

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

EVOLUÇÃO DE GENES DO CLUSTER *HOXD* NA ORDEM CETARTIODACTYLA

BRENDA GODOY ALEXANDRE

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do Grau de Mestre em Genética e Biologia Molecular.

Orientador: Prof. Dr. Thales Renato Ochotorena de Freitas

Co-orientadora: Prof. Dr. Maríndia Deprá

Porto Alegre Julho de 2019

Instituições e fontes financiadoras

Este trabalho foi desenvolvido no Laboratório de Citogenética e Evolução Animal do Departamento de Genética do Instituto de Biociências da Universidade Federal do Rio Grande do Sul. A aluna recebeu bolsa de estudos concedida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Instituições Financiadoras:

CNPq - Conselho Nacional de Pesquisa.

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

FAPERGS - Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul.

Ao meu avô Manoel Alexandre, por ter me ensinado sobre o que é o verdadeiro amor, sobre o valor do trabalho e da dedicação para a construção do caráter, e o poder que temos em nossas próprias mãos para moldar nosso destino. Dedico!

Agradecimentos

Ao meu orientador Thales Renato O. de Freitas agradeço pelo acolhimento e por desde a primeira vez acreditar em mim, por me dar a oportunidade de trabalhar com aquilo que amo e fazer o meu projeto (uma curiosidade de criança) se tornar essa dissertação. Obrigada pelo apoio e pela orientação, que sempre desenvolveu de forma amigável e confiante, e agradeço por me tornar parte dessa família que é o Lab Tucão. Grata a todos os colegas por fazer do ambiente de trabalho um lugar tão descontraído, hospitaleiro e caloroso, não vou citar todos os nomes, mas sou muito agradecida por ter encontrado pessoas e amizades tão especiais, obrigada pelas conversas sempre empáticas e construtivas, me tornando uma cientista e uma pessoa melhor.

A minha co-orientadora Maríndia, que me acompanha desde que eu era uma caloura da biologia, querendo aprender mais sobre evolução e embriões, agradeço por todos os ensinamentos, (que foram tantos) os quais ajudaram a construir a profissional que sou hoje, por sempre ter acreditado em mim e por todo o carinho e apoio desde o primeiro dia. Assim como a todos do laboratório de Drosófila, onde cresci (por dentro né pessoal rs) e passei tanto anos, que me acompanham até hoje, sempre de portas e coração abertos para uma boa conversa e um cafezinho, me ajudando não só na área da pesquisa, mas como amigos queridos. Um agradecimento especial à professora Vera, que foi sempre uma inspiração para mim como pesquisadora, professora e mulher, e foi através de seus ensinamentos que ideias sobre evo devo brotaram na minha cabecinha.

Agradeço a existência e amizade de umas pessoas muito especiais que fizeram parte da construção dessa dissertação, me ajudando com ideias e ensinamentos, me apoiando nos momentos de desespero, me recebendo e me auxiliando sempre que precisei, tornando tudo mais tranquilo e com um resultado muito mais bonito, afinal nada conseguimos sozinhos. Obrigada Rebeca, Natasha, Thays, Gilberto e Rafa por segurarem todas as barras comigo. Não teria conseguido sem vocês, vocês moram no meu coração.

Aos meus amigos, nem sei dizer o que faria sem eles, tenho muita sorte de ser rodeada de

peessoas que me cercam de amor, sempre colocando minha autoestima lá em cima, obrigada pelo apoio, principalmente a algumas amigas-irmãs que me acompanham faz um tempo e que fazem dos meus dias mais leves e felizes, obrigada Nati, Lilith, Ilana, Paqui, Jazz e Kel, por sempre estarem aqui. Amo vocês!

Ao meu companheiro Leon, obrigada por todo incentivo, apoio e carinho. Por sempre me fazer esquecer os problemas e rir muito, grata por toda leveza e amor que tu sempre me dá.

A minha família, que é tão maravilhosa, sou tão sortuda de ter nascido nela, cresci com as pessoas mais inspiradoras ao meu redor. Obrigada as minhas avós Rosalina e Erondina, mulheres tão admiráveis, as mais fortes e amorosas que já conheci, me ensinaram a amar todas as pequenas formas de vida, guerreiras nas quais me espelho todos os dias. Ao meu avô Manoel, meu super-herói, construtor de casas e de amor, construiu essa família linda e fez com as próprias mãos um lar para todos. Obrigada por me ensinarem o valor do suor, e foi através do suor deles, que hoje estou aqui, esse mestrado é nosso. Amo vocês mais do que cabe em mim. Agradeço aos meus primos, tive a sorte de crescer com meus melhores amigos, obrigada por sempre estarem ao meu lado e tornarem tudo mais divertido. Principalmente as minhas primas Ju e Manu, que são minhas irmãs mais velhas, e me deram meus maiores presentes, o Heitor e a Lis, meus filhos de coração.

A minha maninha Carolina, obrigada por mesmo entre trancos e barrancos sempre ser minha maior parceria desde as fraldas. E finalmente aos meus pais, obrigada pelo amor e incentivo sem tamanho. Ao meu pai, pela presença forte, por saber que não importa a situação posso sempre contar contigo, meu companheiro de aventuras, obrigada pelos genes da teimosia e persistência, e por me transmitir aquela que se tornou a minha (quer dizer nossa) maior paixão, ao qual nenhum de nós dois consegue viver longe, que é o mar, o que me fez trilhar esse caminho que me trouxe até aqui. Gratidão a minha mãe, essa pessoa que tem tanto amor e alegria, que nunca deixou de acreditar em mim, minha melhor amiga com certeza. Grata por nos trazer até aqui, que segurou minha mão e lutou do meu lado, não só a primeira Godoy formada, mas também mestre, essa conquista é nossa, e tão tua quanto minha. Amo muito vocês! Muito Obrigada!

Sumário

Agradecimentos	4
Lista de abreviaturas	7
Resumo	8
Abstract	9
INTRODUÇÃO	11
Ecologia e evolução da ordem Cetartiodactyla.....	11
Os genes do desenvolvimento na evolução dos membros.	15
Elementos transponíveis e seu papel na evolução.	20
OBJETIVOS	27
Objetivos Gerais.....	27
Objetivos Específicos	27
MATERIAIS E MÉTODOS	28
Amostragem, extração de DNA, amplificação e sequenciamento	28
Análises evolutivas do <i>HoxD12</i>	32
Identificação de elementos transponíveis e sítios de ligação de fatores de transcrição em região do cluster <i>HoxD</i>	33
RESULTADOS	37
Árvore por inferência Bayesiana	37
Análises evolutivas do <i>HoxD12</i>	40
Elementos transponíveis e sítios de ligação de fatores de transcrição dentro de região do cluster <i>HoxD</i> ..	42
DISCUSSÃO	46
História evolutiva do <i>HoxD12</i> em Cetartiodactyla	46
Elementos transponíveis em região no cluster <i>HoxD</i>	49
Sítios de ligação de fatores de transcrição nos elementos MIR	54
CONSIDERAÇÕES FINAIS	57

Lista de abreviaturas

AIC - *Akaike Information Criterion*

BEB - *Bayes Empirical Bayes*

cDNA – DNA complementar

Evx2 - *Even-Skipped Homeobox 2*

Hox - *Homeobox*

LINES - *long interspersed nuclear element*

Lnp - *Limb and neural pattern*

LRT - *log likelihood ratio test*

LTR - *long terminal repeat*

MIR - *Mammalian-wide Interspersed Repeat*

mRNA – RNA mensageiro

PCR - *Polymerase Chain Reaction*

PP - valores de probabilidade posterior

PRC2 - *Polycomb Repressive Complex 2*

Shh - *Sonic hedhog*

SINE - *short interspersed nuclear element*

TBP - *TATA binding protein* (proteína de ligação ao TATA-box)

TE – *Transposable element*

TFBS - do inglês *transcription fator binding site*

TFs – do inglês *transcription factors*

Resumo

A ordem Cetartiodactyla é composta por dois grupos de mamíferos que até recentemente não eram relacionados, Artiodactyla e Cetacea. Evidências morfológicas e moleculares sugerem esses dois grupos possuem um ancestral em comum. Os ancestrais dos cetáceos passaram por uma transição da vida terrestre para a aquática, adquirindo membros anteriores em forma de nadadeiras e posteriores ausentes. Os genes *Hox* possuem papel central na especificação dessas estruturas corporais. O *cluster HoxD* foi cooptado na evolução de tetrápodes para a função de regular o crescimento e padronização de membros. Também foi demonstrado que a acumulação e atividade transcricional de elementos transponíveis (TEs) inseridos nos genomas entre o *cluster HoxD* pode estar associada com a mudança de planos corporais. Assim, temos como objetivo contribuir para o melhor entendimento da história evolutiva em Cetartiodactyla, avaliando as relações entre os complexos de genes no *cluster HoxD* na evolução dos membros desse grupo. A partir de extrações de DNA de amostras de tecido de espécies ainda não analisadas, como *Hippopotamus sp* (espécie de artiodáctilo mais próximo aos cetáceos), foi amplificado e sequenciado o gene *HoxD12*. Foi construído por inferência Bayesiana uma árvore evolutiva a partir dessas sequências junto com sequências de mamíferos disponíveis no *genbank*, além disso foi analisada as taxas evolutivas em mamíferos e Cetartiodactyla. Também foram analisadas através da plataforma *online RepeatMasker*, sequências obtidas no banco de dados *genbank* entre os genes *HoxD10* e *Evx2* em mamíferos em busca de elementos transponíveis nessa região. Pela plataforma *Alibaba* foram feitas previsões quanto a presença de sítios para fatores de transcrição nos elementos mais encontrados. Foram identificados 32 elementos transponíveis diferentes. Desses elementos, destacou-se os da família SINE/MIR, presentes predominantemente na região intergênica entre *Evx2* e *HoxD13*. E foram encontrados principalmente 3 sítios para fatores de transcrição dentro de elementos MIR. A árvore filogenética correspondeu a dados da literatura, com *Hippopotamus sp* sendo mais relacionado com Cetacea do que Artiodactyla, além disso Cetacea apresentou maiores taxas evolutivas médias que Artiodactyla, apesar de serem em um menor número de sítios.

Abstract

The order Cetartiodactyla is composed of two groups of mammals that until recently were not related: Artiodactyla and Cetacea. Recent morphological and molecular evidence suggests the evolution of Cetacea from a common ancestor with Artiodactyla. Cetacea ancestors had transitioned from terrestrial to aquatic life, acquiring hydrodynamic bodies with fin-shaped hindlimbs and absent forelimbs. *Hox* genes play a central role in the specification of these body structures, and modifications on their expression patterns have been linked to the emergence of vertebrate body plans. The *HoxD* cluster was recently co-opted in evolution, acquiring the function of regulating the growth and standardization of limbs as well as digits and the absence of this loci leads to rudimentary and truncated appendages. It has also been shown that the accumulation and transcriptional activity of transposable elements (TEs) inserted into the genomes, in the regions between the genes *HoxD10* and *Evx2* (even-skipped homeobox 2), may be associated with the change of body plan. Thus, our objective is to contribute to better understanding the Cetartiodactyla evolutionary history, evaluating the relationships between gene complexes in the *HoxD* cluster. DNA was extracted from 10 non-analyzed species, such as *Hippopotamus sp* (the species of artiodactyl closest to cetaceans). The *HoxD12* gene was amplified by Polymerase Chain Reaction and sequenced. These sequences and the mammalian sequences available on *GenBank* were used to generate an evolutionary tree using Bayesian Inference. The evolutionary rates in mammals, and particularly in Cetartiodactyla, were analyzed. Also, sequences from the region between *HoxD10* and *Evx2* from mammals, obtained from the *GenBank* database, were analyzed in the RepeatMasker online platform, searching for TEs in this region. Predictions were made regarding the presence of sites for transcription factors in the most found elements, in *AliBaba2.1* platform. Thirty-two different transposable elements (TE) were identified. Among them, the elements of the SINE / MIR family, predominantly present in the intergenic region between *Evx2* and *HoxD10*, were highlighted. Were found three transcription factors sites within MIR elements. The phylogenetic tree corroborates previous studies, with *Hippopotamus sp* being more related to

Cetacea than to Artiodactyla. Besides, Cetacea presented higher average evolutionary rates than Artiodactyla, despite the smaller number of sites with faster evolution.

INTRODUÇÃO

Ecologia e evolução da ordem Cetartiodactyla

Cetartiodactyla é uma ordem composta por dois grupos que até recentemente eram classificados como duas ordens de mamíferos separadas por sua morfologia que se diferencia de forma contrastante. Esses grupos são: Artiodactyla, que são mamíferos ungulados com número par de dígitos, por exemplo, bovídeos e cervídeos e Cetacea, mamíferos especializados para hábitos de vida aquático, representada pelos animais conhecidos popularmente como baleias e golfinhos. Evidências morfológicas, paleontológicas e moleculares de caráter recente sugerem que Cetacea evoluiu de antepassados em comum com artiodáctilos, tornando Artiodactyla não monofilético, a menos que Cetacea esteja incluída. Especialistas sugerem que o clado monofilético que representa artiodáctilos e cetáceos seja chamado Cetartiodactyla. (Boisserie et al., 2005; Gatesy et al., 1999; Lum et al., 2000; Milinkovitch et al., 1994; Nikaido et al., 1999; Taylor et al., 2007; Thewissen et al., 2001; Zurano et al., 2019). Assim, atualmente a ordem Cetartiodactyla é dividida em 5 subordens: Acodonta, Cetacea, Ruminantia, Suina e Tylopoda (Hassanin et al., 2012).

Os cetáceos são ecomorfológicamente diversos, a maioria das espécies é marinha, vivendo em águas oceânicas e também costeiras, mas existem espécies que ocupam sistemas fluviais (Jefferson et al., 2007). Já que esses animais são mamíferos, isso implica que possuem uma origem em terra. Portanto os ancestrais dos cetáceos passaram por uma transição da vida terrestre para a aquática (Figura 1), adquirindo corpos hidrodinâmicos, com nadadeiras peitorais como membros anteriores, geralmente uma nadadeira dorsal, uma nadadeira caudal e ausência de membros posteriores externos, embora estejam presentes em seus embriões (figura 2) (Thewissen et al., 2006).

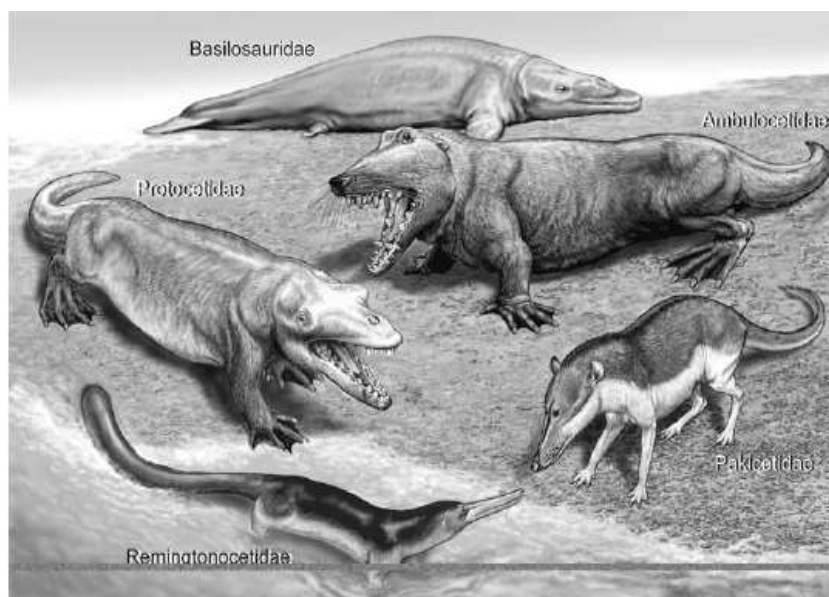


Figura 1. A mudança do plano corporal dos cetáceos durante os primeiros dez milhões de anos de evolução, representados por cinco famílias de arqueocetos (Thewissen et al., 2009).

Em 2001, ossos de tornozelo de dois cetáceos do Eoceno foram descobertos no Paquistão (Gingerich et al., 2001; Thewissen et al., 2001). O astrágalo (também chamado de tálus) é o osso do tornozelo em que o pé articula com a tíbia (osso da canela). A cabeça do astrágalo dos artiodáctilos é única entre os mamíferos com a forma de tróclea (Schaeffer, 1947). Os ossos de tornozelo de cetáceos ancestrais apresentavam o mesmo formato (Gingerich et al., 2001; Thewissen et al., 2001). O que foi corroborado com as filogenias moleculares que propunham que os cetáceos eram parentes próximos dos artiodáctilos (Geisler and Uhen, 2005).

Este achado levou os paleontologistas a procurarem o parente mais próximo dos cetáceos entre os artiodáctilos. Embora evidências moleculares indiquem que o parente moderno mais próximo dos cetáceos seja o hipopótamo, o registro fóssil dessa família remonta apenas a 15 milhões de anos atrás (Boisserie et al., 2005), enquanto os cetáceos mais antigos têm mais de 50 milhões de anos. Os fósseis fragmentários da família Raoellidae, de artiodáctilos endêmicos da Índia, sugerem que estes poderiam estar intimamente relacionados com os cetáceos (Geisler and Uhen, 2005), e um material fóssil mais completo para raoelídeos expôs claramente essa

proximidade (Thewissen et al., 2007). Análises filogenéticas indicam que os raoelídeos são o grupo mais próximo dos cetáceos e que os hipopótamos são o grupo irmão de raoellidae-cetacea combinado (Geisler and Theodor, 2009), conforme pode se observar na figura 3.

