



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR, EMBRIOLOGIA E GENÉTICA
CURSO DE LICENCIATURA EM CIÊNCIAS BIOLÓGICAS

Clarissa Callegaro Zottis

Filogenia e evolução molecular de genes envolvidos na aprendizagem vocal em cetáceos

Florianópolis

2024

Clarissa Callegaro Zottis

Filogenia e evolução molecular de genes envolvidos na aprendizagem vocal em cetáceos

Trabalho de Conclusão de Curso submetido ao curso de Graduação em Ciências Biológicas do Centro Ciências Biológicas da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Licenciada em Ciências Biológicas.

Orientador: Prof. Guilherme de Toledo e Silva, Dr.
Coorientador: Prof. Guilherme Renzo Rocha Brito, Dr.

Florianópolis

2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Zottis, Clarissa Callegaro

Filogenia e evolução molecular de genes envolvidos na
aprendizagem vocal em cetáceos / Clarissa Callegaro Zottis
; orientador, Guilherme de Toledo e Silva, coorientador,
Guilherme Renzo Rocha Brito, 2024.
54 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro de Ciências
Biológicas, Graduação em Ciências Biológicas, Florianópolis,
2024.

Inclui referências.

1. Ciências Biológicas. 2. Aprendizagem vocal. 3.
Filogenia. 4. Evolução molecular. 5. Cetáceos. I. Silva,
Guilherme de Toledo e. II. Brito, Guilherme Renzo Rocha.
III. Universidade Federal de Santa Catarina. Graduação em
Ciências Biológicas. IV. Título.

Clarissa Callegaro Zottis

Filogenia e evolução molecular de genes envolvidos na aprendizagem vocal em cetáceos

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Licenciada e aprovado em sua forma final pelo Curso Ciências Biológicas.

Sala CCB06, 05 de dezembro de 2024.

Coordenação do Curso

Banca examinadora

Prof. Guilherme de Toledo e Silva, Dr.
Orientador(a)

Prof.^a Luísa Damazio Rona Pitaluga, Dra.
Universidade Federal de Santa Catarina

Prof.^a Andrea Rita Marrero, Dra
Universidade Federal de Santa Catarina

Florianópolis, 2024

Este trabalho é dedicado a Universidade Pública, por criar oportunidades de ser e
fazer.

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador, Guilherme de Toledo, por me acolher em tempos de dúvidas e crises sobre meu caminho na biologia, pela paciência, por me ver e entender, e por me mostrar que há formas mais tranquilas e saudáveis de encarar a vida. Aprecio as conversas que tivemos. Foi incrível finalizar essa jornada tendo você como orientador. Obrigada!

Agradeço também ao meu companheiro de vida, Luiz, por todos os momentos bons, risadas, viagens e suporte incondicional durante nosso tempo juntos. Sua presença faz meus dias mais felizes. Te amo! Não posso deixar de agradecer aos nossos amados gatinhos, Canjica e Chuva, que diversas vezes me forneceram serotonina com seus ronrons, e Robin, nossa querida gatinha que faleceu em fevereiro deste ano, mas que estará eternamente em nossos corações.

Sou muito grata à minha família: à minha mãe, pelo amor e carinho incondicionais; ao meu pai, por sempre responder às minhas perguntas curiosas de criança e por me doar suas revistas velhas da Superinteressante. Se hoje permaneço curiosa, é graças a isso. À minha querida avó, por sempre me motivar nos estudos e a ingressar na universidade; às minhas tias, Giovanna, Márcia e Fátima, por serem minhas segundas mães e sempre acreditarem em mim; e aos meus irmãos, Sâmmmer, Nur e Kais, pela parceria de sempre.

Agradeço também às minhas amigas Giulia e Guli, pela época em que dividimos casa, pelo companheirismo e cumplicidade, pelas louças trocadas e roupas emprestadas, pelas risadas e séries assistidas. À minha amiga Bruna, minha dupla de estágio e trabalhos da faculdade, pelos cafés, fofocas e risadas. Essa árdua jornada pelo diploma foi mais leve por ter você trilhando junto.

Por fim, agradeço à Clarissa do passado, por ter tomado as decisões que me trouxeram até aqui e não ter desistido da Ilha e nem da Universidade. Valeu a pena!

*“Encarar a vida pela frente. Sempre.
Encarar a vida pela frente, e vê-la como ela é.
Por fim, entendê-la e amá-la pelo que ela é. E
depois deixá-la seguir. Sempre os anos entre nós,
sempre os anos. Sempre o amor. Sempre a razão.
Sempre o tempo. Sempre... As horas.”.*

Virginia Woolf (1882 – 1941)

RESUMO

Cetáceos, conhecidos por suas complexas comunicações vocais, utilizam a aprendizagem vocal para diversas funções sociais e ecológicas. Essa rara habilidade permite a modificação e produção de vocalizações através da imitação, e evoluiu independentemente em diferentes grupos de vertebrados. O objetivo deste estudo é investigar como genes possivelmente associados à aprendizagem vocal evoluíram na linhagem dos cetáceos e verificar se há regiões e sítios específicos mais suscetíveis à seleção natural. Para isso, realizamos análises filogenéticas para mapear a evolução desses genes e identificar potenciais padrões de evolução convergente relacionados a essa habilidade. Foram analisados sete genes: TSHZ3, ZNF536, GALC, ERBB4, TCF4, CDH4 e FOXP2. Os seis primeiros foram selecionados pela proximidade com regiões de cromatina aberta no córtex motor de morcegos, sugerindo uma possível relevância para a aprendizagem vocal. Já o gene FOXP2 é amplamente estudado devido à sua associação com mutações que impactam a fala. Para este estudo, conduzimos também uma análise de evolução molecular focada em seleção positiva. As análises filogenéticas mostraram que os seis primeiros genes seguem relações evolutivas consistentes com a topologia das espécies, sugerindo baixa convergência evolutiva nas sequências proteicas. Isso indica que a aprendizagem vocal pode estar mais ligada à expressão gênica do que à sequência proteica em si. Em contraste, o gene FOXP2 mostrou uma possível convergência evolutiva entre morcegos e cetáceos, com seis sítios sob seleção positiva, incluindo o sítio 185Q, com suporte significativo ($p < 0,05$). A presença de uma região POLY-Q nessa área sugere que essas variações podem ter desempenhado um papel na evolução da função do FOXP2, por sua influência na frequência de vocalização, pode provavelmente ter influenciado a aprendizagem vocal e outras adaptações fenotípicas.

Palavras-chave: vocalização, baleias, comportamento, bioinformática, genética

ABSTRACT

Cetaceans, known for their complex vocal communications, utilize vocal learning for various social and ecological functions. This rare ability allows the modification and production of vocalizations through imitation and has evolved independently in different groups of vertebrates. The aim of this study is to investigate how genes potentially associated with vocal learning have evolved in the cetacean lineage and to determine whether there are specific regions and sites more susceptible to natural selection. To achieve this, we performed phylogenetic analyses to map the evolution of these genes and identify potential patterns of convergent evolution related to this ability. Seven genes were analyzed: TSHZ3, ZNF536, GALC, ERBB4, TCF4, CDH4, and FOXP2. The first six were selected based on their proximity to open chromatin regions in the motor cortex of bats, suggesting possible relevance for vocal learning. FOXP2, on the other hand, is widely studied due to its association with mutations that affect speech. For this gene, we also conducted a molecular evolution analysis focusing on positive selection. The phylogenetic analyses revealed that the first six genes follow evolutionary relationships consistent with the species' topology, suggesting low evolutionary convergence in protein sequences. This indicates that vocal learning may be more associated with gene expression rather than protein sequence. In contrast, FOXP2 showed possible evolutionary convergence between bats and cetaceans, with six sites under positive selection, including site 185Q, with significant support ($p < 0.05$). The presence of a POLY-Q region in this area suggests that these variations may have played a role in the evolution of FOXP2 function, potentially influencing vocalization frequency and likely contributing to vocal learning and other phenotypic adaptations.

Keywords: vocalization, whales, behavior, bioinformatics, genetics

LISTA DE FIGURAS

Figura 1 – Árvore filogenética da família Cetacea.....	20
Figura 2 - Estrutura tridimensional predita da proteína codificada pelo gene FOXP2 em Orcinus orca, gerada pelo AlphaFold	25
Figura 3 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.mammal com suporte de <i>bootstrap</i> para 1000 réplicas no <i>software</i> IQ-Ta partir do alinhamento de aminoácidos da região transcrita do gene FOXP2.....	32
Figura 4 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.mammal+I+G4 com suporte de <i>bootstrap</i> para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene THSZ3.....	33
Figura 5 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.bird+G4 com suporte de <i>bootstrap</i> para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene ZNF536.....	34
Figura 6 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.plant+I+G4 com suporte de <i>bootstrap</i> para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene GALC	35
Figura 7 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.mammal+I+R2 com suporte de <i>bootstrap</i> para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene CDH4	36
Figura 8 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.insect+F+G4 com suporte de <i>bootstrap</i> para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene TCF4	37
Figura 9 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição JTT+R2 com suporte de <i>bootstrap</i> para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene ERBB4.....	39
Figura 10 - Frequência dos códons CAA e CAG e número de glutaminas consecutivas em espécies analisadas, com relação ao fenótipo de aprendizado vocal	45

LISTA DE TABELAS

Tabela 1 - Aprendizagem contextual x aprendizagem vocal.....	18
Tabela 2 - Descrição dos genes escolhidos	24
Tabela 3 - Modelos de substituição	28
Tabela 4 - Taxa de similaridade entre as sequências.....	40
Tabela 5 - Comparação entre modelos evolutivos para análise de seleção positiva	42

SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	DEFINIÇÕES E CONCEITUAÇÕES DE APRENDIZAGEM VOCAL	17
1.2	MAMÍFEROS COM APRENDIZAGEM VOCAL	18
1.2.1	Cetáceos	19
1.2.1.1	<i>Diferenças evolutivas e morfológicas.....</i>	<i>19</i>
1.2.2	Morcegos.....	22
1.2.3	Pinípedes.....	23
1.2.4	Humanos.....	23
1.3	GENES ANALISADOS	24
2	OBJETIVOS.....	26
2.1	OBJETIVO GERAL	26
2.2	OBJETIVOS ESPECÍFICOS	26
3	METODOLOGIA.....	26
3.1	COLETA DE SEQUÊNCIAS DE PROTEÍNAS	26
3.2	ALINHAMENTO DAS SEQUÊNCIAS.....	27
3.3	REMOÇÃO DE GAPS.....	27
3.4	ANÁLISE FILOGENÉTICA	28
3.5	ANÁLISE DE SELEÇÃO POSITIVA.....	29
4	RESULTADOS E DISCUSSÃO.....	30
4.1	FILOGENIA	31
4.2	EVOLUÇÃO MOLECULAR	41
4.3	TENDÊNCIA EVOLUTIVA <i>POLY-Q</i>	44
5	CONCLUSÃO.....	46
	REFERÊNCIAS.....	48
	APÊNDICE A – Tabela de espécies e fenótipos.....	52
	APÊNDICE B – Tabela com genes e respectivos códigos do NCBI.....	53

1 INTRODUÇÃO

A aprendizagem vocal é a habilidade que envolve a capacidade de ouvir, repetir e contextualizar vocalizações de indivíduos da mesma espécie ou de outra espécie (Janik *et al*, 2000, 2021; Fitch *et al*, 2000; Wirthlin *et al*, 2024). Pode ser definida também como a capacidade baseada na experiência de produzir sinais vocais novos por meio da imitação ou modificação de vocalizações existentes (Vernes *et al*, 2019; Tyack *et al*, 2010). Esse tipo de habilidade é uma forma de aprendizagem social que influencia diretamente a comunicação animal.

A ampla pesquisa sobre este tema foi predominantemente conduzida em aves, com uma atenção comparativamente menor dedicada aos mamíferos, e ainda menor em mamíferos marinhos. Embora alguns pássaros distantes (papagaios, pássaros canoros e beija-flores) sejam capazes de modificar sons associados à aprendizagem e exposição auditiva, capacidades comparáveis só foram recentemente identificadas em mamíferos (Reichmuth *et al*, 2014).

Para mamíferos marinhos, é fundamental considerar que mudanças ambientais severas, como a recolonização do ambiente marinho, impuseram a necessidade de desenvolvimento de mecanismos adaptativos específicos para assegurar o sucesso evolutivo. Os cetáceos, assim como os pinípedes, apresentam um elevado grau de controle voluntário sobre suas emissões sonoras, uma habilidade que provavelmente está correlacionada com as adaptações respiratórias que se adaptaram para a prática do mergulho (Reichmuth *et al*, 2014). A aprendizagem vocal pode também desempenhar um papel especialmente importante na manutenção de vocalizações distintivas individuais para mamíferos marinhos enquanto mergulham no mar (Tyack *et al*, 2010).

Embora a capacidade de vocalizações para comunicação seja comum na maioria dos animais, a habilidade de aprendizagem vocal é rara e evoluiu independentemente entre morcegos, cetáceos, pinípedes, aves e humanos (Janik *et al*, 2000; Wirthlin *et al*, 2024). A evolução do comportamento de aprendizagem vocal tem sido associada à evolução de características anatômicas do cérebro (Wirthlin *et al*, 2024).

Pesquisas sugerem que redes regulatórias compartilhadas para o comportamento vocal e social tendem a se adaptar de maneiras semelhantes quando uma linhagem evolui para

o comportamento de aprendizagem vocal¹. Um estudo identificou um conjunto de proteínas conservadas em aprendizes vocais e descobriu uma região cortical motora vocal no morcego egípcio da fruta (*Rousettus aegyptiacus*). Essas descobertas auxiliaram na identificação de elementos cis regulatórios ativos no córtex motor de aprendizes vocais, destacando as bases moleculares subjacentes à aprendizagem vocal (Wirthlin, 2024).

Entre os 200 genes identificados na literatura científica, o presente estudo focou na análise da evolução molecular de seis genes específicos: *teashirt zinc finger homeobox 3* (TSHZ3), *Zinc Finger Protein 536* (ZNF536), *Galactocerebrosidase* (GALC), *Erb-B2 Receptor Tyrosine Kinase 4* (ERBB4), *Transcription Factor 4* (TCF4), *Cadherin 4* (CDH4). E em adição, o gene *forkhead box P2* (FOXP2).

Este trabalho visa investigar como a evolução molecular desses genes relacionados à aprendizagem vocal em cetáceos ocorreu e como essas mudanças genéticas contribuem para a compreensão da história evolutiva e do funcionamento dos mecanismos de aprendizagem vocal nesses animais.

A justificativa para estudar a aprendizagem vocal reside em sua importância para entender a evolução da comunicação, da cognição e da linguagem humana, além dos mecanismos neurais envolvidos na aquisição de habilidades complexas. Esse fator é relevante para a saúde humana e animal, pois pode ajudar a esclarecer questões de comunicação e abrir caminhos para pesquisas em neurociência e em transtornos que afetam habilidades comunicativas. Como uma característica rara, presente em apenas alguns grupos de vertebrados, a aprendizagem vocal destaca-se por sua singularidade e valor científico. Estudos comparativos sobre essa habilidade podem revelar quais mecanismos estão ligados às pressões evolutivas que moldam a função neurobiológica e anatômica de diferentes espécies.

