UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL INSTITUTO DE BIOCIÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Evolutionary forces shaping genomes: from whales to viruses

(Forças evolutivas moldando genomas: das baleias aos vírus)

Yuri Belisario Yépez Mora

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular)

Orientadora: **Prof.ª Dr.ª Maria Cátira Bortolini**

Porto Alegre, fevereiro de 2024

Institutions and funding sources

This work was conducted at the facilities of the Laboratory of Human and Molecular Evolution of the Department of Genetics at the Institute of Biosciences of the Federal University of Rio Grande do Sul, with funding from the Coordination for the Improvement of Higher Education Personnel (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*, CAPES).

The student received a scholarship granted by the Coordination for the Improvement of Higher Education Personnel.

Dedication

To my mother, for everything.

Acknowledgments

Gostaria de me permitir fazer meus agradecimentos em português…

Quero agradecer por todo o apoio recebido durante o processo de realização desta tese. Em primeiro lugar, agradeço à minha orientadora Cátira por ser a pessoa responsável, em um sentido figurado, por tudo o que vivi nos últimos 5 anos, por ter aberto as portas do seu laboratório e, além da academia, por ter me ajudado tanto como pessoa e por ser uma figura de carinho, respeito e admiração pessoal.

Também quero reconhecer os professores e membros acadêmicos, administrativos e operacionais do programa PPGBM por suas lições, orientações, comentários e incentivos ao longo desta trajetória acadêmica. Um agradecimento especial ao senhor Elmo, por ser uma pessoa tão excepcional em seu trabalho e com um carisma que é contagiante, alegrando as pessoas, e ao Prof. Guilherme Baldo por me permitir passar um tempo fazendo parte de seu grupo de pesquisa, do qual aprendi muito.

Quero agradecer à agência de financiamento CAPES, pela bolsa de doutorado recebida durante minha pesquisa. Às instituições e programas de educação e pesquisa envolvidos: à UFRGS, ao PPGBM e ao HCPA. Também ao Departamento de Genética da UFRGS, e aos laboratórios LEHM, CTG e SBCB por fornecerem recursos, instalações e/ou assistência técnica. E dentro dessas instituições, aos colaboradores, colegas e parceiros de pesquisa que contribuíram para o meu projeto, ao Prof. Márcio Dorn, Bibiana, Camila, Bruno, Bruna, Lucca, entre outros.

Obrigado à minha mãe, meu irmão, minha avó, minhas tias, minha família. Mesmo à distância, celebro vocês. Obrigado aos novos amigos que fiz nesta etapa. Obrigado a Pedro, que desde o início foi um amigo que me ajudou a me inserir em outro país, em outra cultura, tornando tudo mais fácil. Obrigado pelas risadas e pelo espanhol. Obrigado a Luane por ser tão especial, por tudo o que você representa em um nível pessoal. Obrigado a Thaynara pela amizade, os bons momentos e a lasanha de camarão do Pará. Obrigado a todos eles e àqueles que estiveram e estão presentes.

Obrigado a Mariana por ser minha companheira de vida, por estarmos juntos, lado a lado, fazendo os dois o doutorado. Tem sido uma jornada difícil mas no final valiosíssima, daquelas das quais surge o verdadeiro sentido da vida.

Em geral, obrigado a todas as pessoas e instituições pelo apoio que tornou este doutorado possível. Obrigado, Brasil.

Contents

List of abbreviations

ARNT (also termed as HIF1β) - Aryl Hydrocarbon Receptor Nuclear Translocator

ARNT2 (also termed as HIF2β) - Aryl Hydrocarbon Receptor Nuclear Translocator 2

Asn - Asparagine residue

BEAST - Bayesian evolutionary analysis by sampling trees

BEB - Bayes empirical Bayes

bHLH - basic helix-loop-helix protein domain

BLAST - Basic Local Alignment Search Tool

BUSTED - Branch-Site Unrestricted Statistical Test for Episodic Diversification

CAD - C-terminal transactivation domain

CDS - Coding sequences

COVID-19 - Coronavirus disease 2019

CYP8B1 - Cytochrome P450 family 8 subfamily B member 1

dN - Nonsynonymous substitutions per non-synonymous site

dS - Synonymous substitutions per synonymous site

EGLN1 - (also termed as PHD2) Egl-9 Family Hypoxia Inducible Factor 1

EGLN2 - (also termed as PHD1) Egl-9 Family Hypoxia Inducible Factor 2

EGLN3 - (also termed as PHD3) Egl-9 Family Hypoxia Inducible Factor 3

EPAS1 (also termed as HIF2α)

Fe - Iron

FIH1 (also termed as HIF1AN) - Factor inhibiting HIF1

FUBAR - Fast, Unconstrained Bayesian AppRoximation

HIF - Hypoxia-induced factor

HIF pathway - Hypoxia signaling pathway

HIF1α (also termed as HIF1A) - Hypoxia Inducible Factor 1 Subunit Alpha

HIF2α (also termed as HIF2A) - Hypoxia Inducible Factor 2 Subunit Alpha

HIF3α (also termed as HIF3A) - Hypoxia Inducible Factor 3 Subunit Alpha

HRE hypoxia-responsive elements

HyPhy - Hypothesis Testing using Phylogenies

KEGG - Kyoto Encyclopedia of Genes and Genomes

L - Leucine residue

LRT - Likelihood ratio test

m - Metro

MB - Myoglobin

MEME - Mixed Effects Model of Evolution software

NAD - N-terminal transactivation domain

NCBI - National Center for Biotechnology Information

O_2 - Molecular oxygen

ODD - Oxygen-dependent degradation

P - Proline residue

P.1.1 - SARS-CoV-2 variant

P.1.2 - SARS-CoV-2 variant

PAML - Phylogenetic Analysis Using Maximum Likelihood

PAS - Per-Arnt-Sim protein domain

PCOC - Profile Change with One Change

PHD - Prolyl Hydroxylase Domain-Containing Protein

Pro - Proline residue

pVHL - von Hippel-Lindau protein

Q - Glutamine residue

R - Arginine residue

RAG1 - Recombination activating gene 1

S - Serine residue

SARS-CoV-2 - Severe Acute Respiratory Syndrome CoronaVirus 2

T - Threonine residue

UCSF - University of California, San Francisco

V - valine residue

VEGFA - Vascular endothelial growth factor A

VHL - von Hippel-Lindau gene

VOC - Variants of concern

W - tryptophan residue

ΔG_M - mutated protein structure free energy

ΔG_W - wild-type protein structure free energy

$\Delta\Delta G$ - Unfolding *Gibbs* free energy changes

ω - ratio of substitution rates at non-synonymous and synonymous sites

List of Figures and Tables

Resumo

Neste trabalho de tese, integramos investigações sobre as adaptações evolutivas de mamíferos aquáticos à hipóxia e a dinâmica evolutiva do SARS-CoV-2, enfatizando a natureza dinâmica da exploração científica em meio a desafios globais urgentes. Apesar da importância estabelecida de HIF1α e genes associados no mecanismo molecular de resposta à hipóxia, seu papel na adaptação evolutiva de mamíferos aquáticos permanece incompletamente compreendido. Nossa abordagem de genes candidatos analisou 11 genes críticos na via de sinalização de HIF1α em mamíferos aquáticos, revelando evidências de seleção negativa e regimes seletivos relaxados em cetáceos, juntamente com a identificação de uma variante única de glutamina 68 na proteína HIF3α exclusiva de cetáceos. Concomitantemente, investigamos a história evolutiva da cepa Gamma e suas linhagens derivadas (P.1.1 e P.1.2) de SARS-CoV-2, detectando 194 locais sob seleção positiva em 12 genes/ORFs. Notavelmente, alguns locais diagnósticos para Gamma careciam de uma assinatura de seleção positiva, sugerindo evasão da seleção purificadora, enquanto análises de redes revelaram haplótipos em expansão únicos para P.1.1 e P.1.2, incluindo o surgimento de novidades adaptativas como o alelo 3509Y em ORF1a. Discutimos como fenômenos como epistasia e pleiotropia antagônica podem limitar o surgimento de novos alelos em linhagens de SARS-CoV-2, mantendo a infectividade em humanos e fornecendo capacidades de resposta rápida contra as respostas imunes do hospedeiro. Esta abordagem interdisciplinar destaca a interconexão de estudos evolutivos em diversos contextos biológicos e destaca a importância da investigação científica dinâmica no enfrentamento dos desafios globais contemporâneos.

Abstract

In this thesis, we integrate investigations into the evolutionary adaptations of aquatic mammals to hypoxia and the evolutionary dynamics of SARS-CoV-2, emphasizing the dynamic nature of scientific exploration amidst urgent global challenges. Despite the established importance of HIF1α and associated genes in the molecular mechanism of hypoxia response, their role in the evolutionary adaptation of aquatic mammals remains incompletely understood. Our candidate gene approach analyzed 11 critical genes in the HIF1α signaling pathway in aquatic mammals, revealing evidence of negative selection and relaxed selective regimes in cetaceans, along with the identification of a unique glutamine 68 variant in the HIF3α protein exclusive to cetaceans. Concurrently, we investigated the evolutionary history of the Gamma strain and its derived lineages (P.1.1 and P.1.2) of SARS-CoV-2, detecting 194 sites under positive selection across 12 genes/ORFs. Notably, some diagnostic sites for Gamma lacked a signature of positive selection, suggesting evasion of purifying selection, while network analyses unveiled expanding haplotypes unique to P.1.1 and P.1.2, including the emergence of adaptive novelties such as the 3509Y allele in ORF1a. We discuss how phenomena like epistasis and antagonistic pleiotropy may limit the emergence of new alleles in SARS-CoV-2 lineages, maintaining infectivity in humans while providing rapid response capabilities against host immune responses. This interdisciplinary approach underscores the interconnectedness of evolutionary studies in diverse biological contexts and highlights the importance of dynamic scientific inquiry in addressing contemporary global challenges.

Chapter 1. Introduction

1.1. Aquatic mammals: a taxonomic diverse group

Marine mammals encompass a diverse array of species inhabiting oceanic environments, with varying degrees of dependence on aquatic life. This classification is not based on biological or taxonomic grouping. For instance, dolphins are fully reliant on aquatic ecosystems, while sea lions forage underwater but engage in other activities like resting, molting, and breeding on land. Additionally, animals like polar bears and hippopotamuses can alternate between aquatic and terrestrial habitats in pursuit of food. Living marine mammals include representatives from three mammalian orders: Cetacea (whales, dolphins, and porpoises), Sirenia (manatees and dugongs), and some members of Carnivora, such as pinnipeds (seals, sea lions, walruses) and mustelids (sea otters) (Berta et al., 2006; Jefferson et al., 2011). These taxonomic groups (Cetacea, Sirenia, and Carnivora) also include species inhabiting freshwater environments, allowing for a broader category of aquatic mammals that inhabit both ocean and freshwater environments globally.

The large groups of aquatic mammals have independent evolutionary origins from different terrestrial mammal groups (Figure 1). These animal groups share adaptations extremely similar to the aquatic environment that they inhabit despite their separate evolutionary origins. These characteristics include locomotion (such as the modification of the limb shape or the absence of hind limbs in cetaceans and sirenians); the dive ability and anaerobic conditions adaptation; heat exchange and thermal insulation systems; mechanisms of water regulation and modifications of the renal system to conserve water; sensory modifications associated with their ability to live in a variety of aquatic environments (salt, brackish and freshwater; Berta et al., 2006; Foote et al., 2015).

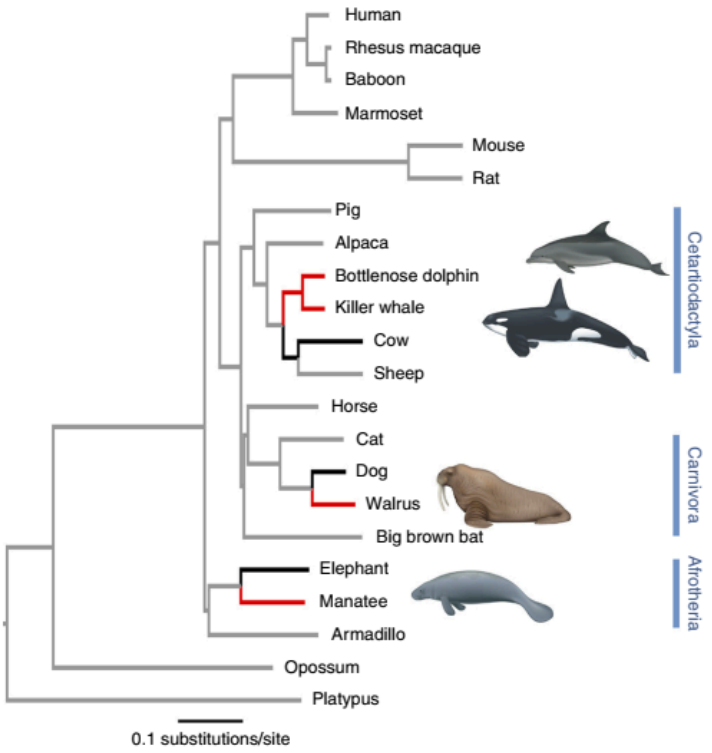

Figure 1. Phylogenetic relations of aquatic mammals within the eutherian group, with a marsupial as an outgroup. Red colored branches representing the independent evolution of aquatic mammal lineages (From Foote et al., 2015).

Specifically, many aquatic mammals are capable of prolonged and deep dives. The capacity of breath-hold at depth depends on the gas exchange, both in the lungs and in the peripheral tissues. Anatomical and physiological adaptations in the respiratory system, cardiovascular system, blood, and peripheral tissues contribute to the remarkable diving capacity of these mammals. For instance, some of these adaptations include slower heart rates, reduced oxygen consumption, and supply of oxygenated blood only to essential organs and tissues, as shown in Table 1. Furthermore, flexible ribs allow the lung to collapse, and thickened tissue in the middle ear helps resist pressures at great depths in pinniped and cetacean species. The final results of these adaptations involve efficient ventilation, better oxygen storage, regulated transport, respiratory gas supply, and tolerance to tremendous environmental pressures, ischemia and extreme tolerance to hypoxia (Berta et al., 2006; Ponganis, 2011); the latter constitutes one of the central points in this thesis.

One interest of this thesis includes the elucidation of some molecular characteristics

(coding portion of genes of interest) potentially involved in macroevolutionary mechanisms that could explain the adaptations of these mammals to the aquatic environment. Specifically, adaptive modifications at the amino acid level are involved in diving capacity and hypoxia tolerance response.

Table 1. Several physiological characteristics of some aquatic mammals in present study. (1) Lung volumes: Total Lung Capacity (TLC); Diving Lung Volume (DLV); (2) Hemoglobin (Hb) Concentrations and Blood Volumes (BV). (3) Myoglobin (Mb) Concentrations and Muscle Mass (% Body Mass). (4) Magnitude and Distribution of O_2 body stores (Modified from Ponganis, 2011).

Species	**TLC** **ml kg^{-1}**	**DLV** **ml kg^{-1}**	**Hb** **g dl^{-1}**	**BV** **ml kg^{-1}**	**Mb** **g 100 g^{-1}**	**Muscle** **%**	**Mass** **kg**	**Total O_2 ml** **kg^{-1}**	**Lungs %**	**Blood %**	**Muscle %**
Sei whale *Balaenoptera borealis*	61-126	-	-	-	0.9	-	-	-	-	-	-
Fin whale *Balaenoptera physalus*	61-126	-	-	-	2.4	-	-	-	-	-	-
Sperm whale *Physeter macrocephalus*	-	28[c]	22	200	-	-	10000	68	4	38	58
Pygmy sperm whale *Kogia breviceps*	-	-	-	-	4.3	-	-	-	-	-	-
Harbor porpoise *Phocoena phocoena*	80-130	-	-	-	-	-	-	-	-	-	-
Bottlenose dolphin *Tursiops truncatus*	50-91	40-50[c]	14	71	3.3	30	200	36	34	27	39

Pacific white-sided dolphin *Lagenorhynchus obliquidens*	-	-	17	108	3.5	-	-	-	-	-	-
Beluga whale *Delphinapeterus leucas*	-	-	21	128	3.4	30	1000	51	17	51	32
Cuvier's beaked whale *Ziphius cavirostris*	-	-	-	-	4.3	-	-	-	-	-	-
Northern fur seal *Callorhinus ursinus*	145	-	17	109	3.5	30	-	-	-	-	-
Steller sea lion *Eumetopias jubata*	110	-	-	120	2.7		238	40	20	45	35
California sea lion *Zalophus californianus*	-	48	18	96-120	2.7-5.4	37	35-87	39-52	21-16	45-41	34-43
Weddell seal *Leptonychotes weddellii*	48	22f-27[g]	-	-	5.4	35	-	-	-	-	-

Northern elephant seal *Mirounga angustirostris*	-	-	25	216	6.5	28	400	87	5	66	29
Walrus *Odonbenus rosmarus*	116	-	16	106	3.0	30	65	38	24	50	26
Weddell seal *Leptonychotes weddellii*	-	-	26	210	5.4	-	400	87	5	66	29
Sea otter *Enhydra lutris*	345	207	17	91	2.6	30	28	55	55	29	16
Manatee *Trichechus manutus*	65	-	15	80	0.4	35	155	21	33	60	7

1.1.1. Order Cetacea

Cetaceans are an order of existing mammals with a fully aquatic life form that includes whales, dolphins, and porpoises. Unlike other groups of aquatic mammals, cetaceans sleep, mate, give birth and breastfeed their young in the water. They are unable to move or support their weight on land. There are approximately 85 living species characterized by about 40 genera. This clade groups two existing suborders and an extinct suborder; the existing suborders are Mysticeti and Odontoceti, with Archaoecete extinct approximately 25 million years ago. Both extant suborders differ in several characteristics associated with their life habits. Odontoceti is a monophyletic group that includes all existing toothed whales; it is the largest and most diverse suborder, with at least 70 species characterized by the presence of teeth, including fish and cephalopod feeders. Mysticeti is also a monophyletic group that includes the 15 species of existing bearded whales, which are fed by filtration of the zooplankton. In some classifications, cetaceans are considered a suborder within a more extensive Order Cetartiodactyla; the suborders Mysticeti and Odontoceti are denoted as infraorders (Gingerich et al., 2001; Folkens & Randall, 2002; Jamieson, 2016).

In overall, cetacean species belong to three families: the rorquals (Balaenopteridae), which represent about 60% of all living mysticetes; and oceanic dolphins (Delphinidae) and beaked whale (Ziphiidae), which represent approximately 50% and 30% of all live odontocetes, respectively (Marx et al., 2016).

Cetaceans vary in size from approximately 1m to more than 30m and are predominantly aquatic, inhabiting both coastal areas and open ocean, and being found in all the world's oceans. Some species also inhabit rivers and freshwater lakes in Asia, South America, and North America, while other species inhabit brackish waters of estuaries and coastal marshes (Berta, 2006). They are an essential part of ocean ecosystems as distributors of large-scale nutrients, top predators, and as a food source for many organisms of the ocean floor (Marx et al., 2016).

From an evolutionary point of view, cetaceans are a highly derived group, that is, they share many derived characters that emerged after the separation of the most recent common ancestor. The evolutionary adaptation to aquatic life led to major modifications in all body parts and their functions such as the loss of hind limbs, profound changes of the skeleton and head, evolution to a fusiform and aerodynamic body, the disappearance of the outer ear and hair, among some of the most extreme modifications (Jamieson, 2016).

In addition to those mentioned above, all cetaceans share general characteristics that are part of a similar body design: paddle-shaped fins (modified forelimbs) that are used to direct, balance and stop, but not to move forward; telescoped skull, that is, the posterior displacement of the facial bones respect the braincase; nasal openings in the upper part of the head; soft and gummy skin; lack of glands and hairs, instead, rely on a thick layer of insulating fat to maintain body temperature; absence of protruding ears; internal reproductive organs; boneless structures in the form of horizontal tail flukes and dorsal fin (secondary loss of the latter in some species), mainly composed of connective tissue (Jefferson et al., 2011; Romero, 2009).

In spite of these shared traits, Mysticeti and Odontoceti species are considerably different. In particular, odontocetes are recognizable by having conical teeth, a single blowhole, and by the echolocation ability, that is, the use of sound to navigate and locate prey. In contrast, mysticetes are often extremely large, have lost any sign of teeth in adulthood, and instead of teeth, possess a series of keratinous, sieve-like baleen plates suspended in two rows from their upper jaw (Marx et al., 2016).

Notwithstanding these modifications, the internal anatomy of cetaceans indicates the ancestry of terrestrial mammals: it is surprisingly similar with some exceptions, such as the presence of a four-chamber stomach and the cartilaginous reinforcement of the airways to the alveolus (Jefferson et al., 2011).

Most morphological and molecular data generally affirm that artiodactyls, specifically hippos (Hippopotamidae), are the closest relatives of cetaceans; these separated from the artiodactyls about 50 million years ago during the Eocene (Graur & Higgins, 1994; Liu &

Miyamoto, 1999; Arnason et al., 2000; Berta et al., 2006). The different synapomorphies and related characteristics between Hippopotamidae and cetaceans have been useful to understand the process of early evolution from a terrestrial environment to the aquatic environment (McGowen et al., 2014).

On the other hand, it is remarkable, the wide range of immersion capacity among the members of this clade. On one end, there are relatively poor divers such as the Yangtze finless porpoises (*Neophocaena asiaeorientalis*) that live in low and middle sections of the Yangtze River in China; this species has a maximum reported depth of 20m, with a maximum immersion time of 2 minutes (Akamatsu et al., 2002). Unlike beaked whales and sperm whales (*Physeter macrocephalus*), on the other hand, they are known as the most outstanding divers in terms of both depth and duration. A sperm whale can dive as deep as 2,000m or more; the most extended dive recording was the 2 hours and 18 minutes by five sperm whales. Their sound clicking was recorded in November 1983, in the Caribbean (Romero, 2009).

Additionally, cetaceans include species that hold several world records, which are often related to gigantic size. The blue whale (*Balaenoptera musculus*), with up to 190 tons, is the most massive animal on Earth, and the largest animal that has ever existed; The sperm whale has the largest brain, up to 8 kg; and the longest of all mammals, the bowhead whale (*Balaena mysticetus*), which can reach an astonishing age of over 200 years (Marx et al., 2016).