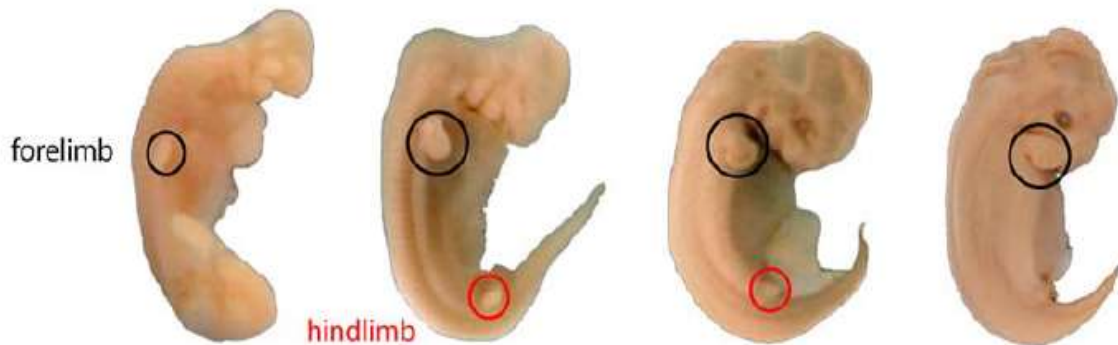


Figura 2. Embriões do golfinho *Stenella attenuata* entre aproximadamente entre as semanas 4 a 9 do desenvolvimento embrionário. O primeiro embrião mostra membros anteriores, mas não membros posteriores (na maioria dos mamíferos, os membros anteriores se desenvolvem antes dos membros posteriores). No segundo embrião, os membros posteriores começaram a se formar, mas seu desenvolvimento cessa e desaparecem lentamente à medida que o embrião se desenvolve (Thewissen et al., 2009).

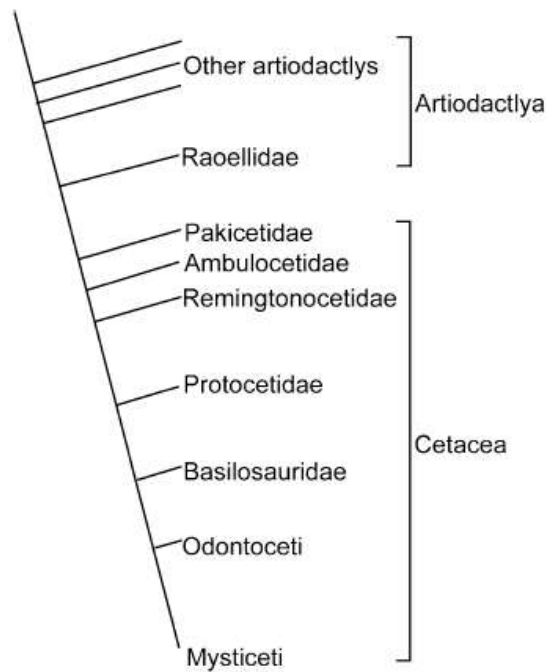


Figura 3. Cladograma mostrando a relação dos cetáceos com seus ancestrais terrestres (Thewissen et al., 2009).

Assim como os cetáceos modernos, os primitivos viviam na água. No entanto, estes foram caminchantes lentos, provavelmente semelhantes ao hipopótamo moderno. É improvável que esses mamíferos fossem nadadores rápidos, porque os ossos dos membros com sua camada cortical grossa aumentariam significativamente a inércia e o arrasto. Em contraste, a maioria dos cetáceos modernos tem osso cortical muito fino e leve, auxiliando em uma natação mais ágil (Grey and Sadoff, 2007).

Segundo Thewissen et al. (2006), durante a evolução dos cetáceos, a hidrodinâmica foi um fator preponderante, assim a forma de torpedo e a perda de membros posteriores foi sendo selecionada por ser mais adaptativa, lhes proporcionando uma natação mais eficiente. A seleção para a perda dessas estruturas deve ter sido forte quando os cetáceos se tornaram predadores de perseguição no Eoceno tardio.

Os membros posteriores dos basilosaurídeos, ancestrais dos cetáceos modernos, retêm os ossos presentes nos mamíferos terrestres: fêmur, tíbia, fíbula, tarsos, metatarsos e falanges

(Gingerich et al., 1990). No entanto, são muito reduzidos em tamanho e a pelve não está presa à coluna vertebral, tornando-os inadequados para suportar o peso corporal desses animais. Não há membros posteriores externos em cetáceos modernos, embora, muito raramente, um indivíduo anômalo nasça com esses membros. Internamente, existem remanescentes em todas as espécies, que podem se relacionar com a origem dos músculos para os órgãos genitais (figura 3). Em algumas espécies, pélvis, fêmur e tíbia estão presentes (Thewissen et al., 2009).

As modificações nos membros e no número de dígitos ocorreram repetidamente durante a evolução dos tetrápodes e geralmente são vistas como respostas adaptativas às mudanças no estilo de vida e no uso do habitat, suprimindo necessidades de diferentes tipos de locomoção (Shubin et al., 2015).

Apesar de existirem outros mamíferos com modificações nos membros para adaptação na vida aquática, como em sirênios, pinípedes, algumas espécies de mustelídeos (como por exemplo, a lontra-marinha) e o urso-polar, a reorganização evolutiva mais dramática ocorreu nos cetáceos. Alterações na morfologia dos membros associadas ao retorno à água seguiram três rotas principais. Primeiro, os dígitos tornaram-se incorporados no tecido da pele (uma rede interdigital), um estado também observado em outros mamíferos. Em segundo lugar, os cetáceos perderam seus membros posteriores (Berta et al., 2006; Thewissen et al., 2006). Em terceiro lugar, os cetáceos mostram uma variação considerável no número de falanges, com os dígitos I e V tipicamente caracterizados por poucas falanges (variando de zero a cinco) e dígitos II-IV com até 13 no segundo raio digital de algumas espécies. De fato, os cetáceos mostram polimorfismos claros espécie-específicos, e a contagem de falange pode até ter variação intraespecífica (Cooper et al., 2007).

Os genes do desenvolvimento na evolução dos membros.

A biologia evolutiva do desenvolvimento (conhecida como Evo-Devo) estuda uma abordagem comparativa dos mecanismos do desenvolvimento embrionário, de modo a esclarecer como os genes poderiam gerar novas formas, funções e comportamentos no curso temporal da evolução (Gellon e McGinnis, 1998; Amores et al, 2004). Os genes responsáveis por regular e

dirigir o desenvolvimento de determinados segmentos e estruturas corporais são chamados de genes homeóticos. A maioria dos genes homeóticos decodifica fatores de transcrição de proteínas que contém um homeodomínio de ligação ao DNA e esses genes são chamados genes *Hox* (Gellon and McGinnis, 1998).

Os genes *Hox* (*homeobox*) são centrais para a especificação de estruturas ao longo do eixo anterior-posterior do corpo (Deschamps, 2005; Kmita et al., 2002) e modificações em sua expressão surgiu paralelamente à diversidade de planos corporais encontrados nos animais (Burke et al., 1995; Gaunt, 1994). Os genes *Hox* são agrupados, e seus domínios de expressão, no tempo e no espaço, refletem suas respectivas ordens genômicas. Em tetrápodes, esses genes são classificados em 13 grupos de paralogia e são fortemente agrupados em quatro loci: *HoxA* a *HoxD*. Embora esses genes sejam bem conservados entre os organismos, sendo encontrados em todos os filos de metazoários (Graham et al., 1989; Gaunt, 1994), recentes estudos têm destacado uma diversidade inesperada no número de genes *Hox*, organização genômica e padrões de expressão (Garcia-Fernández, 2005; Pearson et al., 2005).

Nos últimos anos, foram feitos progressos consideráveis na compreensão das bases moleculares do controle de padrões digitais e na formação dos membros. Muitos genes expressos no botão embrionário de membros demonstraram desempenhar um papel importante na formação desses (Tickle, 2015).

O cluster *HoxD* foi cooptado mais recentemente na evolução para regular o crescimento e padronização de membros e dos dígitos (Gérard et al., 1993). Duas fases da expressão de *HoxD* são importantes no desenvolvimento e padronização de membros (Kmita et al., 2005; Zakany et al., 2007). A primeira fase resulta na expressão dos genes 3'-*HoxD* (*HoxD1-9*). Isso restringe a expressão de 5'-*HoxD* para o lado posterior do broto do membro distal e é necessário para o crescimento do membro, o desenvolvimento do membro proximal, a polaridade anterior-posterior do membro e a expressão posterior do gene *Sonic hedgehog* (*Shh*). Essa colinearidade temporal e espacial de 3'-5' é remanescente da regulação no eixo embrionário principal, que é acompanhada

pela perda progressiva da trimetilação da lisina 27 na histona H3 (H3K27me3) catalisada pela PRC2 (do inglês *Polycomb Repressive Complex 2*) (Montavon et al., 2011). No entanto, também pode exigir elementos 3'-regulatórios ainda indefinidos de *HoxD* (Spitz et al., 2005)

Uma fase posterior da expressão de *HoxD* no membro distal é necessária para a morfogênese digital (Spitz et al., 2003). Esta é caracterizada pela "colinearidade quantitativa", na qual a expressão do gene *HoxD13* com mais 5 genes é inicialmente mais forte no mesênquima distal posterior, com expressão progressivamente menos forte de *HoxD12* a *HoxD10*. O domínio de expressão de *HoxD13*, em seguida, se espalha anteriormente, de modo que é o único gene *HoxD* cuja expressão é robusta o suficiente para ser detectável no lado mais anterior (Montavon et al., 2008). Isso é impulsionado por elementos potenciadores, incluindo uma região de controle global (GCR do inglês *global control region*) de aproximadamente 40 kb localizada 180kb a montante do gene *HoxD13*, e o intensificador *Prox* localizado entre *Evx2* (do inglês *Even-Skipped Homeobox 2*) e *Lnp* (do inglês *Limb and neural pattern*) (Spitz et al., 2003; Tschopp and Duboule, 2011).

Todos os ossos longos do membro requerem a ativação dos genes *Hox* em combinações diferentes, embora parcialmente sobrepostas. Inicialmente, *HoxD9* ao *HoxD11* e *HoxA11* são expressos no membro proximal em desenvolvimento (Figura 4). Posteriormente, em uma segunda fase da ativação transcricional, *HoxD9* ao *HoxD13* e *HoxA13* são expressos nos dígitos. A existência de módulos regulatórios distintos para ossos longos em ambos os lados da articulação mesopodial (o punho e o tornozelo), juntamente com as trajetórias evolutivas separadas desses elementos, apoiou a visão de que os membros dos tetrápodes são geneticamente organizados seguindo um padrão bimodal específico de ossos longos proximais (braço e antebraço) e distais (dígitos) (Tschopp and Duboule, 2011).

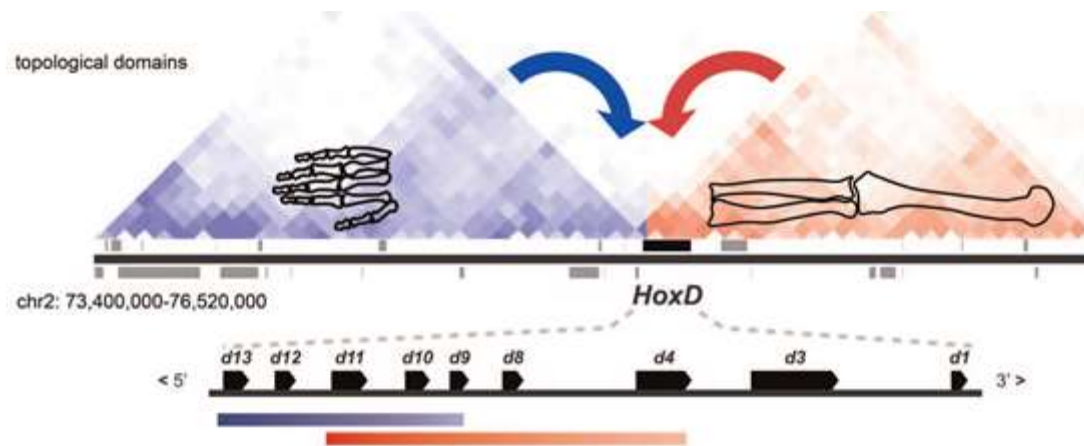


Figura 2. Domínios de expressão do gene *HoxD* proximal (vermelho) e distal (azul) em um membro em desenvolvimento (Woltering et al., 2014)

Entre os genes *Hox* expressos em dígitos durante o desenvolvimento, *HoxA13*, é o único gene do cluster *HoxA* expresso neste local, é igualmente transcrito nos primórdios de todos os dígitos, incluindo o polegar, o que difere dos genes *HoxD*. Os genes *HoxD10*, *HoxD11* e *HoxD12* são expressos exclusivamente nos dígitos de 2 a 5, enquanto excluídos do dígito I (o polegar). Em contraste, *HoxD13*, como *HoxA13*, é expresso em todas as células digitais (Montavon et al., 2008).

Os genes *HoxD12* e *HoxD13* do cluster *HoxD* mostram forte expressão nas células posteriores dos brotos dos membros (Papageorgiou, 2017) e estudos comparativos sugerem que a sua expressão (Freitas et al., 2007), sequências codificadoras e até elementos não codificadores (Lee et al., 2006; Lemons, 2006) são altamente conservados nos vertebrados.

Wang et al. (2008) sugerem que *HoxD12* e *HoxD13* sofreram adaptações moleculares nos cetáceos e que essas mudanças correspondem a grandes rearranjos no número de dígitos (Figura 5). O registro de fósseis de cetáceos fornece um dos exemplos mais espetaculares de transição evolutiva de todos os mamíferos, a partir das pequenas formas de tetrápodes semi aquáticas do Eoceno inferior (Thewissen et al., 2007) às grandes espécies totalmente aquáticas do Eoceno tardio, nas quais os membros posteriores foram sendo reduzidos em tamanho até desaparecerem por completo (Thewissen et al., 2006). A evidência fóssil e a reconstrução do estado ancestral

sugerem que os antepassados dos cetáceos possuíam cinco dígitos (Gingerich et al., 2001) e, portanto, podem ter recuperado um dígito após a divergência com o ancestral em comum com *hipopotamidae* (Nikaido et al., 1999), cujos antebraços têm quatro dígitos (Boisserie, 2005). Um cenário alternativo, que parece mais plausível em termos de morfologia, mas que é menos concordante com o resultado dos testes de seleção, é que o número de dígitos nos primeiros cetáceos é um estado primitivo e que as reduções no número de dígitos ocorreram de forma independente em quatro linhagens ancestrais de Artiodactyla. A perda subsequente de um dígito em alguns mysticetos (Cooper et al., 2007) no Oligoceno tardio (Deméré et al., 2008; Fitzgerald, 2006) lhes conferiu uma nadadeira alongada e mais estreita e bem adaptada para natação (Goldbogen, 2006). Os resultados das análises da evolução molecular de *HoxD12* são, portanto, consistentes com explicações adaptativas para transições em números de digitais associadas à mudança do habitat.

Em trabalhos anteriores foi detectado valores ω aumentados de certos genes *Hox*, onde ω se refere a taxa de substituição de bases (dn/ds), em ancestrais das baleias (*HoxD12*), golfinhos (*HoxD9*), baleias (*HoxC4*) e peixe-boi (*HoxB4*). Os autores também observaram que substituições no *HoxD12* estão associadas a diferentes mudanças adaptativas nos membros em diferentes grupos de mamíferos como, por exemplo, em cetáceos, peixe-boi, tatu-bola e na preguiça (Kui et al., 2018).

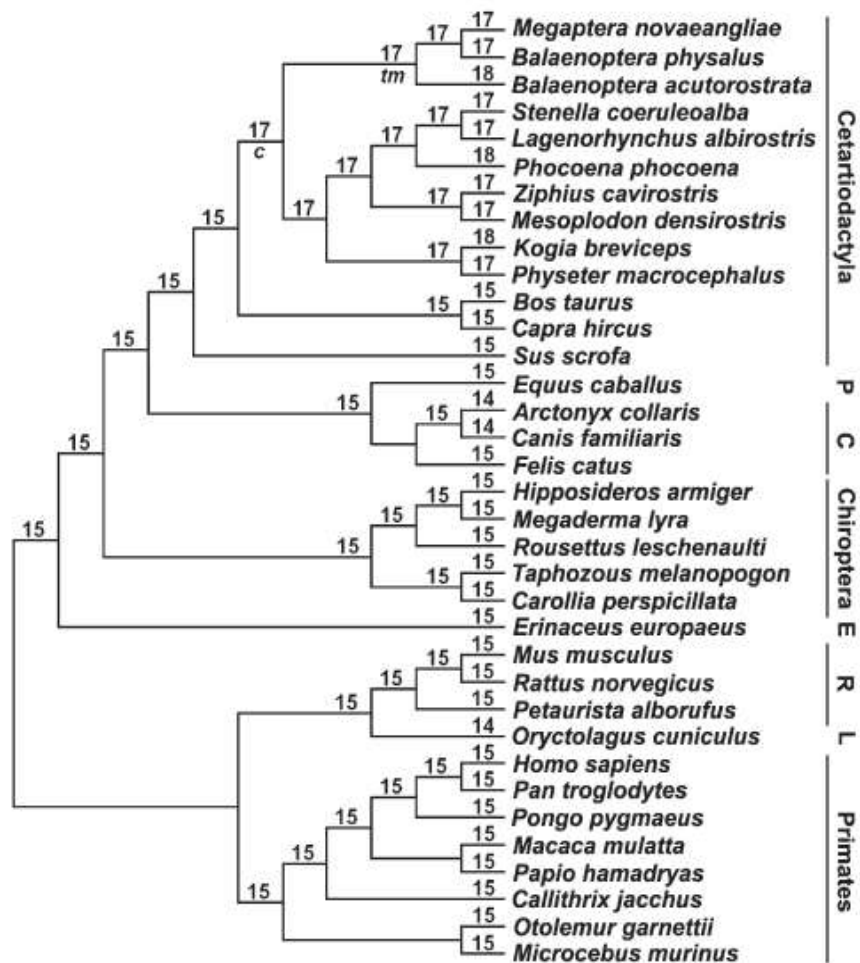


Figura 3. Filogenia de 35 mamíferos utilizados para análise evolutiva de *HoxD12* e *HoxD13* publicada no trabalho de Wang et al. (2008).

Elementos transponíveis e seu papel na evolução.

Elementos transponíveis (TEs do inglês *transposable elements*) são sequências de DNA capazes de mudarem de posição tanto dentro como entre genomas. A descoberta de Barbara McClintock desses "genes saltadores" (McClintock, 1950) lhe rendeu um prêmio Nobel em 1983, a partir de seus experimentos com milho. TEs foram encontrados em praticamente todas as espécies eucarióticas investigadas até o momento. Os TEs afetam o genoma pela sua capacidade de se mover e replicar, gerando assim plasticidade (Wicker et al., 2007).

