpali Moura Estevão	**UNICAMP**
João Pedro Vieira Mariz	**UNICAMP**

Juliana Mitie Pereira **UNICAMP**
Michelle Carneiro Rachid Ribeiro **UNICAMP**
Flávia Akemi Nitta Fernandes **UNICAMP**

Data do Cadastro: **23/02/2018 20:42:30**

Situação do Cadastro: **Concluído**

Conselho de Gestão do Patrimônio Genético

Situação cadastral conforme consulta ao SisGen em **10:18** de **05/12/2018**.

SISTEMA NACIONAL DE GESTÃO DO PATRIMÔNIO GENÉTICO E DO CONHECIMENTO TRADICIONAL ASSOCIADO - SISGEN

Ministério do Meio Ambiente
CONSELHO DE GESTÃO DO PATRIMÔNIO GENÉTICO

SISTEMA NACIONAL DE GESTÃO DO PATRIMÔNIO GENÉTICO E DO CONHECIMENTO TRADICIONAL ASSOCIADO

Comprovante de Cadastro de Acesso

Cadastro nº AB72D32

A atividade de acesso ao Patrimônio Genético, nos termos abaixo resumida, foi cadastrada no SisGen, em atendimento ao previsto na Lei n° 13.123/2015 e seus regulamentos.

Número do cadastro: **AB72D32**
Usuário: **UNICAMP**
CPF/CNPJ: **46.068.425/0001-33**
Objeto do Acesso: **Patrimônio Genético**
Finalidade do Acesso: **Pesquisa**

Espécie

Pontoporia blainvillei

Sotalia Guianensis

Título da Atividade: **Usando genômica comparativa para entender a evolução convergente em mamíferos: em busca das pegadas moleculares da ocupação do ambiente marinho e fluvial**

Equipe

Mariana Freitas Nery	**UNICAMP**
Bruna Cristina Dias	**UNICAMP**
Érica Martinha Silva de Souza	**UNICAMP**
Lucas Eduardo Costa Canesin	**UNICAMP**
Elisa Karen da Silva Ramos	**UNICAMP**
Pedro de Gusmão Ribeiro	**UNICAMP**
Maria Luiza Andreani	**UNICAMP**
Letícia Magpali Moura Estevão	**UNICAMP**

João Pedro Vieira Mariz	**UNICAMP**
Juliana Mitie Pereira	**UNICAMP**
Michelle Carneiro Rachid Ribeiro	**UNICAMP**
Flávia Akemi Nitta Fernandes	**UNICAMP**

Envios de Amostra

Espécie: **Pontoporia blainvillei**

Tipo do Patrimônio Genético: **-**

Forma do Patrimônio Genético: **Amostra líquida em tubo de microcentrífuga (eppendorf)**

Instituicão Destinatária: **Macrogen Inc.**

Sede da Instituição Destinatária: **254 Beotkkot-ro, Seoul, Geumcheon-gu, 08511, Coreia do Sul**

Espécie: **Sotalia guianensis**

Tipo do Patrimônio Genético: **-**

Forma do Patrimônio Genético: **Amostra líquida em tubo de microcentrífuga (eppendorf)**

Instituição Destinatária: **Macrogen Inc.**

Sede da Instituição Destinatária: **254 Beotkkot-ro, Seoul, Geumcheon-gu, 08511, Coreia do Sul**

Data do Cadastro: **01/03/2018 16:56:30**

Situação do Cadastro: **Concluído**