1.1 DEFINIÇÕES E CONCEITUAÇÕES DE APRENDIZAGEM VOCAL

Como mencionado anteriormente, a limitada atenção dedicada à aprendizagem vocal em mamíferos, em comparação com as aves, torna necessário esclarecer e reforçar os conceitos e definições do que caracteriza a aprendizagem vocal. Segundo Janik (2000) em aves “a ocorrência de aprendizagem vocal é tão óbvia que muito pouco esforço tem sido

¹ Embora já existam indícios sobre a importância das regiões regulatórias, focar nas regiões codificantes foi uma escolha prática e viável dentro dos objetivos deste trabalho, além da relevância que essas regiões têm para as investigações de seleção natural. Esse enfoque cria uma boa base, e a análise das regiões regulatórias aparece como um possível próximo passo natural para aprofundar o entendimento das bases genéticas da aprendizagem vocal

focado em definir o que constitui a aprendizagem vocal e classificando as formas que ela pode levar”. Desta forma, Janik elucida as definições de aprendizagens sociais que influenciam comunicação vocal:

Tabela 1 - Aprendizagem contextual x aprendizagem vocal

Forma de Aprendizado	Definição
<i>Aprendizagem contextual</i>	Um sinal existente está associado a um novo contexto como resultado da experiência com o uso de sinais por outros indivíduos
• <i>Aprendizagem de uso</i>	Um sinal existente é produzido em um novo contexto como resultado da experiência com o uso de sinais por outros indivíduos.
• <i>Aprendizagem de compreensão</i>	Um receptor extrai um novo significado de um sinal como resultado da experiência com o uso dos sinais por outros indivíduos.
<i>Aprendizagem de produção vocal</i>	Os sinais são modificados na forma como resultado da experiência com os de outros indivíduos. Isso pode levar a sinais que são semelhantes ou diferentes do modelo

Fonte: Adaptado de Janik, 2000

Com estas definições, Janik conclui que se um animal puder aprender a dar um sinal de alarme na ausência de um predador (aprendizado de uso), ou puder copiar um sinal de alarme de outra espécie (aprendizado de produção), ele poderá se beneficiar de um maior acesso a alimentos abandonados por outros indivíduos em resposta a essa chamada.

Desta forma, a aprendizagem de produção é a aprendizagem vocal, e pode ter evoluído para facilitar o reconhecimento individual ou de grupo (Janik e Slater, 1997, Rendell *et al*, 2019). Compreendendo as definições da aprendizagem vocal, ainda é preciso determinar quais os mecanismos estão envolvidos na produção de um novo sinal.

É importante observar que, mesmo em aprendizes vocais sofisticados, parte do repertório vocal pode ser inata — por exemplo, os humanos mantêm vocalizações inatas como o choro e o riso (Vernes *et al*, 2019).

1.2 MAMÍFEROS COM APRENDIZAGEM VOCAL

1.2.1 Cetáceos

Os cetáceos, um grupo diversificado de mamíferos aquáticos que inclui baleias, golfinhos e botos, exibe uma variedade de comportamentos complexos. A comunicação sofisticada dos cetáceos é um reflexo de sua alta inteligência e complexidade social, e a aprendizagem vocal é um componente crucial dessa comunicação (Janik *et al*, 2021). As dinâmicas comunicativas dos cetáceos envolvem não apenas diferenças entre espécies, mas também variações regionais e dialetais dentro de uma mesma espécie, transmitidas culturalmente (Rendell *et al*, 2012). Podemos aqui traçar um paralelo aos sotaques linguísticos nos humanos.

A adaptação dos cetáceos à vida aquática exerceu pressões seletivas significativas em diversos aspectos: anatômicos, comportamentais, fisiológicos e genéticos. Um padrão semelhante, observado em outros mamíferos que passaram a ocupar ambientes aquáticos novamente ao longo de sua evolução, são as "perdas genéticas". Um exemplo é a desativação do gene MMP12 em cetáceos e manatis (peixes-boi), responsável pela degradação da proteína elastina. Essa desativação foi provavelmente favorecida pela seleção natural, contribuindo para a evolução de pulmões altamente elásticos em ambos os grupos, permitindo-lhes exalar e inalar rapidamente cerca de 90% do volume de seus pulmões quando emergem na superfície (Huelsmann, 2019).

Nesse contexto, a produção de novos e distintos sons emerge como uma adaptação importante. Modular diferentes sons permite compensar as alterações nas características vocais causadas pela pressão da água no trato vocal cheio de ar, especialmente durante mergulhos em diferentes profundidades (Janik e Slater, 1997). Além disso, em ambientes marinhos, o som viaja muito mais rápido e mais longe do que outros sinais, tornando a acústica a principal modalidade sensorial que os cetáceos utilizam ao longo da vida (Hersh *et al*, 2024). É possível que essas pressões seletivas estejam relacionadas ao surgimento de adaptações moleculares no desenvolvimento da aprendizagem vocal.

Além disso, cetáceos alterando seus sinais em resposta ao ruído antropogênico não é incomum e sugere que o aprendizado vocal pode ajudar os animais a lidarem com a poluição sonora (Lesage *et al*, 1999; Janik *et al*, 2014)

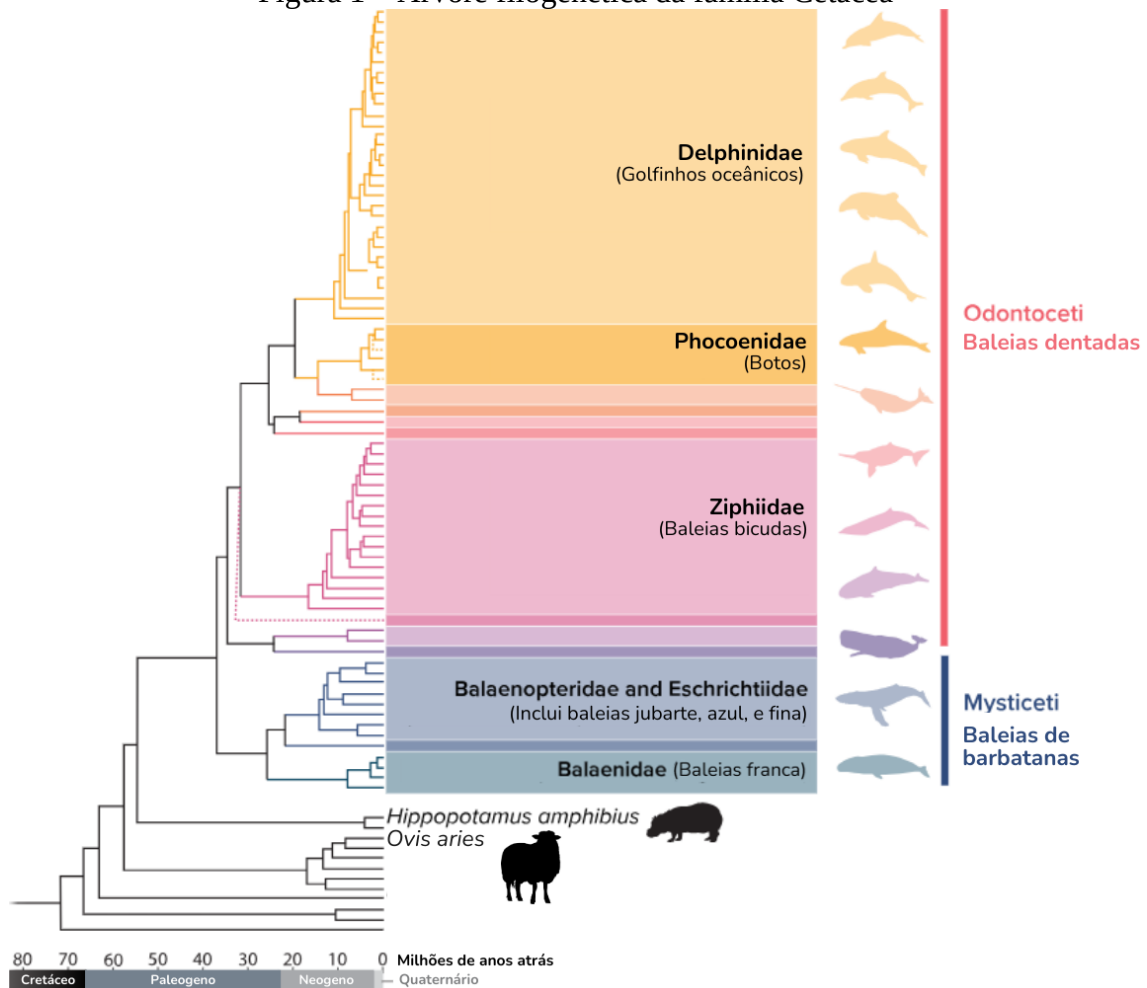
1.2.1.1 Diferenças evolutivas e morfológicas

A capacidade dos cetáceos de aprender e modificar seus sons vocais em resposta ao ambiente possui funções diferentes para os grupos. As baleias jubarte, por exemplo, utilizam

essa habilidade em exibições de canto durante a seleção sexual, enquanto os golfinhos empregam sinais aprendidos para reconhecimento individual e manutenção de relações sociais (Janik *et al*, 2014).

Há também uma diferenciação anatômica e evolutiva entre os grupos de odontocetos, que inclui cetáceos dentados como golfinhos, orcas e cachalotes; e de mysticetos que abrangem as baleias que possuem barbatanas no lugar de dentes, como baleia jubarte, baleia azul, entre outras (veja Figura 1).

Figura 1 – Árvore filogenética da família Cetacea



Fonte: adaptado de Cabrera *et al.*, 2021

Os mysticetos, como as baleias jubarte (*Megaptera novaeangliae*), possuem uma anatomia adaptada para vocalizações e audição de baixa frequência (Hersh *et al*, 2024).

Como descreve Annalisa *et al* (2005), a perda de audição de altas frequências nos mysticetos ocorreu no Oligoceno e está relacionada à transição para a alimentação por predação massiva usando as barbatanas. Essas barbatanas, compostas por filamentos queratinizados ajudam a reter as pequenas presas, como os krills, enquanto a água é expelida.

Essa mudança evolutiva ocorreu sob pressões ambientais, como mudanças no zooplâncton. O registro fóssil mostra que misticetos com e sem barbatanas coexistiram, e a transição para a filtração foi associada a um aumento no tamanho corpóreo, especialmente da cabeça, permitindo uma maior capacidade de filtração (Davis *et al*, 2019).

Misticetos produzem sons por meio da vibração das pregas vocais localizadas na laringe (Hersh *et al*, 2024). As vocalizações dos misticetos são caracterizadas por uma frequência fundamental baixa, geralmente entre 7 a 20 Hz, com harmônicos e uma duração que pode variar de alguns segundos a minutos (Hersh *et al*, 2024). Essas vocalizações não são apenas simples chamadas, mas formas complexas de comunicação que podem incluir canções longas e repetitivas, com uma estrutura hierárquica: as canções são formadas por unidades repetidas (chamadas), que se organizam em frases, temas e, por fim, em canções completas (Hersh *et al*, 2024).

A função das canções nos misticetos está ligada ao comportamento de cortejo, onde, em muitas espécies, como as baleias jubarte, os machos cantam para atrair fêmeas, semelhante ao comportamento de canto nas aves (Hersh *et al*, 2024). A aprendizagem vocal é um aspecto fundamental no desenvolvimento dessas canções. Em algumas espécies, as canções evoluem ao longo do tempo por meio de aprendizado social, como observado nas baleias jubarte, que modificam suas canções com base nas interações com outras baleias, o que indica uma transmissão cultural da vocalização (Hersh *et al*, 2024). Além disso, os misticetos também produzem vocalizações não musicais, como chamadas sociais e sons de forrageamento, embora estas tenham sido menos estudadas (Hersh *et al*, 2024).

Os odontocetos, como golfinhos, orcas e belugas, possuem uma anatomia adaptada para a produção de vocalizações de alta frequência, uma característica distinta em relação aos misticetos. Os odontocetos produzem vocalizações por meio dos lábios fônicos, localizados na cabeça, gerando sons como cliques, sons pulsados e assobios (May-Collado *et al*, 2007; Hersh *et al*, 2024). Os cliques são sons de banda larga, curtos e impulsivos, usados para ecolocalização, enquanto os assobios são sinais de banda estreita e modulados por frequência, geralmente utilizados em contextos sociais e de comunicação (Tyack, 2020; Janik *et al*, 2000). Tanto a recepção e quanto a modulação desses sons são feitas pelo melão, uma estrutura de tecido gorduroso na parte frontal da cabeça, que ajuda a direcionar e amplificar as vocalizações para a ecolocalização e percepção precisa dos sons (May-Collado *et al*, 2007; Tyack *et al*, 2010).

Estudos mostraram que várias espécies de odontocetos, como as orcas e belugas, têm uma notável capacidade de aprender novas vocalizações, principalmente a partir de interações

sociais. Por exemplo, estudos em belugas revelaram que filhotes começam a produzir as vocalizações de seus pais apenas após esses pais serem introduzidos ao grupo, sugerindo que os filhotes aprendem vocalizações de sons específicos através de interação social (Janik *et al*, 2014). Além disso, as orcas demonstraram capacidade de modificar seus repertórios vocais por meio de aprendizagem social. Elas podem imitar vocalizações de outros membros do grupo e até de humanos, como observado em estudos de orcas em cativeiro (Janik *et al*, 2014).

De acordo com Brenowitz (2023), tanto a ecolocalização quanto a aprendizagem vocal exigem controle voluntário sobre a produção vocal e modulação em resposta ao *feedback* auditivo, sendo a ecolocalização uma característica ancestral de cetáceos e morcegos, possivelmente evoluindo junto à aprendizagem vocal, de forma que os circuitos neurais responsáveis pela ecolocalização podem ter facilitado o desenvolvimento do controle neural necessário para a aprendizagem vocal.

Esse ponto é relevante, pois sugere que a aprendizagem vocal pode estar relacionada à ecolocalização (outra característica rara, encontrada apenas em morcegos e odontocetos), indicando uma possível interconexão evolutiva entre esses dois processos.

1.2.2 Morcegos

Assim como em cetáceos dentados (odontocetos), os morcegos também desenvolveram a capacidade de ecolocalização, que é frequente em animais com alta capacidade auditiva e de controle vocal. Essas vocalizações são os “chamados de ecolocalização” (*echolocation calls*) que variam entre as espécies, bem como chamados sociais usados para reuniões entre pais e filhotes, defesa territorial ou manutenção da integração do grupo (Vernes, 2019). Estes animais utilizam a ecolocalização para navegar e caçar insetos no escuro, mas também empregam vocalizações para a comunicação social (Knörnschild, 2014).

Morcegos jovens aprendem a ecolocalizar imitando os adultos, indicando um processo de aprendizagem vocal. Estudos observacionais e experimentais com morcegos-pico-de-lança pálidos, uma espécie tropical do Novo Mundo, demonstram de forma convincente que os filhotes adaptam progressivamente seus chamados de isolamento para corresponder aos chamados diretivos maternos (Esser *et al*, 1989), evidenciando a aprendizagem vocal.

As famílias mais conhecidas e documentadas por aprendizagem vocal são Emballonuridae, Phyllostomidae, Pteropodidae, Rhinolophidae, Molossidae, Vespertilionidae, Mormoopidae (Knörnschild *et al*, 2014; Vernes *et al*, 2020)

Os morcegos-egípcios-da-fruta, *Rousettus aegyptiacus*, uma espécie da família Pteropodidae, utilizam cliques da língua para ecolocalização e possuem um repertório variado de chamados. Eles emitem diferentes vocalizações durante interações agonísticas e conseguem ajustar suas vocalizações de acordo com as que escutam. (Prat *et al*, 2017, Harten *et al*, 2018, Vernes *et al*, 2020). As espécies *Rhinolophus ferrumequinum* e *Hipposideros terasensis* também demonstraram aprendizagem vocal em experimentos similares (Vernes *et al*, 2020)

Assim como em orcas, os morcegos também apresentam variações em seus dialetos. Existe uma hipótese, chamada de 'deriva cultural', que sugere que colônias geograficamente isoladas de morcegos de frequência constante podem diferir em suas frequências de ressonância por acaso (Vernes *et al*, 2019; Esser *et al*, 1989).