Os elementos transponíveis têm o potencial de influenciar a trajetória evolutiva de seu hospedeiro de três maneiras distintas: via alterações da função gênica através da inserção; pela

indução de rearranjos cromossômicos; como uma fonte de material codificante e não-codificante que permite o surgimento da novidade genética (como novos genes e sequências regulatórias) (Feschotte and Pritham, 2007).

Várias décadas depois, os TEs adquiriram os rótulos antropomórficos de “egoístas” e “parasitas” por causa de sua autonomia replicativa e potencial para a ruptura genética (Jangam et al., 2017; Smit, 1999). No entanto, os TEs geralmente existem em genomas eucarióticos em uma forma geneticamente indetectável, reversivelmente inativa, que agora chamamos de "epigeneticamente silenciados", cuja descoberta também pode ser atribuída aos estudos genéticos de McClintock (Fedoroff and Bennetzen, 2013). Apesar dessa visão dos TEs como geradores de mutações prejudiciais aos organismos e causadores de doenças, cada vez mais eles tem se mostrado como importantes fontes de novidades evolutivas (Feschotte 2008; Jangam et al., 2017).

Os TEs são geralmente encontrados em regiões não codificadoras. Contudo, podem ser observados juntos a éxons ou mesmo íntrons. Os TEs podem, por exemplo, fornecer promotores alternativos para expressão de genes; quando inseridos em genes, inativá-los; viabilizar variantes de proteínas através do splicing; ter ação epigenética quando modificam a expressão de um gene sem modificar sua sequência; e ocasionar rearranjos cromossômicos (Gray et al., 2000). Nas regiões não codificantes os TEs podem assumir um papel estrutural estando associados ao centrômero e/ou regiões terminais dos cromossomos (Regner et al., 1996; Oliveira et al., 2013; Bertocchi et al., 2018).

TEs são divididos em duas classes principais: transposons de DNA e retrotransposons (Finnegan, 1989). Os transposons de DNA são capazes de se mover e se inserir em novos locais genômicos (Craig et al., 2002). Os retrotransposons replicam-se formando intermediários de RNA, que são então transcritos reversamente para fazer sequências de DNA e inseridos em novas localizações genômicas (Pace and Feschotte, 2007). Os retrotransposons são classificados em dois grupos: retrotransposons com LTR (*long terminal repeat*) e sem LTR. Os retrotransposon sem LTR ainda são divididos em LINES (*long interspersed nuclear element*) e SINES (*short*

interspersed nuclear element).

Elementos transponíveis possuem grande influência na evolução dos organismos, inclusive em mamíferos, por meio de mutações de inserção, influenciando a taxa de recombinação e distribuição de sequências que podem ser usadas para várias finalidades (McDonald, 1993; Robins, e Samuelson, 1993). Retrotransposons têm sido usados como sequências reguladoras de transcrição (Robins e Samuelson, 1993), sinalizador de poliadenilação (Harendza e Johnson, 1990; Goodchild et al., 1992) ou sítios de *splice* de RNA (Varagona et al., 1992). Muito do que se sabe sobre a utilização de sequências derivadas de retrotransposons no genoma dos mamíferos foi deduzido a partir de inserções que ocorreram no passado relativamente recente. Embora seja seguro supor que os retrotransposons tenham desempenhado um papel similar ao longo da evolução (Bejerano et al., 2006; Feschotte e Pritham, 2007; Bourque et al., 2008; Polak e Domany, 2006; Santangelo et al., 2007; Schmitz e Brosius, 2011), mutações nas sequências de retrotransposons ancestrais tornam sua identificação mais difícil.

Evidência de que TEs comumente adquirem função reguladora vem de fragmentos de elementos que são profundamente conservados entre mamíferos. Primeiro esses elementos tendem a se agrupar em torno de genes que estão envolvidos no desenvolvimento e na regulação transcricional (Lowe et al., 2007). Segundo, eles estão super-representados dentro dos módulos de regulação cis (Gentles et al., 2007), isto é, segmentos genômicos contendo redes densas de sítios de ligação de fatores de transcrição (TFBS, do inglês *transcription fator binding site*). Terceiro, aproximadamente um quinto dos elementos conservados não-codificadores específicos de eutérios fornecem sequências promotoras ou sítios de ligação para proteínas reguladoras (Mikkelsen et al., 2007). Finalmente, há um número crescente de casos individuais de TEs altamente conservados documentados para atuar como intensificadores transcricionais (Bejerano et al., 2006; Santangelo et al., 2007) ou como éxons alternativos de *splicing*. Juntos, esses dados sugerem que os TEs têm sido uma fonte abundante de novas sequências regulatórias ao longo da evolução dos mamíferos (Feschotte, 2008).

Donehower et al. (1989) e Smit e Riggs (1995) caracterizaram os elementos MIRs (*Mammalian-wide Interspersed Repeat*), uma classe de SINEs presente em todos os mamíferos. O elemento MIR foi amplificado dentro do genoma dos primeiros mamíferos há 130 milhões de anos (Jurka et al., 1995). Devido à idade, o número de elementos MIR no genoma dos mamíferos é difícil de determinar, mas estima-se que esse elemento represente 2% do genoma humano (Jurka et al., 1995; Smit e Riggs, 1995). O caráter repetitivo de MIR foi observado pela primeira vez em 1987 (Degen e Davie, 1987) e, posteriormente, em outros estudos (Zhou et al., 1988; Jurka et al., 1993). Jurka et al. (1995) mostram que os MIRs são altamente repetitivos nos genomas de marsupiais e monotremados. Elementos em sítios ortólogos em diferentes espécies de mamíferos evidenciam que a distribuição de MIRs ocorreu antes da radiação dos mamíferos (Donehower et al., 1989). O alto número de cópias e a idade dos elementos MIR fazem deles um excelente objeto de estudo para investigar o papel dos elementos transponíveis na evolução do genoma dos mamíferos.

Donehower et al. (1989) descrevem os MIRs como elementos de 70 pb (pares de base) de comprimento sem as características típicas dos SINEs "genéricos". Foi, portanto, sugerido que os MIRs representam uma classe separada de elementos repetitivos. Smit e Riggs (1995) mostraram que este elemento MIR de 70 pb e duas outras repetições intercaladas não caracterizadas anteriormente relatadas são fragmentos de um SINE derivado de um tRNA clássico de 260 pb.

Os MIRs são geralmente truncados em uma ou ambas as extremidades (Smit e Riggs, 1995). A parte 5' tRNA do MIR é fundida a uma sequência não relacionada ao tRNA, e a extremidade 3' apresenta um fragmento de 50 nucleotídeos que é semelhante à extremidade 3' dos LINEs, um provável local de ligação da transcriptase reversa de um LINE, necessário para retrotransposição (Kapitonov et al., 2004). Além disso, possuem um domínio central conservado incluindo uma sequência central de 15 nucleotídeos. No entanto, não se sabe por que a região central é altamente conservada nos chamados CORE-SINEs, como os MIRs. Possivelmente, os CORE-SINEs ainda não têm uma função genômica bem compreendida (Gilbert e Labuda, 1999;

Krull et al., 2007). Os elementos MIR foram ativos à retroposição antes que os monotremados e os mamíferos placentários divergissem, mas essa atividade foi silenciada após a perda da transcriptase reversa específica de LINEs.

A comparação deste elemento MIR entre humanos e cães ou gatos demonstra que o grau de homologia (aproximadamente 70%) é semelhante aos elementos MIR em outros loci ortólogos, o que levou à conclusão de que a maioria dos elementos MIR não são conservados (Armour et al., 1989; Jurka et al., 1995). Entretanto, a importância dos elementos MIR fica evidente em casos como o do gene do receptor de acetilcolina em humanos, que ganhou um éxon adicional (Beeson et al., 1990), usando um novo local de *splicing* que ocorre dentro da região central conservada da sequência consenso de um elemento MIR (Murnane e Morales, 1995). A presença do elemento MIR, como outros elementos transponíveis (Varagona et al., 1992), pode, portanto, influenciar o *splicing* de RNA. Além disso, como resultado do *splicing* alternativo, aproximadamente 26 pb do elemento MIR foi incorporado na região codificadora da proteína. Esta análise mostra que o elemento MIR foi usado para a evolução de proteínas, como mostrado anteriormente para sequências de Alu (Margalit et al., 1994). Em outros casos de envolvimento do elemento MIR em *splicing* que foram observados, sugerem que esta não é uma ocorrência comum e, portanto, não é provável que explique a prevalência da sequência do núcleo central do elemento MIR no genoma dos mamíferos (Murnane e Morales, 1995).

Murnane e Morales (1995) encontraram uma região de homologia entre a extremidade 3' da sequência de consenso MIR (Smit e Riggs, 1995) com as extremidades 3' de quatro sequências de cDNA. A extremidade 3' do elemento MIR contém um sinal de poliadenilação (AATAAA) que é conservado em todos os quatro cDNAs. As sequências de AATAAA nestes cDNAs estão localizadas 13-23 pb da extremidade 3', correspondendo à posição esperada (10-30 pb da extremidade 3') da sequência de poliadenilação para mRNAs (Proudfoot e Brownlee, 1976). A sequência AATAAA do elemento MIR, portanto, tem sido usada como sinal de poliadenilação para esses genes. A presença desta sequência dentro de quatro genes diferentes sugere que seu uso para

poliadenilação não é incomum. O fim do elemento MIR contendo o sinal de poliadenilação mostrou ser mais prevalente no genoma de mamíferos do que outras porções do elemento MIR (Smit e Riggs, 1995). A prevalência desta extremidade do elemento MIR pode indicar que esta região está sendo conservada; no entanto, outra explicação possível é que a parte final do elemento MIR também está presente em outro elemento repetitivo (Smit e Riggs, 1995).

Segundo análises de Murnane e Morales (1995), foi demonstrado que sequências do elemento MIR, como outros retrotransposons ou retrovírus endógenos, foram usadas para funções específicas dentro do genoma de mamíferos, incluindo sequências de codificação. Essa observação é consistente com a hipótese de que, embora os retrotransposons possam se originar como "DNA egoísta", seu uso no genoma dos mamíferos pode ter um papel importante na evolução. Assim, avaliar a presença desses elementos dentro do genoma e se esses elementos podem estar causando mudanças nessas regiões, pode ajudar a entender melhor e reconstruir a história evolutiva do grupo de seus hospedeiros.

Os TEs podem abrigar muitos sítios de ligação de fatores de transcrição (TFBSs, do inglês *transcription factor binding sites*) e a natureza móvel dos TEs, que lhes permite ocupar quase qualquer local de um genoma, os torna uma rota poderosa para a disseminação de sequências cis-regulatórias prontas para uso (Loreto et al., 2018). A adição de novos TFBSs em regiões reguladoras pode criar novos padrões de expressão gênica. Existem exemplos em diversos organismos de genes em que houve exaptação TE-TFBSs; (Chuong et al., 2016). Nos mamíferos, Polavarapu et al. (2008) descobriram que 7-10% dos TFBSs caracterizados experimentalmente no genoma humano são derivados de TEs. A evolução de organismos multicelulares complexos em vários ramos da árvore da vida foi acompanhada, e talvez facilitada, por uma expansão e diversificação dos fatores de transcrição (TFs do inglês *transcription factors*) (Wray et al., 2003; Gagneux, 2002; Chen e Rajewsky, 2007; Levine e Tijan, 2003). Acredita-se que o surgimento de novos TFs permitiu a elaboração de redes cada vez mais complexas de genes ligados por elementos cis, que são reconhecidos por diferentes conjuntos de TFs (Gagneux, 2002; Levine e Tijan, 2003).

Por isso, é relevante investigar a existência de sítios de transcrição dentro de elementos transponíveis.

OBJETIVOS

Objetivos Gerais

Contribuir para o melhor entendimento da história evolutiva em Cetartiodactyla, avaliando mudanças em uma região no cluster *HoxD* em relação à evolução dos membros desse grupo, assim auxiliando na melhor compreensão dos processos evolutivos e como alterações em genes do desenvolvimento e elementos transponíveis podem influenciar histórias macroevolutivas.

Objetivos Específicos

Analisar as relações entre o gene *HoxD12* na evolução do grupo Cetartiodactyla, incluindo espécies não analisadas em estudos anteriores, como *Hippopotamus sp*, espécies que ocorrem na América do Sul e outros mamíferos também associados com ambientes aquáticos. Avaliar se elementos transponíveis podem ter influenciado essa história evolutiva dentro da região em enfoque, na evolução dos membros desse grupo.

MATERIAIS E MÉTODOS

Amostragem, extração de DNA, amplificação e sequenciamento

As amostras de tecido utilizadas no trabalho foram obtidas de grupos colaboradores: Laboratório ECOMEGA da FURG (Universidade Federal do Rio Grande), no qual foram enviadas amostras coletadas durante saídas de campo do Projeto Talude; além de amostras da UFMG, CECLIMAR e do Museu de Ciências Naturais e Zoológico de Sapucaia, pertencentes à Fundação Zoobotânica do Rio Grande do Sul. Foram feitas extrações de DNA a partir dos tecidos com protocolo de CTAB (Doyle e Doyle, 1987) e no caso do hipopótamo foi utilizado Kit de extração Purelink Genomic DNA Mini Kit da Invitrogen (Carlsbad, EUA) e utilizado para amplificar o gene *HoxD12* via PCR. Foi desenhado um par de primers (forward: 5' –ATG TGT GAG CGC AGT CTC TAC AG- 3'; reverse: 5' –ATA CGC CGG TTC TGG AAC CA- 3'), amplificando aproximadamente 721 pb (89%) do gene *HoxD12*. A reação de amplificação foi feita em um volume de 15 µL contendo 1,5 µL de tampão de PCR 10X, 1 µL de MgCl₂ (25mM), 0,3µL de cada primer (20mM), 1 µL de DNTPs (10ng/ µL), 0,3 µL de Taq (5U/µL) e 1 µL de DNA (com uma concentração mínima de 50ng/ µL). As condições de amplificação foram as seguintes: 3 minutos de desnaturação inicial a 95°C, seguido de 35 ciclos de desnaturação a 94°C por 45 segundos, anelamento a 50°C por 45 segundos, extensão a 72°C por 1 minuto e extensão final a 72°C por 7 minutos. O sequenciamento foi feito em ambas as direções, sendo e anti senso pela empresa Macrogen (Seoul, South Korea).

Os eletroferogramas foram analisados e as sequências consensos foram geradas pelo pacote Staden versão 2.0 (Staden, 1996). As 9 novas sequências, juntamente com 97 sequências do gene *HoxD12* de mamíferos obtidas no GenBank (Tabela 1), foram alinhadas utilizando o algoritmo MUSCLE (Multiple Sequence Comparison by Log-Expectation), implementado no MEGA 7.0 (Kumar et al., 2016). Foram analisados 612 nucleotídeos do gene *HoxD12* da região codificante (aproximadamente 75,5% da região codificante total), o íntron existente entre os dois éxons do

gene foi removido. Todas as ordens dentro das linhagens de mamíferos Euarchontaglires e Laurasiatheria foram representadas, contudo Xenarthra e Afrotheria não possuem representantes de todas as ordens nessa análise.

Tabela 1. Lista de espécies utilizadas nas análises do gene HoxD12, com seu número de acesso do GenBank e novas sequências prospectadas neste trabalho, destacadas com *.

Táxon	Nome popular	Número de acesso GenBank
SUBCLASSE PROTOTHERIA		
Ordem Monotremata		
<i>Ornithorhynchus anatinus</i>	Ornitorrinco	XM_001511444.1
SUBCLASSE THERIA		
INFRACLASSE MARSUPIALIA		
Ordem Didelphimorphia		
<i>Monodelphis domestica</i>	Catita	XM_001368427.3
INFRACLASSE PLACENTALIA		
<u>SUPERORDEM AFROTHERIA</u>		
Ordem Afrosoricida		
<i>Echinops telfairi</i>	Tenreque	XM_004701443.1
Ordem Sirenia		
<i>Trichechus manatus latirostris</i>	Peixe-boi-marinho	XM_004375484.1
Ordem Tubulidentata		
<i>Orycteropus afer afer</i>	Porco formigueiro	XM_007939300.1
<u>SUPERORDEM EUARCHONTOGLIRES</u>		
Ordem Dermoptera		
<i>Galeopterus variegatus</i>	Galeopiteco	XM_008573600.1
Ordem Lagomorpha		
<i>Ochotona princeps</i>	Pika americana	XM_004577035.1
<i>Oryctolagus cuniculus</i>	Coelho europeu	XM_002712302.2
Ordem Primates		
<i>Aotus nancymae</i>	Macaco-da-noite	XM_012471907.1
<i>Carlito syrichta</i>	Társio das filipinas	XM_008061444.1
<i>Cercocebus atys</i>	Macaco do Velho Mundo	XM_012047069.1
<i>Colobus angolensis palliatus</i>	Macaco Colobo	XM_011944012.1
<i>Homo sapiens</i>	Humano	NM_021193.3
<i>Macaca mulatta</i>	Macaco Rhesus	NM_001194772.1
<i>Mandrillus leucophaeus</i>	Dril (babuíno)	XM_011991856.1
<i>Microcebus murinus</i>	Lêmure	XM_012753367.1
<i>Nomascus leucogenys</i>	Gibão-de-bochechas-brancas-do-norte	XM_003253751.3
<i>Pan paniscus</i>	Bonobo	XM_003815542.2
<i>Pan troglodytes</i>	Chimpanzé	XM_515926.5
<i>Pongo abelii</i>	Orangotango-de-sumatra	XM_002812607.2
<i>Propithecus coquereli</i>	Sifaka-de-coquerel	XM_012650742.1
<i>Rhinopithecus bieti</i>	Macaco-preto-de-nariz-arrebitado	XM_017868313.1
<i>Saimiri boliviensis boliviensis</i>	Macaco-de-cheiro	XM_003921842.2