1.1.2. Order Sirenia

The sirenians (manatees and dugongs), as well as cetaceans, are fully adapted to aquatic life and spend their entire lives in the water. They are the only fully aquatic mammals that are herbivores. It is widely accepted that the living relatives, phylogenetically closest to the sirenians, are the elephants. The members of the order Sirenia originated from the diversification of Afrotheria, with a fossil record that corresponds to the early Eocene, around 50 million years ago (Berta et al., 2006). Fossil record deposits are found at scattered points

in Europe, Africa, and western India. The fossils of the earliest sirenians indicate a large body supported by four legs; they had four sacral vertebrae and probably were able to walk on land, although due the position of the nasal openings and the large ribs it suggests an amphibious life (Berta et al., 2006; Vaughan & Czaplewski, 2011).

There are four live species in two genera (*Dugong*, Dugongidae; *Trichechus*, Trichechidae). A fifth species, extinct in 1768, Steller´s sea cow (*Hydrodamalis gigas*, belonging to Dugongidae), was cold-adapted for life in the Bering Sea, in contrast to the other members of Sirenia that are distributed in waters Tropical and subtropical. The dugong (*Dugong dugon*) is distributed from tropical waters of East Africa, throughout the Indo-Australian Archipelago, to the western Pacific and the Indian Ocean. The Amazonian manatee (*Trichechus inunguis*) inhabits the Amazon and Orinoco rivers in South America. *Trichechus manatus*, the western Indian manatee, inhabits the Caribbean and the Atlantic coasts from Brazil to the United States, and the African manatee, *Trichechus senagalensis*, inhabits tropical West Africa. Manatees are unified as a monophyletic clade by skull characteristics, specifically, the ear region. Other derived traits include the reduction of neural spines in the vertebrae, the possible tendency of enlargement of the body, and, at least in *Trichechus*, anteroposterior elongation of thoracic vertebrae (Berta et al., 2006; Vaughan & Czaplewski, 2011).

Sirenians are characterized by having a relatively large, robust body, reaching more than 1500 kilograms. The shape is aerodynamic and almost hairless, except for the bristles in the snout, which is downturned. The skin is thick, rough or finely wrinkled. Regarding their closest terrestrial ancestor, the sirenians have extreme morphological modifications: the forelimbs were modified in the form of fins, the hindlimbs are absent, the tail is a horizontal fluke, the pelvis is vestigial and there is no clavicle. Bones are unusually dense, a condition called pachyostosis. The lungs are unlobed, unusually long, horizontally oriented and separated from the massive intestine by a horizontal diaphragm. The orientation of the lungs and the density and weight of the bones allow the animal to make small adjustments in the lung volume to maintain a horizontal attitude while feeding at various depths. Their nostrils are located at the top of their snouts and are closed by valves (Romero, 2009; Marx et al., 2016).

As obligate herbivores, they feed on a variety of seaweeds and plants such as seagrasses, water weeds, and other aquatic vegetation. The members of both families are social, occur in large aggregations, and frequently interact with each other. Due to its dependence on seagrasses, the sirenians do not need to dive to great depths. They typically stay 1 to 3 meters below the surface and come to the surface to breathe every 2 to 5 minutes. However, the deepest feeding dive recorded is 33 meters, made by dugongs. Also, manatee immersion times greater than 16 minutes have been recorded (Scholander & Irving, 1941; Chilvers et al., 2004; Marsh et al., 2012).

1.1.3. Order Carnivora: pinnipeds and mustelids

Both pinnipeds and mustelids are within the order Carnivora. Four families of this order have members that inhabit the aquatic environment: Phocidae, Otariidae, Odobenidae (all members of the suborder Pinnipedia), and Mustelidae (Romero, 2009). Like all typical carnivores, they have well-developed claws and a pair of specialized teeth in the shape of blades for cutting food; in pinnipeds, typical teeth called carnassials have been modified as premolars.

1.1.3.1. Pinnipedia

Pinnipeds comprise a little more than a quarter (28%) of the diversity of aquatic mammals. Approximately 34 species are distributed worldwide: 19 phocids, 14 to 15 otariids, and the walrus. Around 90% of individuals are phocids, and the remaining 10% are otariids and odobenids. The fossil record indicates that extant pinnipeds represent only a small fraction of what was once a larger and more diverse group, dating back to an origin of 30.6-28 million years ago, in the mid-late Oligocene (Berta et al., 2006; Berta, 2018).

The name pinniped comes from the Latin "pinna" and "pedis", which means "feather feet", referring to the paddle-like fore- and hind limbs. Pinnipeds, like cetaceans and sirenians, have a radically modified body plan regarding terrestrial mammals. However,

unlike those, they have an amphibious way of life and spend part of their life on land or ice, to mate and give birth ((Berta et al., 2006).

Unlike for cetaceans, the early stages of pinniped evolution are not clear. Pinnipeds monophyly is well supported by both morphological and molecular data. The closest relatives of the pinnipeds are the arctoid carnivores, and most of the evidence supports a link between the pinnipeds and the ursids or the pinnipeds and the mustelids (Berta et al., 2018).

Phocidae (true, "hair," and earless seals) is the morphologically most divergent pinnipeds family. Most morphological and molecular data indicate phocid as a monophyletic group and recognize two monophyletic subfamilies, Monachinae (Southern Hemisphere seals) and Phocinae (Northern Hemisphere seals). They are distinguishable from otariids and odobenids based on several features of the ear region. This family is characterized by a fusiform and rotund body, with a short and thick neck; pinna or auricle absent; the rear fins cannot turn forward, and body undulations allow the progression on land; fins with hair on both surfaces; claws on fin tips; skulls almost or completely lack postorbital processes, and the alisphenoid canal is also absent (Romero, 2009; Berta et al., 2018).

Compared to otariids, the phocids inhabit higher latitudes, spend more time in the water, exhibit different locomotion patterns, and have more diverse cranial morphologies, as well as dietary and reproductive strategies. They also have a larger body size compared to otariids, with an average weight of two tons in the northern elephant seal. Additionally, several phocids, especially the sea elephant and the Weddell seal, are spectacular divers that feed on squid and pelagic fish at depths of 1000 meters or more. Most seals feed on fish, squid, octopus, and shellfish, and one species (*Hydrurga leptonyx*) feeds on penguins and other seals (Romero, 2009; Berta et al., 2018).

The Otariidae family (eared seals or fur seals and sea lions) is the second most diverse clade of existing pinnipeds, and today they are represented by 14 to 15 species distributed in cold temperate waters of the North Pacific and the Southern Hemisphere. They are divided into two subfamilies, the Otariinae (sea lions) and the Arctocephalinae (fur seals). Several intergeneric hybrids reveal the close relationship between sea lions and fur seals. For

example, in the wild, *Zalophus californianus* has hybridized with *Eumetopias jubatus* and *Callorhinus ursinus*, and there is information about hybrids between *Arctocephalus townsendi* and *Callorhinus ursinus*; in captivity, the California sea lion *Zalophus californianus* has hybridized with *Arctocephalus pusillus* and *Otaria flavescens* (Romero, 2009; Berta et al., 2018).

Otariids, unlike the phocids or odobenids, are morphologically conserved and lack certain ecological specializations present in the other families. They are more variable in size, with species ranging from the Galapagos sea lion with 27 kg (*Arctocephalus galapagoensis*) to the male Steller's sea lion of a ton. They feed on fish, cephalopods, and crustaceans. General morphological characteristics of this group include the commonly thin body shape with a long neck; pinna present; long, hairless front fins in most species, with rudimentary fingertips nails; skulls bear-like, an alisphenoid canal is present, as are the postorbital and supraorbital processes. In contrast to phocids, they can move their rear fins forward and use them for walking (Romero, 2009; Berta et al., 2018).

Family Odobenidae has only one extant member, the walrus *Odobenus rosmarus*, being the most distinctive member of the pinnipeds. The most characteristic feature of this animal is a pair of elongated and constantly growing upper canine teeth (tusks) found in adults of both sexes. The modern walrus is a shallow diver of a large body, between 800 to 1,200 kg, which feeds mainly on benthic invertebrates, especially mollusks. The fossil record of walruses is considerably more diverse with at least 20 species described in 16 genera. Fossil walruses first appear in the early Miocene. In contrast to the low diversity, ecological specialization and Arctic distribution of Odobenidae at present, extinct walruses were quite diverse, inhabited temperate and subtropical latitudes, showed a wide range of body sizes and included many fish-eaters that lacked tusks (Berta et al., 2009; Berta et al., 2018).

1.1.3.2. Mustelidae: subfamily Lutrinae

On the other hand, Mustelidae is the largest family in the order Carnivora, including 70 species of otters, skunks, weasels, and badgers. Mustelidae is divided into different

subfamilies: Lutrinae, Taxidiinae, Mellivorinae, Melinae, Helictidinae, Mustelinae, Guloninae, and Ictonychinae (Kruuk, 2006). Otters are members of the subfamily Lutrinae, a semiaquatic group that includes 13 currently recognized extant species (de Ferran et al., 2022). They are the smallest aquatic mammals compared to cetaceans, pinnipeds, and sirenians, and in a way, they are the least "aquatic" of aquatic mammals.

Lutrinae represent one of the mammal clades that most recently entered the aquatic environment. Unlike cetaceans and pinnipeds that inhabit the aquatic environment around 30-50 million years ago, otters have been nearly fully aquatic for only 1 to 3 million years (Berta et al., 2006). Due to these differences in evolutionary histories, otters lack certain adaptations typical of the completely aquatic life form shared by cetaceans and pinnipeds. Some of these physiological adaptations include thermal insulation by blubber, which also acts as an energy reserve; a well-developed diving response that allows oxygen conservation during submersion; and a counter-current heat exchange system to dissipate or maintain heat. Thus, vestiges of a more terrestrial existence may have left the Lutriane members with a narrow range of environmental tolerance. The aquatic lifestyle represents a more demanding energy challenge compared to other aquatic mammals, which is compensated through behaviors such as increased daily intake compared to body size (Romero, 2009; Berta et al., 2018).

The two main branches on the otter tree are *Lutra* (Euroasian) with three species, and *Lontra* (American) with four species, the three *Aonyx* as a more recent offshoot of *Lutra*, and the three species odd ones out (the sea otter *Enhydra lutris*, the giant otter *Pteronura brasiliensis* and the smooth-coated *Lutrogale perspicillata*) (Kruuk, 2006). Only two species have a very different lifestyle than the rest, inhabit the sea: the marine otter, *Lontra felina* (the only aquatic species of the genus) and the sea otter (also, only species of this genus) (Romero, 2009; Berta, 2018). The rest of the species inhabit freshwater habitats.

With respect to physical characteristics, the 13 species range in adult size from 0.6 to 1.8 m in length and 1 to 45 kg in weight. To exemplify on one side, *Enhydra lutris* are the

heavier members of the family Mustelidae. Males have an average weight of 37 kg, and both sexes can reach 1.4m length. Like the pinnipeds, the hind legs are webbed and adapted for swimming. The front toes are short and stiff, with retractable claws for grooming and eating. On land, their walk is clumsy and vulnerable and is probably the reason why they are rarely found more than a few meters from the water. They have a very dense coat to trap the air and maintain an insulating thermal layer. They are capable of submerging 60 meters to the sea bottom, and generally, the dives last less than 1 minute, although some individuals can remain submerged for 5 minutes or more. They inhabit shallow waters of the North Pacific (Romero, 2009). The marine otter is much smaller, with a maximum length of 1.15 m and an average weight of 4 kg. The short legs are webbed and have strong nails and a dense coat. They are submerged to feed a short distance from the coast, inhabiting different coastal regions of Chile and Peru.

1.2. Evolution to aquatic life and tolerance to hypoxia

Aquatic mammals need to rise to the surface to breathe air and, therefore, are subjected to hypoxia during the submersion time. This is much more noticeable in cetaceans and sirenians, due to the exclusively aquatic life that characterizes them. Immersion behavior is a specific characteristic of each species that varies both in the immersion time and in-depth reach (Table 2). In the case of cetaceans, as Bi et al. (2015) indicate, the extremely wide range of immersion capabilities suggests that the immersion behavior plays an indispensable role in the evolutionary adaptation to aquatic environments. In turn, it can be thought that the availability and regulation of oxygen, directly related to immersion behavior, is a critical element of this evolutionary adaptation.

Table 2. Dive characteristics of some cetacean and pinniped species in this study: dive time and depth reached (Modified from Ponganis, 2011).

Species	Time (min)		Depth (m)	
	Common	Maximum	Common	Maximum

Harbor porpoise - *Phocoena phocoena*	1	5	14-40	226
Bottlenose dolphin - *Tursiops truncatus*	1	8	20	390
Humpback whale - *Megaptera novaeanglieae*	1-4	-	23-118	~160
Fin whale - *Balaenoptera physalus*	3-5	-	180-200	-
Long-finned pilot whale - *Globicephala melas*	5-15	21	100-800	1019
Cuvier's beaked whale - *Ziphius cavirostris*	58-70	95	1070-1334	1888
Sperm whale - Physeter macrocephalus	40-60	138	400-900	2250
Antarctic fur seal - *Arctocephalus gazella*	<2	5	30	101
Galapagos fur seal - *Arctocephalus galapagoensis*	<2	5	26	115
Steller sea lion - *Eumetopias jubata*	<2	8	9-24	250
Northern fur seal - *Callorhinus ursinus*	2	8	65	256
California sea lion - *Zalophus californianus*	2	10	62	274
New Zealand sea lion - *Phocarctos hookeri*	4	11	123	474
Weddell seal - *Leptonychotes weddellii*	10-12	82	150-400	726
Northern elephant seal - *Mirounga angustirostris*	23	119	437	1581

Specifically, the response before hypoxia implies a molecular mechanism that leads to the expression of genes that increase oxygen supply and regulate metabolic activity in the body. In metazoans, this response is coordinated by a central signaling pathway (Semenza, 2007; Taylor & McElwain, 2010), that involves what has been termed as the "regular master" of cellular oxygen homeostasis: a transcription factor known as hypoxia-induced factor (HIF),

which plays a central role in the transcriptional response to changes in oxygen availability (Taylor & McElwain, 2010).

1.3. Hypoxia signaling pathway

HIF is a basic helix-loop-helix (bHLH) protein heterodimer of the PAS (abbreviation for "Per-Arnt-Sim", first coined in 1991) family. It consists of a stable and constitutive unit, HIF1β (also called ARNT1) and an unstable α subunit inducible by hypoxia (Weidemann & Johnson, 2008). There are three homologous HIFα proteins (HIF1α, HIF2α, HIF3α) that form dimers with the HIFβ subunit, each encoded by different locus genes. The HIF3α protein appears to be present only in mammals.

The underlying cellular mechanism under hypoxic conditions is determined by the deregulation of HIFα, which accumulates and translocates to the cell nucleus, where it dimerizes with HIFβ. The HIFα/β heterodimer binds to a DNA motif (RCGTG, R = Purine) of hypoxia-responsive elements (HRE) within the promoter region of HIF target genes. These target genes are fundamentally involved in the hypoxia response system, such as angiogenesis, erythropoiesis, and cellular responses such as impaired glucose energy metabolism and response to oxidative stress (Schofield & Ratcliffe, 2004).

Under normoxic conditions, protein degradation of the HIFα subunit occurs by hydroxylation. This process involves two domains of oxygen-dependent degradation (ODD), called NODD and CODD, which act independently and are located in the central region of the protein. HIFα also has two transactivation domains: N-terminal transactivation domain or NAD, and a C-terminal transactivation domain or CAD (Figure 2).

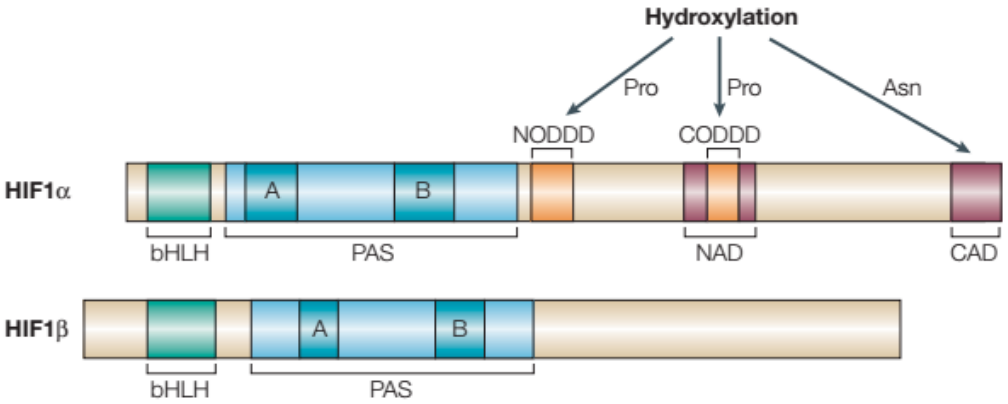

Figure 2. The domain structure of HIF1α and HIF1β. The oxygen responsiveness of HIFα is transmitted by separate protein interactions with three regions of HIFα that transfer the ability to respond to hypoxia (NODD, CODD, and CAD), and these interactions involve oxygen-dependent enzymatic hydroxylation of residues specific in these regions. Basic helix–loop–helix (bHLH) domains; PAS (PER (period circadian protein), the amino-terminal oxygen-dependent degradation domain (NODD); the carboxy-terminal oxygen-dependent degradation domain (CODD); the amino-terminal activation domain (NAD); and the carboxy-terminal activation domain (CAD) (From Schofield & Ratcliffe 2004).

The oxygen-dependent hydroxylation of HIF is catalyzed by three prolyl-4-hydroxylases, PHD1, PHD2, and PHD3 (also called EGLN2, EGLN1, and EGLN3, respectively), of one or two highly conserved prolyl residues near NAD. The hydroxylation of some, or both, of these prolyl residues, generates a binding site for the ubiquitin ligase known as von Hippel-Lindau (pVHL), which is part (with elongin B and C, Cullin 2 and RING-box 1 proteins) of a ubiquitin ligase complex (Figure 3). As a consequence, the polyubiquitination of HIFα is generated, and its consequent proteasomal degradation (Schofield & Ratcliffe, 2004; Kaelin & Ratcliffe 2008). Thus, under normoxic conditions, HIF is maintained in a state of repression by the activity of the HIF hydroxylases.

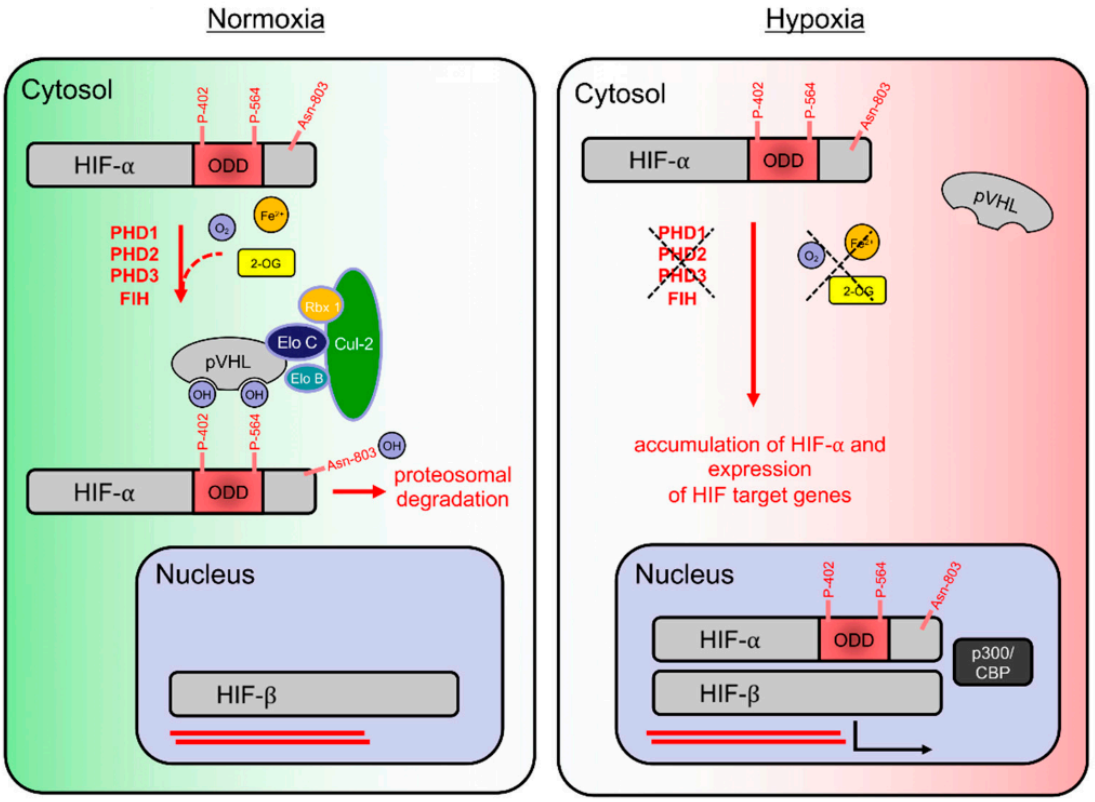

Figure 3. Schematic overview of the HIF regulation by prolyl hydroxylation and proteasomal degradation under normoxia (left) and hypoxia (right). There are three hydroxylation sites in the HIF1α subunit: two prolyl residues (human HIF1α: Pro402 and 564; human HIF2α: Pro405 and Pro531) in the oxygen-dependent degradation domain (ODD) and one asparaginyl residue (human: Asn803) in the C-terminal transactivation domain (C-TAD). Under normoxia, prolyl hydroxylation is catalyzed by hypoxia-inducible factor (HIF) prolyl hydroxylases (PHD1–3) use oxygen (O2), iron (Fe2+), 2-oxoglutarate (2-OG) and ascorbic acid (vitamin C; not shown) as co-substrates, thus triggering recognition by the von Hippel-Lindau tumor suppressor protein (pVHL), which, as part of a E3 ubiquitin ligase complex (also containing elongin [Elo] C and B, cullin-2 [Cul-2] and RING-box protein [Rbx] 1), induces proteasomal degradation. Asparaginyl hydroxylation (Asn803 in human HIF1α; Asn851 in human HIF2α) is catalyzed by the factor-inhibiting HIF (FIH), additionally preventing the binding of the transcriptional co-activator histone acetyltransferase p300/CREB-binding protein (p300/CBP) (left). In contrast, under hypoxia PHDs and FIH are unable to hydroxylate HIFα subunits which accumulate and translocate to the nucleus, heterodimerizes with HIFβ binds to hypoxia-response elements (HREs) in the regulatory regions of numerous HIF target genes (From Strowitzki et al., 2019).

