Globin Evolution in the Australian Lungfish: An Integrated Bioinformatic, Structural and Functional Study

---

Dissertation

with the aim of achieving a
doctoral degree (Dr. rer. nat.)
at the
Faculty of Mathematics, Informatics and Natural Sciences
Department of Biology
University of Hamburg

submitted by
Soraya Levinia Safavi

Hamburg, July 2026

“It is not the strongest or the most intelligent who will survive but those who can best manage change.”

- Leon C. Megginson

1st Evaluator: Prof. Dr. Kathrin Otte

2nd Evaluator: Dr. Andrej Fabrizius

Disputation committee:

Prof. Dr. Tobias Lenz (chair)

Prof. Dr. Baris Tursun

Date of disputation: 1st of Oktober 2026

Zusammenfassung iv
Abstract v
1. Introduction 1
1.1 Oxygen as a biological challenge 1
1.2 Globins as oxygen-binding proteins 1
1.3 Heme coordination and structural basis of globins 2
1.4 Hemoglobin 3
1.5 Myoglobin 4
1.6 Neuroglobin 5
1.7 Cytoglobin 6
1.8 Globin E 8
1.9 Globin Y 9
1.10 Globin X 9
1.11 Androglobin 10
1.12 Evolution of globins 10
1.13 Lungfish 12
1.13.1 Respiration 14
1.13.2 Neoteny and genome size 14
1.13.3 Evolutionary link to tetrapods 15
1.14 Multiple globins in Lungfish 16
1.15 Aim of the thesis 16
2 Chapter 1 - Identification of globin genes in the lungfish 17
2.1 Abstract 17
2.2 Introduction 17
2.3 Results 19
2.3.1 Genome annotation of *N. forsteri* 19
2.3.2 Detection of globin candidates 19
2.3.3 Localization of globin genes within the genome 21
2.3.4 Phylogenetic analysis 21
2.4 Discussion 23
2.4.1 Summary of key findings 23
2.4.2 Annotation quality and methodological considerations 23
2.4.3 Detection, classification and genomic context of globin genes 24
2.4.4 Evolutionary implications of the globin repertoire 24
2.4.5 Limitations 25
2.4.6 Conclusions and perspectives 25

2.5 Material and methods 26
2.5.1 Genome preprocessing and repeat masking 26
2.5.2 Genome annotation 26
2.5.3 Detection of putative globin genes 28
2.5.4 Sequence filtering and classification 28
2.5.5 RNA-seq backmapping for expression validation 28
2.5.6 Phylogenetic analyses 29
2.6 Acknowledgements 30
3 Chapter 2 - Neuroglobin and cytoglobin in lungfish 31
3.1 Abstract 31
3.2 Introduction 31
3.3 Results 33
3.3.1 Identification and tissue-specific expression of Ngb and Cygb 33
3.3.2 Sequence characteristics and conserved functional residues 33
3.3.3 Phylogenetic classification of lungfish Ngb and Cygb 34
3.3.4 Structural prediction and comparison with human globins 35
3.3.5 Spectroscopic analysis of coordination states 37
3.4 Discussion 39
3.4.1 Detection of Ngb and Cygb transcripts in *N. forsteri* 39
3.4.2 Evolutionary significance of lungfish Ngb and Cygb 39
3.4.3 Sequence and structural conservation 39
3.4.4 Structural conservation of Ngb and Cygb 40
3.4.5 Spectroscopic support for hexacoordination 40
3.4.6 Evolutionary implications and functional conservation 41
3.4.7 Limitations and future perspectives 41
3.5 Material and methods 41
3.5.1 Sample material and cDNA generation 41
3.5.2 PCR amplification and sequence verification 42
3.5.3 Sequence analysis and alignments 42
3.5.5 Structural modelling and comparison 43
3.5.6 Recombinant expression and purification of Ngb and Cygb 43
3.5.7 UV-Vis spectroscopy 45
3.6 Acknowledgements 45
4 Chapter 3 - Functional *in vitro* studies of lungfish Neuroglobin and Cytoglobin 46
4.1 Abstract 46
4.2 Introduction 46

4.3 Results 48
4.3.1 H_2O_2 stress experiment 48
4.3.2 Temperature stress experiment 50
4.4 Discussion 51
4.4.1 H_2O_2 stress 51
4.4.2 Temperature stress 52
4.4.3 Outlook and limitations 53
4.5 Material and methods 54
4.5.1 Culturing of CHSE-214 cells 54
4.5.2 Plasmids and transfection 54
4.5.3 Stress experiments 54
4.5.5 Statistical analyses 55
4.6 Acknowledgements 56
5 General Discussion 57
5.1 Lungfish as a key model for vertebrate evolution 57
5.2 Bioinformatic identification and annotation of globin genes in *N. forsteri* 59
5.3 Evolutionary diversification of the lungfish globin repertoire 61
5.4 Functional characterization of lungfish Ngb and Cygb 63
5.4.1 Sequence- and structure-based evidence for conserved globin features 63
5.4.2 Tissue-specific detection and evolutionary significance 65
5.4.3 Stress-related effects in CHSE-214 cells 66
5.5 Limitations and future perspectives 68
5.6 General conclusions 70
6 References 72
7 Supplementary Materials 84
8 Acknowledgments 128
9 Affidavit 129

Zusammenfassung

Im Rahmen dieser Arbeit wurde das Globin-Repertoire des australischen Lungenfisches (*Neoceratodus forsteri*), des nächstlebenden Verwandten der Tetrapoden, untersucht, um Einblicke in die Evolution der Globine in Wirbeltieren während des Übergangs vom Wasser- zum Landleben zu gewinnen. Da das Genom des Lungenfisches außergewöhnlich groß und stark repetitiv ist, lag ein wesentlicher Schwerpunkt der Arbeit auf der Entwicklung und Anwendung komplementärer bioinformatischer Strategien zur zuverlässigen Identifizierung und Annotation von Genen. Mithilfe einer lungenfischspezifischen *repeat*-maskierten Genomannotation in Kombination mit sequenzähnlichkeitbasierten Suchverfahren, phylogenetischen Analysen, transkriptomischer Validierung und Strukturvorhersagen wurde ein vielfältiges Globin-Repertoire erfasst. Dieses beinhaltet Vertreter der wichtigsten vertebraten Globin-Familien.

Insgesamt wurden 103 globinähnliche Kandidaten identifiziert, von denen 18 hochkonfidente Sequenzen für weiterführende Analysen ausgewählt wurden. Dazu zählen multiple Kopien von Myoglobin und Globin E sowie von Globin X, Globin Y, Androglobin und den Hämoglobin-α/β-Komponenten. Besonders hervorzuheben ist die Identifizierung von Neuroglobin (Ngb) und Cytoglobin (Cygb). Diese Arbeit liefert erstmals den Nachweis von Ngb und Cygb in *N. forsteri* und in Lungenfischen insgesamt.

Beide Gene wurden aus cDNA amplifiziert und zeigten eine gewebespezifische Expression im Gehirn bzw. im Herzen. Darüber hinaus gruppierten sie innerhalb der erwarteten Vertebraten-Ngb- und -Cygb-Kladen und wiesen die für hexakoordinierte Globine typischen konservierten strukturellen Merkmale auf. AlphaFold-Modelle und UV-Vis-spektroskopische Analysen stützten zusätzlich die Erhaltung der kanonischen Globin-Faltung und der Hämkoordination.

Ergänzend zur genomweiten und molekularen Charakterisierung wurden explorative Zellkulturexperimente durchgeführt, um den Einfluss von Lungenfisch-Ngb und -Cygb auf zelluläre Stressantworten in einem heterologen Fischzellsystem zu untersuchen. Obwohl sich unter den getesteten Bedingungen kein eindeutiger protektiver Effekt nachweisen ließ, bilden diese Versuche eine Grundlage für zukünftige funktionelle Untersuchungen.

Insgesamt zeigt diese Arbeit, dass Lungenfische über ein komplexes und evolutionär aufschlussreiches Globin-Repertoire verfügen und dass integrierte bioinformatische und experimentelle Ansätze für die Untersuchung von Genen in außergewöhnlich großen, repetitiven Genomen unverzichtbar sind.

Abstract

This thesis investigates the globin repertoire of the Australian lungfish (*Neoceratodus forsteri*), the closest living relative of tetrapods, to gain insight into vertebrate globin evolution during the transition from water to land. The lungfish genome is exceptionally large and highly repetitive, so a major part of the work focused on developing and applying complementary bioinformatic strategies for reliable gene identification and annotation. A diverse globin repertoire was recovered using a lungfish specific repeat-masked genome annotation pipeline combined with sequence similarity-based searches, phylogenetic analyses, transcriptomic validation, and structural prediction. This repertoire includes representatives of the major vertebrate globin families.

A total of 103 globin-like candidates were identified. Of these, 18 high-confidence sequences were selected for further analysis. These included multiple copies of myoglobin and globin E, as well as globin X, globin Y, androglobin and hemoglobin α/β components. Importantly, they also included neuroglobin (Ngb) and cytoglobin (Cygb). This study is the first to confirm the presence of Ngb and Cygb in *N. forsteri* and in lungfishes in general.

Both genes were amplified from cDNA and exhibited tissue-specific expression in the brain and heart. They also clustered within the expected vertebrate Ngb and Cygb clades and retained the conserved structural features characteristic of hexacoordinated globins. AlphaFold models and UV-Vis spectroscopy further supported the conservation of the canonical globin fold and heme coordination.

In addition to genome-wide and molecular characterization, exploratory cell culture experiments were conducted to investigate the impact of lungfish Ngb and Cygb on cellular stress responses within a heterologous fish cell system. While these assays did not reveal a clear protective effect under the tested conditions, they established a foundation for future functional studies.

Together, this thesis shows that lungfish retain a complex and evolutionarily informative globin repertoire and that integrated bioinformatic and experimental approaches are essential for studying genes in exceptionally large, repeat-rich genome.

1. Introduction

1.1 Oxygen as a biological challenge

Oxygen is a chemically reactive yet physiologically indispensable molecule essential to most complex life on Earth (Thannickal, 2009). It connects photosynthetic production in the biosphere with aerobic energy metabolism in animals and many microorganisms, serving as a fundamental currency of biological processes (Cloud & Gibor, 1970; Lane, 2002). In natural systems, molecular oxygen is primarily generated by oxygenic photosynthesis and subsequently consumed through aerobic respiration. As the terminal electron acceptor in mitochondrial and many bacterial electron transport chains, oxygen enables the production of adenosine triphosphate (ATP) from reduced carbon substrates in a highly efficient manner (Salin et al., 2015). This supports energetically demanding processes, such as active locomotion, neural activity, and homeostatic regulation. Oxygen acquisition, distribution, and utilization are fundamental constraints that shape multicellular organisms' physiology, ecology, and evolutionary trajectories (Decker & Van Holde, 2010; Salin et al., 2015).

1.2 Globins as oxygen-binding proteins

Since oxygen is poorly soluble in water and can be toxic at high concentrations, animals have evolved specialized oxygen-binding proteins that transport, store, and buffer it (Childress & Seibel, 1998; Reinhard et al., 2016). Globins are among the most widely distributed respiratory proteins, occurring across vertebrates, invertebrates, prokaryotes, fungi, archaea, protozoa, and plants (Lesk & Lecomte, 2013; Vinogradov et al., 2013). In plants, for example, leghemoglobins play a central role in symbiotic nitrogen fixation, while other plant globins are implicated in nitric oxide homeostasis and stress responses (Singh & Varma, 2017).

Beyond these canonical vertebrate globins, a diverse range of other globins, globin-like and non-globin oxygen-binding proteins exists across metazoans and microbes (Alvarez-Carreno et al., 2016; Estrín et al., 2020), reflecting repeated evolutionary solutions to the challenge of managing a reactive gas that is both essential and potentially harmful.

1.3 Heme coordination and structural basis of globins

Among globins, structural variation in heme coordination has important implications for ligand binding and function (Sugimoto et al., 2004). A key structural feature that distinguishes globins is the coordination state of the heme iron (Fig.1). In hexacoordinate globins, the iron atom in the heme group is coordinated by the proximal histidine on the protein's F helix and by a second, distal histidine acting as an endogenous sixth ligand. In the absence of exogenous ligands such as oxygen, the heme iron is therefore fully coordinated by six ligands. By contrast, pentacoordinate globins possess only the proximal histidine as an intrinsic axial ligand, leaving the sixth coordination site open for reversible binding of external ligands, including O_2, CO, and NO (Pesce et al., 2004; Sawai et al., 2003).

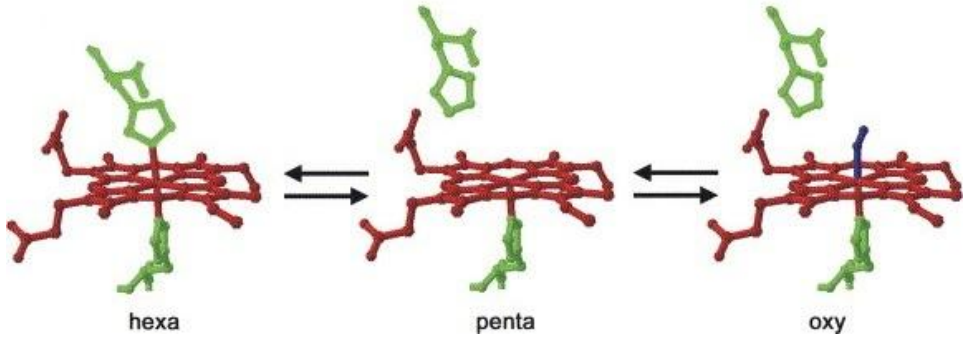


Figure 1: Schematic representation of globin hexa-coordination. In the absence of external ligands, the histidine at position 7 of the E helix (HisE7) occupies the distal sixth coordination site of the heme iron, establishing a hexacoordinate state in which any gaseous ligand (e.g., O_2 or CO) must first displace the internal HisE7 ligand. This competition gives rise to biphasic ligand-binding kinetics in hexacoordinate globins such as neuroglobin (Ngb) and cytoglobin (Cygb), where the rate-limiting step is the slow equilibrium between the hexa- and pentacoordinate forms, followed by a rapid binding phase reflecting the high intrinsic affinity of the heme iron for external ligands once HisE7 is displaced. Colors indicate structural and functional elements: red, heme group; green, interacting histidines; blue, oxygen ligand (Modified from (Hankeln, 2005)).

This difference in coordination has important functional consequences, as it influences ligand affinity, kinetic behavior, and regulatory properties of the protein (Hamdane et al., 2003). Hexacoordination is often associated with slower ligand association and dissociation rates, whereas pentacoordination generally permits more rapid ligand binding. As a result, these two coordination modes are linked to distinct physiological roles, ranging from oxygen transport and storage to redox sensing and enzymatic or signaling functions (Hankeln, 2005).

1.4 Hemoglobin

Hemoglobin (Hb) is the most prominent and extensively studied globin protein and serves as the primary oxygen carrier in the blood of almost all vertebrates, except Antarctic icefish (Corliss et al., 2019). Its main function is the reversible transport of oxygen from the lungs or gills to peripheral tissues and the return transport of carbon dioxide and hydrogen ions to the lungs, contributing to gas exchange and the buffering of blood pH (Hoeger et al., 2020). Beyond this, Hb is a polyfunctional molecule involved in various catalytic processes. It acts as a nitrite reductase, nitric oxide dioxygenase, monooxygenase, alkyl hydroperoxidase, esterase, and lipoxygenase, and participates in nitric oxide metabolism, metabolic reprogramming, pH regulation, and maintenance of redox balance (Hoeger et al., 2020; Kosmachevskaya & Topunov, 2018).

Hb is a tetrameric, iron-containing globular protein composed of four polypeptide chains. In adult hemoglobin A (HbA), these consist of two alpha (α) and two beta (β) chains ($\alpha_2\beta_2$), forming a globular heterotetramer (Marengo-Rowe, 2006). Each subunit contains a heme prosthetic group, so one HbA molecule carries four heme units and can thus bind up to four oxygen molecules (Hoeger et al., 2020). Each heme consists of a protoporphyrin ring with a central ferrous iron ion (Fe^{2+}) coordinated to four nitrogen atoms of the ring and the proximal histidine of the globin chain. At the sixth coordination site, the Fe^{2+} binds molecular oxygen reversibly (Voltarelli et al., 2023). The iron and heme structure also give blood its characteristic red color when oxygenated.

The four subunits of Hb are organized as two $\alpha\beta$ dimers. This quaternary structure enables a conformational switch between a low-affinity "T" (tense) state and a high-affinity "R" (relaxed) state during oxygen binding and release, resulting in cooperative binding (Hoeger et al., 2020; Marengo-Rowe, 2006). This allosteric transition allows hemoglobin to efficiently load oxygen in the lungs or gills and unload it in tissues where it is most needed. Allosteric effectors such as H^+ ions, CO_2, and 2,3-bisphosphoglycerate (2,3-BPG) modulate Hb's oxygen affinity (Wang et al., 2021). The Bohr effect specifically describes the decrease in oxygen affinity with increased CO_2 concentration or lowered pH, as occurs in actively respiring tissues. This right-shift in the hemoglobin-oxygen dissociation curve ensures enhanced oxygen release in metabolically active areas. Conversely, in the lungs or gills, where CO_2 is low and pH is higher, hemoglobin's oxygen affinity is increased, promoting oxygen uptake (Kaufman et al., 2024; Malte & Lykkeboe, 2018). Altogether, the dynamic regulation of oxygen binding and release allows hemoglobin to efficiently supply oxygen throughout the body under varying physiological conditions.

1.5 Myoglobin

Myoglobin (Mb) was the first (globin) protein to be structurally characterized, and is one of the most fundamental oxygen-binding proteins in vertebrates (Kendrew et al., 1958). It is the main oxygen storage molecule in muscle tissue and ensures an immediate supply of oxygen during periods of high metabolic demand or limited oxygen availability, such as during intense exercise or diving (Ordway & Garry, 2004). Unlike Hb, which transports oxygen via the bloodstream, myoglobin facilitates oxygen diffusion within muscle fibers and buffers intracellular oxygen concentration. This maintains aerobic metabolism under hypoxic conditions (Helbo & Fago, 2012).

Mb is a monomeric, iron-containing globular protein composed of a single polypeptide chain containing about 150 amino acids, as well as one heme prosthetic group (Brunori et al., 1971). The heme group consists of a protoporphyrin IX ring with a central ferrous iron ion (Fe^{2+}), which is coordinated by four nitrogen atoms from the porphyrin ring and one nitrogen atom from a proximal histidine residue of the globin. The sixth coordination site of the iron ion binds molecular oxygen reversibly. This heme-iron interaction gives oxygenated muscle tissue its characteristic red color. The tertiary structure of Mb is predominantly α-helical and comprises eight helices, A through H. These helices form a compact hydrophobic pocket that stabilizes the heme part and provides an optimal microenvironment for oxygen binding (Kendrew et al., 1960).

Unlike tetrameric Hb, which exhibits cooperative oxygen binding, Mb displays a simple hyperbolic oxygen-binding curve. This reflects its monomeric nature and non-cooperative binding behavior (Hoeger et al., 2020). This enables Mb to achieve high oxygen affinity under low partial pressures, which is crucial for oxygen extraction in tissues where Hb's affinity begins to decline. Thus, myoglobin functions as a molecular bridge that facilitates oxygen transport from capillaries to mitochondria and plays a key role in maintaining intracellular oxygen homeostasis (Wittenberg & Wittenberg, 2003). Additionally, Mb exhibits enzymatic and protective functions. It acts as a nitrite reductase and nitric oxide scavenger, contributing to redox regulation and cytoprotection under oxidative or nitrosative stress (Flögel et al., 2010; Helbo & Fago, 2012).

Overall, Mb's structure and biochemical properties enable it to serve as a critical component in muscular oxygen storage, diffusion, and metabolic resilience, particularly in species adapted to hypoxic or challenging environments.

Interestingly, several species adapted to hypoxic environments possess multiple copies of the Mb gene, or isoforms (Helfenrath et al., 2021; Koch et al., 2016). This is believed to be an

evolutionary adaptation that optimizes oxygen storage and delivery to tissues under limited oxygen availability. In hypoxia-tolerant fish, such as carp, reedfish and lungfish, multiple Mb genes have been identified and may help maintain aerobic metabolism during prolonged periods of oxygen shortage or air breathing. The diversification of Mb indicates its significant role in physiological adaptation to fluctuating or chronically low oxygen conditions (Helfenrath et al., 2021; Lüdemann, 2019; Roesner, 2007).

1.6 Neuroglobin

Neuroglobin (Ngb) is a globin protein that diverged early in metazoan evolution, likely before the split between *Protostomia* and *Deuterostomia* (Burmester et al., 2000). It is a monomeric, intracellular globin protein that is structurally similar to Mb and Hb. However, in contrast to Hb and Mb it is distinguished by a six-coordinate heme-Fe atom with bis-histidyl ligation. Ngb binds oxygen more tightly than Hb does, though its physiological ligands include not only oxygen (O_2), but nitric oxide (NO) and carbon monoxide (CO). It is predominantly expressed in neural tissues, where its expression is upregulated under hypoxic conditions. Its structure is adapted to reversible gas binding and cellular protection in hypoxic or ischemic environments (Ascenzi et al., 2016; Burmester & Hankeln, 2009; Exertier et al., 2022). The major proposed intracellular functions of Ngb are summarized in Figure 2.

Multiple mechanisms contribute to Ngb's neuroprotective effects. Ngb acts as an intrinsic antioxidant that directly detoxifies reactive oxygen and nitrogen species (ROS/RNS), such as superoxide, hydrogen peroxide, nitric oxide, and peroxynitrite. This protects neurons from oxidative damage during hypoxia and ischemia-reperfusion (Herold et al., 2004; Li et al., 2010). Ngb also localizes to mitochondria, where it interacts with and rapidly reduces cytochrome c. This inhibits apoptosis by maintaining cytochrome c in its reduced state, supporting mitochondrial membrane potential, and promoting ATP production under oxidative stress (Fago et al., 2006, 2008; Raychaudhuri et al., 2009; Tiwari et al., 2018; Wen et al., 2018). These proposed cytoprotective mechanisms are illustrated schematically in Figure 2.

Additionally, oxidized Ngb functions as a guanine nucleotide dissociation inhibitor (GDI), binding to guanosine diphosphate (GDP)-bound Gαi/o subunits and inhibiting GDP/GTP exchange. This releases G$\beta\gamma$ subunits, which activate pro-survival signaling pathways. This GDI activity is oxygen-dependent, enabling Ngb to sense and respond to hypoxic conditions (Wakasugi et al., 2003; Wakasugi & Morishima, 2005). Ngb is widely expressed throughout the central and peripheral nervous systems. In mammals, the highest expression occurs in the

hypothalamus (Fabrizius et al., 2016) and is also highly abundant in the retina (Solar Fernandez et al., 2021).

In ray-finned and cartilaginous fishes, Ngb orthologs are generally present and are assumed to contribute to neuronal oxygen supply, NO metabolism, and protection from oxidative stress, broadly similar to the situation in mammals (Gallagher & Macqueen, 2017; Hankeln et al., 2005). In lungfish, however, a clear Ngb ortholog has so far not been convincingly identified. This is likely due in part to the extremely large and complex genome, which has only recently been assembled and annotated; earlier searches relied mainly on transcriptomic data and failed to detect an unambiguous Ngb transcript. On this basis, it has been proposed that other globins (for example Mb-, Cygb- or Hb-like paralogs) may have functionally replaced classical Ngb-type roles in neuronal oxygen delivery, NO signaling, and redox protection in lungfish (Lüdemann, 2019).

1.7 Cytoglobin

Cytoglobin (Cygb) is a ubiquitously expressed vertebrate globin that was first identified as a distinct family member based on its unique gene structure and amino acid sequence (Burmester et al., 2002; Schmidt et al., 2004). Cygb is primarily localized in fibroblast-like cells of connective tissues, including dermal fibroblasts, hepatic stellate cells, and certain neuronal populations, where it associates with the cytoplasmic face of the endoplasmic reticulum and nuclear membranes. The predominantly intracellular and membrane-associated localization suggests that Cygb contributes mainly to cellular homeostasis and stress protection rather than to systemic gas transport. (Hai & Kawada, 2020; Schmidt et al., 2004).

At the molecular level, Cygb is a single-domain globin carrying a heme prosthetic group that can reversibly bind oxygen and other diatomic ligands, similar to other vertebrate globins. Compared to hemoglobin and myoglobin, its oxygen affinity is relatively low, and it is usually hexacoordinate in the deoxygenated state, with both proximal and distal histidines ligating the heme iron (Smagghe et al., 2008). Transient dissociation of the distal histidine opens the sixth coordination site for ligand binding and allows Cygb to engage in redox-related reactions, including peroxidase-like activity and NO metabolism(Ascenzi et al., 2014; Nishi et al., 2011). Cygb expression is upregulated under oxidative and hypoxic stress and exerts protective effects in several organs. In the liver and lungs, Cygb protects against fibrosis by scavenging reactive oxygen species and modulating collagen deposition, likely via regulation of prolyl hydroxylase-dependent signaling (Kawada, 2015). In the central nervous system, Cygb expression correlates

with resistance to hypoxic-ischemic injury, and its overexpression reduces neuronal death following oxygen-glucose deprivation, indicating a neuroprotective role (Nishi et al., 2011; Schmidt et al., 2004). Moreover, Cygb has been implicated as a tumor suppressor in various cancers (De Backer & Hoogewijs, 2024; Shivapurkar et al., 2008). The major proposed intracellular functions of Cygb are summarized in Figure 2.

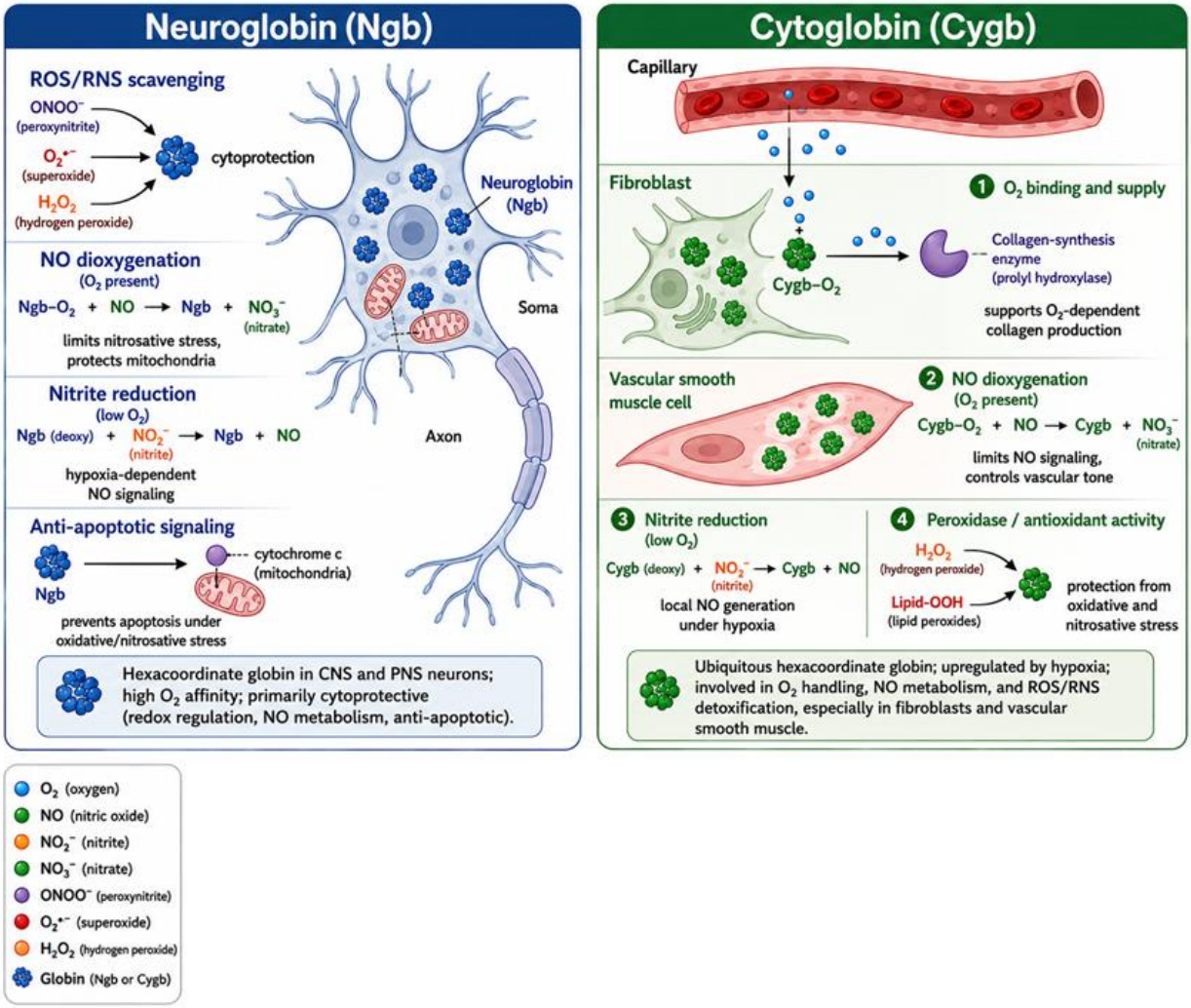


Figure 2: Proposed Intracellular Functions of Ngb and Cygb in Vertebrate Tissues. This schematic overview illustrates the proposed intracellular functions of Ngb and Cygb in vertebrate tissues. Neuroglobin is present in neurons of the central and peripheral nervous systems and is localized in the cytoplasm in close proximity to mitochondria. Its proposed functions include high-affinity oxygen binding, scavenging of reactive oxygen and nitrogen species (ROS/RNS), dioxygenation of nitric oxide (NO) to nitrate under normoxic conditions, reduction of nitrite to NO during hypoxia, and interaction with cytochrome c to limit apoptosis and preserve mitochondrial function. Cytoglobin is depicted as a broadly expressed globin in fibroblasts and vascular smooth muscle cells. Its proposed functions include O_2 binding and delivering O_2 intracellularly to support O_2-dependent collagen synthesis. It also dioxygenates NO under normoxic conditions, reduces nitrite to NO under hypoxic conditions, and exhibits peroxidase activity toward hydrogen peroxide and lipid peroxides. This contributes to antioxidant defense and the regulation of vascular tone. Together, these hexacoordinate globins complement the functions of hemoglobin and myoglobin by integrating O_2 handling, NO metabolism, and redox protection in non-erythroid tissues. Schematic illustration generated with the assistance of ChatGPT (OpenAI, GPT-5.5) and subsequently modified by the author.

Several teleost fishes possess two distinct Cygb genes: cygb1 and cygb2. These genes likely originated from a gene duplication event that occurred during the teleost-specific whole-genome duplication (Burmester & Hankeln, 2014; Fuchs et al., 2005). These paralogs are expressed in different tissues and developmental stages, which suggests that they have undergone subfunctionalization or partial redundancy in their roles in protecting against oxidative stress and maintaining connective tissue homeostasis (Fuchs et al., 2005; Nishi et al., 2011). The presence of two cytoglobin isoforms in many teleosts, in contrast to the single cygb gene in most mammals and other vertebrates, highlights the plasticity of globin gene evolution and the potential for lineage-specific functional diversification (Fuchs et al., 2005).
Similarly to Ngb, Cygb has not yet been detected at the transcriptomic level in lungfish (Lüdemann, 2019).

1.8 Globin E

Globin E (GbE), also referred to as eye-globin, is a vertebrate globin predominantly expressed in the retina. It is thought to contribute to local oxygen homeostasis in highly active visual tissues. In birds, GbE is highly enriched in photoreceptor cells and is believed to support oxygen supply in the avian retina, which lacks a dense capillary network despite its high oxygen demand (Blank, Kiger, et al., 2011; Schwarze et al., 2015). These findings suggest that GbE may function as a specialized respiratory protein that facilitates intracellular oxygen delivery and buffering rather than systemic oxygen transport.
GbE is a pentacoordinate iron-containing heme protein that exhibits one of the highest levels of sequence conservation among globins, comparable to Mb and Hb (Blank, Kiger, et al., 2011). The low oxygen affinity (P50=5.8 Torr at 25 °C) and slow autoxidation rate ($t_{1/2}$≈10 h at 25 °C) of the recombinant chicken GbE are consistent with stable, reversible oxygen binding suitable for respiratory function. Phylogenetically, GbE clusters closely with Mb, which supports its proposed role as a specialized oxygen-binding protein in ocular tissues (Blank, Kiger, et al., 2011). In lungfish, GbE exhibits remarkable functional specialization. Unlike in birds, it is exclusively and highly expressed in oocytes rather than in retinal tissue (Lüdemann et al., 2019). Multiple GbE gene copies - up to 18 across studied lungfish species - have undergone lineage-specific amplification, likely enhancing the oxygen supply to developing embryos during periods of hypoxia. Recombinant lungfish GbE1 displays pentacoordinate heme iron with Mb-like oxygen-binding properties, suggesting an analogous role to that of embryonic Hb in supporting aerobic metabolism under oxygen-limited conditions. This oocyte-specific

expression pattern represents a striking evolutionary divergence of GbE function in hypoxia-adapted vertebrates (Lüdemann et al., 2019).

1.9 Globin Y

Globin Y (GbY) is a penta-coordinated monomeric globin with a characteristic globin fold. However, it exhibits a highly irregular phylogenetic distribution among vertebrates. Notably, it is absent from higher mammals *(Placentalia* and *Marsupialia*), birds, and ray-finned fishes (*Actinopterygii*) (Burmester & Hankeln, 2014). GbY was first discovered in *Xenopus laevis* and occurs in diverse species, including platypuses, lizards, ghost sharks, coelacanths, and turtles (Fuchs et al., 2006; Schwarze et al., 2015).

GbY generally exhibits low expression levels, but shows peak mRNA abundance in the ovary, kidney, eye, and lung tissues. This suggests that it plays specialized roles in oxygen delivery, buffering, or protection within these metabolically active and environmentally exposed organs (Fuchs et al., 2006; Schwarze et al., 2015). During embryonic development, distinct ontogenetic phases exhibit transiently elevated GbY expression, indicating a temporary respiratory support role similar to that of embryonic hemoglobins in other vertebrate lineages (Fuchs, 2008; Schwarze & Burmester, 2013). The combination of GbY's structural conservation, restricted phylogenetic presence, and tissue- and stage-specific expression positions it as a distinctive, niche-adapted member of the vertebrate globin repertoire.

1.10 Globin X

Globin X (GbX) is an ancient member of the globin family found in fish, amphibians, and some reptiles, but not in birds and mammals. Its presence in diverse vertebrate lineages suggests that it has an important cellular function that differs from the classical role of oxygen transport (Blank, Wollberg, et al., 2011; Roesner et al., 2005). Instead of primarily acting as an oxygen carrier, GbX is now understood to be a globin involved in redox regulation, oxidative stress protection, and nitrite-dependent signaling (Corti et al., 2016; Koch & Burmester, 2016).

GbX is distinguished from many other vertebrate globins by its membrane association. GbX is N-terminally acylated, which anchors the protein to intracellular membranes, such as the plasma membrane, and appears to be important for its function. This localization enhances GbX's ability to protect cells from ROS; membrane-bound GbX reduces oxidative stress more

effectively than soluble forms lacking these acyl groups (Koch & Burmester, 2016). In addition to its antioxidant properties, GbX participates in NO metabolism.

Corti *et al.* demonstrated that GbX is a six-coordinate globin found in fish red blood cells that can reduce nitrite to nitric oxide more rapidly than human Hb, Ngb, or Cygb. This reaction contributes to NO signaling and inhibits platelet activation in fish red blood cells, suggesting that GbX is functionally involved in redox and signaling processes (Corti et al., 2016).

1.11 Androglobin

Androglobin (Adgb) is a conserved chimeric globin characterized by a multidomain architecture with an N-terminal calpain-like region and an internal circularly permuted globin domain (Hoogewijs et al., 2012; Reeder et al., 2024). In mammals, Adgb is predominantly expressed in the testes and is associated with late stages of spermatogenesis and ciliogenesis. This suggests that Adgb plays a specialized role in male germ cell development rather than in classical oxygen transport (Hoogewijs et al., 2012).

In mice, loss of Adgb causes male infertility and defects in sperm formation, indicating that Adgb is required for proper spermatogenesis and sperm maturation (Keppner et al., 2022). At the biochemical level, the globin domain of Adgb shows atypical heme stabilization and NO interaction, and it appears to participate in NO homeostasis and calcium signaling rather than in oxygen delivery (Reeder et al., 2024). Altogether, these findings suggest that Adgb is a functionally specialized globin with roles in reproductive biology and redox-related signaling (Hoogewijs et al., 2012).

1.12 Evolution of globins

The globin gene family has a long and complex evolutionary history, shaped by gene duplication, divergence, and lineage-specific gene loss. Comparative analyses indicate that the deuterostome common ancestor possessed at least four distinct globin paralogs that were differentially retained in subsequent lineages (Schuster et al., 2024). The functional diversification of the globin superfamily in vertebrates was strongly influenced by two rounds of whole-genome duplication in the stem lineage (Hoffmann, Opazo, & Storz, 2012). This provided the genetic basis for the evolution of major vertebrate globin lineages (Hoffmann, Opazo, Hoogewijs, et al., 2012; Storz et al., 2013).

It is thought that these duplication events enabled an early division of oxygen transport, storage, and intracellular protective functions. The retained proto-Hb and proto-Mb genes in jawed vertebrates, in particular, supported the separation of blood-borne oxygen transport by Hb and intracellular oxygen storage by Mb. A subsequent tandem duplication in the Hb lineage gave rise to the proto-α and proto-β globin genes, enabling the formation of heteromeric, tetrameric Hbs with cooperative oxygen binding. Cygb, GbE, and GbY also evolved from vertebrate-specific duplication events, though their evolutionary trajectories differ among lineages (Hoffmann, Opazo, & Storz, 2012; Storz et al., 2013).

Not all globin lineages were retained in every vertebrate clade (Fig. 3). Repeated gene loss has also shaped the current globin repertoire. GbE and GbY, for example, have a more restricted phylogenetic distribution and have undergone lineage-specific diversification or secondary loss in several groups. Conversely, Adgb, Ngb and GbX are older globin lineages with ancient origins and specialized cellular roles. These patterns demonstrate that globin evolution was not a simple, linear process but rather a dynamic combination of duplication, functional divergence, and differential retention across vertebrate lineages (Hoffmann, Opazo, & Storz, 2012; Hoffmann, Opazo, Hoogewijs, et al., 2012; Keppner et al., 2020; Storz et al., 2013).