Ordem Rodentia		
<i>Castor canadensis</i>	Castor-americano	XM_020184341.1
<i>Cavia porcellus</i>	Porquinho-da-índia	XM_003478568.2
<i>Chinchilla lanigera</i>	Chinchila comum	XM_005373416.1
<i>Fukomys damarensis</i>	Rato-toupeira de Damaralândia	XM_010607050.1
<i>Heterocephalus glaber</i>	Rato-toupeira-pelado	XM_004857719.2
<i>Ictidomys tridecemlineatus</i>	Esquilo terrestre de treze listras	XM_005324628.2
<i>Jaculus jaculus</i>	Jerboa egípcio	XM_004660316.1
<i>Marmota marmota marmota</i>	Marmota	XM_015478991.1
<i>Meriones unguiculatus</i>	Esquilo-da-mongólia	XM_021650438.1
<i>Octodon degus</i>	Roedor	XM_004634733.1
<i>Oligoryzomys stramineus*</i>	Rato do mato	Presente estudo
<i>Petaurista alborufus</i>	Esquilo-voador-gigante-vermelho-e-branco	FJ455474.1
<i>Urocitellus parryii</i>	Esquilo do ártico	XM_026405656.1
<i>Wiedomys purrehorhinus*</i>	Rato de fava	Presente estudo
Ordem Scandentia		
<i>Tupaia chinensis</i>	Tupaia	XM_006151106.1
<u>SUPERORDEM</u> <u>LAURASIATHERIA</u>		
Ordem Carnivora		
<i>Ailuropoda melanoleuca</i>	Panda gigante	XM_011236972.2
<i>Arctonyx collaris</i>	Texugo	FJ455467.1
<i>Canis lupus</i>	Lobo	XM_545535.2
<i>Enhydra lutris kenyonii</i>	Lontra-marinha	XM_022501107.1
<i>Eumetopias jubatus</i>	Leão-marinho-de-Steller	XM_028121416.1
<i>Felis catus</i>	Gato	FJ455468.1
<i>Leptonychotes weddellii</i>	Foca de wedellii	XM_006739536.1
<i>Mustela putorius furo</i>	Furão	XM_004774396.1
<i>Neomonachus schauinslandi</i>	Foca-monge-do-havaí	XM_021703217.1
<i>Odobenus rosmarus divergens</i>	Morsa	XM_004409010.1
<i>Panthera pardus</i>	Leopardo	XM_019460347.1
<i>Panthera tigris altaica</i>	Tigre siberiano	XM_007074656.1
<i>Ursus arctos horribilis</i>	Urso-cinzento	XM_026490561.1
<i>Ursus maritimus</i>	Urso polar	XM_008687516.1
<i>Vulpes vulpes</i>	Raposa Vermelha	XM_025994257.1
Ordem Cetartiodactyla		
Subordem Ancodonta		
<i>Hippopotamus amphibious*</i>	Hipopótamo	Presente estudo
Subordem Cetacea		
<i>Balaenoptera acutorostrata</i>	Baleia Minke	FJ455457.1
<i>Balaenoptera acutorostrata scammoni</i>	Baleia-minke-antártica-anã	XM_007190425.1
<i>Balaenoptera physalus</i>	Baleia Fin	FJ455456.1
<i>Delphinapterus leucas</i>	Beluga	XM_022570198.1
<i>Delphinus delphis*</i>	Golfinho comum	Presente estudo
<i>Kogia breviceps</i>	Cachalote pigmeu	FJ455458.1
<i>Lagenorhynchus albirostris</i>	Golfinho-de-bico-branco	FJ455464.1
<i>Lipotes vexillifer</i>	Golfinho lacustre chinês	XM_007451084.1
<i>Megaptera novaeangliae</i>	Baleia jubarte	FJ455455.1
<i>Mesoplodon densirostris</i>	Baleia bicuda de Blainville	FJ455461.1
<i>Monodon moncerus</i>	Narval	XM_029225730.1
<i>Neophocaena asiaeorientalis</i>	Boto do Índico	XM_024732888.1
<i>Orcinus orca</i>	Orca	XM_004267343.1

<i>Phocoena phocoena</i>	Toninha comum	FJ455462.1
<i>Physeter catodon</i>	Cachalote	FJ455459.1
<i>Stenella coeruleoalba</i>	Golfinho listrado	FJ455463.1
<i>Sotalia guianensis*</i>	Boto-cinza	Presente estudo
<i>Tursiops truncatus</i>	Golfinho nariz de garrafa	XM_004314724.2
<i>Ziphius cavirostris</i>	Baleia bicuda de Cuvier	FJ455460.1
<i>Subordem Ruminantia</i>		
<i>Bison bison bison</i>	Bisão	XM_010859792.1
<i>Bos mutus</i>	Iaque	XM_005910403.1
<i>Bos taurus</i>	Boi	NM_001205871.1
<i>Bubalus bubalis</i>	Búfalo-asiático	XM_006066094.1
<i>Capra hircus</i>	Cabra	FJ455465.1
<i>Mazama nana*</i>	Veado-bororó	Presente estudo
<i>Mazama guazoubira*</i>	Veado-catingueiro	Presente estudo
<i>Odocoileus virginianus texanus</i>	Veado da Virgínia	XM_020886146.1
<i>Ovis aries</i>	Ovelha	XM_012143651.2
<i>Pantholops hodgsonii</i>	Chiru	XM_005980837.1
<i>Subordem Suina</i>		
<i>Sus scrofa</i>	Javali	FJ455466.1
<i>Subordem Tylopoda</i>		
<i>Camelus bactrianus</i>	Camelo-bactriano	XM_010960720.1
<i>Camelus dromedarius</i>	Dromedário	XM_010992003.1
<i>Camelus ferus</i>	Camelo	XM_006189921.1
<i>Vicugna pacos</i>	Alpaca	XM_006204890.1
Ordem Chiroptera		
<i>Eptesicus fuscus</i>	Morcego marrom gigante	XM_008138531.1
<i>Hipposideros armiger</i>	Morcego	FJ455470.1
<i>Megaderma lyra</i>	Morcego	FJ455471.1
<i>Pteropus alecto</i>	Raposa voadora	XM_006921226.1
<i>Pteropus vampyrus</i>	Raposa gigante voadora	XM_011357157.1
<i>Rhinolophus sinicus</i>	Morcego	XM_019739283.1
<i>Rousettus aegyptiacus</i>	Morcego egípcio	XM_016135242.1
<i>Rousettus leschenaultii</i>	Morcego	FJ455469.1
<i>Taphozous melanopogon</i>	Morcego	FJ455472.1
Ordem Eulipotyphla		
<i>Condilura cristata</i>	Toupeira-nariz-de-estrela	XM_004674503.1
<i>Erinaceus europaeus</i>	Ouriço terrestre	XM_007524686.1
<i>Sorex araneus</i>	Musaranho comum	XM_004601177.1
Ordem Perissodactyla		
<i>Ceratotherium simum simum</i>	Rinoceronte branco	XM_004426653.2
<i>Equus asinus</i>	Burro	XM_014834036.1
<i>Equus caballus</i>	Cavalo	XM_001500102.5
<i>Equus przewalskii</i>	Cavalo-de-przewalski	XM_008533049.1
Ordem Pholidota		
<i>Manis javanica</i>	Pangolim malaio	XM_017675311.1
<u>SUPERORDEM XENARTHRA</u>		
Ordem Cingulata		
<i>Dasyus novemcinctus*</i>	Tatu galinha	Presente estudo

Análises evolutivas do *HoxD12*

Os padrões de substituição nucleotídica foram avaliados utilizando-se o modelo sugerido pelo AIC (Akaike Information Criterion) obtidos no MrModelTest (Nylander, 2004), que foi o modelo GTR+G+I ($G = 1.1586$; $I = 0.3032$). A árvore filogenética foi construída pelo método de Inferência Bayesiana, executado no Mr.Bayes 3.1.2. (Ronquist e Huelsenbeck, 2003).

Foram executadas ao menos 2.000.000 de gerações de cadeias de Markov (MCMC), amostrando árvores a cada 100 gerações e usando quatro cadeias (3 quentes e uma fria), seguindo o padrão. As corridas foram interrompidas quando os valores de convergência diminuíram abaixo de 0,01 e foi aplicado um *burn-in* de 25% das gerações antes de sumarizar as árvores e os parâmetros. Os valores de suporte nos nós refletem as probabilidades posteriores de cada clado. A árvore foi enraizada utilizando como grupo externo o monotremado *Ornithorhynchus anatinus*.

A árvore consenso gerada foi visualizada e editada no Interactive Tree of Life (iTOLv.4) (Letunic e Bork, 2019).

Para investigar uma possível pressão seletiva que tenha moldado a evolução do gene *HoxD12* em cetáceos, foi realizado um teste *branch-site* (Zhang et al., 2005) usando o software CODEML (Yang, 2007) do pacote PAML. Esse pacote testa seleção positiva por comparação de uma série de hipóteses alternativas que diferem em como a variação da proporção de dN/dS (dS significa a taxa de substituição sinônima e dN é a taxa de substituição não sinônima) pode mudar em diferentes ramos e códons, quando $dN/dS > 1$ indica seleção positiva e quando $dN/dS < 1$ indica seleção purificadora, devido a restrição seletiva ao nível de códon.

Com esses testes, as árvores filogenéticas foram separadas em ramos *foreground*, nos quais foi testado a seleção positiva, e ramos *background*, representado pelas outras linhagens. Nos testes foram usados dois modelos: MA (modelo alternativo) nos quais os diferentes ramos são permitidos ter diferentes valores de ω ; e o M0 (modelo nulo), que permite apenas um valor de ω para todos os ramos da árvore filogenética. Os ramos que apresentaram ramos com valores de distância igual a zero foram removidos das análises. Os testes MA e M0 foram comparados e os sítios candidatos

com seleção positiva foram identificados usando o *Bayes Empirical Bayes* (BEB). Foi usado o *log likelihood ratio test* (LRT) para inferir seleção positiva quando esses valores resultam em valores significantes. A significância do teste LRT foi verificada por uma distribuição qui-quadrado, para valores < 0.05 . As análises de taxas evolutivas por sítio do gene *HoxD12* foram conduzidas no programa MEGA 7.0, utilizando o modelo evolutivo para aminoácidos JTT + G + I.

Identificação de elementos transponíveis e sítios de ligação de fatores de transcrição em região do cluster *HoxD*

Foram obtidas as sequências entre o *HoxD10* e *Evx2* (*HoxD10*, *HoxD11*, *HoxD12*, *HoxD13* e *Evx2*) incluindo as regiões intergênicas de 32 espécies com genomas depositados no *genbank* (Tabela 2). Foram selecionados 32 genomas de maneira que representasse todas as ordens de mamíferos disponíveis no banco de dados e uma espécie da Classe Aves, *Gallus gallus*, como grupo externo.

Para identificação dos TE foi utilizado o *software* RepeatMasker disponível *online* através do link (<http://www.repeatmasker.org/>) e foram utilizados os parâmetros padrão (Smit et al., 2015). Os TEs MIRs analisados foram obtidos no banco de dados Repbase e alinhados juntamente com as sequências gênicas através da plataforma MAFFT, parâmetros *default*, e refinados manualmente (Bao et al., 2015) A estimativa da conservação nucleotídica das cópias identificadas de MIR foram analisadas através de matriz de similaridade utilizando os parâmetros *default* do software UGENE (Katoh e Standley, 2013).

Por fim, a predição de possíveis sítios de fator de transcrição nas regiões de inserção dos elementos MIR foi estimada através da plataforma Alibaba 2.1 (<http://gene-regulation.com/pub/programs/alibaba2/index.html>), utilizando os seguintes parâmetros: 64 sítios conhecidos por parcimônia, com o número mínimo de 5 sítios, 80% conservação mínima na sequência e nível 4 da classe do fator (RAR), os outros parâmetros foram utilizados os valores *default* (Tabela 3).

Tabela 2. Lista de espécies utilizadas nas análises de elementos transponíveis, com número de acesso, elementos e fatores de transcrições que foram preditos por nossas análises.

Táxon	Nome popular	Número de acesso	Família do TE	TE	Início (pb)	Fim (pb)
CLASSE AVE						
<i>Gallus gallus</i>	Galo	NC_006094.5				
INFRACLASSE MARSUPIALIA						
ORDEM DIDELPHIMORPHIA						
<i>Monodelphis domestica</i>	Catita	NC_008804.1	LINE/RTE-BovB LTR/ERV1 LTR/ERVK LTR/ERVK SINE/tRNA SINE/MIR DNA/hAT-Charlie	BovB_Me LTR18F_ML RLTR31D_MM RLTR17B_Mm Ves2_ML MIR_Mars OposCharlie1	1214 2567 2626 6603 2702 11618 42828	1272 2696 2701 6676 2717 11687 42873
INFRACLASSE PLACENTALIA						
ORDEM CINGULATA						
<i>Dasyopus novemcinctus</i>	Tatu-galinha	NW_004491115	SINE/MIR	MIR	6743	6894
ORDEM PROBOSCIDAEE						
<i>Loxodonta africana</i>	Elefante-africano	NW_003573423	SINE/MIR SINE/MIR SINE/MIR LINE/CR1	MIR MIR MIRc MIRb L3	6743 6946 27303 27565 35803	6894 7134 27522 27600 35857
ORDEM CHIROPTERA						
<i>Eptesicus fuscus</i>	Morcego grande marrom	NW_007370658	SINE/MIR LINE/L1 LINE/L1	THER1_MD L1-4a_EF L1-2a_EF	6935 32219 32236	7101 32327 32349
ORDEM PERISSODACTYLA						
<i>Equus caballus</i>	Cavalo	NC_009161.3	LTR/ERV1 SINE/MIR	MdoERV2a-int MIRb	1563 10111	1661 10271
ORDEM CARNIVORA						
<i>Acinonyx jubatus</i>	Guepardo	NW_020834737.1	LINE/L1 SINE/MIR SINE/tRNA	Lx10 MIRb SINEC_old	1207 9762 29254	1277 9922 29298
<i>Ursus maritimus</i>	Urso polar	NW_020834737.1	LTR/ERVK SINE/MIR	RLTR17B_Mm MIR_Mars	1314 9155	1445 9314
<i>Leptonychotes weddellii</i>	Foca-de-weddell	NW_006384104	SINE/MIR LINE/CR1	MIR_Mars L3_Mars	6857 8677	7017 8775
<i>Callorhinus ursinus</i>	Lobo-marinho-do-norte	NW_020312786	SINE/MIR	MIR_Mars	6928	7088
<i>Zalophus californianus</i>	Lobo-marinho-da-califórnia	NW_020884869	SINE/MIR	MIR_Mars	7164	7324
<i>Odobenus rosmarus divergens</i>	Morsa	NW_004450672.1	SINE/MIR	MIR_Mars	6997	7157
<i>Mustela putorius furo</i>	Furão	NW_004569339.1	LTR/ERVK	RLTR17B_Mm	73	182

			SINE/MIR snRNA SINE/MIR	MIR_Mars U4 MIRb	8390 8491 8558	8489 8556 8624
ORDEM CETARTIODACTYLA						
<i>Capra hircus</i>	Cabra	NC_032651.1	SINE/MIR	MIR	6816	6970
<i>Bos indicus</i>	Zebu	NC_032651.1	LINE/RTE-BovB SINE/MIR SINE/tRNA	BovB_Oa MIR TUBE2	385 6984 13116	738 7048 13242
<i>Bos mutus</i>	Iaaque doméstico	NW_005396625.1	SINE/MIR SINE/tRNA	MIR TUBE2	6645 12332	6799 12457
<i>Bos taurus</i>	Gado	NC_037329.1	SINE/MIR	MIR	6791	6945
<i>Bubalus bubalis</i>	Búfalo-asiático	NC_037546.1	SINE/MIR	MIR	6967	7121
<i>Pantholops hodgsonii</i>	Chiru	NW_005816213.1	SINE/MIR	MIR	6334	6488
<i>Odocoileus virginianus texanus</i>	Veado-da-virgínia	NW_018343797.1	SINE/MIR	MIR	6752	6906
<i>Sus scrofa</i>	Javali	NC_010457.5	SINE/MIR SINE/tRNA SINE/tRNA	MIR_Mars Pre0_SS Ves2_ML	6828 7055 38587	6993 7297 38686
<i>Camelus ferus</i>	Camelo	NW_006211459.1	DNA/hAT- Tip100 SINE/MIR	MER45A MIR	1858 13312	1956 13474
<i>Physeter catodon</i>	Cachalote	NC_041215.1	SINE/MIR	MIR_Mars	6973	7133
<i>Delphinapterus leucas</i>	Baleia branca	NW_019160862.1	SINE/MIR	MIR_Mars	6976	7137
<i>Neophocaena asiaeorientalis</i>	Boto-do-índico	NW_020175831.1	SINE/MIR	MIR_Mars	6396	6557
<i>Lagenorhynchus obliquidens</i>	Golfinho-de-laterais-brancas- do-pacífico	NW_020837946.1	LTR/ERV1 SINE/MIR	LTR39B2_ML MIR_Mars	5258 7445	5326 7605
<i>Orcinus orca</i>	Orca	NW_004438435.1	SINE/MIR	MIR_Mars	6961	7121
<i>Lipotes vexillifer</i>	Golfinho-lacustre-chinês	NW_006777439.1	SINE/MIR	MIR_Mars	6898	7058
<i>Tursiops truncatus</i>	Golfinho-nariz-de-garrafa	NW_017844625.1	SINE/MIR	MIR_Mars	5549	5709
<i>Balaenoptera acutorostrata</i>	Baleia-anã	NW_006730567.1	SINE/MIR	MIR_Mars	6699	6859
ORDEM RODENTIA						
<i>Mus musculus</i>	Camundongo	NC_000068.7	LTR/ERVK DNA/hAT- Charlie LTR/ERV1 SINE/Alu	RLTR17B_Mm hAT-2_TS MdoERV2a-int PB1D7	1390 1415 4764 12786	1474 1484 4866 12881
ORDEM PRIMATA						
<i>Pan troglodytes</i>	Chimpanzé-comum	NC_036881.1	LTR/ERVK LTR/ERV1 SINE/tRNA LTR/ERVK SINE/MIR SINE/Alu	RLTR17B_Mm LTR44B_MD Ves2_ML RNLTR15A2 MIRb AluSg	402 6819 6880 6914 9073 9431	566 6931 6962 6982 9222 9720

Tabela 3. Lista de sítios de fatores de transcrição detectados em todos os elementos MIRs identificados por esse trabalho.