Transactivation repression of HIFα is also carried out by another enzyme called factor-inhibiting HIF or FIH (Mahon et al., 2001) (Figure 3). FIH under conditions of normoxia hydroxylates an asparagine residue within CAD. This hydroxylation generates a steric effect that prevents the binding of the transcriptional coactivators p300 and CBP with HIFα. This mechanism inhibits the transactivation of HIF that escapes degradation.

Like prolyl-4-hydroxylases, FIH is a type of enzyme that metabolizes molecular oxygen, called dioxygenase, characterized by incorporating both oxygen atoms into its products. In turn, they belong to the superfamily of 2-oxoglutarate-dependent-oxygenase, a large group of enzymes that use 2-oxoglutarate (an intermediate of the Krebs cycle) as a co-substrate. During enzymatic catalysis, one oxygen atom is incorporated into the amino acid residue of HIFα, and the other oxygen atom is incorporated into the oxidative decarboxylation of 2-oxoglutarate producing succinate and carbon dioxide. This enzymatic reaction is also dependent on iron (Fe+2) and ascorbic acid (Schofield & Ratcliffe, 2004; Kulshreshtha et al., 2007).

Chordates possess the 3 *PHD* paralogous genes, while flies and worms possess only one member of the *PHD* family, called *Egl9* (Kaelin & Ratcliffe, 2008). The three genes are widely expressed in the former, and the cellular location varies from one another: PHD1 is found exclusively in the nucleus, PHD2 is found mainly in the cytoplasm, and PHD3 is both cytoplasmic and nuclear (Metzen et al., 2003). Among the three PHD enzymes, PHD2 (EGLN1) appears to be the primary prolyl-hydroxylase of HIFα, contributing to the majority of the HIF prolyl hydroxylase activity in cells with normal oxygen levels, and therefore, establishing normoxic levels of HIF1α (Berra et al., 2003).

In addition to directly activating transcription through HREs in the target gene locus, the HIF pathway is extended both by secondary transcription factors under HIF transcriptional control, through non-coding microRNA activity (Kulshreshtha et al., 2007), and by the HIFs interaction with other signaling pathways such as *Notch*, *Myc*, and *Wnt* pathways, through different mechanisms not mediated by HRE (Koshiji et al., 2004; Gustafsson et al., 2005; Kaidi et al., 2007). The conservation of the three HIFα paralogs, along with evidence *in vivo* and in vitro, suggests that its gene products have essential non-redundant functions although

with some functional overlap, which varies between cell types (Kaelin & Ratcliffe, 2008; Weidemann & Johnson 2008).

1.4. Aquatic mammals HIF

The hypoxia signaling pathway is highly conserved among species with varying levels of complexity; It is found both in mammals and more ancient taxa as insects, nematodes, and corals (Hampton‐Smith & Peet, 2009). The importance of this system is illustrated by embryonic lethality in mice when generating homozygous knockout of any *HIF1α*, *HIF2α*, *PHD2*, or *VHL* (Kaelin & Ratcliffe 2008).

On the other hand, the HIF response has been studied in the mammal's adaptation to hypoxia, including humans, mainly in high-altitudes and fossorial animals (Shams et al., 2004; Schmidt et al., 2017; Witt & Huerta-Sánchez, 2019). Multiple investigations, mainly carried out in human populations, with a considerable amount of evidence, show that tolerance and evolutionary adaptation to hypoxia are coordinated by the HIF system. Mainly two genes, *HIF2α* and *PHD2* (also known as *EPAS1* and *EGLN1*, respectively), have been identified in this regard (Peng et al., 2010;Yi et al., 2010; Simonson et al., 2010; Hanaoka et al., 2012, Jeong et al., 2014; Witt & Huerta-Sánchez, 2019).

Positive selection signals and convergent substitutions in hypoxia pathway genes across high-altitude, subterranean and aquatic mammals, including cetaceans, suggest a critical role for HIF1α during mammalian adaptation to hypoxic conditions (Zhu et al., 2018). Specifically, among cetaceans, single substitutions were found: 578T in HIF1α and 451S in HIF2α were distributed within the ODDD/N-TAD domain, and 140S in PHD3 was within the β2β3 loop, which participates in binding to target prolines (Zhu et al., 2018).

Another study in cetaceans, characterizes animal diving behaviors and molecular functions of key factors of the hypoxia signaling pathway in three species: sperm whale, beluga whale, and Yangtze finless porpoise. Based on HIF1α amino acid sequence differences, it is suggested that the response of sperm whale HIF1α to different oxygen tensions is more variable than that of the two other HIF1α species. Thus, indicating that the

rate of degradation of HIF1α may be more variable in long-duration divers than in short-duration diver cetaceans (Bi et al., 2015).

Regarding other aquatic mammals, studies indicate that seals possess a *HIF1α* gene similar in sequence to terrestrial mammal gene, but with several amino acid differences in the oxygen-dependent degradation domain (Johnson et al., 2005); it is suggested that HIF expression has a protective role against oxidative damage in seal tissues (Johnson et al., 2004). Additionally, have been demonstrated the HIF1α activation in elephant seal muscle in response to prolonged fasting and repetitive sleep apneas (Vázquez-Medina et al., 2011; Soñanez-Organis et al., 2013).

Nevertheless, despite the efforts that resulted in the scientific papers cited above, there is still a significant knowledge gap about the diversity of genes related to hypoxia tolerance in cetaceans and other aquatic mammals. In part, this is because these studies involved a small number of species and also a limited set of candidate genes. Here, we outline a study that initially aims to investigate genes of the more restrict HIF signaling pathway and network (e.g., *HIF1α, HIF2α, PHD2, VHL EGLN2, EGLN1,* and *EGLN3, PHD2, PHD3*), for later advances to investigations regarding genes from other networks and correlated pathways. In other words, describing how a group of genes and its taxon-specific variants are connected with hypoxia adaptation in cetaceans and other aquatic mammals.

1.5. SARS-COVID-2: an urgent demand of investigation

The global outbreak of COVID-19, which emerged in late 2019, had profound impacts on societies and human health worldwide. In response to this crisis, the academic community has rallied like never before to address the challenges posed by the virus and explore potential solutions (Haghani & Bliemer, 2020; Fassin, 2021). This mobilization has included an unprecedented effort in genomic surveillance, bolstered by advancements in sequencing technology and the availability of research resources. Consequently, there has been a surge in scientific publications covering various aspects of the virus, including its basic biology, epidemiology, evolution, and clinical and public health implications. A similar surge in research activity has been observed among groups involved in viral genomic surveillance in

Brazil.

In this setting, while the primary focus of this thesis remains on the theoretical aspects outlined regarding the Hypoxia signaling pathway and its potential implications in the evolution and adaptation of aquatic mammals, the pressing needs of the COVID-19 pandemic context prompted a parallel inquiry into the subject of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This investigation unfolded concurrently with ongoing research in an unrelated area, underscoring the dynamic nature of scientific exploration amidst urgent global challenges.

1.6. COVID-19: introduction to the disease

COVID-19, stemming from SARS-CoV-2 infection, was first documented in December 2019. By March 2020, the virus had reached 197 countries, prompting the World Health Organization (WHO) to declare it a global pandemic. To date, the pandemic remains a significant concern, with over 770 million reported COVID-19 cases and 7.0 million reported deaths worldwide (WHO, 2024a).

The incubation period ranges from 2 to 14 days, during which asymptomatic individuals can spread the virus to others. Symptoms typically manifest five days after exposure. While most SARS-CoV-2 infections result in mild illness, affecting primarily upper airway cells, some individuals develop severe pneumonia, driven by immunopathology in the lower respiratory tract. Those with severe COVID-19 often exhibit predispositions leading to inadequate immune responses, particularly in type I or type III interferon reactions. Alveolar damage may occur directly from AT2 cell infection or indirectly from local inflammation. This inflammation induces a 'leaky state' in both the epithelium and endothelium, exacerbating inflammation and coagulation, with monocytes, macrophages, and neutrophils playing pivotal roles in amplifying pro-inflammatory and/or profibrotic responses (Lamers & Haagmans, 2022; WHO, 2024b).

Uncontrolled inflammation ultimately contributes to severe immunopathology characteristic of COVID-19. Typically, individuals seek hospital care around the eleventh day

of infection, experiencing symptoms such as fever, fatigue, muscle pain, nausea, diarrhea, headache, weakness, nasal discharge, loss of smell, and taste. Pneumonia may develop, leading to SARS and potentially requiring intubation, depending on disease severity. Severe cases may also involve damage to the liver, kidneys, heart, and occasionally the nervous system (Lamers & Haagmans, 2022; Menezes et al., 2022; WHO, 2024b).

1.7. SARS-CoV-2

SARS-CoV-2 is responsible for the COVID-19 pandemic. Belonging to the Coronaviridae family, specifically as a sarbecovirus with an RNA genome, it can cause a range of illnesses from mild cold-like symptoms to severe respiratory syndromes similar to SARS and Middle East Respiratory Syndrome (Lamers & Haagmans, 2022; Tosta et al., 2023). The virus is believed to have originated from bats and might have involved an intermediary host in its zoonotic transmission (Benvenuto et al., 2020).

The SARS-CoV-2 RNA genome encodes various structural and non-structural proteins, and mutations in these proteins result in diverse molecular alterations, leading to differences among virus variants (Arya et al., 2021; Sun et al., 2022).

1.7.1. SARS-CoV-2 RNA genome and proteins

SARS-CoV-2 carries a single-stranded RNA genome spanning approximately 30 kilobases (kb), exhibiting around 80% sequence homology with its counterpart, SARS-CoV. Among its genetic repertoire are structural proteins like the membrane, nucleocapsid, envelope, and spike glycoproteins, alongside non-structural proteins, primarily involved in viral replication and transcription, and accessory proteins. These structural components, in conjunction with a host-derived lipid bilayer, compose the enveloped virion, facilitating the transport of viral genomic RNA into host cells. While accessory proteins aren't vital for replication, they contribute to other facets of the virus's lifecycle (Lamers & Haagmans, 2022).

SARS-CoV-2's genomic RNA encompasses 14 open reading frames (ORFs), with two

major ORFs, ORF1a and ORF1b, overlapped by a (-1) ribosomal frame-shift mechanism. These ORFs, constituting a significant portion of the genome, give rise to polyproteins pp1a and pp1ab, respectively. The subsequent cleavage of these polyproteins by viral proteases, including the papain-like and main proteases, generates a range of non-structural proteins (Nsps), from Nsp1 to Nsp16. Some of these Nsps, along with host factors, assemble into a replication-transcription complex housed within double-membrane vesicles, serving as the epicenter for viral genome replication and transcription. The remaining segment of the genome harbors overlapping ORFs responsible for encoding essential structural proteins such as spike (S), membrane (M), envelope (E), and nucleocapsid (N), as well as various accessory proteins like ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF9, and ORF10 (Arya et al., 2021).

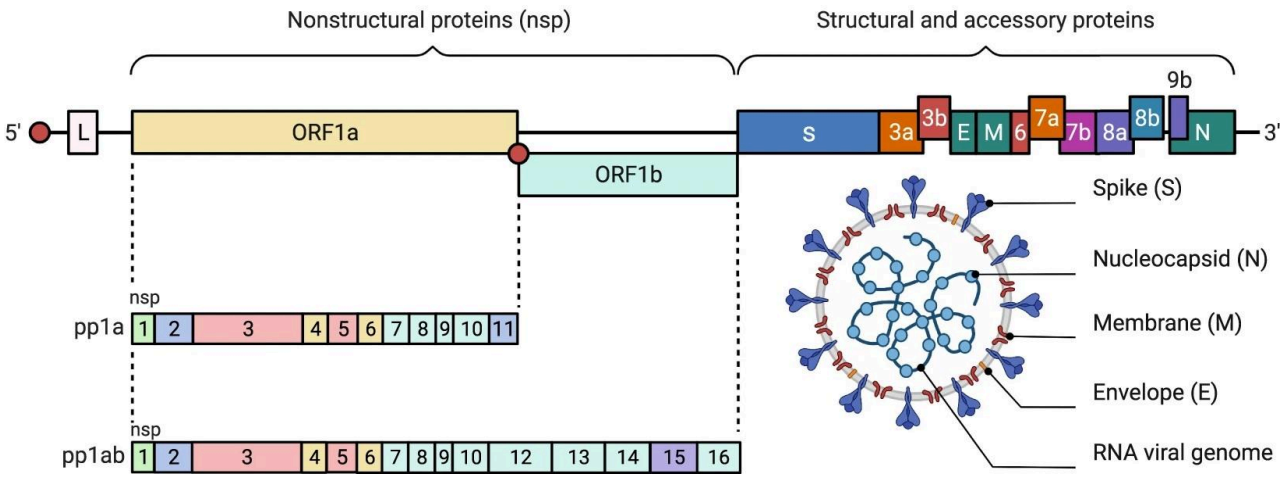

Figure 4. Linear genome architecture of encoded viral protein and structural overview of SARS-CoV-2 (From Sun et al., 2022).

1.7.2. SARS-CoV-2 infection mechanisms

During viral infection, SARS-CoV-2 introduces its genome into the host cell via either endosomes or direct fusion of the viral envelope with the host cell membrane. This process is facilitated by the binding of the spike (S) protein to the human angiotensin-converting enzyme 2 (ACE2) on the cell surface (steps 1–2, Figure 5). Upon entry into the host cell, the viral genomic RNA (gRNA) is released into the cytoplasm and translated by host ribosomes (steps 3–4). Sixteen non-structural proteins (Nsps), derived from polyproteins pp1a and

pp1ab, are released through proteolytic cleavage by cysteine proteases located within Nsp3 (papain-like protease; PLpro) and Nsp5 (chymotrypsin-like protease) (step 5). Nsps, along with other factors, assemble to form the replication-transcription complex (RTC) within the infected host cell. While some Nsps play supportive roles, others provide essential enzymatic functions for viral genome replication and transcription within the RTC. Initially, the positive RNA strand is replicated to a negative strand, which is then utilized for either further replication or transcription of sub-genomic mRNAs (steps 6–7). These sub-genomic mRNAs encode structural proteins (S, M, E, N) and accessory proteins (step 8). The S, M, and E proteins localize to the endoplasmic reticulum (ER), while the N protein binds to the genomic RNA to form a nucleoprotein complex. This complex, along with structural proteins, migrates to the ER-Golgi intermediate compartment (ERGIC) where virions assemble, mature, and bud off from the Golgi in vesicles (steps 9–11). Subsequently, these vesicles travel to the host cell membrane and are released into the extracellular region via exocytosis (step 12, Figure 5). Released virions then infect new cells, perpetuating disease progression (Arya et al., 2021; V'kovski et al., 2021).

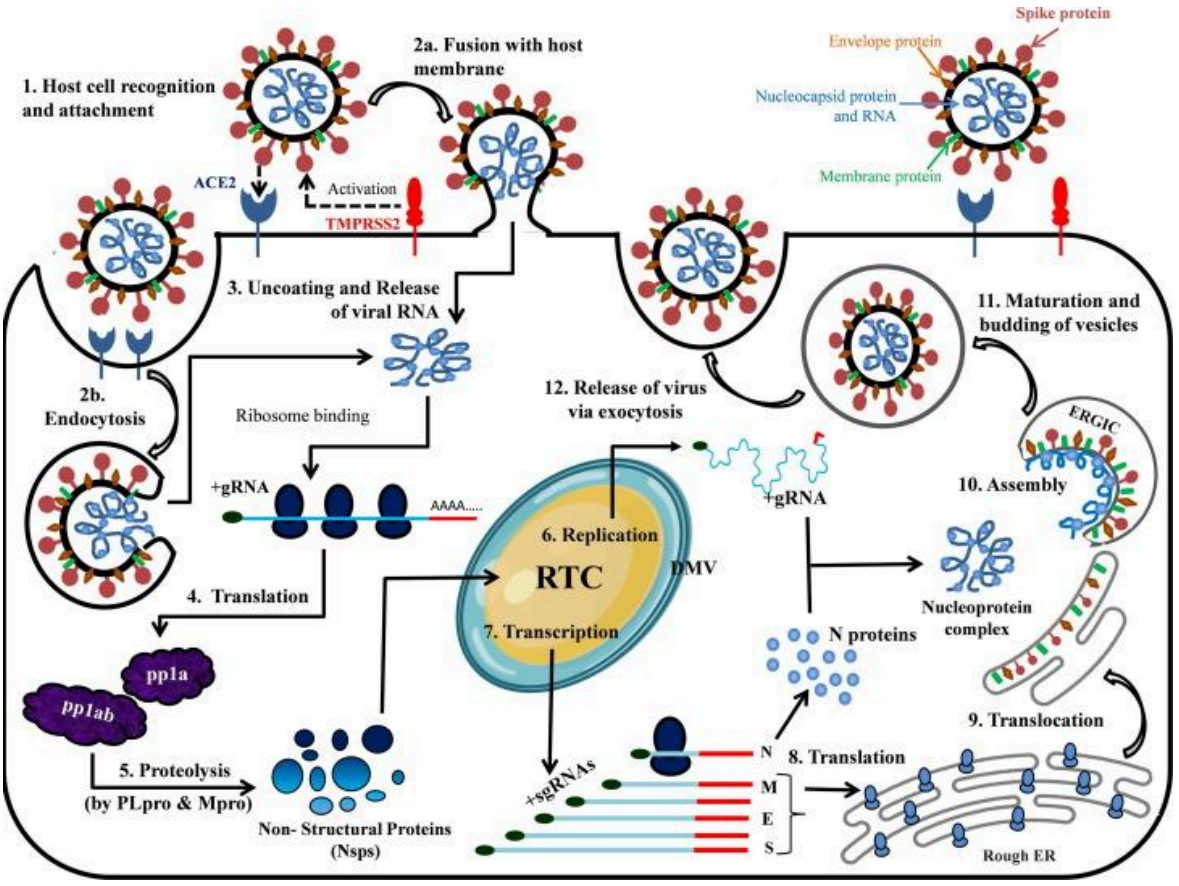

Figure 5. The infection cycle of SARS-CoV-2 inside the host cell. The sequence of events, from host cell recognition through the release of new virion, is represented graphically as steps 1 to 12 (From Arya et al. 2021).

1.7.2.1. Spike protein

The spike protein of SARS-CoV-2 plays a pivotal role in facilitating efficient human-to-human transmission due to its essential properties. This major structural protein is incorporated into the viral membrane in a homo-trimeric configuration. It comprises two subunits: the S1 subunit, which binds to the host entry receptor angiotensin-converting enzyme 2 (ACE2), and the S2 subunit, responsible for mediating membrane fusion. The S1–S2 site, separating these subunits, contains a furin cleavage motif and undergoes cleavage in the virus-producing cell (Sun et al., 2022; WHO, 2024b).

Upon binding to ACE2 on the target cell, the spike protein is cleaved at the S2′ site by the transmembrane serine protease TMPRSS2. This cleavage activates the S2 subunit trimers, facilitating the fusion of viral and host lipid bilayers and releasing the viral ribonucleoprotein complex into the cell. Alternatively, the virus may enter cells via endosomes, where cathepsins can cleave the spike protein, although this pathway is less efficient in primary epithelial cells (WHO, 2024b).

Critical regions within the distal S1 subunit include the receptor binding domain (RBD) and the N-terminal domain (NTD). The RBD, specifically, serves as the binding site for ACE2, making it a key determinant of virus-host interaction and a vulnerable target for antibody neutralization. Most neutralizing antibodies or vaccines are designed to target the RBD to prevent or inhibit viral infection. Additionally, the binding of RBD with ACE2 necessitates conformational changes, and a spike protein that readily transitions from a "closed" to "open" conformation facilitates viral infection. Consequently, mutations in the spike protein of SARS-CoV-2 variants can significantly impact the conformational structure of the spike protein and alter its interaction with ACE2 or neutralizing antibodies (Sun et al., 2022).

1.7.2. SARS-CoV-2 variants and transmission

In late 2020, after circulating within the human population for nearly a year, SARS-CoV-2 underwent a significant evolutionary shift, resulting in highly mutated forms with increased transmission rates compared to earlier variants. These variants were identified as 'variants of concern' (VOCs) (Sun et al., 2022). Phylogenetic classification serves as a primary method for categorizing emergent SARS-CoV-2 strains, with systems such as the clade-nomenclature system utilized by the Global Initiative of Sharing All Influenza Data (GISAID; Hamed et al., 2021) or the Pango lineage system established by the Pango Network (Rambaut et al., 2020; O'Tool et al., 2022). However, due to the exponential growth in sequenced genomes and broader observation of variant distribution, a more concise naming convention for critical variants was necessary to inform global anti-virus strategies. Consequently, the World Health Organization and the Centers for Disease Control and Prevention (CDC) proposed using the Greek alphabet to designate significant SARS-CoV-2 clades or Pango lineages and introduced the concept of Variant of Concern (VOC), Variant of Interest (VOI), and Variant of High Consequence (VOHC) as part of a comprehensive classification system (Sun et al., 2022). Designated as Alpha, Beta, Gamma, Delta, and Omicron, these VOCs emerged independently and swiftly became dominant, either regionally or globally, supplanting preceding variants (Carabelli et al., 2023).

Five major SARS-CoV-2 variants have been identified as variants of concern (VOCs) by the World Health Organization and national public health agencies due to their substantially altered transmissibility or immune evasion capabilities, necessitating close monitoring. Each VOC demonstrated transmission advantages over earlier variants and gained dominance either regionally, as observed in the cases of Alpha (PANGO lineage10 B.1.1.7), Beta (B.1.351), and Gamma (P.1) originating from the UK, South Africa, and Brazil, respectively, or globally, as seen in Delta (B.1.617.2/AY sublineages) and numerous Omicron sublineages (B.1.1.529/BA sublineages, such as BA.1, BA.2, and BA.5) (Carabelli et al., 2023).

These variants were characterized by an increased number of non-synonymous mutations, primarily in the spike protein, along with distinct phenotypic properties such as

altered transmissibility and antigenicity, enabling evasion of a primed immune response. This dynamic interaction has driven evolutionary pressure, akin to the classic Red Queen competition, where viruses and hosts engage in a continual struggle for survival, leading to the emergence of variants with heightened transmissibility and/or reduced sensitivity to neutralizing antibodies (Jacob et al., 2023).