1.2.3 Pinípedes

Os pinípedes, que incluem focas, leões-marinhos e morsas, também demonstram capacidades de aprendizagem vocal. Há registros de aprendizagem vocal em espécies desse grupo tanto em condições de cativeiro quanto em seus habitats naturais (Duengen *et al*, 2023).

Para a produção de sons fora da água, a laringe está comumente associada, e para produção de sons subaquáticos, possuem os chamados típicos sem liberação de ar (Reichmuth *et al*, 2014). Estudos sugerem que focas também utilizam sons semelhantes a cliques para localizar presas, embora as evidências para isso sejam principalmente indiretas (Brenowitz *et al*, 2023).

Além disso, as morsas e muitas focas possuem tecidos orais especialmente carnudos e lábios utilizados para alimentação por sucção subaquática (e, em alguns casos, para amamentação subaquática), o que aumenta ainda mais a diversidade de suas emissões sonoras através da vibração ou manipulação das estruturas supralaríngeas (Reichmuth *et al*, 2014).

1.2.4 Humanos

Os humanos são os únicos primatas conhecidos por sua capacidade de aprendizagem vocal. Contrariamente ao que se poderia esperar, outros primatas não demonstraram essa habilidade. Embora os chimpanzés possam aprender linguagem de sinais, evidenciando a aplicação do conceito de aprendizagem contextual, eles não conseguem vocalizar palavras.

Além disso, a capacidade de ecolocalização já foi documentada em humanos. Algumas pessoas cegas foram capazes de "perceber" a distância de objetos ou obstáculos com base na audição de sons produzidos, como cliques (Thaler *et al*, 2016).

Em humanos, o autismo frequentemente está associado a atrasos na fala e diferenças no comportamento social, ambos os quais podem estar relacionados ao traço evolutivo da capacidade de aprendizagem vocal. Vários genes associados ao autismo estavam próximos dos ortólogos humanos dos elementos regulatórios associados à aprendizagem vocal (Wirthlin *et al*, 2024).

O gene mais proeminente envolvido em distúrbios da fala humana é o FOXP2. A interrupção desse gene leva à apraxia de fala infantil (CAS), em que a fala das crianças afetadas é severamente prejudicada (Vernes *et al*, 2020; Lai *et al*, 2001).

1.3 GENES ANALISADOS

Os genes TSHZ3, ZNF536, GALC, ERBB4, TCF4, CDH4 (Tabela 2) foram recentemente foco de estudos pois são proximais a regiões de cromatina aberta do córtex motor de morcegos da espécie *Rousettus aegyptiacus* (Wirthlin *et al*, 2024). Essas regiões de cromatina aberta, podem estar diretamente relacionadas a fatores epigenéticos de aprendizagem vocal.

Tabela 2 - Descrição dos genes escolhidos

Abreviação	Tradução para o Português	Código NCBI
TSHZ3	Homeobox de Dedo de Zinco Teashirt 3	57616
ZNF536	Proteína de Dedo de Zinco 536	9745
GALC	Galactocerebrosidase	2581
ERBB4	Receptor Quinase Erb-B2 Tipo Tirosina 4	2066
TCF4	Fator de Transcrição 4	6925
CDH4	Caderina 4	1002
FOXP2	Caixa Forkhead P2	93986

Fonte: autora

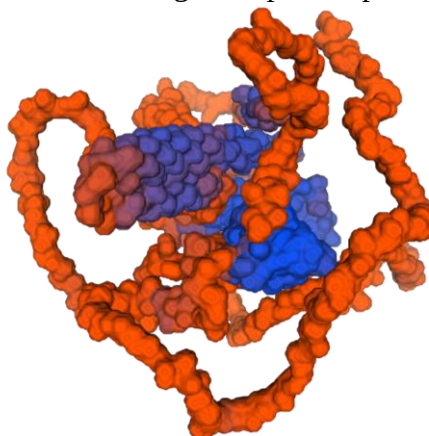
O gene TSHZ3 está envolvido na formação de circuitos corticostriatais, que desempenham um papel central no comportamento de aprendizagem vocal em mamíferos e aves, e sua interrupção na população humana está associada a uma forma de autismo que inclui aquisição de fala atrasada ou interrompida (Kaplow *et al*, 2023; Chabbert *et al*, 2019; Caubit *et al*, 2016; Jarvis *et al*, 2004).

Em humanos, mutações no TCF4 causam uma desordem do espectro autista conhecida como síndrome de Pitt–Hopkins, caracterizada por uma variedade de fenótipos anômalos, incluindo deficiência intelectual severa, ausência de fala, desenvolvimento cognitivo e motor atrasado, e traços dismórficos (Teixeira *et al*, 2021).

Estudos anteriores exploraram um possível papel das caderinas no controle vocal (Matsunaga *et al*, 2011), e sugerem que a regulação negativa de CDH4 influencia esse processo (Lovell *et al*, 2018).

Já o gene FOXP2 (Figura 2) é conhecido pela sua função no desenvolvimento adequado das regiões da fala e da linguagem do cérebro durante a embriogênese e pode estar envolvido em uma variedade de vias biológicas e cascatas que podem, em última instância, influenciar o desenvolvimento da linguagem (Refseq, 2010). Mutações neste gene causam distúrbio de fala e linguagem 1 (SPCH1), também conhecido como distúrbio autossômico dominante de fala e linguagem com dispraxia orofacial (Morgan *et al*, 2016).

Figura 2 - Estrutura tridimensional predita da proteína codificada pelo gene FOXP2 em *Orcinus orca*, gerada pelo AlphaFold



Fonte: elaborado pela autora.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Analisar a filogenia e evolução molecular de genes candidatos envolvidos na aprendizagem vocal entre espécies de cetáceos, e investigar divergências ou convergências evolutivas entre esses genes.

2.2 OBJETIVOS ESPECÍFICOS

- Obter sequências de nucleotídeos e aminoácidos de TSHZ3, ZNF536, ERB4, GALC, TCF4, CDH4 das espécies selecionadas e realizar alinhamento múltiplo para cada conjunto obtido;
- Analisar a evolução na sequência de proteínas associado ao comportamento de aprendizagem vocal nos genes TSHZ3, ZNF536, ERB4, GALC, TCF4, CDH4.
- Realizar análise filogenética usando métodos de máxima verossimilhança para cada gene.
- Realizar a análise de seleção positiva para o gene FOXP2.

3 METODOLOGIA

3.1 COLETA DE SEQUÊNCIAS DE PROTEÍNAS

A obtenção das sequências gênicas foi realizada utilizando o banco de dados *GenBank*, que é acessível publicamente através do *National Center for Biotechnology Information* (NCBI), acessado em maio de 2024. Inicialmente, foram localizadas as sequências correspondentes aos genes de interesse, como TSHZ3, ZNF536, GALC, ERBB4, TCF4, CDH4 e FOXP2.

Em seguida, utilizou-se a ferramenta *NCBI Orthologs* (<https://www.ncbi.nlm.nih.gov/gene/93986/ortholog/?scope=7742&term=FOXP2>), disponível no próprio site do NCBI. Esta ferramenta foi fundamental para a identificação e seleção de ortólogos dos genes de interesse em diferentes espécies. A *NCBI Orthologs* fornece uma comparação entre genes de diversas espécies, facilitando a escolha das

sequências ortólogas para as espécies foco da pesquisa, facilitando a análise comparativa e filogenética subsequente.

Foram selecionadas todas as espécies de cetáceos disponíveis na plataforma *NCBI Orthologs*, totalizando 18 táxons. Além disso, foram incluídos táxons de grupos externos, abrangendo seis espécies de aves, seis de morcegos, oito de carnívoros (incluindo cães, gatos e pinípedes), três primatas (incluindo humanos), além de um representante de cada um dos seguintes animais: cobra-cega, hipopótamo, peixe-boi, ovelha e marmota. No total, foram selecionadas 48 espécies, das quais 33 são aprendizes vocais e 13 não são aprendizes vocais. A tabela completa com a lista de espécies pode ser visualizada no Apêndice 1.

3.2 ALINHAMENTO DAS SEQUÊNCIAS

As sequências gênicas obtidas foram submetidas a um processo de alinhamento múltiplo utilizando a ferramenta computacional *COBALT (Constraint-Based Alignment Tool)* (Papadopoulos *et al*, 2007), disponível no *NCBI Orthologs* (acessado em maio de 2024). Esta etapa permitiu a comparação das sequências, identificando regiões conservadas e variáveis ao longo dos genes. As regiões conservadas são aquelas que mantêm uma similaridade alta entre as diferentes sequências, enquanto as regiões variáveis apresentam variações entre as sequências, o que pode indicar divergência ou convergência evolutiva.

3.3 REMOÇÃO DE GAPS

Para garantir a precisão e a confiabilidade dos nossos dados filogenéticos, optamos por adotar uma abordagem conservadora na remoção de gaps. A exclusão completa dos gaps é uma estratégia que visa minimizar a introdução de artefatos na análise filogenética que poderiam resultar de alinhamentos imprecisos. Para implementar essa abordagem, utilizamos o *software Gblocks* (Dereeper *et al*, 2010; Dereeper *et al*, 2008; Castresana *et al*, 2000). O *software* está disponível na versão *online* para navegadores (http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=gblocks), facilitando o acesso e a aplicação do método sem a necessidade de instalação local. Utilizamos a versão *online* (acessada em maio de 2024, através do Phylogeny.fr) do *Gblocks* para garantir que o processo de limpeza fosse eficiente e compatível com os padrões mais recentes de análise de dados filogenéticos. Utilizamos os parâmetros padrão do *GBlocks* (sem itens marcados) para os alinhamentos. No entanto, quando os alinhamentos resultantes mantinham menos de 40% da

quantidade original de aminoácidos ou apresentaram menos de 200 aminoácidos por sequência, aplicamos uma seleção menos rigorosa, ajustando dois parâmetros: permitir blocos finais menores e a permitir inclusão de posições com lacunas nos blocos finais. Caso contrário, seria estatisticamente inviável realizar uma filogenia com sequências tão curtas.

3.4 ANÁLISE FILOGENÉTICA

Para a realização das análises filogenéticas dos alinhamentos, utilizamos o *software* IQ-TREE (Nguyen *et al*, 2015), que é uma ferramenta avançada para inferência de árvores filogenéticas. O IQ-TREE é conhecido por sua capacidade de realizar análises robustas e eficientes, oferecendo uma variedade de modelos de substituição de métodos de inferência. A escolha do IQ-TREE permitiu a aplicação de métodos estatísticos modernos para avaliar a confiabilidade das árvores filogenéticas e ajustar os modelos de substituição de acordo com a complexidade dos dados.

A primeira etapa da análise filogenética envolve a identificação do melhor modelo de substituição para o alinhamento de aminoácidos. Para isso, utilizou-se o comando no prompt do IQ-TREE:

[iqtree] -s [arquivo] -m MFP

O parâmetro MFP, referente ao pacote *ModelFinder* (Kalyaanamoorthy *et al*, 2017), facilita a seleção do modelo de melhor ajuste (*best-fit model*) para análises filogenéticas, aprimorando a precisão das estimativas evolutivas.

Tabela 3 - Modelos de substituição

Gene ID	Nome completo	Modelo de melhor ajuste
TSHZ3	teashirt zinc finger homeobox 3	Q.mammal+I+G4
ZNF536	zinc finger protein 536	Q.bird+G4
GALC	galactosylceramidase	Q.plant+I+G4
ERBB4	erb-b2 receptor tyrosine kinase 4	JTT+R2
TCF4	transcription factor 4	Q.insect+F+G4
CDH4	cadherin 4	Q.mammal+I+R2
FOXP2	forkhead box P2	Q.mammal

Fonte: autora

Após determinar o modelo de substituição mais adequado, o próximo passo foi realizar o processamento filogenético com *bootstrap* de 1000 réplicas. Para essa etapa, o comando utilizado foi:

[iqtree] -s [arquivo] -m [bestfitmodel] -bb 1000 -redo

Os arquivos de alinhamento originalmente recebidos apresentavam nomes extensos e carregados de informações adicionais, o que tornava a organização e manipulação dos arquivos um processo trabalhoso. Para simplificar e automatizar essa tarefa, foram utilizados scripts em *Python* para renomear os arquivos, reduzindo os nomes apenas ao nome da espécie correspondente. Além disso, foram desenvolvidos dois outros scripts: um para converter os arquivos para o formato “.tree” e outro para organizar os resultados em pastas específicas. Todos os scripts utilizados estão disponíveis no repositório GitHub (https://github.com/ifoxy404/TCC_dados.git), facilitando a reprodução e o acesso aos métodos empregados neste trabalho.

Após a construção das árvores filogenéticas com o IQ-TREE, utilizamos o *software ChiPlot* (Xie *et al*, 2023) para a visualização e edição das árvores.

3.5 ANÁLISE DE SELEÇÃO POSITIVA

Para conduzir a análise de seleção positiva, a sequência nucleotídica do gene FOXP2 coletada anteriormente, foi submetida à busca por *open reading frames* (ORFs)² utilizando o *software mEmboss* (Rice *et al*, 2000), a fim de garantir que seus códons fossem alinhados corretamente.

O alinhamento das sequências nucleotídicas e suas traduções em aminoácidos foi realizado utilizando o *TranslatorX* (Abascal *et al*, 2010) na versão disponível para navegadores, garantindo a correspondência adequada entre as substituições nas sequências de códons e aminoácidos. Para garantir um alinhamento preciso de códons e evitar incoerências, é recomendável realizar o alinhamento de nucleotídeos e proteínas de forma simultânea.

Para a etapa de obtenção do alinhamento de códons, utilizamos o *software PAL2NAL* (Suyama *et al*, 2006) disponível para navegadores, que converte alinhamentos de proteínas e

² As ORFs (*Open Reading Frames* ou "Frames de Leitura Abertos") são sequências de nucleotídeos em um DNA ou RNA que podem ser traduzidas em uma proteína. Uma ORF começa com um códon de iniciação (geralmente o códon AUG, que codifica a metionina) e termina com um códon de terminação (como UAA, UAG ou UGA)

nucleotídeos em um formato de alinhamento de códons. Esse processo é fundamental, pois permite a análise detalhada da correspondência entre as sequências de nucleotídeos e as respectivas traduções em aminoácidos. Além disso, o *PAL2NAL* é útil para identificar possíveis erros nos alinhamentos de aminoácidos e nucleotídeos, assegurando a integridade dos dados. Na execução deste procedimento, foram utilizados os parâmetros padrão do *software*.

A remoção dos gaps foi realizada utilizando o programa *BioEdit* 7.2 (Hall *et al*, 2011), mantendo apenas as colunas em que algumas espécies apresentaram gaps, enquanto as demais continham nucleotídeos.

Em seguida, construímos uma árvore filogenética a partir das sequências alinhadas, utilizando o *software* RAxML 8.0 (Stamatakis *et al*, 2014). Os parâmetros utilizados para a construção da árvore estão disponíveis na pasta do GitHub deste projeto, acompanhados de descrições e explicações detalhadas. A árvore não foi enraizada, e então exportada para o formato *Newick* com *FigTree*.