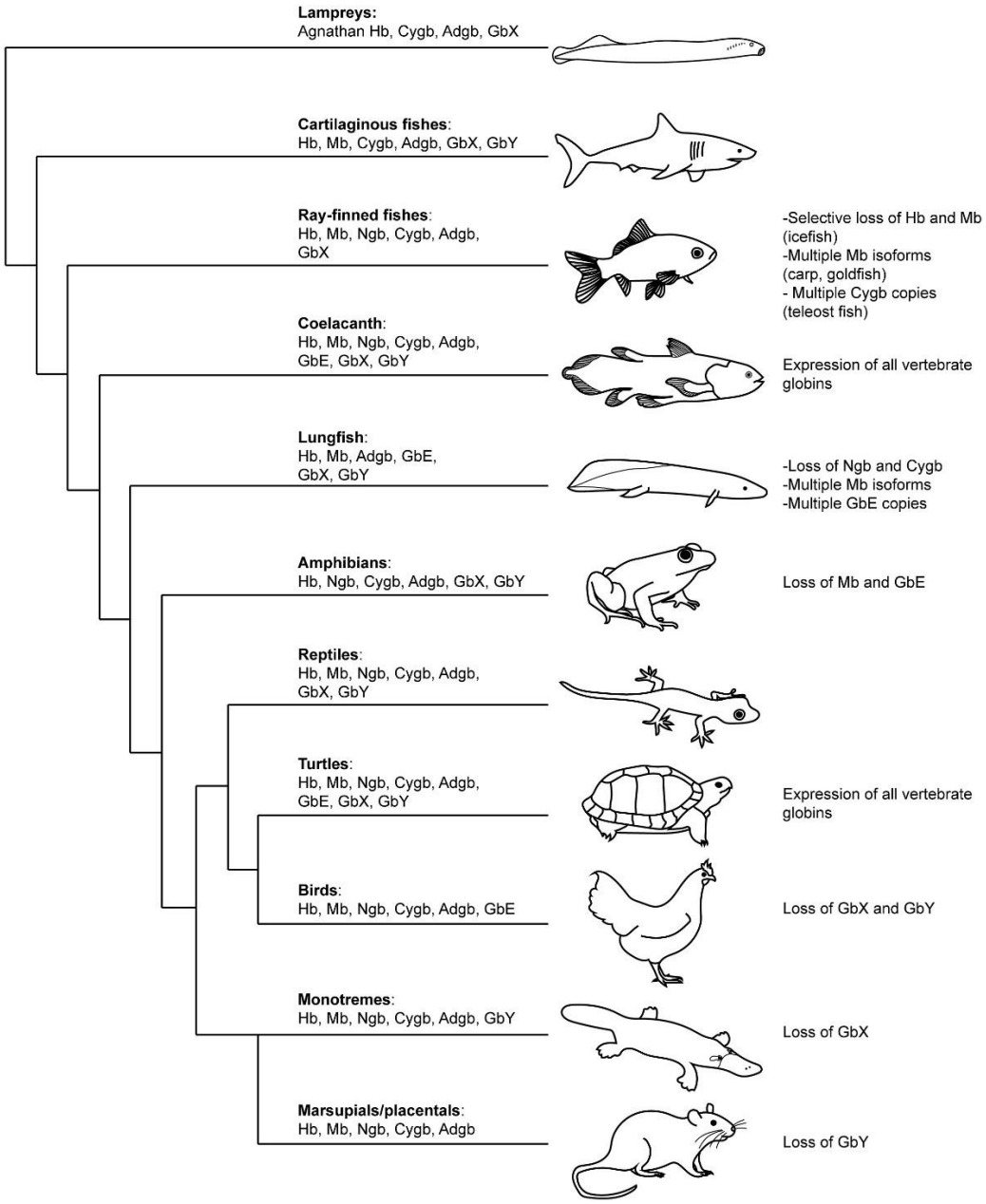


Figure 3: Phylogenetic overview of the globin gene family across vertebrates. Key variations in gene copy number and expression patterns are indicated to the right of each branch, such as the loss of Hb and Mb in certain ray-finned fish lineages, as well as Mb and GbE duplications in lungfish (Keppner et al., 2020).

Note that the phylogenetic overview of the globin gene family (Fig. 3) indicates that Ngb has been lost in lungfish (Keppner et al., 2020; Lüdemann, 2019). This annotation will be examined at the genome level in the present study.

1.13 Lungfish

Lungfishes (*Dipnoi*) are a remarkable lineage of sarcopterygian fishes and the closest living relatives of *tetrapods* (Meyer et al., 2021). This makes them key organisms for studying the

evolutionary transition from water to land. Currently, only six species remain, distributed across three lineages: four species of African lungfishes (*Protopterus*); one species of South American lungfish (*Lepidosiren*); and one species of Australian lungfish (*Neoceratodus*) (Jorgensen & Joss, 2016) (Fig. 4).

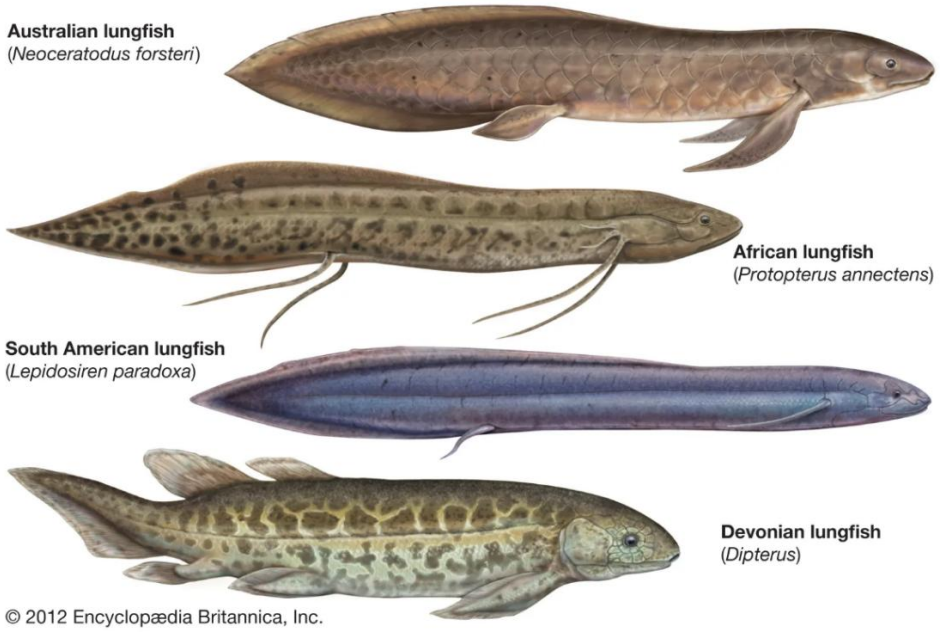


Figure 4: Living and fossil forms of *Dipnoi* fishes. The upper three taxa represent the three living lungfish lineages, while the lower taxon, the Devonian lungfish, represents an ancestral fossil form from the Devonian period. Encyclopædia Britannica, Inc. Available at: https://www.britannica.com/animal/lungfish#/media/1/351582/106876 (accessed July 17, 2026).

Despite their relatively low species diversity, lungfishes exhibit several distinctive ancestral and derived traits that set them apart from other fishes. These characteristics include radically different dental patterns with tooth plates rather than marginal tooth rows, a trait that has been conserved for 360 million years (Kemp et al., 2017; Smith et al., 2002); highly effective dual gas exchange using both air and water (Burggren & Johansen, 1987); and the largest genomes of any vertebrate, with the largest genome being roughly 30 times larger than the human genome (Sousa et al., 2024). Notably, they have up to 40 times the deoxyribonucleic acid (DNA) content of mammalian cells (Thomson, 1972). Lungfishes are evolutionarily distinct because they combine ancient, conserved features with continued adaptive capacity. Their protein-coding genes evolve at remarkably slow rates, reflecting their “living fossil” status, while transposable elements remain highly active and diverse, driving genome expansion (Biscotti et al., 2016; Metcalfe et al., 2012; Sousa et al., 2024).

The combination of their phylogenetic position, morphological specializations, genomic characteristics, and developmental innovations makes lungfishes important for understanding the evolutionary transition from aquatic to terrestrial lifeforms in vertebrates. Given their dual mode of respiration and ability to cope with fluctuating oxygen availability, lungfishes provide an important model for investigating the function and evolution of globin proteins.

1.13.1 Respiration

Lungfishes have highly effective dual gas exchange capabilities and use both air and water for respiration. Their circulatory system displays several plesiomorphic features, including separate atrial chambers, a vertical septum in the ventricle, pulmonary veins, conal valves, and a twisted *bulbus cordis*. The *Lepidosirenidae* (*Protopterus* and *Lepidosiren*) are obligate air-breathers within lungfishes and possess derived features that distinguish them from the facultative air-breather *Neoceratodus*, including paired lungs, reduced anterior gill arches, and a well-developed spiral valve in the conus (Burggren & Johansen, 1987).

In vertebrates, ventilation is stimulated by the partial pressure of O_2 rather than by the total amount of O_2 present. This principle also applies to *Lepidosiren*. When subjected to elevated temperatures, lungfishes and amphibians become increasingly reliant on the lung for the purpose of CO_2 elimination. The Australian lungfish relies primarily on gill breathing but uses air breathing during elevated physiological activity, such as courtship, or in hypoxic environments(Jorgensen & Joss, 2016).

1.13.2 Neoteny and genome size

Neoteny has been hypothesized to play a role in the evolutionary transition between lungfishes and tetrapods. These animals retain juvenile characteristics, such as larval fin structure and partially inactive thyroid regulation, even as adults. During the Devonian period, some lungfish species may have reached reproductive maturity without fully undergoing metamorphosis into the specialized adult form. These neotenic lungfishes may have resembled the ancestral lobe-finned fishes from which tetrapods evolved. Modern lungfishes exhibit several larval characteristics, which suggests evidence for *paedomorphosis*, indicating that their developmental trajectory is similar to that of neotenic amphibians (Joss, 2006). Lungfishes, such as *Lepidosiren paradoxa* and *Neoceratodus forsteri*, have the largest sequenced genomes of any animal, ranging from 40 to over 91 gigabase pairs (Gb), which is vastly larger than that of any other vertebrates (Meyer et al., 2021; Sousa et al., 2024). This genome gigantism is primarily due to the massive accumulation of transposable elements, especially Long

Interspersed Nuclear Element (LINEs), long terminal repeats (LTRs), and short interspersed nuclear elements (SINEs), which make up 67-90% of their DNA (Metcalfe et al., 2012; Meyer et al., 2021; Sousa et al., 2024). Many of these transposable elements are still active, with evidence of ongoing insertion and expression. This contributes to further genome growth. The high repeat content and extremely large intron sizes (some introns up to 5.8 megabase pairs) distinguish lungfish genomes from those of other vertebrates. These features facilitate chromosomal rearrangement while maintaining ancestral karyotype structure (Fig. 5) (Sousa et al., 2024).

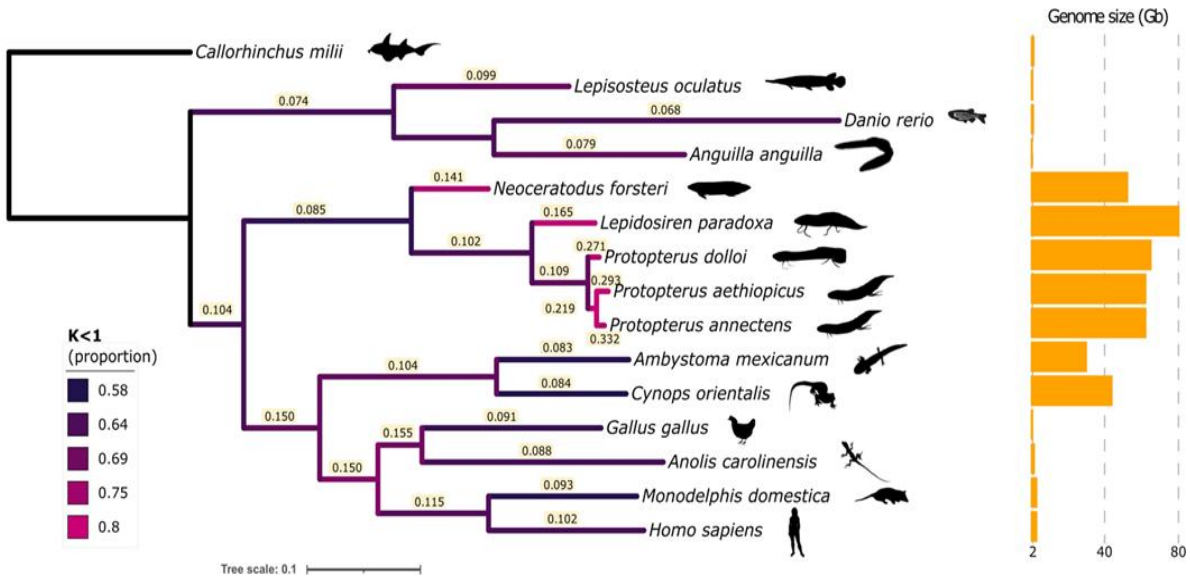


Figure 5: Branch-specific selection pressure (ω and k) and relaxation signatures across vertebrate lineages. Maximum-likelihood phylogeny of 15 vertebrate species (including 5 lungfish) based on 1.790 orthologous genes; all branches have 100% bootstrap support. Scale bar: expected nucleotide substitutions per site. Branch labels show maximum-likelihood ω values (CODEML free-ratio model). Branches are colored by proportion of genes with relaxed selection ($k\omega$_bg ∩ RELAX k < 1; "Relaxed genes") or conserved selection (CODEML ω_fg ≤ 1; "Constrained genes") (Fuselli et al., 2023).

This extreme genome expansion also presents a major practical challenge for gene identification and annotation, as exceptionally long introns and high repeat content can complicate assembly, alignment, and the detection of coding regions by standard bioinformatic tools. Therefore, multiple complementary approaches were used in the present study to investigate the target gene.

1.13.3 Evolutionary link to tetrapods

Lungfishes are genetically and morphologically the closest living relatives to tetrapods. Their limb-like fins and air-breathing capabilities are similar to those of early land vertebrates (Meyer

et al., 2021; Sousa et al., 2024). Studying extant lungfish lineages helps us to uncover the genomic and developmental mechanisms that underlie the transition from aquatic to terrestrial life. This study will also help us to reconstruct the crucial role that lungfishes played in the conquest of land by vertebrates. The combination of their phylogenetic position, morphological specializations, genomic characteristics, and developmental innovations makes lungfishes essential for understanding this evolutionary transition. Beyond their relevance for vertebrate evolution more broadly, lungfishes are of particular interest because they possess an unusually diverse and lineage-specific globin repertoire.

1.14 Multiple globins in Lungfish

Lungfishes exhibit a remarkably diverse globin repertoire characterized by multiple paralogous genes for several globin types. This diversity likely resulted from lineage-specific gene duplications. The expansion is especially evident in Mb, with seven copies, and GbE, with 5-18 copies, whereas other globins show single-copy retention or the apparent loss of Ngb and Cygb (Helfenrath et al., 2021; Koch et al., 2016; Lüdemann et al., 2019).

1.15 Aim of the thesis

The aim of this thesis is to investigate the genomic architecture and globin repertoire of the Australian lungfish in order to better understand how their unique genetic properties may have contributed to vertebrate adaptation during the water-to-land transition. Because lungfish genomes are exceptionally large and highly repetitive, a major challenge of this study is to identify globin genes within this genomic background and to evaluate which bioinformatic approaches are most suitable for detecting individual genes in such complex assemblies. In particular, this thesis focuses on the putative presence, genomic context, and evolutionary conservation of Ngb and Cygb, which have not yet been conclusively identified in lungfish. If these genes are present, further questions concern their transcriptional evidence and predicted heme coordination, as well as their putative functional properties. In addition, exploratory cell culture-based experiments were initiated as a preliminary step toward future functional analyses. Together, these approaches aim to contribute to a better understanding of globin evolution in lungfish and its relevance to vertebrate adaptation during the water-to-land transition.

2 Chapter 1 - Identification of globin genes in the lungfish

2.1 Abstract

The Australian lungfish (*Neoceratodus forsteri*) is the closest living relative of tetrapods. It possesses an exceptionally large (>40 Gb) and repeat-rich genome, which poses significant challenges for functional gene annotation. Here, the BRAKER3 pipeline has been used, which integrates *ab initio* predictions, RNA-seq evidence and protein homology, to generate a high-confidence gene annotation. Repetitive elements were soft-masked using RepeatMasker with a custom lungfish repeat library to enhance accuracy. Using this annotation, putative globin genes have been systematically identified through iterative BLAST and Double Index Alignment of Next-generation sequencing Data (DIAMOND) searches against broad vertebrate protein databases and a globin-specific reference set. This yielded 103 globin gene candidates. Eighteen high-confidence globin sequences were selected from these candidates for phylogenetic analysis and validation of the expression on transcriptome level via RNA-seq backmapping. These findings provide the first comprehensive overview of globin diversity in *N. forsteri* and highlight its evolutionary expansion in early sarcopterygians. This work establishes a robust genomic framework for future comparative studies on oxygen transport adaptations in vertebrates.

2.2 Introduction

The globin family is a conserved group of proteins found in nearly all organisms that plays a central role in intracellular oxygen transport and storage (Keppner et al., 2020; Wajcman et al., 2009). Globins are of particular evolutionary interest because they illustrate how a conserved protein family can diversify into distinct functional roles, including oxygen transport, oxygen storage, and cellular stress responses, thereby providing a useful framework for studying adaptation across major vertebrate lineages (Vinogradov & Moens, 2008). Comparative analysis of globin genes in early tetrapod relatives provides valuable insight into globin evolution and functional adaptation under diverse physiological conditions (Keppner et al., 2020; Storz et al., 2013). This includes challenges associated with the transition from aquatic to aerial respiration in early sarcopterygians, where variable oxygen levels likely promoted functional diversification within the globin family (Hsia et al., 2013).

The Australian lungfish, *N. forsteri*, holds a unique phylogenetic position as the closest living relative of tetrapods (Irisarri & Meyer, 2016; Meyer et al., 2021). With a genome exceeding 40 Gb and an exceptionally high repeat content of approximately 90%, it is one of the largest and most complex vertebrate genomes described to date (Metcalfe et al., 2012; Meyer et al., 2021; Schartl et al., 2024). The predominance of repetitive elements, largely derived from transposable elements, further complicates genome assembly and functional annotation by increasing fragmentation and ambiguity in gene-rich regions (Schartl et al., 2024).

These genomic peculiarities complicate assembly and functional annotation, particularly for gene families such as the globins, whose frequent duplications, short exon-intron structures, and high sequence similarity increase the likelihood of fragmented or misassigned gene models in repeat-rich regions. In addition, the unusually large genome size exceeds the technical limits of many bioinformatic tools, making it necessary to split the genome into smaller segments for analyses; this also restricts the ability to visualize the genome or gene structures in standard genome browsers. Together, these constraints underscore the need for adapted bioinformatic approaches and substantial computational resources.

Despite its high evolutionary relevance, a systematic characterization of the globin repertoire in *N. forsteri* has not yet been established, and the number, classification, and evolutionary relationships of putative globin genes remain insufficiently resolved (Irisarri & Meyer, 2016; Meyer et al., 2021). Previous investigations have primarily examined phylogenetic relationships or individual globin variants, mostly in African lungfish species (Lüdemann, 2019), rather than establishing a comprehensive globin repertoire for *N. forsteri.*

A genome-wide identification has so far been hindered by the exceptional size and repeat-rich structure of the lungfish genome, which complicates the reconstruction of complete gene models and increases the likelihood of fragmented or misassigned paralogs. Therefore, a comprehensive identification and classification of the globin repertoire in the Australian lungfish genome has yet to be conducted.

Recent annotation pipelines integrating *ab initio* prediction with transcriptomic and homology-based evidence can improve gene model accuracy, even in complex genomes. This makes them particularly suitable for non-model species, such as *N. forsteri* (Gabriel et al., 2024; Hoff et al., 2019).

Therefore, the aim of this study was to generate a high-quality genome annotation for *N. forsteri* and to use this as the basis for a systematic identification, classification, and validation of globin genes. The BRAKER3 pipeline has been employed, which combines *ab initio* predictions with RNA-seq evidence and protein homology. Repetitive elements were soft-masked using a

species-specific repeat library to reduce annotation artifacts (Smit et al., 2015a; Tempel, 2019). Putative globin candidates were identified through BLAST- and DIAMOND-based similarity searches against vertebrate protein databases and a custom globin sequence set (Buchfink et al., 2015; Wolfsberg & Madden, 1999). Candidate sequences were subsequently validated through multiple sequence alignment, domain inspection, phylogenetic inference, and RNA-seq backmapping to confirm expression and structural completeness.

Together, these steps provide the first comprehensive assessment of the globin gene repertoire in the Australian lungfish and establish a foundation for reconstructing globin evolution in early sarcopterygians.

2.3 Results

2.3.1 Genome annotation of *N. forsteri*

Annotation of the genome was performed using BRAKER3 on a RepeatMasker-processed assembly. On chromosome 1 alone, BRAKER3 predicted a total of 130,481 gene models, comprising annotated features such as genes, transcripts, exons, introns, and CDS (coding sequences) generated by AUGUSTUS. Multi-exon gene structures were often fragmented. Gene counts for all remaining chromosomes are provided in the supplementary material.

The unusually large genome size (>40 Gb) introduced several technical constraints during the annotation process. For example, memory limitations required RNA-seq mapping to be performed chromosome-wise and region-specific inspection. Similarly, genome visualization tools such as the Integrative Genomics Viewer (IGV) was not able to load full chromosome sequences, requiring manual inspection of extracted genomic regions around the predicted globin genes. The resulting annotation dataset was used for globin identification in subsequent analyses.

2.3.2 Detection of globin candidates

Before the genome annotation was completed, a nucleotide BLAST search found nine possible globin-like regions in the genome assembly. After annotation, DIAMOND BLASTp searches of the predicted proteins against the Universal Protein Resource (UniProt) vertebrate database revealed ten further candidates. A targeted DIAMOND search using a custom globin-specific reference database revealed 103 globin-like sequences (supplementary material) spanning the entire genome (Fig. 6). Many of these candidates consisted of truncated open reading frames

(ORFs), short isolated exons, or duplicated hits originating from highly similar genomic regions. Eighteen high-confidence globin gene models were retained for downstream classification and phylogenetic analyses based on alignment quality, domain completeness, and the presence of conserved globin motifs (Tab. 1).

Candidate screening was performed on a chromosome-by-chromosome basis. In addition to the loci associated with chromosomes, several globin candidates were located on unplaced scaffolds. Notably, Cygb was found on scaffold 17, and GbY on scaffold 18.

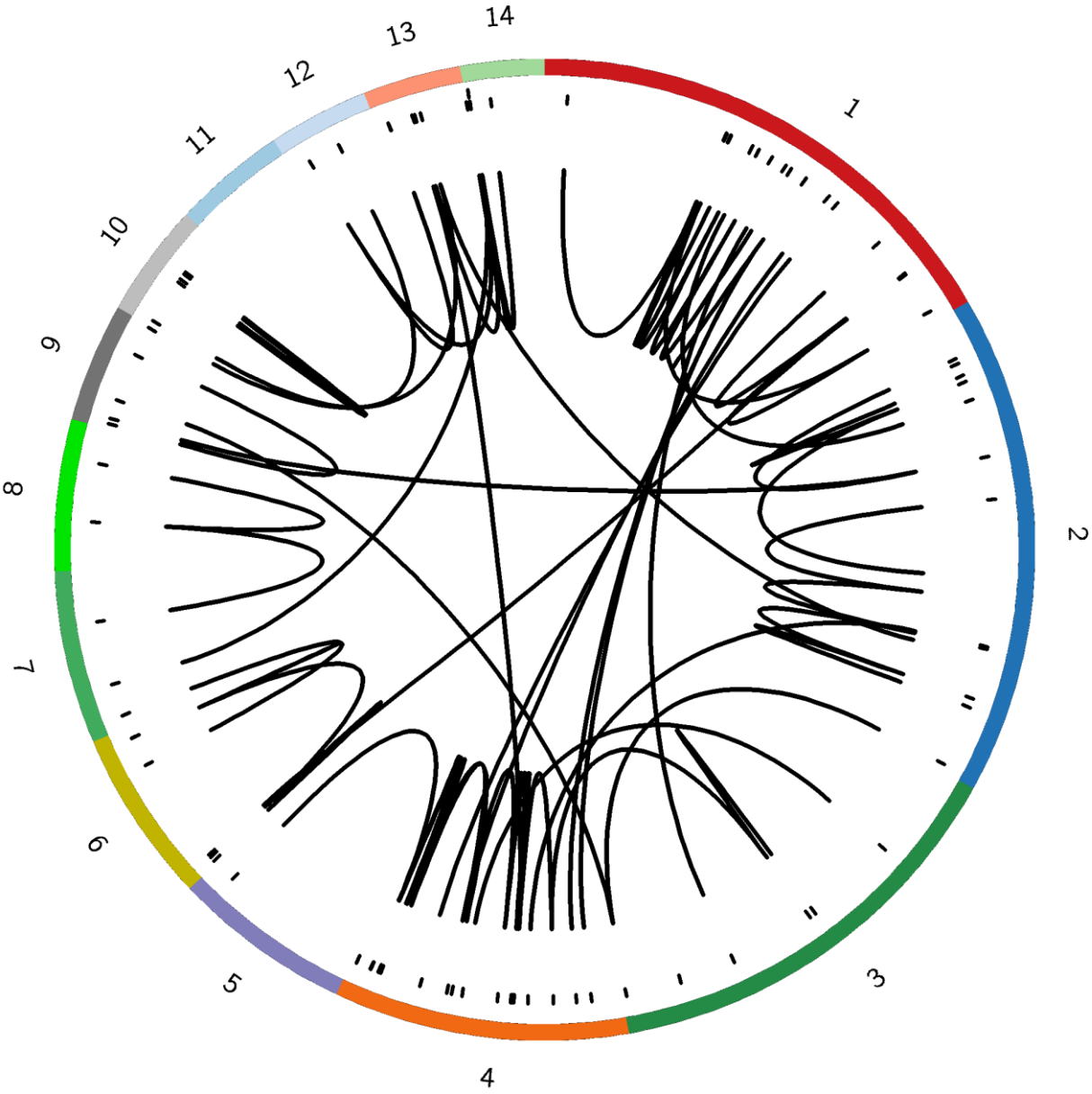


Figure 6: The distribution of the different globin candidates across chromosomes 1-14. The relative locations of the identified globin sequences are marked with black ticks. The 14 chromosomes are shown in different colors on the outer circle. Black lines in the inner circle indicate inter- and intrachromosomal connections between the same globin types. The figure was generated using the CIRCOS (v0.69-8) tool on the GALAXY (v galaxy10) server of the University of Hamburg, Department of Biology.

2.3.3 Localization of globin genes within the genome

To characterize the spatial organization of globin candidates, their genomic coordinates were mapped onto assembled chromosomes and unplaced scaffolds (Fig. 6). Globin genes were distributed across multiple chromosomes, and several local gene clusters were present. The largest number of globin candidates was located on chromosome 1, the largest chromosome of *N. forsteri* (4.9 Gb). Additional globin sequences were found on unassigned scaffolds, including Cygb on scaffold 17 and GbY on scaffold 18. This is consistent with the presence of unplaced microchromosomal regions in the current genome assembly.

Table 1: Globin candidates that were used for further investigation. Overview of the utilized names, the location of the genes (numbers of chromosomes and chromosome parts) and the predicted protein length (number of AA).

Name	Globin	Location	Length (AA)
Nfo_Cygb	Cygb	scaffold 17	175
NfoG2	GbE1	1_2	74
NfoG12	GbE2	1_2	151
NfoG13	GbE3	1_2	151
NfoG57	GbE4	1_1	177
NfoG3	Mb1	1_2	154
NfoG14	Mb2	1_2	159
NfoG5	Mb3	1_2	153
NfoG6	Mb4	1_2	153
NfoG15	Mb5	1_2	154
NfoG4	Mb6	1_2	157
NfoG1	Ngb	10	161
NfoG11	GbX	1_2	222
NfoG46	GbY	scaffold 18	179
NfoG18	Adgb (partial)	1_2	107
NfoG110	Hb1 (β)	14	153
NfoG111	Hb2 (β)	14	211
NfoG112	Hb1 (α)	14	124

2.3.4 Phylogenetic analysis

A Bayesian phylogenetic analysis was performed to classify the 18 high-confidence globin candidates using reference sequences representing major vertebrate globin types (Fig. 7). The alignment included 93 sequences and 318 amino acid positions. The reconstructed phylogeny revealed well-supported clades that corresponded to established globin families. The *N. forsteri* sequences grouped with the expected vertebrate globin families, including Mb, GbE, GbX,

Ngb, Cygb, and GbY lineages, as well as the Hb *α/β* components. Posterior probability values were obtained for all nodes, which support the placement of the lungfish sequences within these globin groups.

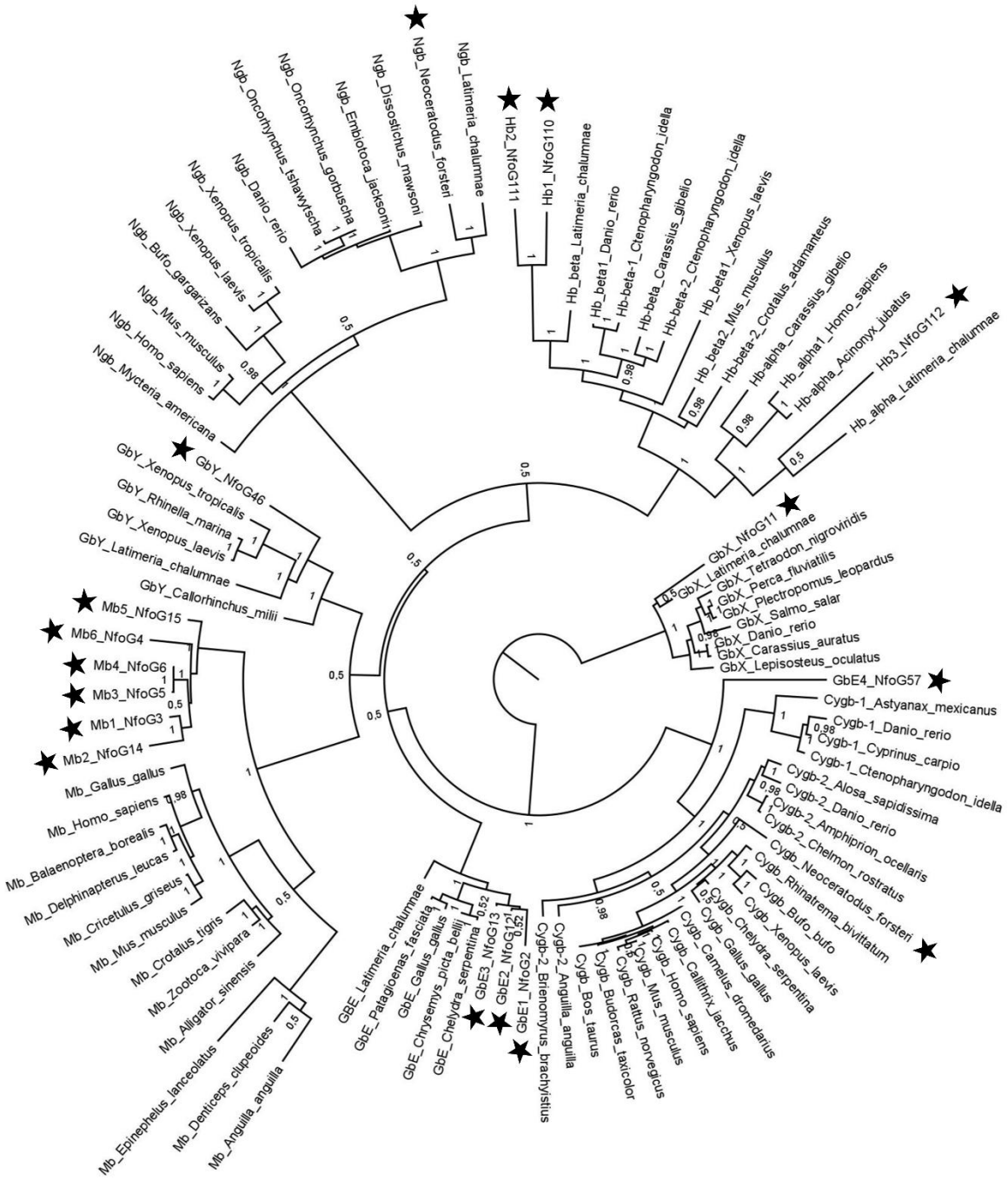


Figure 7: Bayesian phylogenetic tree of representative vertebrate globin gene family members and the *N. forsteri* globin genes (NfoG). *N. forsteri* sequences are marked with a star. Posterior probability support values are displayed on every reconstructed node. A total of 93 sequences were analyzed with 318 aligned AA sites (supplementary material) using the Jones-Taylor-Thornton matrix with gamma distribution. Note that the Adgb candidate was excluded from the analysis due to its permuted globin domain.

2.4 Discussion

2.4.1 Summary of key findings

This study provides the first comprehensive, annotation-based assessment of the globin gene repertoire in *N. forsteri*. Through the integration of species-specific repeat masking, evidence-guided gene prediction, and multiple similarity-based approaches, 103 globin candidates were identified. Eighteen high-confidence sequences were retained for subsequent analyses. Phylogenetic inference and RNA-seq expression evidence further validated classifying these candidates into the major vertebrate globin families: Mb, GbE, GbX, Ngb, Cygb, and GbY, as well as the Hb *α*/*β* components. These results reveal that *N. forsteri* possesses a complete and diverse globin repertoire, offering new insights into globin evolution in early sarcopterygians.

2.4.2 Annotation quality and methodological considerations

BRAKER3 was effective in generating gene models in the exceptionally large and repeat-rich *N. forsteri* genome. Integrating RNA-seq evidence, protein homology, and *ab initio* prediction reduced fragmentation in many loci compared to unassisted approaches. Nevertheless, annotating this genome presented considerable challenges. Gene counts at the chromosomal level were markedly inflated, exceeding biologically plausible values for vertebrates. Previous studies have estimated that there are approximately 21,000 protein-coding genes in *N. forsteri* (Schartl et al., 2024), and vertebrate genomes typically contain around 19,000 to 25,000 protein-coding genes (Clamp et al., 2007; Prachumwat & Li, 2008). The extremely high numbers observed here are therefore best explained by overprediction, a well-documented limitation of gene annotation tools applied to large, repeat-rich genomes.

In genomes dominated by transposable elements, short open reading frames, repetitive motifs, and fragmented alignments may be misclassified as independent genes during *ab initio* prediction. BRAKER3 reduces these effects by integrating external evidence, but the complexity of the lungfish genome means that some overannotation is unavoidable. These technical limitations underscore the importance of integrating transcriptomic data, protein evidence, and domain-specific information, such as Protein Profile eXtension (PPX) files, to enhance the accuracy of annotations in non-model organisms.

2.4.3 Detection, classification and genomic context of globin genes

The identification of globin genes in *N. forsteri* demonstrated the value of different search strategies. Initial BLASTn queries on the unannotated genome detected only nine globin-like regions. Annotation-based DIAMOND BLASTp searches increased this number to 18. The most extensive recovery came from a targeted DIAMOND search against a custom database of verified vertebrate globin sequences alone, yielding 103 candidates.

Many of these corresponded to incomplete, truncated, or duplicated gene fragments, which is a common characteristic of gene families located within repeat-rich genomic regions (Zhi et al., 2006). After filtering based on motif presence, domain completeness, and alignment length, 18 high-confidence globin genes remained.

Mapping the genomic positions of these sequences revealed their dispersed distribution across chromosomes, with several local clusters. Consistent with its size and gene density, chromosome 1 harbored the largest number of candidates. Additional globin genes, including Cygb and GbY, were found on unplaced scaffolds. This suggests that some of the globin repertoire is located within unresolved chromosomal regions of the current assembly.

2.4.4 Evolutionary implications of the globin repertoire

The composition of the globin repertoire in *N. forsteri* provides insight into the evolution of oxygen-binding proteins in early sarcopterygians. The recovery of multiple Mb and GbE variants resembles patterns described in the West African lungfish (*Protopterus annectens*) (Koch et al., 2016; Lüdemann et al., 2019), suggesting lineage-specific expansions in lungfish. The presence of Cygb, Ngb, and GbX indicates the conservation of the core vertebrate globin families that are involved in cellular oxygen homeostasis, signaling, and protection against oxidative stress.

N. forsteri is a facultative air breather that relies on both gills and lungs. This dual respiratory physiology may have influenced globin diversification. Fluctuating oxygen availability in freshwater habitats, particularly during seasonal hypoxia, may have driven the functional specialization of globin paralogs. Therefore, the broad representation of globin families in *N. forsteri* may reflect ancient sarcopterygian traits as well as adaptive responses to ecological pressures experienced by modern lungfish.

2.4.5 Limitations

There are several limitations that need to be considered. First, the current lungfish genome assembly contains numerous unplaced scaffolds and unresolved regions, particularly among microchromosomes. Consequently, the chromosomal context of some globin genes is incomplete. Additionally, transcriptomic data were restricted to available public RNA-seq datasets, which may not capture rare or stage- and tissue-specific globin expression patterns. Reliance on reference databases introduces potential bias against highly divergent or lineage-specific globins. Some of the initial 103 candidates likely represent pseudogene fragments or annotation artifacts, which highlights the need for improved genome assemblies and long-read transcriptomics. Phylogenetic analyses may also be affected by limited taxon sampling or sequence divergence among ancient globin families.

2.4.6 Conclusions and perspectives

Despite these challenges, this study provides the most comprehensive overview to date of globin diversity in the Australian lungfish. The identification of a complete vertebrate globin repertoire, supported by gene annotation, phylogenetic inference and expression data, establishes *N. forsteri* as a valuable model for studying the evolution of oxygen-binding proteins in early tetrapods.

2.5 Material and methods

2.5.1 Genome preprocessing and repeat masking

Prior to gene prediction, the *N. forsteri* genome assembly (National Center for Biotechnology Information (NCBI) BioProject: PRJNA644903) was screened for repetitive elements using RepeatMasker (v4.1.4 with Database of repetitive DNA families (Dfam)-v3.7). A species-specific repeat library (supplementary material) was previously generated with RepeatModeler (v2.0.3) and used to annotate repetitive sequences. The genome was soft-masked to preserve genomic context while preventing repetitive regions from being misinterpreted as coding sequence during downstream annotation. Gene prediction was further supported by protein evidence processed with ProtHint, which converted homologous protein alignments of related vertebrates into annotation hints for BRAKER3/AUGUSTUS. This approach helps reduce false-positive exon predictions and improves overall annotation accuracy in repeat-rich genomes.

2.5.2 Genome annotation

Genome annotation was performed using BRAKER3 (v3.0.1), which integrates *ab initio* prediction with RNA-seq and protein evidence. RNA-seq datasets from BioProject PRJNA605733 were downloaded and used to generate external hints. Reads were aligned chromosome-wise due to memory limitations imposed by the >40 Gb genome size.

GeneMark-ETP (v4.71) and AUGUSTUS (v3.7.0) were guided by RNA-seq-derived intron, exon, start- and stop-codon hints. AUGUSTUS was trained specifically for *N. forsteri* using high-confidence gene models generated during the initial prediction phase. PPX files derived from InterPro domain searches were incorporated to improve recognition of globin-specific structural features during training. The final annotation consisted of AUGUSTUS predictions informed by both transcriptomic and protein homology evidence. An overview of the bioinformatic pipeline is shown in figure 8.