Taxon	TE	Fator de transcrição
INFRACLASSE MARSUPIALIA		
ORDEM DIDELPHIMORPHIA		
<i>Monodelphis domestica</i>	MIR_Mars	Oct-1

INFRACLASSE PLACENTALIA		
ORDEM PROBOSCIDAE		
<i>Loxodonta africana</i>	MIR MIRc	Oct-1 Sp1 Sp1 CPC1 AP-1
ORDEM CHIROPTERA		
<i>Eptesicus fuscus</i>	THER1_MD	Oct-1 TBP
ORDEM PERISSODACTYLA		
<i>Equus caballus</i>	MIRb	Oct-1 Sp1 Odd
ORDEM CARNIVORA		
<i>Acinonyx jubatus</i>	MIRb	Oct
<i>Ursus maritimus</i>	MIR_Mars	Oct-1 Sp1 Ttk-88k
<i>Leptonychotes weddellii</i>	MIR_Mars	Oct-1 Ttk-88k
<i>Callorhinus ursinus</i>	MIR_Mars	Oct-1
<i>Zalophus californianus</i>	MIR_Mars	Oct-1
<i>Odobenus rosmarus divergens</i>	MIR_Mars	Oct-1
<i>Mustela putorius furo</i>	MIR_Mars MIRb	Oct-1 Ttk-88k
ORDEM CETARTIODACTYLA		
<i>Bos mutus</i>	MIR	Oct-1
<i>Bos taurus</i>	MIR	Oct-1
<i>Bubalus bubalis</i>	MIR	Oct-1
<i>Pantholops hodgsonii</i>	MIR	Oct-1
<i>Odocoileus virginianus texanus</i>	MIR	Oct-1
<i>Sus scrofa</i>	MIR_Mars	Oct-1
<i>Camelus ferus</i>	MIR	Oct-1 C/EBPalfa
<i>Physeter catodon</i>	MIR_Mars	Oct-1 NF1 GATA-3 HSE-bind
<i>Delphinapterus leucas</i>	MIR_Mars	Oct-1 NF1 GATA-3
<i>Neophocaena asiaeorientalis</i>	MIR_Mars	Oct-1 NF1 GATA-3
<i>Lagenorhynchus obliquidens</i>	MIR_Mars	Oct-1 NF1 GATA-3
<i>Orcinus orca</i>	MIR_Mars	Oct-1 NF1 GATA-3
<i>Lipotes vexillifer</i>	MIR_Mars	Oct-1 NF1 GATA-3
<i>Tursiops truncatus</i>	MIR_Mars	Oct-1 NF1

		GATA-3
<i>Balaenoptera acutorostrata</i>	MIR_Mars	NF1 GATA-3
ORDEM PRIMATA		
<i>Pan troglodytes</i>	MIRb	NF1 Zta

RESULTADOS

Árvore por inferência Bayesiana

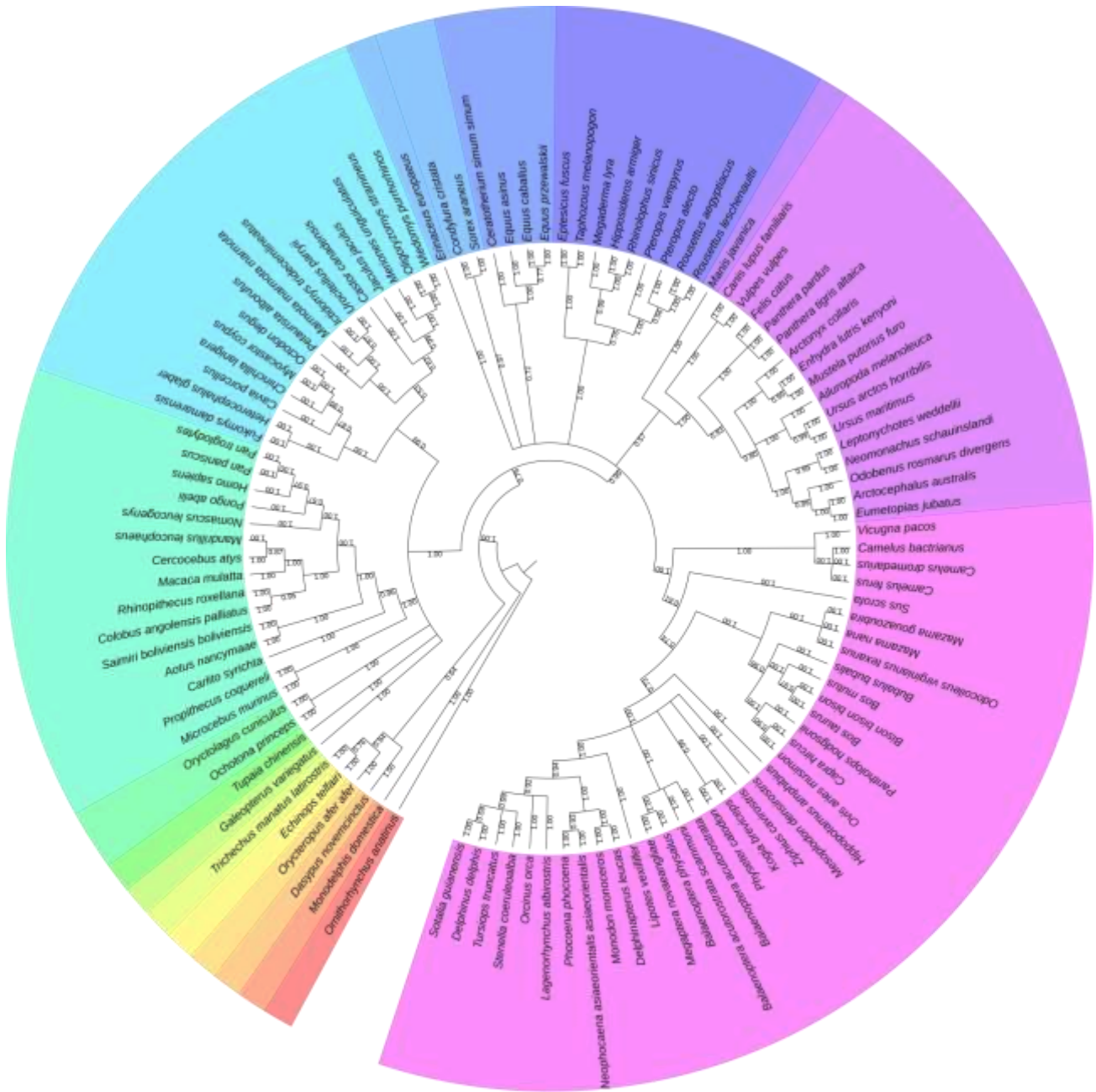
A árvore consenso gerada a partir das sequências parciais do gene *HoxD12* (Figura 6) apresentou altos valores de suporte na maioria dos ramos e recuperou a monofilia dos placentários. Os monotremados, aqui representado pelo ornitorrinco, foram os primeiros a divergir, seguidos por Marsupialia, representada pelo *Monodelphis domestica*.

Dentro de placentários, o ramo Afrotheria (Tubulidentata + Afrosorida + Sirenia) + Xenarthra (com um representante da ordem Cingulata) apresentou um valor PP=0,64, sendo Xenarthra mais basal em relação à Afrotheria. Euarchontoglires (Primata + Roedores + Scadentia + Dermoptera) ficou bem estabelecido (PP=1), assim como Laurasiatheria (Carnivora, Cetartiodactyla, Chiroptera, Eulipotyphla, Perissodactyla e Pholidota) (PP=0,99). Todas as ordens são recuperadas como clados monofiléticos, com exceção de Eulipotyphla. No entanto, as relações entre as ordens, tanto dentro da superordem Euchantoglires quanto Laurasiatheria apresentam politomia.

As relações entre as espécies dentro das ordens não são bem estabelecidas, o que pode ser observado pela quantidade de politomias encontradas dentro das ordens. A exceção a esse fato são as ordens Carnivora, Lagomorpha, Perissodactyla, Chiroptera e Primata, cujas relações entre espécies não apresentam politomias.

A ordem Cetartiodactyla é recuperada como um grupo monofilético (PP=1) e as relações entre as principais linhagens (Tylopoda (Suoidea, (Ruminantia (Cetancodonta))) são bem

suportadas, ainda que existam politomias dentro das linhagens Cetacodonta e Ruminantia). Nossos resultados também apoiam a monofilia de Cetancodonta (Cetacea + Ancodonta) e Cetruminantia (Cetancodonta + Ruminantia), o que está de acordo com outras evidências moleculares, morfológicas e paleontológicas (Zhou et al., 2011; Hassanin et al., 2012; Vislobokova, 2013).



Order	Color
Monotremata	Red
Didelphimorpha	Orange
Cingulata	Yellow
Tubulidentata	Light Green
Afrosoricida	Yellow-Green
Sirenia	Green
Dermoptera	Light Blue
Scandentia	Medium Green
Lagomorpha	Teal
Primates	Light Cyan
Rodentia	Cyan
Eulipotyphla	Blue
Perissodactyla	Dark Blue
Chiroptera	Purple
Pholidota	Magenta
Carnivora	Pink
Cetartiodactyla	Light Pink

Figura 4. Árvore nucleotídica por inferência Bayesiana gerada a partir de sequências do gene *HoxD12* em mamíferos, os valores de probabilidade posterior (PP) estão representados nos ramos. Na imagem há representantes de todas as ordens dentro das linhagens de mamíferos Euarchontaglires e Laurasiatheria, somente as linhagens Xenarthra e Afrotheria não possuem representantes de todas as ordens nessa análise (Foley et al., 2016).

Análises evolutivas do *HoxD12*

O pacote MEGA 7.0 foi utilizado para computar a taxa evolutiva média (relativa) das sequências de *HoxD12* (Figura 7). As taxas evolutivas são escalonadas de tal forma que a taxa média evolutiva em todos os locais é igual a 1. Isso significa que sítios com taxa < 1 estão evoluindo mais lentamente que a média, e aqueles com taxa > 1 estão evoluindo mais rapidamente que a média.

De maneira geral, o gene *HoxD12* está sob restrições evolutivas na maioria de seus sítios. Quando são levadas em conta todas as sequências de mamíferos aqui analisadas, encontram-se 87 sítios evoluindo mais rapidamente do que a média dos demais sítios (valores > 1) (Figura 7 A). Uma análise detida ao ramo Cetartiodactyla, verifica-se que Artiodactyla apresenta 36 sítios com valores de taxa evolutiva média > 1 (Figura 7 B), enquanto cetácea apresenta 26 sítios evoluindo mais rapidamente que a média. Apesar de Artiodactyla apresentar mais sítios sob rápida evolução, os valores não estão muito acima de 1 (média igual à 1.03), enquanto que em cetácea a média dos 26 sítios é de 1.56 (Figura 7C).

Foram empregados dois testes através do programa CODEML. O primeiro onde o ramo de *Hippopotamus amphibius* seria o *foreground*, a fim de averiguar possível sinal de seleção na respectiva linhagem. Como resultado, apenas um sítio apresentou sinal de seleção positiva (Ala-146), porém o valor LTR não foi significativo. Quando o ramo de divergência entre as linhagens de Artiodactyla e Cetaceae foi utilizado como *foreground*, obteve-se dois sítios com sinal de seleção

positiva (Asp-151 e Arg-171). Entretanto, novamente o valor de LTR não foi significativo.

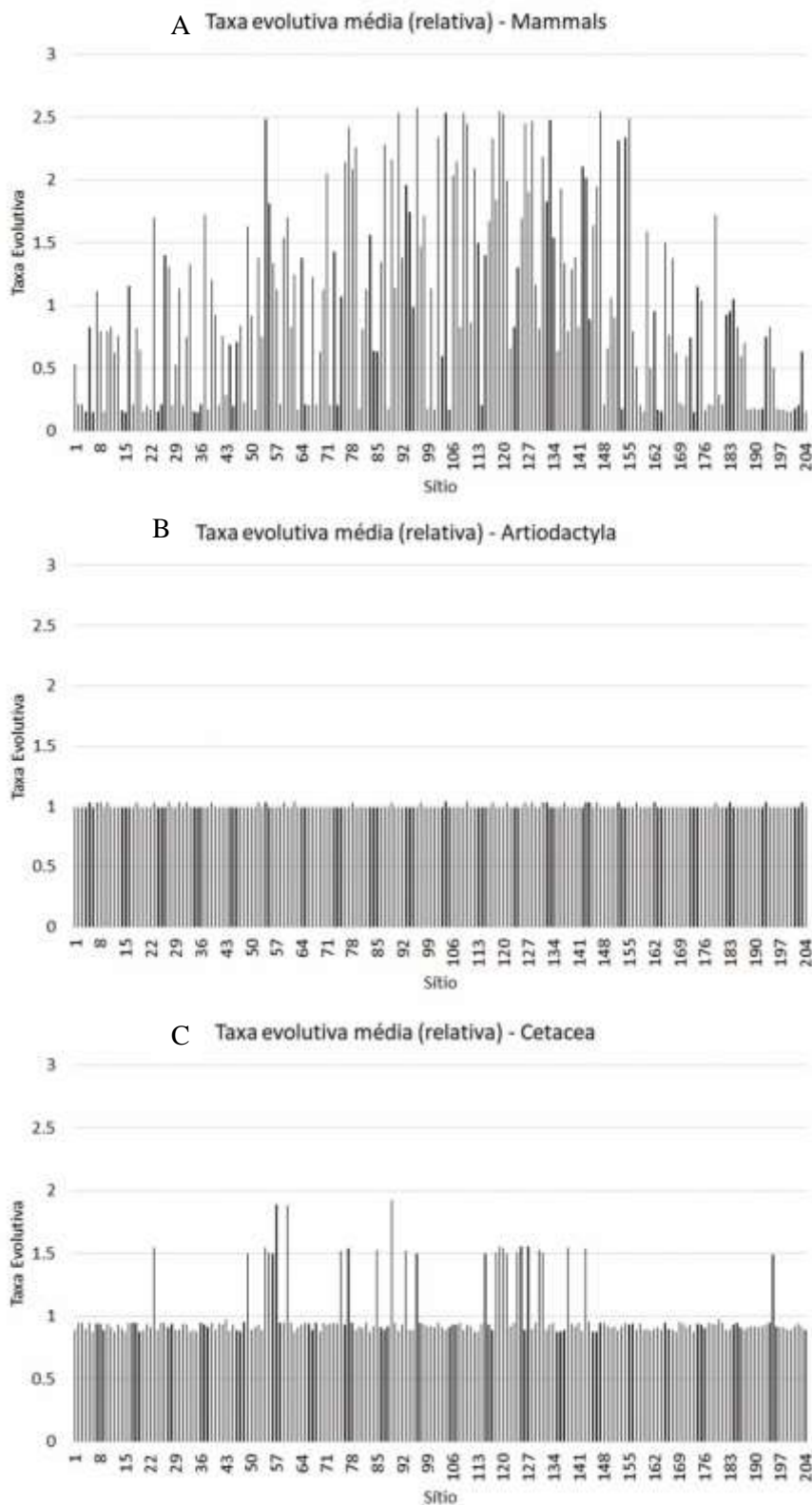


Figura 7. Taxas evolutivas das seqüências do gene HoxD12 de todos os mamíferos analisados no presente trabalho (A), de Artiodactyla (B) e Cetacea (C) separadamente. Todas as posições

contendo lacunas e dados em falta foram eliminados.

Elementos transponíveis e sítios de ligação de fatores de transcrição dentro de região do cluster *HoxD*

Os genomas selecionados representam as ordens da Classe Mammalia disponíveis no *Genbank*, porém da ordem Cetartiodactyla e mamíferos aquáticos foram selecionadas todas as espécies com sequências disponíveis.

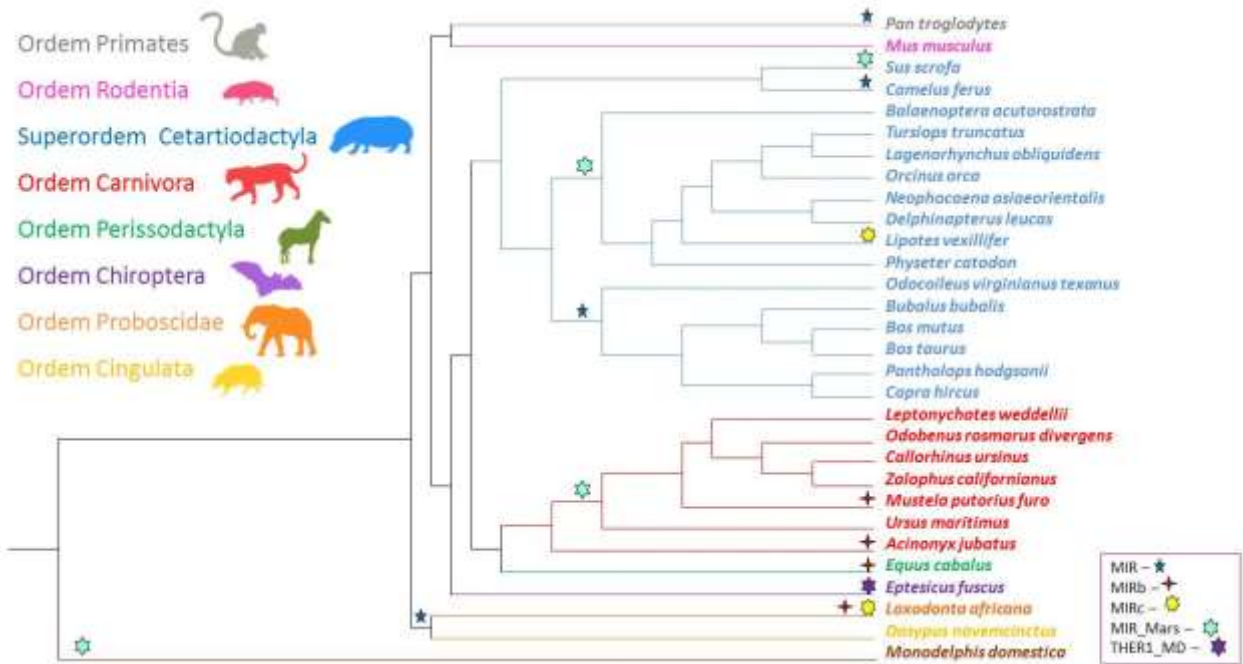
Foram identificados 30 elementos transponíveis (Tabela 2) nas sequências gênicas *HoxD10-Evx2* de mamíferos, contudo no grupo externo, genoma de *G. gallus*, não foi observado nenhum elemento. Dos elementos identificados apenas três eram de DNA (classe II) (Tabela 2). O restante dos TEs identificados são retrotransposon de classe I, onde 8 dos TEs são com LTRs e 19 sem LTR. Entre os retrotransposons sem LTR destacaram-se os SINEs, com 12 representantes, dentre esses os elementos MIR foram observados em 30 espécies de mamíferos.