Para testar a presença de seleção positiva, foi utilizado o *software* CODEML, do pacote PAML 4.7 (Yang, 1997). Foram testados diferentes modelos evolutivos, como M0, M1, M2, M7 e M8, comparando a taxa de substituições não sinônimas (dN) e sinônimas (dS), o que indicaria a presença de seleção positiva em certas regiões do gene. A comparação entre os modelos foi feita por meio de testes de verossimilhança ($2\Delta\ln L$), sendo a análise entre os modelos M1 vs M2 e M7 vs M8 determinante para identificar sítios potencialmente sob seleção positiva, seleção neutra ou seleção purificadora.

4 RESULTADOS E DISCUSSÃO

A análise filogenética de um gene oferece uma visão detalhada das relações evolutivas entre as sequências gênicas, permitindo identificar padrões de similaridade que podem não coincidir com a filogenia das espécies. Quando essas discrepâncias ocorrem, uma hipótese plausível é a presença de convergências evolutivas, onde diferentes espécies desenvolvem características semelhantes de forma independente devido a pressões seletivas similares. Essas convergências podem surgir como resultado de funções adaptativas específicas, levando à preservação de sequências gênicas semelhantes, apesar da divergência evolutiva entre as espécies.

Esse tipo de análise no presente trabalho pode ajudar a compreender a complexidade dos processos evolutivos relacionados à aprendizagem vocal e o papel que fatores como a seleção natural a nível molecular desempenham na evolução dessa característica tão rara.

4.1 FILOGENIA

Na árvore filogenética do gene FOXP2 (Figura 3), observa-se que todas as espécies de cetáceos estão agrupadas em clados com suporte de *bootstrap* variando de moderado a alto, o que confere uma razoável confiabilidade à representação das relações evolutivas entre esses táxons. O nó ancestral mais basal dos cetáceos apresenta um valor de *bootstrap* de 68, sugerindo a existência de um ancestral comum entre cetáceos e um clado específico de morcegos, composto por *Miniopterus natalensis* e *Myotis davidii*. Essa aproximação sugere uma relação filogenética ou talvez a presença de características moleculares compartilhadas entre esses grupos, e uma possível convergência evolutiva.

Um nó altamente suportado (*bootstrap* de 95) marca uma divergência significativa entre *Balaenoptera musculus* (baleia-azul) e os demais cetáceos, destacando essa espécie como um grupo isolado dentro dos misticetos e sugerindo uma diferenciação molecular relevante para o gene FOXP2. Logo após, observa-se um clado de cachalotes (*Kogia breviceps* e *Physeter catadon*) com suporte de 80, que aparece como grupo irmão de um clado maior contendo os demais cetáceos, com um suporte de 59. Esse arranjo sugere uma relação evolutiva mais próxima entre os cachalotes e distingue-os dos outros cetáceos restantes. Odontocetos e misticetos estão agrupados sem grandes distâncias filogenéticas entre seus subgrupos, evidenciando uma relativa conservação do gene FOXP2 entre eles. Essa proximidade entre os subgrupos indica poucas diferenças evolutivas a nível molecular para FOXP2, sugerindo que esse gene pode ter mantido funções semelhantes ao longo da evolução dos cetáceos.

As espécies de morcegos *Pteropus vampyrus* e *Rousettus aegyptiacus* formaram um grupo irmão com a ovelha (*Ovis aries*). A proximidade deste clado com os cetáceos é sustentada por um nó com valor de *bootstrap* de 40, indicando um suporte estatístico moderado. *Hipposideros armiger* e *Rhinolophus ferrumequinum* em vez de se agruparem próximas aos outros quirópteros, foram associadas aos carnívoros, a humanos (*Homo sapiens*) e à marmota (*Marmota flaviventris*), sugerindo uma divergência filogenética em relação ao restante dos morcegos.

A proximidade entre *Ovis aries* e os cetáceos é coerente com a filogenia conhecida desses grupos, que indica uma ancestralidade comum entre artiodátilos e cetáceos (veja Figura 1), conforme corroborado pelo agrupamento observado. Essa relação sugere que o gene FOXP2 pode apresentar padrões de conservação molecular que refletem essa proximidade filogenética entre os táxons analisados.

Por fim, o grupo de aves permaneceu significativamente distante como grupo externo, logo após o táxon da cobra-cega (*Rhinatrema bivittatum*).

Figura 3 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.mammal com suporte de *bootstrap* para 1000 réplicas no *software* IQ-Ta partir do alinhamento de aminoácidos da região transcrita do gene FOXP2.

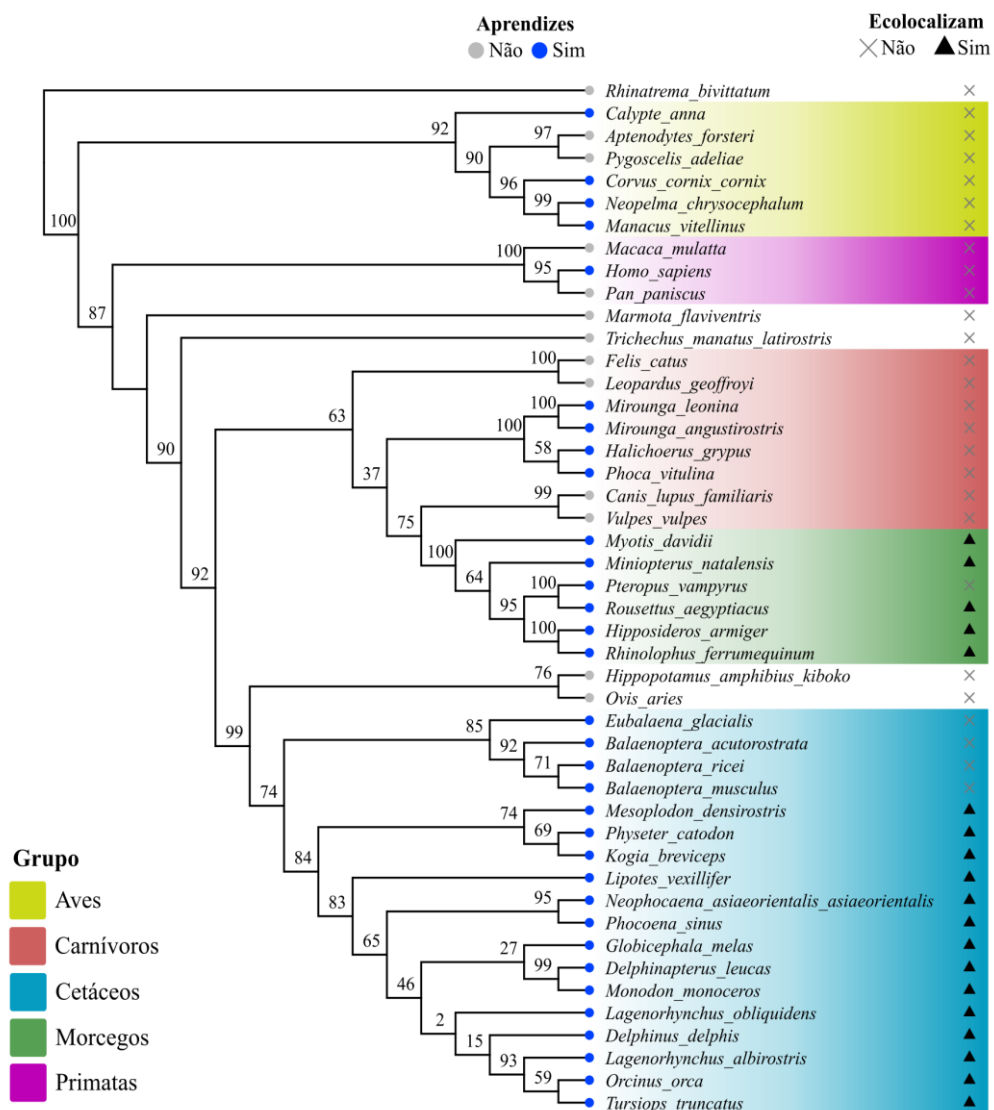


Fonte: autora

Na árvore filogenética do gene TSHZ3 (Figura 4), a maioria dos valores de *bootstrap* é superior a 70. Os cetáceos formam um grupo monofilético bem suportado (bb 99), com

nítidas diferenças entre odontocetos e mysticetos. Subgrupos, como as baleias cachalotes, destacam-se pela distinção significativa em relação aos demais, apresentando um valor de *bootstrap* de 74. Nos outros grupos, nota-se a proximidade entre aves e primatas, e *Rhinatrema bivittatum* encontra-se como esperado, sendo o grupo mais externo, dada sua distância filogenética em relação às demais espécies. Os morcegos e pinípedes também formam grupos monofiléticos, enquanto os carnívoros, que incluem pinípedes, canídeos e felídeos, constituem um grupo parafilético. Esse cenário, em que os táxons refletem a proximidade evolutiva das espécies, sem correlações entre táxons com fenótipos semelhantes, sugere a ausência de convergência evolutiva nas sequências alinhadas.

Figura 4 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.mammal+I+G4 com suporte de *bootstrap* para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene THSZ3

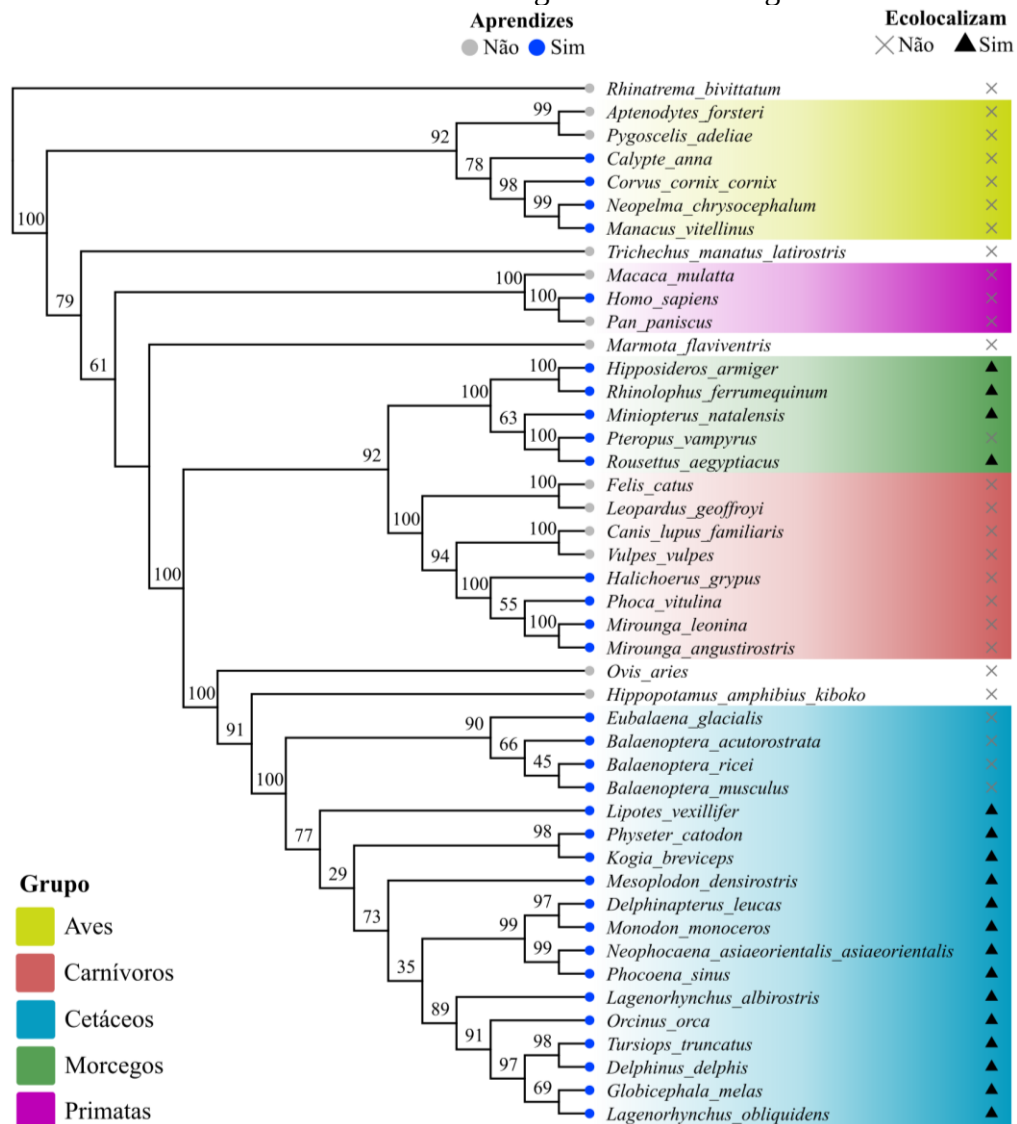


Fonte: autora

Na árvore filogenética do gene ZNF536 (Figura 5) observa-se que todos os grupos de espécies próximas filogeneticamente, formaram grupos monofiléticos. Os cetáceos formam um grupo monofilético com altos valores de *bootstrap* (bb 100 e 91) e são claramente subdivididos em odontocetos e misticetos. Essa subdivisão reflete uma divergência evolutiva esperada e está em consonância com as expectativas baseadas na história evolutiva desses grupos.

Com um suporte estatístico de 92, morcegos e carnívoros emergem como grupos irmãos e são os táxons mais próximos dos cetáceos. Essa proximidade sugere uma afinidade evolutiva significativa entre esses grupos.

Figura 5 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.bird+G4 com suporte de *bootstrap* para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene ZNF536

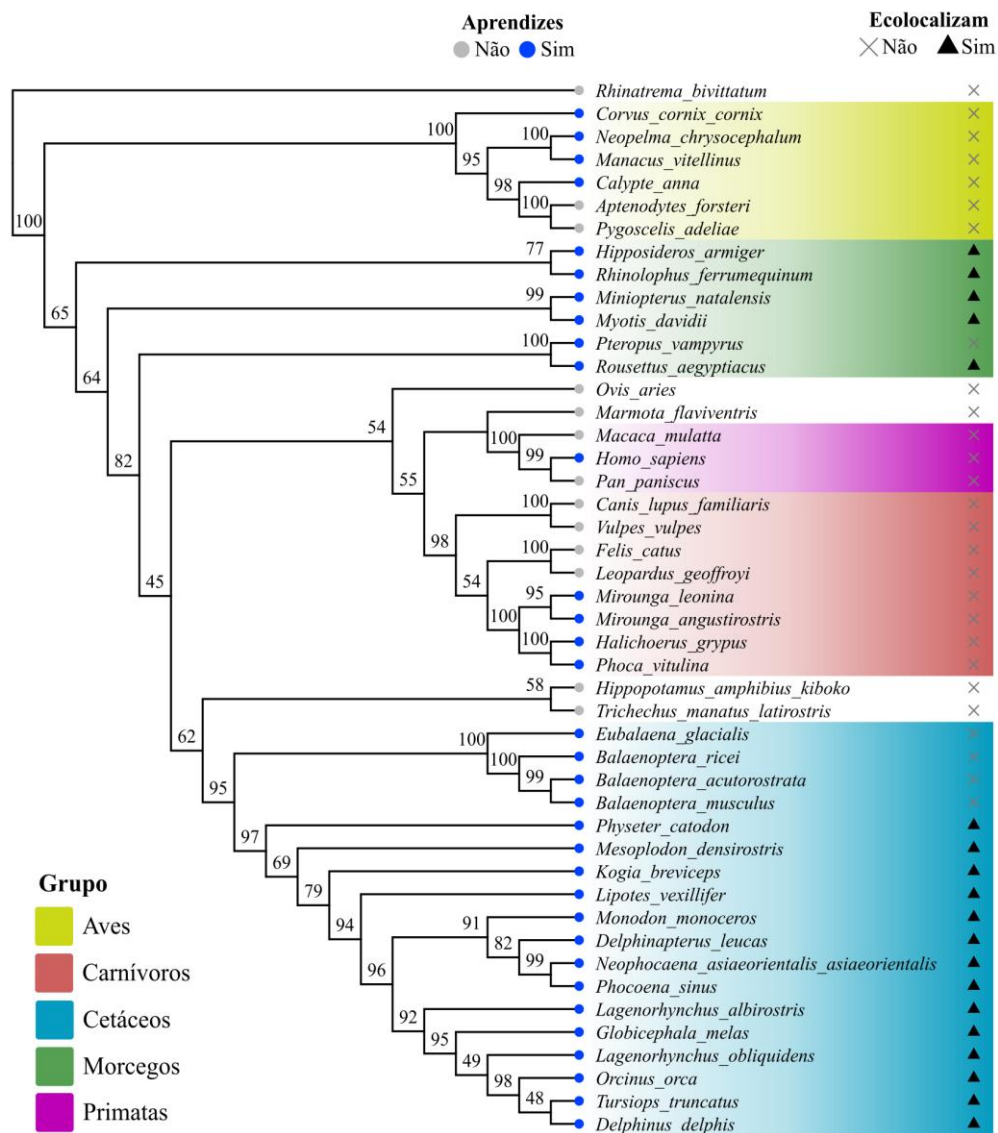


Fonte: autora

Na árvore filogenética do gene GALC (Figura 6) todos os grupos analisados são monofiléticos, exceto pelos morcegos, que se apresentam como um grupo parafilético. As aves formam um grupo irmão de *Rhinatrema bivittatum*, assim como observado na árvore do gene ZNF536 (Figura 5), com um forte suporte estatístico, refletido em valores de *bootstrap* muito altos (bb 100). Os carnívoros também se organizam como grupos irmãos.