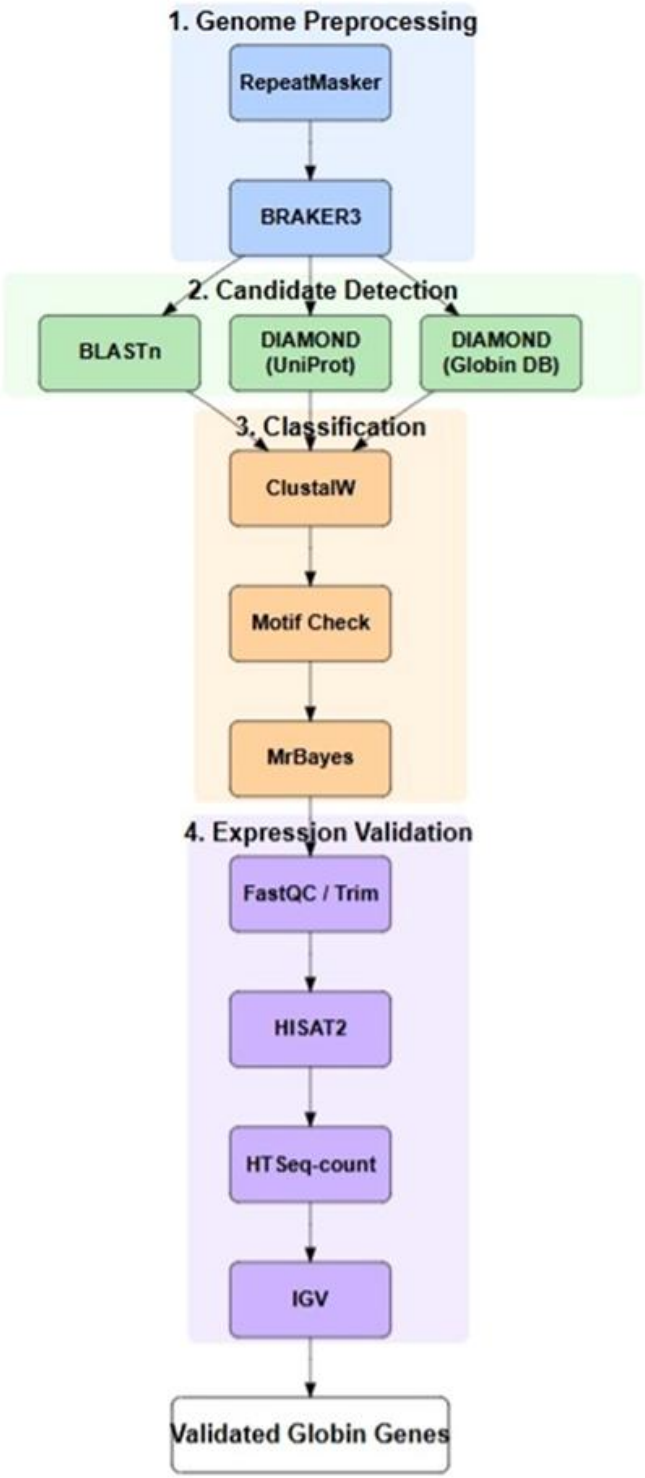


Figure 8: Schematic overview of the bioinformatic pipeline for globin gene identification and validation. Genome preprocessing (RepeatMasker, BRAKER3) is followed by candidate detection (BLASTn, DIAMOND), sequence classification (ClustalW, motif analysis, MrBayes), and RNA-seq-based expression validation (FastQC/Trim, HISAT2, HTSeq-count, IGV), resulting in a set of validated globin genes. The figure was generated in R using ggplot2.

2.5.3 Detection of putative globin genes

Putative globin genes were identified using a multi-step similarity-based workflow.

First, a nucleotide BLAST search was performed on the unannotated genome assembly to detect initial globin-like regions, using different globin gene sequences of related vertebrate species as query. After annotation, predicted proteins were queried with DIAMOND BLASTp against the UniProt vertebrate protein database.

A more targeted search was carried out using a custom globin reference database containing verified globin sequences (supplementary material) representing all major vertebrate globin families. This DIAMOND search identified 103 putative globin candidates across chromosomes and unplaced scaffolds. All candidates were retained for downstream screening.

2.5.4 Sequence filtering and classification

Candidate sequences were filtered based on globin domain completeness, alignment length, presence of conserved motifs and clustering behavior with known reference sequences. Multiple sequence alignments were generated using standard alignment tools, and sequences lacking characteristic globin features were excluded.

The remaining 18 high-confidence sequences were assigned to specific globin families based on conserved residues, structural features and phylogenetic clustering. Chromosomal locations and scaffold assignments were extracted from the BRAKER3 annotation and recorded for each candidate.

2.5.5 RNA-seq backmapping for expression validation

The following bioinformatic tools were used in the respective order for the specified purposes.

- FastQC (v0.12.1): Was employed to generate an initial quality control report of raw sequencing reads by assessing measures such as per-base sequence quality, GC content, and adapter contamination (Andrews & Krueger, 2010).

- Trimmomatic (v0.39): Served to remove adapter sequences and low-quality bases from reads to improve data quality for downstream analysis (Bolger et al., 2014).

- FastQC (second run): Re-run after trimming to assess the effectiveness of the trimming step and confirm an improvement in overall sequence quality.

- HISAT2: Was used to index the reference genome (either the entire genome or selected regions) and align the processed reads to the reference, generating mapping files in SAM/BAM format (Zhang et al., 2021).

- Samtools (v1.16.1): Was utilized to manipulate alignment files in SAM/BAM format, including sorting, indexing, and converting between formats as needed for downstream analyses (Danecek et al., 2021).

- HTseq (v0.13.5): Was applied to count the number of reads mapping to each gene, producing raw gene count data for differential expression or downstream analyses (Anders et al., 2015).

- IGV (Integrative Genomics Viewer; v2.13.2): Was used to visually inspect and explore mapped reads and alignment results in the context of the reference genome to facilitate quality control and interpretation of mapping results (Thorvaldsdóttir et al., 2013).

2.5.6 Phylogenetic analyses

Phylogenetic analyses were conducted using MrBayes. The alignment comprised 318 AA positions, which were treated as a single data partition. The Jones amino acid substitution model was applied (aamodelpr=fixed[jones]). Rate heterogeneity across sites was modeled using a gamma distribution (rates=gamma). Bayesian inference was performed using four parallel Markov chain Monte Carlo (MCMC) chains, consisting of one heated and three cold chains (nchains=4). The chain temperature was set to 0.1 to facilitate chain swapping. The MCMC run was carried out for 5,000,000 generations (ngen=5000000), sampling every 500 generations (samplefreq=500). Status updates were printed every 2000 generations (printfreq=2000), and convergence diagnostics were calculated every 5000 generations (diagnfreq=5000). Upon completion of the runs, posterior parameter distributions (sump) and the consensus tree with posterior probabilities (sumt) were summarized. Phylogenetic trees were visualized and edited using FigTree (v1.4.5).

2.6 Acknowledgements

I would like to thank my former colleague Dr. Raphael Koll for introducing me to UNIX and for sharing his scripts and troubleshooting advice.

For this work the HPC-cluster Hummel-2 at University of Hamburg was used. The cluster was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 498394658.

3 Chapter 2 - Neuroglobin and cytoglobin in lungfish

3.1 Abstract

Neuroglobin (Ngb) and cytoglobin (Cygb) are two conserved, hexacoordinated vertebrate globins, but the evolutionary origins and functions of these proteins in early sarcopterygians remain unclear. As lungfish are the closest living relatives of tetrapods, the Australian Lungfish *Neoceratodus forsteri* (*N. forsteri*) is a key model for investigating the evolution of vertebrate globins. In this study, we identified and characterized Ngb and Cygb homologues from *N. forsteri* at sequence, structural, phylogenetic and spectroscopic levels. Both genes were successfully amplified from cDNA, Ngb expression was detected in brain tissue and Cygb expression in heart tissue. Sequence alignments confirmed the typical globin fold and the presence of the conserved HisE7 and HisF8 residues characteristic of hexacoordinated globins. Phylogenetic analyses unambiguously placed the lungfish sequences within the vertebrate Ngb and Cygb clades. AlphaFold structural predictions showed high similarity to human reference structures and indicated the presence of canonical hexacoordination. UV-visible spectroscopic analyses experimentally confirmed this coordination state. Overall, our results demonstrate that both Ngb and Cygb are present, expressed and structurally conserved in *N. forsteri*. These findings support the hypothesis that these globins were already present in the last common ancestor of lungfish and tetrapods, and they contribute to our understanding of the early functional diversification of vertebrate globins.

3.2 Introduction

Globins are an ancient, functionally diverse protein family widely distributed across vertebrates. In addition to the classic oxygen-transport proteins hemoglobin and myoglobin, vertebrates possess several other globin types, including neuroglobin (Ngb) and cytoglobin (Cygb) (Reuss, 2025). Both are hexacoordinated heme proteins that play a role in intracellular oxygen-related processes and cellular stress responses (Reeder, 2023; Thuy et al., 2026). Ngb is primarily found in neural tissue and is involved in maintaining neuronal oxygen homeostasis and protecting against hypoxia (Yañez et al., 2025), whereas Cygb is more broadly expressed

and may contribute to defending against oxidative stress, handling oxygen, and tissue-specific metabolic functions (Burmester & Hankeln, 2014; Reeder, 2023).

From an evolutionary perspective, Ngb and Cygb are of particular interest because they represent divergent globin lineages with distinct phylogenetic histories. Comparative analyses have shown that Ngb belongs to an ancient branch related to invertebrate nerve globins. In contrast, Cygb is more closely related to myoglobin-like globins (Burmester & Hankeln, 2014; Keppner et al., 2020; Reuss, 2025). These findings suggest that the functional diversification of tissue globins occurred early in vertebrate evolution. However, the exact physiological roles of Ngb and Cygb remain unclear.

Lungfishes (Dipnoi) are crucial for studying the evolution of vertebrate globin genes because they are the closest extant relatives of tetrapods (Meyer et al., 2021). The Australian lungfish (*N. forsteri*) is especially significant due to its phylogenetic position, as well as its conserved morphology and physiology (Schartl et al., 2024). Previous studies on lungfish globins have mainly focused on Hbs (Rasmussen et al., 2009; Weber & Jensen, 2011) and Mbs (Koch et al., 2016). To date, neither Ngb nor Cygb has been detected in lungfish. It has been proposed that other globins in this lineage may instead fulfill their functions (Lüdemann, 2019). With the publication of lungfish genome data in 2021 (Meyer et al., 2021), it has become feasible to reexamine this question. Due to the phylogenetic position of lungfish as the closest living relatives of tetrapods, identifying the presence or absence of Ngb and Cygb in *N. forsteri* is crucial for understanding the evolution and potential functional diversification of vertebrate globins. Therefore, determining whether these globins are present, expressed, and structurally conserved in *N. forsteri* is essential to clarifying their evolutionary history in sarcopterygians.

This study investigates Ngb and Cygb homologs in *N. forsteri* at sequence, phylogenetic, structural, and spectroscopic levels. Through a combination of transcript amplification, sequence alignment, phylogenetic reconstruction, structural modeling, and UV-visible spectroscopy, we aim to determine if the lungfish sequences align with the typical vertebrate Ngb and Cygb sequences and to evaluate their coordination properties. These analyses are intended to clarify the evolutionary status and potential functional conservation of Ngb and Cygb in lungfish.

3.3 Results

3.3.1 Identification and tissue-specific expression of Ngb and Cygb

We verified the presence and expression of the previously predicted Ngb and Cygb genes in *N. forsteri* (chapter 1) by amplifying both transcripts from cDNA derived from various tissues. Successful PCR amplification confirmed that both genes are expressed in *N. forsteri*. Ngb transcripts were amplified from brain cDNA, whereas Cygb transcripts were amplified from heart cDNA. PCR products were of the expected size and sequence verification confirmed their identity as *N. forsteri* Ngb and Cygb.

3.3.2 Sequence characteristics and conserved functional residues

To confirm the identity of the obtained sequences, we analyzed the secondary structure of the predicted *N. forsteri* Ngb and Cygb amino acid (AA) sequences (Fig. 9, 10).

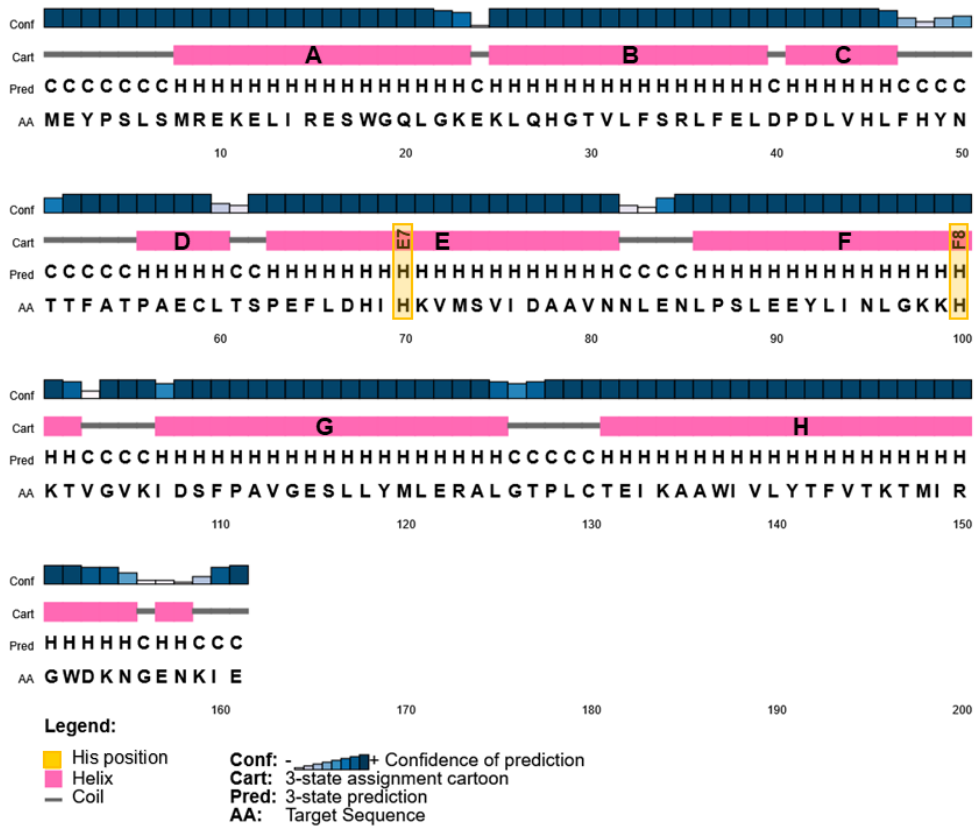


Figure 9: Secondary structure prediction of the *N. forsteri* Ngb protein using PSIPRED. The prediction was performed using sequence profile searches against the UniClust30 (2018_08) and UniRef90 databases with HHblits and PSI-BLAST, followed by secondary structure assignment with PSIPRED (v4.1) and PSIPASS2 (v4.1). The resulting plot shows the position of the proximal and distal histidines and the predicted distribution of alpha-helices and coils along the protein sequence.

The analysis revealed the expected overall sequence architecture typical for vertebrate Ngb and Cygb. Conserved residues characteristic for each globin type, including the proximal (HisF8) and distal (HisE7) histidines, were identified in both lungfish sequences, supporting the correct annotation of the respective globin. The total protein length of *N. forsteri* Ngb is 161 AA, which is comparable to the Ngb length in other sarcopterygians. *N. forsteri* Cygb comprises 180 AA, consistent with the size range observed for vertebrate cytoglobins. Both sequences show the highest similarity to tetrapod homologs, indicating strong evolutionary conservation.

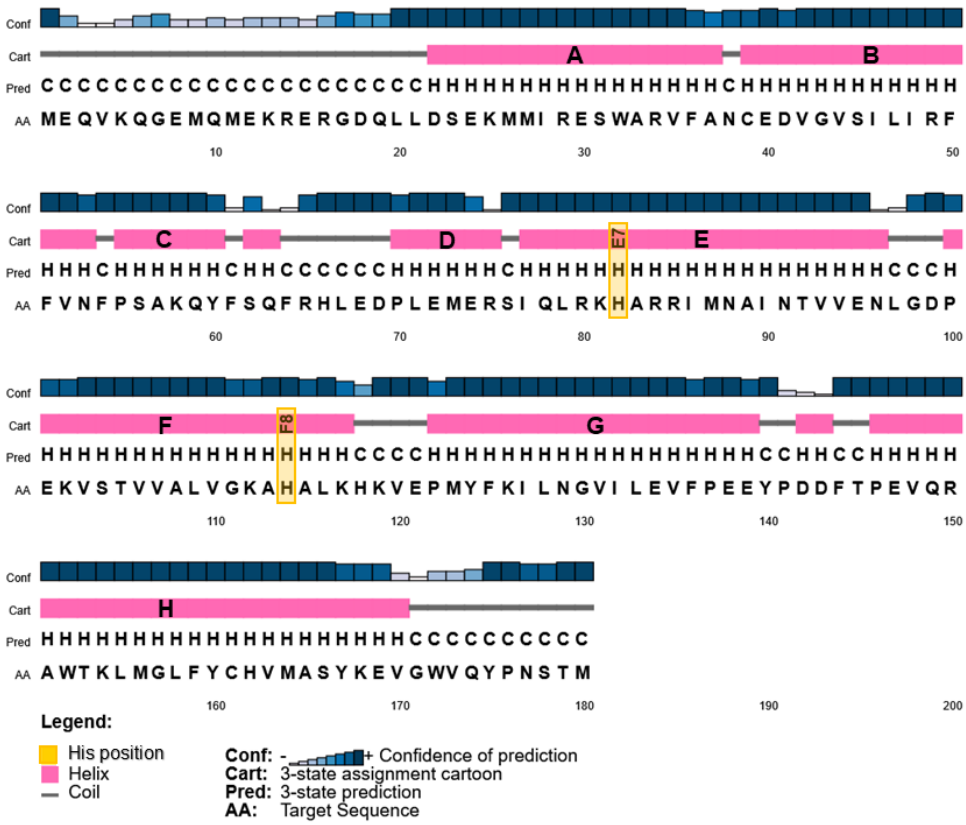


Figure 10: Secondary structure prediction of the *N. forsteri* Cygb protein using PSIPRED. The prediction was performed using sequence profile searches against the UniClust30 (2018_08) and UniRef90 databases with HHblits and PSI-BLAST, followed by secondary structure assignment with PSIPRED (v4.1) and PSIPASS2 (v4.1). The resulting plot shows the position of the proximal and distal histidines and the predicted distribution of alpha-helices and coils along the protein sequence.

3.3.3 Phylogenetic classification of lungfish Ngb and Cygb

To further validate the classification of the two lungfish globins, we performed Bayesian phylogenetic analyses on the aligned datasets. The resulting phylogenetic trees (Fig. 11) clearly placed *N. forsteri* Ngb within the established vertebrate neuroglobin clade. Similarly, *N. forsteri* Cygb clustered unambiguously within the cytoglobin lineage, forming a sister relationship with

tetrapod Cygb sequences. Both globins were separated from hemoglobin, myoglobin and other vertebrate globin lineages with high posterior probability values (PP=1). These results confirm the correct assignment of the *N. forsteri* sequences and suggest that Ngb and Cygb were present in the last common ancestor of lungfish and tetrapods.

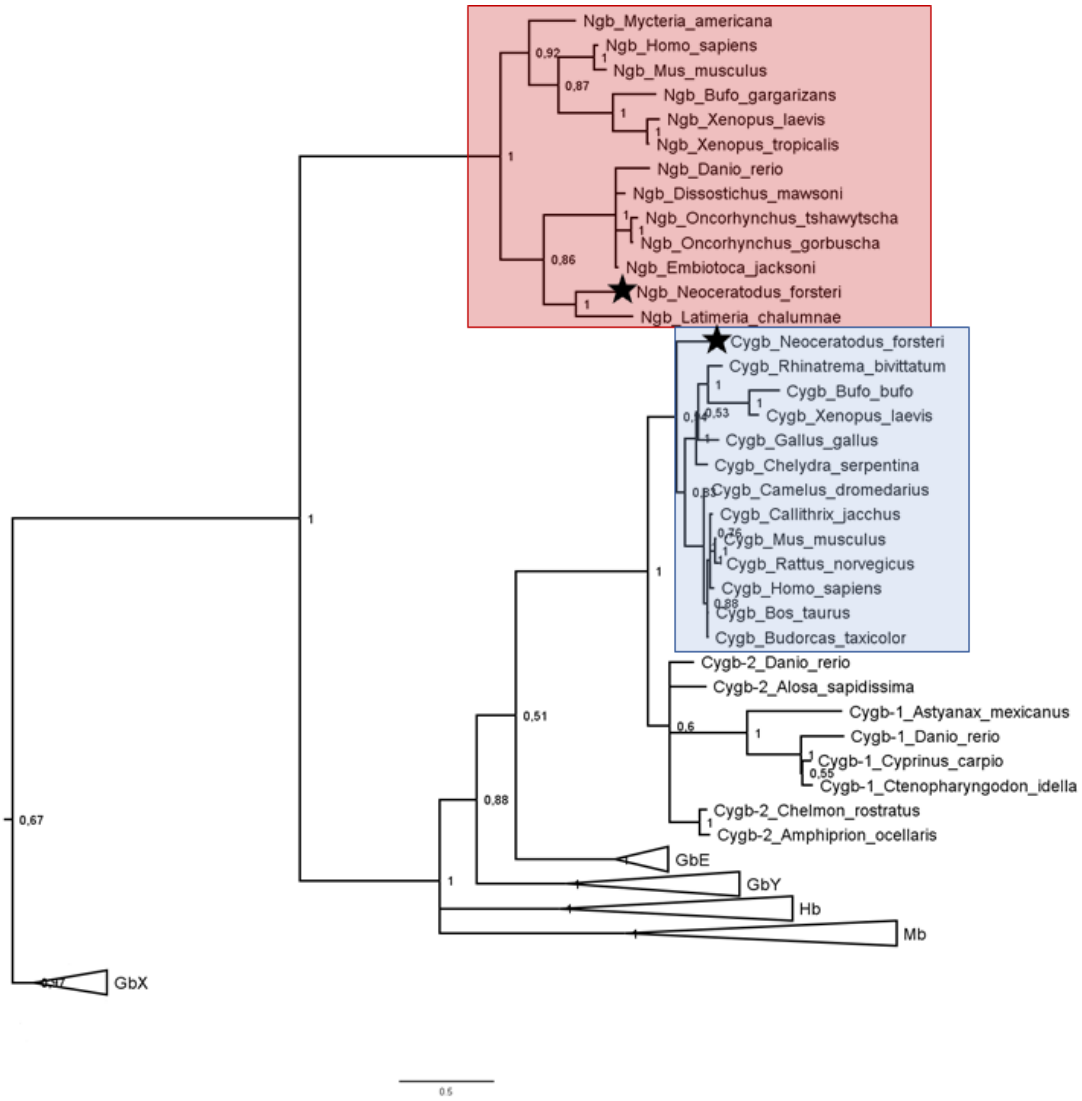


Figure 11: Bayesian phylogenetic tree of representative vertebrate globin gene family members and the *N. forsteri* globin genes (NfoG). Posterior probability support values are displayed on every reconstructed node. Ngb sequences are highlighted in red and Cygb in blue. *N. forsteri* sequences are marked with a star. A total of 50 sequences were analyzed with 226 aligned AA sites (supplementary material) using the Jones-Taylor-Thornton matrix with gamma distribution.

3.3.4 Structural prediction and comparison with human globins

Structural predictions generated using AlphaFold produced high-confidence 3D models of both lungfish globins (Fig. 12). Overall, the predicted three-dimensional structures of *N.forsteri* Ngb and Cygb both exhibited the characteristic globin fold, consisting predominantly of α-helices

surrounding a central heme-binding pocket. Structural overlays with the corresponding human reference proteins revealed a high degree of structural similarity, particularly in the heme-binding core region.

Aligning *N. forsteri* Ngb with the human Ngb structure (PDB: 1OJ6) revealed strong similarities in the arrangement of the globin helices and the heme-binding cavity (Fig. 12A). The proximal histidine residue (His100) and the distal histidine residue (His68) were positioned in an arrangement consistent with hexacoordinated heme binding, a characteristic of vertebrate Ngb.

Similarly, *N. forsteri* Cygb colocalized with the human Cygb structure (PDB: 2DC3); however, it has a C-terminal extension (Figure 12B). The proximal histidine residue (His114) and the distal histidine residue (His82) were positioned in an arrangement consistent with hexacoordinated heme binding, which is characteristic for vertebrate Cygb.

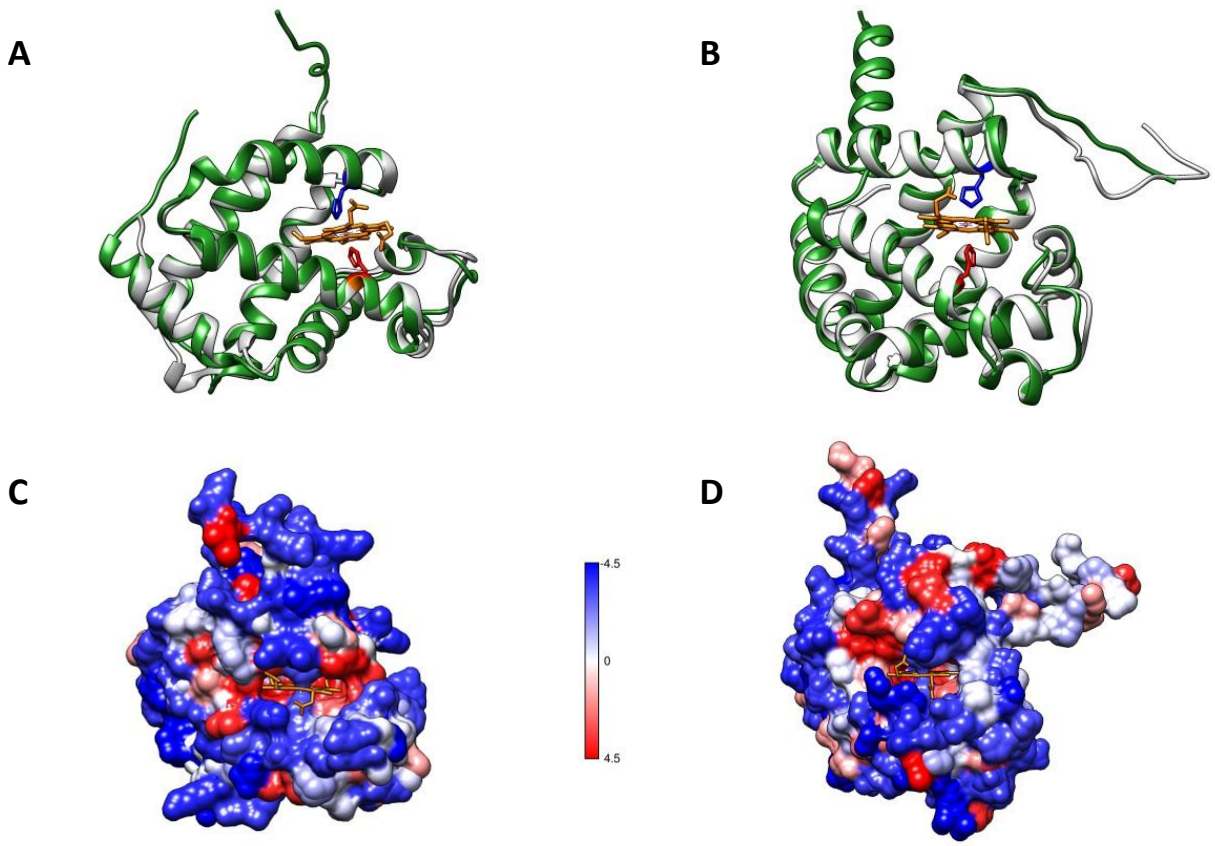


Figure 12: Structural comparison and surface hydrophobicity analysis of *N. forsteri* Ngb and Cygb. (A) Structural overlay of the predicted *N. forsteri* Ngb (green) and the human Ngb (grey, PDB: 1OJ6). The heme group is shown as orange sticks; the proximal histidine (His100) is shown as blue sticks; and the distal histidine (His68) is shown as red sticks. (B) Structural overlay of the predicted *N. forsteri Cygb* (green) and the human Cygb (grey; PDB: 2DC3). The heme group is shown as orange sticks; the proximal histidine (His114) is shown as blue sticks; and the distal histidine (His82) is shown as red sticks. (C) Surface hydrophobicity of *N. forsteri* Ngb. Blue indicates hydrophilic regions and red indicates hydrophobic regions. The heme group is shown as orange sticks. (D) Surface hydrophobicity of *N. forsteri* Cygb. Blue indicates hydrophilic regions and red indicates hydrophobic regions. The heme group is shown as orange sticks. The images were generated using Chimera (v1.19).

Analysis of surface hydrophobicity revealed that both Ngb and Cygb possess strongly hydrophobic heme-binding pockets (Figure 12C, D). The region surrounding the heme group in both proteins was predominantly composed of hydrophobic surface areas, indicating a well-shielded ligand-binding environment. This hydrophobic character is consistent with the stabilization of the heme prosthetic group within the globin fold. The largely hydrophobic environment around the heme group may facilitate efficient ligand binding and contribute to the regulation of oxygen and nitric oxide interactions. The partially buried positioning of the heme group, together with the conserved proximal and distal histidines, supports the typical functional properties of vertebrate globins and indicates that the ligand-binding mechanisms are conserved in *N. forsteri* Ngb and Cygb.

3.3.5 Spectroscopic analysis of coordination states

UV-visible (UV-Vis) absorption spectra were recorded for the purified *N. forsteri* Ngb and Cygb proteins in both oxygenated (oxy) and deoxygenated (deoxy) states (Fig. 13, 14).

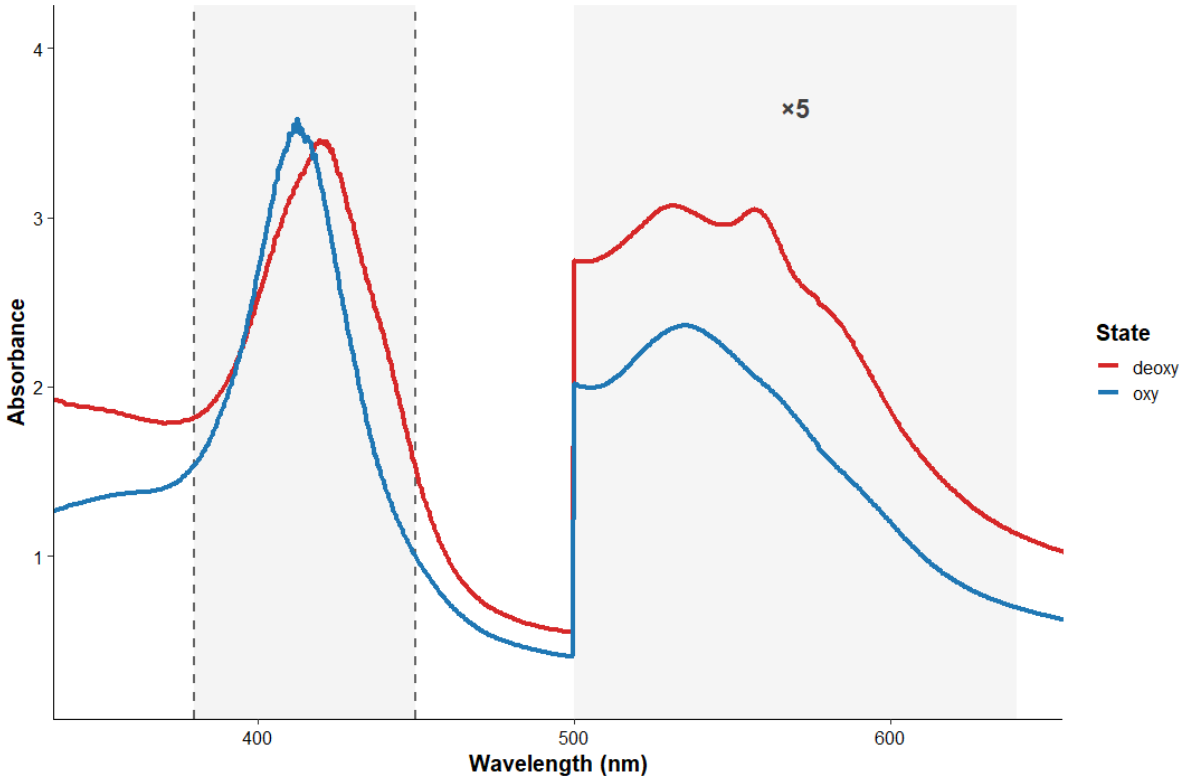


Figure 13: UV-Vis absorption spectra of *N. forsteri* Ngb in different (oxy and deoxy) coordination states. The Soret region (380-450 nm) is highlighted in gray and bounded by dashed vertical lines. Spectra were recorded over the full wavelength range of the photometer. The visible spectral region above 500 nm is shown with a 5-fold expanded y-scale for improved visibility. Figures was generated in R using ggplot2.

Both proteins exhibited a pronounced Soret band in the range of approximately 410-440 nm, along with weaker absorbance features in the visible region above 500 nm. The deoxygenated samples showed higher absorbance in the visible range than the oxygenated samples in both cases.

For Ngb, the Soret maximum was observed at approximately 412 nm in the oxy state and 420 nm in the deoxy state. The visible spectrum showed clear differences between the two states, with the deoxygenated form exhibiting increased absorbance across the Q-band region (Fig. 13). The overall spectral profile was consistent with that of a hexacoordinated heme protein.

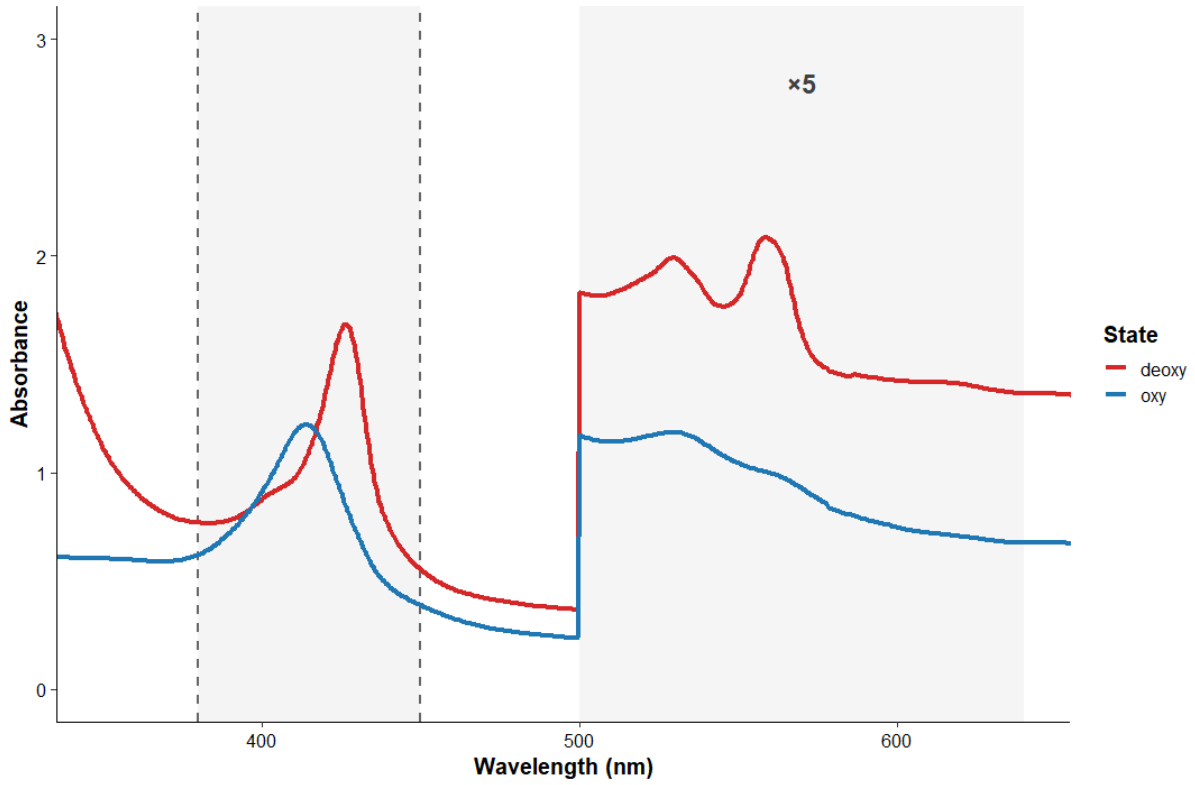


Figure 14: UV-Vis absorption spectra of *N. forsteri* Cygb in different (oxy and deoxy) coordination states. The Soret region (380-450 nm) is highlighted in gray and bounded by dashed vertical lines. Spectra were recorded over the full wavelength range of the photometer. For improved visualization, the visible region (>500 nm) is displayed with a 5-fold expanded y-scale. Figures was generated in R using ggplot2.

Cygb showed a similar spectral pattern, with a Soret maximum at approximately 416 nm in the oxy state and 426 nm in the deoxy state. As for Ngb, the deoxygenated form exhibited stronger absorbance in the visible region than the oxygenated form (Fig. 14). These features align with the expected spectral behavior of a hexacoordinated globin absorbance in the visible region than the oxygenated form (Fig. 14).

3.4 Discussion

3.4.1 Detection of Ngb and Cygb transcripts in *N. forsteri*

The present study provides evidence that *N. forsteri* possesses both Ngb and Cygb, and that both genes are transcribed in a tissue-specific manner. Ngb was amplified from brain cDNA, consistent with its known enrichment in neural tissue, whereas Cygb was amplified from heart cDNA, in line with its comparatively broad tissue distribution and particularly high expression in the heart (Keppner et al., 2020). Because only conventional PCR was used, these data demonstrate transcript presence rather than quantitative expression levels; nevertheless, the observed tissue specificity is consistent with the established expression patterns of these globins in other vertebrates (Burmester & Hankeln, 2014; Hankeln et al., 2004; Hankeln, Ebner, Fuchs, Gerlach, Haberkamp, Laufs, Roesner, Schmidt, Weich, Wystub, et al., 2005; Keppner et al., 2020).

3.4.2 Evolutionary significance of lungfish Ngb and Cygb

From an evolutionary perspective, the detection of Ngb and Cygb in lungfish is highly significant because lungfish occupy a key phylogenetic position as the closest living relatives of tetrapods (Meyer et al., 2021). The placement of the *N. forsteri* sequences within the vertebrate Ngb and Cygb clades with strong posterior support indicates that these proteins are authentic homologs of the canonical vertebrate tissue globins. The phylogenetic reconstruction consistently grouped the lungfish sequences with the expected globin lineages, and the decisive nodes separating Ngb from Cygb, as well as these clades from other vertebrate globins, were supported by posterior probabilities of 1 or at least above 0.8. This robust topological support strengthens the conclusion that Ngb and Cygb were already present in the lineage leading to lungfish and tetrapods (Hankeln, Ebner, Fuchs, Gerlach, Haberkamp, Laufs, Roesner, Schmidt, Weich, Wystub, et al., 2005; Keppner et al., 2020; Wystub et al., 2004).

3.4.3 Sequence and structural conservation

The results further suggest that the structural and functional characteristics of these globins have been strongly conserved over evolutionary time. Sequence analysis confirmed the presence of the proximal and distal histidines that are characteristic for hexacoordinated

vertebrate globins. These residues are essential for heme coordination and are a hallmark of the Ngb and Cygb families (Burmester & Hankeln, 2014; Pesce et al., 2004; Sawai et al., 2003). In addition, the predicted protein lengths of both lungfish globins fall within the expected range for vertebrate homologues, supporting their annotation as canonical Ngb and Cygb orthologues (Hankeln, Ebner, Fuchs, Gerlach, Haberkamp, Laufs, Roesner, Schmidt, Weich, Wystub, et al., 2005). The conservation of these features suggests that the globin fold and heme-binding chemistry of Ngb and Cygb have remained under strong evolutionary constraint.

3.4.4 Structural conservation of Ngb and Cygb

The AlphaFold models confirmed the sequence-based interpretation by showing that both lungfish proteins adopt the typical globin fold with a largely alpha-helical architecture surrounding the heme pocket. Structural superposition with the human reference proteins revealed a high degree of similarity, particularly in the core heme-binding region. The predicted arrangement of the conserved histidines is consistent with hexacoordination and therefore with the functional properties expected for vertebrate Ngb and Cygb (Hankeln, Ebner, Fuchs, Gerlach, Haberkamp, Laufs, Roesner, Schmidt, Weich, Wystub, et al., 2005). Together, the sequence and structural data indicate that the lungfish proteins retain the canonical structural framework of tissue globins.

3.4.5 Spectroscopic support for hexacoordination

The spectroscopic results provide additional support for this conclusion. Both proteins showed Soret maxima in the range of approximately 410 to 440 nm, consistent with hexacoordinated heme proteins and with spectra typically reported for vertebrate Ngb and Cygb (Hankeln, 2005). Although the spectra do not by themselves prove the exact coordination geometry, they are compatible with internal histidine ligation in the deoxy state and with ligand-dependent spectral transitions characteristic of hexacoordinated globins (Craik et al., 1980; Uppal et al., 2016). The observed Q-band features in the visible region further support this interpretation, as these weak absorption bands (typically 500-600 nm) are sensitive to the coordination state of the heme iron (Turilli-Ghisolfi et al., 2023). Together with the structural predictions, these data support the assignment of the Ngb and Cygb lungfish proteins as hexacoordinated globins.