Nas 31 sequências gênicas analisadas, apenas a espécie *Mus musculus*, não apresentou nenhuma inserção de elementos *MIR-like*. Nas outras espécies foram observadas cinco subfamílias distintas de MIR, inseridas na região entre *Evx2* e *HoxD13* (Figura 8A, 8B). A maior parte desses elementos possuem entre 150 à 160 pb.

Foram identificados cinco subfamílias de elementos MIR: MIR, Mars, MirB, MirC e THER1_MD, sendo MirC encontrado apenas em *Lipotes vexillifer* e THER1_MD apenas em *Eptesicus fuscus*. Enquanto Mars foi identificado em todos os cetáceos, no artiodáctilo *Sus scrofa*, nos carnívoros pinípedes, no *Mustela putorius furo*, no *Ursus maritimus*, e no marsupial *Monodelphis domestica*. (Figura 8A, 8B).



A



B

Figura 8A. Esquema da região de *HoxD*, incluindo os genes *HoxD10* (laranja), *HoxD11* (rosa), *HoxD12* (amarelo), *HoxD13* (verde) e *Exx2* (vermelho). Em cor cinza, está representada a região intergênicas, em preto e azul são os elementos transponíveis, sendo que em azul são elementos do tipo MIR.

B. Esquema filogenético das espécies utilizadas na análise, construído a partir de filogenias de trabalhos anteriores (Nyakatura e Bininda-emonds, 2012; Price et al., 2005; Song et al., 2012), mostrando quais espécies compartilham os diferentes tipos de MIR, lembrando que o elemento MIR se inseriu no genoma no início da radiação de mamíferos.

As sequências de MIR-Mars encontradas nesses genomas estão em torno de 160pb, *M. domestica* apresentou uma sequência menor, com 69 pb, e possuindo uma semelhança variando de 51-56% com os outros Mars (Figura 9). A similaridade entre as sequências de Mars dentro de Cetacea varia entre 96- 99%, sendo assim um elemento bem conservado dentre os cetáceos. Porém, a sequência homóloga de Mars em *Sus scrofa* possui uma similaridade aproximada de 85% com os elementos encontrados em Cetacea (Figura 9). A sequência homóloga ao elemento Mars de Cetacea apresentou uma similaridade de 74-80% com o observado Mars de Carnívora (Figura 9).

Os elementos da subfamília *MIR-like* em Artiodactyla apresentaram similaridade entre 92-100% e quando comparado às sequências entre Cetacea e Artiodactyla, apresentam similaridade que varia entre 76% e 87% (Figura 9).

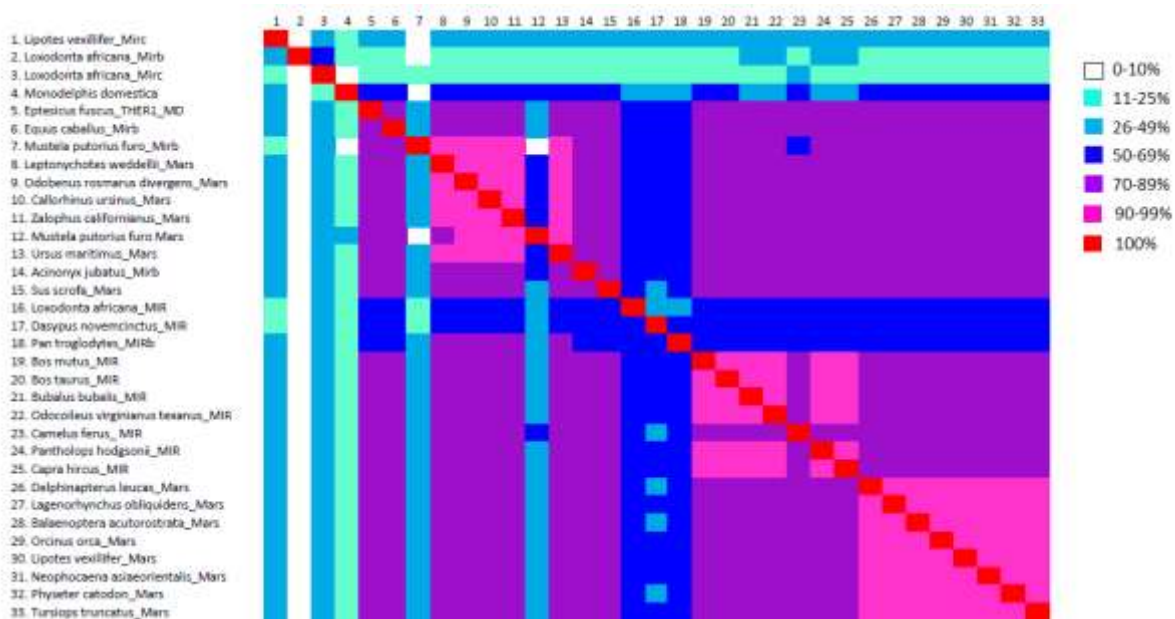


Figura 9. Figura gerada pelo software UGENE, indicando os níveis de similaridade entre as sequências dos elementos *MIR-like*.

A espécie *Camelus ferus* apresentou diferenças significativas, o elemento em *C. ferus* possui 162 pb, como observado nas sequências homólogas das outras espécies do grupo, porém a sequência possui similaridade de aproximadamente 73% quando comparando com as outras espécies de Artiodactyla (Figura 9).

Apenas os elementos da subfamília MIR foram investigados quanto à presença de sítios de fatores de transcrição, a predição feita pelo AliBaba 2.1 apontou a existência de doze TFBSs (Tabela 3), sendo seis desses sítios encontrados em Cetartiodactyla. O sítio de fator de transcrição Oct-1 foi o mais representativo, encontrado em 23 elementos e presente em quase todas as espécies de Cetartiodactyla analisadas. Dentro de Cetacea, todas as espécies com exceção de *Balaenoptera acutorostrata*, apresentaram três tipos de sítios em seu elemento Mars, os sítios: Oct-1, NF1 e GATA-3. Já em Artiodactyla, todas as espécies parecem possuir Oct-1, e *Camelus ferus* também parece possuir o sítio C/EBPalpha em sua sequência.

DISCUSSÃO

História evolutiva do *HoxD12* em Cetartiodactyla

Trabalhos anteriores forneceram poderosas evidências dos papéis dos genes *Hox* durante o desenvolvimento do plano do corpo, padrão esquelético e na formação dos membros (Kui et al., 2018; Pearson et al., 2005). O *HoxD12* foi apontado em diversos trabalhos como um gene importante na rota de membros dentro de vertebrados (Darbellay et al., 2019; Gonzalez et al., 2007; Kmita et al., 2002; Kui et al., 2018; Liang et al., 2013; Montavon et al., 2011, 2008; Nery et al., 2016; Papageorgiou, 2012; Villavicencio-Iorini et al., 2010; Williamson et al., 2012). Assim, avaliar esse gene na evolução de um dos grupos com as adaptações mais extremas, como os cetáceos, que tiveram seus membros terrestres anteriores transformados em nadadeiras e a perda dos posteriores é pertinente e instigante.

Apesar de existirem alguns trabalhos que mostram que o gene *HoxD12* sofreu alterações dentro de cetáceos (Kui et al., 2018; Liang et al., 2013; Nery et al., 2016), existe apenas uma referência na literatura (Wang et al., 2008) que constrói a história evolutiva de Cetartiodactyla utilizando o gene *HoxD12*, na qual utilizaram dez sequências de cetáceos e três de artiodáctilos, sendo dois ruminantes e um suíno. O acréscimo do hipopótamo, espécie existente mais próxima de cetáceos (Boisserie et al., 2005; Vislobokova, 2013) devido a seus hábitos e características intermediárias entre Artiodactyla e Cetacea, e provavelmente com hábito e características similares a cetáceos ancestrais (Thewissen et al., 2009), é oportuno e apropriado. Neste trabalho reconstruímos a árvore filogenética do gene *HoxD12* dentro de mamíferos (Figura 6), tendo como enfoque o grupo Cetartiodactyla. Também fizemos análises para detectar se há seleção desse gene, além de avaliar possíveis elementos causadores de alterações gênicas na região 5' do cluster *HoxD*, onde o *HoxD12* está inserido, assim contribuindo para entendermos melhor uma das histórias evolutivas mais curiosas e inusitadas entre vertebrados

A árvore filogenética por inferência Bayesiana (Figura 6) correspondeu à história evolutiva

das diferentes ordens de mamíferos de forma concordante com as filogenias anteriores (Foley et al., 2016; dos Reis et al., 2012; Nery et al., 2012; Nikaido et al., 1999; Song et al., 2012; Springer et al., 2004; Zhou et al., 2011). Contudo o estabelecimento das relações das espécies dentro dessas diferentes ordens nem sempre conseguiu ser bem resolvido. Isso provavelmente ocorreu devido ao alto grau de conservação do gene *HoxD12*, assim a maior parte das alterações nesse gene deve ter ocorrido antes ou no início da separação desses grupos, havendo pouca divergência entre as sequências entre as famílias de uma mesma ordem, resultando nesses casos nas politomias. Isso pode indicar que os genes do desenvolvimento por serem tão bem conservados, não sejam ideais para a construção de filogenias. Contudo, é pertinente estudar como genes tão importantes para determinar os diferentes planos corporais em animais evoluíram durante a história dos grupos, podendo, por exemplo, contribuir para as diferentes adaptações dos membros encontradas entre os mamíferos (Kui et al., 2018).

Estudos recentes discutem certa instabilidade na base raiz da filogenia de Cetartiodactyla, variando entre Suoidea ou Tylopoda como a primeira linhagem a divergir (Figura 6) (Lum et al., 2000; Vislobokova, 2013). Os nossos dados mostram uma posição basal de Tylopoda (PP = 1.0), que é concordante com outros trabalhos com abordagens genômicas, utilizando Citocromo B e também elementos transponíveis para as construções das filogenias (Leduc et al., 1999; Meredith et al., 2011; Song et al., 2012; Zhou et al., 2011). No entanto, inferências filogenéticas anteriores utilizando DNA mitocondrial apontam Suidae com uma divergência anterior a Tylopoda (Hassanin et al., 2012; Zurano et al., 2019). Esses diferentes resultados podem estar relacionados a diferenças nos métodos de inferência filogenética, marcadores moleculares, amostragem de táxons e escolha de grupos externos. Nossos resultados suportam a monofilia de Cetancodonta (Cetacea + Ancodonta) e também de Cetruminantia (Cetancodonta + Ruminantia), ou seja, que Hippopotamidae é o grupo irmão de Cetacea, o que está de acordo com outras evidências moleculares, morfológicas e paleontológicas (Hassanin et al., 2012; Vislobokova, 2013; Zhou et al., 2011; Zurano et al., 2019).

A família Hippopotamidae é semelhante aos cetáceos em certas características morfológicas relacionadas com adaptações ao ambiente aquático: por exemplo, a alta posição das órbitas no crânio e osteosclerose. Essas adaptações se desenvolveram convergentemente em diferentes grupos de vertebrados (por exemplo, em crocodilos). As sinapomorfias presumíveis do clado Cetacodonta (Hippopotamidae + Cetacea) (região facial alongada, a borda dos nasais posicionada posterior à margem orbital anterior, ausência quase completa de pelos e ausência de glândulas sebáceas (Geisler e Uhen, 2005) são consideradas não confiáveis por alguns autores, que discutem se elas podem ter surgido por convergência (Pickford, 2008; Vislobokova, 2013). Nossos dados moleculares suportam a hipótese de que Hippopotamidae é mais relacionado à Cetacea, e que compartilham um ancestral em comum, evidenciando que essas sinapomorfias são confiáveis e não resultado de convergência, e que muito provavelmente esse ancestral era um animal anfíbio com essas características, e se deslocava dentro da água como caminantes de fundos (Thewissen et al., 2009), assim como os hipopótamos.

Nas análises evolutivas do *HoxD12* (Figura 7), foi detectado que a maioria dos sítios está sobre forte restrições evolutivas, o que é esperado para os genes *Hox*, por serem genes conservados sob efeito de uma forte seleção purificadora (Nei, 2007; Pascual-Anaya et al., 2008). Contudo contabilizando todos os mamíferos analisados, foram encontrados 87 sítios evoluindo mais rapidamente do que a média dos demais sítios, entretanto mesmo sequências de evolução neutra podem ter um aumento no número de substituições por causa da variação na taxa de evolução ao longo de grandes escalas de tempo (Figura 7A) (Gillespie, 1988; Stolyarova et al., 2019), no caso de mamíferos estamos falando em cerca de 166,2 milhões de anos desde sua origem (Bininda-Emonds et al., 2007).

Comparando Cetacea e Artiodactyla (Figura 7B e 7C), podemos observar dois padrões diferentes. Em Artiodactyla (Figura 7B) apesar de apresentar 36 sítios com valores de taxa evolutiva média > 1 , os valores não estão muito acima de 1 (média igual à 1.03). O que sugere que essa restrição no número de mutações se deve a seleção purificadora normalmente observada

dentro de genes homeóticos. Enquanto Cetacea (Figura 7C) que apresenta menos sítios (26) evoluindo mais rapidamente que a média, teve valores médios de 1.56. Apesar da taxa evolutiva mais alta, não podemos afirmar que essas alterações de aminoácidos podem modificar a função da proteína resultante. Estudos estruturais e evolutivos sobre a localização desses aminoácidos seriam muito pertinentes.

Nas análises de seleção na região gênica analisada, apesar de sinais de seleção terem sido encontrados em alguns sítios, não obtiveram valores significativos, mas isso não descarta a possibilidade que em uma análise da sequência completa seja encontrado algum sítio que esteja sobre seleção de forma significativa, análises futuras mais robustas são necessárias.

Elementos transponíveis em região no cluster *HoxD*

Os genes homeóticos são conhecidos por serem altamente conservados entre as diferentes espécies, mutações nesses genes e também alterações na sua regulação podem resultar em modificações drásticas na morfologia nesses animais (Wagner et al., 2016). Por esse motivo elementos de caráter repetitivo, reconhecidos por serem fontes de reorganização gênica, são geralmente ausentes dentro desse *loci* em tetrápodes, com exceção da espécie de lagarto *Anolis carolinensis* (Di-Poï et al., 2009). Neste estudo identificamos diferentes sequências homólogas à TEs associadas aos genes *Hox* de mamíferos, diferentemente do que é observado em outras espécies de tetrápodes. Em Di-Poï et al. (2010) observou-se relaxamento na seleção purificadora nesses clusters na ordem Squamata, o que resultou em modificações nos padrões de expressão dos genes *Hox* e a presença de um grande número de elementos transponíveis entre os genes *HoxD10* e *Evx2*, comparado com outros grupos de tetrápodes, o que foi relacionado com o alongamento corporal e redução dos membros observados no plano corporal desses animais.

Embora algumas regiões extensas de baixa densidade de transposons tenham sido previamente observadas, particularmente na vizinhança dos genes *Hox* (Lander et al. 2001; Chinwalla et al. 2002; Wagner et al. 2003), existem poucos trabalhos avaliando a incidência desses

elementos nessas regiões (Simons et al., 2005).

Segundo Kui et al. (2018), embora não haja diferença significativa nos elementos repetitivos e sua distribuição nos cluster de *Hox* entre os diferentes mamíferos, há uma sutil diferença na densidade de elementos repetitivos (o comprimento de todas as repetições e no comprimento total do cluster *Hox*) em mamíferos marinhos quando comparado com as espécies mais relacionadas terrestres. Uma densidade aumentada de elementos repetitivos foi determinada em grupos *HoxC* em pinípedes, *HoxB* e *HoxC* em cetáceo e *HoxD* de sirênios. As mutações específicas da linhagem que sofreram alterações nas propriedades dos aminoácidos radicais das proteínas *Hox* são concentradas em cetáceos (*HoxA3*, *HoxB9*, *HoxD1* e *HoxD12*) e em pinípedes (*HoxA2*, *HoxB9*). Mutações nos genes *HoxB9* e *HoxD12* resultam em defeitos na porção terminal da coluna e na caixa torácica e nos membros, respectivamente (Alexander et al. 2009; Mallo et al. 2010). Assim, inserções de elementos repetitivos nessas sequências, como os elementos transponíveis podem acarretar modificações morfológicas na história evolutiva desses animais. Além disso, já foi visto que quando é inserido artificialmente dois transposons em torno do gene *Evx2* causa defeitos nos membros e afeta a transcrição dos genes *HoxD* a jusante (Monge et al. 2003).

Kui et al. (2018) apontaram no seu trabalho que os elementos transponíveis ERV, sem LTR (CR1, LINE e SINE) e DNA/hAT (Tabela 2) geralmente se repetem em mamíferos, que foram os elementos mais representativos da nossa análise. Além disso, esses elementos concentraram-se em regiões semelhantes entre diferentes mamíferos nos cluster de genes *Hox*; ou seja, repetições detectadas a partir do cluster *HoxA* localizado no 5'terminal *Evx1-HoxA13*, *HoxB* em *HoxB13-HoxB9* e *HoxD* na região de *HoxD9* a 3'terminal (Kui et al., 2018). Contudo em nosso trabalho foi analisada a porção do *HoxD10* até *Evx2* na porção 5'terminal, onde encontramos elementos transponíveis em todas as sequências de mamíferos analisadas (Figura 8 A e B).

Os elementos do tipo MIR tiveram destaque em nosso trabalho por serem os mais representativos, presentes em praticamente todos os mamíferos analisados, com exceção de *Mus*

musculus. Este dado corrobora com a literatura, onde foi demonstrada a ausência do elemento MIR em roedores comparando com outros mamíferos (Yu e Zhang, 2005). Segundo Jurka et al. (1995) este efeito pode ser explicado pela maior taxa de substituição em roedores do que em outros mamíferos, causando, assim, um decaimento mais rápido de segmentos genômicos não selecionados nessas espécies. Todas as outras espécies de mamíferos analisadas possuem um elemento MIR aproximadamente na mesma posição, entre o gene *Evx2* e o gene *HoxD13*, e num grau de conservação considerável, principalmente dentro de Cetacea (Figura 9).