Além disso, os táxons hipopótamo e peixe-boi formam um grupo irmão e são os mais próximos dos cetáceos. Os cetáceos se organizam de forma muito fidedigna à filogenia evolutiva das espécies, como evidenciado nas análises dos genes ZNF536 (Figura 5) e TSHZ3 (Figura 4). Nesse formato, não se espera uma convergência evolutiva significativa nas seqüências de proteínas do alinhamento.

Figura 6 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.plant+I+G4 com suporte de *bootstrap* para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene GALC

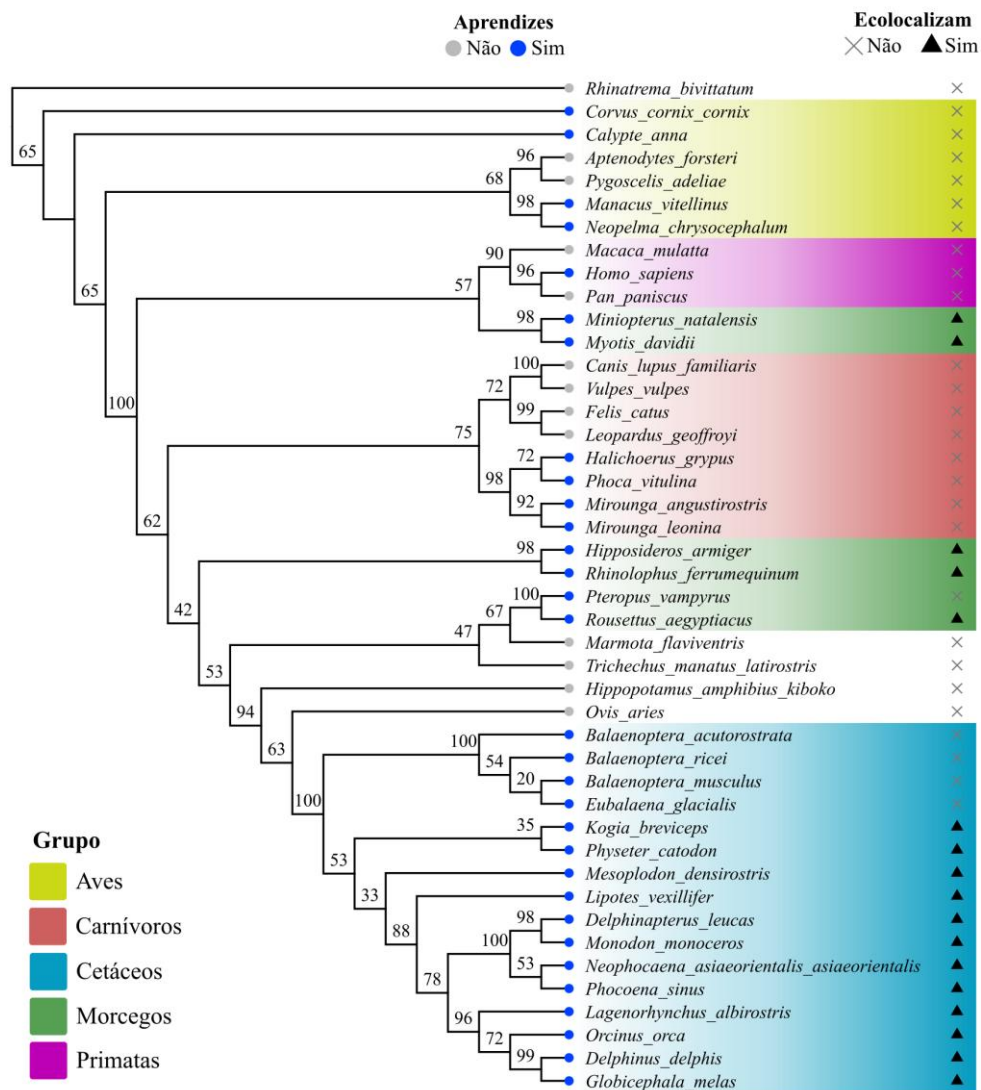


Fonte: autora

Na árvore filogenética do gene CDH4 (Figura 7) pode-se observar que os cetáceos permaneceram num grupo monofilético, como já visto em genes analisados anteriormente. No entanto, o grupo dos morcegos se dividiu três vezes nessa árvore.

O primeiro grupo a se diferenciar inclui as espécies *Pteropus vampyrus* e *Rousettus aegyptiacus*, que se posicionam como o grupo mais externo em relação a todas as outras espécies. Em seguida, vemos que os morcegos *Rhinolophus ferrumequinum* e *Hipposideros armiger* ficam próximos aos demais grupos, que incluem carnívoros, outros morcegos, primatas e aves. A terceira divisão dos morcegos envolve as espécies *Miniopterus natalensis* e *Myotis davidii*, que aparecem como grupo irmão dos primatas, com um valor de *bootstrap* de 57.

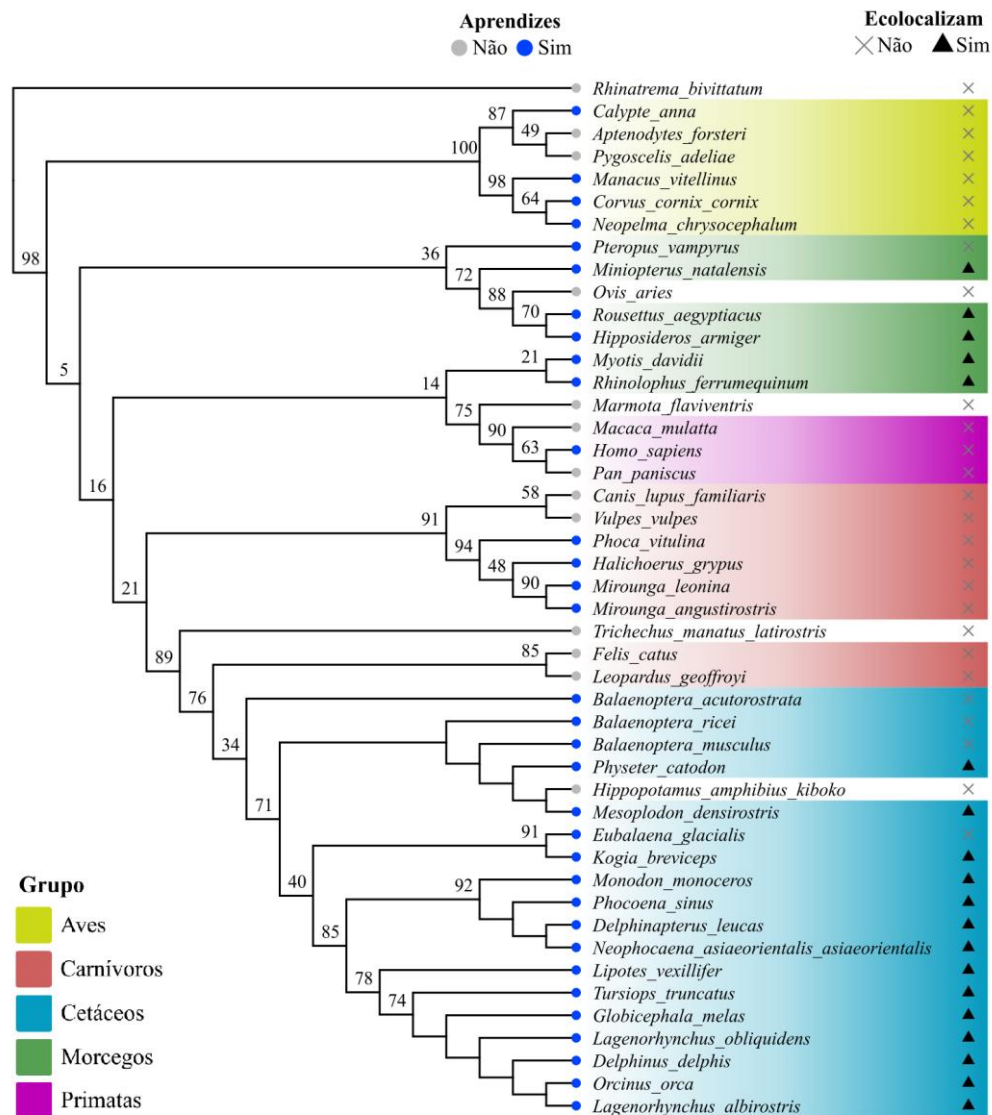
Figura 7 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.mammal+I+R2 com suporte de *bootstrap* para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene CDH4



Fonte: autora

Na análise do gene TCF4 (Figura 8), os cetáceos pela primeira vez formam um grupo monofilético pouco suportado, com valor de *bootstrap* 34%. Observa-se que muitas das espécies possui suporte de *bootstrap* significativo em seus agrupamentos internos, apesar de valores baixos para os ramos externos.

Figura 8 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição Q.insect+F+G4 com suporte de *bootstrap* para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene TCF4



Fonte: autora

O ramo que separa os morcegos dos demais mamíferos apresentou um suporte de *bootstrap* de apenas 5%, considerado extremamente baixo, indicando uma relação próxima com a ovelha (*Ovis aries*). Além disso, dois táxons de morcegos foram posicionados como grupo-irmão dos primatas e da marmota, seguindo uma topologia que diverge

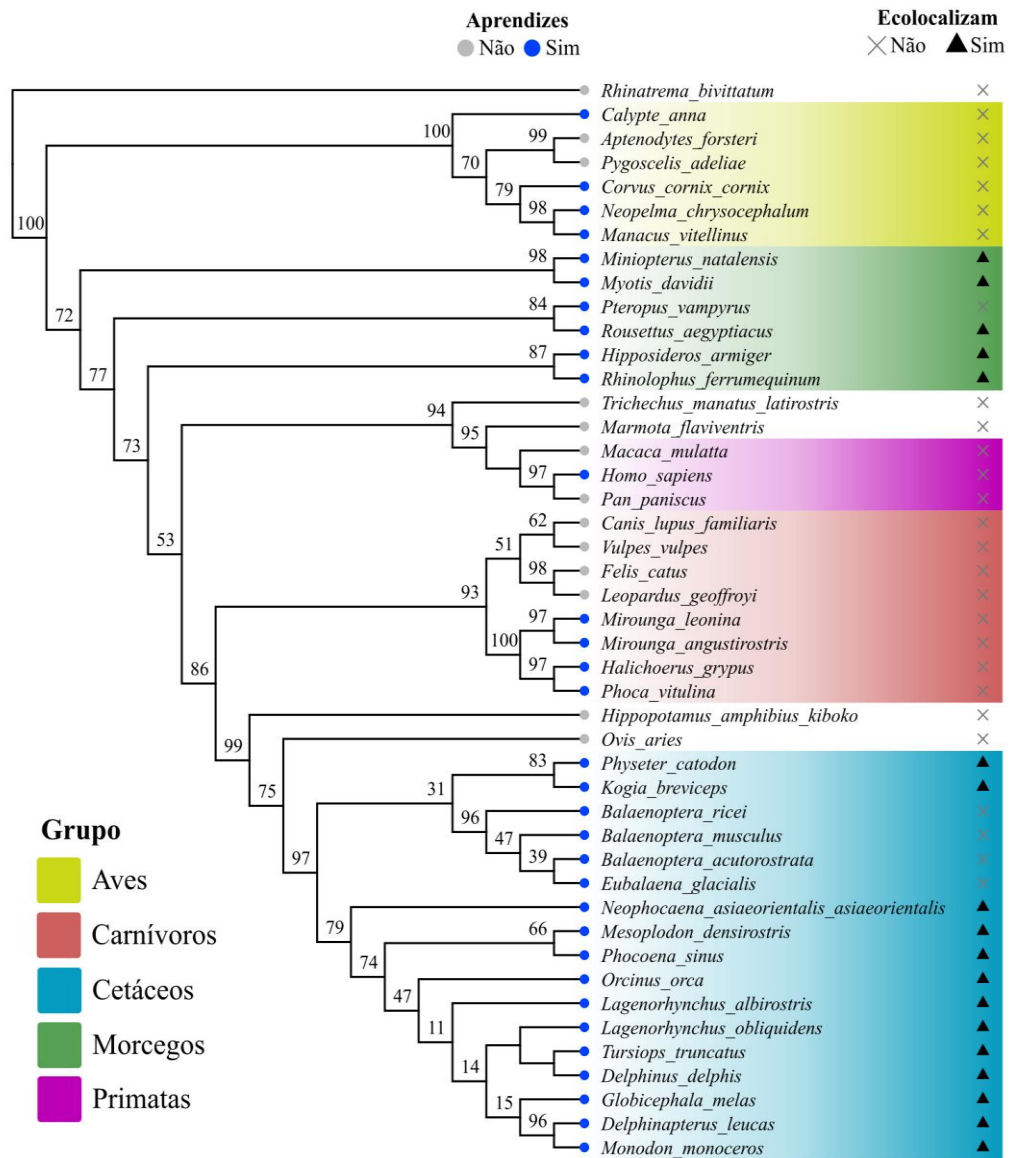
significativamente das relações esperadas para essas espécies. O grupo dos cetáceos foi agrupado como grupo-irmão dos felinos, enquanto o táxon do hipopótamo foi alocado junto aos mysticetos, com um suporte de *bootstrap* de 71%. Diferentemente dos demais genes, em TCF4 os valores de *bootstrap* foram em maioria baixos, e a topologia obtida se mostrou consideravelmente divergente da topologia das espécies.