3.4.6 Evolutionary implications and functional conservation

Taken together, the findings of this study extend the known occurence of Ngb and Cygb into *N. forsteri* and support the view that both globin lineages were already established before the divergence of lungfish and tetrapods (Burmester & Hankeln, 2014; Meyer et al., 2021). This interpretation is consistent with the broader evolutionary scenario in which vertebrate tissue globins diversified early to fulfil specialized intracellular roles in oxygen handling, redox balance, and stress protection (Burmester & Hankeln, 2014; Hankeln, Ebner, Fuchs, Gerlach, Haberkamp, Laufs, Roesner, Schmidt, Weich, Wystub, et al., 2005). The presence of Ngb in brain and Cygb in heart tissue further suggests that the tissue-specific deployment of these globins was already present in sarcopterygian vertebrates (Gallagher & Macqueen, 2017; Hundahl et al., 2013; Keppner et al., 2020; Mammen et al., 2006). Accordingly, the present results do not support a scenario of complete functional replacement of Ngb and Cygb by other globins in lungfish. Rather, they indicate that both globin lineages were retained in *N. forsteri* and may preserve conserved tissue-specific functions.

3.4.7 Limitations and future perspectives

A limitation of the present study is that the expression analyses were based on conventional PCR and therefore only demonstrate presence or absence of transcripts, not relative expression levels. In addition, the structural models are computational predictions, and the spectroscopic data, while informative, do not alone establish detailed kinetic behavior or ligand-binding mechanisms. Future studies should therefore combine quantitative expression analyses like qRT-PCR, biochemical characterization, and ideally comparative work across additional lungfish species to further clarify the functional evolution of Ngb and Cygb.

3.5 Material and methods

3.5.1 Sample material and cDNA generation

Tissue samples of *N. forsteri* were obtained from material stored at -80 °C in RNAlater from previous projects. Based on their expected tissue-specific expression patterns and in reference to the expression sites shown in Keppner et al. 2020, Fig. 3, brain tissue was used to amplify Ngb and heart tissue to amplify Cygb. Total RNA was isolated using the Crystal RNA Mini Kit

(Biolabs) according to the manufacturer's instructions, with an additional DNase digestion step due to the high level of DNA present in the lungfish samples. The integrity and concentration of the RNA were assessed using a TapeStation system, yielding RNA Integrity Number (RIN) values between 6.6 and 7.6. First-strand cDNA was synthesised from the extracted RNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific).

3.5.2 PCR amplification and sequence verification

The Ngb and Cygb coding sequences were amplified by conventional PCR from tissue-derived cDNA using gene-specific primers (supplementary material). The PCR reactions were performed in a total volume of 50 µl, containing Taq DNA polymerase, 10× PCR buffer, 25 mM $MgCl_2$, dNTPs, primers and 1,000 ng/µl of the template cDNA. The cycling conditions consisted of an initial denaturation step at 95 °C/4 min, followed by 35 cycles of denaturation at 95 °C, annealing at 59 °C and 10 min extension at 72 °C. This was followed by a final extension step at 72 °C for 11 minutes.

The PCR products were analyzed by agarose gel electrophoresis and purified enzymatically prior to sequencing. For purification, 2.25 µl of Exonuclease I and 9 µl of Shrimps Alkaline Phosphatase (SAP) (FastAP) were mixed with 48.75 µl of high-performance liquid chromatography (HPLC)-grade water. Then, 25 µl of the PCR product was combined with the ExoI/SAP mixture (6 µl) and incubated in a thermocycler for 25 min at 37 °C, followed by 15 min at 72 °C. The purified PCR products were sent for Sanger sequencing to Eurofins Genomics (formerly GATC Biotech). Sequence chromatograms were inspected manually and the sequences were assembled using SnapGene 8.2.2.

3.5.3 Sequence analysis and alignments

Translated Ngb and Cygb AA sequences were aligned with representative vertebrate globin sequences retrieved from NCBI and UniProt. Multiple sequence alignments were generated using MUSCLE v5.3 (Edgar, 2022) with default parameters. Conserved globin residues and structure elements were examined based on the alignment and visualized using SnapGene (v8.2.2).

The secondary structure of the Ngb and Cygb protein sequences of *N. forsteri* was predicted using PSIPRED (Buchan et al., 2024) which carries out sequence profile searches using HHblits and PSI-BLAST against the UniClust30 (2018_08) and UniRef90 databases to generate the input profile for predicting secondary structures. The resulting predicted secondary structure

elements were used to support the interpretation of conserved globin features, complement sequence alignment analyses and assess conserved His positions.

3.5.4 Phylogenetic analysis

The multiple sequence alignments were converted to Nexus format using seqmagick (v0.8.6). Phylogenetic analyses were performed with MrBayes (v3.2.7a) (Ronquist et al., 2012) using respective protein sequences. The amino acid substitution model was specified with a Jones model (`aamodelpr = fixed(jones)`), and rates across sites were modelled as gamma-distributed (`rates = gamma`). Bayesian inference was run for 5,000,000 generations with four Markov chains per run (`nchains = 4`), sampling every 500 generations (`samplefreq = 500`). Convergence diagnostics were computed during the run (`diagnfreq = 5000`). The burn-in phase was removed prior to summarizing the posterior distribution of trees (first 25% of sampled trees). Convergence was assessed using standard diagnostics from the MrBayes output (e.g., average standard deviation of split frequencies and effective sample size). Posterior probabilities were mapped onto the consensus tree.

3.5.5 Structural modelling and comparison

The three-dimensional structures of *N. forsteri* Ngb and Cygb were predicted using AlphaFold v3. These models were then structurally compared with those of human Ngb (PDB: 1OJ6) and human Cygb (PDB: 2DC3). Structural alignments and visualizations were performed in Chimera v1.19 (Pettersen et al., 2004). Heme cofactors were modeled based on corresponding reference structures and surface hydrophobicity was calculated to characterize the proteins' physicochemical properties.

3.5.6 Recombinant expression and purification of Ngb and Cygb

The coding sequences of *N. forsteri* Ngb and Cygb were synthesized by BioCat GmbH and provided in pET-15b expression vectors that encode an N-terminal 6×His tag. The recombinant plasmids were transformed into *Escherichia coli* (Jm109) expression cells. Overnight starter cultures were then grown at 37 °C in LB medium supplemented with ampicillin and chloramphenicol.

To express the recombinant protein, one liter of LB medium was inoculated with two milliliters of an overnight culture and supplemented with ampicillin, chloramphenicol, and 1-aminolevulinic acid (ALA), which serves as a heme precursor. The cultures were then

incubated at 37 °C until an OD_{600} of 0.4-0.8 was reached. Protein expression was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) to reach a final concentration of 0.4 mM. Following induction, the cultures were incubated overnight at 27 °C while shaking.

The cells were harvested by centrifugation at 4 °C, then resuspended in a buffer containing a protease inhibitor, 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 1 mM Pefabloc, 1 mM $MgCl_2$, DNase, and RNase. Cell disruption was performed with three freeze-thaw cycles in liquid nitrogen and a 37 °C water bath, followed by sonication on ice. The insoluble material was then removed by centrifugation at 4 °C. The clarified supernatant was then subjected to immobilized metal affinity chromatography using His60-Ni columns. After equilibration and washing, the bound proteins were eluted in three fractions with an imidazole-containing buffer. The eluted fractions containing Ngb or Cygb were pooled and stored at 4 °C or -80 °C until further use.

The Ngb coding sequence was provided in a pET-15b expression vector encoding an N-terminal 6×His-tag followed by a thrombin cleavage site:

catcatcatcatcatcacagcagcggcctggtgccgcgcggcagccatatggaatacccgagcctgtctatgagagaaaaggagctg
atcagagagagctgggggcaactgggcaaggagaaacttcagcacggcactgtgttgttctccaggctgtttgaattggacccggatc
tggtgcatctcttccattacaacaccactttcgctacacctgccgagtgtctcacctccccggagttcctggaccatatccacaaggtgat
gagtgtgattgatgctgccgtgaacaacctagagaacctgccctccctggaggaatacctgataaacttgggcaagaagcacaagacc
gttggagtgaaaattgattcttttccggctgttggcgagtcgctgctctatatgttggagagagcacttggcactccgttgtgcactgagat
caaagctgcctggattgtgctctacaccttcgtgacaaaaaccatgatccgaggctgggacaagaatggggaaaacaagatagagtga

The Cygb coding sequence was provided in a pET-15b expression vector encoding an N-terminal 6×His-tag followed by a thrombin cleavage site:

catcatcatcatcatcacagcagcggcctggtgccgcgcggcagccatatggctctctcagatgctgaggtccagaatgcccgtggtg
cctgggcaaagctctatgccaatgcagaggacaatggagcaatcgtgctagccatgttcacagactaccctgacaccaagtcctacttc
acccatttcaaggggatggggtctgctgcagagatggaggggtcagcccaggttcggggccatgggaagaaggtctttgcttctctga
atgacatggtacagcacctgaacaacatggatgccttaaatgggattatcagcccgctaggaaagaaacacgctacgcagctcaaagt
ggaccccaaaaacttcagagtaatggaagagaaatttggtggggatgccagagcagcttttgataaagcgaccaatgtggtgtgtgcc
caaatgaacgccgcctacaaggaggctggttggtaaggatcc

3.5.7 UV-Vis spectroscopy

Purified Ngb and Cygb proteins were analyzed by UV-Vis spectroscopy using a VWR UV-3100PC spectrophotometer. Spectra were recorded from in the range 350-650 nm at room temperature in Tris-HCl buffer pH 8.5. The deoxy-form was obtained by adding a tip of a spatula of sodium dithionite. Protein concentrations were not measured but diluted until a light orange color was obtained. Spectra were recorded in 1 mm path length cuvette and baseline-corrected against Tris-HCl buffer. Plots were generated in RStudio (v2026.01.0+392) (Kronthaler & Zöllner, 2020).

3.6 Acknowledgements

I would like to thank my Master's student Lasse Roggenkamp and my colleague Dr. Cäcilia Plate for their valuable assistance with the laboratory work.

4 Chapter 3 - Functional *in vitro* studies of lungfish Neuroglobin and Cytoglobin

4.1 Abstract

The Australian lungfish *Neoceratodus forsteri (N. forsteri)* holds a significant evolutionary position among sarcopterygians, making it a valuable model for studying the functional properties of vertebrate globins. Neuroglobin (Ngb) and cytoglobin (Cygb), which have been associated with oxygen homeostasis, redox regulation and cytoprotection under stress conditions, have recently been identified in this species. However, their functional roles in lungfish remain poorly understood. This study investigated whether transient expression of *N. forsteri* Ngb and Cygb affects the viability of CHSE-214 fish cells under oxidative and thermal stress. Cells were transiently transfected with plasmids encoding lungfish Ngb, Cygb, or eGFP as a transfection control, with untransfected cells serving as baseline reference. Under oxidative stress, H_2O_2 reduced cell viability in all groups, with the strongest effect in Ngb-expressing cells. Under thermal stress, no significant effects of construct, temperature, or their interaction on viability were detected.

Taken together, these findings provide an initial functional characterization of *N. forsteri* Ngb and Cygb in a heterologous fish cell system. Although no clear protective effect was observed, the results suggest construct-dependent differences in stress response and indicate the need for further experiments with biological replication and refined stress conditions.

4.2 Introduction

In order to investigate the functional roles of the neuroglobin (Ngb) and cytoglobin (Cygb) described in previous chapters, the corresponding plasmids were transiently expressed in the CHSE-214 fish cell line (Lannan et al., 1984). The effects of Ngb and Cygb on cell viability were then assessed under two stress conditions: oxidative stress induced by hydrogen peroxide (H_2O_2), and an elevated temperature. Oxidative stress is a major cause for cellular damage, impairing cell survival by increasing levels of reactive oxygen species (ROS). H_2O_2 is a common experimental model for inducing oxidative stress and evaluating the ability of stress-

responsive proteins to support cell viability under damaging conditions (Ransy et al., 2020). Ngb and Cygb are of particular interest in this context because both globins have been associated with protection against oxidative damage in vertebrate systems (Thuy et al., 2026).
In addition to oxidative stress, temperature is an important environmental factor for ectothermic organisms, such as fish. Their physiological processes are strongly influenced by ambient temperature, with even moderate changes affecting protein stability, metabolism, redox balance and cell viability (Alfonso et al., 2021). Therefore, thermal stress experiments can provide insight into cellular stress tolerance and potential protective mechanisms mediated by globins.
Globins such as Ngb and Cygb have been associated with cytoprotective capabilities in response to hypoxic and oxidative stress in various vertebrate systems (Reuss, 2025; Thuy et al., 2026). Neuroglobin is predominantly expressed in neural tissues, whereas cytoglobin exhibits broader tissue distribution and has been associated with oxygen binding, redox regulation, and cellular stress responses. However, the precise functions of these proteins remain unclear (Reuss, 2025). In fish, globins may be particularly relevant given that aquatic ectotherms regularly encounter fluctuations in temperature and oxygen availability (Crawshaw, 1977).
To test whether lungfish Ngb and Cygb modulate cellular stress tolerance, CHSE-214 cells were transiently transfected with constructs encoding Ngb, Cygb, or enhanced green fluorescent protein (eGFP) as a transfection control. Untransfected cells served as the baseline reference.
CHSE-214 cells were chosen as a suitable model because they are a well-established and robust fish cell line that can be maintained under highly controlled conditions and transiently transfected with high reproducibility (Lannan et al., 1984). As such, they provide a useful system to assess whether heterologous expression of lungfish Ngb or Cygb affects cellular tolerance to defined stressors, including oxidative stress and elevated temperature. Cell viability was then quantified following exposure to H_2O_2 or an elevated temperature. For the temperature stress experiment, the cells were incubated at either 20 °C or 23 °C for 24 hours, and for the oxidative stress experiment, the cells were treated with H_2O_2 and subsequently analyzed for survival using a fluidlab R300 viability assay.

The aim of this study was to determine whether the expression of the newly identified lungfish Ngb or Cygb influences cell survival under oxidative or heat stress and whether these effects differ from those observed in control transfections.

4.3 Results

To monitor cell density and transfection efficiency during the stress experiments, CHSE-214 cells were observed at different time points after transfection. Figure 15 provides an overview of the culture setup, including the optimal confluence for transfection, the starting point for stress induction, and representative examples of eGFP expression used as a transfection control. These images illustrate the progression of the cultures during the experimental workflow and confirm that the cells reached the desired confluence before stress treatment.

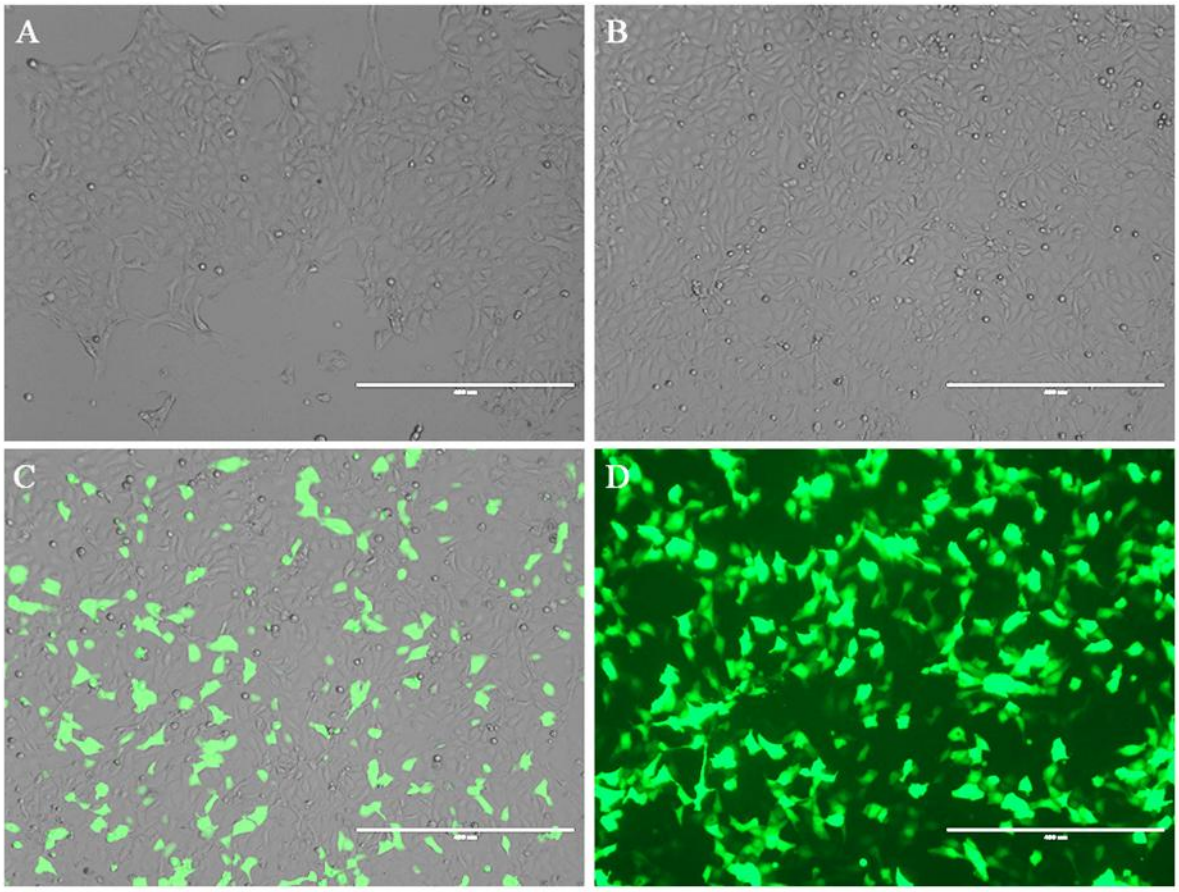


Figure 15: Overview of the CHSE-214 culture setup for stress experiments. (A) Optimal condition for transfection at 70% confluence. (B) Starting point for stress induction and beginning of the experiments at >90% confluence. (C) eGFP control 48h post-transfection with >90% confluence and transfection efficiency >35%. (D) eGFP control 72 h post-transfection as a transfection efficiency reference with >50% transfected cells.

4.3.1 H_2O_2 stress experiment

Cell viability was assessed following exposure to oxidative stress using a fluidlab R300 cell counter. H_2O_2 treatment reduced viability in all conditions compared with the corresponding untreated controls. Ngb-expressing cells exhibited a significant decrease in viability under

stress, with viability values of around 51% compared to approximately 80% in the corresponding untreated condition. A similar trend was observed in Cygb-expressing cells, with viability decreasing from approximately 97% in the untreated condition to around 66% after H_2O_2 treatment. A more moderate decrease in viability was exhibited by eGFP-expressing cells, with viability decreasing from approximately 87% in the untreated condition to around 74% under stress. Viability was also reduced in untransfected cells, decreasing from approximately 85% in the untreated condition to around 70% after oxidative stress (Fig. 16).

Statistical analysis using a binomial generalized linear model revealed a significant overall effect of the construct on cell viability (likelihood ratio test, $p < 2.2 \times 10^{-16}$; (supplementary material). Pairwise comparisons confirmed that H_2O_2 treatment significantly reduced viability within each construct relative to the corresponding untreated control. Additionally, under oxidative stress, cells expressing Ngb, Cygb and eGFP differed significantly from untransfected cells. Of the tested constructs, Ngb exhibited the most significant reduction in viability following H_2O_2 treatment, while Cygb and eGFP demonstrated comparatively higher viability.

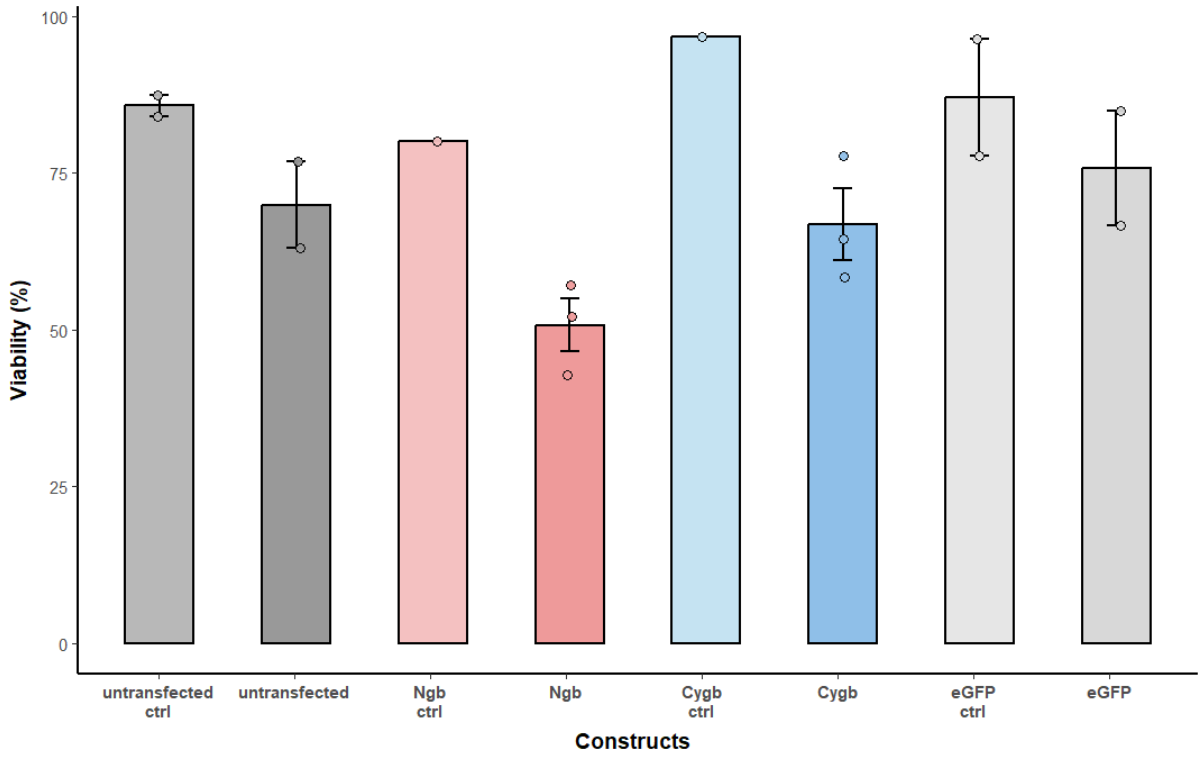


Figure 16: Cell viability after oxidative stress in different constructs. Cell viability was assessed using a fluidlab R300 cell counter after treatment with H_2O_2, or under the corresponding untreated conditions. Untransfected cells were used as the baseline reference. The bars represent the mean viability of the technical replicates and the error bars indicate the standard error of the mean (SEM). Individual data points represent single measurements. The conditions shown are the untransfected control, Ngb, Cygb and eGFP, with the untreated and H_2O_2-treated samples displayed in pairs. The plot was generated in R using the ggplot2 package.

4.3.2 Temperature stress experiment

The survival of CHSE-214 cells transfected with Ngb, Cygb, eGFP, or the untransfected control was assessed 24 h after incubation at 20 °C and 23 °C using the fluidlab R300 viability assay. Based on the analyzed technical replicates, Cygb and Ngb exhibited slightly higher mean survival at 23 °C than at 20 °C; however, eGFP and the untransfected control did not demonstrate a consistent temperature-dependent pattern (Fig. 17). However, two-way ANOVA did not detect a significant effect of construct ($F(3, 12) = 0.304$, $p = 0.822$), temperature ($F(1, 12) = 0.087$, $p = 0.773$) or interaction between the two ($F(3, 12) = 1.924$, $p = 0.180$).

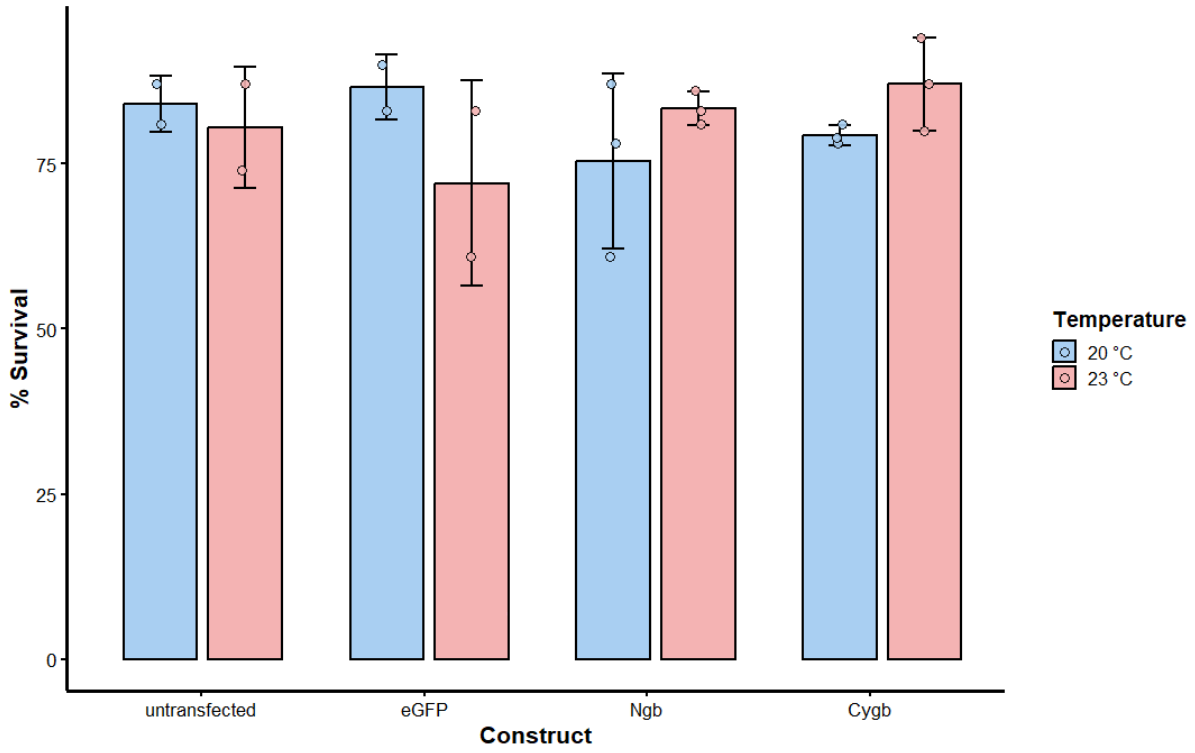


Figure 17: Survival of CHSE-214 cells transfected with Ngb, Cygb, eGFP, or untransfected control after 24 h incubation at 20 °C or 23 °C. Cell viability was measured using the fluidlab R300 viability assay. Bars indicate mean % survival, error bars show SD, and individual data points represent technical replicates. The 20 °C condition is shown in blue and the 23 °C condition in red. Statistical analysis by two-way ANOVA did not detect significant effects of construct, temperature, or their interaction. Plot was generated in R using ggplot2.

As these data are based on technical rather than biological replicates, the results should be interpreted as exploratory.

4.4 Discussion

A transfection efficiency of 30-35% in the eGFP control was sufficient for the present experiments and allowed a comparison of construct-dependent effects on cell viability. However, this value was determined using a fluorescent reporter construct and may not fully reflect the expression efficiency of the globin plasmids. Because the globin constructs are smaller than the eGFP control construct, they may have been taken up and expressed with similar or slightly higher efficiency, although this cannot be inferred directly from the present dataset (Pezzoli et al., 2017). Overall, the experimental design was suitable for an initial assessment of whether lungfish Ngb and Cygb influence stress tolerance in CHSE-214 cells, but additional experiments with independent biological replicates will be necessary to confirm these findings.

4.4.1 H_2O_2 stress

This study examined the impact of the expression of lungfish Ngb, Cygb on the viability of CHSE-214 cells subjected to H_2O_2-induced oxidative stress. Overall, H_2O_2 treatment caused a significant decrease in viability for all constructs, confirming the effectiveness of the treatment in inducing cellular stress and damage. This is consistent with hydrogen peroxide's well-established role as a common experimental inducer of oxidative stress, which leads to the accumulation of reactive oxygen species and impairment of cell survival (Ransy et al., 2020).
Among the constructs that were tested, the ones expressing Ngb showed the strongest reduction in viability after H_2O_2 exposure. This suggests that Ngb expression does not provide an obvious protective effect under these conditions. By contrast, Cygb- and eGFP-expressing cells exhibited higher viability than Ngb-expressing cells when subjected to oxidative stress. The relatively higher viability of Cygb-expressing cells compared to Ngb-expressing cells indicates that the two globins have different capabilities to detoxify oxidative stress induced via H_2O_2. However, these findings should be interpreted with caution, as the experiment was based on few technical replicates only and therefore cannot inform us about biological variability.
Further analysis of the control conditions showed that the different constructs were associated with distinct baseline viability values, even in the absence of H_2O_2. This suggests that transfection and/or construct expression may influence cell survival independently of oxidative stress. These effects are important to consider in stress response experiments because differences observed during H_2O_2 treatment may reflect the direct effects of oxidative stress and construct-specific baseline differences.

Using untransfected cells as a reference baseline was important for distinguishing construct-associated effects from general treatment effects. In the present dataset, all constructs differed significantly from untransfected cells, and likewise from sham-transfected cells (i.e., cells transfected with an empty vector), under both control and stress conditions. This indicates that both construct expression and treatment condition contribute to the observed viability patterns (Chong et al., 2021). eGFP, which is commonly used as a transfection control, also showed altered viability relative to untransfected cells. This highlights the fact that even control constructs can affect cell physiology (Konstantinov, 1996).

A limitation of the study is the lack of independent biological replication, meaning independent cell cultures prepared on different days / from different passages. Although the statistical differences were highly significant, the results should be interpreted primarily as evidence of assay sensitivity and construct-dependent differences under the specific experimental conditions used. Future experiments including biological replicates, additional stress intensities and time-course analyses would help to determine whether the observed patterns are robust and can be generalized to other situations (Lazic et al., 2018).

In summary, H_2O_2 treatment strongly reduced cell viability in all groups and the response of the tested constructs to oxidative stress differed. Ngb-expressing cells exhibited the lowest viability under stress, whereas Cygb and eGFP retained comparatively higher viability. These findings suggest that the constructs may modulate cellular stress responses differently, although additional experiments are required to confirm their biological significance.

4.4.2 Temperature stress

In this study, we investigated whether the transient expression of lungfish Ngb or Cygb affects the survival of CHSE-214 fish cells. We compared this with the standard cultivation temperature of 20 °C. This was under mild heat-stress conditions (Lannan et al., 1984). Although descriptive data indicated slightly higher viability in Ngb- and Cygb-transfected cells at 23 °C, particularly in the case of Cygb, statistical analysis revealed no significant effects of construct, temperature, or their interaction. Under the conditions tested here, therefore, no clear protective effect of Ngb or Cygb on cell survival could be demonstrated.

One possible explanation is that the temperature shift from 20 °C to 23 °C was too small to induce a measurable stress response in CHSE-214 cells (Lannan et al., 1984). Fish cells may tolerate such a modest increase in temperature without substantial loss of viability, which would

make it difficult to detect differences dependent on the construct (Katsuji & Lagerspetz, 1990). Additionally, transient transfection itself can introduce variability in cellular physiology and viability, potentially masking the subtle effects of Ngb or Cygb expression (Stepanenko & Heng, 2017).

4.4.3 Outlook and limitations

An important limitation of the present dataset is that only few technical rather than biological replicates were used. Although technical replicates are useful for evaluating measurement variability, they do not fully capture biological variation between independent experiments (Chan & Teo, 2020). Consequently, the statistical power of the analysis was limited and the findings should be regarded as exploratory. Future experiments should therefore include independent biological replicates and, ideally, a larger sample size, to determine whether Ngb or Cygb can confer a reproducible survival advantage under heat stress (Tarone et al., 1983). Also, other viability assays like WST-1 or CellTiter-Glo should be considered for the future.

The variability observed in some groups, particularly for eGFP and Ngb, may reflect differences in transfection efficiency, cell density, or assay performance. To quantify transfection efficiency, qRT-PCR could be applied prior stress application and ideally the creation of a stable transfection line should be considered.

In this context, eGFP is useful as a transfection control to assess successful plasmid uptake and to evaluate the effects of the transfection procedure. However, it cannot replace a mock-transfected control. Therefore, including an additional mock-transfected control group in future experiments would help to separate the effects of transfection from those of temperature stress (Konstantinov, 1996; McConnell et al., 2022).

Overall, the data suggest that Ngb and Cygb do not significantly affect survival under the specific temperature conditions tested, or that any potential effect is too subtle to be detected by the current experimental design. Further experiments using biological replicates, a broader temperature range and stronger stress conditions, if possible, will be necessary to evaluate whether these globins exert a protective function in fish cells.

In conclusion, these results provide the first insights into the stress-related functions of the lungfish proteins Ngb and Cygb *in vitro*. However, their biological significance remains to be established.

4.5 Material and methods

4.5.1 Culturing of CHSE-214 cells

CHSE-214 cells (obtained from AG Fischgenetik, Universität Rostock) were cultured in T75 flasks containing 10 ml of minimum essential medium (MEM) supplemented with 10% fetal bovine serum (FBS), 1% non-essential amino acids (NEAA) and 2 mM L-glutamine. The medium was changed every 48 hours or three times per week, after washing the cells with Phosphate Buffered Saline (PBS). Once the cells reached 90% confluence, they were passaged as follows: first, the cells were washed with 10 mL of PBS. Then, 0.05% trypsin/EDTA was added for 5 minutes to detach the cells and the reaction was stopped by the addition of a medium containing FBS. The cell suspension was then centrifuged at 500×g for 5 minutes and the pellet resuspended in 1-3 ml of fresh medium. Cell numbers were determined using a fluidlab R300 cell counter, after which the cells were seeded into new T75 flasks containing fresh medium. Cells were maintained at 20 °C in an atmosphere with 5% CO_2.

4.5.2 Plasmids and transfection

Plasmids encoding the lungfish proteins Ngb and Cygb were purchased from BioCat GmbH, Heidelberg. The AA sequences were synthesized into the pcDNA3.1(+) plasmid backbone (supplementary material for plasmid maps).

The following mixture was prepared for transfection: 25 µL of serum-free medium, 0.25 µg of plasmid DNA and 0.75 µL of X-tremeGENE™ HP DNA Transfection Reagent. This mixture was incubated at room temperature for 20 minutes before being added to cells in 24-well plates containing 1 ml of medium per well. After 4 h of incubation, 1 ml of double-concentrated medium was added to each well to restore the final culture conditions.

4.5.3 Stress experiments

Two stress experiments were performed: an oxidative stress experiment using H_2O_2 and a temperature stress experiment.

For the oxidative stress experiment, cells were seeded in 24-well plates at a density of approximately 19,000 cells per well. This resulted in 60-70% confluence after 24 hours.

After three hours, 0.5 ml of double-concentrated medium was added to support cell growth by increasing the nutrient concentration. Three technical replicates were used for each condition, along with an untransfected control.

Twenty-four hours after transfection, H_2O_2 was added to the respective wells at a final concentration of 400 µM. The cells were then incubated for an additional 24h under standard culture conditions. After treatment, the cells were examined and documented under the microscope (EVOS FL Digital Inverted Fluorescence Microscope). The cells were then trypsinised using trypsin/EDTA, mechanically detached and resuspended in a medium containing FBS to stop the reaction.

Cell viability was then determined using a fluidlab R300 cell counter in viability mode, which distinguishes between live and dead cells. The number of viable and non-viable cells was recorded for each sample.

4.5.4 Temperature stress experiment

For this experiment, cells were seeded in 24-well plates at a density of approximately 19,000 cells per well. After 24 hours, the cells had reached 60-70% confluence. Three technical replicates were used for Ngb and Cygb and two for eGFP and the untransfected control.

Twenty-four hours after transfection, the cells were incubated for 24 hours either at 20 °C (the control condition) or at 23 °C (the heat stress condition).

eGFP was used as a transfection control. After the incubation period, the cells were examined and documented under a microscope.

For harvesting, the cells were trypsinised using trypsin/EDTA for 6-7 minutes, mechanically detached and resuspended in a medium containing FBS. Cell viability was subsequently measured using a fluidlab R300 cell counter in viability mode.

4.5.5 Statistical analyses

Statistical analyses were performed in R. For the oxidative stress experiment, a binomial generalized linear model (GLM) with a logit link function was used. Live and dead cell counts were analyzed as binomial data based on the combined counts for each measurement. Pairwise comparisons between experimental groups were performed using estimated marginal means, and p-values were adjusted for multiple testing using the Holm method.

For the temperature stress experiment, a two-way analysis of variance (ANOVA) was used to assess the effects of construct and temperature, as well as their interaction. *Post hoc* comparisons were performed where appropriate. Data visualization was carried out in R using the ggplot2 package (v4.0.3). A p-value of < 0.05 was considered statistically significant.

4.6 Acknowledgements

I would like to thank my colleague Dr. Cäcilia Plate for our collaboration, my colleague Michelle Wisniewsky for helping out with the cells and AG Fischgenetik / Alexander Rebl (Universität Rostock) for kindly sharing their CHSE-214 cell line.

5 General Discussion

As illustrated in Figure 18, this thesis employs genomic, molecular, structural, and functional analyses to study the globin repertoire of the Australian lungfish, *Neoceratodus forsteri*. Against this background, the evolutionary position of lungfish makes them a particularly informative model for studying the origins of key vertebrate traits associated with the transition to life on land.

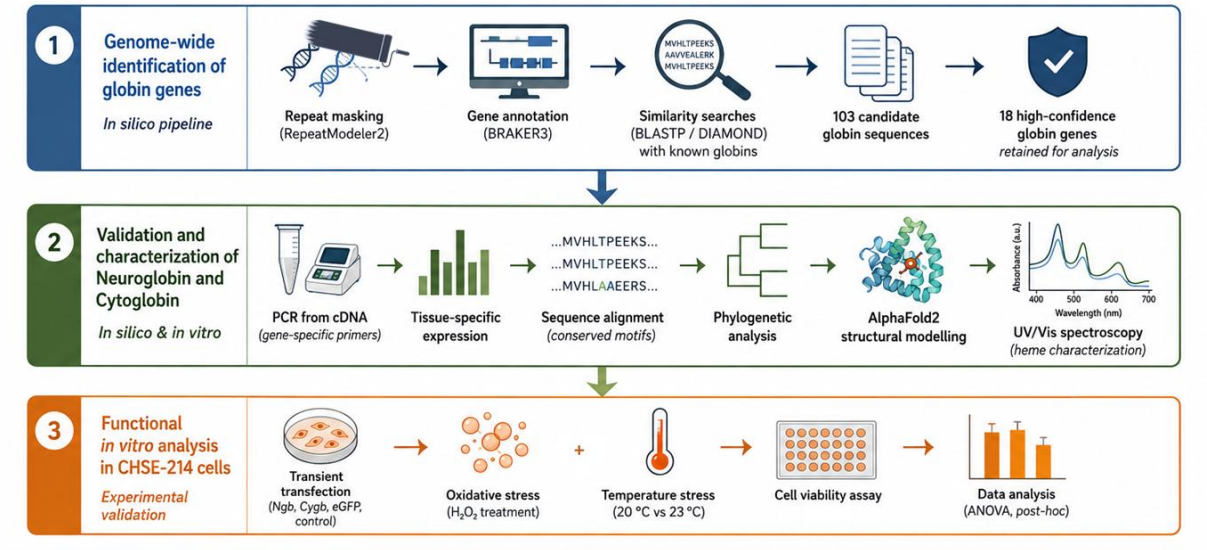


Figure 18: Integrated overview of the thesis and main findings. This schematic summarizes the three complementary levels of analyses employed in this study to investigate globins in Neoceratodus forsteri: genome-wide identification of globin genes, molecular and structural validation of selected candidates, and exploratory functional analysis in a heterologous cell system. Together, these approaches provide a multi-level framework for evaluating the composition, conservation, and potential functional significance of lungfish globins within an evolutionary context. Schematic illustration based on the author's conceptual design, generated with the assistance of OpenAI ChatGPT (GPT-5.5), and subsequently edited and finalized by the author.