1.7.2.1. VOCs in Brazil

According to a recapitulation by Menezes et al. (2022), SARS-CoV-2 genomic surveillance in Brazil began in São Paulo in February 2020, following the return of international travelers. However, a study detected SARS-CoV-2 mRNA sewage in Santa Catarina in November 2019, indicating community transmission had already begun. Strains from clade B.1, characterized by the D614G mutation in the Spike protein, dominated prevalence studies during the first wave in 2020, notably B.1.1.28 and B.1.1.33. The earliest re-infection cases were reported in São Paulo and Rio de Janeiro, with co-infections documented in Rio Grande do Sul in 2020. In Brazil, new lineages surged in November-Dezember 2020, coinciding with the World Health Organization's identification of Variants of Interest (VOIs) and Variants of Concern (VOCs). The VOI Zeta was first detected in Rio de Janeiro in July 2020 and became the most prevalent variant until February 2021, associated with co-infections in multiple states. The VOC gamma emerged in January 2021, coinciding with a surge in COVID-19 mortality. VOC delta was first confirmed in Brazil in August 2021, replacing Gamma and leading to a substantial decrease in cases and deaths until December 2021. Other variants, including Mu, Lambda, Alpha, and Beta, were also detected but did not circulate widely. Studies in 2021 focused on re-infections and vaccine breakthrough infections, with reports linking them to specific variants. The emergence of VOC Omicron was observed at the end of November 2021, with significant circulation reported by December 2021. In 2022, Brazil experienced its third and fourth waves, with daily reported cases exceeding 250,000 in January. Numerous studies characterized Omicron circulation across various states.

Chapter 2. Justificative and Objectives

Justificative and general objective

The groundbreaking discovery of the Hypoxia-Inducible Factor (HIF) and its subsequent recognition with the 2019 Nobel Prize in Physiology marked a significant advancement in our understanding of cellular responses to hypoxia. That work aims to delve into the intricate molecular mechanisms governing how animal cells sense and respond to fluctuations in oxygen levels, particularly within the context of hypoxia tolerance in vertebrates. While hypoxia-tolerant species like aquatic mammals have emerged as valuable models for studying adaptive responses to low oxygen levels, the era of genomics presents unprecedented potential to uncover the genetic foundations of hypoxia tolerance. Therefore, this study seeks to explore the molecular evolution of genes within the HIF signaling pathway, focusing on aquatic mammals, to unravel their associations with hypoxia adaptation and tolerance. Additionally, given the unique context of the global pandemic caused by the SARS-CoV-2 virus, a portion of this thesis research will be allocated to examining variants or lineages derived from the virus prevalent in Brazil during a specific time period.

Specific objectives

1. To examine genetic variability within the coding regions of the hypoxia signaling pathway genes in cetaceans and other aquatic mammals.
2. To infer molecular evolutionary patterns of hypoxia signaling pathway genes in cetaceans and other aquatic mammals.
3. To identify if there is molecular convergence between species from different taxa, related to hypoxia tolerance in aquatic environments.
4. To identify the possible relationship between the functional variability of several gene groups and adaptation to the aquatic environment.
5. To determine in greater detail as possible the evolutionary trajectory of SARS-CoV-2 VOC Gamma (P.1) in Brazil and its derived sub-lineages, P.1.1, P.1.2.

Chapter 3. Results - Adaptive Strategies of Aquatic Mammals: Exploring the Role of the HIF Pathway and Hypoxia Tolerance

Manuscript submitted to and accepted by the *Genetics and Molecular Biology Journal*
https://doi.org/10.1590/1678-4685-GMB-2023-0140

Yuri Yépez[1], Mariana Marcano-Ruiz[1], Maria Cátira Bortolini[1].

[1]Laboratório de Evolução Humana e Molecular, Departamento de Genética,Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Corresponding author:
Maria Cátira Bortolini, ORCID 0000-0003-0598-3854
Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Departamento de Genética. Av. Bento Gonçalves, 9500. Prédio 43323, Porto Alegre, RS, Brazil.
E-mail: maria.bortolini@ufrgs.br.

Key words: HIF Pathway, HIF3A, HIF3α, Aquatic Mammals, hypoxia tolerance.

Abstract

Aquatic mammals (marine and freshwater species) share significant and similar adaptations, enabling them to tolerate hypoxia during regular breath-hold diving. Despite the established importance of *HIF1A*, a master regulator in the molecular mechanism of hypoxia response, and other associated genes, their role in the evolutionary adaptation of aquatic mammals is not fully understood. In this study, we investigated this topic by employing a candidate gene approach to analyze 11 critical genes involved in the *HIF1A* signaling pathway in aquatic mammals. Our gene analyses included evaluating positive and negative selection, relaxation or constriction of selection, and molecular convergence compared to other terrestrial mammals, including subterranean mammals. Evidence of selection suggested a significant role of negative selection, as well as relaxation of the selective regime in

cetaceans for most of these genes. We found that the glutamine 68 variant in the HIF3α protein is unique to cetaceans and initial evaluations indicated a destabilizing effect on protein structure. However, further analyses are necessary to evaluate its functional impact and adaptive relevance in this taxon.

Introduction

Aquatic mammals, including cetaceans, sirenians, and carnivores such as pinnipeds and some mustelids, have developed significant and similar adaptations to live in aquatic environments, despite their distinct evolutionary origins (Berta et al., 2015; Fordyce, 2018; Berta et al., 2018). These adaptations, including anatomical, physiological, and biochemical changes, allow them to tolerate low oxygen tension or hypoxia, which is necessary for regular breath-hold diving, and permits them to withstand lower blood oxygen tension than most terrestrial mammals (Ponganis, 2011).

The physiological and biochemical mechanisms underlying the immersion behavior of aquatic mammals have been extensively explored in previous studies (Allen and Vázquez-Medina, 2019; McKnight et al., 2019). Moreover, genomic investigations encompassing individual species and cross-lineage comparisons among marine mammals have unveiled various evolutionary adaptations to the aquatic environment. These adaptations encompass hypoxia tolerance, specialized energy metabolism, responses to oxidative stress, modifications in body plan, and anatomical changes (Yim et al., 2014; Fan et al., 2019; Foote et al., 2015; Zhou et al., 2015; Chikina et al., 2016).

Despite these advancements, our comprehension of the genetic mechanisms governing hypoxia tolerance in aquatic mammals remains incomplete. To shed light on this aspect, it is essential to consider the role of the transcription factor HIF (Hypoxia Induced Factor), which assumes a central position in the hypoxia response across metazoans (Schofield and Ratcliffe, 2004; Kaelin and Ratcliffe, 2008). HIF consists of one of three α subunits (HIF1α, HIF2α, and HIF3α) and a common β subunit (HIF1β or ARNT), encoded by the genes *HIF1A*, *HIF2A*,

HIF3A, and *ARNT*, respectively. HIF's primary function is to activate the transcription of genes that enhance oxygen delivery or facilitate metabolic adaptation to hypoxic conditions.

Under normal oxygen levels, specific prolyl residues and one asparaginyl residue within HIFα undergo hydroxylation (Schofield and Ratcliffe, 2004; Kaelin and Ratcliffe, 2008). Three HIF prolyl-hydroxylases (PHDs 1–3), encoded by the genes *EGLN1*, *EGLN2*, and *EGLN3*, hydroxylate the prolyl residues, while the factor inhibiting HIF (FIH1), encoded by the *HIF1AN* gene, hydroxylates the asparaginyl residue. Hydroxylated prolyl residues serve as binding sites recognized by the von Hippel-Lindau protein (encoded by *VHL*), leading to the degradation of HIFs through the ubiquitin-proteasome pathway. The hydroxylation of the asparaginyl residue causes steric hindrance, preventing the binding of specific transcriptional co-activators. However, under hypoxic conditions, the hydroxylation of HIFα by PHDs is inhibited, resulting in the accumulation of HIF. This accumulated HIF translocates into the cell nucleus, where it forms dimers with HIF1β and binds to specific DNA motifs of target genes. This binding activates the transcription of these target genes, like *VEGFA*, which encodes a growth factor that plays a crucial role in promoting vascular endothelial cell proliferation (see Figure S1 for a visual representation of this process).

Multiple studies in human populations, and high altitude plateau and fossorial mammals have provided robust evidence that the hypoxia signaling pathway (also HIF pathway) plays a crucial role in hypoxia tolerance in low oxygen environments (Xiao et al., 2017; Witt and Huerta-Sánchez, 2019). As a result, it is not surprising that HIFs and other proteins that interact with these transcription factors are involved in mammalian adaptation to aquatic environments. Convergent evolution has been observed in three genes of HIF pathway (*HIF1A*, *HIF2A*, and *EGLN3*), as demonstrated in a comparison between five cetacean species and high-altitude mammals (Zhu et al., 2018).

Further insight into the functional adaptations of pathway proteins in cetacean species has been gleaned from in vitro studies, revealing variable responses to different oxygen tensions. For example, differences in HIF1α were observed in three cetacean species (Bi et al., 2015), and unique functions of VHL associated with HIF2α degradation were reported in

beluga whales (Bi et al., 2017). In vivo studies in elephant seals (*Mirounga angustirostris*) have shown HIF1α stabilization in response to fasting and sleep apneas, further supporting the significance of the HIF pathway in hypoxia tolerance (Vázquez-Medina et al., 2011).

As we have access to an expanding list of aquatic mammal genomes, we can gain a more comprehensive understanding of the genes that are involved in the adaptation to hypoxia associated with an aquatic lifestyle in mammals. To this end, we propose a study that examines the molecular evolutionary patterns of the coding region of 11 genes that encode proteins of the HIF pathway, by comparing the aquatic mammal sequences (marine and freshwater species) with those of terrestrial mammals, including hypoxia tolerant fossorial species.

Materials and Methods

Hypoxia signaling pathway genes and genomic data collection

We searched for 11 genes involved in the evolutionarily conserved cellular signaling pathway of responses to hypoxia, *ARNT (*also termed as *HIF1β*), *ARNT2* (*HIF2β*), *HIF1AN* (*FIH1*), *HIF1A* (*HIF1α*), *EPAS1* (*HIF2α*), *HIF3A (HIF3α*), *EGLN1*, *EGLN2*, *EGLN3*, *VHL*, and *VEGFA* (Supplementary Table S1; Supplementary Figure S1). We gathered genetic and functional information from the GeneCards Database (2023) and consulted the gene interaction prediction program STRING v11.0 (2023) to analyze gene interactions.

We obtained the orthologous coding sequences (CDS) of 11 genes from the NCBI Orthologs database. We utilized data from 21 NCBI annotated genomes and CDS from 35 terrestrial mammal species (Supplementary Table S2). Additionally, we incorporated assembled genome data from 19 aquatic mammal species, which lacked annotation in the NCBI database. To enhance our dataset, we included draft genome assemblies from 11 aquatic mammal species that were publicly available, as provided by the DNA ZOO consortium (Dudchenko et al., 2017).

To retrieve the orthologous CDS, we employed a BLAST-based approach. Subsequently, we utilized the Geneious software (Kearse et al., 2012) to map the sequences from these genomes with the CDS references (Supplementary Table S3), resulting in contiguous consensus sequences.

Sequences alignments

We used GUIDANCE2 (Landan and Graur, 2008; Sela et al., 2015) with the PRANK codon alignment algorithm to align the CDS of the genes. We eliminated any sequences or columns with uncertain confidence scores (threshold <6) from the original dataset before conducting further analyses.

Phylogenetic trees

We reconstructed phylogenetic trees from the 11 HIF pathway gene alignments using Bayesian inference with BEAST v.1.10.4 (Suchard et al., 2018) for the subsequent selection analyses.

Selection analyses

To identify natural selection signals, we employed several statistical tests based on ω values. Site-level model analyses (M2 and M8 models) using PAML v.4.9 CODEML software (Yang, 2007) were conducted, as well as the CODEML branch-site Test 2 to detect positive selection along a specific branch of the tree. The study also used HyPhy 2.5.31 software (Kosakovsky et al., 2020) with the RELAX method, FUBAR, MEME, and BUSTED to identify positive selection in genes where neither the site nor the branch-site methods detected positive selection signals.

Selection Pressure Analysis in Cetacean Clades with Contrasting Diving Abilities

We analyzed selection pressure in cetacean clades with different diving abilities using the nested branch model approach proposed by Tian et al. (2016). Species of cetacean and pinnipeds were classified into deep-divers (≥300m of common depth and ~40 minutes of average dive time) and shallow-divers (common depth around 100m in shallow waters; Supplementary Table S4).

Molecular convergent analysis

To identify convergent amino acid substitutions in aquatic mammals, we employed the Profile Change with One Change (PCOC) method (Rey et al., 2018). This method employs a criterion that characterizes convergent shifts as substitutions occurring on all branches where a change in phenotype has occurred, and these substitutions align with a modification in the preferred type of amino acid at that specific position.

Structural analyses

We accessed predicted three-dimensional protein structures of HIF pathway proteins from the AlphaFold database (Varadi et al., 2022). To visualize and map positively selected sites onto these three-dimensional protein structures, we employed the UCSF ChimeraX 1.5 software (Pettersen et al., 2021). Additionally, to predict the impact of specific variants on protein stability, we conducted ΔΔG (Gibbs free energy of unfolding) analyses using the DDGun3D software (Montanucci et al., 2019).

Additional information

For a more comprehensive explanation of these and previous analyses, and detailed information about genes and sequences, please refer to the Supplementary materials and methods section. We also utilized OpenAI's GPT-3.5 model to proofread and enhance the English phrasing and grammar in our manuscript.

Results

We conducted a candidate gene approach to analyze 11 HIF pathway genes CDS across over 86 mammalian species in order to improve our understanding of their evolution. Our analysis compared species with aquatic lifestyles, including 35 cetacean species, 10 pinniped species, 6 Lutrinae species, and one manatee species, with those with terrestrial habits, such as subterranean species of rodents.

Relaxation analysis

To investigate the possible changes in selective pressures that occurred in HIF pathway genes during the transition to an aquatic lifestyle, we used the HyPhy RELAX method. Initially, we performed a joint analysis of the four aquatic mammal taxa, followed by separate analyses of each taxa. Our results provide strong evidence of changes in selective regimes, including both relaxation ($K < 1$) and intensification ($K > 1$) of selection. When considering the set of aquatic lineages, we observed relaxation in four genes: *ARNT*, *ARNT2*, *EGLN2*, and *HIF1AN* (Supplementary Table S5). Further analysis by taxa revealed relaxation in six genes, including the four genes mentioned above, *EPAS1* and *EGLN1*, and evidence of intensification in *VEGFA*, all in cetaceans. In the Pinnipedia clade, evidence of relaxation was detected in one gene, *HIF1AN*. The Lutrinae group, on the other hand, exhibited an intensification of the selective pressure in *EGLN3*. Interestingly, no evidence of change in selective pressures was observed in any HIF pathway genes for *Trichechus manatus*.

Signatures of positive selection

To gain insight into the impact of selection on the evolution of HIF pathway genes in aquatic mammal lineages, we utilized a gene-wide test for episodic diversification, BUSTED, with aquatic taxa specified as foreground branches. Positive results from this test would indicate that at least one site was positively selected in at least one period of time on the specified branches. The BUSTED analysis revealed evidence of selection at low frequencies

of sites (0.01% to 1.55%) in only three genes (*ARNT, ARNT2, EPAS1*) among aquatic mammals (see Supplementary Table S6).

Although BUSTED is technically a Branch-site test, it is not designed to detect individual sites under selection. Therefore, we implement different methods based on codon substitution to detect positive selection at the site level. Our analysis using CODEML models M2 and M8 showed that most genes fit better with the null models M1a and M7, respectively. However, positive selection was detected at seven sites in *HIF3A* and *VHL* by the M2 model, while the less conservative M8 model identified 21 sites with positive selection signature in *ARNT*, *HIF1AN*, *HIF1A*, and *EPAS1*. The sites under positive selection as identified by the BEB analysis are presented in Supplementary Table S7.

Using the MEME and FUBAR, we identified a total of 136 codons from 10 genes with evidence of positive selection (Supplementary Table S8). Specifically, FUBAR identified 16 positively selected sites in 7 genes, while MEME detected 134 sites in 10 genes (except *HIF1AN*). Notably, MEME is designed to detect both episodic and pervasive selection, whereas FUBAR only identifies sites under pervasive positive selection (Berta et al., 2015). Thus, the difference in the number of sites detected by each method is likely due to a large proportion of sites detected by MEME being associated with episodic selection events. We identified 24 sites from 7 genes (5 in *HIF3A*, 7 in *VHL*, 3 in *ARNT*, 1 in *HIF1A*, 3 in *EPAS1*, 3 in *EGLN2*, 2 in *VEGFA*) as robustly under positive selection, as they were identified by at least two methods at the site level (Table 1). We identified large portions of each gene under negative selection using FUBAR, refer to Supplementary Table S9 for details.

To confirm positive selection and investigate specific signals in individual sites across aquatic mammal taxa Cetacea (order Artiodactyla), Pinnipedia (order Carnivora), and Lutrinae (Mustelidae; Carnivora) compared to their terrestrial counterparts, we employed a Branch-site method known as Test 2 or the branch-site test of positive selection, implemented in CODEML. Our analysis identified only one site, at position 68, of HIF3α protein under episodic selection in the cetacean lineage (Table 1). This site was also detected by MEME, with support from a branch, and it has an exclusive arginine residue in this group. No amino

acid residues unique to aquatic mammal lineages were observed at sites with evidence of positive selection found by more than one method, except for this case.

Additionally, we found two different sites, 437 and 195, in HIF3α and VHL, respectively, that are unique to cetaceans and hippopotamuses, both of which are members of the suborder Whippomorpha. These sites did not exhibit amino acid residues exclusive to subterranean Rodentia species (e.g., *Fukomys damarensis*, *Nannospalax galili*, and *Heterocephalus glaber*), or even of the other Artiodactyla species, such as *Vicugna pacos* (family Camilidae) adapted to live in Andean high-altitude, where the hypoxic environment is also well documented.

Branch models and diving contrasts

To investigate whether the selective pressures acting on HIF pathway genes differed between cetaceans and pinnipeds with distinct abilities to dive and remain submerged, we employed CODEML branch models using the nested design of Tian et al. (2016). In cetaceans, we classified the species from the taxonomic families Balenopteridae, Balaenidae, Eschrichtiidae, Lipotidae, Phocoenidae, Iniidae, Pontoporiidae, and three species from the family Delphinidae as deep-divers, and species from the families Ziphiidae, Physeteridae, Monodontidae, and three additional species from Delphinidae as shallow-divers. In pinnipeds, we classified species from the family Phocidae as shallow-divers, and those from the Otariidae and Odobenidae families as deep-divers (Supplementary Table S4).

The LRTs showed significant differences between the 1ω vs. 2ω models for four genes in cetaceans and eight genes in pinnipeds, indicating that these two aquatic taxa have experienced different selective pressures compared to their terrestrial counterparts. Supplementary Table S10 provides the corresponding values. Notably, pinnipeds displayed significant changes in selective pressures towards higher values of ω when compared to terrestrial lineages and when comparing deep-diving and shallow-diving pinnipeds. The *ART2* analysis revealed a ω value of 0.0188 for deep-diving pinnipeds, which is considerably higher than that of shallow-diving pinnipeds (Figure 1). However, there were no significant

differences when considering ancestral and descendant branches (see Supplementary Table S10).

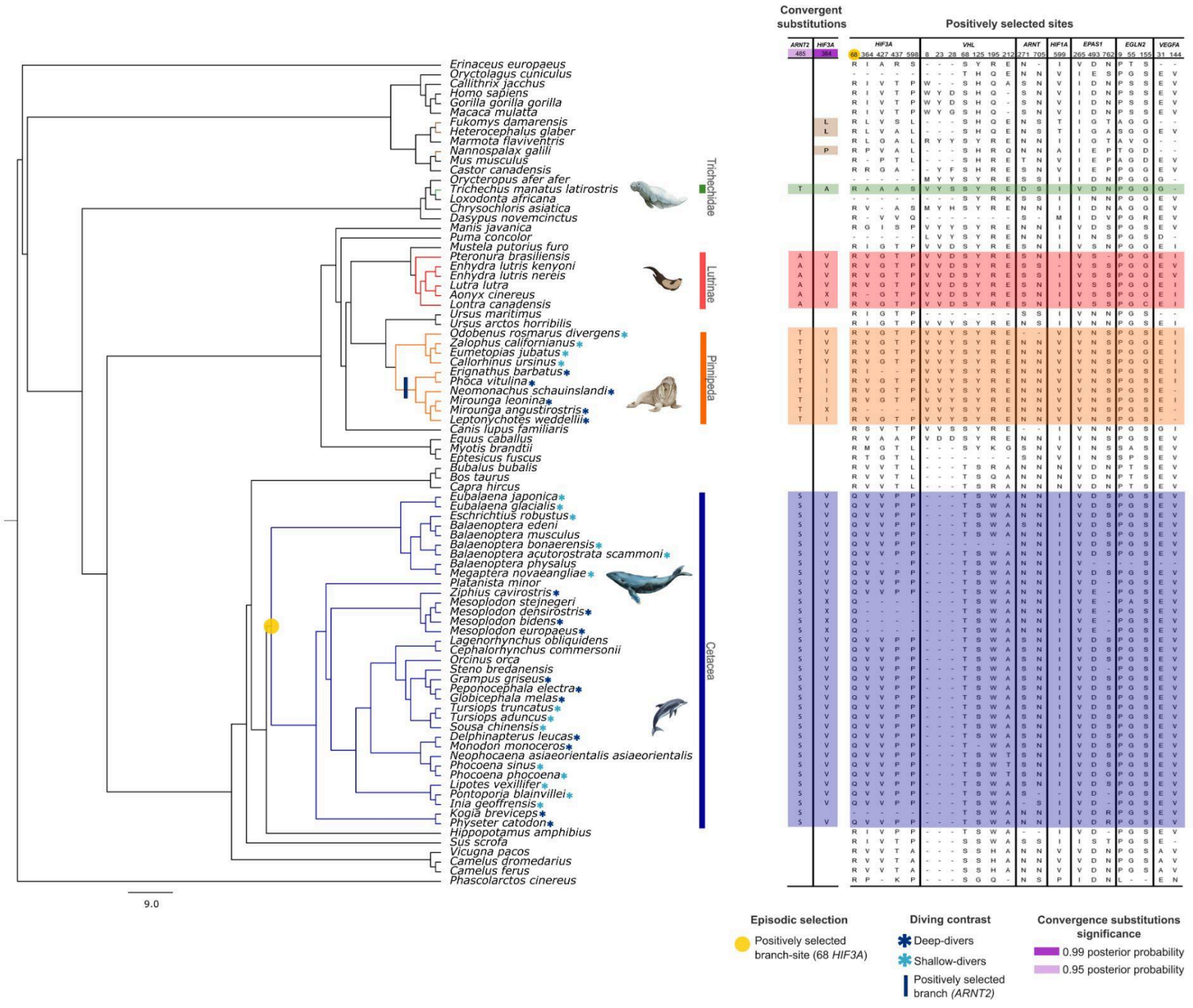
Convergent substitutions
Positively selected sites
ARNT2
HIF3A
HIF3A
VHL
ARNT
HIF1A
EPAS1
EGLN2
VEGFA
Erinaceus europaeus
Oryctolagus cuniculus
Callithrix jacchus
Homo sapiens
Gorilla gorilla gorilla
Macaca mulatta
Fukomys damarensis
Heterocephalus glaber
Marmota flaviventris
Nannospalax galili
Mus musculus
Castor canadensis
Orycteropus afer afer
Trichechus manatus latirostris
Loxodonta africana
Chrysochloris asiatica
Dasypus novemcinctus
Manis javanica
Puma concolor
Mustela putorius furo
Pteronura brasiliensis
Enhydra lutris kenyoni
Enhydra lutris nereis
Lutra lutra
Aonyx cinereus
Lontra canadensis
Ursus maritimus
Ursus arctos horribilis
Odobenus rosmarus divergens
Zalophus californianus
Eumetopias jubatus
Callorhinus ursinus
Erignathus barbatus
Phoca vitulina
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Leptonychotes weddellii
Canis lupus familiaris
Equus caballus
Myotis brandtii
Eptesicus fuscus
Bubalus bubalis
Bos taurus
Capra hircus
Eubalaena japonica
Eubalaena glacialis
Eschrichtius robustus
Balaenoptera edeni
Balaenoptera musculus
Balaenoptera bonaerensis
Balaenoptera acutorostrata scammoni
Balaenoptera physalus
Megaptera novaeangliae
Platanista minor
Ziphius cavirostris
Mesoplodon stejnegeri
Mesoplodon densirostris
Mesoplodon bidens
Mesoplodon europaeus
Lagenorhynchus obliquidens
Cephalorhynchus commersonii
Orcinus orca
Steno bredanensis
Grampus griseus
Peponocephala electra
Globicephala melas
Tursiops truncatus
Tursiops aduncus
Sousa chinensis
Delphinapterus leucas
Monodon monoceros
Neophocaena asiaeorientalis asiaeorientalis
Phocoena sinus
Phocoena phocoena
Lipotes vexillifer
Pontoporia blainvillei
Inia geoffrensis
Kogia breviceps
Physeter catodon
Hippopotamus amphibius
Sus scrofa
Vicugna pacos
Camelus dromedarius
Camelus ferus
Phascolarctos cinereus
Trichechidae
Lutrinae
Pinnipeda
Cetacea
9.0
Episodic selection
Positively selected branch-site (68 HIF3A)
Diving contrast
Deep-divers
Shallow-divers
Positively selected branch (ARNT2)
Convergence substitutions significance
0.99 posterior probability
0.95 posterior probability