A filogenia do gene ERBB4 (Figura 9) apresenta agrupamentos bem suportados por valores de *bootstrap* elevados, indicando confiança nas relações evolutivas inferidas entre as espécies. Nos cetáceos, observa-se um forte suporte para diversos clados: *Physeter catodon* e *Kogia breviceps* formam um clado com *bootstrap* de 83%, enquanto *Balaenoptera musculus*, *Balaenoptera acutorostrata* e *Eubalaena glacialis* estão agrupados com suporte de 96%. A monofilia do clado dos cetáceos é novamente reforçada nesta filogenia, onde os cetáceos e a ovelha formam grupos irmãos fortemente sustentado por um *bootstrap* de 99%.

Há um forte suporte para a separação do grupo mais externo *Rhinatrema bivittatum*, com *bootstrap* de 100%. Os primatas apresentam *bootstrap* de 97% para o clado formado por *Homo sapiens* e *Macaca mulatta*, e de 94% para o clado entre *Macaca mulatta* e *Pan paniscus*. Entre os carnívoros, o clado formado por *Canis lupus familiaris* e *Vulpes vulpes* tem um valor de *bootstrap* de 51%, enquanto *Felis catus* e *Leopardus geoffroyi* compartilham um valor de 86%. No grupo dos morcegos, destaca-se o suporte de 87% para o clado entre *Pteropus vampyrus* e *Rousettus aegyptiacus*, e de 98% para *Miniopterus natalensis* e *Myotis davidii*.

Vemos que a árvore revela uma organização que se alinha com as filogenias tradicionais das espécies, agrupando todos os grupos como monofiléticos, a não ser pelos morcegos, que se separam em topologias com *bootstrap* com média de 70%. Essa similaridade entre topologia molecular e das espécies, minimiza a hipótese de convergência molecular.

Figura 9 - Árvore filogenética construída pelo método de máxima verossimilhança utilizando o modelo de substituição JTT+R2 com suporte de *bootstrap* para 1000 réplicas, a partir do alinhamento de aminoácidos da região transcrita do gene ERBB4



Fonte: autora

Analisar a similaridade das sequências permite identificar quais genes apresentam maior conservação genética ou até mesmo ortologia, onde uma maior similaridade indica uma maior conservação ao longo da evolução. Em contraste, uma menor similaridade pode indicar divergências evolutivas, que podem ocorrer devido a adaptações a diferentes pressões ambientais ou maior velocidade de mutações ao longo do tempo.

Na Tabela 4, podemos visualizar uma grande discrepância entre o gene de maior similaridade, FOXP2 com 90,29%, e o de menor similaridade, ERBB4, com 0,63%.

Tabela 4 - Taxa de similaridade entre as sequências

Gene	Taxa de Similaridade
FOXP2	90,29%
THSZ3	73,13%
GALC	51,54%
ZNF536	57,70%
CDH4	58,23%
TCF4	6,03%
ERBB4	0,63%

Fonte: autora

Embora o gene ERBB4 apresente baixa similaridade, sua filogenia reflete a topologia esperada das espécies, mantendo uma organização consistente. Em contraste, TCF4 exibe uma organização dos grupos que se desvia significativamente da topologia das espécies, o que pode ser atribuído à baixa similaridade da proteína.

ERBB4 pertence à mesma família de receptores associados ao EGF, um gene que foi pseudogenizado nos mysticetos, com sua perda de função ligada à adaptação ao ambiente aquático e à perda de dentes nesse grupo (Silva *et al*, 2024). No entanto, ERBB4 manteve-se funcional, destacando sua importância em processos essenciais. Sua possível relação direta com a aprendizagem vocal parece ser mais complexa, e a filogenia não indicou evidências de convergência evolutiva entre cetáceos e outras espécies nesse contexto.

De maneira geral, a análise filogenética dos genes revela uma clara monofilia dos cetáceos. Na análise dos genes TSHZ3, ZNF536, GALC e CDH4 os cetáceos, continuam a mostrar sua monofilia, com uma subdivisão clara entre odontocetos e mysticetos. Por se manterem próximos a filogenia das espécies, sugere que essas sequências realmente refletem a história evolutiva sem sinais de convergência dentro do grupo dos cetáceos.

Dentro do grupo dos morcegos, observou-se uma notável variação, com algumas espécies se posicionando próximas e outras, em momentos diferentes, distantes do clado cetáceo. Na literatura, já foram evidenciadas convergências evolutivas a nível molecular entre cetáceos e morcegos. Um exemplo é o gene *Prestin*, que apresentou substituições convergentes de aminoácidos em golfinhos com ecolocalização, agrupando-se com morcegos ecolocalizadores em uma árvore filogenética, somados a evidências de que essas mudanças foram impulsionadas pela seleção natural (Liu *et al*, 2010).

Na análise do gene FOXP2, duas espécies de morcegos se agruparam como grupo irmão dos cetáceos, apresentando um alto suporte de *bootstrap*, o que sugere uma possível convergência evolutiva entre os clados.

Diversos estudos têm apontado para semelhanças notáveis em adaptações morfológicas e comportamentais, sugerindo que as pressões ambientais semelhantes enfrentadas por esses animais aquáticos e voadores podem ter levado ao desenvolvimento de características semelhantes, apesar de suas distantes linhagens evolutivas (Foote *et al*, 2015; Liu *et al*, 2010). As mudanças moleculares convergentes podem ser a base das adaptações fenotípicas (Parker *et al*, 2013).

Estudos evolutivos, mostram que gene FOXP2 não é altamente conservado em morcegos, e apresenta grande diversidade em morcegos com ecolocalização. Inclusive, apresentam possível seleção divergente no FOXP2 sugerindo um possível papel do gene na evolução da ecolocalização (Liu *et al*, 2010).

O FOXP2 é o primeiro gene identificado especificamente envolvido no desenvolvimento da fala e da linguagem em humanos. Além disso, as sequências genéticas de baleias, golfinhos e morcegos com aprendizagem vocal não compartilham as substituições únicas encontradas em humanos (Webb *et al*, 2005). Isso torna plausível a hipótese de que a aprendizagem vocal evoluiu de maneiras distintas entre os mamíferos, com possíveis casos de convergência evolutiva entre morcegos e cetáceos.

Além disso, é importante considerar a possibilidade de pleiotropia. Essa característica, na qual um único gene pode influenciar várias características fenotípicas, pode ter um papel significativo nas adaptações dos cetáceos. Genes como FOXP2 e TSHZ3 estão ligados a processos neurológicos e desenvolvimentais que podem afetar a morfologia e o comportamento dos cetáceos. Isso sugere que as pressões evolutivas que moldaram esses organismos são complexas e interligadas, e podem estar fortemente relacionadas à expressão genética e epigenética ao invés de apenas as diferenças moleculares.

4.2 EVOLUÇÃO MOLECULAR

Considerando a reconhecida importância do gene FOXP2 em aspectos relacionados à fala e à comunicação, e a possibilidade de convergência evolutiva observada entre cetáceos e morcegos na análise filogenética anterior, aprofundamos nossa investigação com uma análise de seleção positiva nas regiões codificantes do gene FOXP2.

Os resultados obtidos a partir das análises de modelos de seleção utilizando o CODEML do pacote PAML revelaram importantes *insights* sobre as pressões seletivas atuando sobre o gene em estudo. Para avaliar a presença de seleção positiva no gene, foram

testados diferentes modelos evolutivos, comparando a razão entre as taxas de substituições não sinônimas (dN) e sinônimas (dS), conhecida como ω (ou dN/dS).

Tabela 5 - Comparação entre modelos evolutivos para análise de seleção positiva

Modelo	Np	Estimativas de parâmetros	InL	Comparação de modelos	2 Δ InL	Sites selecionados positivamente
M0	89	omega (dN/dS) = 0.05808	-9072,384304	-	-	Nenhum
M1	90	p: 0.95105 0.04895 w: 0.02812 1.00000	-8997,153712	M1 vs M2	-0,000008	Nenhum
M2	92	p: 0.95105 0.02956 0.01939 w: 0.02812 1.00000 1.00000	-8997,153716			185Q, 186Q, 187P, 192G
M7	90	p = 0.10531 q = 1.40620	-8991,866190	M7 vs M8	8,707462	Nenhum
M8	92	p0 = 0.99627 p = 0.12284 q = 1.87252 (p1 = 0.00373) w = 1.88310	-8987,512459			185Q** , 186Q, 187P, 190H, 191P, 192G

Np: número de parâmetros estimados; InL: log likelihood score; 2 Δ InL duas vezes o log likelihood score dos modelos comparados.

** Sites positivamente selecionados com $p < 0,05$

Fonte: autora

O modelo M0 é considerado a hipótese nula porque assume que não há variação na pressão seletiva ao longo do gene, ou seja, que todas as partes do gene estão sujeitas à mesma taxa de substituição. Isso é usado como ponto de partida para comparar outros modelos que podem considerar variações na seleção. A estimativa de $\omega = 0,05808$ em M0 (Tabela 5) sugere que o gene está, de fato, sob seleção purificadora³.

³ Quando $\omega > 1$, isso indica uma predominância de substituições não sinônimas, sugerindo seleção positiva. Se $\omega \cong 1$, não há evidências de evolução, enquanto $\omega < 1$ indicam que as substituições sinônimas superam as não sinônimas, sugerindo que o gene está sob seleção purificadora, que atua para conservar a integridade do gene.

Os modelos M1 e M2 introduzem uma maior complexidade ao permitir classes de sítios sob diferentes regimes seletivos. O M1 distingue entre sítios sob seleção purificadora ($\omega < 1$) e sítios neutros ($\omega = 1$). Com 95,1% dos sítios estimados sob seleção purificadora ($\omega = 0.02812$) e 4,89% sob neutralidade ($\omega = 1$), o modelo reforça a ideia de que a maioria dos sítios do gene são altamente conservados ou sofrendo mutações de manutenção

O modelo M2, que admite a possibilidade de seleção positiva, identificou uma pequena fração de sítios com $\omega > 1$. Entre os sítios apontados como sujeitos à seleção positiva, destacam-se as posições 185Q ($p = 0.875$), 186Q ($p = 0.571$), 187P ($p = 0.521$) e 192G ($p = 0.695$). No entanto, a comparação dos valores de verossimilhança entre os modelos M1 e M2, utilizando o teste de $2\Delta\ln L$, não evidenciou diferenças significativas, apresentando variação baixíssima. Isso sugere que adicionar uma classe de seleção positiva ao modelo M2 não melhora significativamente o ajuste em comparação com o modelo M1. Assim, a evidência de seleção positiva neste contexto não é conclusiva.

O modelo M7, que assume uma distribuição beta para ω , confirmou a ausência de seleção positiva, com todos os sítios estimados dentro do intervalo $0 < \omega < 1$. Isso sugere que, mais uma vez, a seleção purificadora é predominante no gene.

Já o modelo M8, que permite $\omega > 1$, revelou a presença de uma pequena fração de sítios com $\omega = 1.88310$, indicando seleção positiva nesses sítios. A comparação entre os modelos M7 e M8 usando o teste de $2\Delta\ln L$ resultou em um valor significativo, sugerindo que o modelo M8 se ajusta melhor aos dados. Isso fornece evidências de que alguns sítios no gene estão sob pressão seletiva positiva, possivelmente contribuindo para a adaptação evolutiva da proteína.

Para M7 e M8, ainda podemos comparar o valor 8.707 com uma distribuição χ^2 com 2 graus de liberdade. O valor crítico para χ^2 com 2 graus de liberdade e um nível de significância de 5% ($p = 0,05$) é 5,991. Como $8,707 > 5,991$, isso indica que a hipótese nula, de que M7 explica os dados tão bem quanto M8, é rejeitada. Assim, há evidências estatísticas de que M8 fornece uma explicação melhor que M7, indicando a presença de seleção positiva com base nesse teste.

Os sítios identificados como sujeitos à seleção positiva no modelo M8 incluem as posições 185Q, 186Q, 187P, 190H, 191P e 192G, com destaque para o sítio 185Q, que foi marcado com uma forte evidência de seleção ($p < 0,05$).

De maneira geral, os resultados indicam que o gene estudado é altamente conservado, com evidências predominantes de seleção purificadora. No entanto, a análise mais detalhada utilizando o modelo M8 sugere a presença de seleção positiva em alguns sítios específicos, possivelmente refletindo adaptações recentes. Esses achados são consistentes com a ideia de

que genes essenciais podem conter regiões funcionais conservadas, ao mesmo tempo que retêm uma pequena fração de sítios que evoluem sob seleção positiva, permitindo variação adaptativa ao longo do tempo. Esses sítios podem estar relacionados a mudanças funcionais importantes, talvez conferindo vantagens seletivas em certos contextos ecológicos ou adaptativos.

4.3 TENDÊNCIA EVOLUTIVA *POLY-Q*

Homorrepetições são sequências repetitivas de aminoácidos idênticos em uma proteína, onde um único aminoácido é repetido várias vezes em sequência. As sequências Poly-Q, formadas por repetições do aminoácido glutamina, são uma das repetições homopoliméricas mais comuns nos eucariotos (Faux *et al*, 2005) e estão presentes em diversas proteínas, variando amplamente entre espécies, organismos e até populações de uma mesma espécie. Essa variação pode impactar a evolução de proteínas e organismos de diferentes formas. Além disso, a expansão anômala dessas repetições, resultado de expansões dos trinucleotídeos CAG, está associada a pelo menos nove doenças neurodegenerativas hereditárias em humanos (Fan *et al*, 2014).

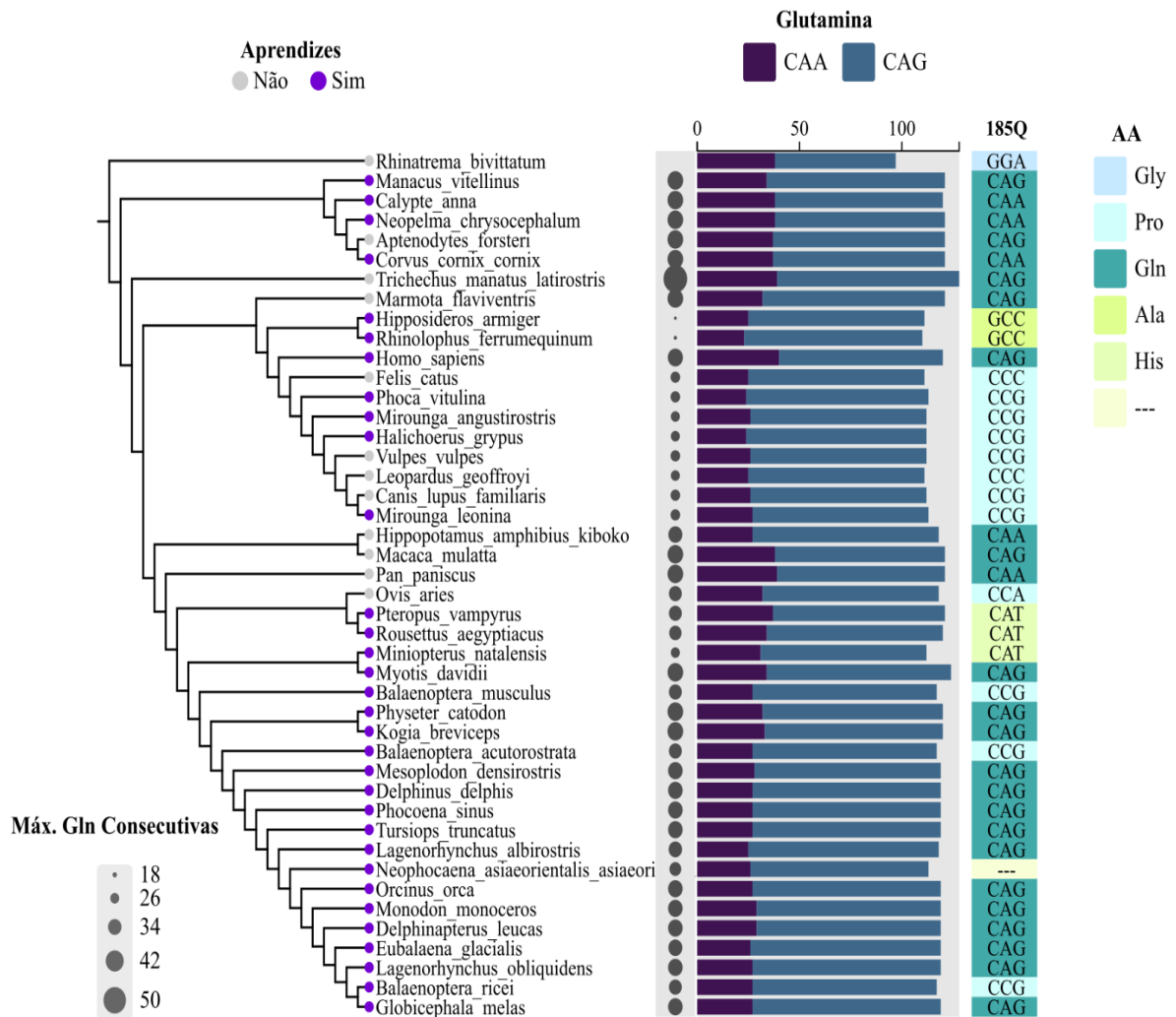
Ao realizar a análise filogenética e evolutiva do gene FOXP2, percebemos uma enorme região de homorrepetição do aminoácido glutamina, sendo a maior sequência de glutamina consecutivas com 52 repetições de glutamina no táxon do peixe-boi (*Trichechus manatus latirostris*). Dentro desta mesma região de Poly-Q, foi onde foi indicado três dos possíveis sítios em seleção positiva, entre eles o sítio com maior suporte estatístico 185Q (Figura 10).