5.1 Lungfish as a key model for vertebrate evolution

Lungfishes occupy a pivotal phylogenetic position as the closest living relatives of tetrapods and therefore provide important insights into early sarcopterygian evolution (Meyer et al., 2021). Their evolutionary significance is not only based on their position in the vertebrate tree,

but also on their combination of ancestral and derived traits, which makes them particularly informative for questions concerning the emergence of terrestrial vertebrates during the Devonian period (Irisarri & Meyer, 2016; Kemp et al., 2017). As one of the few extant lineages of lobe-finned fishes, lungfishes represent an especially valuable comparative system for investigating how key vertebrate traits may have been assembled, modified, or retained during the transition from aquatic to more terrestrially adapted life forms (Kemp et al., 2017). In this sense, they do not merely occupy a basal position in a phylogenetic framework, but also provide a living point of reference for interpreting the evolution of morphological and physiological traits that are otherwise accessible only through the fossil record (Brownstein et al., 2023). Their present-day diversity also reflects a much broader evolutionary history, since fossil evidence indicates that lungfish lineages were once far more diverse and that the extant species represent only the surviving remnants of a larger post-Devonian radiation (Kemp et al., 2017). Compared with African and South American lungfishes, *N. forsteri* retains more ancestral physiological traits, including a single lung, greater dependence on gill respiration, facultative air breathing, and the absence of aestivation (Jorgensen & Joss, 2016; Joss, 2006). In addition, it possesses fleshy, lobed paired fins, which are often interpreted as retaining a more ancestral sarcopterygian morphology (Brownstein et al., 2023). In this context, the Australian lungfish is especially interesting because it retains several features considered primitive among sarcopterygians while at the same time displaying physiological adaptations that may have supported survival in freshwater environments with fluctuating oxygen availability (Brownstein et al., 2023; Germon et al., 2012). These characteristics make *N. forsteri* particularly useful for addressing questions not only of evolutionary conservatism, but also of the functional constraints and ecological conditions that may have shaped the persistence of ancestral traits in a lineage that diverged early from the tetrapod stem.

Beyond their evolutionary position, lungfishes are also of considerable interest because of their exceptionally large genomes, which create substantial challenges for genomic and bioinformatic analyses. Genome size expansion and high repeat content can interfere with gene prediction, reduce annotation quality, and complicate the recovery of complete coding sequences (Carducci et al., 2021). In lungfishes, these genomic properties are not merely technical inconveniences, but fundamental features that shape the way genomic data can be generated, interpreted, and compared across species (Metcalfe et al., 2012; Schartl et al., 2024). This is particularly relevant because large repeat-rich genomes often increase the likelihood of fragmented assemblies, collapsed paralogous regions, and incomplete or ambiguous gene

models, all of which can complicate downstream analyses (Blommaert, 2020). As a result, lungfishes are not only biologically relevant, but also methodologically demanding study organisms. This makes them a valuable system for evaluating how modern computational approaches can be applied to non-model vertebrate genomes and for testing strategies to generate biologically meaningful annotations despite substantial technical obstacles. Moreover, the combination of unusual amount of transposable elements, of which many are still active and their phylogenetic position renders lungfishes an important test case for assessing how assembly and annotation biases might influence comparative genomic inferences about gene family evolution and the reconstruction of ancestral states (Metcalfe et al., 2012; Schartl et al., 2024). This is especially important for multigene families such as globins, where copy number variation, sequence divergence, and local genome complexity can all affect whether loci are detected, correctly annotated, and interpreted in an evolutionary context (Koch et al., 2016; Lüdemann, 2019; Lüdemann et al., 2019).

A further reason why lungfishes are relevant in this context is that they allow the comparison of genome organization and gene repertoire between deeply diverged vertebrate lineages while still retaining a phylogenetic position close to the origin of tetrapods (Damas et al., 2021; Sacerdot et al., 2018). Conducting different approaches for gene identification makes it possible to distinguish broadly conserved features from lineage-specific innovations and to assess whether observed differences reflect genuine evolutionary change or are partly the result of technical limitations in genome assembly and annotation. In studies of globin genes, such distinctions are particularly important because these genes are often embedded in complex genomic regions and may be affected by duplication, partial assembly, or low-confidence prediction (Carducci et al., 2021; Metcalfe et al., 2012). For this reason, the analysis of globin repertoires in *N. forsteri* offers not only insight into the evolutionary history of these genes, but also an opportunity to evaluate the extent to which current annotation strategies are able to recover biologically meaningful results from exceptionally challenging vertebrate genomes.

5.2 Bioinformatic identification and annotation of globin genes in *N. forsteri*

The identification of globin genes in *N. forsteri* demonstrates that robust annotation of a large and repetitive genome requires an integrated workflow combining multiple complementary approaches (Tørresen et al., 2019). In the present study, species-specific repeat masking,

evidence-guided gene prediction, homology-based searches, and transcriptomic support were combined to identify candidate globin loci (Gabriel et al., 2024; Smit et al., 2015b). This multi-layered workflow was necessary because no single approach was sufficient to resolve the complete globin repertoire with confidence. In particular, the complexity of the lungfish genome and the fragmented nature of some genomic regions meant that the annotation process had to rely on the convergence of several independent lines of evidence. In a genome of this size and complexity, apparent gene predictions cannot be interpreted in isolation, but must instead be evaluated in light of structural plausibility, sequence conservation, and supporting transcriptional evidence (Tørresen et al., 2019).

One significant finding of the analysis was the substantial discrepancy between the initial set of candidates and the final group of high-confidence globin genes (chapter 1; Fig. 6 + Tab. 1). This pattern highlights the strengths and limitations of similarity-based gene discovery in challenging genomes. Broad similarity searches are highly sensitive and can detect divergent or partial loci that might otherwise be missed (Mistry et al., 2021). However, these searches also retrieve truncated sequences, duplicated fragments, and potential annotation artifacts, particularly in repeat-rich regions where gene prediction is error-prone. In the case of lungfish, the abundance of repetitive elements and the fragmentation of some genomic regions likely contributed to the recovery of many partial or ambiguous hits (Lerat, 2010). Therefore, it was essential to filter the candidate sequences based on domain completeness, alignment quality, conserved motifs, and phylogenetic placement (Nishimura, 2000; Smit et al., 2015a). In practical terms, this step was necessary not only to reduce false positives, but also to avoid overestimating the size and complexity of the globin repertoire.

In addition, the use of multiple search strategies revealed that annotation-based approaches were more informative than direct searches on the unannotated genome like Genome BLAST searches (chapter 1; 2.3.2). This underlines the importance of reliable gene models for downstream analyses, particularly in non-model species whose genomes are difficult to assemble and annotate. When annotation information is incomplete or absent, direct similarity searches may overlook loci with disrupted gene structure, weak conservation, or atypical exon-intron organization, even if those loci are biologically relevant. By contrast, annotation-guided approaches can integrate structural and transcriptomic information in a way that improves the recovery of complete and interpretable gene models (Ayub et al., 2026).

Another important aspect of the analysis was the influence of genome architecture on annotation quality. In a genome as large and repeat-rich as *N. forsteri's*, overprediction is a major concern. Notably, previous studies have shown that many transposable elements are still active in the lungfish genome, which may contribute to its exceptional size and repetitive complexity (Fuselli et al., 2023; Germon et al., 2012). The large number of predicted gene models likely reflects a combination of true biological signals and technical inflation caused by short ORFs, repetitive motifs, and fragmented alignments being interpreted as separate genes. This is a known challenge in large genomes and underscores the importance of exercising caution when interpreting raw annotation output (Ayub et al., 2026).
The final set of 18 high-confidence globin sequences (Chapter 1; tab. 1) represents a more biologically plausible subset supported by sequence conservation, structural features, and phylogenetic evidence. Notably, this subset includes four GbE and six Mb candidates, consistent with previous findings suggesting an expansion of these globin lineages in lungfish (Koch et al., 2016; Lüdemann, 2019; Lüdemann et al., 2019). This pattern may reflect lineage-specific duplication events or the retention of ancestral globin copies in the lungfish genome. Both scenarios are plausible in the context of an evolutionarily dynamic, TE-rich genome (Lüdemann, 2019).

Overall, the bioinformatic workflow used in this study demonstrates that meaningful gene family reconstruction is possible even in exceptionally challenging vertebrate genomes, provided that multiple independent lines of evidence are combined. This approach is not only relevant for lungfish, but also broadly applicable to other non-model organisms with large, repetitive genomes. In that sense, the present work contributes both to globin evolution and to methodological development in comparative genomics.

5.3 Evolutionary diversification of the lungfish globin repertoire

The globin repertoire recovered from *N. forsteri* provides important insight into the evolution of oxygen-binding proteins in early sarcopterygians. The presence of representatives from the major vertebrate globin families suggests that lungfish possess a diverse and evolutionarily significant set of globins. The detection of multiple copies of myoglobin and globin E suggests lineage-specific expansion (Lüdemann, 2019). The presence of Ngb, Cygb, GbX, and Hb components supports the conservation of the core vertebrate globin toolkit (Keppner et al.,

2020). This combination of duplication and retention indicates that the evolution of lungfish globins has been influenced by ancient vertebrate genome dynamics like the three whole genome duplication events in vertebrates (3R) (Qi et al., 2024) as well as more recent lineage-specific events.

The expansion of Mbs in *N. forsteri* is especially noteworthy. Mb is typically associated with oxygen storage and diffusion in muscle tissue (Vanek & Kohli, 2019). Multiple copies may provide functional diversification under fluctuating oxygen conditions. In a facultative air-breathing species such as *N. forsteri*, where oxygen availability varies considerably depending on environmental conditions, tissue demand, and respiratory mode, retaining several myoglobin paralogs could be an adaptation to maintain an efficient intracellular oxygen supply (Koch et al., 2016). Similarly, the expression of GbE may indicate lineage-specific diversification related to oxygen management in tissues with high metabolic demands or specialized developmental contexts. Although the exact physiological roles of these copies are unclear, their presence suggests that globin diversification in lungfish is not merely a neutral consequence of genome expansion, but rather, it may also have functional significance. A previous study (Lüdemann et al., 2019) of the West African lungfish supports this interpretation. That study showed that multiple GbE gene copies are present, yet they are expressed only in the ovaries, unlike in other taxa where they are expressed in the eyes (Blank, Kiger, et al., 2011). Unfortunately, the tissue-specific expression of GbE was not investigated in greater detail in the Australian lungfish. This leaves open the question of whether a similar expression pattern occurs in this lineage.

The detection of Ngb, Cygb, and GbX suggests that lungfish possess ancient tissue globins, which are believed to play a role in intracellular oxygen homeostasis, redox regulation, and cellular protection (Burmester & Hankeln, 2014; Keppner et al., 2020). These globins are more closely linked to stress responses than to systemic oxygen transport (Schuster et al., 2024). Their presence in *N. forsteri* suggests that the fundamental cellular functions attributed to these proteins were established early in vertebrate evolution. This is particularly relevant because lungfish are phylogenetically close to the origin of tetrapods (Meyer et al., 2021). This implies that the ancestral sarcopterygian globin repertoire may have included proteins specialized for transport, storage, and intracellular protection.

The distribution of globin genes across multiple chromosomes, together with the presence of some loci on unplaced scaffolds, has evolutionary implications as well. This indicates that parts

of the repertoire are embedded in genomic regions that are not yet fully resolved in the current assembly. This is not surprising, given the enormous size and repetitive nature of the lungfish genome. However, this means that the full genomic organization of the globin family may be incomplete. Therefore, it is possible that additional globin-related sequences or paralogs remain undetected in unresolved regions. Future improvements in genome assembly may reveal additional globin genes.

From an evolutionary perspective, the lungfish globin repertoire can be best explained as a result of deep conservation and lineage-specific diversification. Maintaining several ancient globin lineages indicates a strong functional constraint on core oxygen-related processes. Meanwhile, the expansion of particular families suggests that gene duplication provided the raw material for functional specialization This interplay between retention and diversification is characteristic of many multigene families and illustrates how genome evolution can shape physiological potential over long evolutionary timescales. (Hoffmann, Opazo, & Storz, 2012; Schuster et al., 2024). In lungfish, this pattern may reflect the ecological demands of living in freshwater habitats with variable oxygen conditions, where both respiratory flexibility and cellular resilience are likely to be advantageous.

5.4 Functional characterization of lungfish Ngb and Cygb

5.4.1 Sequence- and structure-based evidence for conserved globin features

Sequence and structural analyses of lungfish Ngb and Cygb strongly support their classification as canonical vertebrate tissue globins. Both proteins retained the conserved histidine residues characteristic of hexacoordinated globins, and their predicted lengths were consistent with those of known homologs from other vertebrates (Burmester & Hankeln, 2014). These findings suggest that the core structural framework of these proteins has remained intact despite the unusual genomic background of *N. forsteri*. The conservation of these key residues is important because they directly coordinate heme and are essential for the biochemical behavior of Ngb and Cygb.

The AlphaFold models supported the sequence-based interpretation, showing that both proteins have the typical globin structure, which consists of mostly alpha helices surrounding the hydrophobic heme pocket (Iritani et al., 2023). The structural similarity to human reference proteins (Chapter 2; Fig. 12) was especially evident in the heme-binding region, where the

proximal and distal histidines occupied positions consistent with hexacoordination. This level of conservation is striking because, despite hundreds of millions of years of divergence, it suggests that the structural core of these tissue globins has remained highly constrained. In other words, the proteins recovered from *N. forsteri* do not appear to be degenerate globin-like sequences but rather genuine orthologs that preserve the canonical architecture expected of functional Ngb and Cygb.

The spectroscopic data further support this conclusion (chapter 2; Fig. 13, 14). Both proteins exhibited absorption patterns consistent with hexacoordinated globins. These patterns included Soret maxima in the typical range of 415 nm for Ngb and Cygb, as well as visible-region features that changed upon deoxygenation (Sawai et al., 2003). Although UV-Vis spectroscopy alone cannot definitively prove the precise coordination state, the observed spectral behavior is consistent with internal histidine ligation in the deoxy state. Together with the structural predictions, these results strongly suggest that the structural and biochemical properties (chapter 2; Fig. 16) of lungfish Ngb and Cygb are similar to those of vertebrate tissue globins.

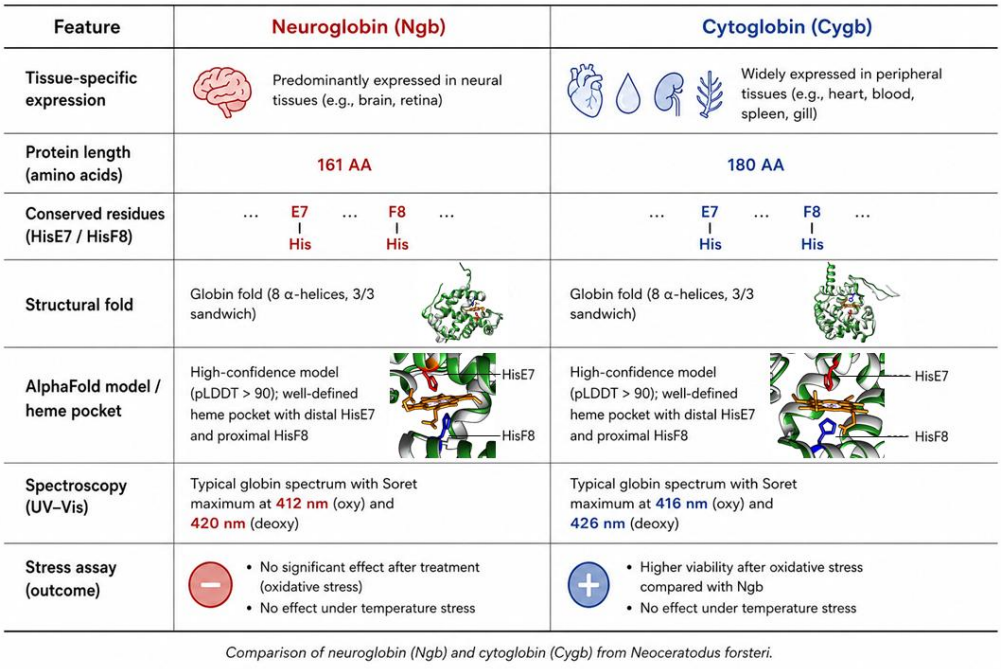


Figure 16: Comparative schematic overview of *N. forsteri* Ngb and Cygb. Schematic illustration summarizing key structural and functional properties of Ngb and Cygb, including protein length, spectral characteristics (Soret maxima in oxy- and deoxy-states), and relative responses to stress conditions. Ngb and Cygb differ in protein length (161 vs. 180 amino acids) and exhibit distinct Soret peak positions (Ngb: 412 nm oxy / 420 nm deoxy; Cygb: 416 nm oxy / 426 nm deoxy). Functional assays indicate no significant effect of treatment on Ngb viability under oxidative stress, whereas Cygb shows increased viability under oxidative stress conditions. No temperature-dependent effects were observed for either protein. The figure was adapted and modified by the author based on an AI-generated schematic (ChatGPT-5.5, OpenAI).

From a broader perspective, these results illustrate how multiple lines of evidence can be combined to strengthen functional annotation in difficult genomes. In large and repeat-rich assemblies, sequence similarity alone is often insufficient, since partial matches or fragmented loci may appear plausible at first glance but do not necessarily encode intact proteins [ref]. Structural prediction can therefore provide an additional layer of validation, especially when the inferred fold matches well-characterized vertebrate globins and supports a conserved protein architecture (Schneider, 2023). The inclusion of spectroscopic evidence adds yet another independent criterion, making it possible to distinguish between incomplete predictions and genuine globin genes with preserved biochemical properties (Sawai et al., 2003; Uno et al., 2004). In this respect, the lungfish Ngb and Cygb sequences are particularly informative because they exemplify the degree to which deeply conserved protein architecture can persist even in a lineage with highly unusual genome organization.

5.4.2 Tissue-specific detection and evolutionary significance

The detection of Ngb transcripts in brain tissue and Cygb transcripts in heart tissue aligns with the observed expression patterns of these globins in other vertebrates (Keppner et al., 2020). While the RT-PCR-based approach used here cannot yield quantitative results, it clearly shows that both genes are transcribed in a tissue-specific manner (supplementary material). This finding is biologically significant because it suggests that the tissue-specific expression of these globins was established early in the lineage leading to lungfish and tetrapods.

Ngb is generally associated with neural tissues, where it contributes to oxygen homeostasis and protects against hypoxic or oxidative stress (Exertier et al., 2022). Therefore, its detection in the brain cDNA of *N. forsteri* is consistent with its expected function. By contrast, Cygb is more broadly expressed, often showing particularly strong expression in connective and cardiovascular tissues (Reeder, 2023). The amplification of Cygb from heart cDNA suggests that its expression in *N. forsteri* follows a conserved vertebrate pattern. Although the current data do not reveal expression levels, the tissue specificity is an important finding because it indicates the conservation of functional expression domains.

Phylogenetic analyses confirmed that the lungfish sequences fall within the established Ngb and Cygb clades (chapter 2; Fig. 11). The strong support values that separate these lineages from each other and from other globin families provide a robust framework for orthology assignment (Seppey et al., 2019). This is essential because globin family evolution is complex and accurate classification requires phylogenetic and structural evidence. The placement of the

lungfish sequences within the expected vertebrate tissue globin lineages suggests that Ngb and Cygb existed prior to the divergence of lungfish and tetrapods. This finding improves our understanding of the early evolution of tissue globins and strengthens the idea that they became specialized for function before the transition from water to land.

Taken together, the expression, sequence, structural, and phylogenetic data suggest that the lungfish Ngb and Cygb genes are highly conserved orthologs that likely have retained tissue-specific functions. The results do not support the idea that these globins were lost or replaced entirely by other globin types in lungfish. Rather, they suggest that ancestral vertebrate tissue globins remained present and functional in this lineage.

5.4.3 Stress-related effects in CHSE-214 cells

The heterologous expression experiments provided an initial indication of how lungfish Ngb and Cygb may affect cellular stress responses, although the results (chapter 3; Fig. 16, 17) must be interpreted cautiously. Under oxidative stress induced by hydrogen peroxide, cell viability decreased across all constructs, confirming that the treatment was effective and that the assay was able to detect stress-induced damage (Bhattacharjee, 2019). Among the tested constructs, cells expressing Ngb showed the strongest reduction in viability after exposure to H_2O_2, whereas cells expressing Cygb or eGFP retained comparatively higher viability. At first glance, these results suggest that Ngb does not confer an obvious protective effect under these conditions and that Cygb may perform better in this assay. However, the small sample size and the lack of biological replicates means these results cannot be generalized with confidence. They should therefore be viewed as exploratory rather than definitive. Transient transfection experiments can introduce substantial variability in expression levels, cellular physiology, and apparent stress sensitivity, which can mask subtle functional effects.

The different constructs showed distinct baseline viability values in untreated conditions, which complicates interpretation. This suggests that construct expression itself or the transfection procedure may independently influence cell survival, apart from oxidative stress. This is an important consideration because the differences observed after H_2O_2 treatment may reflect not only the stress response, but also preexisting differences in cellular state. Therefore, including untransfected controls was useful for establishing a reference baseline. However, future experiments should also include mock-transfected controls to more precisely separate the effects of transfection reagents from plasmid expression (Chong et al., 2021).

The temperature stress experiment revealed no significant effects of construct, temperature, or their interaction. Descriptively, cells expressing Ngb and Cygb appeared to show slightly higher survival at 23 °C than at 20 °C, but this trend was not statistically significant. One possible explanation is that the temperature shift was too mild to induce a strong stress response in the cells. Fish cells may tolerate such a small increase in temperature without experiencing significant changes in viability, particularly over a relatively short incubation period (Lannan et al., 1984). Another possibility is that any protective effect of Ngb or Cygb was too subtle to detect under the tested conditions, particularly given the limited replication and variability associated with transient transfection.

It is important to note that the experimental design relied on technical rather than biological replicates. Although technical replicates are useful for evaluating consistency in measurements, they do not capture biological variation between independent experimental runs involving different cell passage numbers (Lazic et al., 2018). Consequently, the statistical significance observed in the oxidative stress experiment should not be overinterpreted. The temperature experiment likewise remains preliminary. Future studies should incorporate independent biological replication, including cells from different passages and a broader range of stress intensities. Ideally, they would also include time-course analyses to determine whether Ngb and Cygb exhibit consistent protective effects. A short-term assay may not capture more gradual protective effects, especially if these globins act indirectly through redox buffering, oxygen handling, or modulation of downstream stress pathways rather than through an immediate impact on viability (Keller 4th et al., 2022). Accordingly, the functional relevance of these globins in cellular stress protection remains an open question that will require additional experiments with stronger stress regimes, quantitative expression analyses, and independent biological replication. Future work should also include more controlled comparison of transfection efficiency and protein abundance in addition to the eGFP control since these variables are likely to be critical for interpreting any phenotype associated with globin expression.

Overall, the *in vitro* experiments do not provide strong evidence of a clear cytoprotective effect of lungfish Ngb or Cygb under the tested conditions. However, the results suggest that the constructs may influence cellular stress responses differently. The ability to transiently express the proteins in fish cells and test them in a heterologous system provides a useful foundation for future functional studies. More refined experiments are needed to determine whether the

apparent differences observed here reflect true biological effects or are primarily due to technical variability.

5.5 Limitations and future perspectives

Several limitations should be considered when interpreting the present results. On the bioinformatic side, the genome assembly still contains some unresolved regions and unplaced scaffolds, which likely means that parts of the globin repertoire remain incompletely characterized. In addition, reliance on available transcriptomic datasets may have missed low-abundance or tissue-specific expression, and annotation-based searches may still be biased against highly divergent sequences. These limitations are typical for analyses of non-model organisms with large and repeat-rich genomes, but they nevertheless affect the completeness of the conclusions that can be drawn from the present data (Tørresen et al., 2019; Zhi et al., 2006).

A second limitation involves transcriptomic evidence. While RNA-seq backmapping and PCR validation were useful for confirming the expression of selected loci, the available transcriptomic data was not exhaustive. Therefore, rare, developmental, or strongly tissue-specific expression patterns may have been overlooked (Keller 4th et al., 2022). This is particularly relevant for globin genes, which can exhibit highly specialized or tissue-dependent expression (Keppner et al., 2020). Future work should include broader sampling of transcripts across tissues, developmental stages, and physiological conditions to obtain a more complete picture of globin expression in *N. forsteri*.

The functional experiments are limited by transient transfection, small sample size, and the absence of biological replicates. Although the CHSE-214 cell system provided a practical platform for initial assays (Lannan et al., 1984), it is not an ideal substitute for endogenous expression in lungfish tissues. Furthermore, expressing a heterologous globin in fish cells may not fully recapitulate its native function. Therefore, future studies should explore additional model systems, such as stable cell lines, heterologous recombinant expression systems in bacteria, yeast, or insect cells to enable more controlled expression and more extensive biochemical characterization of purified proteins. Further experimental work should also address the underlying mechanisms of globin function. For Ngb and Cygb, this could entail studying oxygen-binding kinetics, redox activity, nitric oxide interactions, and cellular

localization. Comparative analyses across additional lungfish species would also be valuable, especially since different lungfish lineages vary in respiratory mode and ecological specialization (Jorgensen & Joss, 2016). These comparisons could help distinguish ancestral features from lineage-specific ones and clarify whether the functions observed in *N. forsteri* are representative of lungfish in general.

Finally, the present study underscores the importance of combining bioinformatic and experimental methods. In complex genomes, such as that of the lungfish, no single method can establish the identity, evolutionary history, and function of genes with complete confidence. The combination of annotation, phylogenetics, structural modeling, spectroscopy, and cell-based assays employed in this study provides a useful framework for the future study of challenging gene families in non-model organisms. Expanding this framework through more extensive replication and additional functional assays will likely provide deeper insight into the biology of lungfish globins and their role in vertebrate evolution.

In addition to further functional studies, comparative analyses across lungfish species and other vertebrates would be highly valuable for placing the present findings into a broader evolutionary context.

A comparative analysis of globins in other extant lungfish species, as well as in a broader range of vertebrates, would strengthen the evolutionary interpretation of the presented results. Since the Australian lungfish is the only surviving member of the once more diverse *Neoceratodontidae* family, its globin repertoire provides only a partial view of lungfish globin evolution. Including the African and South American lungfish species would allow us to distinguish features likely ancestral to extant lungfish from those specific to *N. forsteri*. This is particularly relevant because living lungfish species differ markedly in ecology and respiratory physiology. The African and South American lungfishes are obligate air-breathers that can aestivate, whereas the Australian lungfish retains a more ancestral respiratory strategy and does not aestivate (Jorgensen & Joss, 2016). These differences may be reflected in the retention, expansion, or functional diversification of globin genes, particularly those involved in oxygen storage, sensing, and protection against cellular stress.

Comparisons with other vertebrates would also be valuable for reconstructing the evolutionary history of the vertebrate globin repertoire. Lungfishes are the closest living relatives of tetrapods and thus occupy a particularly informative position for inferring the genomic and functional features of their last common ancestor (Meyer et al., 2021). In this respect, lungfish are often

more relevant than ray-finned fish for understanding the origins of vertebrate traits associated with the transition to land. The close genetic relationship between lungfish and tetrapods suggests that similarities in globin repertoires, expression patterns, and structural properties reflect shared ancestry rather than independent evolution (Irisarri & Meyer, 2016). This is particularly significant for tissue globins, such as Ngb and Cygb, which appear to be highly conserved across vertebrates and may represent ancient components of the sarcopterygian globin toolkit.

At the same time, making broader comparisons across vertebrate lineages could help identify which globin features are highly conserved and which evolved independently in response to lineage-specific physiological demands. Contrasts with teleosts, amphibians, reptiles, birds, and mammals, for example, could clarify the extent to which globin gene duplication, loss, and functional specialization have shaped vertebrate globin system diversity. In this context, lungfish could be an important evolutionary link between the more advanced globin repertoires of tetrapods and the frequently highly rearranged globin loci of ray-finned fish. Therefore, their genomes provide a unique opportunity to study how globin families evolved in response to changing oxygen levels and how their functions changed during the evolution of air breathing and terrestrial adaptation.

Overall, comparative studies involving additional lungfish species and a broader sampling of vertebrates would be highly valuable for contextualizing the present findings within a broader evolutionary framework. These analyses could clarify whether the patterns observed in *N. forsteri* reflect ancestral sarcopterygian conditions or lineage-specific adaptations. They would also provide deeper insight into globin gene evolution in vertebrates.

5.6 General conclusions

This thesis offers a multi-level characterization of the globin repertoire in the Australian lungfish, *N. forsteri*, and advances our understanding of globin evolution in early sarcopterygians. Bioinformatic analyses identified a broad set of globin candidates and revealed that *N. forsteri* retains representatives of all major vertebrate globin families. In accordance with the findings in the West African lungfish *Protopterus annectens*, Mb and GbE appear to have undergone lineage-specific expansion (Koch et al., 2016; Lüdemann et al., 2019).

Additionally, Ngb, Cygb, GbX, GbY, Adgb, and Hb components were recovered, which supports the presence of a complex and evolutionarily informative globin repertoire.

A significant finding of the study is the identification and validation of Ngb and Cygb in *N. forsteri*. The detection of these proteins in tissue-specific expression analyses, along with strong phylogenetic support, structural conservation, and spectroscopic evidence for canonical hexacoordination, suggests that they are authentic vertebrate tissue globins (Pesce et al., 2004; Sawai et al., 2003). These findings expand the known distribution of Ngb and Cygb to include lungfish, supporting the view that these globins were already present in the lineage leading to tetrapods. Thus, the study provides evidence of the deep evolutionary conservation of tissue-specific globin functions across sarcopterygian vertebrates.

This work also demonstrates the methodological challenges of analyzing a very large, repetitive genome. Successful recovery of biologically meaningful globin sequences from such a challenging genomic background shows that integrated annotation and validation strategies can produce reliable results, even in non-model organisms. This is relevant not only for lungfish biology but also for comparative genomics more generally, as similar challenges are encountered in other species with complex genome architecture (Ayub et al., 2026).

The functional *in vitro* analyses provided an initial, exploratory view of how lungfish Ngb and Cygb may affect cellular stress responses. While no clear protective effect was demonstrated under the tested conditions, the experiments laid the groundwork for future research and underscored the importance of replication, assay design, and physiological context.

Overall, the results of this thesis support the conclusion that lungfish are a key model for understanding the evolution and potential function of vertebrate globins.

6 References

Alfonso, S., Gesto, M., & Sadoul, B. (2021). Temperature increase and its effects on fish stress physiology in the context of global warming. *Journal of Fish Biology*, *98*(6), 1496–1508.

Alvarez-Carreno, C., Becerra, A., & Lazcano, A. (2016). Molecular evolution of the oxygen-binding hemerythrin domain. *PloS One*, *11*(6), e0157904.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169.

Andrews, S., & Krueger, F. (2010). *FastQC: a quality control tool for high throughput sequence data.* https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ascenzi, P., Leboffe, L., Pesce, A., Ciaccio, C., Sbardella, D., Bolognesi, M., & Coletta, M. (2014). Nitrite-reductase and peroxynitrite isomerization activities of Methanosarcina acetivorans protoglobin. *PLoS One*, *9*(5), e95391.

Ayub, J., Bibi, S., Jabbir, F., Ullah, M., & Ahmad, H. I. (2026). Current Trends, Challenges, and Optimization Strategies in Bioinformatic Pipelines for Whole Genome Sequencing of Non-Model Species. *SciNex Journal of Advanced Sciences*, *1*(1), 66–82.

Bhattacharjee, S. (2019). ROS and Oxidative Modification of Cellular Components. In *Reactive Oxygen Species in Plant Biology*. https://doi.org/10.1007/978-81-322-3941-3_4

Biscotti, M. A., Gerdol, M., Canapa, A., Forconi, M., Olmo, E., Pallavicini, A., Barucca, M., & Schartl, M. (2016). The lungfish transcriptome: A glimpse into molecular evolution events at the transition from water to land. *Scientific Reports*, *6*(February), 1–12. https://doi.org/10.1038/srep21571

Blank, M., Kiger, L., Thielebein, A., Gerlach, F., Hankeln, T., Marden, M. C., & Burmester, T. (2011). Oxygen supply from the bird's eye perspective: globin E is a respiratory protein in the chicken retina. *Journal of Biological Chemistry*, *286*(30), 26507–26515.

Blank, M., Wollberg, J., Gerlach, F., Reimann, K., Roesner, A., Hankeln, T., Fago, A., Weber, R. E., & Burmester, T. (2011). A membrane-bound vertebrate globin. *PloS One*, *6*(9), e25292.

Blommaert, J. (2020). Genome size evolution: towards new model systems for old questions. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1933), 20201441.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brownstein, C. D., Harrington, R. C., & Near, T. J. (2023). The biogeography of extant

lungfishes traces the breakup of Gondwana. *Journal of Biogeography*, *50*(7), 1191–1198. https://doi.org/10.1111/jbi.14609

Buchan, D. W. A., Moffat, L., Lau, A., Kandathil, S. M., & Jones, D. T. (2024). Deep learning for the PSIPRED protein analysis workbench. *Nucleic Acids Research*, *52*(W1), W287–W293.

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60.

Burggren, W. W., & Johansen, K. (1987). *Circulation and Respiration in Lungfishes (Dipnoi). 236*(1986), 217–236.

Burmester, T., Ebner, B., Weich, B., & Hankeln, T. (2002). Cytoglobin: a novel globin type ubiquitously expressed invertebrate tissues. *Molecular Biology and Evolution*, *19*(4), 416–421.

Burmester, T., & Hankeln, T. (2014). Function and evolution of vertebrate globins. *Acta Physiologica*, *211*(3), 501–514.

Carducci, F., Carotti, E., Gerdol, M., Greco, S., Canapa, A., Barucca, M., & Biscotti, M. A. (2021). Investigation of the activity of transposable elements and genes involved in their silencing in the newt Cynops orientalis, a species with a giant genome. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-94193-6

Chan, J.-W., & Teo, A. K. K. (2020). Replicates in stem cell models—How complicated! *Stem Cells*, *38*(9), 1055–1059.

Childress, J. J., & Seibel, B. A. (1998). Life at stable low oxygen levels: adaptations of animals to oceanic oxygen minimum layers. *Journal of Experimental Biology*, *201*(8), 1223–1232.

Chong, Z. X., Yeap, S. K., & Ho, W. Y. (2021). Transfection types, methods and strategies: a technical review. *PeerJ*, *9*, e11165.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K., & Lander, E. S. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences*, *104*(49), 19428–19433.

Cloud, P., & Gibor, A. (1970). The oxygen cycle. *Scientific American*, *223*(3), 110–123.

Corliss, B. A., Delalio, L. J., Iv, T. C. S. K., Keller, A. S., & Berne, R. M. (2019). Vascular Expression of Hemoglobin Alpha in Antarctic Icefish Supports Iron Limitation as Novel Evolutionary Driver. *Frontiers in Physiology*, *10*(November). https://doi.org/10.3389/fphys.2019.01389

Corti, P., Xue, J., Tejero, J., Wajih, N., Sun, M., Stolz, D. B., Tsang, M., Kim-Shapiro, D. B., & Gladwin, M. T. (2016). Globin X is a six-coordinate globin that reduces nitrite to nitric

oxide in fish red blood cells. *Proceedings of the National Academy of Sciences*, *113*(30), 8538–8543.

Craik, C. S., Buchman, S. R., & Beychok, S. (1980). Characterization of globin domains: heme binding to the central exon product. *Proceedings of the National Academy of Sciences*, *77*(3), 1384–1388.

Crawshaw, L. I. (1977). Physiological and behavioral reactions of fishes to temperature change. *Journal of the Fisheries Board of Canada*, *34*(5), 730–734.

Damas, J., Corbo, M., & Lewin, H. A. (2021). Vertebrate Chromosome Evolution. *Annual Review of Animal Biosciences*, *9*, 1–27. https://doi.org/10.1146/annurev-animal-020518-114924

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), giab008.

De Backer, J., & Hoogewijs, D. (2024). The cytoglobin-dependent transcriptome in melanoma indicates a protective function associated with oxidative stress, inflammation and cancer-associated pathways. *Scientific Reports*, *14*(1), 1–15. https://doi.org/10.1038/s41598-024-69224-7

Decker, H., & Van Holde, K. E. (2010). *Oxygen and the Evolution of Life*. Springer Science & Business Media.

Edgar, R. C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications*, *13*(1), 6968.

Estrín, D. A., Luque, F. J., Ilangovan, G., & Zweier, J. L. (2020). Dioxygen binding and sensing proteins. *Antioxidants & Redox Signaling*, *32*(16), 1151–1154.

Exertier, C., Montemiglio, L. C., Freda, I., Gugole, E., Parisi, G., Savino, C., & Vallone, B. (2022). Neuroglobin, clues to function and mechanism. *Molecular Aspects of Medicine*, *84*(November 2021). https://doi.org/10.1016/j.mam.2021.101055

Fuchs, C. (2008). Das Globingen-Repertoire von Amphibien und Teleostiern. *Dissertation*.

Fuchs, C., Burmester, T., & Hankeln, T. (2006). The amphibian globin gene repertoire as revealed by the Xenopus genome. *Cytogenetic and Genome Research*, *112*(3–4), 296–306. https://doi.org/10.1159/000089884

Fuchs, C., Luckhardt, A., Gerlach, F., Burmester, T., & Hankeln, T. (2005). Duplicated cytoglobin genes in teleost fishes. *Biochemical and Biophysical Research Communications*, *337*(1), 216–223. https://doi.org/10.1016/j.bbrc.2005.08.271

Fuselli, S., Greco, S., Biello, R., Palmitessa, S., Lago, M., Meneghetti, C., Mcdougall, C.,

Trucchi, E., Rota Stabelli, O., Biscotti, A. M., Schmidt, D. J., Roberts, D. T., Espinoza, T., Hughes, J. M., Ometto, L., Gerdol, M., & Bertorelle, G. (2023). Relaxation of Natural Selection in the Evolution of the Giant Lungfish Genomes. *Molecular Biology and Evolution*, *40*(9), 1–6. https://doi.org/10.1093/molbev/msad193

Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., & Stanke, M. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research*, *34*(5), 769–777.

Gallagher, M. D., & Macqueen, D. J. (2017). Evolution and expression of tissue globins in ray-finned fishes. *Genome Biology and Evolution*, *9*(1), 32–47. https://doi.org/10.1093/gbe/evw266

Germon, I., Joss, J., Casane, D., Metcalfe, C. J., & File, J. (2012). *Evolution of the Australian Lungfish (Neoceratodus forsteri) Genome : A Major Role for CR1 and L2 LINE Elements Research article*. *29*(11), 3529–3539. https://doi.org/10.1093/molbev/mss159

Hai, H., & Kawada, N. (2020). Role of cytoglobin, a novel radical scavenger, in stellate cell activation and hepatic fibrosis. *Clinical and Molecular Hepatology*, *26*(3), 280.

Hamdane, D., Kiger, L., Dewilde, S., Green, B. N., Pesce, A., Uzan, J., Burmester, T., Hankeln, T., Bolognesi, M., & Moens, L. (2003). The redox state of the cell regulates the ligand binding affinity of human neuroglobin and cytoglobin. *Journal of Biological Chemistry*, *278*(51), 51713–51721.

Hankeln, T., Ebner, B., Fuchs, C., Gerlach, F., Haberkamp, M., Laufs, T. L., Roesner, A., Schmidt, M., Weich, B., & Wystub, S. (2005). Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *Journal of Inorganic Biochemistry*, *99*(1), 110–119.