Figure 1. Positive selection and convergence patterns of HIF pathway genes have been analyzed across the mammalian phylogeny. Aquatic mammal clades have been highlighted. A Bayesian inferred phylogeny of concatenated loci is depicted, with changes across positively selected sites confirmed by two or more natural selection inference analyses. Corresponding amino acids for each species are listed on the right side of the figure. *ARNT2* and *HIF3A* genes have also been shown to exhibit signals of molecular convergence with the PCOC method under two convergence scenarios: considering only aquatic mammals and including aquatic mammals and subterranean rodent species. PCOC model posterior probability values of convergent amino acids are indicated. Asterisks in the species names indicate their diving capabilities classified as deep or shallow divers, for representative species and clades. The blue vertical line indicates evidence of a positively selected branch on *ARNT2* detected using the nested branch model implemented in CODEML, and the yellow circle indicates episodic selection in the cetacean lineage in *HIF3A* detected using MEME and CODEML branch-site Test 2.

Convergence analysis

We utilized the PCOC method to evaluate potential cases of adaptive convergent amino acid evolution across the 11 HIF pathway proteins. This method calculates the posterior probabilities of convergent evolution at each site under the "PCOC" model. We considered two scenarios: the first encompassing general adaptations to an aquatic lifestyle across all aquatic mammal groups, and the second scenario including subterranean rodent species due to their shared phenotypes of hypoxia tolerance due their respective habitats.

In the first scenario, we detected two candidate sites for convergent substitutions in two genes: site 485 in ARNT2 and site 356 in HIF3α (position 364 in human reference sequence). Both sites had a posterior probability greater than 0.95 under the PCOC model (Figure 1). Additionally, site 364 in HIF3α showed positive selection signals detected by three methods. Most amino acids at this site had hydrophobic side chains. In the second scenario that included subterranean rodent species (*Fukomys damarensis*, *Heterocephalus glaber*, and

Nannospalax galili) with a similar phenotype of hypoxia tolerance, we found only one significant convergent substitution, also at site 356 in HIF3α.

Protein structures

We proceeded with the mapping of amino acid changes indicated by positive selection in the predicted 3D protein structures. When we marked putative selected sites detected by more than one codon substitution method, we found that most of these sites were located in disordered regions of the proteins. However, approximately 22% (5/23) of these sites were within or close to domains of these proteins.

Based on the available protein structure and sequence information, we tested the effect of three amino acid changes, R68Q and T437P in HIF3α and Q195W in VHL, on protein stability (Table 2). R68Q, a change unique to cetaceans, occurs in the second alpha helix motif of the bHLH domain of HIF3α (Supplementary Figure S2). This substitution replaces an arginine (R) residue, which has a cationic side chain, with a glutamine (Q) residue, which has a polar uncharged side chain, and was predicted to decrease protein stability ($\Delta\Delta G = -0.1$) by both the sequence-based and structural-based methods DDGun and DDGun3D, respectively. In contrast, we did not observe any specific changes in protein stability due to the T437P substitution (Supplementary Figure S2). Finally, the Q195W substitution, which changes a glutamine to a tryptophan (W) residue with a hydrophobic side chain in VHL, was predicted to stabilize the protein ($\Delta\Delta G = 0.1$) by both methods (Supplementary Figure S3).

Discussion

Relaxed Selection signatures

Shifts in the distribution of selection pressure are often associated with large biological transitions, such as the transition from land to water, which can lead to phenotypic plasticity and adaptive changes (Lahti et al., 2009). In the case of lipid metabolism gene

CYP8B1, for example, signs of relaxation of selection pressure were observed in the common ancestor of cetaceans (Shinde et al., 2019), while the *MB* gene showed relaxation of constraints, reflected in the protein stability, in shallow-diving cetacean species (Holm et al., 2016). The *RAG1* gene also showed a relaxation of selective pressure in cetaceans (Dias and Nery, 2020).

Relaxation of selection pressure can lead to gene pseudogenization by allowing the accumulation of stop codons or other changes in the nucleotide sequence that disrupt the reading frame, leading to gene loss events that may constitute an evolutionary mechanism of adaptation to different niches (Sharma et al., 2018). In cetaceans, gene loss in ancestral species may indeed have contributed to adaptation to the aquatic environment (Huelsmann et al., 2019; Espregueira et al., 2020). Besides, relaxation of selective pressure is also a finding compatible with a scenario of conditional neutrality, a way for changes to accumulate and escape the action of purifying selection, resulting in variability (Raman et al., 2016). Furthermore, there are various ways that selection can be relaxed, including the removal of the a functional constraint, and reduction in efficiency of selection (*e.g.*, increased relevance of genetic drift due to a reduction in effective population size; Charlesworth and Charlesworth 2010, Wertheim et al. 2015).

We found that cetaceans are the representative group of aquatic mammals with the largest number of genes under a relaxation regime (6 of 11 HIF pathway genes studied). Interestingly, when considering all aquatic mammals in relation to their terrestrial relatives, fewer genes with relaxation or constriction signals were evident, and these signals were mainly observed in cetaceans.

One of the genes that exhibits relaxed selection in pinnipeds is *HIF1AN*, also known as *FIH1*. This gene plays a crucial role in directly regulating HIFα, adding an additional layer of control to the HIF response (Baik and Jain, 2020). It complements the function of PHD proteins, which regulate HIFα levels through prolyl hydroxylation. This observation suggests that there might be a less stringent control mechanism in place for regulating HIF1A in cetaceans and pinnipeds.

In a broader context, these findings indicate that the proteins in cetaceans may have initially evolved under different functional constraints compared to those in pinnipeds and Lutrinae. This implies that relaxed selection could have played a significant role in shaping the evolutionary path of HIF pathway genes in cetaceans, particularly.

Adaptive evolution

While several studies have explored various aspects of the evolution of marine mammals, including their genetic and physiological adaptations to aquatic life (Chikina et al., 2016; Yuan et al., 2021), they have typically focused on marine species, leaving freshwater species understudied. Overall, aquatic mammals have evolved a range of adaptations to suit their lifestyles in different environments (Berta et al., 2015). Cetaceans, such as whales and dolphins, are fully aquatic and have streamlined bodies and flippers or flukes for propulsion. Depending on the species, they feed on a variety of organisms like fish, plankton, and other mammals. Sirenians, such as manatees, are also fully aquatic and have large, rounded bodies adapted for buoyancy. They feed on seagrasses and other aquatic plants and are found in shallow coastal waters and rivers. Pinnipeds, including seals and sea lions, are semi-aquatic and divide their time between land and sea, using their four flippers to swim and dive. Lutrinae, such as otters, are semi-aquatic and live in both freshwater and marine environments, using their streamlined bodies and webbed feet for swimming. The last two clades are carnivores and feeding on fish, crustaceans, and mollusks.

Adaptations to their environments have shaped the different lifestyles of aquatic mammals, and identifying the common genetic elements that underlie these adaptations is a major challenge. Coping with hypoxic environments is a crucial adaptation that all aquatic mammals have evolved, enabling them to thrive in their marine habitats. For instance, cetaceans and sirenians experience hypoxia regularly throughout their lives, which has led many cetacean species to develop the ability to hold their breath for extended periods. This adaptation allows them to explore the depths of the ocean and remain submerged for extended periods (Ponganis, 2011).

Similar selective pressures could potentially give rise to comparable molecular patterns at the molecular level. Therefore, the investigation of the HIF pathway genes in aquatic mammals holds significant importance in understanding their evolutionary adaptations and survival strategies in challenging environments. This is due to the pathway's previously established significance at the cellular level. Our analysis of the natural selection regime revealed predominantly purifying selection on the coding regions that characterize these proteins, particularly as transcription factors, which correspond to highly conserved domains within these genes. Among all HIFα proteins, a common feature is the presence of an N-terminal bHLH (basic helix-loop-helix) domain located upstream of two PAS (Per-Arnt-Sim) domains. Mutations occurring in the bHLH and PAS domains are typically disruptive, leading to loss of function (Pamenter et al., 2020). Thus, our findings of the pervasive effect of purifying selection on the HIF pathway genes align with prior research (Rytkönen et al., 2011).

On the other hand, our analysis revealed that most of the sites detected under a positive selection regime suggest probable episodic selection events in different clades across the mammalian phylogeny. Notably, we found evidence pointing to a significant episodic event of positive selection in *HIF3A*, specifically a Glutamine residue at site 68, which is present in all cetaceans, differing from the Arginine present in the rest of the phylogeny. While this represents a conservative replacement, whose effect is predicted to be neutral (Li et al., 1984), our initial approximation approach to assess the structural effect of this variant and its potential impact on HIF3α folding and function using ΔΔG protein analysis, predicted a decrease in the stability of the HIF3α structure due to the Arg68Gln change.

HIF3α position 68 is located within the bHLH domain, composed mainly of basic amino acid residues, such as Arginine, that facilitate binding to the DNA phosphate backbone. The change to glutamine, which is a charge-neutral amino acid, may alter the electrostatic potential of the bHLH domain, likely decreasing the affinity to DNA. HIF3α is a transcription factor involved in the cellular response to hypoxia or low oxygen levels, although its precise role is not yet fully understood. Some studies have suggested that HIF3α may act as a negative regulator of the HIF1A pathway (Suzuki et al., 2017; Cai et al., 2020), which is critical for oxygen homeostasis in cells. Based on the above, we cannot predict or propose the

effect of this variant on the cetacean clade. Thus, further functional studies *in silico* or *in vitro* will be necessary to evaluate in detail the structural role of this change and its adaptive effect in cetaceans to hypoxic environments.

On the other hand, our study differs from Zhu et al. (2018) in several aspects, such as the number and selection of specific sequences used, as well as the number of genes analyzed. Despite these discrepancies, both studies reached a similar conclusion that positively selected sites are located in the same regions within the proteins and that purifying selection had a significant role in the evolution of HIF pathway genes. We confirm that the 140S amino acid of EGLN3 is exclusive to cetaceans, although the protein stability analysis based on ΔΔG did not predict any stability changes.

Deep-diving species

The depth and duration of immersion attained by aquatic mammals are closely linked to their diving capacity (Panneton, 2013). Previous research has shown that positive selection in certain genes is linked to diving abilities in cetaceans (Tian et al., 2016; Tian et al., 2021). We proposed that this hypothesis can be extended to other genes and taxonomic clades. For instance, certain representatives of Ziphiidae and Monodontidae (both cetacean families) and the genus *Mirounga sp*. (Pinnipeda) included in this study exhibit extreme deep-diving foraging behaviors, reaching depths of over 1000 m. The most extreme deep-diver is the beaked whale *Ziphius cavirostris*, which holds the record among mammals, reaching a depth of at least 2992 m and lasting 137.5 minutes (Schorr et al., 2014). Yuan et al. (2021) found evidence of multiple genes related to HIFα were under positive selection or had unique amino acid substitutions in cetacean deep-diving species. Refining our analysis of the HIF pathway genes, we did not find any associations of deep diving in cetaceans. However, we did observe statistically significant dN/dS values on the *ARNT2* gene in pinnipeds, particularly in the deep-diving Phocidae branch, as compared to shallow-diving species. These findings suggest that the common ancestor of Phocidae species may have already possessed this selective trait, possibly indicating their ability for deep-diving.

Phenotypic pattern convergence

The emergence of similar phenotypes in different species may suggest shared evolutionary paths, however, these similarities may also occur in phylogenetically distant species through convergent evolution, as seen in the emergence of hypoxia tolerance in aquatic and subterranean mammals. At the molecular level, parallel molecular evolution refers to independent substitutions at a site resulting in the same derived amino acid in different species, while convergent substitutions refer to independent changes resulting in the same derived amino acid from different ancestral amino acids (Zhang and Kumar, 1997; Sackton and Clark, 2019).

There are various strategies for investigating convergent evolution, and we utilized the "Profile Change with One Change" (PCOC) method proposed by Rey et al. (2018). This method compares species with a convergent phenotype to closely related species with a different ancestral phenotype to identify genetic signatures responsible for convergent evolution. The method analyzes changes in nucleotide or amino acid positions specifically in the branches where the convergent phenotype emerged. The same amino acid does not need to be present in species with similar adaptive traits, as the convergent evolution responsible for genetic changes may involve different amino acids that perform a similar functional role.

Using the PCOC method, we identified two convergent sites: site 485 in ARNT2 and site 364 in HIF3α. Analysis of ARNT2 amino acids in branches corresponding to Pinnipeda, *Trichechus manatus*, and Cetacea revealed the presence of threonine and serine, both having uncharged polar side chains, unlike the hydrophobic alanine present in the remaining phylogeny. Meanwhile, the side chains of amino acids at site 364 in HIF3α were classified as hydrophobic, with the valine amino acid (V) found in most aquatic mammals, isoleucine (I) in otariids (Pinnipedia), and leucine (L) in *Heterocephalus glaber*. The positive selection detected at the HIF3α 364 site by three methods further supports its possible role in convergence among aquatic mammalian lineages and this subterranean rodent species under hypoxia tolerance scenarios.

It should be noted that the inclusion of a broad range of genomes allows for a more comprehensive analysis of the HIF pathway genes involved in hypoxia tolerance. A larger number of species analyzed reduces the probability of inferring molecular convergence between species of aquatic mammals (Thomas et al., 2017). This is because changes or alleles that are observed in only a few species within small datasets may not be exclusive to these groups, creating a sampling issue. Such variants may be present in a larger number of samples or species, making it, at least, uncertain to associate specific variants with adaptive phenotypes. By expanding the range of genomes analyzed, this study aimed to enhance its power to analyze and interpret data from aquatic mammals species. The inclusion of a broad range of genomes enables a more robust analysis of the HIF pathway genes underlying hypoxia tolerance. As an example, the functional characterization of *HIF1A* variants by Bi et al. (2015) was claimed to be exclusive to three cetacean species. However, a broader comparison of this site that includes a greater number of mammal species is enough to disprove this claim.

Regulation mechanisms and hypoxia tolerance

The HIF pathway genes may be involved in various levels of cellular regulation related to hypoxia control. Our study focused solely on analyzing their respective coding sequences and examining how they exhibit selective regimes associated with different capacities for hypoxia tolerance. However, the cellular control of hypoxia may also involve the participation of these genes at other levels of regulation. Hindle (2020) has pointed out that gene regulatory mechanisms affecting the abundance or timing of gene expression could be responsible for physiological differences observed in marine mammal species. This is consistent with various previous studies that indicate that the HIF gene pathway is not "enriched" at the genome level in analyses of convergent phenotypes among marine mammals (Chikina et al., 2016; Yuan et al., 2021).

Therefore, it is possible that regulatory elements of gene expression at the cis or trans level, such as miRNAs, play a crucial role in the precise and fine-tune regulation of the hypoxic response in aquatic mammals. Several studies have highlighted the importance of

"master controllers" of miRNAs in physiological regulation and adaptation to hypoxia (Hadj-Moussa et al., 2022). For instance, miR-204 was found to downregulate the *VEGF* in Nile tilapia, a fish species, by directly acting on its 3′-UTR (Zhao et al., 2014). Another research demonstrated that during hypoxic events, the upregulation of miR-24, which inhibits apoptosis, combined with the downregulation of miR-210, an indirect stabilizer of *HIF1*, could potentially promote neuronal preservation and trigger an adaptive hypoxic response in naked mole rats (Logan et al. 2020). In the case of aquatic mammals, Penso-Dolfin et al. (2020) investigated the role of miRNAs in the adaptive evolution of diving capacities in Weddell seals (*Leptonychotes weddelli*). Through differential expression analysis, they identified potential protective mechanisms in individual tissues, particularly relevant to hypoxia tolerance and anti-apoptotic pathways. Further studies are required to comprehend the role of these molecules in the overall adaptation of mammals to the aquatic environment.

Conclusions

Our investigation of the HIF pathway genes in aquatic mammals attempted to shed light on the molecular adaptations that these animals have undergone to survive in hypoxic environments. The analysis of the natural selection regime revealed predominantly purifying selection on the coding regions that characterize these proteins, which correspond to highly conserved domains within these genes. However, we also found evidence of positive selection events in different clades across the mammalian phylogeny. We suggest that relaxed selection could have played a significant role in the evolutionary trajectory of HIF pathway genes in cetaceans. Notably, we identified an important episodic event of positive selection in *HIF3A*, which is present in all cetaceans. Further studies are needed to better understand the functional implications of these genetic changes and their role in the adaptation of aquatic mammals to their environment. We conclude that there is no unique "adaptive signature" in the evolution of diverse mammalian groups facing comparable selective pressures and that analyzing coding sequences alone may not fully elucidate the complex molecular scenario of adaptive processes, but our research adds a valuable piece to this puzzle. We can illustrate this concept, comparing different genetic elements such as coding changes, gene regulation and other factors to chess pieces. So, the infinite combinations of pieces movements on the chess

board would resemble the multitude of possible combinations of the elements that can arise in the evolution of various mammalian taxa under comparable selective pressures.

Acknowledgments

This research was financially supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We would like to thank Professors Sandro Bonatto and Márcio Dorn for encouraging our research.

Conflict of Interest

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

Authors Contributions

Y.Y. Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing.
M.M.-R. Data curation, Investigation, Writing.
M.C.B. Conceptualization, Investigation, Funding acquisition, Supervision, Project administration, Writing.

References

Allen KN and Vázquez-Medina JP (2019) Natural tolerance to ischemia and hypoxemia in diving mammals: a review. Front Physiol 10: 1199.

Baik AH and Jain IH (2020) Turning the oxygen dial: balancing the highs and lows. Trends in cell biology 30: 516-536.

Berta A, Churchill M and Boessenecker RW (2018) The origin and evolutionary biology of pinnipeds: seals, sea lions, and walruses. Annu Rev Earth Planet Sci 46: 203–228.

Berta A, Sumich JL and Kovacs KM (2015) Marine mammals: evolutionary biology: 3rd edition. Mar Mamm Evol Biol 726.

Bi J, Hu B, Wang J, Liu X, Zheng J, Wang D and Xiao W (2017) Beluga whale pVHL enhances HIF-2α activity via inducing HIF-2α proteasomal degradation under hypoxia. Oncotarget 8:42272–42287.

Bi J, Hu B, Zheng J, Wang J, Xiao W and Wang D (2015) Characterization of the hypoxia-inducible factor 1 alpha gene in the Sperm whale, Beluga whale, and Yangtze finless porpoise. Mar Biol 162:1201–1213.

Cai X, Zhou Z, Zhu J, Liao Q, Zhang D, Liu X, Wang J, Ouyang G and Xiao W (2020) Zebrafish Hif3α modulates erythropoiesis via regulation of gata1 to facilitate hypoxia tolerance. Development 147.

Charlesworth B, Charlesworth D (2010) Elements of evolutionary genetics. Roberts and Company Publishers, Greenwood Village.

Chikina M, Robinson JD and Clark NL (2016) Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. Mol Biol Evol 33:2182–2192.

Dias BC and Nery MF (2020) Analyses of RAG1 and RAG2 genes suggest different evolutionary rates in the Cetacea lineage. Mol Immunol 117:131–138.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP et al. (2017) De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356:92–95.

Espregueira Themudo G, Alves LQ, Machado AM, Lopes-Marques M, da Fonseca RR, Fonseca M, Ruivo R and Castro LFC (2020) Losing genes: the evolutionary remodeling of Cetacea skin. Front Mar Sci 7:912.

Fan G, Zhang Y, Liu X, Wang J, Sun Z, Sun S, Zhang H, Chen J, Lv M, Han K et al. (2019) The first chromosome‐level genome for a marine mammal as a resource to study ecology and evolution. Mol Ecol Resour 19:944–956.

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.

Fordyce RE (2018) Cetacean evolution. In: Wursig B, JGM T and Kovacs K (eds) Encyclopedia of marine mammals. 3rd edition. Academic Press 180-185.

Hadj-Moussa H, Hawkins LJ and Storey KB (2022) Role of microRNAs in extreme animal survival strategies. Methods Mol Biol 2257:311–347.