Uma proposta de modelo evolutivo sugere que a alta frequência de CAG leva à instabilidade e crescimento dessas regiões, contribuindo para a variabilidade evolutiva em organismos (Mier *et al*, 2018). Logo, essa associação entre a presença de sítios sob seleção positiva e a instabilidade causada pelas expansões de CAG reforça a hipótese de que essas regiões têm potencial de influenciar a função e a evolução da proteína FOXP2.

O número máximo de glutaminas consecutivas varia amplamente entre as espécies, como indicado pelo tamanho das bolhas no gráfico, que vão de 18 a 52 repetições. Cetáceos mysticetos do gênero *Balaenoptera* apresentam uma característica única entre os cetáceos, com uma prolina substituindo a glutamina na sequência Poly-Q, enquanto o táxon *Eubalaena* manteve a Glutamina,

As espécies com as menores sequências de poly-Q incluem a cobra-cega, um grupo externo, com 10 repetições, e, surpreendentemente, duas espécies de morcegos: *Hipposideros armiger* (15) e *Rhinolophus ferrumequinum* (16). Ambas as espécies de morcegos apresentam também uma substituição de alanina no sítio 185Q, em vez da glutamina típica. Outros táxons de morcegos também exibiram substituições de aminoácidos nessa posição, com exceção de *Myotis davidii*, que manteve a glutamina. O grupo dos morcegos foi o com maior variação no sítio. Os primatas foram os que demonstraram maior quantidade de códons CAA na região.

Figura 10 - Frequência dos códons CAA e CAG e número de glutaminas consecutivas em espécies analisadas, com relação ao fenótipo de aprendizado vocal



Fonte: autora

Algumas espécies intimamente relacionadas compartilham padrões semelhantes no número de glutaminas consecutivas e na proporção de códons de glutamina (como por exemplo o caso dos carnívoros), sugerindo uma possível herança evolutiva comum. No

entanto, observa-se variação dentro de grupos aparentados, indicando que a frequência e o número de glutaminas consecutivas podem ter evoluído conforme a filogenia das espécies, sem uma relação direta com o fenótipo. Além disso, não foi identificado um padrão claro associado ao fenótipo de aprendizagem vocal, tanto aprendizes quanto não aprendizes apresentando quantidades semelhantes de glutamina e as mesmas substituições. Isso sugere que o número de glutaminas e as substituições observadas podem não estar diretamente ligadas ao aprendizado vocal.

Evidências mostram que as repetições de poliQ, assim como outras repetições de aminoácidos, modulam a estrutura, as interações e a função das proteínas de maneira dependente do comprimento, com implicações evolutivas significativas. Especificamente, a variação no comprimento das repetições de poliQ (glutamina) ao longo da evolução regula a estrutura, o processo de separação de fases líquido-líquido e a função do gene FOXP2 (Vaglietti *et al*, 2023).

Como visto por Vaglietti (2023) em um estudo recente, revelou que a relação entre os comprimentos das repetições de poliQ (Q1/Q2) no gene FOXP2 pode estar diretamente associada aos intervalos de frequência das vocalizações dos animais, desde as infrassônicas até as ultrassônicas. Especificamente, espécies que emitem vocalizações ultrassônicas, como os morcegos, possuem uma relação Q1/Q2 baixa, enquanto espécies que emitem vocalizações em frequências mais baixas (infrassônicas), como elefantes e baleias, apresentam uma relação Q1/Q2 mais alta.

O estudo também observa que, as mudanças no gene FOXP2 que ocorrem em morcegos não são as mesmas que ocorrem em cetáceos odontocetos. Isso sugere que, embora ambos os grupos possuam sistemas de vocalização especializados, esses sistemas evoluíram de maneiras distintas, com modificações genéticas independentes.

Embora não exista um padrão diretamente relacionado à convergência do fenótipo de aprendizagem vocal, há evidências substanciais de que essas mutações, sujeitas à seleção positiva, impactam o funcionamento da proteína e, consequentemente, influenciam a frequência vocal ao longo da evolução. A alteração na frequência vocal sugere que a aprendizagem vocal possa também ser moldada por pressões seletivas em nível molecular.

5 CONCLUSÃO

As análises filogenéticas dos genes FOXP2, TSHZ3, ZNF536, GALC, CDH4, TCF4, e ERBB4 nos oferecem uma visão filogenética das relações evolutivas entre cetáceos,

morcegos, primatas, carnívoros e aves. Filogeneticamente, apenas o gene FOXP2 demonstrou uma possibilidade de convergência evolutiva do grupo dos cetáceos com os morcegos.

Na análise de evolução molecular do gene FOXP2, o sítio 185Q foi identificado como apresentando forte evidência estatística de estar sob seleção positiva, com $p > 0,05$. Além disso, esse sítio, juntamente com outros dois que também podem estar sujeitos a seleção positiva, compõe uma longa sequência de glutaminas, conhecida como poly-Q. Essa associação entre a presença de sítios sob seleção positiva, a instabilidade decorrente das expansões de CAG e a influência da taxa entre Q1/Q2 na frequência de vocalização, sugere um papel crucial dessas regiões na função e evolução do FOXP2. Embora essas variações nas regiões poly-Q impactem a frequência da vocalização, é importante ressaltar que frequência de vocalização não é equivalente à aprendizagem vocal. Além disso, as mutações sujeitas à seleção positiva no FOXP2 podem ter implicações além do aprendizado vocal, como na regulação do desenvolvimento neural e nos processos de interação social.

Este estudo contribui para a compreensão das adaptações evolutivas do gene FOXP2 nos cetáceos, destacando as diferenças nas regiões das sequências de proteínas desse gene entre diversos grupos de mamíferos, com ênfase na possibilidade de convergência entre morcegos e cetáceos. Principalmente sobre a possibilidade de sítios específicos sob seleção positiva que podem desempenhar um papel crucial nas adaptações fenotípicas associadas a este gene.

Além disso, a predominância de pesquisas centradas em aves e humanos limita uma visão mais abrangente sobre as linhagens de mamíferos marinhos. Portanto, é essencial realizar mais estudos focados nesses animais para expandir o entendimento sobre suas particularidades evolutivas. Para superar esses desafios, investigações futuras devem explorar as inter-relações de elementos ativadores e regulatórios de expressão gênica, a fim de minimizar vieses de pleiotropia.

REFERÊNCIAS

- ABASCAL, F.; ZARDOYA, R.; TELFORD, M. J. **TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations**. *Nucleic Acids Research*, v. 38, n. suppl_2, p. W7–W13, 1 jul. 2010.
- ANNALISA, B.; SUMICH, J. L.; KOVACS, K. M. **Marine mammals: evolutionary biology**. 2a edição ed. [s.l.] Elsevier, 2005.
- BRENOWITZ, E. A.; BEECHER, M. D. **An ecological and neurobiological perspective on the evolution of vocal learning**. *Frontiers in Ecology and Evolution*, v. 11, 30 jun. 2023.
- CABRERA, A. A. et al. **A Genetic Perspective on Cetacean Evolution**. *Annual Review of Ecology, Evolution, and Systematics*, v. 52, n. 1, p. 131–151, 2 nov. 2021.
- CASTRESANA, J. **Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis**. *Molecular Biology and Evolution*, v. 17, n. 4, p. 540–552, 1 abr. 2000.
- CAUBIT, X. et al. **TSHZ3 deletion causes an autism syndrome and defects in cortical projection neurons**. *Nature Genetics*, v. 48, n. 11, p. 1359–1369, 1 nov. 2016.
- CHABBERT, D. et al. **Postnatal Tshz3 Deletion Drives Altered Corticostriatal Function and Autism Spectrum Disorder-like Behavior**. *Biological Psychiatry*, v. 86, n. 4, p. 274–285, ago. 2019.
- DAVIS, R. W. **Marine Mammals**. Cham: Springer International Publishing, 2019.
- DEREEPER, A. et al. **Phylogeny.fr: robust phylogenetic analysis for the non-specialist**. *Nucleic Acids Research*, v. 36, n. Web Server, p. W465–W469, 19 maio 2008.
- DEREEPER, A. et al. **BLAST-EXPLORER helps you building datasets for phylogenetic analysis**. *BMC Evolutionary Biology*, v. 10, n. 1, p. 8, 12 dez. 2010.
- DUENGEN, D.; FITCH, W. T.; RAVIGNANI, A. **Hoover the talking seal**. *Current Biology*, v. 33, n. 2, p. R50–R52, 23 jan. 2023.
- ESSER, K. -H; SCHMIDT, U. **Mother-Infant Communication in the Lesser Spear-nosed Bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae) — Evidence for Acoustic Learning**. *Ethology*, v. 82, n. 2, p. 156–168, 1989.
- FAN, H.-C. et al. **Polyglutamine (PolyQ) Diseases: Genetics to Treatments**. *Cell Transplantation*, v. 23, n. 4–5, p. 441–458, 1 maio 2014.
- FAUX, N. G. et al. **Functional insights from the distribution and role of homopeptide repeat-containing proteins**. *Genome Research*, v. 15, n. 4, p. 537–551, abr. 2005.

- FITCH, W. T. **The evolution of speech: a comparative review.** Trends in Cognitive Sciences, v. 4, n. 7, p. 258–267, jul. 2000.
- FOOTE, A. D. et al. **Convergent evolution of the genomes of marine mammals.** Nature Genetics, v. 47, n. 3, p. 272–275, 26 mar. 2015.
- HALL, T.; BIOSCIENCES, I.; CARLSBAD, C. J. G. B. B. **BioEdit: an important software for molecular biology.** GEF Bull Biosci 2.1 , p. 60–61, 2011.
- HARTEN, L. et al. **Persistent producer-scrounger relationships in bats.** Science Advances, v. 4, n. 2, 2 fev. 2018.
- HERSH, T. A.; RAVIGNANI, A.; WHITEHEAD, H. **Cetaceans are the next frontier for vocal rhythm research.** Proceedings of the National Academy of Sciences, v. 121, n. 25, 18 jun. 2024.
- JANIK, V. M. **Cetacean vocal learning and communication.** Current Opinion in NeurobiologyElsevier Ltd, , 2014.
- JANIK, V. M.; KNÖRNSCHILD, M. **Vocal production learning in mammals revisited.** Philosophical Transactions of the Royal Society B: Biological SciencesRoyal Society Publishing, , 25 out. 2021.
- JANIK, V. M.; SLATER, P. J. B. **The different roles of social learning in vocal communication.** Animal BehaviourAcademic Press, , 2000.
- JARVIS, E. D. **Learned Birdsong and the Neurobiology of Human Language.** Annals of the New York Academy of Sciences, v. 1016, n. 1, p. 749–777, 16 jun. 2004.
- JARVIS, E. D. **Evolution of vocal learning and spoken language.** Science, v. 366, n. 6461, p. 50–54, 4 out. 2019.
- KALYAANAMOORTHY, S. et al. **ModelFinder: fast model selection for accurate phylogenetic estimates.** Nature Methods, v. 14, n. 6, p. 587–589, 8 jun. 2017.
- KAPLOW, I. M. et al. **Relating enhancer genetic variation across mammals to complex phenotypes using machine learning.** Science, v. 380, n. 6643, 28 abr. 2023.
- KNÖRNSCHILD, M. **Vocal production learning in bats.** Current Opinion in NeurobiologyElsevier Ltd, , 2014.
- LAI, C. S. L. et al. **A forkhead-domain gene is mutated in a severe speech and language disorder.** Nature, v. 413, n. 6855, p. 519–523, 4 out. 2001.
- LIU, Y. et al. **Convergent sequence evolution between echolocating bats and dolphins.** Current Biology, v. 20, n. 2, p. R53–R54, jan. 2010.
- LOVELL, P. V. et al. **The constitutive differential transcriptome of a brain circuit for vocal learning.** BMC Genomics, v. 19, n. 1, p. 231, 3 dez. 2018.

- MATSUNAGA, E. et al. **Dynamic Expression of Cadherins Regulates Vocal Development in a Songbird**. PLoS ONE, v. 6, n. 9, p. e25272, 20 set. 2011.
- MAY-COLLADO, L. J.; AGNARSSON, I.; WARTZOK, D. **Phylogenetic review of tonal sound production in whales in relation to sociality**. BMC Evolutionary Biology, 2007.
- MIER, P.; ANDRADE-NAVARRO, M. A. **Glutamine Codon Usage and polyQ Evolution in Primates Depend on the Q Stretch Length**. Genome Biology and Evolution, v. 10, n. 3, p. 816–825, 1 mar. 2018.
- MORGAN, A. et al. **FOXP2-Related Speech and Language Disorder**. [s.l: s.n.].
- NGUYEN, L.-T. et al. **IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies**. Molecular Biology and Evolution, v. 32, n. 1, p. 268–274, jan. 2015.
- NOWICKI, S.; SEARCY, W. A. **The evolution of vocal learning**. Current Opinion in Neurobiology, v. 28, p. 48–53, out. 2014.
- PAPADOPOULOS, J. S.; AGARWALA, R. **COBALT: constraint-based alignment tool for multiple protein sequences**. Bioinformatics, v. 23, n. 9, p. 1073–1079, 1 maio 2007.
- PARKER, J. et al. **Genome-wide signatures of convergent evolution in echolocating mammals**. Nature, v. 502, n. 7470, p. 228–231, 4 out. 2013.
- PRAT, Y. et al. **Crowd vocal learning induces vocal dialects in bats: Playback of conspecifics shapes fundamental frequency usage by pups**. PLoS Biology, v. 15, n. 10, 31 out. 2017.
- REICHMUTH, C.; CASEY, C. **Vocal learning in seals, sea lions, and walruses**. Current Opinion in NeurobiologyElsevier Ltd, , 2014.
- RENDELL, L. et al. **Can genetic differences explain vocal dialect variation in sperm whales, physeter macrocephalus?** Behavior Genetics, v. 42, n. 2, p. 332–343, mar. 2012.
- RICE, P.; LONGDEN, I.; BLEASBY, A. **EMBOSS: The European Molecular Biology Open Software Suite**. Trends in Genetics, v. 16, n. 6, p. 276–277, jun. 2000.
- SILVA, F. A. et al. **The molecular evolution of genes previously associated with large sizes reveals possible pathways to cetacean gigantism**. Scientific Reports, v. 13, n. 1, p. 67, 19 jan. 2023.
- STAMATAKIS, A. **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. Bioinformatics, v. 30, n. 9, p. 1312–1313, 1 maio 2014.
- SUYAMA, M.; TORRENTS, D.; BORK, P. **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments**. Nucleic Acids Research, v. 34, n. Web Server, p. W609–W612, 1 jul. 2006.