Hankeln, T., Ebner, B., Fuchs, C., Gerlach, F., Haberkamp, M., Laufs, T. L., Roesner, A., Schmidt, M., Weich, B., Wystub, S., Saaler-Reinhardt, S., Reuss, S., Bolognesi, M., Sanctis, D. De, Marden, M. C., Kiger, L., Moens, L., Dewilde, S., Nevo, E., … Burmester, T. (2005). Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *Journal of Inorganic Biochemistry*, *99*(1), 110–119. https://doi.org/10.1016/j.jinorgbio.2004.11.009

Hankeln, T., Wystub, S., Laufs, T., Schmidt, M., Gerlach, F., Saaler-Reinhardt, S., Reuss, S., & Burmester, T. (2004). The Cellular and Subcellular Localization of Neuroglobin and Cytoglobin-a Clue to Their Function? *IUBMB Life*, *56*(11-12), 671–679.

Helfenrath, K., Sauer, M., Kamga, M., Wisniewsky, M., Burmester, T., & Fabrizius, A. (2021).

The More, the Merrier? Multiple Myoglobin Genes in Fish Species, Especially in Gray Bichir (Polypterus senegalus) and Reedfish (Erpetoichthys calabaricus). *Genome Biology and Evolution*, *13*(7), 1–12. https://doi.org/10.1093/gbe/evab078

Hoeger, U., Ahmed, M. H., Ghatge, M. S., & Safo, M. K. (2020). *Vertebrate and Invertebrate Respiratory Proteins , Lipoproteins and other Body Fluid Proteins* (Subcellula).

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. In *Gene prediction: methods and protocols* (pp. 65–95). Springer.

Hoffmann, F. G., Opazo, J. C., Hoogewijs, D., Hankeln, T., Ebner, B., Vinogradov, S. N., Bailly, X., & Storz, J. F. (2012). Evolution of the globin gene family in deuterostomes: lineage-specific patterns of diversification and attrition. *Molecular Biology and Evolution*, *29*(7), 1735–1745.

Hoffmann, F. G., Opazo, J. C., & Storz, J. F. (2012). Whole-genome duplications spurred the functional diversification of the globin gene superfamily in vertebrates. *Molecular Biology and Evolution*, *29*(1), 303–312.

Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F. G., Fabrizius, A., Moens, L., Burmester, T., Dewilde, S., Storz, J. F., & Vinogradov, S. N. (2012). Androglobin: a chimeric globin in metazoans that is preferentially expressed in mammalian testes. *Molecular Biology and Evolution*, *29*(4), 1105–1114.

Hsia, C. C. W., Schmitz, A., Lambertz, M., Perry, S. F., & Maina, J. N. (2013). Evolution of air breathing: oxygen homeostasis and the transitions from water to land and sky. *Comprehensive Physiology*, *3*(2), 849–915.

Hundahl, C. A., Kelsen, J., & Hay-Schmidt, A. (2013). Neuroglobin and cytoglobin expression in the human brain. *Brain Structure and Function*, *218*(2), 603–609.

Irisarri, I., & Meyer, A. (2016). The identification of the closest living relative (s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. *Systematic Biology*, *65*(6), 1057–1075.

Iritani, Y., Ishikawa, H., Mizuno, M., & Mizutani, Y. (2023). Heme pocket structure and its functional implications in an ancestral globin protein. *Biochemistry*, *62*(18), 2727–2737.

Jorgensen, J. M., & Joss, J. (2016). *The biology of lungfishes*. CRC Press.

Joss, J. M. P. (2006). *Lungfish evolution and development*. *148*, 285–289. https://doi.org/10.1016/j.ygcen.2005.10.010

Katsuji, T., & Lagerspetz, K. Y. H. (1990). Direct adaptation of cells to temperature: membrane fluidity of goldfish cells cultured in vitro at different temperatures. *Comparative Biochemistry and Physiology Part A: Physiology*, *96*(1), 57–60.

Kaufman, D. P., Kandle, P. F., Murray, I. V, & Dhamoon, A. S. (2024). Physiology, Oxyhemoglobin Dissociation Curve. 2023 Jul 31. *StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing*.

Kawada, N. (2015). Cytoglobin as a marker of hepatic stellate cell-derived myofibroblasts. *Frontiers in Physiology*, *6*, 329.

Keller 4th, T. C. S., Lechauve, C., Keller, A. S., Brooks, S., Weiss, M. J., Columbus, L., Ackerman, H., Cortese-Krott, M. M., & Isakson, B. E. (2022). The role of globins in cardiovascular physiology. *Physiological Reviews*.

Kemp, A., Cavin, L., & Guinot, G. (2017). Evolutionary history of lungfishes with a new phylogeny of post-Devonian genera. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *471*, 209–219. https://doi.org/10.1016/j.palaeo.2016.12.051

Keppner, A., Correia, M., Santambrogio, S., Koay, T. W., Maric, D., Osterhof, C., Winter, D. V, Clerc, A., Stumpe, M., & Chalmel, F. (2022). Androglobin, a chimeric mammalian globin, is required for male fertility. *Elife*, *11*, e72374.

Keppner, A., Maric, D., Correia, M., Koay, T. W., Orlando, I. M. C., Vinogradov, S. N., & Hoogewijs, D. (2020). Lessons from the post-genomic era: Globin diversity beyond oxygen binding and transport. *Redox Biology*, *37*, 101687. https://doi.org/10.1016/j.redox.2020.101687

Koch, J., & Burmester, T. (2016). Membrane-bound globin X protects the cell from reactive oxygen species. *Biochemical and Biophysical Research Communications*, *469*(2), 275–280.

Koch, J., Lüdemann, J., Spies, R., Last, M., Amemiya, C. T., & Burmester, T. (2016). Unusual diversity of myoglobin genes in the lungfish. *Molecular Biology and Evolution*, *33*(12), 3033–3041.

Konstantinov, K. B. (1996). Monitoring and control of the physiological state of cell cultures. *Biotechnology and Bioengineering*, *52*(2), 271–289.

Kosmachevskaya, O. V, & Topunov, A. F. (2018). Alternate and additional functions of erythrocyte hemoglobin. *Biochemistry (Moscow)*, *83*(12), 1575–1593.

Kronthaler, F., & Zöllner, S. (2020). R and RStudio. In *Data Analysis with RStudio: An Easygoing Introduction* (pp. 1–12). Springer.

Lane, N. (2002). *Oxygen: the molecule that made the world*. Oxford University Press.

Lannan, C. N., Winton, J. R., & Fryer, J. L. (1984). Fish cell lines: establishment and characterization of nine cell lines from salmonids. *In Vitro*, *20*(9), 671–676.

Lazic, S. E., Clarke-Williams, C. J., & Munafò, M. R. (2018). What exactly is 'N'in cell culture

and animal experiments? *PLoS Biology*, *16*(4), e2005282.

Lerat, E. (2010). Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity*, *104*(6), 520–533.

Lesk, A. M., & Lecomte, J. T. J. (2013). The globin family. *Protein Families: Relating Protein Sequence, Structure, and Function*, 207–235.

Lüdemann. (2019). Functional Diversification of Multiple Lungfish Globins. *Dissertation*, *June*.

Lüdemann, J., Verissimo, K. M., Dreger, K., Fago, A., Schneider, I., & Burmester, T. (2019). Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes. *Scientific Reports*, *9*(1), 280.

Malte, H., & Lykkeboe, G. (2018). The Bohr / Haldane effect : a model-based uncovering of the full extent of its impact on O 2 delivery to and CO 2 removal from tissues. *J Appl Physiol (1985)*, 916–922. https://doi.org/10.1152/japplphysiol.00140.2018

Mammen, P. P. A., Shelton, J. M., Ye, Q., Kanatous, S. B., McGrath, A. J., Richardson, J. A., & Garry, D. J. (2006). Cytoglobin is a stress-responsive hemoprotein expressed in the developing and adult brain. *Journal of Histochemistry & Cytochemistry*, *54*(12), 1349–1361.

Marengo-Rowe, A. J. (2006). Structure-function relations of human hemoglobins. *Baylor University Medical Center Proceedings*, *19*(3), 239–245.

McConnell, R. E., Youniss, M., Gnanasambandam, B., Shah, P., Zhang, W., & Finn, J. D. (2022). Transfection reagent artefact likely accounts for some reports of extracellular vesicle function. *Journal of Extracellular Vesicles*, *11*(10), e12253.

Metcalfe, C. J., Filée, J., Germon, I., Joss, J., & Casane, D. (2012). Evolution of the Australian lungfish (neoceratodus forsteri) genome: A major role for CR1 and L2 LINE elements. *Molecular Biology and Evolution*, *29*(11), 3529–3539. https://doi.org/10.1093/molbev/mss159

Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J. M., Irisarri, I., Wong, W. Y., Nowoshilow, S., Kneitz, S., Kawaguchi, A., Fabrizius, A., Xiong, P., Dechaud, C., Spaink, H. P., Volff, J. N., Simakov, O., Burmester, T., Tanaka, E. M., & Schartl, M. (2021). Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature*, *590*(7845), 284–289. https://doi.org/10.1038/s41586-021-03198-8

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., & Richardson, L. J. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419.

Nishi, H., Inagi, R., Kawada, N., Yoshizato, K., Mimura, I., Fujita, T., & Nangaku, M. (2011). Cytoglobin, a novel member of the globin family, protects kidney fibroblasts against oxidative stress under ischemic conditions. *The American Journal of Pathology*, *178*(1), 128–139.

Nishimura, D. (2000). RepeatMasker. *Biotech Software & Internet Report*, *1*(1–2), 36–39.

Pesce, A., Dewilde, S., Nardini, M., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T., & Bolognesi, M. (2004). The human brain hexacoordinated neuroglobin three-dimensional structure. *Micron*, *35*(1–2), 63–65.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612.

Pezzoli, D., Giupponi, E., Mantovani, D., & Candiani, G. (2017). Size matters for in vitro gene delivery: investigating the relationships among complexation protocol, transfection medium, size and sedimentation. *Scientific Reports*, *7*(1), 44134.

Prachumwat, A., & Li, W.-H. (2008). Gene number expansion and contraction in vertebrate genomes with respect to invertebrate genomes. *Genome Research*, *18*(2), 221–232.

Qi, M., Clark, J., Moody, E. R. R., Pisani, D., & Donoghue, P. C. J. (2024). Molecular dating of the teleost whole genome duplication (3R) is compatible with the expectations of delayed rediploidization. *Genome Biology and Evolution*, *16*(7), evae128.

Ransy, C., Vaz, C., Lombès, A., & Bouillaud, F. (2020). Use of H2O2 to cause oxidative stress, the catalase issue. *International Journal of Molecular Sciences*, *21*(23), 9149.

Rasmussen, J. R., Wells, R. M. G., Henty, K., Clark, T. D., & Brittain, T. (2009). Characterization of the hemoglobins of the Australian lungfish Neoceratodus forsteri (Krefft). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *152*(2), 162–167.

Reeder, B. J. (2023). Insights into the function of cytoglobin. *Biochemical Society Transactions*, *0*(July), 1907–1919. https://doi.org/10.1042/BST20230081

Reeder, B. J., Deganutti, G., Ukeri, J., Atanasio, S., Svistunenko, D. A., Ronchetti, C., Mobarec, J. C., Welbourn, E., Asaju, J., & Vos, M. H. (2024). The circularly permuted globin domain of androglobin exhibits atypical heme stabilization and nitric oxide interaction. *Chemical Science*, *15*(18), 6738–6751.

Reinhard, C. T., Planavsky, N. J., Olson, S. L., Lyons, T. W., & Erwin, D. H. (2016). Earth's oxygen cycle and the evolution of animal life. *Proceedings of the National Academy of Sciences*, *113*(32), 8933–8938.

Reuss, S. (2025). Neuroglobin and Cytoglobin in Mammalian Nervous Systems: About Distribution, Regulation, Function, and Some Open Questions. *Brain Sciences*, *15*(8), 784. https://doi.org/10.3390/brainsci15080784

Roesner, A., Fuchs, C., Hankeln, T., & Burmester, T. (2005). A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals. *Molecular Biology and Evolution*, *22*(1), 12–20.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & 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. https://doi.org/10.1093/sysbio/sys029

Sacerdot, C., Louis, A., Bon, C., Berthelot, C., & Roest Crollius, H. (2018). Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biology*, *19*(1), 1–15. https://doi.org/10.1186/s13059-018-1559-1

Salin, K., Auer, S. K., Rey, B., Selman, C., & Metcalfe, N. B. (2015). Variation in the link between oxygen consumption and ATP production, and its relevance for animal performance. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151028.

Sawai, H., Kawada, N., Yoshizato, K., Nakajima, H., Aono, S., & Shiro, Y. (2003). Characterization of the heme environmental structure of cytoglobin, a fourth globin in humans. *Biochemistry*, *42*(17), 5133–5142.

Schartl, M., Woltering, J. M., Irisarri, I., Du, K., Kneitz, S., Pippel, M., Brown, T., Franchini, P., Li, J., & Li, M. (2024). The genomes of all lungfish inform on genome expansion and tetrapod evolution. *Nature*, *634*(8032), 96–103.

Schmidt, M., Gerlach, F., Avivi, A., Laufs, T., Wystub, S., Simpson, J. C., Nevo, E., Saaler-Reinhardt, S., Reuss, S., & Hankeln, T. (2004). Cytoglobin is a respiratory protein in connective tissue and neurons, which is up-regulated by hypoxia. *Journal of Biological Chemistry*, *279*(9), 8063–8069.

Schneider, T. B. (2023). *Investigation of the Function and Evolution of the Globin Superfamily*. Saint Louis University.

Schuster, C. D., Salvatore, F., Moens, L., & Martí, M. A. (2024). Globin phylogeny, evolution and function, the newest update. *Proteins: Structure, Function and Bioinformatics*, *92*(6), 720–734. https://doi.org/10.1002/prot.26659

Schwarze, K., & Burmester, T. (2013). Conservation of globin genes in the “living fossil” Latimeria chalumnae and reconstruction of the evolution of the vertebrate globin family.

Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, *1834*(9), 1801–1812.

Schwarze, K., Singh, A., & Burmester, T. (2015). The full globin repertoire of turtles provides insights into vertebrate globin evolution and functions. *Genome Biology and Evolution*, *7*(7), 1896–1913.

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. In *Gene prediction: methods and protocols* (pp. 227–245). Springer.

Shivapurkar, N., Stastny, V., Okumura, N., Girard, L., Xie, Y., Prinsen, C., Thunnissen, F. B., Wistuba, I. I., Czerniak, B., & Frenkel, E. (2008). Cytoglobin, the newest member of the globin family, functions as a tumor suppressor gene. *Cancer Research*, *68*(18), 7448–7456.

Singh, S., & Varma, A. (2017). Structure, function, and estimation of leghemoglobin. In *Rhizobium biology and biotechnology* (pp. 309–330). Springer.

Smagghe, B. J., Trent III, J. T., & Hargrove, M. S. (2008). NO dioxygenase activity in hemoglobins is ubiquitous in vitro, but limited by reduction in vivo. *PloS One*, *3*(4), e2039.

Smit, A. F. A., Hubley, R., & Green, P. (2015a). *RepeatMasker Open-4.0. 2013–2015*. Seattle, USA.

Smit, A. F. A., Hubley, R., & Green, P. (2015b). RepeatModeler Open-1.0. 2008–2015. *Seattle, USA: Institute for Systems Biology*, *1*, 2018.

Smith, M. M., Krupina, N. I., & Joss, J. (2002). Developmental Constraints Conserve Evolutionary Pattern in an Osteichthyan Dentition. *Connective Tissue Research*, *43*(2–3), 113–119. https://doi.org/10.1080/03008200290000925

Sousa, J. D. F., Amaral, D. B., & America, S. (2024). *The genomes of all lungfish inform on genome expansion and tetrapod evolution*. *July 2023*. https://doi.org/10.1038/s41586-024-07830-1

Stepanenko, A. A., & Heng, H. H. (2017). Transient and stable vector transfection: Pitfalls, off-target effects, artifacts. *Mutation Research/Reviews in Mutation Research*, *773*, 91–103.

Storz, J. F., Opazo, J. C., & Hoffmann, F. G. (2013). Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Molecular Phylogenetics and Evolution*, *66*(2), 469–478.

Sugimoto, H., Makino, M., Sawai, H., Kawada, N., Yoshizato, K., & Shiro, Y. (2004). Structural basis of human cytoglobin for ligand binding. *Journal of Molecular Biology*, *339*(4), 873–885.

Tarone, R. E., Scudiero, D. A., & Robbins, J. H. (1983). Statistical methods for in vitro cell

survival assays. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *111*(1), 79–96.

Tempel, S. (2019). *Chapter 2 Using and Understanding RepeatMasker*. *January 2012*. https://doi.org/10.1007/978-1-61779-603-6

Thannickal, V. J. (2009). Oxygen in the evolution of complex life and the price we pay. In *American journal of respiratory cell and molecular biology* (Vol. 40, Issue 5, pp. 507–510). American Thoracic Society.

Thomson, K. S. (1972). An Attempt to Reconstruct Evolutionary Changes in the Cellular DNA Content of Lungfish. *Journal of Experimental Zoology*, *180*(3), 363–371.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192.

Thuy, L. T. T., Hai, H., Anh, P. T., Chi, N. B. T., Bao, T. Van, Huyen, T. D. A., Sang, N. T. Q., & Hermiston, M. L. (2026). Therapeutic Potential of Cytoglobin and Neuroglobin in Oxidative Stress-Driven Liver Diseases. *Antioxidants*, *15*(4), 485.

Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., Gruca, A., Grynberg, M., Kajava, A. V, & Promponas, V. J. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Research*, *47*(21), 10994–11006.

Turilli-Ghisolfi, E. S., Lualdi, M., & Fasano, M. (2023). Ligand-based regulation of dynamics and reactivity of hemoproteins. *Biomolecules*, *13*(4), 683.

Uno, T., Ryu, D., Tsutsumi, H., Tomisugi, Y., Ishikawa, Y., Wilkinson, A. J., Sato, H., & Hayashi, T. (2004). Residues in the distal heme pocket of neuroglobin: implications for the multiple ligand binding steps. *Journal of Biological Chemistry*, *279*(7), 5886–5893.

Uppal, S., Kumar, A., Shandilya, M., Mukhi, N., Singh, A. K., Kateriya, S., Kaur, J., & Kundu, S. (2016). Penta-and hexa-coordinate ferric hemoglobins display distinct pH titration profiles measured by Soret peak shifts. *Analytical Biochemistry*, *510*, 120–128.

Vanek, T., & Kohli, A. (2019). Biochemistry, myoglobin. *StatPearls Publishing*.

Vinogradov, S. N., & Moens, L. (2008). Diversity of globin function: enzymatic, transport, storage, and sensing. *Journal of Biological Chemistry*, *283*(14), 8773–8777.

Vinogradov, S. N., Tinajero-Trejo, M., Poole, R. K., & Hoogewijs, D. (2013). Bacterial and archaeal globins—A revised perspective. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1834*(9), 1789–1800.

Voltarelli, V. A., Alves de Souza, R. W., Miyauchi, K., Hauser, C. J., & Otterbein, L. E. (2023).

Heme: the lord of the iron ring. Antioxidants 12: 1074.

Wajcman, H., Kiger, L., & Marden, M. C. (2009). Structure and function evolution in the superfamily of globins. *Comptes Rendus Biologies*, *332*(2–3), 273–282.

Wang, G., Huang, Y., Zhang, N., Liu, W., Wang, C., Zhu, X., & Ni, X. (2021). Hydrogen sulfide is a regulator of hemoglobin oxygen-carrying capacity via controlling 2, 3-BPG production in erythrocytes. *Oxidative Medicine and Cellular Longevity*, *2021*(1), 8877691.

Weber, R. E., & Jensen, F. B. (2011). Respiratory adaptations in lungfish blood and hemoglobin. *The Biology of Lungfishes*, 283–303.

Wolfsberg, T. G., & Madden, T. L. (1999). Sequence similarity searching using the BLAST family of programs. *Current Protocols in Protein Science*, *15*(1), 2–5.

Wystub, S., Ebner, B., Fuchs, C., Weich, B., Burmester, T., & Hankeln, T. (2004). Interspecies comparison of neuroglobin, cytoglobin and myoglobin: Sequence evolution and candidate regulatory elements. *Cytogenetic and Genome Research*, *105*(1), 65–78. https://doi.org/10.1159/000078011

Yañez, I. M., Torres-Cuevas, I., & Corral-Debrinski, M. (2025). Neuroglobin: A promising candidate to treat neurological diseases. *Neural Regeneration Research*, *21*(4), 1292.

Zhang, Y., Park, C., Bennett, C., Thornton, M., & Kim, D. (2021). Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Research*, *31*(7), 1290–1295.

Zhi, D., Raphael, B. J., Price, A. L., Tang, H., & Pevzner, P. A. (2006). Identifying repeat domains in large genomes. *Genome Biology*, *7*(1), R7.

7 Supplementary Materials

N. forsteri **RepeatMasker summary**

```
Repeat Classes
==============
Total Sequences: 46
Total Length: 34557647948 bp
Class                   Count         bpMasked    %masked
=====                   =====         ========     =======
Unspecified             65950586      24439207411   70.72%
                       ---------------------------------
    total interspersed  65950586      24439207411   70.72%

Simple_repeat           3930681       167130649    0.48%
-----------------------------------------------------------
Total                   69881267      24606338060   71.20%
```

N. forsteri **genome annotation**

The BRAKER3 annotation files are provided at the Forschungsdatenrepositorium - FDM - UHH and accessible under:

https://doi.org/10.25592/uhhfdm.18981

N. forsteri **globin candidates**

>NfoCygb(Cygb)
MEQVKQGEMQMEKRERGDQLLDSEKMMIRESWARVFANCEDVGVSILIRFFVNFPSAKQYFSQFRHLE
DPLEMERSIQLRKHARRIMNAINTVVENLGDPEKVSTVVALVGKAHALKHKVEPMYFKILNGVILEVFP
EEYPDDFTPEVQRAWTKLMGLFYCHVMASYKEVGWVQYPNSTM
>NfoG2(GbE)
MALSDAEVQNACGAWAKLYANAEDNGAIMLASRMFTDYSDIKSYFTQFKGMGSTVEMEGSAQVRSH
GVGLYFSE
>NfoG12(GbE)
MALSDAEVQNARGAWAKLYAKSEDNGAIVLARMFTDYPDTKSYFTNFKGMGSAAEMEGSAQIRGHG
KKVFASLNDMVQHLDSMDALTGIISPLGKKHATQLKVDPKNFRIICDIILQLLEEKFGGDARAAFDKVT
NVVCAQMNTAYREAGW
>NfoG13(GbE)
MALSDAEVQNARGAWAKLYANAEDNGATVLARMFTDYPDTKSYFANFKGMGSAAEMEGSAQIRGH
GKKVFASLNDMVQHLDSMDALTGIISPLGKKHATQLKVDPKNFRIICDIILQLLEEKFGGDARAAFDKV
TNVVCAQMNTAYKEAGW
>NfoG57(GbE)
MEGASSAPEGSNLEGRQSILHVLSEAEKELIVASWEQIYQNCEDVGVQVLTRYFLEFPDAKEHFRQFRH
LQSGEEMGQCPQLRRHACRIMNIMRDVIKNLEDCGRLSATVRVVGKSHALTHNVSFHFHQMLTNTML
EVFTTDFPHLFSDEAQEAWTKLMDIVCSQIAEAYEEVGQHV
>NfoG3(Mb)
MATLSDAQWNNVLTFWDSKIEPSLANHGHEVLIRMFLNYEPTKEKFPKFQHLKTEAEMRSSDELKKHG
NTVFTALGKVLKLKGNHENELYPMAESHANKHKIPLNNFNIICEVIDDYLQKTYPADYVADTRGAMKT
ALALTVQTLTKLYQEVEK
>NfoG14(Mb)
MATLSDAQWNNVLTFWDSKIEPSLANHGHEVLIRMFLNYEPTKEKFPKFQHLKTEAEMRSSDELKKHG
NTVFTALGKVLKLKGNHENELYPMAESHANKHKIPLNNFNIICEVIDDYLQKTYPADYVADTRGSSRGL
DIELKDLREEMADIYNQINQSG
>NfoG5(Mb)
MATLSDAQWNNVFEFWDSKIEPSLANHGHEVLVRRMFLNYKPTKEYFPKFQNLVTEAEMRSSDEMKK
FGNTVFTALGKVLKLKGNHENDLYPVAKSHAIEHKVPLNNIICTVIVDYLHETYPAEYGADTRDAMKT
ALALIVKSMTKVYKEVQK
>NfoG6(Mb)
MATLSDAQWNNVFEFWDSKIEPSLANHGHEVLVRRMFLNYKPTKEYFPKFQNLVTEAEMRSSDEMKK
FGNTVFTALGKVLKLKGNHENDLYPVAKSHAIEHKVPLNNIICTVIVDYLHETYPAEYGADTRDAMKT
ALALIVKSMTKVYKEVQK
>NfoG15(Mb)
MATLSDAQWNNILTFWDSKIEPSLANHGHEVLVRMFVNYKATKEFFPKFQHLATEAEMRSSDDVKKH
GNVVFTALGKALKLKGNHENDLYPMAESHANKHKIPLNNFDIICKVIDEYLHEKYPAEYGADTQGAM
KTALALVVKTLNELYKKVQK

>NfoG4(Mb)
MATLSDAQWNNVLKFWDSKIAPNSANHGREVLIRMFLNYKTSKEYFPKFQNLVTEAEMRSSEELKKH
GDITLISLGKVLKLKGKPETELNPMTELHATNLKVPLIAFDDIICKVIDDYLHETYPADYKADTRDAMKT
ALALMIQAMNKRYKQIEHKA
>NfoG1(Ngb)
MEYPSLSMREKELIRESWGQLGKEKLQHGTVLFSRLFELDPDLVHLFHYNTTFATPAECLTSPEFLDHIH
KVMSVIDAAVNNLENLPSLEEYLINLGKKHKTVGVKIDSFPAVGESLLYMLERALGTPLCTEIKAAWIV
LYTFVTKTMIRGWDKNGENKIE
>NfoG11(GbX)
MRCALSGPGITPGKSAPDVSDSGGETGPETGSGILKSNPLPLTESQKELIHESWKILHQDIARVGIIMFIRL
FETHPECKDVFFLFRDIDDFQQLRMSKELQAHGLRCFRVMSFIEKSVARLEQEEKLEQIAFELGRSHCRY
NAPPKYYEQYVGIQFISAVKPILKEKWTSDVEEAWESLFKYLITMMKKGFQEEEKKHPINKSVYPKKQN
SSGMSSVQNSV
>NfoG46(GbY)
MLDFCYTKSPTITSTLPCIPLPYFSHSGTDPLAIKAFPVIQISVSNINTLYVLSVMGILHASFTFMNDTNQQI
TQASFTQSKCRQVLFLKFLFCTLKKLFLLLLLCSGNVELNPEKYKATTLPLPAPIKISTDNITTTSKKDILN
LLNNHFAMAVLDFDSSPASMLCSSLLNSDTSSPH
>NfoG18(Adgb)
MNYKKVDYNDMHFYDLSKFTITTFLFHLIMIDQSEAETIEVRKDTERMDEIKAMKQAWEAAEPGRAVK
ALQSRLQFINKCLRKMSVEPTNETNGVITTAVPGDSGTL
>NfoG110(Hb)
MVHWTERERKLTEDVWIHLHIPEVGCHILARLLYVYPWTRRCFHKFNLGSDFRSIYGNSHVREHGIIVL
RLVAMAADSLGDVKNVFKLSKLYFEEFNVEPHTFKLLGQIVVIVLAQRLGQKFTPEVQKSTEKFVSIYV
DAMSRDPTDPYWQRG
>NfoG111(Hb)
MVEWTAEEKKIIQSLWKDLNVADVGHRVLARLLFVYPWTRRYFKNFNLGHDFNSIYGNSHVTSHGNV
VLTEVDKAVHHLGGIKELLHNLSLEHSEKLGVDPHNFKLIGQIIIIVLAQDLGKKFTADAQEHFEKFVSE
LENPIFHAPKTGDTRVDCLAYANDLVLISTSVLGLQRKCNYFS
KLCDTHKLRINMSKTKILIFGNTKNILNHYG
>NfoG112(Hb)
MQCCALFSASERRKGTKAANMVFTSADSTNIQNAWHAIKAIPNAGGESLARCFESFPGSKAYFHHFGH
DLSASSAKVKEHGQKVFAALDLAVNNLHSLESALSSLSDVHATKLMVDPVNFGVST

The complete dataset containing all globin candidates is provided at the Forschungsdatenrepositorium - FDM - UHH and accessible under: https://doi.org/10.25592/uhhfdm.18979