Nossos dados indicam que a conservação dos elementos *MIR-like* na mesma região em praticamente todos os grandes grupos de mamíferos, e em adicional em outra posição do *cluster HoxD* em cetáceos pode indicar papel relevante na regulação desses genes. Visto que muitos trabalhos vêm demonstrando que os elementos MIRs podem desempenhar papéis importantes na regulação genética em mamíferos (Testori et al., 2012). Retrotransposons são bem conhecidos por serem cooptados pelo genoma do hospedeiro para regulação e até codificação de proteínas, por processo de exaptação (Krull et al., 2007). Especialmente em mamíferos, existem grandes variedades e quantidades de diferentes elementos e processos que levam a novos exóons, novos locais de *splicing* internos e pontos de ramificação que podem levar a variações do *splicing* (Schmitz e Brosius, 2011). Normalmente longos tempos evolutivos são necessários para adquirir e estabelecer essas novas condições. O elemento MIR foi amplificado dentro do genoma dos primeiros mamíferos há 130 milhões de anos (Jurka et al., 1995). Portanto, elementos como MIRs com uma longa história dentro de mamíferos são mais frequentemente encontrados exonizados (Lin et al., 2009; Schmitz e Brosius, 2011). Yu e Zhang (2005) sugerem que há uma forte tendência de famílias SINEs específicas se inserirem dentro ou muito próximas das MIRs preexistentes, levando-nos a concluir que o elemento MIR pode ser um *hotspot* para integrações sucessivas de outros SINEs.

Em geral, apesar de algum avanço em nosso conhecimento de como a presença massiva de MIRs impacta a biologia do genoma, sua contribuição para o transcriptoma ainda é pouco

explorada, assim como seu mecanismo de transcrição e regulação. Em particular, os elementos de ação *cis* que controlam a transcrição de MIR nunca foram sistematicamente estudados. Apesar de os elementos MIR serem inativos, eles são amplamente transcritos e, portanto, contribuem para o transcriptoma como foi visto em células humanas (Carnevali et al., 2017).

Apesar de o elemento MIR estar presente em praticamente todas as espécies analisadas, houve diferenças encontradas nos tipos de MIRs entre os grupos, e existe um grau de conservação dentro dos diferentes grupos desse elemento, o que pode ser um indicativo que talvez esse elemento possa ter alguma função na regulação desses genes. Os cetáceos apresentaram o elemento Mars do tipo MIR, com um alto nível de homologia entre as sequências, sua similaridade varia de 96 a 99% (Figura 9), diferenciando da maioria dos artiodáctilos que apresentaram o elemento MIR do tipo MIR, somente o artiodáctilo *Sus scrofa* apresentou o mesmo elemento que Cetacea, todavia as filogenias apontam que Ruminantia sendo o clado mais próximo de Cetacea do que Suidae, que é grupo ao qual pertence *Sus scrofa* (Lum et al., 2000; Zhou et al., 2011). Diferenças nas sequências de *Sus scrofa* podem ter convergido para uma sequência mais similar com a encontrada em cetáceos. Ou ao contrário, já que Mars também é encontrado em marsupial, ele pode ser um elemento mais basal, e o elemento dessas outras espécies de artiodáctilos é que divergiu ao decorrer do tempo, mas análises subsequentes são necessárias para testar esse hipótese.

O elemento Mars que foi encontrado no marsupial *Monodelphis domestica*, apesar de ser apontado como o mesmo elemento pelo RepeatMasker, possui um tamanho muito menor e com uma divergência bem alta em sua sequência comparado aos outros Mars. Além de *Monodelphis domestica*, o elemento Mars foi encontrado em outros mamíferos associados com ambientes aquáticos, todos pertencentes à ordem Carnivora. Apesar de também apresentarem divergências nas suas sequências (similaridade de 74-80%) (Figura 9) é instigante observar que esses animais possuem o mesmo tipo de elemento, e seria muito interessante investigar se existe algum tipo de convergência entre essas sequências.

para sua forma hidrodinâmica convergente, a fim de se adaptar ao ambiente aquático. Considerando a sobreposição funcional de diferentes genes *Hox*, as assinaturas moleculares adaptativas desses genes também podem sugerir uma rota evolutiva diferente para as morfologias de membros semelhantes em baleias e pinípedes. Isto parece conformar-se com a suposição de que a convergência fenotípica de mamíferos marinhos pode ser alcançada através da seleção de diferentes genes *Hox* com funções equivalentes propostas por Nery et al. (2016). Embora a evolução fenotípica convergente possa resultar da evolução molecular convergente, esses casos são raros, e a evolução utiliza com mais frequência vias moleculares diferentes para alcançar um resultado fenotípico similar. (Foote et al., 2015). No entanto, Nery et al. (2016) detectaram apenas uma pequena minoria de genes sobre seleção positiva com funções sobrepostas em mamíferos aquáticos, o que pareceu ser insuficientemente convincente no que diz respeito às notáveis características fenotípicas convergentes de mamíferos aquáticos.

Kui et al. (2018) detectou a substituição paralela de aminoácidos de *HoxD12* em cetáceos e no peixe-boi. As análises evolutivas esclareceram que a evolução adaptativa de *HoxD12* contribuiu para a origem e diversificação dos cetáceos (Wang et al. 2008). Assim, a substituição paralela de aminoácidos de *HoxD12* pode estar ligada à origem das nadadeiras de baleias e também no peixe-boi. Uma substituição paralela de aminoácidos (sítio: 164) no *HoxD12* foi encontrada em pinípedes e em *Xenarthra*. Os membros dos pinípedes se especializaram em nadadeiras curtas e planas para produzir propulsão enquanto nadam (Pierce et al. 2011). Em contraste, os tatus modificaram seus membros para escavação, com grandes garras e longos braços de alavanca para facilitar a linha de ação dos principais músculos (Vizcaino et al. 2006). Além disso, as preguiças desenvolveram uma forma semelhante a um gancho com membros longos e delgados, equipados com garras longas, curvadas e pontiagudas (Britton, 1941).

Assim, a substituição paralela de aminoácidos pode estar associada à sua adaptação a estilos de vida diferentes, apesar de não haver convergência morfológica de membros entre essas duas linhagens. Substituições paralelas de aminoácidos de alguns genes *Hox* também foram

descobertas entre mamíferos marinhos e seus contemporâneos terrestres mais próximos. Estes resultados foram consistentes com a inferência de que morfologias similares podem ser derivadas por evolução molecular independente (Foote et al. 2015).

Sítios de ligação de fatores de transcrição nos elementos MIR

A composição característica de uma célula é determinada pelo seu padrão de expressão gênica. O primeiro passo na produção de uma proteína é a transcrição do DNA em um RNA mensageiro que será então traduzido, modificado e transportado para o local apropriado na célula. A transcrição por ser o primeiro passo na síntese de proteínas, é também um importante ponto de regulação. A expressão apropriada dos genes específicos requeridos por um tipo de célula definido é determinada pela estrutura da cromatina que pode ser mais ou menos permissiva para a maquinaria de transcrição acessar ao local de início da transcrição de um gene. A transcrição começa pelo recrutamento de fatores de transcrição do núcleo para o promotor para formar um complexo de pré-iniciação da transcrição (Pance, 2016).

Nos últimos anos, várias análises mostraram que os TEs do genoma humano são significativamente enriquecidos com sítios de ligação para fatores de transcrição (Schmid e Bucher, 2010; Wang et al., 2007). Esta associação entre fatores de transcrição e elementos transponíveis sugere que os eventos de transposição poderiam ter desempenhado um papel central na emergência e sucesso da regulação combinatória de genes em eucariotos complexos. De acordo com esta hipótese, quando um elemento transponível com uma “boa” combinação de sequências de ligação é transposto perto da região reguladora de um gene alvo, é “exaptado”, ou seja, utilizado para alguma função adaptativa pelo hospedeiro (Gould, Jay e Vrba, 2013) e assim conservado pela evolução. A mutação estocástica da sequência transponível original pode então ajustar o módulo regulador eliminando ou adicionando motivos de ligação adequados. De fato, há evidências claras de que várias famílias de TEs estão sob forte seleção purificadora (Kamal et al., 2006; Lowe et al.,

2007) e que a maioria dos elementos conservados não codificadores conhecidos estão localizadas em elementos transponíveis laterais (Gentles et al., 2007).

Uma característica interessante é que a maioria das famílias de TEs enriquecidas com sítios de fatores de transcrição são transposons antigos, isto é particularmente verdade para as sequências tipo MIR (Testori et al., 2012). Foi sugerido que essa notável conservação poderia ter uma origem regulatória e que esse núcleo compartilhado poderia conter as sequências de ligação de várias proteínas de ligação ao DNA (Sironi et al., 2006; Smit e Riggs, 1995). Análises sugerem que, quando existe uma associação específica entre um fator de transcrição e uma classe de transposons, o sucesso evolutivo desse fator pode estar relacionado ao sucesso do elemento associado (Testori et al., 2012).

As predições feitas nas análises para possíveis sítios de fatores de transcrição indicaram que quase todos os elementos MIRs (Tabela 3) possuem o sítio para o fator Oct-1. A família de fatores de transcrição POU inclui mais de uma dúzia de proteínas que foram implicadas em uma variedade de funções. Vários membros são específicos do tipo celular e, portanto, de expressão limitada, enquanto Oct-1 é onipresente nos tecidos e de expressão constitutiva (Ferraris et al., 2011). Embora muitos papéis tenham sido atribuídos a Oct-1, sua função permanece um tanto obscura, mas sua ampla distribuição sugere que ela desempenha um papel fundamental na homeostase celular (Pance, 2016). Consistente com essa propriedade, Oct-1 foi implicado na regulação transcricional basal (Kim e Peterson, 1995). No entanto, também foi descrito respondendo à sinalização através de outros fatores de transcrição indutíveis (Xie, 1997).

Estes achados implicam que Oct-1 tem uma função dupla de manter a expressão de algumas proteínas gerais, bem como controlar a expressão de genes indutíveis. Também foi reconhecido que Oct-1 facilita o recrutamento de TBP (do inglês *TATA binding protein*, proteína de ligação ao *TATA box*) nos promotores através de uma interação direta (Bertolino e Singh, 2002). Além disso, foi recentemente relatado que Oct-1 pode modular o contexto da cromatina em torno dele para regular a expressão de genes alvo (Shakya et al., 2011). Assim, o Oct-1 foi descrito

por dimerizar com vários TFs (Donner et al., 2007), contudo em muitos desses casos as consequências transcricionais dessas interações não foram exploradas (Pance, 2016).

As análises indicaram a presença de sítios para fatores de transcrição na sequência de todos os cetáceos examinados neste trabalho, com exceção de *Balaenoptera acutorostrata*, uma possível presença de três TFBS: Oct-1, GATA-3 e NF1. A família de fatores de transcrição GATA é composta por seis membros, incluindo o GATA1-6, e foi originalmente considerada por seus papéis em células hematopoiéticas (Viger et al., 2008). Atualmente, os fatores GATA são conhecidos por serem expressos em uma ampla variedade de tecidos, onde atuam como reguladores críticos da expressão gênica específica celular. Entre eles, GATA-3 tem sido convencionalmente considerado como um fator de transcrição que regula a diferenciação e proliferação de células T (Hoyler et al., 2012). Recentemente, nota-se que GATA-3 também funciona em células não hematopoiéticas (Wan, 2014). Por exemplo, GATA-3 é encontrado no sistema nervoso com papéis importantes para controlar o desenvolvimento neuronal e sobrevivência de células neuronais (Tsarovina et al., 2010; Zheng et al., 2019). Já o NF1 é constitutivamente expresso na maioria dos tecidos humanos e de camundongos e atua como um importante regulador do crescimento e diferenciação celular (Mancini et al., 1999). Para se confirmar a presença desses sítios nas sequências MIR encontradas e como esses fatores de transcrição podem estar atuando e quais genes eles podem estar regulando, outras análises necessitam ser feitas.

CONSIDERAÇÕES FINAIS

Assim, este trabalho contribui para um melhor entendimento da evolução do cluster *HoxD* dentro de Cetartiodactyla, reconstruindo a história filogenética de um dos seus genes (*HoxD12*) e verificando sinais de seleção, que apesar de não detectados nesse presente estudo, não descarta a possibilidade de existirem, necessitando análises mais aprofundadas de toda a sequência desse gene. Também temos que levar em conta que, mesmo que não haja alterações na região gênica, pode haver diferenças nas regiões regulatórias entre esses Cetacea e Artiodactyla que explique as alterações drásticas na morfologia desses membros, além de que outros genes *Hox* também atuam no desenvolvimento de membros, podendo assim ser vários fatores que juntos são responsáveis por essas modificações que diferenciam Cetacea e Artiodactyla. A inclusão do Hipopótamo nessas análises é muito pertinente, já que este animal apesar de morfologicamente se assemelhar mais com Artiodactyla, todos os estudos moleculares até o momento mostram que ele é mais relacionado com Cetacea, portanto é um caso bastante apropriado para entendermos melhor a relação da evolução molecular com a evolução morfológica. Além disso, detectamos não só uma família de TEs presente nessas sequências, mas também uma possível presença de sítios de ligação de fatores de transcrição. Estudos futuros para avaliar a influência desses elementos nessa região são relevantes e adequados. Portanto, é interessante e conveniente ampliar o nosso conhecimento sobre genes do desenvolvimento na história evolutiva de Cetartiodactyla.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alexander T, Nolte C and Krumlauf R (2009) Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol* 25:431–456.
- Amores A, Suzuki T, Yan Y-L, Pomeroy J, Singer A, Amemiya C and Postlethwait JH (2004) Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res* 14:1–10.
- Armour JAL, Wong Z, Wilson V, Royle NJ and Jeffreys AJ (1989) Sequences flanking the repeat arrays of human minisatellites: association with tandem and dispersed repeat elements. *Nucleic Acids Res* 17:4925–4936.
- Bao W, Kojima KK and Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11.
- Beeson D, Morris A, Vincent A and Newsom-Davis J (1990) The human muscle nicotinic acetylcholine receptor alpha-subunit exist as two isoforms: a novel exon. *EMBO J* 9:2101–2106.
- Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ and Haussler D (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441:87–90.
- Berta A, L. Sumach J and Kovacs K (2007) *Marine mammals: evolutionary biology* (second edition). *Polar Res* 560pp.
- Bertocchi NA, de Oliveira TD, Del Valle Garnero A, Coan RLB, Gunski RJ, Martins C and Torres FP (2018) Distribution of CR1-like transposable element in woodpeckers (Aves Piciformes): Z sex chromosomes can act as a refuge for transposable elements. *Chromosome Res* 26:333–343.
- Bertolino E and Singh H (2002) POU/TBP Cooperativity: A Mechanism for Enhancer Action from a Distance. *Mol Cell* 10:397–407.

- Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL and Purvis A (2007) The delayed rise of present-day mammals. *Nature* 446:507-512.
- Boisserie J-R, Lihoreau F and Brunet M (2005) The position of Hippopotamidae within Cetartiodactyla. *Proc Natl Acad Sci U S A* 102:1537-1541.
- Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew J-L, Ruan Y, Wei C-L, Ng HH et al. (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* 18:1752–1762.
- Britton SW (1941) Form and Function in the Sloth. *Q Rev Biol* 16:13-34.
- Burke AC, Nelson CE, Morgan BA and Tabin C (1995) Hox genes and the evolution of vertebrate axial morphology. *Development* 121:333-346.
- Carnevali D, Conti A, Pellegrini M and Dieci G (2017) Whole-genome expression analysis of mammalian-wide interspersed repeat elements in human cell lines. *DNA Res* 24:59–69.
- Chen K and Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8:93-103.
- Chinwalla AT, Cook LL, Delehaunty KD, Fewell GA, Fulton LA, Fulton RS, Graves TA, Hillier LW, Mardis ER, McPherson JD et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520-562.
- Chuong EB, Elde NC and Feschotte C (2016) Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 18:71-86.
- Cooper LN, Berta A, Dawson SD and Reidenberg JS (2007) Evolution of hyperphalangy and digit reduction in the cetacean manus. *Anat Rec (Hoboken)* 290:654–672.
- Darbellay F, Bochaton C, Lopez-Delisle L, Mascrez B, Tschopp P, Delpretti S, Zakany J and Duboule D (2019) The constrained architecture of mammalian Hox gene clusters. *Proc Natl Acad Sci* 116:13424-13433.
- Degen SJ and Davie EW (1987) Nucleotide sequence of the gene for human prothrombin.

Biochemistry 26:6165–6177.

Deméré TA, McGowen MR, Berta A and Gatesy J (2008) Morphological and Molecular Evidence for a Stepwise Evolutionary Transition from Teeth to Baleen in Mysticete Whales. *Syst Biol* 57:15–37.

Deschamps J and van Nes J (2005) Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* 132:2931–2942.

Di-poï N, Montoya-burgos JI and Duboule D (2009) Atypical relaxation of structural constraints in Hox gene clusters of the green anole lizard. *Genome Res* 19:602–610.

Di-Poï N, Montoya-Burgos JI, Miller H, Pourquié O, Milinkovitch MC and Duboule D (2010) Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* 464:99-103.

Donner AL, Episkopou V and Maas RL (2007) Sox2 and Pou2f1 interact to control lens and olfactory placode development. *Dev Biol* 303:784–799.

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15.

Du Toit Z, Grobler JP, Kotzé A, Jansen R, Brettschneider H and Dalton DL (2014) The complete mitochondrial genome of Temminck's ground pangolin (*Smutsia temminckii*; Smuts, 1832) and phylogenetic position of the Pholidota (Weber, 1904). *Gene* 551:49–54.

Fedoroff N V and Bennetzen JL (2013) Transposons, Genomic Shock, and Genome Evolution. *Plant Transposons Genome Dyn Evol* 181–201.

Ferraris L, Stewart AP, Kang J, DeSimone AM, Gemberling M, Tantin D and Fairbrother WG (2011) Combinatorial binding of transcription factors in the pluripotency control regions of the genome. *Genome Res* 21:1055–1064.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9:397-405.

Feschotte C and Pritham EJ (2007) DNA Transposons and the Evolution of Eukaryotic Genomes.

Annu Rev Genet 41:331–368.

Fitzgerald EM (2006) A bizarre new toothed mysticete (Cetacea) from Australia and the early evolution of baleen whales. *Proc Biol Sci* 273:2955–2963.

Foley NM, Springer MS and Teeling EC (2016) Mammal madness: is the mammal tree of life not yet resolved? *Philos Trans R Soc Lond B Biol Sci* 371.

Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V et al. (2015) Convergent evolution of the genomes of marine mammals. *Nat Genet* 47:272–275.

Gagneux P and Diego S (2002) The Evolution of Developmental Pathways. *J Hered* 93:460–461.