Hindle AG (2020) Diving deep: understanding the genetic components of hypoxia tolerance in marine mammals. J Appl Physiol 128:1439–1446.

Holm J, Dasmeh P and Kepp KP (2016) Tracking evolution of myoglobin stability in cetaceans using experimentally calibrated computational methods that account for generic protein relaxation. Biochim Biophys Acta 1864:825–834.

Huelsmann M, Hecker N, Springer MS, Gatesy J, Sharma V and Hiller M (2019) Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. Sci Adv 5:6671–6696.

Kaelin WG and Ratcliffe PJ (2008) Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol Cell 30:393–402.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647.

Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A et al. (2020) HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol 37:295–299.

Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K and Foster SA (2009) Relaxed selection in the wild. Trends Ecol Evol 24:487–496.

Landan G and Graur D (2008) Local reliability measures from sets of co-optimal multiple sequence alignments. Pac Symp Biocomput 15–24.

Li WH, Wu CI and Luo CC (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71.

Logan SM, Szereszewski KE, Bennett NC, Hart DW, van Jaarsveld B, Pamenter ME and Storey KB (2020) The brains of six African mole-rat species show divergent responses to hypoxia. J Exp Biol.

McKnight JC, Bennett KA, Bronkhorst M, Russell DJF, Balfour S, Milne R, Bivins M, Moss SEW, Colier W, Hall AJ et al. (2019) Shining new light on mammalian diving physiology using wearable near-infrared spectroscopy. PLOS Biol 17:e3000306.

Montanucci L, Capriotti E, Frank Y, Ben-Tal N and Fariselli P (2019) DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. BMC Bioinformatics 20:1–10.

Pamenter ME, Hall JE, Tanabe Y and Simonson TS (2020) Cross-species insights into genomic adaptations to hypoxia. Front Genet 11:743.

Panneton WM (2013) The mammalian diving response: an enigmatic reflex to preserve life? Physiology (Bethesda) 28: 284-97.

Penso-Dolfin L, Haerty W, Hindle A and Di Palma F (2020) MicroRNA profiling in the Weddell seal suggests novel regulatory mechanisms contributing to diving adaptation. BMC Genomics 21:1–17.

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH and Ferrin TE (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 30:70–82.

Ponganis PJ (2011) Diving mammals. Compr Physiol 1: 447–465.

Raman AS, White K I and Ranganathan R (2016) Origins of allostery and evolvability in proteins: a case study. Cell, 166: 468–481.

Rey C, Guéguen L, Sémon M and Boussau B (2018) Accurate detection of convergent amino-acid evolution with PCOC. Mol Biol Evol 35:2296–2306.

Rytkönen KT, Williams TA, Renshaw GM, Primmer CR and Nikinmaa M (2011) Molecular evolution of the metazoan PHD-HIF oxygen-sensing system. Mol Biol Evol 28:1913–1926.

Sackton TB and Clark N (2019) Convergent evolution in the genomics era: new insights and directions. Philos. Trans. R. Soc. B 374: 20190102.

Schofield CJ and Ratcliffe PJ (2004). Oxygen sensing by HIF hydroxylases. Nature reviews Molecular cell biology, 5(5), 343-354.

Schorr GS, Falcone EA, Moretti DJ and Andrews RD (2014) First long-term behavioral records from Cuvier's Beaked whales (Ziphius cavirostris) reveal record-breaking dives. PLoS One 9:e92633.

Sela I, Ashkenazy H, Katoh K and Pupko T (2015) GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res 43:W7.

Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE and Hiller M (2018) A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. Nat Commun 2018 91 9:1–9.

Shinde SS, Teekas L, Sharma S and Vijay N (2019) Signatures of relaxed selection in the CYP8B1 gene of birds and mammals. J Mol Evol 87:209–220.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ and Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol.

Suzuki N, Gradin K, Poellinger L and Yamamoto M (2017) Regulation of hypoxia-inducible gene expression after HIF activation. Exp Cell Res 356:182–186.

Thomas GWC, Hahn MW and Hahn Y (2017) The effects of increasing the number of taxa on inferences of molecular convergence. Genome Biol Evol 9:213–221.

Tian R, Geng Y, Guo H, Yang C, Seim I and Yang G (2021) Comparative analysis of the superoxide dismutase gene family in Cetartiodactyla. J Evol Biol 34:1046–1060.

Tian R, Wang Z, Niu X, Zhou K, Xu S and Yang G (2016) Evolutionary genetics of hypoxia tolerance in cetaceans during diving. Genome Biol Evol 8:827–839.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50:D439–D444.

Vázquez-Medina JP, Zenteno-Savín T, Tift MS, Forman HJ, Crocker DE and Ortiz RM (2011) Apnea stimulates the adaptive response to oxidative stress in elephant seal pups. J Exp Biol 214:4193–4200.

Wertheim JO, Murrell B, Smith MD, Kosakovsky SL, and Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. Molecular biology and evolution, 32(3), 820-832.

Witt KE and Huerta-Sánchez E (2019) Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos Trans R Soc B Biol Sci 374:20180235.

Xiao B, Wang S, Yang G, Sun X, Zhao S, Lin L, Cheng J, Yang W, Cong W, Sun W et al. (2017) HIF-1α contributes to hypoxia adaptation of the naked mole rat. Oncotarget 8:109941–109951.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591.

Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee J-H, Yang EC, Kwon KK et al. (2014) Minke whale genome and aquatic adaptation in cetaceans. Nat Genet 46:88–92.

Yuan Y, Zhang Y, Zhang P, Liu C, Wang J, Gao H, Rus Hoelzel A, Seim I, Lv M, Lin M et al. (2021) Comparative genomics provides insights into the aquatic adaptations of mammals. Proc Natl Acad Sci USA 118:e2106080118.

Zhang J and Kumar S. (1997) Detection of convergent and parallel evolution at the amino acid sequence level. Mol. Biol. Evol. 14: 527–536.

Zhao Y, Zhu CD, Yan B, Zhao JL and Wang ZH (2014) miRNA-directed regulation of VEGF in tilapia under hypoxia condition. Biochem Biophys Res Commun 454:183–188.

Zhou X, Seim I and Gladyshev VN (2015) Convergent evolution of marine mammals is associated with distinct substitutions in common genes. Sci Rep 5:16550.

Zhu K, Ge D, Wen Z, Xia L and Yang Q (2018) evolutionary genetics of hypoxia and cold tolerance in mammals. J Mol Evol 86:618–634.

Internet resources

UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://www.uniprot.org/ (April 14, 2023).

Supplementary material - the following online material is available for this article:

Table S1 – Hypoxia signaling pathway genes summary.

Table S2 – Species names and NCBI accession numbers for each sequence.

Table S3 – Coding sequences of phylogenetic closely related species used as references to mapping genomes without annotation.

Table S4 – Aquatic mammals classification according to diving ability.

Table S5 – Presence of selection relaxation and constraint on aquatic mammals.

Table S6 – Parameter estimates of BUSTED method for HIF pathway genes.

Table S7 – Selective pressure analyses of HIF pathway genes by PAML site model method.

Table S8 – Positively selected sites from FUBAR and MEME.

Table S9 – FUBAR evidence of site purifying selection.

Table S10 – Nested Branch-level model.

Figure S1 – HIF signaling pathway (Kegg: ko04066).

Figure S2 – Molecular visualization of mutations (sites 68 and 437) in the AlphaFold three-dimensional model of HIF3A.

Figure S3 – Molecular visualization of mutation (site 195) in the AlphaFold three-dimensional model of VHL.

This material is available as part of the online article from http://www.scielo.br/gmb

Supplementary Information

Material and methods

Hypoxia signaling pathway genes

We searched for 11 genes involved in the evolutionarily conserved cellular signaling pathway of responses to hypoxia (Kegg: ko04066), *ARNT (*also termed as *HIF1β*), *ARNT2* (*HIF2β*), *HIF1AN* (*FIH1*), *HIF1A* (*HIF1α*), *EPAS1* (*HIF2α*), *HIF3A (HIF3α*), *EGLN1*, *EGLN2*, *EGLN3*, *VHL*, and *VEGFA* (Supplementary Table S1; Supplementary Figure S1). *EGLN2*, *EGLN1*, and *EGLN3* encode the proline hydroxylases PHD1, PHD2, and PHD3, respectively. *ARNT2* encodes a transcription factor that forms heterodimers with HIFα in the nucleus of the cell and belongs to the basic helix-loop-helix (bHLH)/PAS (bHLH-PAS) protein superfamily (O'Leary et al., 2016). To better understand the role of hypoxia in vessel growth and the control over angiogenic pathways, we also screened the Vascular endothelial growth factor-A gene (*VEGFA*), a growth factor involved in vascular endothelial cell proliferation and migration that binds to heparin (O'Leary et al., 2016).

Genomic data collection

We downloaded the orthologous coding sequences (CDS) of 11 genes from the hypoxia signaling pathway available from the NCBI Orthologs database. We used the sequences of 21 annotated genomes, including species such as Baiji (*Lipotes vexillifer*), California sea lion (*Zalophus californianus*), and Florida manatee (*Trichechus manatus latirostris*), among others. We also included CDS from 35 species of terrestrial mammals, such as mole rats (*Heterocephalus glaber*) and Middle East blind mole-rat (*Nannospalax galili*), which are known for their tolerance to hypoxic environments. A detailed list of the species names and NCBI accession numbers for each sequence can be found in Supplementary Table S2.

We used nucleotide sequences that were comparable in length and had the highest protein domain conservation to the corresponding human canonical sequences reported in Uniprot among the gene sequences corresponding to the alternative transcripts. We performed an initial comparison by aligning the amino acid sequences using COBALT (Papadopoulos and Agarwala, 2007).

To retrieve orthologous sequences for aquatic mammals with assembled genomes lacking NCBI annotation, we used the BLAST server. BLASTN and TBLASTX were employed to search for sequences in the phylogenetically closest organisms, and we selected sequences with alignment scores of at least 200, percent identity greater than 95%, and statistically significant e-values. We obtained gene sequences for 19 mostly cetacean species, including fin whale (*Balaenoptera physalus*), Antarctic minke whale (*Balaenoptera bonaerensis*), humpback whale (*Megaptera novaeangliae*), gray whale (*Eschrichtius robustus*), North Pacific right whale (*Eubalaena japonica*), pygmy sperm whale (*Kogia breviceps*), Northern elephant seal (*Mirounga angustirostris*), sea otter (*Enhydra lutris nereis*), Boutu (*Inia geoffrensis*), Franciscana (*Pontoporia blainvillei*), Indus river dolphin (*Platanista minor*), Indo-Pacific bottlenose dolphin (*Tursiops aduncus*), Cuvier's beaked whale (*Ziphius cavirostris*), Sowerby's beaked whale (*Mesoplodon bidens*), Indo-Pacific humpbacked dolphin (*Sousa chinensis*), Harbour porpoise (*Phocoena phocoena*), Eurasian river otter (*Lutra lutra*), the giant otter (*Pteronura brasiliensis*), and hippopotamus

(*Hippopotamus amphibius*), which is the phylogenetically closest terrestrial relative to cetaceans.

In addition, we included draft genome assemblies of 11 aquatic mammal species generated by the DNA ZOO consortium and publicly available (Dudchenko et al., 2017). These species are Blainville's beaked whale (*Mesoplodon densirostris*), Gervais' beaked whale (*Mesoplodon europaeus*), Stejneger's beaked whale (*Mesoplodon stejnegeri*), Commerson's dolphin (*Cephalorhynchus commersonii*), melon-headed whale (*Peponocephala electra*), rough-toothed dolphin (*Steno bredanensis*), Risso's dolphin (*Grampus griseus*), Bryde's whale (*Balaenoptera edeni*), North Atlantic right whale (*Eubalaena glacialis*), bearded seal (*Erignathus barbatus*), and Asian small-clawed otter (*Aonyx cinereus*). We downloaded these genomic sequences and added them to our database. We then used the BLAST+ 2.2.29 tool (Camacho et al., 2009) to run MEGABLAST and TBLASTX on a local server and search for orthologous sequences of HIF genes in these species.

To visualize the aligned sequences obtained from the BLAST search, we used Geneious v.7.1.3 (Kearse et al., 2012). Geneious was used to map the sequences to reference sequences, using default parameters. The coding sequences of closely related species were used as reference sequences, which are listed in Supplementary Table S3.

Sequences alignments

To align the CDS of the hypoxia signaling pathway genes, we utilized GUIDANCE2 (Landan and Graur, 2008; Sela et al., 2015). This software provided a measure of confidence for the alignments to ensure that they were reliable for phylogenetic analysis and identifying positive selection signals. We used the PRANK codon alignment algorithm and obtained alignment confidence scores, which helped us identify and remove ambiguous regions.

Phylogenetic trees

We reconstructed phylogenetic trees from the 11 *HIF* signaling pathway genes using both Bayesian inference with BEAST v.1.10.4 (Suchard et al., 2018) and maximum likelihood

with RAxML 8.2.12 (Stamatakis, 2014). To determine the best-fitting nucleotide substitution models for each dataset, we used Jmodeltest v.2.1.10 (Darriba et al., 2012) with the Akaike information criterion (AIC). For the Bayesian analysis, we used the Birth and Death model as the tree prior (Gernhard, 2008) and ran 20,000,000 generations with the Markov Chain Monte Carlo (MCMC) algorithm (Brooks et al., 2011). We discarded the first 2,000,000 trees as the burn-in and generated a "Maximum clade credibility" tree (MCC tree) using TreeAnnotator, which is part of the BEAST package. We visualized and analyzed the MCMC outputs using Tracer v.1.7.1 (Rambaut et al., 2018b). For the maximum likelihood analysis, we used RAxML with default parameters (GTR+GAMMA) and 1000 bootstraps, which were performed using the CIPRES Science Gateway (Miller et al, 2010). Finally, we viewed and edited the resulting trees using FigTree v1.4.4 (Rambaut and Drummond, 2018a).

The two methods used to reconstruct gene trees, Bayesian inference and Maximum likelihood, largely produced congruent topological relationships between species for most genes. However, the Bayesian inference method was found to produce more phylogenetically consistent topology for the HIF pathway genes when compared to the species tree obtained from TimeTree (Kumar et al., 2017). It should be noted that some of the species analyzed in this study were not included in the TimeTree available species phylogeny. Therefore, the concatenated gene tree generated by Bayesian inference was utilized for the selection analyses.

Selection analyses

To identify natural selection signals in the protein-coding sequences of HIF pathway genes in marine mammals, we employed several statistical tests based on ω values. ω represents the ratio of the rate of non-synonymous substitutions (dN) to the rate of synonymous substitutions (dS), providing a measure of the intensity and direction of natural selection. A value of $\omega < 1$ indicates purifying selection, where natural selection reduces the fixation of deleterious mutations, $\omega = 1$ indicates neutral evolution, where nonsynonymous substitutions are neutral, and $\omega > 1$ indicates positive selection, where advantageous mutations become fixed.

Selection analyses were conducted using the PALM v.4.9 (Yang, 2007) CODEML program with the PAMLX graphical interface (Xu and Yang, 2013), as well as four statistical methods from the HyPhy 2.5.31 software (Weaver et al., 2018; Kosakovsky et al., 2020): RELAX to detect relaxed selection in a codon-based phylogenetic framework (Wertheim et al., 2015), FUBAR (Fast, Unconstrained Bayesian AppRoximation for Inferring Selection) (Murrell et al., 2013), MEME for detection of individual sites subject to episodic diversifying selection (Murrel et al., 2012), and BUSTED (Branch-site Unrestricted Statistical Test of Episodic Diversification (Murrel et al., 2015).

First, a site-level model analysis was conducted using the CODEML program. This involved comparing neutral models with alternative models that allow positive selection and performing a pairwise comparison of two likelihood ratio tests (LRTs) with a chi-square distribution. Specifically, the M1 (nearly neutral) vs M2 (positive selection) and M7 (beta distribution, $\omega > 1$ disallowed) vs M8 (beta distribution, $\omega > 1$ allowed) models were compared at a significance level of p-value 0.05. The Bayes empirical Bayes (BEB) approach was used to test the posterior probability of each site in the alternative model and to select potentially positively selected sites. Sites with posterior probabilities greater than 90% were considered to be under positive selection.

In addition to the site-level models, we utilized the CODEML branch-site Test 2 to identify instances of positive selection along specific branches of the phylogenetic tree. This analysis involved comparing Model A (model = 2, Nsites = 2, fix_omega = 0, and omega = 0) with the null model (model = 2, Nsites = 2, fix_omega = 1, and omega = 1) to ascertain whether positive selection was occurring along a particular branch. For each individual test, we assigned Cetacea, Pinnipedia, and Lutrinae as the foreground branches, respectively, and contrasted them with terrestrial lineages. To assess the statistical significance, we applied Likelihood Ratio Test (LRT).

Using Hyphy, various statistical approaches were employed to assess selection hypotheses. The first method utilized was the RELAX method, which aimed to evaluate selective pressure and detect signals of relaxation or intensification of selection strength in HIF signaling pathway genes. Since it focused on testing those branches corresponding to aquatic species, the following lineages were defined as test branches: Cetacea, Pinnipedia,

Lutrinae, and the only studied member of Sirenia, *Trichechus manatus*. The set of branches corresponding to the species of terrestrial mammals was established as reference branches. We use Phylotree to label the test and reference branches in the phylogeny. The K statistic was used to measure the strength of selection in each tested branch. A K value less than 1 indicated a relaxation of selective pressure, where the test ω values trend towards neutrality compared to the reference branches, while a K value greater than 1 indicated intensification of natural selection along the tested branches. Statistical significance was assessed using the LRTs.

Additionally, we implemented two HyPhy site-level methods: FUBAR and MEME. FUBAR is a Bayesian method that detects pervasive selection at the site-level, assuming that the selection pressure for each site is constant along the entire phylogeny. It uses a large number of predefined dN and dS values, which cover the range of negative, neutral, and positive selection scenarios, and which are applied across sites. FUBAR identifies sites that are positively or negatively selected if the posterior probabilities are high enough (≥ 0.9). On the other hand, MEME uses a mixed-effects maximum likelihood approach to calculate, for each site, two ω rate classes and corresponding weights that represent the probability that the site evolves under each rate class at a given branch. Statistical significance between null and alternative model fits is obtained using LRTs. MEME is designed to detect not only pervasive selection but also transient or episodic selection signals. BUSTED, a gene-wide method, was also used to detect some evidence of positive selection in genes where neither the site nor the branch-site methods detected signals, likely because the signals, although present, were weak.

It is worth noting that the choice of the appropriate method for each analysis should depend on the study's specific research question and data set. While HyPhy offers a wide range of methods, the simultaneous use of all available methods is not always advisable. Although a consensus between different methods can help to reduce the likelihood of false positives, using less sensitive methods can reduce statistical power. In this study, we aimed to provide a comprehensive list of sites that may be under positive selection, and differences in the sites detected between FUBAR and MEME could indicate sites that may be subject to episodic selection. Overall, the application of multiple complementary methods can provide a

more reliable assessment of selective pressures acting on a gene or genes associated with a specific cellular pathway.

Selection Pressure Analysis in Cetacean Clades with Contrasting Diving Abilities

We built upon the nested branch model approach proposed by Tian et al. (2016) to analyze selection pressure in cetacean clades with different diving abilities. We extended the study to include a larger number of cetacean species as well as pinnipeds. Our methodology involved employing variable model 2 of CODEML branch models, which allows for multiple ω ratios. This enabled us to create distinct branch groups with different ω values, leading to the '2ω ratio,' '3ω ratio,' and '5ω ratio.' For investigations focused specifically on clades with contrasting diving abilities, we employed two separate trees for each analysis. These trees excluded branches corresponding to Lutrinae and Sirenia. In one tree, cetaceans were designated as the foreground lineage, while in the other, pinnipeds took on this role. Terrestrial relatives were considered background lineages in both trees.

Our analyses entailed comparing several models: one ratio model, which enforces the same ω ratio for all lineages; 2ω model, allowing one ω ratio for all terrestrial mammal branches and another for all branches of aquatic families; 3ω model, assuming independent rates for ancestral and descendant branches of long-diving and short-diving aquatic clades, with a third ω value for all terrestrial mammal branches; 5ω model, estimating independent rates for descendant branches from their ancestral branches in long-diving and short-diving families, compared with the 2ω model. Positive selection inference was carried out through Likelihood Ratio Tests (LRT), where we compared two pairs of models.

Our classification of species was based on their diving abilities, with deep-divers defined as those capable of reaching depths of nearly 300m or more, with an average dive time of 40 minutes, and shallow-divers as those that stay in shallow waters and typically dive to depths of around 100m. The classification can be located within the Supplementary Table S4, where we have documented four parameters for each species: common immersion time, maximum immersion time, common depth, and maximum depth. Based on the lack of information to support differential classification of species into two categories of diving

ability, the taxonomic group Lutrinae and *Trichechus manatus* were not included in the analysis, as all members were classified as shallow-divers.

Molecular convergent analysis

To identify convergent amino acid substitutions in aquatic mammals, we employed the Profile Change with One Change (PCOC) (Rey et al., 2018) Bayesian approach. Unlike methods that look for independent substitutions for the same amino acid, PCOC is more permissive and considers similar amino acids as convergent. This method defines convergent shifts based on amino acid profiles. Besides, PCOC compares species with a convergent phenotype to closely related species with a different phenotype, analyzing changes in nucleotide or amino acid positions specifically in the branches where the convergent phenotype emerged (Rey et al., 2018). We assessed whether an amino acid site evolved convergently by evaluating the posterior probabilities.

Structural analysis

We utilized the AlphaFold database (Varadi et al., 2022) to access predicted three-dimensional protein structures of *HIF* pathway proteins with sites detected on positive selection by at least 2 methods of natural selection and with residues unique to aquatic mammal taxonomic clades. In order to assess the potential functional significance of the selected sites, we mapped them onto the three-dimensional protein structures and visualized them using UCSF ChimeraX software (Pettersen et al., 2021).

The impact of amino acid sequence variations on protein stability is typically quantified by the Gibbs free energy of unfolding (ΔΔG). This measure is obtained by calculating the difference between the free energy of the mutated protein structure and that of its wild-type form ($\Delta\Delta G = \Delta G_M - \Delta G_W$). We used ΔΔG analyses to predict the effect of presumably unique residues of aquatic mammalian taxonomic clades at sites indicated with positive selection. We employed the DDGun3D software (Montanucci et al., 2019) to perform the predictions, using human canonical sequences as wild-type and mutated sequences with

the respective changes noted above. To check the Uniprot accession numbers referred to using the AlphaFold database, see Table S1.