Alignment Fig. 7

```
Hb3_NfoG112                MQCC----------------------------ALF-----S-A-SE-RR
Mb_Anguilla_anguilla       -------------------------------------------------
Mb_Denticeps_clupeoides    -------------------------------------------------
Mb_Epinephe...lanceolatus  -------------------------------------------------
Hb2_NfoG111                MV-------------------------------------------------
Hb1_NfoG110                MV-------------------------------------------------
Hb_beta1_Danio_rerio       MV-------------------------------------------------
Hb-beta-1_C...odon_idella  MV-------------------------------------------------
Hb_beta2_Mus_musculus      MV-------------------------------------------------
Hb-beta_Carassius_gibelio  MV-------------------------------------------------
Hb-beta-2_C...odon_idella  MV-------------------------------------------------
Hb-beta-2_C..._adamanteus  MV-------------------------------------------------
Hb_beta1_Xenopus_laevis    M-------------------------------------------------
Hb_beta_Lat...a_chalumnae  V-------------------------------------------------
GbE1_NfoG2                 M-------------------------------------------------
Mb2_NfoG14                 MA------------------------------------------------
Hb_alpha_La...a_chalumnae  -------------------------------------------------
Hb-alpha_Ca...ius_gibelio  M-------------------------------------------------
Hb_alpha1_Homo_sapiens     M-------------------------------------------------
Hb-alpha_Acinonyx_jubatus  M-------------------------------------------------
Mb3_NfoG5                  MA------------------------------------------------
Mb4_NfoG6                  MA------------------------------------------------
Mb6_NfoG4                  MA------------------------------------------------
Mb1_NfoG3                  MA------------------------------------------------
Mb5_NfoG15                 MA------------------------------------------------
Mb_Alligator_sinensis      M-------------------------------------------------
Mb_Mus_musculus            M-------------------------------------------------
Mb_Delphinapterus_leucas   M-------------------------------------------------
Mb_Balaenoptera_borealis   M-------------------------------------------------
Mb_Homo_sapiens            M-------------------------------------------------
Mb_Zootoca_vivipara        M-------------------------------------------------
Mb_Cricetulus_griseus      M-------------------------------------------------
Mb_Gallus_gallus           M-------------------------------------------------
Mb_Crotalus_tigris         M-------------------------------------------------
GbE_Chelydra_serpentina    M-------------------------------------------------
GbE_Chrysem...icta_bellii  M-------------------------------------------------
GbE3_NfoG13                M-------------------------------------------------
GbE_Gallus_gallus          M-------------------------------------------------
GbE_Patagioenas_fasciata   M-------------------------------------------------
```

```
GbE2_NfoG12                M-----------------------------------------------
GBE_Latimeria_chalumnae    M-----------------------------------------------
GbE4_NfoG57                MEGA--------------------------SSAPEG--S-N-LE-GR
GbY_Callorhinchus_milii    MT----------------------------------------------
GbY_Latimeria_chalumnae    MA----------------------------------------------
GbY_NfoG46                 MP----------------------------------------------
GbY_Xenopus_laevis         MA----------------------------------------------
GbY_Rhinella_marina        MA----------------------------------------------
GbY_Xenopus_tropicalis     MA----------------------------------------------
Cygb-1_Astyanax_mexicanus  MEGN----------------------------------G-GMQE-QM
Cygb-2_Alosa_sapidissima   MDSG----------------------------------R-D-EE-PR
Cygb-1_Danio_rerio         MEGD----------------------------------G-G-VQ-LT
Cygb-1_Cyprinus_carpio     MEGD----------------------------------G-D-VE-LT
Cygb-1_Cten...odon_idella  MEGD----------------------------------R-G-VE-LT
Cygb_Neocer...us_forsteri  MEQVK--------------------------QGE-----M-Q-ME-KR
Cygb-2_Anguilla_anguilla   MEEKE-------------------------QGE-----R-E-ME-HR
Cygb_Rhinat..._bivittatum  MEKV--------------------------QGE-----M-E-MD-RW
Cygb_Camelus_dromedarius   MEKV--------------------------PGE-----M-E-IE-RR
Cygb_Callithrix_jacchus    MEKV--------------------------AGE-----M-E-IE-RR
Cygb_Bos_taurus            MEKV--------------------------PGE-----M-E-IE-RR
Cygb_Homo_sapiens          MEKV--------------------------PGE-----M-E-IE-RR
Cygb_Mus_musculus          MEKV--------------------------PGD-----M-E-IE-RR
Cygb_Rattus_norvegicus     MEKV--------------------------PGD-----M-E-IE-RR
Cygb_Budorcas_taxicolor    MEKV--------------------------PGE-----M-E-IE-RR
Cygb_Bufo_bufo             MEKV--------------------------QEE-----N-E-TE-RW
Cygb-2_Brie...rachyistius  MEEV--------------------------RGD-----G-A-ME-QR
Cygb-2_Danio_rerio         MEKE--------------------------RED-----E-E-TE-GR
Cygb_Chelydra_serpentina   MEKV--------------------------QGE-----M-E-IE-RW
Cygb_Xenopus_laevis        MEKV--------------------------QGE-----N-D-ME-RW
Cygb_Gallus_gallus         MEKV--------------------------QGE-----M-E-IE-RW
Cygb-2_Amph...n_ocellaris  MSRRESPPPPSPPPP--LLGL--Q---------RGE-----V-Q-CEERP
Cygb-2_Chelmon_rostratus   MSRRESPPPPSPPPPPQLLGV--R---------RGE-----V-E-CEDRP
GbX_NfoG11                 MRCALSGPGIT---------P--GKSAPDVSDSGGE-TGPETGS-GI-LK
GbX_Latimeria_chalumnae    MGCAISGLSWR---------AVKG---------PGE-EGK-G-K-EG-EA
GbX_Tetraod...igroviridis  MGCAISSLGAE---------A--E---------FGD-----R-S-AE-EE
GbX_Perca_fluviatilis      MGCAISGLAAK---------A--E---------FGD-----R-T-TE-EA
GbX_Plectro...s_leopardus  MGCAISGLAAK---------A--E---------FGE-----R-T-TE-DA
Ngb_Mycteria_americana     MES---------------------------------------------
Ngb_Homo_sapiens           M-----------------------------------------------
Ngb_Mus_musculus           M-----------------------------------------------
Ngb_Bufo_gargarizans       MES---------------------------------------------
```

```
Ngb_Xenopus_laevis          MEK---------------------------------------------
Ngb_Xenopus_tropicalis      MEK---------------------------------------------
GbX_Salmo_salar             MGCAISVLA-P---------K--T---------AGE-----C----K-TE
Ngb_Danio_rerio             ME----------------------------------------------
Ngb_Dissostichus_mawsoni    ME----------------------------------------------
Ngb_Oncorhy...tshawytscha   ME----------------------------------------------
Ngb_Oncorhy...s_gorbuscha   ME----------------------------------------------
Ngb_Embiotoca_jacksoni      ME----------------------------------------------
Ngb_Neoceratodus_forsteri   MEY---------------------------------------------
Ngb_Latimeria_chalumnae     MEK---------------------------------------------
GbX_Lepisosteus_oculatus    MGCAISGLGLA---------P-------------KQ-----I-D-AT-EE
GbX_Danio_rerio             MGCAISGSGLT---------A--G---------APE-----I-R-PG-EE
GbX_Carassius_auratus       MGCAISGSGLT---------A--R---------APE-----I-R-AG-EE

Hb3_NfoG112                 KGTK-AANMVFTSADSTNIQNAWH-AIKA-IPNAGGESL-A-RCFESFPG
Mb_Anguilla_anguilla        ------------MTDFELVLKAWK-PIEADLKGNGGVVL-T-RLFQEHPE
Mb_Denticeps_clupeoides     ------------MADFDMVLNCWG-KVEADYTGTGGEVL-T-RLFKEHPE
Mb_Epinephe...lanceolatus   ------------MADYDAVLKCWS-AMEADYSGIGNLVM-R-RLFKEFPQ
Hb2_NfoG111                 ---------EWTAEEKKIIQSLWK-DLN--VADVGHRVL-A-RLLFVYPW
Hb1_NfoG110                 ---------HWTERERKLTEDVWI-HLH--IPEVGCHIL-A-RLLYVYPW
Hb_beta1_Danio_rerio        ---------EWTDAERTAILGLWG-KLN--IDEIGPQAL-S-RCLIVYPW
Hb-beta-1_C...odon_idella   ---------EWTDAERTAILGLWG-KLN--IDEIGPQAL-S-RCLIVYPW
Hb_beta2_Mus_musculus       ---------HLTDAEKSAVSCLWA-KVN--PDEVGGEAL-G-RLLVVYPW
Hb-beta_Carassius_gibelio   ---------EWTDAERKAIKDVWG-KLK--GDIIGPQAL-A-RCLIVYPW
Hb-beta-2_C...odon_idella   ---------EWTEFERTTIQDIFS-KMN--YEVVGPQAL-A-RCLIVYPW
Hb-beta-2_C..._adamanteus   ---------HWTPEEKQLITALWG-KVD--VPEVGAATL-G-KLLVVYPW
Hb_beta1_Xenopus_laevis     ---------GLTAHDRQLINSTWG-KLC--AKTIGQEAL-G-RLLWTYPW
Hb_beta_Lat...a_chalumnae   ---------HWTETERATIETVYQ-KLH--LDEVGREAL-T-RLFIVYPW
GbE1_NfoG2                  ---------ALSDAEVQNACGAWA-KLYANAEDNGAIMLAS-RMFTDYSD
Mb2_NfoG14                  ---------TLSDAQWNNVLTFWDSKIEPSLANHGHEVL-I-RMFLNYEP
Hb_alpha_La...a_chalumnae   ---------GLTAADKTLIKSIWG-KVEKETEAIGVEAL-V-RLFKCFPQ
Hb-alpha_Ca...ius_gibelio   ---------SLSDKDKAVVKALWA-KIGSRADEIGAEAL-G-RMLTVYPQ
Hb_alpha1_Homo_sapiens      ---------VLSPADKTNVKAAWG-KVGAHAGEYGAEAL-E-RMFLSFPT
Hb-alpha_Acinonyx_jubatus   ---------VLSAADKSNVKACWG-KIGSHAGEYGAEAL-E-RTFCSFPT
Mb3_NfoG5                   ---------TLSDAQWNNVFEFWDSKIEPSLANHGHEVL-VRRMFLNYKP
Mb4_NfoG6                   ---------TLSDAQWNNVFEFWDSKIEPSLANHGHEVL-VRRMFLNYKP
Mb6_NfoG4                   ---------TLSDAQWNNVLKFWDSKIAPNSANHGREVL-I-RMFLNYKT
Mb1_NfoG3                   ---------TLSDAQWNNVLTFWDSKIEPSLANHGHEVL-I-RMFLNYEP
Mb5_NfoG15                  ---------TLSDAQWNNILTFWDSKIEPSLANHGHEVL-V-RMFVNYKA
Mb_Alligator_sinensis       ---------QLSDQEWKHVLDIWA-KVESKLPEHGHEVI-I-RLLQEHPE
Mb_Mus_musculus             ---------GLSDGEWQLVLNVWG-KVEADLAGHGQEVL-I-GLFKTHPE
```

```
Mb_Delphinapterus_leucas   ---------GLSEGEWQLVLNVWG-KVEADLAGHGQDVL-I-RLFKGHPE
Mb_Balaenoptera_borealis   ---------VLSDAEWQLVLNIWA-KVEADVAGHGQDIL-I-RLFKGHPE
Mb_Homo_sapiens            ---------GLSDGEWQLVLNVWG-KVEADIPGHGQEVL-I-RLFKGHPE
Mb_Zootoca_vivipara        ---------GLSDQEWQKVVDIWG-KVEPDLPAHGQEVI-I-RMFQKHPE
Mb_Cricetulus_griseus      ---------GLSDGEWQKVLNVWG-KVEADLAGHGQEVL-I-RLFKNHPE
Mb_Gallus_gallus           ---------GLSDQEWQQVLTIWG-KVEADIAGHGHEVL-M-RLFHDHPE
Mb_Crotalus_tigris         ---------GLTEQEWQKVMDIWE-KVEPDLSLHGQEVI-I-RMFQNHPE
GbE_Chelydra_serpentina    ---------AFSEAEVQRARGAWE-KMYANAEDNGTTVL-V-RMFTEHPD
GbE_Chrysem...icta_bellii  ---------AFSEAEVQRARGAWE-KMYANAEDNGTTVL-V-RMFTEHPD
GbE3_NfoG13                ---------ALSDAEVQNARGAWA-KLYANAEDNGATVL-A-RMFTDYPD
GbE_Gallus_gallus          ---------SFSEAEVQSARGAWE-KMYVDAEDNGTAVL-V-RMFTEHPD
GbE_Patagioenas_fasciata   ---------SFSEAEVQSARGAWE-KMYADAEDNGTTVL-V-RMFTEHPD
GbE2_NfoG12                ---------ALSDAEVQNARGAWA-KLYAKSEDNGAIVL-A-RMFTDYPD
GBE_Latimeria_chalumnae    ---------ALSDAEVQTARDVWG-QIYANAEENGTIIL-V-RMFTEHPD
GbE4_NfoG57               Q----SILHVLSEAEKELIVASWE-QIYQNCEDVGVQVL-T-RYFLEFPD
GbY_Callorhinchus_milii    ---------GITEADKENIHFIWE-KLYENPEENGKTIV-L-RMFTDYPE
GbY_Latimeria_chalumnae    ---------ALTEADKQNIRGIWK-TVFENAEENGRTIV-I-RLFEKYPE
GbY_NfoG46                 ---------SLTETDVQNIRGIWG-VICQNPEENGRTVV-I-RLFIDYPE
GbY_Xenopus_laevis         ---------DLTAADIENINEIWC-KIYANPEESGKTVV-I-RLFTTYPQ
GbY_Rhinella_marina        ---------DLTAADIENINEIWC-KIYANPEESGKTVV-I-RLFTTYPQ
GbY_Xenopus_tropicalis     ---------DLTGADIENINEVWS-KIYANPEESGRTVV-I-SLFLTYPQ
Cygb-1_Astyanax_mexicanus  E-----RPEMLTDGEQAIIQNTWT-KVFQNKEAAGVAVL-M-RLFTSFPS
Cygb-2_Alosa_sapidissima   E-----RLEPLSDTELAIIQDTWA-RVYENCEDVGVTIL-I-RFFVNFPS
Cygb-1_Danio_rerio        Q-----SPDSLTEEDVCVIQDTWK-PVYAERDNAGVAVL-V-RFFTNFPS
Cygb-1_Cyprinus_carpio    Q-----SPETLTEDDVCVIQDTWR-PVYENRENAGVAVL-I-RFFSNFPS
Cygb-1_Cten...odon_idella Q-----SPETLTEDDVCVIQDTWR-PVYENRESAGVAVL-I-RFFGNFPS
Cygb_Neocer...us_forsteri  E-----RGDQLLDSEKMMIRESWA-RVFANCEDVGVSIL-I-RFFVNFPS
Cygb-2_Anguilla_anguilla   E-----RAEPLSEAERGMILDTWA-RVYENCEDVGVSIL-I-RFFVNFPS
Cygb_Rhinat..._bivittatum  E-----RSEEISDAEKRVIQETWS-RVYTNCEDVGVSIL-I-RFFVNFPS
Cygb_Camelus_dromedarius   E-----RSEELSEAERKAVQATWA-RLYANCEDVGVAIL-V-RFFVNFPS
Cygb_Callithrix_jacchus    E-----RSEELSEAERKAVQATWA-RLYAHCEDVGVAIL-V-RFFVNFPS
Cygb_Bos_taurus            E-----RSEELSEAERKAVQATWA-RLYANCEDVGVAIL-V-RFFVNFPS
Cygb_Homo_sapiens          E-----RSEELSEAERKAVQAMWA-RLYANCEDVGVAIL-V-RFFVNFPS
Cygb_Mus_musculus          E-----RSEELSEAERKAVQATWA-RLYANCEDVGVAIL-V-RFFVNFPS
Cygb_Rattus_norvegicus     E-----RNEELSEAERKAVQATWA-RLYANCEDVGVAIL-V-RFFVNFPS
Cygb_Budorcas_taxicolor    E-----RSEELSEAERKAVQATWA-RLYANCEDVGVAIL-V-RFFVNFPS
Cygb_Bufo_bufo             E-----RPEEITESERDIIKETWA-RLYTNCEDVGVSIL-I-RFFVNFPK
Cygb-2_Brie...rachyistius  E-----RADQLSEAETGMIQDTWA-RVYENCEDVGVAIL-I-RFFVNFPS
Cygb-2_Danio_rerio         E-----RPEPLTDVERGIIKDTWA-RVYASCEDVGVTIL-I-RFFVNFPS
Cygb_Chelydra_serpentina   E-----RSEELSDAERKVIQETWS-RVYTNCEDVGVSIL-I-RFFVKFPS
Cygb_Xenopus_laevis        E-----RLEDITESERGVIKETWA-RVYTNCEDVGVSIL-I-RFFVNFPS
```

```
Cygb_Gallus_gallus          E-----RSEEISDAEKKVIQETWS-RVYANCEDVGVSIL-I-RFFVNFPS
Cygb-2_Amph...n_ocellaris   E-----RAEPLSDAEREIIQDSWG-HVYKNCEDVGVSVL-I-RFFVNFPS
Cygb-2_Chelmon_rostratus    E-----RAEPLSDAEREIIQDTWG-HVYKNCEDVGVSVL-I-RFFVNFPS
GbX_NfoG11                  S-----NPLPLTESQKELIHESWK-ILHQDIARVGIIMF-I-RLFETHPE
GbX_Latimeria_chalumnae     E-----ALPAISQLQIHLIQESWK-LIQEDIAKVGIIMF-V-RLFETHPE
GbX_Tetraod...igroviridis   DAAAAAAVVYPREDQIQMIKDSWK-VIRDDIAKVGIIMF-V-RLFETHPE
GbX_Perca_fluviatilis       A-----GAAYPSEDQIQMIKESWK-VIQDDIAKVGIIMF-V-RLFETHPE
GbX_Plectro...s_leopardus   ------AAVFPSEDQIQMIKESWK-VIRDDIAKVGIIMF-V-RLFETHPE
Ngb_Mycteria_americana      -------GLLLSGRQRALIRESWQ-RVSGNPVQHGLVLF-T-RLFDLDPD
Ngb_Homo_sapiens            -----------ERPEPELIRQSWR-AVSRSPLEHGTVLF-A-RLFALEPD
Ngb_Mus_musculus            -----------ERPESELIRQSWR-VVSRSPLEHGTVLF-A-RLFALEPS
Ngb_Bufo_gargarizans        --------LQLPEPQKELIRDTWR-IVGQEQIHHGTVLF-T-RLFQLEPE
Ngb_Xenopus_laevis          --------DQLSGPQKELIRESWQ-TVSQDQLHHGTVLF-S-RLFELEPE
Ngb_Xenopus_tropicalis      --------DQLSGPQKELIRESWQ-TVSQDQLHHGTVLF-S-RLFELEPE
GbX_Salmo_salar             D-----AVVHLSKEQIEMIKESWK-VIEEDIAKVGIIMF-V-RLFETHPE
Ngb_Danio_rerio             ---------KLSEKDKGLIRDSWE-SLGKNKVPHGIVLF-T-RLFELDPA
Ngb_Dissostichus_mawsoni    ---------KLSEKDKELIRGSWE-SLGKNKVPHGVVMF-S-RLFELDPE
Ngb_Oncorhy...tshawytscha   ---------KLTEKDKELIRGSWE-SLGKNKVPHGVVMF-S-RLFELEPA
Ngb_Oncorhy...s_gorbuscha   ---------KLTEKDKELIRGSWE-SLGKNKVPHGVVMF-S-RLFELEPA
Ngb_Embiotoca_jacksoni      ---------KLSGKDKELIRGSWE-SLGKNKVPHGVVMF-S-RLFELDPA
Ngb_Neoceratodus_forsteri   --------PSLSMREKELIRESWG-QLGKEKLQHGTVLF-S-RLFELDPD
Ngb_Latimeria_chalumnae     --------HNFSVRSKELIRESWD-RLGKNKLPHGTVMF-T-RLFELDPD
GbX_Lepisosteus_oculatus    E-----ALPHLSDHHIEQIKDSWK-VIQEDIAKVGIIMF-V-RLFETHPE
GbX_Danio_rerio             E-----TPAGLTTNHIRLIKESWR-LIQEDIAKVGIIMF-V-RLFETHPE
GbX_Carassius_auratus       E-----TPAGLTANHIRLIKESWR-LTQEDIAKVGIIMF-V-RLFETHPE

Hb3_NfoG112                 SKAYFHH-FGHDLS----ASSAKVKEHGQK---VFAALDLAVNNLHSLE-
Mb_Anguilla_anguilla        TQQLFPK-FAAIAP-GDLAGNAAISEHGCT---VLTKLGDLLHAKGNHA-
Mb_Denticeps_clupeoides     TLKLFPK-FAGIAQ-GDLAGNAAVCDHGAT---VLKKLGELLKAKGKHA-
Mb_Epinephe...lanceolatus   SLNHFPK-FDGMSQ-ADLASNGDLTAQGVT---VLTKLGEVLKTRGNHG-
Hb2_NfoG111                 TRRYFKN-FNLGHDFNSIYGNSHVTSHGNV---VLTEVDKAVHHLGGIK-
Hb1_NfoG110                 TRRCFHK-FNLGSDFRSIYGNSHVREHGII---VLRLVAMAADSLGDVK-
Hb_beta1_Danio_rerio        TQRYFAT-FGNLSSPAAIMGNPKVAAHGRT---VMGGLERAIKNMDNVK-
Hb-beta-1_C...odon_idella   TQRYFAT-FGNLSSPAAIIGNPKVAAHGKT---VMGGLERAIKNLDNIK-
Hb_beta2_Mus_musculus       TQRYFDS-FGDLSSASAIMGNPKVKAHGKK---VITAFNEGLKNLDNLK-
Hb-beta_Carassius_gibelio   TQRYFGS-FGSLSDAAAIMANKKVAAHGKV---VLNGLGRAVQSMDDIK-
Hb-beta-2_C...odon_idella   TQRYFGK-FGNLYNAAAILGNPMVAAHGTV---VLHGLDRAVKNMDNIK-
Hb-beta-2_C..._adamanteus   TQRFFAH-FGNLSGPSAICANPLVRAHGKK---VMNSFAEAVKHMDNIK-
Hb_beta1_Xenopus_laevis     TQRYFSS-FGNLNSADAVFHNEAVAAHGEK---VVTSIGEAIKHMDDIK-
Hb_beta_Lat...a_chalumnae   TTRYFKS-FGDLSSSKAIASNPKVTEHGLK---VMNKLTEAIHNLDHIK-
GbE1_NfoG2                  IKSYFTQ-FKGMGSTVEMEGSAQVRSHGVG---LYFS-------------
```

```
Mb2_NfoG14                TKEKFPK-FQHLKTEAEMRSSDELKKHGNT---VFTALGKVLKLKGNHE-
Hb_alpha_La...a_chalumnae SKVYFDH-FTDLS-----PSSQKLHAHAKV---VLGALTKAVNHLDNIT-
Hb-alpha_Ca...ius_gibelio TKTYFSH-WSDLS-----PGSGPVKKHGKT---IMGAVGDAVSKIDDLV-
Hb_alpha1_Homo_sapiens    TKTYFPH-FDL-S-----HGSAQVKGHGKK---VADALTNAVAHVDDMP-
Hb-alpha_Acinonyx_jubatus TKTYFPH-FDL-S-----HGSAQVKAHGQK---VADALTQAVAHMDDLP-
Mb3_NfoG5                 TKEYFPK-FQNLVTEAEMRSSDEMKKFGNT---VFTALGKVLKLKGNHE-
Mb4_NfoG6                 TKEYFPK-FQNLVTEAEMRSSDEMKKFGNT---VFTALGKVLKLKGNHE-
Mb6_NfoG4                 SKEYFPK-FQNLVTEAEMRSSEELKKHGDI---TLISLGKVLKLKGKPE-
Mb1_NfoG3                 TKEKFPK-FQHLKTEAEMRSSDELKKHGNT---VFTALGKVLKLKGNHE-
Mb5_NfoG15                TKEFFPK-FQHLATEAEMRSSDDVKKHGNV---VFTALGKALKLKGNHE-
Mb_Alligator_sinensis     TQERFEK-FKHMKTADEMKSSEEMKKHGTN---VFAALGNILKQKGNHA-
Mb_Mus_musculus           TLDKFDK-FKNLKSEEDMKGSEDLKKHGCT---VLTALGTILKKKGQHA-
Mb_Delphinapterus_leucas  TLEKFDK-FKHLKTEAEMKASEDLKKHGNT---VLTALGGILKKKGHHD-
Mb_Balaenoptera_borealis  TLEKFDK-FKHLKTEAEMKASEDLKKHGNT---VLTALGGILKKKGHHE-
Mb_Homo_sapiens           TLEKFDK-FKHLKSEDEMKASEDLKKHGAT---VLTALGGILKKKGHHE-
Mb_Zootoca_vivipara       TQERFDK-FKNLKTVDDMKSSEELKKHGTT---VLTALGKILKQKGNHE-
Mb_Cricetulus_griseus     TLDKFDK-FKNLKSEDEMKASEDLKKHGCT---VLTALGVILKKKGQHA-
Mb_Gallus_gallus          TLDRFDK-FKGLKTPDQMKGSEDLKKHGAT---VLTQLGKILKQKGNHE-
Mb_Crotalus_tigris        TQERFVK-FKNFKTLDEMKNSEELKKHGTT---VLSALGKILKQKGSHE-
GbE_Chelydra_serpentina   TKSYFTH-FKGMGTAEEMEQSDQVRSHGKR---VLTTINDLVQHLDSTDA
GbE_Chrysem...icta_bellii TKSYFTH-FKGMGTAEEMEQSDQVRSHGKR---VLTTINDLVQHLDSTDA
GbE3_NfoG13               TKSYFAN-FKGMGSAAEMEGSAQIRGHGKK---VFASLNDMVQHLDSMDA
GbE_Gallus_gallus         TKSYFTH-FKGMDSAEEMKQSDQVRGHGKR---VFTAINDMVQHLDNTEA
GbE_Patagioenas_fasciata  TKSYFTH-FKGMDSAEEMKQSDQVRGHGKK---VFTAINDMVQHLDNSEA
GbE2_NfoG12               TKSYFTN-FKGMGSAAEMEGSAQIRGHGKK---VFASLNDMVQHLDSMDA
GBE_Latimeria_chalumnae   TKSYFGN-FKGMGSAAEMEQSAQVRTHGKK---IFSALNDMIQHLDSTDA
GbE4_NfoG57               AKEHFRQ-FRHLQSGEEMGQCPQLRRHACR---IMNIMRDVIKNLEDCGR
GbY_Callorhinchus_milii   TKMYFQH-FKNISTLEEMKKSPQIKRHGKI---VMSALNKLIANLDNGEE
GbY_Latimeria_chalumnae   TKVYFKN-FKNISTMEEMQKNEQIRIHGLR---VMNSLNQVIQNIDNLNE
GbY_NfoG46                TRKYFKN-FKNIDTEEGIKQSSQVRDHGRR---VVTLLNKVIENLGDWNE
GbY_Xenopus_laevis        TKVYFKN-LKNIATLEEMQVNPGIRAHGKR---VMGALNQVIQNLNDWEV
GbY_Rhinella_marina       TKVYFKN-LKNIATLEEMQVNPGIRAHGKG---VMGALNQVIQNLNDWEV
GbY_Xenopus_tropicalis    TKIYFKN-LKNISTLQEMQDNAGIRAHGKR---VMGALNHVIENLKDWDA
Cygb-1_Astyanax_mexicanus SKQFFSQ-FRDTEDPEELKSSVQLKKHALR---VMSALNTLVENVNDEEK
Cygb-2_Alosa_sapidissima  AKQYFSQ-FQDMEDPEEMEKSSQLRQHARR---VMNAINSVVENLSDADK
Cygb-1_Danio_rerio        AKQYFEH-FRELQDPAEMQQNAQLKKHGQR---VLNALNTLVENLRDADK
Cygb-1_Cyprinus_carpio    AKQYFDQ-FRDMHDPEEMKQSVQLKKHALR---VMTALNTLVENLRDGDK
Cygb-1_Cten...odon_idella AKQYFSQ-FRDIEDPEEMKQSVQLKKHALR---VMTALNTLVENLRDGDN
Cygb_Neocer...us_forsteri AKQYFSQ-FRHLEDPLEMERSIQLRKHARR---IMNAINTVVENLGDPEK
Cygb-2_Anguilla_anguilla  AKQYFSQ-FRDMDDPEEMERSTQLRKHAGR---VMNAINTVLENLHDPDK
Cygb_Rhinat..._bivittatum AKQYFSQ-FKHMEDPLEMERSLQLRKHARR---VMGAVNNVVENVNDMDK
```

```
Cygb_Camelus_dromedarius   AKQYFSQ-FKHMEEPLEMERSPQLRKHACR---VMGALNTVVENLHDPEK
Cygb_Callithrix_jacchus    AKQYFSQ-FKHMEDPLEMERSPQLRKHACR---VMGALNTVVENLHDPDK
Cygb_Bos_taurus            AKQYFSQ-FKHMEEPLEMERSPQLRKHACR---VMGALNTVVENLHDPEK
Cygb_Homo_sapiens          AKQYFSQ-FKHMEDPLEMERSPQLRKHACR---VMGALNTVVENLHDPDK
Cygb_Mus_musculus          AKQYFSQ-FRHMEDPLEMERSPQLRKHACR---VMGALNTVVENLHDPDK
Cygb_Rattus_norvegicus     AKQYFSQ-FKHMEDPLEMERSPQLRKHACR---VMGALNTVVENLHDPDK
Cygb_Budorcas_taxicolor    AKQYFSQ-FKHMEEPLEMERSPQLRKHACR---VMGALNTVVENLHDPEK
Cygb_Bufo_bufo             AKQYFSQ-FKDMEDPLEMERSTQLRKHGRR---VMGAINSVVENLGDPEK
Cygb-2_Brie...rachyistius  AKQYFSQ-FQDMEDMEEMERSTQLRKHAQR---VMNALNTVVENLHDPDK
Cygb-2_Danio_rerio         AKQYFSQ-FQDMEDPEEMEKSSQLRKHARR---VMNAINTVVENLHDPEK
Cygb_Chelydra_serpentina   AKQYFSQ-FKHMEDPLEMERTPQLRKHARR---VMGAVNTVVENINDAEK
Cygb_Xenopus_laevis        AKQHFSQ-FKHMEDPLEMEGSMQLRKHARR---VMGAVNSVVENLGDPEK
Cygb_Gallus_gallus         AKQYFSQ-FKHMDDTLEMERSLQLRKHAQR---VMGAINTVVENLDDPEK
Cygb-2_Amph...n_ocellaris  AKQYFSQ-FQDMEDPEEMERSSQLRHHARR---VMNAINTVVENLHDPEK
Cygb-2_Chelmon_rostratus   AKQYFSQ-FQDMEDPEEMERSSQLRHHARR---VMNAINTVVENLHDPEK
GbX_NfoG11                 CKDVFFL-FRDIDDFQQLRMSKELQAHGLRCFRVMSFIEKSVARLEQEEK
GbX_Latimeria_chalumnae    CKDVFFL-FRDVDDLQRLRTSKELRAHGLR---VMSFVEKSVARLEHSER
GbX_Tetraod...igroviridis  CKDVFFL-FRDVEDLERLRSSRELRAHGLR---VMSFIEKSVARLDQQDR
GbX_Perca_fluviatilis      CKDVFFL-FRDVEDLERLRTSRELRAHGLR---VMSFIEKSVARLDQLER
GbX_Plectro...s_leopardus  CKDVFFL-FRDVEDLERLRTSRELRAHGLR---VMSFIEKSVARLDQLER
Ngb_Mycteria_americana     LLPLFQYNCKQFASPQECLSTPEFLDHIRK---VMLVIDAAVSHLENLSC
Ngb_Homo_sapiens           LLPLFQYNCRQFSSPEDCLSSPEFLDHIRK---VMLVIDAAVTNVEDLSS
Ngb_Mus_musculus           LLPLFQYNGRQFSSPEDCLSSPEFLDHIRK---VMLVIDAAVTNVEDLSS
Ngb_Bufo_gargarizans       LLFLFHYNSAHFSSVQECLSSSEFTEHIRK---VMTVIDAAVRSLDSLPS
Ngb_Xenopus_laevis         LIFFFQYNSSHFSEVQDCLSSAEFTEHIRK---VMTVIDAAVSSLDCLAS
Ngb_Xenopus_tropicalis     LVFLFQYNSSHFSKVQDCLSSAEFTEHIRK---VMTVIDAAVSSLDCLSS
GbX_Salmo_salar            CKDVFFL-FRDVEDRDRLRISRELRAHGLR---VMSFIEKSVARLEQLER
Ngb_Danio_rerio            LLTLFSY-STNCGDAPECLSSPEFLEHVTK---VMLVIDAAVSHLDDLHT
Ngb_Dissostichus_mawsoni   LLTLFHY-TTNCGSTQDCLSSPEFLEHVTK---VMLVIDAAVSNLDDLPS
Ngb_Oncorhy...tshawytscha  LLNLFHY-NTNCSPTQDCLSSPEFLDHVTK---VMLVIDAAVSHLDDLHT
Ngb_Oncorhy...s_gorbuscha  LLNLFHY-NTNCSPTQDCLSSPEFLDHVTK---VMLVIDAAVSHLDNLHT
Ngb_Embiotoca_jacksoni     LLSLFHY-STNCGSTQDCLSSPEFLDHVTK---VMLVIDAAVSHLDDLHS
Ngb_Neoceratodus_forsteri  LVHLFHY-NTTFATPAECLTSPEFLDHIHK---VMSVIDAAVNNLENLPS
Ngb_Latimeria_chalumnae    LMHLFNY-NSSFPSQAGCLTSPEFIDHIHK---VMMVVDAAVNSLDNLLS
GbX_Lepisosteus_oculatus   CKDVFFL-FRDVEDLERLRASKELRAHGLK---VMSFIEKTVARLDQLDR
GbX_Danio_rerio            CKDVFFL-FRDVEDLERLRTSRELRAHGLR---VMSFIEKSVARLDQLER
GbX_Carassius_auratus      CKDVFFL-FRDVEDLERLRTSRELRAHGLR---VMSFIEKSVARLDQLER

Hb3_NfoG112                --SALSSLSDVHATKLMVDPVNFG-V-S----------------------
Mb_Anguilla_anguilla       --DILKPLAKTHATQHKIKLQNFQ-L-ITEVIVKLMGEKGV----DA-AG
Mb_Denticeps_clupeoides    --DILKPLATRHANTHKIGLNNFK-L-ISEVIVKVMAEKAG-L--DG-AG
```

```
Mb_Epinephe...lanceolatus  --EIIKNLSKSHANQHKVPIGHFK-V-MAEILGKVMEEKAG-L--DG-PG
Hb2_NfoG111                --ELLHNLSLEHSEKLGVDPHNFK-L-IGQIIIIVLAQDLG-KKFTA-DA
Hb1_NfoG110                --NVF-KLSKLYFEEFNVEPHTFK-L-LGQIVVIVLAQRLG-QKFTP-EV
Hb_beta1_Danio_rerio       --NTYAALSVMHSEKLHVDPDNFR-L-LADCITVCAAMKFGQAGFNA-DV
Hb-beta-1_C...odon_idella  --ATYSALSVMHSEKLHVDPDNFR-L-LADCITVCAAMKFGPSGFNA-DV
Hb_beta2_Mus_musculus      --GTFASLSELHCDKLHVDPENFR-L-LGNAIVIVLGHHLG-KDFTP-AA
Hb-beta_Carassius_gibelio  --NTYAPLSVLHSEKLHVDPDNFR-L-LGDCLTIVVASQLG-RSFTP-DI
Hb-beta-2_C...odon_idella  --ATYAELSVLHSEKLHVDPDNFR-L-LADCLTVVVAGQLG-ASFTP-EV
Hb-beta-2_C..._adamanteus  --ETFSKLSELHCEKLHVDPENFR-L-LGNILIIVLAGHHG-KEFIP-ST
Hb_beta1_Xenopus_laevis    --GYYAQLSKYHSETLHVDPLNFK-R-FGGCLSIALARHFH-EEYTP-EL
Hb_beta_Lat...a_chalumnae  --DLFHKLSEKHFHELHVDPQNFK-L-LSKCLIIVLATKLG-KQLTP-DV
GbE1_NfoG2                 --------------------------------------------------
Mb2_NfoG14                 --NELYPMAESHANKHKIPLNNFN-I-ICEVIDDYLQKTYP-ADYVA-DT
Hb_alpha_La...a_chalumnae  --DTLHDISLVHAKKLLVDPVNFE-L-LGHCLEVALAAHFA-TDFTP-EV
Hb-alpha_Ca...ius_gibelio  --GALSSLSELHAFKLRIDPANFK-I-LAHNVIVVIGMLFP-GDFTP-EV
Hb_alpha1_Homo_sapiens     --NALSALSDLHAHKLRVDPVNFK-L-LSHCLLVTLAAHLP-AEFTP-AV
Hb-alpha_Acinonyx_jubatus  --NALSALSDLHAYKLRVDPVNFK-F-LSHCLLVTLACHHP-AEFTP-AV
Mb3_NfoG5                  --NDLYPVAKSHAIEHKVPLN--N-I-ICTVIVDYLHETYP-AEYGA-DT
Mb4_NfoG6                  --NDLYPVAKSHAIEHKVPLN--N-I-ICTVIVDYLHETYP-AEYGA-DT
Mb6_NfoG4                  --TELNPMTELHATNLKVPLIAFDDI-ICKVIDDYLHETYP-ADYKA-DT
Mb1_NfoG3                  --NELYPMAESHANKHKIPLNNFN-I-ICEVIDDYLQKTYP-ADYVA-DT
Mb5_NfoG15                 --NDLYPMAESHANKHKIPLNNFD-I-ICKVIDEYLHEKYP-AEYGA-DT
Mb_Alligator_sinensis      --EVLKPLAESHALKHKIPVKYLE-F-ISEIIVKVIAEKYP-VDFGA-DS
Mb_Mus_musculus            --AEIQPLAQSHATKHKIPVKYLE-F-ISEIIIEVLKKRHS-GDFGA-DA
Mb_Delphinapterus_leucas   --AELKPLAQSHATKHKIPIKYLE-F-ISEAIIHVLHSRHP-GDFGA-DA
Mb_Balaenoptera_borealis   --AELKPLAQSHATKHKIPIKYLE-F-ISDAIIHVLHSRHP-GDFGA-DA
Mb_Homo_sapiens            --AEIKPLAQSHATKHKIPVKYLE-F-ISECIIQVLQSKHP-GDFGA-DA
Mb_Zootoca_vivipara        --AELAPLSQTHANTHKIPVKYLE-F-ICEIIVGVIAEKHS-ADFGA-DS
Mb_Cricetulus_griseus      --AEIQPLAQSHATKHKIPVKYLE-F-ISESIIEVLKSRHS-GDFGD-DA
Mb_Gallus_gallus           --SELKPLAQTHATKHKIPVKYLE-F-ISEVIIKVIAEKHA-ADFGA-DS
Mb_Crotalus_tigris         --AELSPLAQTHANRHKIPVKYLE-F-ISEVIIDVIAEKHS-ADFGA-DS
GbE_Chelydra_serpentina    FLGIVNPLGKKHAMQLKVDPKNFR-I-ICDIILQLMEEKYG-----G-DC
GbE_Chrysem...icta_bellii  FLGIVNPLGKKHAMQLKVDPKNFR-I-ICDIILQLMEEKYG-----G-DC
GbE3_NfoG13                LTGIISPLGKKHATQLKVDPKNFR-I-ICDIILQLLEEKFG-----G-DA
GbE_Gallus_gallus          FLGILNPLGQKHATQLKIDPKNFR-I-ICDIILQLMEEKFG-----G-DC
GbE_Patagioenas_fasciata   FLGIVNPLGKKHATQLKIDPKNFR-I-ICDIILQLMEEKFG-----G-DC
GbE2_NfoG12                LTGIISPLGKKHATQLKVDPKNFR-I-ICDIILQLLEEKFG-----G-DA
GBE_Latimeria_chalumnae    LLGVVNPLGKKHATQLKVDPKNFK-I-ICNILLQVLDEKFG-----G-DA
GbE4_NfoG57                LSATVRVVGKSHALTHNVSFHFHQ-M-LTNTMLEVFTTDFP-HLFSD-EA
GbY_Callorhinchus_milii    LSSLLAKMAERHINVHKVDLHNFQ-I-IFNIIIAILEETFG-NAFTP-EI
GbY_Latimeria_chalumnae    VYSILTHLAKRHQYVHRVDVHNFK-L-IFGVIIKILKEALG-ATFTE-EI
```

```
GbY_NfoG46                AAKLLSKLADRHRDYHKVDVFNFQ-L-LFAALDSVYVDAFG-PAFTP-EI
GbY_Xenopus_laevis        VSSALTHLAQRHQDVHKVGVNNFQ-L-LFLVILTIFKEALG-ADFTP-EH
GbY_Rhinella_marina       VSSALTHLAQRHQDVHKVGVNNFQ-L-LFLVILTIFKEALG-ADFTP-EH
GbY_Xenopus_tropicalis    VCSALSHLAKRHQDVHKVEVNNFE-L-LFLVIISVFKEALG-SGFTP-EQ
Cygb-1_Astyanax_mexicanus TAAVLKVVAKSHAIKHSVEPRYFK-L-LTGVILEVLVEAFP-ETFGV-EA
Cygb-2_Alosa_sapidissima  VSSVLSLVGKAHALKHKVEPVYFK-I-LSGVILEVLADDFG-ECFTT-DV
Cygb-1_Danio_rerio        LNTIFNQMGKSHALRHKVDPVYFK-I-LAGVILEVLVEAFP-QCFSPAEV
Cygb-1_Cyprinus_carpio    LNTIFQQMGKSHALKHKVDPLYFK-I-LAGVILEVLVEAFP-QCFSPASV
Cygb-1_Cten...odon_idella LNTIFNQMGKSHALRHKVDPVYFK-I-LAGVILEVLVEAFP-QSFSPASV
Cygb_Neocer...us_forsteri VSTVVALVGKAHALKHKVEPMYFK-I-LNGVILEVFPEEYP-DDFTP-EV
Cygb-2_Anguilla_anguilla  VSSVLGLVGKAHALKHKVEPMYFK-I-LSGVILEVLSEDYP-DCFTA-EV
Cygb_Rhinat..._bivittatum VSTILALVGKSHALKHKVEPVYFK-I-LTGVMLEVIAEEYA-NDFTP-EV
Cygb_Camelus_dromedarius  VSSVLALVGKAHALKHKVEPVYFK-I-LSGVILEVIAEEFA-NDFPP-ET
Cygb_Callithrix_jacchus   VSSVLALVGKAHALKHKVEPVYFK-I-LSGVILEVIAEEFA-NDFPP-ET
Cygb_Bos_taurus           VSSVLSLVGKAHALKHKVEPVYFK-I-LSGVILEVIAEEFA-NDFPP-ET
Cygb_Homo_sapiens         VSSVLALVGKAHALKHKVEPVYFK-I-LSGVILEVVAEEFA-SDFPP-ET
Cygb_Mus_musculus         VSSVLALVGKAHALKHKVEPMYFK-I-LSGVILEVIAEEFA-NDFPV-ET
Cygb_Rattus_norvegicus    VSSVLALVGKAHALKHKVEPMYFK-I-LSGVILDVIAEEFA-NDFPV-ET
Cygb_Budorcas_taxicolor   VSSVLALVGKAHALKHKVEPVYFK-I-LSGVILEVIAEEFA-SDFPP-ET
Cygb_Bufo_bufo            VATVLAIVGKSHAVKHKVD----------------------------
Cygb-2_Brie...rachyistius VASILNLVGKAHAIKHKVEPMYFK-I-LSGVILEILSEDCP-ESFTP-DV
Cygb-2_Danio_rerio        VSSVLVLVGKAHAFKYKVEPIYFK-I-LSGVILEILAEEFG-ECFTP-EV
Cygb_Chelydra_serpentina  VSSVLALVGKAHALKHKVEPVYFK-F-FTGVMLEVIAEEYA-NDFTP-EV
Cygb_Xenopus_laevis       VTTVLSIVGKSHALKHKVEPVYFK-I-LTGVMLEVIAEEYA-KDFTP-DV
Cygb_Gallus_gallus        VSSVLALVGKAHALKHKVEPVYFK-K-LTGVMLEVIAEAYG-NDFTP-EA
Cygb-2_Amph...n_ocellaris VSSVLALVGKAHAIKHQVEPIYFK-I-LSGVMLEVLSEDFP-EFFTA-EV
Cygb-2_Chelmon_rostratus  VSSVLALVGKAHAIKHKVEPMYFK-I-LSGVMLEVFAEDFP-EFFTA-EV
GbX_NfoG11                LEQIAFELGRSHCR-YNAPPKYYE-QYVGIQFISAVKPILK-EKWTS-DV
GbX_Latimeria_chalumnae   LEQLALELGKSHYR-YNAPPKYYQ-Y-VGIEFVCAVQPILK-ERWTP-EV
GbX_Tetraod...igroviridis LEALAVELGKSHYH-YNAPPKYYS-Y-VGAEFICAVQPILK-ERFTS-EL
GbX_Perca_fluviatilis     LEALALELGKSHYH-YNAPPKYYS-Y-VGAEFICAVQPILK-ERWTA-EL
GbX_Plectro...s_leopardus LEALALELGKSHYH-YNAPPKYYS-Y-VGAEFICAVQPILK-ERWTA-EL
Ngb_Mycteria_americana    LEEYLCNLGKKHQA-VGVKVESFS-T-VGESLLYMLEKCLG-TAFSP-DM
Ngb_Homo_sapiens          LEEYLASLGRKHRA-VGVKLSSFS-T-VGESLLYMLEKCLG-PAFTP-AT
Ngb_Mus_musculus          LEEYLTSLGRKHRA-VGVRLSSFS-T-VGESLLYMLEKCLG-PDFTP-AT
Ngb_Bufo_gargarizans      LEDYLTSLGRKHRA-TGVKLESFK-M-VGEALLYALESGLG-ATFTS-DA
Ngb_Xenopus_laevis        LDEYLTSLGRKHKA-VGVKPESFK-T-VGESLLFALESGLG-DAFTS-DT
Ngb_Xenopus_tropicalis    LDEYLTSLGRKHRA-VGVKLESFN-T-VGESLLFALESCLG-DAFTS-DT
GbX_Salmo_salar           LEVLAVELGRSHYP-YNAPPKYYS-Y-VGAEFICAVQPILK-ERWTP-EL
Ngb_Danio_rerio           LEDFLLNLGRKHQA-VGVNTQSFA-L-VGESLLYMLQSSLG-PAYTT-SL
Ngb_Dissostichus_mawsoni  LEDFLLNLGRKHQA-VGVNTQSFA-E-VGESLLHMLQCSLG-QAYTA-PL
```

```
Ngb_Oncorhy...tshawytscha  LEDFLLSLGKKHQA-VGVNTQSFA-V-VGESLLYMLQCSLG-HGYTG-PL
Ngb_Oncorhy...s_gorbuscha  LEDFLLNLGKKHQA-VGVNTQSFA-V-VGESLLYMLQCSLG-HGYTG-PL
Ngb_Embiotoca_jacksoni     LEDFLLNLGRKHQA-VGVNTQSFA-V-VGESLLYMLQCSLG-QAYTA-PL
Ngb_Neoceratodus_forsteri  LEEYLINLGKKHKT-VGVKIDSFP-A-VGESLLYMLERALG-TPLCT-EI
Ngb_Latimeria_chalumnae    LEDYLINLGKKHQA-AGVKMESFP-V-VGESLLYMLEHGLD-SAFSV-EV
GbX_Lepisosteus_oculatus   LDQLALELGKSHYR-YNAPPKYYG-Y-VGTEFICAVQPILK-EKWTS-EV
GbX_Danio_rerio            LETLALELGKSHYR-YNAPPKYYG-Y-VGAEFICAVRPILK-DRWTP-EL
GbX_Carassius_auratus      LETLALELGKSHYR-YNAPPKYYG-Y-VGAEFICAVRPILK-DRWTP-EL

Hb3_NfoG112                ------------------------------------------------
Mb_Anguilla_anguilla       QEAVRKV---MQAVIGDIDNF----------------------------
Mb_Denticeps_clupeoides    QDALRRV---LGIVINDIDIY----------------------------
Mb_Epinephe...lanceolatus  QQALKRL---MTAVIADIEAN----------------------------
Hb2_NfoG111                QEHFEKF---VSELENPIFHAPKTGDTRVDCLAYANDLVLISTSVLGLQR
Hb1_NfoG110                QKSTEKF---VSIYVDAMSRD----------------------------
Hb_beta1_Danio_rerio       QEAWQKF---LAVVVSALCRQ----------------------------
Hb-beta-1_C...odon_idella  QEAWQKF---LSVVVSALCRQ----------------------------
Hb_beta2_Mus_musculus      QAAFQKV---VAGVATALAHK----------------------------
Hb-beta_Carassius_gibelio  QAAWQKF---LAVVVSALGKQ----------------------------
Hb-beta-2_C...odon_idella  QAAFQKF---IAVVISALGRQ----------------------------
Hb-beta-2_C..._adamanteus  HAAFQKL---TNVVAHALAYR----------------------------
Hb_beta1_Xenopus_laevis    HAAYEHL---FDAIADALGKG----------------------------
Hb_beta_Lat...a_chalumnae  QATWEKL---LSVVVAALSRE----------------------------
GbE1_NfoG2                 ------------------------------------------------
Mb2_NfoG14                 RGSSRGLDIELKDLREEMADI----------------------------
Hb_alpha_La...a_chalumnae  HLAIDKF---LYEVEKALFET----------------------------
Hb-alpha_Ca...ius_gibelio  HMSVDKF---FQNLALALSEK----------------------------
Hb_alpha1_Homo_sapiens     HASLDKF---LASVSTVLTSK----------------------------
Hb-alpha_Acinonyx_jubatus  HASLDKF---FSAVSTVLTSK----------------------------
Mb3_NfoG5                  RDAMKTA---LALIVKSMTKV----------------------------
Mb4_NfoG6                  RDAMKTA---LALIVKSMTKV----------------------------
Mb6_NfoG4                  RDAMKTA---LALMIQAMNKR----------------------------
Mb1_NfoG3                  RGAMKTA---LALTVQTLTKL----------------------------
Mb5_NfoG15                 QGAMKTA---LALVVKTLNEL----------------------------
Mb_Alligator_sinensis      QAAMRKA---LELFRNDMASK----------------------------
Mb_Mus_musculus            QGAMSKA---LELFRNDIAAK----------------------------
Mb_Delphinapterus_leucas   QGAMNKA---LELFRKDIAAK----------------------------
Mb_Balaenoptera_borealis   QAAMNKA---LELFRKDIAAK----------------------------
Mb_Homo_sapiens            EGAMNKA---LELFRKDMASN----------------------------
Mb_Zootoca_vivipara        QAAMRKA---LELFRNDMASK----------------------------
Mb_Cricetulus_griseus      QGAMSKA---LELFRNDIAAK----------------------------
```

```
Mb_Gallus_gallus            QAAMKKA---LELFRNDMASK----------------------------
Mb_Crotalus_tigris          QEAMRKA---LELFRNDMASK----------------------------
GbE_Chelydra_serpentina     KASFEKV---TNEICTRLNNA----------------------------
GbE_Chrysem...icta_bellii   KASFEKV---TNEICTRLNNA----------------------------
GbE3_NfoG13                 RAAFDKV---TNVVCAQMNTA----------------------------
GbE_Gallus_gallus           KASFEKV---TNEICTHLTNI----------------------------
GbE_Patagioenas_fasciata    KSSFEKV---TNEICSHLNNV----------------------------
GbE2_NfoG12                 RAAFDKV---TNVVCAQMNTA----------------------------
GBE_Latimeria_chalumnae     RAGFEKV---TDVLCTHLNHA----------------------------
GbE4_NfoG57                 QEAWTKL---MDIVCSQIAEA----------------------------
GbY_Callorhinchus_milii     RETWTKL---FGVIYACLESH----------------------------
GbY_Latimeria_chalumnae     CTSWQKM---LSFTYDYLVSC----------------------------
GbY_NfoG46                  QGSWQKL---YSLVYEYLEKC----------------------------
GbY_Xenopus_laevis          CKSWEKL---FSITYNFLDSC----------------------------
GbY_Rhinella_marina         CKSWEKL---FSITYNFLGSC----------------------------
GbY_Xenopus_tropicalis      SKSWEKL---FSITYKYLESC----------------------------
Cygb-1_Astyanax_mexicanus   QGAWSKL---MEVVHWQVMQV----------------------------
Cygb-2_Alosa_sapidissima    ARAWTKL---MGALYWHVTGA----------------------------
Cygb-1_Danio_rerio          QSSWSKL---MGILYWQMNRV----------------------------
Cygb-1_Cyprinus_carpio      QSAWSKL---LGVLYWQMNKV----------------------------
Cygb-1_Cten...odon_idella   QGAWSKL---MGVLYWQMNKV----------------------------
Cygb_Neocer...us_forsteri   QRAWTKL---MGLFYCHVMAS----------------------------
Cygb-2_Anguilla_anguilla    QRAWTKL---MGLLYWHVTGA----------------------------
Cygb_Rhinat..._bivittatum   QRAWTKL---KNLIYSHVTAT----------------------------
Cygb_Camelus_dromedarius    QRAWAKL---RGLIYSHVTAA----------------------------
Cygb_Callithrix_jacchus     QRAWAKL---RGLIYSHVTAA----------------------------
Cygb_Bos_taurus             QRAWAKL---RGLIYSHVTAA----------------------------
Cygb_Homo_sapiens           QRAWAKL---RGLIYSHVTAA----------------------------
Cygb_Mus_musculus           QKAWAKL---RGLIYSHVTAA----------------------------
Cygb_Rattus_norvegicus      QKAWTKL---RGLIYSHVTAA----------------------------
Cygb_Budorcas_taxicolor     QRAWAKL---RGLIYSHVTAA----------------------------
Cygb_Bufo_bufo              -----------------------------------------------
Cygb-2_Brie...rachyistius   RRAWTKL---MGLVYWHVTGA----------------------------
Cygb-2_Danio_rerio          QTSWSKL---MAALYWHITGA----------------------------
Cygb_Chelydra_serpentina    QRAWTKM---KSLIYTHVTAA----------------------------
Cygb_Xenopus_laevis         QLAWNKL---RSLIYSQVQSA----------------------------
Cygb_Gallus_gallus          HGAWTKM---RTLIYTHVTAA----------------------------
Cygb-2_Amph...n_ocellaris   QMVWTKL---MGAVYWHVTGA----------------------------
Cygb-2_Chelmon_rostratus    QMLWAKL---MGAVYWHVTGA----------------------------
GbX_NfoG11                  EEAWESL---FKYLITMMKKG----------------------------
GbX_Latimeria_chalumnae     EEAWQTL---FKYVTEVMKRG----------------------------
```

```
GbX_Tetraod...igroviridis  EEAWKTL---FQYVTGLMRKG----------------------------
GbX_Perca_fluviatilis      EEAWKTM---FQFVTSLMKKG----------------------------
GbX_Plectro...s_leopardus  EEAWKTM---FQYLTGLMKQG----------------------------
Ngb_Mycteria_americana     REAWSKL---YGAVVKAMRRG----------------------------
Ngb_Homo_sapiens           RAAWSQL---YGAVVQAMSRG----------------------------
Ngb_Mus_musculus           RTAWSRL---YGAVVQAMSRG----------------------------
Ngb_Bufo_gargarizans       RAAWSQL---YAIVVSAMSRG----------------------------
Ngb_Xenopus_laevis         REAWSLL---YATVVQSMSRG----------------------------
Ngb_Xenopus_tropicalis     REAWSLL---YANVVQSMSRG----------------------------
GbX_Salmo_salar            EEAWKTL---FLYLTRLMNRG----------------------------
Ngb_Danio_rerio            RQAWLTM---YSIVVSAMTRG----------------------------
Ngb_Dissostichus_mawsoni   RQAWLNL---YSIVVAAMSQG----------------------------
Ngb_Oncorhy...tshawytscha  RQAWLNM---YTIVVAAMSRG----------------------------
Ngb_Oncorhy...s_gorbuscha  RQAWLNM---YTIVVAAMSRG----------------------------
Ngb_Embiotoca_jacksoni     RQAWLNM---YSIVVAAMSRG----------------------------
Ngb_Neoceratodus_forsteri  KAAWIVL---YTFVTKTMIRG----------------------------
Ngb_Latimeria_chalumnae    RHAWIAL---YSFVVETMARG----------------------------
GbX_Lepisosteus_oculatus   EEAWQTL---FLYVTRIMKRG----------------------------
GbX_Danio_rerio            EEAWKTL---FQYVTSIMREG----------------------------
GbX_Carassius_auratus      EEAWKTL---FQYVTSIMREG----------------------------

Hb3_NfoG112                ------------------------------------------------
Mb_Anguilla_anguilla       ----------------------------------YKE-FGF--------
Mb_Denticeps_clupeoides    ----------------------------------YKE-LGF--------
Mb_Epinephe...lanceolatus  ----------------------------------YKE-LGF--------
Hb2_NfoG111                KCNYFSKLCDTHKLRINMSKTKILIFGNTKNILNHY--------------
Hb1_NfoG110                ----------------------------------PTD-PYW--------
Hb_beta1_Danio_rerio       ----------------------------------Y--------------
Hb-beta-1_C...odon_idella  ----------------------------------Y--------------
Hb_beta2_Mus_musculus      ----------------------------------Y--------------
Hb-beta_Carassius_gibelio  ----------------------------------Y--------------
Hb-beta-2_C...odon_idella  ----------------------------------Y--------------
Hb-beta-2_C..._adamanteus  ----------------------------------Y--------------
Hb_beta1_Xenopus_laevis    ----------------------------------Y--------------
Hb_beta_Lat...a_chalumnae  ----------------------------------Y--------------
GbE1_NfoG2                 ------------------------------------------------
Mb2_NfoG14                 ----------------------------------YNQ-INQ--------
Hb_alpha_La...a_chalumnae  ----------------------------------YR-------------
Hb-alpha_Ca...ius_gibelio  ----------------------------------YR-------------
Hb_alpha1_Homo_sapiens     ----------------------------------YR-------------
Hb-alpha_Acinonyx_jubatus  ----------------------------------YR-------------
```

```
Mb3_NfoG5                 ----------------------------------YKE-V----------
Mb4_NfoG6                 ----------------------------------YKE-V----------
Mb6_NfoG4                 ----------------------------------YKQ-IEH--------
Mb1_NfoG3                 ----------------------------------YQE-V----------
Mb5_NfoG15                ----------------------------------YKK-V----------
Mb_Alligator_sinensis     ----------------------------------YKE-FGY--------
Mb_Mus_musculus           ----------------------------------YKE-LGF--------
Mb_Delphinapterus_leucas  ----------------------------------YKE-LGF--------
Mb_Balaenoptera_borealis  ----------------------------------YKE-LGF--------
Mb_Homo_sapiens           ----------------------------------YKE-LGF--------
Mb_Zootoca_vivipara       ----------------------------------YKE-HGF--------
Mb_Cricetulus_griseus     ----------------------------------YKE-LGF--------
Mb_Gallus_gallus          ----------------------------------YKE-FGF--------
Mb_Crotalus_tigris        ----------------------------------YKE-LGF--------
GbE_Chelydra_serpentina   ----------------------------------YKE-AGW--------
GbE_Chrysem...icta_bellii ----------------------------------YKE-AGW--------
GbE3_NfoG13               ----------------------------------YKE-AGW--------
GbE_Gallus_gallus         ----------------------------------YKE-AGW--------
GbE_Patagioenas_fasciata  ----------------------------------YKE-AGW--------
GbE2_NfoG12               ----------------------------------YRE-AGW--------
GBE_Latimeria_chalumnae   ----------------------------------YKE-AGW--------
GbE4_NfoG57               ----------------------------------YEE-VGQ--------
GbY_Callorhinchus_milii   ----------------------------------YKD-AGF--------
GbY_Latimeria_chalumnae   ----------------------------------YHKSS----------
GbY_NfoG46                ----------------------------------YHK-VDVF-------
GbY_Xenopus_laevis        ----------------------------------YTK-S----------
GbY_Rhinella_marina       ----------------------------------YTK-S----------
GbY_Xenopus_tropicalis    ----------------------------------YAN-T----------
Cygb-1_Astyanax_mexicanus ----------------------------------YSE-IGWTSA-----
Cygb-2_Alosa_sapidissima  ----------------------------------YAE-VGWVKV-----
Cygb-1_Danio_rerio        ----------------------------------YAE-VGWE-------
Cygb-1_Cyprinus_carpio    ----------------------------------YAE-VGWDSI-----
Cygb-1_Cten...odon_idella ----------------------------------YAE-VGWESI-----
Cygb_Neocer...us_forsteri ----------------------------------YKE-VGWVQY-----
Cygb-2_Anguilla_anguilla  ----------------------------------YAE-VGWVQL-----
Cygb_Rhinat..._bivittatum ----------------------------------YKE-VGW--------
Cygb_Camelus_dromedarius  ----------------------------------YKE-VGWVQQVPN--
Cygb_Callithrix_jacchus   ----------------------------------YKE-VGWVQQVPNAT
Cygb_Bos_taurus           ----------------------------------YKE-VGWVQQVPNAT
Cygb_Homo_sapiens         ----------------------------------YKE-VGWVQQVPNAT
Cygb_Mus_musculus         ----------------------------------YKE-VGWVQQVPNTT
```

```
Cygb_Rattus_norvegicus     ----------------------------------YKE-VGWVQQVPNTT
Cygb_Budorcas_taxicolor    ----------------------------------YKE-VGWVQQVPNAT
Cygb_Bufo_bufo             --------------------------------------------------
Cygb-2_Brie...rachyistius  ----------------------------------YTE-SGWVQL-----
Cygb-2_Danio_rerio         ----------------------------------YTE-VGWVKL-----
Cygb_Chelydra_serpentina   ----------------------------------YKE-VGWVQY-----
Cygb_Xenopus_laevis        ----------------------------------YKE-VGWTQY-----
Cygb_Gallus_gallus         ----------------------------------YKE-AGWVSY-----
Cygb-2_Amph...n_ocellaris  ----------------------------------YTE-VGWLQV-----
Cygb-2_Chelmon_rostratus   ----------------------------------YTE-VGWLQV-----
GbX_NfoG11                 ----------------------------------FQE-EEKKHPI-NKS
GbX_Latimeria_chalumnae    ----------------------------------YLQ-EEKIKRT-MET
GbX_Tetraod...igroviridis  ----------------------------------HQE-EGSRQRH--LA
GbX_Perca_fluviatilis      ----------------------------------YEE-ESNRQRH--LA
GbX_Plectro...s_leopardus  ----------------------------------YEE-ESNRQRH--LA
Ngb_Mycteria_americana     ----------------------------------WET-L----------
Ngb_Homo_sapiens           ----------------------------------WD-------------
Ngb_Mus_musculus           ----------------------------------WD-------------
Ngb_Bufo_gargarizans       ----------------------------------WWR------------
Ngb_Xenopus_laevis         ----------------------------------WHR-DSQE-------
Ngb_Xenopus_tropicalis     ----------------------------------WHR-DSQE-------
GbX_Salmo_salar            ----------------------------------YQE-RNRHQQV---G
Ngb_Danio_rerio            ----------------------------------WAK-NGE--------
Ngb_Dissostichus_mawsoni   ----------------------------------WAK-NGE--------
Ngb_Oncorhy...tshawytscha  ----------------------------------WAK-KGE--------
Ngb_Oncorhy...s_gorbuscha  ----------------------------------WAK-NGE--------
Ngb_Embiotoca_jacksoni     ----------------------------------WAK-NGE--------
Ngb_Neoceratodus_forsteri  ----------------------------------WDK-NGE--------
Ngb_Latimeria_chalumnae    ----------------------------------WGA-NGE--------
GbX_Lepisosteus_oculatus   ----------------------------------YQE-EEKSKRN-NVV
GbX_Danio_rerio            ----------------------------------FLE-EERNKRS-NTQ
GbX_Carassius_auratus      ----------------------------------FLE-EERNKRS-NTQ

Hb3_NfoG112                -----------------T
Mb_Anguilla_anguilla       ----------------QG
Mb_Denticeps_clupeoides    ----------------AG
Mb_Epinephe...lanceolatus  ----------------SG
Hb2_NfoG111                -----------------G
Hb1_NfoG110                ---------------QRG
Hb_beta1_Danio_rerio       -----------------H
Hb-beta-1_C...odon_idella  -----------------H
```

```
Hb_beta2_Mus_musculus      ----------------H
Hb-beta_Carassius_gibelio  ----------------H
Hb-beta-2_C...odon_idella  ----------------H
Hb-beta-2_C..._adamanteus  ----------------H
Hb_beta1_Xenopus_laevis    ----------------H
Hb_beta_Lat...a_chalumnae  ----------------H
GbE1_NfoG2                 ----------------E
Mb2_NfoG14                 ----------------SG
Hb_alpha_La...a_chalumnae  ------------------
Hb-alpha_Ca...ius_gibelio  ------------------
Hb_alpha1_Homo_sapiens     ------------------
Hb-alpha_Acinonyx_jubatus  ------------------
Mb3_NfoG5                  ----------------QK
Mb4_NfoG6                  ----------------QK
Mb6_NfoG4                  ----------------KA
Mb1_NfoG3                  ----------------EK
Mb5_NfoG15                 ----------------QK
Mb_Alligator_sinensis      ----------------QG
Mb_Mus_musculus            ----------------QG
Mb_Delphinapterus_leucas   ----------------HG
Mb_Balaenoptera_borealis   ----------------QG
Mb_Homo_sapiens            ----------------QG
Mb_Zootoca_vivipara        ----------------QG
Mb_Cricetulus_griseus      ----------------QG
Mb_Gallus_gallus           ----------------QG
Mb_Crotalus_tigris         ----------------QG
GbE_Chelydra_serpentina    ------------------
GbE_Chrysem...icta_bellii  ------------------
GbE3_NfoG13                ------------------
GbE_Gallus_gallus          ------------------
GbE_Patagioenas_fasciata   ------------------
GbE2_NfoG12                ------------------
GBE_Latimeria_chalumnae    ------------------
GbE4_NfoG57                ----------------HV
GbY_Callorhinchus_milii    ----------------YP
GbY_Latimeria_chalumnae    ----------------GP
GbY_NfoG46                 ---------------NFK
GbY_Xenopus_laevis         ----------------DS
GbY_Rhinella_marina        ----------------DS
GbY_Xenopus_tropicalis     ----------------DS
Cygb-1_Astyanax_mexicanus  ------------NSATE
```

```
Cygb-2_Alosa_sapidissima   ------------SSSAV
Cygb-1_Danio_rerio         -------------NSKK
Cygb-1_Cyprinus_carpio     ------------KNSAK
Cygb-1_Cten...odon_idella  ------------KNSAK
Cygb_Neocer...us_forsteri  ------------PNSTM
Cygb-2_Anguilla_anguilla   ------------SSSAV
Cygb_Rhinat..._bivittatum  ----------------AQ
Cygb_Camelus_dromedarius   --------------ATT
Cygb_Callithrix_jacchus    TPP-------ATLPSSGP
Cygb_Bos_taurus            TPP-------ATLPSSGP
Cygb_Homo_sapiens          TPP-------ATLPSSGP
Cygb_Mus_musculus          TPP-------ATLPSSGP
Cygb_Rattus_norvegicus     TLP-------ATLPSSGP
Cygb_Budorcas_taxicolor    TPP-------ATLPSSGP
Cygb_Bufo_bufo             ------------------
Cygb-2_Brie...rachyistius  ------------SSSAV
Cygb-2_Danio_rerio         ------------SSSAV
Cygb_Chelydra_serpentina   ------------PNSTM
Cygb_Xenopus_laevis        ------------PSNSV
Cygb_Gallus_gallus         ------------PSATL
Cygb-2_Amph...n_ocellaris  ------------SSSAV
Cygb-2_Chelmon_rostratus   ------------SSSAV
GbX_NfoG11                 VYPKKQNSSGMSSVQNSV
GbX_Latimeria_chalumnae    VEPKQA----SRRANNAI
GbX_Tetraod...igroviridis  LPPKDG----PEKRTSAL
GbX_Perca_fluviatilis      LSPKER----PDKKNTAL
GbX_Plectro...s_leopardus  LSPKER----PDKRNTAL
Ngb_Mycteria_americana     -------------PEGD
Ngb_Homo_sapiens           ----------------GE
Ngb_Mus_musculus           ----------------GE
Ngb_Bufo_gargarizans       ----------------DP
Ngb_Xenopus_laevis         ------------QQEGI
Ngb_Xenopus_tropicalis     ------------QREGI
GbX_Salmo_salar            A----S----PKERNIAL
Ngb_Danio_rerio            -------------HKSN
Ngb_Dissostichus_mawsoni   -------------DKAD
Ngb_Oncorhy...tshawytscha  -------------HKTD
Ngb_Oncorhy...s_gorbuscha  -------------HKTD
Ngb_Embiotoca_jacksoni     -------------DKAD
Ngb_Neoceratodus_forsteri  -------------NKIE
Ngb_Latimeria_chalumnae    -------------NKLN
```

```
GbX_Lepisosteus_oculatus   IASKER----PEKKGTAI
GbX_Danio_rerio            TSSRER----PDKRSTAI
GbX_Carassius_auratus      TSSRER----PDKRSTAI
//
```