Garcia-Fernàndez J (2005) Hox, ParaHox, ProtoHox: facts and guesses. *Heredity (Edinb)* 94:145–152.

Gatesy J, Milinkovitch M, Waddell V and Stanhope M (1999) Stability of Cladistic Relationships between Cetacea and Higher-Level Artiodactyl Taxa. *Syst Biol* 48:6–20.

Geisler JH and Theodor JM (2009) Hippopotamus and whale phylogeny. *Nature* 458:E1–E4.

Geisler JH and Uhen MD (2005) Phylogenetic Relationships of Extinct Cetartiodactyls: Results of Simultaneous Analyses of Molecular, Morphological, and Stratigraphic Data. *J Mamm Evol* 12:145–160.

Gellon G and McGinnis W (1998) Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *Bioessays* 20:116–125.

Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, Pollock DD and Jurka J (2007) Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res* 17:992–1004.

Gérard M, Duboule D and Zákány J (1993) Structure and activity of regulatory elements involved in the activation of the *Hoxd-11* gene during late gastrulation. *EMBO J* 12:3539–3550.

Gilbert N and Labuda D (1999) CORE-SINES: Eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci* 96:2869–2874.

- Gillespie JH (1988) More on the overdispersed molecular clock. *Genetics* 118:385–388.
- Gingerich PD, Smith BH and Simons EL (1990) Hind limbs of eocene basilosaurus: evidence of feet in whales. *Science* 249:154–157.
- Gingerich PD, Haq M ul, Zalmout IS, Khan IH and Malkani MS (2001) Origin of Whales from Early Artiodactyls: Hands and Feet of Eocene Protocetidae from Pakistan. *Science* 293:2239-2242.
- Goldbogen JA, Calambokidis J, Shadwick RE, Oleson EM, McDonald MA and Hildebrand JA (2006) Kinematics of foraging dives and lunge-feeding in fin whales. *J Exp Biol* 209:1231-1244.
- Gonzalez F, Duboule D and Spitz F (2007) Transgenic analysis of Hoxd gene regulation during digit development. *Dev Biol.* 306:847–859.
- Goodchild NL, Wilkinson DA and Mager DL (1992) A human endogenous long terminal repeat provides a polyadenylation signal to a novel, alternatively spliced transcript in normal placenta. *Gene* 121:287–294.
- Gould SJ and Vrba ES (1982) Exaptation-A Missing Term in the Science of Form. *Paleobiology* 8:4–15.
- Graham A, Papalopulu N and Krumlauf R (1989) The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* 57:367–378.
- Gray YH (2000) It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet* 16:461–468.
- Grey D and Sadoff CW (2007) Sink or Swim? Water Security for Growth and Development. *Water Policy* 9:545-571.
- Harendza CJ and Johnson LF (1990) Polyadenylation Signal of the Mouse Thymidylate Synthase Gene was Created by Insertion of an L1 Repetitive Element Downstream of the Open Reading Frame. *Proc Natl Acad Sci U S A* 87:2531–2535.
- Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, Ruiz-Garcia M,

- Catzefflis F, Areskoug V, Nguyen TT et al. (2012) Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol* 335:32–50.
- Hoyler T, Klose CSN, Souabni A, Turqueti-Neves A, Pfeifer D, Rawlins EL, Voehringer D, Busslinger M and Diefenbach A (2012) The Transcription Factor GATA-3 Controls Cell Fate and Maintenance of Type 2 Innate Lymphoid Cells. *Immunity* 37:634–648.
- Jefferson TA, Webber MA and Pitman RL (2015) *Marine Mammals of the World: A Comprehensive Guide to Their Identification*. Academic Press, 616 pp.
- Jurka J, Kaplan DJ, Duncan CH, Walichiewicz J, Milosavljevic A, Murali G and Solus JF (1993) Identification and characterization of new human medium reiteration frequency repeats. *Nucleic Acids Res* 21:1273–1279.
- Jurka J, Zietkiewicz E and Labuda D (1995) Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. *Nucleic Acids Res* 23:170–175.
- Kamal M, Xie X and Lander ES (2006) A large family of ancient repeat elements in the human genome is under strong selection. *Proc Natl Acad Sci U S A* 103:2740-2745.
- Kapitonov VV and Jurka J (2004) RandI-1, a family of RandI non-LTR retrotransposons from the *Chlamydomonas reinhardtii* genome. *Repbase Reports* 4:196.
- Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780.
- Kim MH and Peterson DO (1995) Stimulation of basal transcription from the mouse mammary tumor virus promoter by Oct proteins. *J Virol* 69:4717–4726.
- Kmita M and Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301:331–333.
- Kmita M, Tarchini B, Zàkány J, Logan M, Tabin CJ and Duboule D (2005) Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. *Nature* 435:1113–1116.
- Krull M, Petrusma M, Makalowski W, Brosius J and Schmitz J (2007) Functional persistence of

exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res* 17:1139–1145.

Kumar S, Stecher G and Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870–1874.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K et al. (2006) Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* 125:301–313.

Lemons D and McGinnis W (2006) Genomic evolution of Hox gene clusters. *Science* 313:1918–1922.

Letunic I and Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47:W256–W259.

Levine M and Tjian R (2003) Transcription regulation and animal diversity. *Nature* 424:147–151.

Li K, Sun X, Chen M, Sun Y, Tian R, Wang Z, Xu S and Yang G (2018) Evolutionary changes of Hox genes and relevant regulatory factors provide novel insights into mammalian morphological modifications. *Integr Zool* 13:21–35.

Liang L, Shen Y-Y, Pan X-W, Zhou T-C, Yang C, Irwin DM and Zhang Y-P (2013) Adaptive Evolution of the Hox Gene Family for Development in Bats and Dolphins. *PLoS One* 8:e65944.

Loreto ELS, Depra M, Diesel JF, Panzera Y and Valente-Gaiesky VLS (2018) *Drosophila* relics hobo and hobo-MITEs transposons as raw material for new regulatory networks. *Genet Mol Biol* 41:198–205.

Lowe CB, Bejerano G and Haussler D (2007) Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci* 104:8005–8010.

- Lum JK, Nikaido M, Shimamura M, Shimodaira H, Shedlock AM, Okada N and Hasegawa M (2000) Consistency of SINE insertion topology and flanking sequence tree: quantifying relationships among cetartiodactyls. *Mol Biol Evol* 17:1417-1424.
- Mallo M, Wellik DM and Deschamps J (2010) Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* 344:7–15.
- Mancini DN, Singh SM, Archer TK and Rodenhiser DI (1999) Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. *Oncogene* 18:4108–4119.
- Margalit H, Nadir E and Ben-Sasson SA (1994) A complete Alu element within the coding sequence of a central gene. *Cell* 78:173–174.
- Mario dos R, Jun I, Masami H, J. AR, J. DPC and Ziheng Y (2012) Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B Biol Sci* 279:3491–3500.
- McClintock B (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* 36:344–355.
- McDonald JF (1993) Evolution and consequences of transposable elements. *Curr Opin Genet Dev* 3:855–864.
- Meredith RW, Gatesy J, Cheng J and Springer MS (2011) Pseudogenization of the tooth gene enamelysin (MMP20) in the common ancestor of extant baleen whales. *Proceedings Biol Sci* 278:993–1002.
- Milinkovitch MC, Meyer A and Powell JR (1994) Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. *Mol Biol Evol* 11:939–948. doi: 10.1093/oxfordjournals.molbev.a040164
- Monge I, Kondo T and Duboule D (2003) An enhancer-titration effect induces digit-specific regulatory alleles of the *hoxd* cluster. *Dev Biol* 256:212–220.
- Montavon T, Le Garrec J-F, Kerszberg M and Duboule D (2008) Modeling Hox gene regulation in

- digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev* 22:346–359.
- Montavon T, Soshnikova N, Mascres B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F and Duboule D (2011) A Regulatory Archipelago Controls *Hox* Genes Transcription in Digits. *Cell* 147:1132–1145.
- Murnane JP and Morales JF (1995) Use of a mammalian interspersed repetitive (MIR) element in the coding and processing sequences of mammalian genes. *Nucleic Acids Res* 23:2837–2839.
- Nei M (2007) The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci* 104:12235–12242.
- Nery MF, González DJ, Hoffmann FG and Opazo JC (2012) Resolution of the laurasiatherian phylogeny: Evidence from genomic data. *Mol Phylogenet Evol* 64:685–689.
- Nery MF, Borges B, Dragalzew AC and Kohlsdorf T (2016) Selection on different genes with equivalent functions: the convergence story told by *Hox* genes along the evolution of aquatic mammalian lineages. *BMC Evol Biol* 16:113.
- Nikaido M, Rooney AP and Okada N (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci* 96:10261–10266.
- Nishihara H, Smit AFA and Okada N (2006) Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res* 16:864–874.
- Nyakatura K and Bininda-Emonds ORP (2012) Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. *BMC Biol* 10:12.
- Nylander J (2004) MrModeltest V2. Program Distributed by the Author.
- O’Leary MA and Geisler JH (1999) The Position of Cetacea Within Mammalia: Phylogenetic Analysis of Morphological Data from Extinct and Extant Taxa. *Syst Biol* 48:455–490.
- Oliveira SG, Cabral-de-Mello DC, Moura RC and Martins C (2013) Chromosomal organization and evolutionary history of Mariner transposable elements in Scarabaeinae coleopterans. *Mol*

Cytogenet 6:54.

Pace 2nd JK and Feschotte C (2007) The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res* 17:422–432.

Pance A (2016) Oct-1, to go or not to go? That is the PolIII question. *Biochim Biophys Acta - Gene Regul Mech* 1859:820–824.

Papageorgiou S (2009) A biophysical mechanism may control the collinearity of Hoxd genes during the early phase of limb development. *Hum Genomics* 3:275–280.

Papageorgiou S (2017) Physical Forces May Cause the HoxD Gene Cluster Elongation. *Biology (Basel)* 6:32.

Pascual-Anaya J, D’Aniello S and Garcia-Fernández J (2008) Unexpectedly large number of conserved noncoding regions within the ancestral chordate Hox cluster. *Dev Genes Evol* 218:591–597.

Pearson JC, Lemons D and McGinnis W (2005) Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* 6:893-904.

Pierce SE, Clack JA and Hutchinson JR (2011) Comparative axial morphology in pinnipeds and its correlation with aquatic locomotory behaviour. *J Anat* 219:502–514.

Polak P and Domany E (2006) Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7:133.

Polavarapu N, Mariño-Ramírez L, Landsman D, McDonald JF and Jordan IK (2008) Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. *BMC Genomics* 9:226.

Price SA, Bininda-Emonds ORP and Gittleman JL (2005) A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). *Biol Rev Camb Philos Soc* 80:445–473.

Proudfoot NJ and Brownlee GG (1976) 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature* 263:211–214.

- Robins DM and Samuelson LC (1992) Retrotransposons and the evolution of mammalian gene expression. *Genetica* 86:191–201.
- Ronquist F and Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Santangelo AM, de Souza FSJ, Franchini LF, Bumashny VF, Low MJ and Rubinstein M (2007) Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene. *PLOS Genet* 3:e166.
- Schaeffer B (1947) Notes on the origin and function of the artiodactyl tarsus. *Notes Orig. Funct. artiodactyl tarsus. Am. Museum Novit* 1356.
- Schmid CD and Bucher P (2010) MER41 Repeat Sequences Contain Inducible STAT1 Binding Sites. *PLoS One* 5:e11425.
- Schmitz J and Brosius J (2011) Exonization of transposed elements: A challenge and opportunity for evolution. *Biochimie* 93:1928–1934.
- Shakya A, Kang J, Chumley J, Williams MA and Tantin D (2011) Oct1 Is a Switchable, Bipotential Stabilizer of Repressed and Inducible Transcriptional States. *J Biol Chem* 286:450–459.
- Shubin NH and Marshall CR (2000) Fossils, Genes, and the Origin of Novelty. *Paleobiology* 26:324–340.
- Simons C, Pheasant M, Makunin I V and Mattick JS (2006) Transposon-free regions in mammalian genomes. *Genome Res* 16:164–172.
- Sironi M, Menozzi G, Comi GP, Cereda M, Cagliani R, Bresolin N and Pozzoli U (2006) Gene function and expression level influence the insertion/fixation dynamics of distinct transposon families in mammalian introns. *Genome Biol* 7:R120.
- Smit A, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Smit AFA and Riggs AD (1995) MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res* 23:98–102.

- Smit AFA (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 9:657–663.
- Song S, Liu L, Edwards S V and Wu S (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci* 109:14942-14947.
- Spitz F, Gonzalez F and Duboule D (2003) A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the Hox Cluster. *Cell* 113:405–417.
- Springer MS, Stanhope MJ, Madsen O and De Jong WW (2004) Molecules consolidate the placental mammal tree. *Trends Ecol Evol* 19:430–438.
- Stolyarova A V, Bazykin GA, Neretina T V, Kondrashov AS, V. SA, A. BG, V. NT and S. KA (2019) Bursts of amino acid replacements in protein evolution. *R Soc Open Sci* 6:181095.
- Testori A, Caizzi L, Cutrupi S, Friard O, De Bortoli M, Cora' D and Caselle M (2012) The role of Transposable Elements in shaping the combinatorial interaction of Transcription Factors. *BMC Genomics* 13:1–16.
- Thewissen JGM, Cohn MJ, Stevens LS, Bajpai S, Heyning J and Horton WE (2006) Developmental basis for hind-limb loss in dolphins and origin of the cetacean bodyplan. *Proc Natl Acad Sci* 103:8414–8418.
- Thewissen JGM, Cooper LN, Clementz MT, Bajpai S and Tiwari BN (2007) Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature* 450:1190–1194.
- Thewissen JGM, Cooper LN, George JC and Bajpai S (2009) From Land to Water: The Origin of Whales, Dolphins, and Porpoises. *Evol Educ Outreach* 2:272–288.
- Thewissen JGM, Williams EM, Roe LJ and Hussain ST (2001) Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. *Nature* 413:277–281.
- Tsarovina K, Reiff T, Stubbusch J, Kurek D, Grosveld FG, Parlato R, Schutz G and Rohrer H (2010) The Gata3 Transcription Factor Is Required for the Survival of Embryonic and Adult Sympathetic Neurons. *J Neurosci* 30:10833–10843.

- Tschopp P and Duboule D (2011) A Genetic Approach to the Transcriptional Regulation of Hox Gene Clusters . *Annu Rev Genet* 45:145–166.
- Varagona MJ, Purugganan M and Wessler SR (2007) Alternative Splicing Induced by Insertion of Retrotransposons into the Maize waxy Gene. *Plant Cell* 4:811.
- Viger RS, Guittot SM, Anttonen M, Wilson DB and Heikinheimo M (2008) Role of the GATA Family of Transcription Factors in Endocrine Development, Function, and Disease. *Mol Endocrinol* 22:781–798.
- Villavicencio-Lorini P, Kuss P, Friedrich J, Haupt J, Farooq M, Turkmen S, Duboule D, Hecht J and Mundlos S (2010) Homeobox genes d11-d13 and a13 control mouse autopod cortical bone and joint formation. *J Clin Invest* 120:1994–2004.
- Vislobokova IA (2013) On the origin of Cetartiodactyla: Comparison of data on evolutionary morphology and Molecular biology. *Paleontol J* 47:321–334.
- Vizcaíno SF, Bargo MS, Kay RF and Milne N (2006) The armadillos (Mammalia, Xenarthra, Dasypodidae) of the Santa Cruz Formation (early-middle Miocene): An approach to their paleobiology. *Palaeogeogr Palaeoclimatol Palaeoecol* 237:255-269.
- Wagner GP, Amemiya C and Ruddle F (2003) Hox cluster duplications and the opportunity for evolutionary novelties. *Proc Natl Acad Sci U S A* 100:14603–14606.
- Wagner GP, Amemiya C, Ruddle F, Wagner GP, Amemiyat C and Ruddle F (2016) Linked references are available on JSTOR for this article : Hox cluster duplications and the opportunity for evolutionary novelties. *Proc Natl Acad Sci U S A* 100:14603–14606.
- Wan YY (2014) GATA3: A master of many trades in immune regulation. *Trends Immunol* 35:233–242.
- Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK and Haussler D (2007) Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci* 104:18613–18618.
- Wang Z, Yuan L., Rossiter SJ, Zuo X, Ru B, Zhong H, Han N, Jones G, Jepson PD and Zhang S

- (2009) Adaptive Evolution of 5'HoxD Genes in the Origin and Diversification of the Cetacean Flipper. *Mol Biol Evol* 26:613-622.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973–982.
- Williamson I, Eskeland R, Lettice LA, Hill AE, Boyle S, Grimes GR, Hill RE and Bickmore WA (2012) Anterior-posterior differences in HoxD chromatin topology in limb development. *Development* 139:3157–3167.
- Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman M V. and Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* 20:1377–1419.
- Xie QW (1997) A novel lipopolysaccharide-response element contributes to induction of nitric oxide synthase. *J Biol Chem* 272:14867–14872.
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
- Yu L and Zhang YP (2005) Evolutionary implications of multiple SINE insertions in an intronic region from diverse mammals. *Mamm Genome* 16:651–660.
- Zakany J, Zacchetti G and Duboule D (2007) Interactions between HOXD and Gli3 genes control the limb apical ectodermal ridge via Fgf10. *Dev Biol* 306:883–893.
- Zhang J, Nielsen R and Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479.
- Zheng Y, Sun Y, Yang Y, Zhang S, Xu T, Xin W, Wu S and Zhang X (2019) GATA3-dependent epigenetic upregulation of CCL21 is involved in the development of neuropathic pain induced by bortezomib. *Mol Pain* 15:174480691986329.
- Zhou X, Xu S, Yang Y, Zhou K and Yang G (2011) Phylogenomic analyses and improved resolution of Cetartiodactyla. *Mol Phylogenet Evol* 61:255–264.
- Zhou YZ, Slagle BL, Donehower LA, vanTuinen P, Ledbetter DH and Butel JS (1988) Structural

analysis of a hepatitis B virus genome integrated into chromosome 17p of a human hepatocellular carcinoma. *J Virol* 62:4224–4231.

Zurano JP, Magalhães FM, Asato AE, Silva G, Bidau CJ, Mesquita DO and Costa GC (2019) Cetartiodactyla: Updating a time-calibrated molecular phylogeny. *Mol Phylogenet Evol* 133:256–262.