Finally, we utilized OpenAI's GPT-3.5 model (OpenAI, 2021) to proofread and enhance the English phrasing and grammar in our manuscript.

References

Akbar M, Chaudhry AA and Arshed MJ (2004) Effect of water depth and river width on Indus dolphin population. Pak J Life Soc Sci 2:33-35.

Baumgartner MF and Mate BR (2003) Summertime foraging ecology of North Atlantic right whales. Mar Ecol Prog Ser 264:123–135.

Bird DJ, Hamid I, Fox-Rosales L and Van Valkenburgh B (2020) Olfaction at depth: Cribriform plate size declines with dive depth and duration in aquatic arctoid carnivorans. Ecol Evol 10:6929–6953.

Black NA (1994) Behavior and ecology of Pacific white-sided dolphins (Lagenorhynchus obliquidens) in Monterey Bay, California. Doctoral dissertation, San Francisco State University.

Bloodworth BE and Odell DK (2008) Kogia breviceps (Cetacea: Kogiidae). Mamm Species 819:1–12.

Bodkin JL, Esslinger GG, and Monson DH (2004) Foraging depths of sea otters and implications to coastal marine communities. Mar Mammal Sci 20:305–321.

Brooks S, Gelman A, Jones GL and Meng XL (2011) Handbook of Markov Chain Monte Carlo. Handb Markov Chain Monte Carlo 1–592.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Coscarella MA, Pedraza SN and Crespo EA (2010) Behavior and seasonal variation in the relative abundance of Commerson’s dolphin (Cephalorhynchus commersonii) in northern Patagonia, Argentina. J Ethol 28:463–470.

Crespo FA and De Cidre L (2005) Functional significance of bronchial sphincters in two Southwestern Atlantic Dolphins: Pontoporia blainvillei and Lagenorhynchus obscurus - a comparative approach. Mammalia 69:233–238.

Croll DA, Acevedo-Gutiérrez A, Tershy BR and Urbán-Ramírez J (2001) The diving behavior of blue and fin whales: is dive duration shorter than expected based on oxygen stores? Comp Biochem Physiol A Mol Integr Physiol 129:797–809.

Darriba D, Taboada GL, Doallo R and Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 2012 98 9:772–772.

de Ferran, V., Figueiró, H. V., de Jesus Trindade, F., Smith, O., Sinding, M. H. S., Trinca, C. S., ... & Eizirik, E. (2022). Phylogenomics of the world's otters. *Current Biology*, *32*(16), 3650-3658.

DeAngelis AI, Valtierra R, Van Parijs SM and Cholewiak D (2017) Using multipath reflections to obtain dive depths of beaked whales from a towed hydrophone array. J Acoust Soc Am 142:1078–1087.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP et al. (2017) De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356:92–95.

Duplaix N (1980) Observations on the ecology and behavior of the Giant River Otter Pteronura brasiliensis in Suriname. Rev d'Écologie (La Terre La Vie) 34:495–620.

Edwards HH, Martin J, Deutsch CJ, Muller RG, Koslovsky SM, Smith AJ and Barlas ME (2016) Influence of manatees' diving on their risk of collision with watercraft. PLoS One.

Friedlaender AS, Goldbogen JA, Nowacek DP, Read AJ, Johnston D and Gales N (2014) Feeding rates and under-ice foraging strategies of the smallest lunge filter feeder, the Antarctic minke whale (Balaenoptera bonaerensis). J Exp Biol 217:2851–2854.

Gernhard T (2008) The conditioned reconstructed process. J Theor Biol 253:769–778.

Gillespie D, Dunn C, Gordon J, Al ;, Claridge D, Embling C and Boyd I (2009) Field recordings of Gervais' beaked whales Mesoplodon europaeus from the Bahamas. J Acoust Soc Am 125:3428–3433.

Gonçalves LR, Augustowski M and Andriolo A (2016) Occurrence, distribution and behaviour of Bryde's whales (Cetacea: Mysticeti) off south-east Brazil. J Mar Biol Assoc United Kingdom 96:943–954.

Gregr EJ and Coyle KO (2009) The biogeography of the North Pacific right whale (Eubalaena japonica). Prog Oceanogr 80:188–198.

Hamilton CD, Kovacs KM and Lydersen C (2018) Individual variability in diving, movement and activity patterns of adult bearded seals in Svalbard, Norway. Sci Reports 81 8:1–17.

Heyning J, and Dahlheim M (1988) Orcinus orca. Mamm Species 304: 1-9.

Hooker SK, and Baird RW (1999) Observations of Sowerby's beaked whales, Mesoplodon bidens, in the Gully, Nova Scotia. Canadian Field-Naturalist, 113: 273-277.

Hung N and Law CJ (2016) Lutra lutra (Carnivora: Mustelidae). Mamm Species 48:109–122.

Ishii M, Murase H, Fukuda Y, Sawada K, Sasakura T, Tamura T, Bando T, Matsuoka K, Shinohara A, Nakatsuka S et al. (2017) Diving behavior of Sei Whales Balaenoptera borealis relative to the vertical distribution of their potential prey. Mammal Study 42: 1-9.

Jacobsen KO, Marx M and Øien N (2004) Two-way trans-atlantic migration of a north atlantic right whale (Eubalaena glacialis). Mar Mammal Sci 20:161–166.

Jay CV, Farley SD and Garner GW (2006) Summer diving behavior of male walruses in Bristol bay, Alaska. Mar Mammal Sci 17:617–631.

Kato H and Perrin WF (2018) Bryde's Whales: Balaenoptera edeni/brydei. In: Wursig B, JGM T and Kovacs K (eds) Encyclopedia of marine mammals. 3rd edition. Academic Press 143-145.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647.

Kenney RD (2009) Right Whales: Eubalaena glacialis, E. japonica, and E. australis. Encycl Mar Mamm 817–822.

Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A et al. (2020) HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol 37:295–299.

Kruuk H, Wansink D and Moorhouse A (1990) Feeding patches and diving success of otters, Lutra lutra, in Shetland. Oikos 57:68.

Kruuk, H. (2006). *Otters: ecology, behaviour and conservation*. Oxford University Press, USA.

Kumar S, Stecher G, Suleski M and Blair Hedges S (2017) TimeTree: A resource for timelines, timetrees, and divergence times. Mol Biol Evol 34:1812–1819.

Landan G and Graur D (2008) Local reliability measures from sets of co-optimal multiple sequence alignments. Pac Symp Biocomput 15–24.

Miller MA, Pfeiffer W and Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew Comput Environ Work GCE 2010.

Montanucci L, Capriotti E, Frank Y, Ben-Tal N and Fariselli P (2019) DDGun: An untrained method for the prediction of protein stability changes upon single and multiple point variations. BMC Bioinformatics 20:1–10.

Morales López M, Olivera-Gómez LD and Zenteno-Ruiz CE (2012) Intervalo respiratorio y desplazamientos de manatíes antillados Trichechus manatus manatus (Sirenia): Comparación entre las temporadas seca y lluviosa en la Laguna Aislada. Mastozoología Neotrop 19:117–126.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL and Scheffler K (2013) FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Mol Biol Evol 30:1196–1205.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM et al. (2015) Gene-wide identification of episodic selection. Mol Biol Evol 32:1365–1371.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K and Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764.

Noren SR, Jay C V., Burns JM and Fischbach AS (2015) Rapid maturation of the muscle biochemistry that supports diving in Pacific walruses (Odobenus rosmarus divergens). J Exp Biol 218:3319–3329.

Norris TA, Littnan CL, Gulland FMD, Baker JD and Harvey JT (2017) An integrated approach for assessing translocation as an effective conservation tool for Hawaiian monk seals. Endanger Species Res 32:103–115.

Papadopoulos JS and Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 23:1073–1079.

Perrin WF and Wursig B (2009) Encyclopedia of marine mammals. Academic Press.

Perryman WL and Danil K (2018) Melon-headed whale: Peponocephala electra. In: Wursig B, JGM T and Kovacs K (eds) Encyclopedia of marine mammals. 3rd edition. Academic Press, pp 593-595.

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH and Ferrin TE (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 30:70–82.

Pfeiffer P and Culik BM (1998) Energy metabolism of underwater swimming in river-otters (Lutra lutra L.). J Comp Physiol B 168:143–148.

Ponganis PJ (2011) Diving mammals. Compr Physiol 1: 447–465.

Rambaut A, Drummond AJ (2018a) FigTree v1. 4.4. Institute of Evolutionary Biology. University of Edinburgh, Edinburgh.

Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA (2018b) Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904.

Reed-Smith J (ed),Petrini K, Spelman L, Meyers G, Crissey S, Harshaw, Carter A, McFeely M, Muzzy M, Smeeton C, Partridge J et al. (2008) North American river otter husbandry notebook. 4rd Edition. John Ball Zoo, Grand Rapids, Michigan.

Renjun L, Gewalt W, Neurohr B and Winkler A (1994) Comparative studies on the behaviour of Inia geoffrensis and Lipotes vexillifer in artificial environments. Aquat Mamm 20:39–45.

Rey C, Guéguen L, Sémon M and Boussau B (2018) Accurate detection of convergent amino-acid evolution with PCOC. Mol Biol Evol 35:2296–2306.

Schorr GS, Falcone EA, Moretti DJ and Andrews RD (2014) First long-term behavioral records from Cuvier's Beaked whales (Ziphius cavirostris) Reveal Record-Breaking Dives. PLoS One 9:e92633.

Sela I, Ashkenazy H, Katoh K and Pupko T (2015) GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res 43:W7.

Silber GK, Newcomer MW and Barros GJ (1988) Observations on the behavior and ventilation cycles of the vaquita, Phocoena sinus. Mar Mammal Sci 4:62–67.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Steiner L, Silva MA, Zereba J and Leal MJ (2008) Bryde's whales, Balaenoptera edeni, observed in the Azores: a new species record for the region. Mar Biodivers Rec 1:e66.

Stelle LL, Megill WM and Kinzel MR (2008) Activity budget and diving behavior of gray whales (Eschrichtius robustus) in feeding grounds off coastal British Columbia. Mar Mammal Sci 24:462–478.

Stensland E, Carlén I, Särnblad A, Bignert A and Berggren P (2006) Population size, distribution, and behavior of indo-pacific bottlenose (Tursiops aduncus) and humpback (Sousa chinensis) dolphins off the south coast of Zanzibar. Mar Mammal Sci 22:667–682.

Stockin K, Fairbairns R, Parsons E and Sims D (2001) Effects of diel and seasonal cycles on the dive duration of the minke whale (Balaenoptera acutorostrata). J. Mar. Biol. Assoc. U. K. 81: 189-190.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ and Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol.

Tian R, Wang Z, Niu X, Zhou K, Xu S and Yang G (2016) Evolutionary genetics of hypoxia tolerance in cetaceans during diving. Genome Biol Evol 8:827–839.

Urbán J, Rojas-Bracho L, Pérez-Cortés H, Gómez-Gallardo A, Swartz S and Brownell RL (2003) A review of gray whales (Eschrichtius robustus) on their wintering grounds in Mexican waters.J Cetacean Res Manage 5:281–295.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50:D439–D444.

Wang JY, Chu Yang S, Hung SK and Jefferson TA (2007) Distribution, abundance and conservation status of the eastern Taiwan Strait population of Indo-Pacific humpback dolphins, Sousa chinensis. Mammalia 71:157–165.

Weaver S, Shank SD, Spielman SJ, Li M, Muse S V. and Kosakovsky Pond SL (2018) Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol 35:773–777.

Wells RS, Manire CA, Byrd L, Smith DR, Gannon JG, Fauquier D and Mullin KD (2009) Movements and dive patterns of a rehabilitated Risso's dolphin, Grampus griseus, in the Gulf of Mexico and Atlantic Ocean. Mar Mammal Sci 25:420–429.

Wertheim JO, Murrell B, Smith MD, Pond SLK and Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol 32:820–832.

West KL, Mead JG and White W (2011) Steno bredanensis (Cetacea: Delphinidae). Mamm Species 43:177–189.

Wilson K, Littnan C and Read AJ (2017) Movements and home ranges of monk seals in the main Hawaiian Islands. Mar Mammal Sci 33:1080–1096.

Xu B and Yang Z (2013) PAMLX: a graphical user interface for PAML. Mol Biol Evol 30:2723–2724.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591.

Zhou KY, Pilleri G and Li YM (1980) Observations on Baiji (Lipotes vexillifer) and finless porpoise (Neophocaena asiaeorientalis) in the lower reaches of the Chang Jiang. Scientia sinica 23: 785-794.

Internet Resources

GeneCards Database (2023) https://www.genecards.org (April 14, 2022).

Kyoto encyclopedia of genes and genomes (2023) http://www.genome.jp/kegg (April 14, 2022).

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, et al. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 4;44(D1):D733-45 (May 15, 2022).

OpenAI (2021) GPT-3.5 model. https://chat.openai.com/ (February, 2023).

Phylotree (2023) https://phylotree.hyphy.org (April 14, 2022).

Ropert-Coudert Y, Kato A, Robbins A, Humphries GRW (2018) The penguiness book. World Wide Web electronic publication. http://www.penguiness.net (April 14, 2022)

String Consortium (2023) STRING v11.0. https://string-db.org (April 14, 2022).

Chapter 4. Results - Evolutionary history of the SARS-CoV-2 Gamma variant of concern (P.1): a perfect storm

Research article published in Genetics and Molecular Biology Journal

Genetics and Molecular Biology, 45, 1, e20210309 (2022)

DOI: https://doi.org/10.1590/1678-4685-GMB-2021-0309

Research Article
Human and Medical Genetics

Evolutionary history of the SARS-CoV-2 Gamma variant of concern (P.1): a perfect storm

Yuri Yépez[1*], Mariana Marcano-Ruiz[1*], Rafael S Bezerra[2*], Bibiana Fam[1], João PB Ximenez[2], Wilson A Silva-Jr[2,3] and Maria Cátira Bortolini[1]

[1]*Universidade Federal do Rio Grande do Sul, Departamento de Genética, Laboratório de Evolução Humana e Molecular, Porto Alegre, RS, Brazil.*
[2]*Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, Departamento de Genética, Ribeirão Preto, SP, Brazil.*
[3]*Instituto de Pesquisa do Câncer de Guarapuava, Guarapuava, PR, Brazil.*

Abstract

Our goal was to describe in more detail the evolutionary history of Gamma and two derived lineages (P.1.1 and P.1.2), which are part of the arms race that SARS-CoV-2 wages with its host. A total of 4,977 sequences of the Gamma strain of SARS-CoV-2 from Brazil were analyzed. We detected 194 sites under positive selection in 12 genes/ORFs: *Spike, N, M, E, ORF1a, ORF1b, ORF3, ORF6, ORF7a, ORF7b, ORF8,* and *ORF10*. Some diagnostic sites for Gamma lacked a signature of positive selection in our study, but these were not fixed, apparently escaping the action of purifying selection. Our network analyses revealed branches leading to expanding haplotypes with sites under selection only detected when P.1.1 and P.1.2 were considered. The P.1.2 exclusive haplotype H_5 originated from a non-synonymous mutational step (H3509Y) in H_1 of *ORF1a*. The selected allele, 3509Y, represents an adaptive novelty involving *ORF1a* of P.1. Finally, we discuss how phenomena such as epistasis and antagonistic pleiotropy could limit the emergence of new alleles (and combinations thereof) in SARS-COV-2 lineages, maintaining infectivity in humans, while providing rapid response capabilities to face the arms race triggered by host immuneresponses.

Keywords: Gamma, P.1, evolution, COVID-19.

Received: September 30, 2021; Accepted: December 29, 2021.

Introduction

According to the World Health Organization (WHO) and multiple researchers, the estimated average mortality rate, considering detectable/reported cases, for COVID-19 is lower (2.72%) than the disease caused by MERS-CoV (34.4%) and SARS-CoV (9.6%) (Xiao *et al*., 2020; ECDC, 2021a,b; Krishnamoorthy *et al*., 2021; Awadasseid *et al*., 2021). This number remains low even considering that the number of deaths caused by COVID-19 may be underestimated by 50%, as seen in Wuhan, China (Liu, J. *et al*., 2021). Despite this relatively low mortality rate, SARS-CoV-2 infection has led to the deaths of 4,219,578 people (WHO, 2021a). Over the past 18 months since the first reported COVID-19 case, the WHO still recognizes that the global public health risks associated with COVID-19 remain very high (WHO, 2021b). Comparatively, SARS-CoV and MERS-CoV infected 8,098 and 2,566 people and killed 774 and 866 people, respectively (WHO, 2003; Alfaraj *et al*., 2019; WHO, 2020; Petersen *et al*., 2020).

SARS-CoV, MERS-CoV, and SARS-Cov-2 have high mutation rates ($0.80–2.38 \times 10^{-3}$ substitutions per site per year (Zhao *et al*., 2004; Cotten *et al*., 2014; Li R *et al*., 2020). These mutation rates are of the same order of magnitude as other RNA viruses, and can lead to the acquisition of enhanced virulence and high evolvability, favoring changes in the host and rapid dispersion. A successful zoonotic spillover also depends on the vulnerability of the new host's defenses, and ecological and climatic conditions. Human populations also have cultural habits, with some of them facilitating the transmission of pathogens, *i.e*., hugs, kisses, sharing food (Olival *et al*., 2017; Duffy, 2018). Thus, *Homo sapiens* has become a potentially easy target of new pathogens in modern times because of its large population size, urbanization, ease of mobility of people between cities, countries, and continents, and close contact with wild, semi-wild, and domesticated animals. These conditions were in place for SARS-CoV, MERS-CoV, and SARS-CoV-2 so that the interspecific barriers were overcome, and related diseases have been reported (Kan *et al*., 2005; Zaki *et al*., 2012; Hedman *et al*., 2021).

However, there is a notable difference in the outbreak trajectories associated with these β-COVs, specializing in infecting humans and causing severe respiratory syndrome symptoms, as mentioned above. No complete scenario explaining such differences is well understood. However, it is possible to suggest that certain potential drivers, shaped by microevolutionary phenomena, can turn a local epidemic into a global pandemic, as found with SARS-CoV-2: stronger tropism involving host cells, high transmissibility, elevated transmission rates from asymptomatic individuals, substantial viral load, and relatively low lethality all powerful triggers for the emergence of evolutionarily successful viral lineages. All these conditions/factors together represent a "perfect storm".

Send correspondence to Maria Cátira Bortolini. Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Departamento de Genética. Av. Bento Gonçalves, 9500. Prédio 43323, Porto Alegre, RS, Brazil. E-mail: maria.bortolini@ufrgs.br.

*These authors contributed equally.

Chapter 5. Final conclusions

In the genomics era, we have unprecedented opportunities to delve into the refined details of genetic foundations and patterns underlying many biological processes. This includes elucidating elements, from an evolutionary perspective, that support the adaptation of species from diverse taxonomic groups living in various environments, thus subject to varied selective pressures. In this context, it is intriguing, for instance, to contemplate the essential role of evolution in shaping the mechanisms governing the sensing and adaptation to variations in oxygen levels.

Animals encounter environments with varying levels of oxygen, resulting in spatial and temporal fluctuations in tissue oxygen levels. These variations, including those triggered by normal physiological events like physical exertion-induced drops in oxygen levels in skeletal muscle, regulate important adaptive responses in cells and tissues by modulating gene transcription. The 2019 Nobel Prize-winning research in physiology elucidated the molecular mechanisms through which animal cells adjust to fluctuations in oxygen supply, emphasizing the pivotal role of the Hypoxia Inducible Factor (HIF) in governing oxygen-dependent responses (Kaelin & Ratcliffe, 2008). These mechanisms modulate cellular metabolism and control fundamental processes related to development, regeneration, and defense, highlighting the critical significance of animal cells' capability to detect oxygen levels and modify their gene expression patterns. The Oxygen-activated signaling pathway (also named in our work as the HIF signaling pathway) which regulates approximately 300 genes across different genetic networks, impacts a wide range of physiological processes.

It was biological evolution, through its intricate mechanisms and processes, that shaped the ability of animals to sense and respond to different oxygen levels in diverse ways. For instance, over time, natural selection favored genetic variations that confer advantages in oxygen sensing and adaptation to hypoxic environments (Sackton and Clark, 2019). Organisms inhabiting high-altitude regions, for example, underwent adaptive changes in oxygen sensing mechanisms to cope with lower oxygen levels (Graham and McCracken, 2019; Witt and Huerta-Sánchez, 2019). This indicates that variants in these genes may be

involved not only in microevolutionary processes within populations of the same species but also in macroevolutionary processes across different species and taxonomic groups, particularly in adapting to environments with lower oxygen levels. Thus, we hypothesized testing genetic variation that could be selected for by evolution in specific taxonomic groups, which could indicate evolutionary adaptability. Aquatic mammals, a polyphyletic group, were chosen for their lifestyle characterized by the dive response capacity and their potential as a model for understanding evolutionary adaptability, despite some of them not being phylogenetically close.

The title of the thesis, "*Evolutionary forces shaping genomes: from whales to viruses*" accurately reflects the broad scope of the study, which utilizes genomics to explore various evolutionary processes across species of vastly different order of sizes. From massive creatures like whales to biological entities at the scale of viruses, these researches aim to illuminate the diverse mechanisms driving evolution. This breadth of inquiry is further exemplified in another study conducted and published during this PhD (Landau et al., 2021, Annex I).

Amidst investigations into unrelated topics in aquatic animal evolution, the emergence of research on SARS-CoV-2 was prompted by the urgent demands of the COVID-19 pandemic. This unexpected shift underscores the adaptability and multidisciplinary nature of research, demonstrating the capacity to address pressing global challenges even in the face of unforeseen circumstances.

We capitalized on the increasing availability of genomes, examining, on the one hand, a significant number of aquatic mammal species and, on the other hand, genomes of SARS-CoV-2 Gamma variant, a virus variant of particular importance in Brazil, that emerged towards the end of 2020. This enabled us to conduct genome comparisons and pinpoint genetic variations contributing to evolutionary shifts. The data analysis methods showed similarity up to a certain point. In this manner, our analysis involved assessing evolutionary rates at the molecular level, considering both macroevolutionary changes within mammalian phylogeny and microevolutionary dynamics, particularly relevant for viruses given their rapid replication rates and short generation times. Additionally, we explored phenomena such as

amino acid convergence in mammals, potentially linked to adaptation to diverse low-oxygen environments or lifestyles, and reconstructed the spatial trajectories of virus variants.