- TEIXEIRA, J. R. et al. **Transcription factor 4 and its association with psychiatric disorders**. *Translational Psychiatry*, v. 11, n. 1, p. 19, 5 jan. 2021.
- THALER, L.; GOODALE, M. A. **Echolocation in humans: an overview**. *WIREs Cognitive Science*, v. 7, n. 6, p. 382–393, 19 nov. 2016.
- TYACK, P. L. **Vocal Learning and Auditory-Vocal Feedback**. In: [s.l.: s.n.]. p. 261–295.
- TYACK, P. L. **A taxonomy for vocal learning**. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 375, n. 1789, p. 20180406, 6 jan. 2020.
- TYACK, P. L.; SAYIGH, L. S. **Vocal learning in cetaceans**. In: *Social Influences on Vocal Development*. [s.l.] Cambridge University Press, 2010. p. 208–233.
- VERNES, S. C. et al. **Vocal learning in animals and humans**. *Philosophical Transactions of the Royal Society B: Biological Sciences* Royal Society Publishing, , 25 out. 2021.
- VERNES, S. C.; WILKINSON, G. S. **Behaviour, biology and evolution of vocal learning in bats**. *Philosophical Transactions of the Royal Society B: Biological Sciences* Royal Society Publishing, , 6 jan. 2020.
- WEBB, D. M.; ZHANG, J. **FoxP2 in Song-Learning Birds and Vocal-Learning Mammals**. *Journal of Heredity*, v. 96, n. 3, p. 212–216, 1 maio 2005.
- WHITE, S. A. **Genes and vocal learning**. *Brain and Language*, v. 115, n. 1, p. 21–28, out. 2010.
- WIRTHLIN, M. E. et al. **Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements**. *Science*, v. 383, n. 6690, 29 mar. 2024.
- XIE, J. et al. **Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees**. *Nucleic Acids Research*, v. 51, n. W1, p. W587–W592, 5 jul. 2023.
- YANG, Z. **PAML: a program package for phylogenetic analysis by maximum likelihood**. *Computer applications in the biosciences* 13.5 , p. 555–556, 1997.

APÊNDICE A – TABELA DE ESPÉCIES E FENÓTIPOS

Classificação	Espécie	Aprendizes Vocais	Ecolocalizam
Primata	<i>Pan paniscus</i>	Não	Não
Primata	<i>Homo sapiens</i>	Sim	Não
Caninos	<i>Canis lupus familiaris</i>	Não	Não
Caninos	<i>Vulpes vulpes</i>	Não	Não
Felinos	<i>Felis catus</i>	Não	Não
Felinos	<i>Leopardus geoffroyi</i>	Não	Não
Cobra Cega	<i>Rhinatrema bivittatum</i>	Não	Não
Hipopótamo	<i>Hippopotamus amphibius kiboko</i>	Não	Não
Beija-Flor	<i>Calypte anna</i>	Sim	Não
Marmota	<i>Marmota flaviventris</i>	Não	Não
Megabats	<i>Pteropus vampyrus</i>	Sim	Não
Megabats	<i>Rousettus aegyptiacus</i>	Sim	Sim
Microbats	<i>Hipposideros armiger</i>	Sim	Sim
Microbats	<i>Miniopterus natalensis</i>	Sim	Sim
Microbats	<i>Myotis davidii</i>	Sim	Sim
Microbats	<i>Rhinolophus ferrumequinum</i>	Sim	Sim
Misticetos	<i>Balaenoptera ricei</i>	Sim	Não
Misticetos	<i>Balaenoptera acutorostrata</i>	Sim	Não
Misticetos	<i>Balaenoptera musculus</i>	Sim	Não
Misticetos	<i>Eubalaena glacialis</i>	Sim	Não
Odontocetos	<i>Orcinus orca</i>	Sim	Sim
Odontocetos	<i>Physeter catodon</i>	Sim	Sim
Odontocetos	<i>Tursiops truncatus</i>	Sim	Sim
Odontocetos	<i>Lagenorhynchus albirostris</i>	Sim	Sim
Odontocetos	<i>Delphinapterus leucas</i>	Sim	Sim
Odontocetos	<i>Delphinus delphis</i>	Sim	Sim
Odontocetos	<i>Globicephala melas</i>	Sim	Sim
Odontocetos	<i>Kogia breviceps</i>	Sim	Sim
Odontocetos	<i>Lagenorhynchus obliquidens</i>	Sim	Sim
Odontocetos	<i>Lipotes vexillifer</i>	Sim	Sim
Odontocetos	<i>Mesoplodon densirostris</i>	Sim	Sim
Odontocetos	<i>Monodon monoceros</i>	Sim	Sim
Odontocetos	<i>Neophocaena asiaeorientalis</i>	Sim	Sim
Odontocetos	<i>Phocoena sinus</i>	Sim	Sim
Primata	<i>Macaca mulatta</i>	Não	Não
Ovelha	<i>Ovis aries</i>	Não	Não
Peixe-boi	<i>Trichechus manatus latirostris</i>	Não	Não
Penguim	<i>Aptenodytes forsteri</i>	Não	Não
Penguim	<i>Pygoscelis adeliae</i>	Não	Não
Pínípedes	<i>Halichoerus grypus</i>	Sim	Não
Pínípedes	<i>Mirounga leonina</i>	Sim	Não
Pínípedes	<i>Mirounga angustirostris</i>	Sim	Não
Pínípedes	<i>Phoca vitulina</i>	Sim	Não
Songbird	<i>Corvus cornix cornix</i>	Sim	Não
Songbird	<i>Manacus vitellinus</i>	Sim	Não
Songbird	<i>Neopelma chrysocephalum</i>	Sim	Não

APÊNDICE B – TABELA COM GENES E RESPECTIVOS CÓDIGOS DO NCBI

Espécie	CDH4	ERBB4	FOXP2	GALC	TCF4	TSHZ3	ZNF536
<i>Aptenodytes forsteri</i>	XM_019470656.1	XM_019472536.1	XM_009286823.2	XM_009280938.2	XM_019470871.1	XM_009281053.2	XM_019472532.1
<i>Calypste anna</i>	XM_030463155.1	XM_030453944.1	XM_030450312.1	XM_030452366.1	XM_030467039.1	XM_030457973.1	XM_008491665.2
<i>Corvus cornix cornix</i>	XM_039563524.1	XM_039554924.1	XM_039571412.1	XM_039552461.1	XM_039567222.1	XM_039558430.1	XM_010396527.4
<i>Manacus vitellinus</i>	XM_025895228.1	XM_018071743.3	XM_008932480.2	XM_018082924.2	XM_029965664.1	XM_008933325.3	XM_018086169.3
<i>Neopelma chrysocephalum</i>	XM_027697374.1	XM_027678006.1	XM_027682191.1	XM_027673070.1	XM_027701987.1	XM_027697664.1	XM_027697630.1
<i>Pygoscelis adeliae</i>	XM_009326875.1	XM_009329670.1	XM_009324060.1	XM_009319517.1	XM_009323115.1	XM_009320281.1	XM_009320282.1
<i>Hipposideros armiger</i>	XM_019637155.1	XM_019656860.1	XM_019624737.1	XM_019642363.1	XM_019640659.1	XM_019644226.1	XM_019642730.1
<i>Miniopterus natalensis</i>	XM_016203289.1	XM_016196749.1	XM_016204589.1	XM_016210117.1	XM_016225105.1	XM_016212250.1	XM_016211453.1
<i>Myotis davidii</i>	XM_015567722.1	XM_006778963.2	XM_015569384.1	XM_006754754.2	XM_006764465.2	XM_015568584.1	-
<i>Pteropus vampyrus</i>	XM_023524845.1	XM_011358660.2	XM_011372403.1	XM_023528837.1	XM_011361114.2	XM_011380963.2	XM_023523054.1
<i>Rhinolophus ferrumequinum</i>	XM_033094686.1	XM_033113116.1	XM_033099057.1	XM_033109610.1	XM_033087721.1	XM_033129452.1	XM_033128581.1
<i>Rousettus aegyptiacus</i>	XM_036229218.1	XM_036236253.1	XM_036231023.1	XM_016123818.2	XM_036230062.1	XM_036231208.1	XM_036231210.1
<i>Canis lupus familiaris</i>	XM_038434262.1	XM_003640190.5	XM_038439870.1	XM_038672844.1	XM_038653403.1	XM_005616783.4	XM_038529518.1
<i>Felis catus</i>	XM_023251048.2	XM_019838610.3	XM_023240626.2	XM_023255913.2	XM_019815315.3	XM_023245049.2	XM_019819656.3
<i>Halichoerus grypus</i>	XM_036105351.1	XM_036089853.1	XM_036068397.1	XM_036106029.1	XM_036080727.1	XM_036122785.1	XM_036070663.1
<i>Leopardus geoffroyi</i>	XM_045444030.1	XM_045481106.1	XM_045495376.1	XM_045451778.1	XM_045456815.1	XM_045442579.1	XM_045442581.1
<i>Mirounga leonina</i>	XM_035025504.1	XM_035007223.1	XM_035020960.1	XM_034993021.1	XM_034995773.1	XM_035023537.1	XM_035023538.1
<i>Mirounga angustirostris</i>	XM_064573410.1	XM_064585839.1	XM_045864381.3	XM_045901355.3	XM_045886655.3	XM_045891196.3	XM_064580337.1
<i>Phoca vitulina</i>	XM_032387709.1	XM_032402953.1	XM_032401732.1	XM_032405717.1	XM_032422839.1	XM_032390424.1	XM_032390474.1
<i>Vulpes vulpes</i>	XM_026015148.1	XM_026002035.1	XM_026012156.1	XM_026012504.1	XM_025997639.1	XM_026009960.1	XM_026010189.1
<i>Balaenoptera ricei</i>	XM_059898535.1	XM_059928620.1	XM_059933563.1	XM_059914777.1	XM_059895580.1	XM_059903879.1	XM_059904661.1
<i>Balaenoptera acutorostrata</i>	XM_057530175.1	XM_007187863.2	XM_057550248.1	XM_057543201.1	XM_007191942.3	XM_057534656.1	XM_007165032.2
<i>Balaenoptera musculus</i>	XM_036824123.1	XM_036858277.1	XM_036862984.1	XM_036843476.1	XM_036874757.1	XM_036835060.1	XM_036834929.1
<i>Eubalaena glacialis</i>	XM_061207811.1	XM_061187117.1	XM_061197955.1	XM_061182780.1	XM_061170093.1	XM_061172615.1	XM_061173614.1
<i>Orcinus orca</i>	XM_033410223.2	XM_049712234.1	XM_049715204.1	XM_004262307.3	XM_004266116.4	XM_004284860.4	XM_033415005.2
<i>Physeter catodon</i>	XM_024128443.2	XM_028481562.2	XM_055084871.1	XM_028496287.2	XM_007124266.4	XM_024132550.3	XM_055079085.1
<i>Tursiops truncatus</i>	-	XM_033859813.1	XM_019925131.2	XM_019919695.2	XM_019937088.2	XM_033844615.1	XM_033845145.1
<i>Lagenorhynchus albirostris</i>	XM_060122263.1	XM_060152176.1	XM_060156819.1	XM_060125106.1	XM_060121916.1	XM_060132143.1	XM_060130858.1

Delphinapterus leucas	XM_022597400.2	XM_022567593.2	XM_022582869.1	XM_022576369.2	XM_022584418.2	XM_022585874.2	XM_022585875.2
Delphinus delphis	XM_060033317.1	XM_060016164.1	XM_060020757.1	XM_060004092.1	XM_060029094.1	XM_060000816.1	XM_059999941.1
Globicephala melas	XM_060284119.1	XM_030852949.2	XM_060305385.1	XM_030883662.2	XM_030867984.2	XM_030874504.2	XM_060288590.1
Kogia breviceps	XM_059036614.1	XM_059053751.1	XM_059073722.1	XM_059057001.1	XM_059041424.1	XM_059044731.1	XM_059044309.1
Lagenorhynchus obliquidens	-	XM_027122298.1	XM_027102140.1	XM_027120234.1	XM_027118567.1	XM_027079288.1	XM_027079287.1
Lipotes vexillifer	XM_007462285.1	-	-	XM_007471611.1	XM_007459680.1	XM_007470932.1	XM_007470931.1
Mesoplodon densirostris	XM_060120473.1	XM_060107375.1	XM_060108608.1	XM_060097961.1	XM_060119606.1	XM_060084597.1	XM_060084531.1
Monodon monoceros	XM_029219931.1	XM_029224991.1	XM_029234320.1	XM_029239517.1	XM_029217298.1	XM_029235256.1	XM_029208684.1
Neophocaena asiaorientalis	XM_024750190.1	XM_024763203.1	XM_024757744.1	XM_024736268.1	XM_024756812.1	XM_024734051.1	XM_024734107.1
Phocoena sinus	XM_032606699.1	XM_032638162.1	XM_032642873.1	XM_032622584.1	XM_032654530.1	XM_032614763.1	XM_032613631.1
Rhinatrema bivittatum	XM_029611831.1	XM_029606129.1	XM_029617020.1	XM_029597724.1	XM_029579771.1	XM_029608432.1	XM_029608434.1
Homo sapiens	NM_001794.5	NM_005235.3	NM_014491.4	NM_000153.4	NM_001083962.2	NM_020856.4	NM_014717.3
Pan paniscus	XM_034947466.3	XM_003812504.7	XM_034963386.3	XM_003832736.6	XM_055102433.2	XM_003816213.7	XM_055103793.2
Macaca mulatta	XM_028828556.1	XM_015110886.2	XM_028844760.1	NM_001044262.2	XM_028838241.1	XM_001103809.3	XM_015123485.2
Hippopotamus amphibius	XM_057703945.1	XM_057744756.1	XM_057730881.1	XM_057732360.1	XM_057698757.1	XM_057713597.1	XM_057712666.1
Ovis aries	XM_027976984.3	XM_042244993.2	XM_060414734.1	XM_060418682.1	XM_042239261.1	XM_027978176.3	XM_042231923.2
Marmota flaviventris	XM_027954314.1	XM_027941813.1	XM_027926404.1	XM_027931575.1	XM_027949012.1	XM_027941435.1	XM_034636659.1
Trichechus manatus latirostris	XM_023741423.1	XM_023731191.1	XM_004382567.2	XM_023731678.1	XM_004379246.3	XM_004371621.3	XM_023726956.1