In this context, the present Thesis aimed to address issues within the genomic approaches that ended in the published works titled: "Adaptive Strategies of Aquatic Mammals: Exploring the Role of the HIF Pathway and Hypoxia Tolerance" (Chapter 3), and "Evolutionary History of the SARS-CoV-2 Gamma variant of concern (P.1)" (Chapter 4).

Chapter 3 delves into the primary topic of this thesis, examining the molecular-level evolution of genes within the HIF signaling pathway. Within this chapter, the objectives outlined as 1 through 4 of the Thesis were addressed. For objective 1 and 2, we examined genetic variability within the coding regions of the hypoxia signaling pathway genes. Unlike our work, several studies have explored the different aspects of marine mammal evolution, focusing mainly on marine species and leaving freshwater species understudied. Coping with hypoxic environments is a crucial adaptation shared by all aquatic mammals, enabling them to thrive in their habitats. Our results revealed relaxed selection patterns in six out of the eleven genes associated with the HIF pathway in aquatic mammals, particularly cetaceans, suggesting that HIF pathway proteins in cetaceans may have initially evolved under different functional constraints compared to those in pinnipeds and Lutrinae, highlighting the role of relaxed selection in shaping the evolutionary trajectory of HIF pathway genes in aquatic mammals.

Our analysis revealed predominantly purifying selection, indicating highly conserved domains within these genes, as we expected. However, we also detected episodic positive selection events in different aquatic clades, particularly in HIF3α, where a Glutamine residue at site 68 was found exclusively in cetaceans. This residue change may affect HIF3α structure and function, potentially influencing its role as a regulator of the HIF1α pathway. Our study differs from previous research in sequence selection and gene analysis, but confirms similar conclusions regarding the role of purifying selection in HIF pathway gene evolution. Also, we did not find any accelerated evolutionary rates in deep diving species except on the *ARNT2* in pinnipeds, particularly in the deep-diving Phocidae branch suggesting a possible deep-diving

association in their common ancestor. Finally, we identified a cetacean-exclusive amino acid in EGLN3, although structural stability analysis did not predict significant stability changes.

In Objective 3 our aim was to investigate molecular pattern convergence behind a shared phenotype of hypoxia tolerance in mammals. We identified two convergent sites: site 485 in ARNT2 and site 364 in HIF3α. Positive selection detected at the HIF3α 364 site supports its role in convergence among aquatic mammalian lineages and a subterranean rodent species under hypoxia tolerance scenarios.

Regarding Objective 4, the preceding results have contributed to highlighting the potential relationship between genetic variability and adaptation. Specifically, we successfully calculated the effect of certain variants exhibiting positive selection signals on the stability of their respective proteins. However, further investigations focusing on functional analysis are needed to comprehensively understand the effect and contribution of these variants to the adaptation of aquatic mammals to their environment.

In Chapter 4, responding to Objective 5, our study presents insightful findings regarding the evolutionary dynamics and adaptive novelties of SARS-CoV-2 Gamma variants. Thus, we identified 213 positively selected sites, with a notable proportion located within the Spike protein, consistent with prior observations. Notably, ancestral alleles were detected at specific sites, albeit in varying combinations, suggesting the potential for evolvability in the viral Spike protein modulated by factors such as epistasis and pleiotropy. Our investigation also revealed the presence of potentially advantageous alleles, exemplified by the identification of allele 3509Y in ORF1a among P.1.2 sub-lineage sequences, emphasizing the importance of further functional studies to ascertain their fitness implications. Furthermore, our study sheds light on the stochastic nature of lineage expansion and the interplay between positive selection and purifying selection across different genomic sites. However, it's imperative to acknowledge the limitations of our study, including the possibility of false-positive results and the inherent complexity of analyzing positive selection regimes using then-available population genomic data. Overall, our findings have contributed to a slightly deeper understanding of the evolutionary mechanisms driving the emergence and

spread of SARS-CoV-2 variants. We hope that they have provided valuable information for surveillance, prevention, or control efforts carried out during the COVID-19 pandemic.

In conclusion, present Thesis has explored the intricate interplay between genetic evolution and environmental adaptation, spanning from the molecular mechanisms of oxygen sensing in aquatic mammals to the evolutionary dynamics of SARS-CoV-2 variants. Through the utilization of genomic approaches, we believe we have elucidated important details of the genetic foundations governing diverse biological processes and shed light on the adaptive strategies employed by organisms to thrive in their respective environments.

References

Akamatsu, T., Wang, D., Wang, K., Wei, Z., Zhao, Q., & Naito, Y. (2002). Diving behaviour of freshwater finless porpoises (Neophocaena phocaenoides) in an oxbow of the Yangtze River, China. ICES Journal of Marine Science, 59(2), 438-443.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Antonov, A. V., Schmidt, E. E., Dietmann, S., Krestyaninova, M., & Hermjakob, H. (2010).

Arnason, U., Gullberg, A., Gretarsdottir, S., Ursing, B., & Janke, A. (2000). The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. Journal of Molecular Evolution, 50(6), 569-578.

Arya, R., Kumari, S., Pandey, B., Mistry, H., Bihani, S. C., Das, A., ... & Kumar, M. (2021). Structural insights into SARS-CoV-2 proteins. Journal of molecular biology, 433(2), 166725.

Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., & Pupko, T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic acids research, 40(W1), W580-W584.

Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S., Angeletti, S., & Ciccozzi, M. (2020). The 2019‐new coronavirus epidemic: evidence for virus evolution. Journal of medical virology, 92(4), 455-459.

Berra, E., Benizri, E., Ginouvès, A., Volmat, V., Roux, D., & Pouysségur, J. (2003). HIF prolyl‐hydroxylase 2 is the key oxygen sensor setting low steady‐state levels of HIF‐1α in normoxia. The EMBO journal, 22(16), 4082-4090.

Berta, A. (2018). Pinniped evolution. In Encyclopedia of marine mammals. Academic Press.

Berta, A., Churchill, M., & Boessenecker, R. W. (2018). The origin and evolutionary biology of pinnipeds: seals, sea lions, and walruses. Annual Review of Earth and Planetary Sciences, 46(), 203-228.

Berta, A., Sumich, J. L., & Kovacs, K. M. (2006). Marine mammals: evolutionary biology. Elsevier.

Bi, J., Hu, B., Zheng, J., Wang, J., Xiao, W., & Wang, D. (2015). Characterization of the hypoxia-inducible factor 1 alpha gene in the sperm whale, beluga whale, and Yangtze finless porpoise. Marine Biology, 162(6), 1201-1213.

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

Carabelli, A. M., Peacock, T. P., Thorne, L. G., Harvey, W. T., Hughes, J., COVID-19 Genomics UK Consortium de Silva Thushan I. 6, ... & Robertson, D. L. (2023). SARS-CoV-2 variant biology: immune escape, transmission and fitness. Nature Reviews Microbiology, 21(3), 162-177.

Castro, M. C., Kim, S., Barberia, L., Ribeiro, A. F., Gurzenda, S., Ribeiro, K. B., ... & Singer, B. H. (2021). Spatiotemporal pattern of COVID-19 spread in Brazil. Science, 372(6544), 821-826.

Chilvers, B. L., Delean, S., Gales, N. J., Holley, D. K., Lawler, I. R., Marsh, H., & Preen, A. R. (2004). Diving behaviour of dugongs, Dugong dugon. Journal of Experimental Marine Biology and Ecology, 304(2), 203-224.

Choi Y, Chan AP (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics , 31(16), 2745–2747.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: database for annotation, visualization, and integrated discovery. Genome biology, 4(6), R60.

Fassin, Y. (2021). Research on Covid-19: a disruptive phenomenon for bibliometrics. Scientometrics, 126(6), 5305-5319.

Ferrareze, P. A. G., Cybis, G. B., de Oliveira, L. F. V., Zimerman, R. A., Schiavon, D. E. B., Peter, C., & Thompson, C. E. (2024). Intense P. 1 (Gamma) diversification followed by rapid Delta substitution in Southern Brazil: a SARS-CoV-2 genomic epidemiology study. Microbes and Infection, 26(1-2), 105216.

Folkens, Pieter A., and Randall R. Reeves. (2002). Guide to marine mammals of the world. National Audubon Society.

Foote, A. D., Liu, Y., Thomas, G. W., Vinař, T., Alföldi, J., Deng, J., ... & Khan, Z.

(2015). Convergent evolution of the genomes of marine mammals. Nature genetics, 47(3), 272.

Gingerich, P. D., ul Haq, M., Zalmout, I. S., Khan, I. H., & Malkani, M. S. (2001). Origin of whales from early artiodactyls: hands and feet of Eocene Protocetidae from Pakistan. Science, 293(5538), 2239-2242.

Giovanetti, M., Slavov, S. N., Fonseca, V., Wilkinson, E., Tegally, H., Patané, J. S. L., ... & Covas, D. T. (2022). Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. Nature Microbiology, 7(9), 1490-1500.

Graham, A. M., & McCracken, K. G. (2019). Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes EGLN1 and EPAS1 in high-altitude ducks. *Heredity*, *122*(6), 819-832.

Graur, D., & Higgins, D. G. (1994). Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. Molecular Biology and Evolution, 11(3), 357-364.

Gustafsson, M. V., Zheng, X., Pereira, T., Gradin, K., Jin, S., Lundkvist, J., ... & Bondesson, M. (2005). Hypoxia requires notch signaling to maintain the undifferentiated cell state. Developmental cell, 9(5), 617-628.

Haghani, M., & Bliemer, M. C. (2020). Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCov literature. Scientometrics, 125, 2695–2726.

Hamed, S. M., Elkhatib, W. F., Khairalla, A. S., & Noreddin, A. M. (2021). Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. Scientific reports, 11(1), 8435.

Hampton‐Smith, R. J., & Peet, D. J. (2009). From Polyps to People: A HI ghly F amiliar Response to Hypoxia. Annals of the New York Academy of Sciences, 1177(1), 19-29.

Hanaoka, M., Droma, Y., Basnyat, B., Ito, M., Kobayashi, N., Katsuyama, Y., ... & Ota, M. (2012). Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. PloS one, 7(12), e50566.

Hoff, K. J., & Stanke, M. (2019). Predicting genes in single genomes with augustus. Current protocols in bioinformatics, 65(1), e57.

Jacob, J. J., Fletcher, G. J., Priya, T. M., Veeraraghavan, B., & Mutreja, A. (2021). Relevance of immune response and vaccination strategies of SARS-CoV-2 in the phase of

viral red queen dynamics. Indian Journal of Medical Microbiology, 39(4), 417-422.

Jamieson, Barrie G. M. (2016). Reproductive Biology and Phylogeny of Cetaceans. CRC Press.

Jefferson, T. A., Webber, M. A., & Pitman, R. L. (2011). Marine mammals of the world: a comprehensive guide to their identification. Elsevier.

Jefferson, Thomas A., Marc A. Webber, and Robert L. Pitman (2011). Marine mammals of the world: a comprehensive guide to their identification. Elsevier.

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., ... & Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. Nature communications, 5(), 3281.

Johnson, P., Elsner, R., & Zenteno-Savín, T. (2004). Hypoxia-inducible factor in ringed seal (Phoca hispida) tissues. Free radical research, 38(8), 847-854.

Johnson, P., Elsner, R., & Zenteno-Savín, T. (2005). Hypoxia-inducible factor 1 proteomics and diving adaptations in ringed seal. Free Radical Biology and Medicine, 39(2), 205-212.

Kaelin Jr, W. G., & Ratcliffe, P. J. (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Molecular cell, 30(4), 393-402.

Kaidi, A., Williams, A. C., & Paraskeva, C. (2007). Interaction between β-catenin and HIF-1 promotes cellular adaptation to hypoxia. Nature cell biology, 9(2), 210.

Koshiji, M., Kageyama, Y., Pete, E. A., Horikawa, I., Barrett, J. C., & Huang, L. E. (2004). HIF‐1α induces cell cycle arrest by functionally counteracting Myc.. The EMBO journal, 23(9), 1949-1956.

Kulshreshtha, R., Ferracin, M., Wojcik, S. E., Garzon, R., Alder, H., Agosto-Perez, F. J., ... & Calin, G. A. (2007). A microRNA signature of hypoxia. Molecular and cellular biology, 27(5), 1859-1867.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35(6), 1547-1549.

Lamers, M. M., & Haagmans, B. L. (2022). SARS-CoV-2 pathogenesis. Nature reviews microbiology, 20(5), 270-284.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., &

Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics, 27(12), 1739-1740.

Liu, F. G. R., & Miyamoto, M. M. (1999). Phylogenetic assessment of molecular and morphological data for eutherian mammals. Systematic Biology, 48(1), 54-64.

Mahon, P. C., Hirota, K., & Semenza, G. L. (2001). FIH-1: a novel protein that interacts with HIF-1α and VHL to mediate repression of HIF-1 transcriptional activity. Genes & development, 15(20), 2675-2686.

Marsh, H., O'Shea, T. J., Reynolds, J. E., & Reynolds III, J. E. (2012). Ecology and conservation of the Sirenia: dugongs and manatees (Vol. 18). Cambridge University Press.

Marx, F. G., Lambert, O., & Uhen, M. D. (2016). Cetacean paleobiology. John Wiley & Sons.

McGowen, M. R., Gatesy, J., & Wildman, D. E. (2014). Molecular evolution tracks macroevolutionary transitions in Cetacea. Trends in ecology & evolution, 29(6), 336-346.

Menezes, D., Fonseca, P. L. C., de Araújo, J. L. F., & Souza, R. P. D. (2022). SARS-CoV-2 Genomic Surveillance in Brazil: A Systematic Review with Scientometric Analysis. Viruses, 14(12), 2715.

Metzen, E., Berchner-Pfannschmidt, U., Stengel, P., Marxsen, J. H., Stolze, I., Klinger, M., ... & Acker, H. (2003). Intracellular localisation of human HIF-1α hydroxylases: implications for oxygen sensing. Journal of cell science, 116(7), 1319-1326.

O'Toole, Á., Pybus, O. G., Abram, M. E., Kelly, E. J., & Rambaut, A. (2022). Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. BMC genomics, 23(1), 1-13.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., ... & Su, B. (2010). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Molecular biology and evolution, 28(2), 1075-1081.

Ponganis, P. J. (2011). Diving mammals. Comprehensive Physiology, 1(1), 447-465.

R spider: a network-based analysis of gene lists by combining signaling and metabolic pathways from Reactome and KEGG databases.. Nucleic acids research, 38(suppl_2), W78--W83.

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., ... & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature microbiology, 5(11), 1403-1407.

Romero, A (2009). The Biology of Marine Mammals. Arkansas State University.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology, 61(3), 539-542.

Schmidt, H., Malik, A., Bicker, A., Poetzsch, G., Avivi, A., Shams, I., & Hankeln, T. (2017). Hypoxia tolerance, longevity and cancer-resistance in the mole rat Spalax–a liver transcriptomics approach. Scientific reports, 7(1), 14348.

Schofield, C. J., & Ratcliffe, P. J. (2004). Oxygen sensing by HIF hydroxylases. Nature reviews Molecular cell biology, 5(5), 343.

Scholander, P. T., & Irving, L. (1941). Experimental investigations on the respiration and diving of the Florida manatee. Journal of Cellular and Comparative Physiology, 17(2), 169-191.

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., ... & Willerslev, E. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX.. .

Semenza, G. L. (2007). Life with oxygen. Science, 318(5847), 62-64.

Shams, I., Avivi, A., & Nevo, E. (2004). Hypoxic stress tolerance of the blind subterranean mole rat: expression of erythropoietin and hypoxia-inducible factor 1α. Proceedings of the National Academy of Sciences, 101(26), 9698-9703.

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., ... & Prchal, J. T. (2010). Genetic evidence for high-altitude adaptation in Tibet. v, 329(5987), 72-75.

Soñanez-Organis, J. G., Vázquez-Medina, J. P., Crocker, D. E., & Ortiz, R. M. (2013). Prolonged fasting activates hypoxia inducible factors-1α,-2α and-3α in a tissue-specific manner in northern elephant seal pups. Gene, 526(2), 155-163.

Strowitzki, M. J., Cummins, E. P., & Taylor, C. T. (2019). Protein Hydroxylation by Hypoxia-Inducible Factor (HIF) Hydroxylases: Unique or Ubiquitous?. Cells, 8(5), 384.

Sun, C., Xie, C., Bu, GL. et al. Molecular characteristics, immune evasion, and impact

of SARS-CoV-2 variants. Sig Transduct Target Ther 7, 202 (2022).

Taylor, C. T., & McElwain, J. C. (2010). Ancient atmospheres and the evolution of oxygen sensing via the hypoxia-inducible factor in metazoans. Physiology, 25(5), 272-279.

Tosta, S., Moreno, K., Schuab, G., Fonseca, V., Segovia, F. M. C., Kashima, S., ... & Giovanetti, M. (2023). Global SARS-CoV-2 genomic surveillance: What we have learned (so far). Infection, Genetics and Evolution, 105405.

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews Microbiology, 19(3), 155-170.

Vaughan, T., Ryan, J., & Czaplewski, N. (2011). Mammalogy. Jones & Bartlett Learning.

Vázquez-Medina, J. P., Zenteno-Savín, T., Tift, M. S., Forman, H. J., Crocker, D. E., & Ortiz, R. M. (2011). Apnea stimulates the adaptive response to oxidative stress in elephant seal pups. Journal of Experimental Biology, 214(24), 4193-4200.

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

Weidemann, A., & Johnson, R. S. (2008). Biology of HIF-1α. Cell death and differentiation, 15(4), 621.

Witt, K. E., & Huerta-Sánchez, E. (2019). Convergent evolution in human and domesticate adaptation to high-altitude environments. Philosophical Transactions of the Royal Society B, 374(1777), 20180235.

World Health Organisation (15 January, 2024). WHO COVID-19 dashboard. Retrieved from: https://data.who.int/dashboards/covid19.

World Health Organization (16 January, 2024). Transmission of SARS-CoV-2: implications for infection prevention precautions: Scientific Brief. Retrieved from: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

World Health Organization (WHO). SARS-CoV-2. https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

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

Yeates, L. C., Williams, T. M., & Fink, T. L. (2007). Diving and foraging energetics of the smallest marine mammal, the sea otter (Enhydra lutris). Journal of Experimental Biology, 210(11), 1960-1970.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., ... & Zheng, H. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science, 329(5987), 75-78.

Zhu, K., Ge, D., Wen, Z., Xia, L., & Yang, Q. (2018). Evolutionary Genetics of Hypoxia and Cold Tolerance in Mammals. Journal of molecular evolution, 86(9), 618-634.

Annex I. Evolutionary analysis of the anti-viral *STAT2* gene of primates and rodents: Signature of different stages of an arms race

Research article published in the Infection, Genetics and Evolution Journal

Infection, Genetics and Evolution 95 (2021) 105030

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Evolutionary analysis of the anti-viral *STAT2* gene of primates and rodents: Signature of different stages of an arms race

Luane Jandira Bueno Landau [a], Bibiana Sampaio de Oliveira Fam [a], Yuri Yépez [a], Gabriela Barreto Caldas-Garcia [a], Alcides Pissinatti [b], Tiago Falótico [c], Guillermo Reales [a,d], Lavínia Schüler-Faccini [d], Vinicius Albuquerque Sortica [a], Maria Cátira Bortolini [a,*]

[a] Laboratório de Evolução Humana e Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
[b] Rio de Janeiro's Primatology Center (RJPC – INEA), Rio de Janeiro, RJ, Brazil
[c] School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, SP, Brazil
[d] Instituto Nacional de Genética Médica Populacional, Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil

ARTICLE INFO

Keywords:
STAT2
SH2 domain
Molecular evolution
Primates
Flavivirus infection
Rodentia
Arms race

ABSTRACT

STAT2 plays a strategic role in defending viral infection through the signaling cascade involving the immune system initiated after type I interferon release. Many flaviviruses target the inactivation or degradation of STAT2 as a strategy to impair this host's line of defense. Primates are natural reservoirs for a range of disease-causing flaviviruses (e.g., Zika, Dengue, and Yellow Fever virus), while rodents appear less susceptible. We analyzed the STAT2 coding sequence of 28 Rodentia species and 49 Primates species. Original data from 19 Platyrrhini species were sequenced for the SH2 domain of STAT2 and included in the analysis. STAT2 has many sites whose variation can be explained by positive selection, measurement by two methods (PALM indicated 12, MEME 61). Both evolutionary tests significantly marked sites 127, 731, 739, 766, and 780. SH2 is under evolutionary constraint but presents episodic positive selection events within Rodentia: in one of them, a moderately radical change (serine > arginine) at position 638 is found in *Peromyscus* species, and can be implicated in the difference in susceptibility to flaviviruses within Rodentia. Some other positively selected sites are functional such as 5, 95, 203, 251, 782, and 829. Sites 251 and 287 regulate the signaling mediated by the JAK-STAT2 pathway, while 782 and 829 create a stable tertiary structure of STAT2, facilitating its connection with transcriptional co-activators. Only three positively selected sites, 5, 95, and 203, are recognized members who act on the interface between STAT2 and flaviviruses NS5 protein. We suggested that due to the higher evolutionary rate, rodents are, at this moment, taking some advantage in the battle against infections for some well-known *Flaviviridae*, in particular when compared to primates. Our results point to dynamics that fit with a molecular evolutionary scenario shaped by a thought-provoking virus-host arms race.

1. Introduction

1.1. The first line of defense against viral infection and the mammalian STAT2 role

During the animals' evolutionary trajectory, efficient strategies to combat a large diversity of pathogens have evolved. For instance, upon virus infection, type I interferons (IFNs) are expressed in the vertebrate hosts, triggering a complex signaling molecular cascade that inhibits viral replication, containing the disease associated with it (Xu et al., 2013; Jennings and Sang, 2019; Wang. B et al., 2020; Boudinot et al., 2021).

In mammals, in addition to IFNs such as IFNα, αω, β, δ, ε, ζ, κ, τ, ω, many other gene products are involved in the sophisticated activation of the host antiviral defense program, among which members of the signal transducer and activator of transcription (STAT) protein family (Jennings and Sang, 2019; Boudinot et al., 2021). Based on in silico, in vitro, and in vivo studies with humans and animal models, once released on extracellular space, type I IFNs bind to their membrane receptors (IFNAR), leading to the receptors' oligomerization. Posteriorly, Janus

* Corresponding author at: Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil.
E-mail address: maria.bortolini@ufrgs.br (M.C. Bortolini).

https://doi.org/10.1016/j.meegid.2021.105030
Received 28 April 2021; Received in revised form 24 July 2021; Accepted 6 August 2021
Available online 9 August 2021