Alignment Fig. 11

```
GbX_Latimeria_chalumnae    MGCAISGLSW---------RAVKGPGEEGKGK-EG-EAE-----ALPAIS
GbX_Tetraod...igroviridis  MGCAISSLGA---------EA--EFGD---RS-AE-EEDAAAAAAVVYPR
GbX_Perca_fluviatilis      MGCAISGLAA---------KA--EFGD---RT-TE-EAA-----GAAYPS
GbX_Plectro...s_leopardus  MGCAISGLAA---------KA--EFGE---RT-TE-DA------AAVFPS
Ngb_Mycteria_americana     MES----------------------------------------GLLLS
Ngb_Homo_sapiens           M------------------------------------------------E
Ngb_Mus_musculus           M------------------------------------------------E
Ngb_Bufo_gargarizans       MES----------------------------------------LQLP
Ngb_Xenopus_laevis         MEK----------------------------------------DQLS
Ngb_Xenopus_tropicalis     MEK----------------------------------------DQLS
GbX_Salmo_salar            MGCAISVLAP----------K--TAGE---C---K-TED-----AVVHLS
Ngb_Danio_rerio            ME-----------------------------------------KLS
Ngb_Dissostichus_mawsoni   ME-----------------------------------------KLS
Ngb_Oncorhy...tshawytscha  ME-----------------------------------------KLT
Ngb_Oncorhy...s_gorbuscha  ME-----------------------------------------KLT
Ngb_Embiotoca_jacksoni     ME-----------------------------------------KLS
Ngb_Neoceratodus_forsteri  MEY----------------------------------------PSLS
Ngb_Latimeria_chalumnae    MEK----------------------------------------HNFS
GbX_Lepisosteus_oculatus   MGCAISGLGL---------AP----KQ---ID-AT-EEE-----ALPHLS
GbX_Danio_rerio            MGCAISGSGL---------TA--GAPE---IR-PG-EEE-----TPAGLT
GbX_Carassius_auratus      MGCAISGSGL---------TA--RAPE---IR-AG-EEE-----TPAGLT
Mb_Anguilla_anguilla       -------------------------------------------------
Mb_Denticeps_clupeoides    -------------------------------------------------
Mb_Epinephe...lanceolatus  -------------------------------------------------
Hb_beta1_Danio_rerio       MV-----------------------------------------EWT
Hb-beta-1_C...odon_idella  MV-----------------------------------------EWT
Hb_beta2_Mus_musculus      MV-----------------------------------------HLT
Hb-beta_Carassius_gibelio  MV-----------------------------------------EWT
Hb-beta-2_C...odon_idella  MV-----------------------------------------EWT
Hb-beta-2_C..._adamanteus  MV-----------------------------------------HWT
Hb_beta1_Xenopus_laevis    M------------------------------------------GLT
Hb_beta_Lat...a_chalumnae  V------------------------------------------HWT
GbE_Chelydra_serpentina    M------------------------------------------AFS
```

```
GbE_Chrysem...icta_bellii  M-------------------------------------------AFS
GBE_Latimeria_chalumnae    M-------------------------------------------ALS
GbE_Gallus_gallus          M-------------------------------------------SFS
GbE_Patagioenas_fasciata   M-------------------------------------------SFS
Cygb-2_Amph...n_ocellaris  MSRRESPPPPSPPPP--LLGL--QRGE---VQ-CEERPE-----RAEPLS
Cygb-2_Chelmon_rostratus   MSRRESPPPPSPPPPPQLLGV--RRGE---VE-CEDRPE-----RAEPLS
Cygb-1_Astyanax_mexicanus  MEG----------------------N---GGMQE-QME-----RPEMLT
Cygb-2_Alosa_sapidissima   MDS----------------------G---RD-EE-PRE-----RLEPLS
Cygb-1_Danio_rerio         MEG----------------------D---GG-VQ-LTQ-----SPDSLT
Cygb-1_Cyprinus_carpio     MEG----------------------D---GD-VE-LTQ-----SPETLT
Cygb-1_Cten...odon_idella  MEG----------------------D---RG-VE-LTQ-----SPETLT
Cygb_Neocer...us_forsteri  MEQVK------------------QGE---MQ-ME-KRE-----RGDQLL
Cygb_Rhinat..._bivittatum  MEKV-------------------QGE---ME-MD-RWE-----RSEEIS
Cygb_Camelus_dromedarius   MEKV-------------------PGE---ME-IE-RRE-----RSEELS
Cygb_Callithrix_jacchus    MEKV-------------------AGE---ME-IE-RRE-----RSEELS
Cygb_Bos_taurus            MEKV-------------------PGE---ME-IE-RRE-----RSEELS
Cygb_Homo_sapiens          MEKV-------------------PGE---ME-IE-RRE-----RSEELS
Cygb_Mus_musculus          MEKV-------------------PGD---ME-IE-RRE-----RSEELS
Cygb_Rattus_norvegicus     MEKV-------------------PGD---ME-IE-RRE-----RNEELS
Cygb_Budorcas_taxicolor    MEKV-------------------PGE---ME-IE-RRE-----RSEELS
Cygb_Bufo_bufo             MEKV-------------------QEE---NE-TE-RWE-----RPEEIT
Cygb-2_Danio_rerio         MEKE-------------------RED---EE-TE-GRE-----RPEPLT
Cygb_Chelydra_serpentina   MEKV-------------------QGE---ME-IE-RWE-----RSEELS
Cygb_Xenopus_laevis        MEKV-------------------QGE---ND-ME-RWE-----RLEDIT
Cygb_Gallus_gallus         MEKV-------------------QGE---ME-IE-RWE-----RSEEIS
GbY_Callorhinchus_milii    MT-------------------------------------------GIT
GbY_Latimeria_chalumnae    MA-------------------------------------------ALT
GbY_Xenopus_laevis         MA-------------------------------------------DLT
GbY_Rhinella_marina        MA-------------------------------------------DLT
GbY_Xenopus_tropicalis     MA-------------------------------------------DLT
Mb_Alligator_sinensis      M--------------------------------------------QLS
Mb_Mus_musculus            M--------------------------------------------GLS
Mb_Delphinapterus_leucas   M--------------------------------------------GLS
Mb_Balaenoptera_borealis   M--------------------------------------------VLS
Mb_Homo_sapiens            M--------------------------------------------GLS
Mb_Zootoca_vivipara        M--------------------------------------------GLS
Mb_Cricetulus_griseus      M--------------------------------------------GLS
Mb_Gallus_gallus           M--------------------------------------------GLS
Mb_Crotalus_tigris         M--------------------------------------------GLT
Hb_alpha_La...a_chalumnae  ---------------------------------------------GLT
Hb-alpha_Ca...ius_gibelio  M--------------------------------------------SLS
```

```
Hb_alpha1_Homo_sapiens     M---------------------------------------------VLS
Hb-alpha_Acinonyx_jubatus  M---------------------------------------------VLS

GbX_Latimeria_chalumnae    QLQIHLIQESWKLIQEDIAKVGIIMFVRLFETHPECKDVFFL-FRDVDDL
GbX_Tetraod...igroviridis  EDQIQMIKDSWKVIRDDIAKVGIIMFVRLFETHPECKDVFFL-FRDVEDL
GbX_Perca_fluviatilis      EDQIQMIKESWKVIQDDIAKVGIIMFVRLFETHPECKDVFFL-FRDVEDL
GbX_Plectro...s_leopardus  EDQIQMIKESWKVIRDDIAKVGIIMFVRLFETHPECKDVFFL-FRDVEDL
Ngb_Mycteria_americana     GRQRALIRESWQRVSGNPVQHGLVLFTRLFDLDPDLLPLFQYNCKQFASP
Ngb_Homo_sapiens           RPEPELIRQSWRAVSRSPLEHGTVLFARLFALEPDLLPLFQYNCRQFSSP
Ngb_Mus_musculus           RPESELIRQSWRVVSRSPLEHGTVLFARLFALEPSLLPLFQYNGRQFSSP
Ngb_Bufo_gargarizans       EPQKELIRDTWRIVGQEQIHHGTVLFTRLFQLEPELLFLFHYNSAHFSSV
Ngb_Xenopus_laevis         GPQKELIRESWQTVSQDQLHHGTVLFSRLFELEPELIFFFQYNSSHFSEV
Ngb_Xenopus_tropicalis     GPQKELIRESWQTVSQDQLHHGTVLFSRLFELEPELVFLFQYNSSHFSKV
GbX_Salmo_salar            KEQIEMIKESWKVIEEDIAKVGIIMFVRLFETHPECKDVFFL-FRDVEDR
Ngb_Danio_rerio            EKDKGLIRDSWESLGKNKVPHGIVLFTRLFELDPALLTLFSY-STNCGDA
Ngb_Dissostichus_mawsoni   EKDKELIRGSWESLGKNKVPHGVVMFSRLFELDPELLTLFHY-TTNCGST
Ngb_Oncorhy...tshawytscha  EKDKELIRGSWESLGKNKVPHGVVMFSRLFELEPALLNLFHY-NTNCSPT
Ngb_Oncorhy...s_gorbuscha  EKDKELIRGSWESLGKNKVPHGVVMFSRLFELEPALLNLFHY-NTNCSPT
Ngb_Embiotoca_jacksoni     GKDKELIRGSWESLGKNKVPHGVVMFSRLFELDPALLSLFHY-STNCGST
Ngb_Neoceratodus_forsteri  MREKELIRESWGQLGKEKLQHGTVLFSRLFELDPDLVHLFHY-NTTFATP
Ngb_Latimeria_chalumnae    VRSKELIRESWDRLGKNKLPHGTVMFTRLFELDPDLMHLFNY-NSSFPSQ
GbX_Lepisosteus_oculatus   DHHIEQIKDSWKVIQEDIAKVGIIMFVRLFETHPECKDVFFL-FRDVEDL
GbX_Danio_rerio            TNHIRLIKESWRLIQEDIAKVGIIMFVRLFETHPECKDVFFL-FRDVEDL
GbX_Carassius_auratus      ANHIRLIKESWRLTQEDIAKVGIIMFVRLFETHPECKDVFFL-FRDVEDL
Mb_Anguilla_anguilla       MTDFELVLKAWKPIEADLKGNGGVVLTRLFQEHPETQQLFPK-FAAIAP-
Mb_Denticeps_clupeoides    MADFDMVLNCWGKVEADYTGTGGEVLTRLFKEHPETLKLFPK-FAGIAQ-
Mb_Epinephe...lanceolatus  MADYDAVLKCWSAMEADYSGIGNLVMRRLFKEFPQSLNHFPK-FDGMSQ-
Hb_beta1_Danio_rerio       DAERTAILGLWGKLNI--DEIGPQALSRCLIVYPWTQRYFAT-FGNLSSP
Hb-beta-1_C...odon_idella  DAERTAILGLWGKLNI--DEIGPQALSRCLIVYPWTQRYFAT-FGNLSSP
Hb_beta2_Mus_musculus      DAEKSAVSCLWAKVNP--DEVGGEALGRLLVVYPWTQRYFDS-FGDLSSA
Hb-beta_Carassius_gibelio  DAERKAIKDVWGKLKG--DIIGPQALARCLIVYPWTQRYFGS-FGSLSDA
Hb-beta-2_C...odon_idella  EFERTTIQDIFSKMNY--EVVGPQALARCLIVYPWTQRYFGK-FGNLYNA
Hb-beta-2_C..._adamanteus  PEEKQLITALWGKVDV--PEVGAATLGKLLVVYPWTQRFFAH-FGNLSGP
Hb_beta1_Xenopus_laevis    AHDRQLINSTWGKLCA--KTIGQEALGRLLWTYPWTQRYFSS-FGNLNSA
Hb_beta_Lat...a_chalumnae  ETERATIETVYQKLHL--DEVGREALTRLFIVYPWTTRYFKS-FGDLSSS
GbE_Chelydra_serpentina    EAEVQRARGAWEKMYANAEDNGTTVLVRMFTEHPDTKSYFTH-FKGMGTA
GbE_Chrysem...icta_bellii  EAEVQRARGAWEKMYANAEDNGTTVLVRMFTEHPDTKSYFTH-FKGMGTA
GBE_Latimeria_chalumnae    DAEVQTARDVWGQIYANAEENGTIILVRMFTEHPDTKSYFGN-FKGMGSA
GbE_Gallus_gallus          EAEVQSARGAWEKMYVDAEDNGTAVLVRMFTEHPDTKSYFTH-FKGMDSA
GbE_Patagioenas_fasciata   EAEVQSARGAWEKMYADAEDNGTTVLVRMFTEHPDTKSYFTH-FKGMDSA
Cygb-2_Amph...n_ocellaris  DAEREIIQDSWGHVYKNCEDVGVSVLIRFFVNFPSAKQYFSQ-FQDMEDP
```

```
Cygb-2_Chelmon_rostratus    DAEREIIQDTWGHVYKNCEDVGVSVLIRFFVNFPSAKQYFSQ-FQDMEDP
Cygb-1_Astyanax_mexicanus   DGEQAIIQNTWTKVFQNKEAAGVAVLMRLFTSFPSSKQFFSQ-FRDTEDP
Cygb-2_Alosa_sapidissima    DTELAIIQDTWARVYENCEDVGVTILIRFFVNFPSAKQYFSQ-FQDMEDP
Cygb-1_Danio_rerio          EEDVCVIQDTWKPVYAERDNAGVAVLVRFFTNFPSAKQYFEH-FRELQDP
Cygb-1_Cyprinus_carpio      EDDVCVIQDTWRPVYENRENAGVAVLIRFFSNFPSAKQYFDQ-FRDMHDP
Cygb-1_Cten...odon_idella   EDDVCVIQDTWRPVYENRESAGVAVLIRFFGNFPSAKQYFSQ-FRDIEDP
Cygb_Neocer...us_forsteri   DSEKMMIRESWARVFANCEDVGVSILIRFFVNFPSAKQYFSQ-FRHLEDP
Cygb_Rhinat..._bivittatum   DAEKRVIQETWSRVYTNCEDVGVSILIRFFVNFPSAKQYFSQ-FKHMEDP
Cygb_Camelus_dromedarius    EAERKAVQATWARLYANCEDVGVAILVRFFVNFPSAKQYFSQ-FKHMEEP
Cygb_Callithrix_jacchus     EAERKAVQATWARLYAHCEDVGVAILVRFFVNFPSAKQYFSQ-FKHMEDP
Cygb_Bos_taurus             EAERKAVQATWARLYANCEDVGVAILVRFFVNFPSAKQYFSQ-FKHMEEP
Cygb_Homo_sapiens           EAERKAVQAMWARLYANCEDVGVAILVRFFVNFPSAKQYFSQ-FKHMEDP
Cygb_Mus_musculus           EAERKAVQATWARLYANCEDVGVAILVRFFVNFPSAKQYFSQ-FRHMEDP
Cygb_Rattus_norvegicus      EAERKAVQATWARLYANCEDVGVAILVRFFVNFPSAKQYFSQ-FKHMEDP
Cygb_Budorcas_taxicolor     EAERKAVQATWARLYANCEDVGVAILVRFFVNFPSAKQYFSQ-FKHMEEP
Cygb_Bufo_bufo              ESERDIIKETWARLYTNCEDVGVSILIRFFVNFPKAKQYFSQ-FKDMEDP
Cygb-2_Danio_rerio          DVERGIIKDTWARVYASCEDVGVTILIRFFVNFPSAKQYFSQ-FQDMEDP
Cygb_Chelydra_serpentina    DAERKVIQETWSRVYTNCEDVGVSILIRFFVKFPSAKQYFSQ-FKHMEDP
Cygb_Xenopus_laevis         ESERGVIKETWARVYTNCEDVGVSILIRFFVNFPSAKQHFSQ-FKHMEDP
Cygb_Gallus_gallus          DAEKKVIQETWSRVYANCEDVGVSILIRFFVNFPSAKQYFSQ-FKHMDDT
GbY_Callorhinchus_milii     EADKENIHFIWEKLYENPEENGKTIVLRMFTDYPETKMYFQH-FKNISTL
GbY_Latimeria_chalumnae     EADKQNIRGIWKTVFENAEENGRTIVIRLFEKYPETKVYFKN-FKNISTM
GbY_Xenopus_laevis          AADIENINEIWCKIYANPEESGKTVVIRLFTTYPQTKVYFKN-LKNIATL
GbY_Rhinella_marina         AADIENINEIWCKIYANPEESGKTVVIRLFTTYPQTKVYFKN-LKNIATL
GbY_Xenopus_tropicalis      GADIENINEVWSKIYANPEESGRTVVISLFLTYPQTKIYFKN-LKNISTL
Mb_Alligator_sinensis       DQEWKHVLDIWAKVESKLPEHGHEVIIRLLQEHPETQERFEK-FKHMKTA
Mb_Mus_musculus             DGEWQLVLNVWGKVEADLAGHGQEVLIGLFKTHPETLDKFDK-FKNLKSE
Mb_Delphinapterus_leucas    EGEWQLVLNVWGKVEADLAGHGQDVLIRLFKGHPETLEKFDK-FKHLKTE
Mb_Balaenoptera_borealis    DAEWQLVLNIWAKVEADVAGHGQDILIRLFKGHPETLEKFDK-FKHLKTE
Mb_Homo_sapiens             DGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDK-FKHLKSE
Mb_Zootoca_vivipara         DQEWQKVVDIWGKVEPDLPAHGQEVIIRMFQKHPETQERFDK-FKNLKTV
Mb_Cricetulus_griseus       DGEWQKVLNVWGKVEADLAGHGQEVLIRLFKNHPETLDKFDK-FKNLKSE
Mb_Gallus_gallus            DQEWQQVLTIWGKVEADIAGHGHEVLMRLFHDHPETLDRFDK-FKGLKTP
Mb_Crotalus_tigris          EQEWQKVMDIWEKVEPDLSLHGQEVIIRMFQNHPETQERFVK-FKNFKTL
Hb_alpha_La...a_chalumnae   AADKTLIKSIWGKVEKETEAIGVEALVRLFKCFPQSKVYFDH-FTDLS--
Hb-alpha_Ca...ius_gibelio   DKDKAVVKALWAKIGSRADEIGAEALGRMLTVYPQTKTYFSH-WSDLS--
Hb_alpha1_Homo_sapiens      PADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPH-FDL-S--
Hb-alpha_Acinonyx_jubatus   AADKSNVKACWGKIGSHAGEYGAEALERTFCSFPTTKTYFPH-FDL-S--

GbX_Latimeria_chalumnae     QRLRTSKELRAHGLRVMSFVEKSVARLEHSERLEQLALELGKSHYR-YNA
GbX_Tetraod...igroviridis   ERLRSSRELRAHGLRVMSFIEKSVARLDQQDRLEALAVELGKSHYH-YNA
```

```
GbX_Perca_fluviatilis       ERLRTSRELRAHGLRVMSFIEKSVARLDQLERLEALALELGKSHYH-YNA
GbX_Plectro...s_leopardus   ERLRTSRELRAHGLRVMSFIEKSVARLDQLERLEALALELGKSHYH-YNA
Ngb_Mycteria_americana      QECLSTPEFLDHIRKVMLVIDAAVSHLENLSCLEEYLCNLGKKHQA-VGV
Ngb_Homo_sapiens            EDCLSSPEFLDHIRKVMLVIDAAVTNVEDLSSLEEYLASLGRKHRA-VGV
Ngb_Mus_musculus            EDCLSSPEFLDHIRKVMLVIDAAVTNVEDLSSLEEYLTSLGRKHRA-VGV
Ngb_Bufo_gargarizans        QECLSSSEFTEHIRKVMTVIDAAVRSLDSLPSLEDYLTSLGRKHRA-TGV
Ngb_Xenopus_laevis          QDCLSSAEFTEHIRKVMTVIDAAVSSLDCLASLDEYLTSLGRKHKA-VGV
Ngb_Xenopus_tropicalis      QDCLSSAEFTEHIRKVMTVIDAAVSSLDCLSSLDEYLTSLGRKHRA-VGV
GbX_Salmo_salar             DRLRISRELRAHGLRVMSFIEKSVARLEQLERLEVLAVELGRSHYP-YNA
Ngb_Danio_rerio             PECLSSPEFLEHVTKVMLVIDAAVSHLDDLHTLEDFLLNLGRKHQA-VGV
Ngb_Dissostichus_mawsoni    QDCLSSPEFLEHVTKVMLVIDAAVSNLDDLPSLEDFLLNLGRKHQA-VGV
Ngb_Oncorhy...tshawytscha   QDCLSSPEFLDHVTKVMLVIDAAVSHLDDLHTLEDFLLSLGKKHQA-VGV
Ngb_Oncorhy...s_gorbuscha   QDCLSSPEFLDHVTKVMLVIDAAVSHLDNLHTLEDFLLNLGKKHQA-VGV
Ngb_Embiotoca_jacksoni      QDCLSSPEFLDHVTKVMLVIDAAVSHLDDLHSLEDFLLNLGRKHQA-VGV
Ngb_Neoceratodus_forsteri   AECLTSPEFLDHIHKVMSVIDAAVNNLENLPSLEEYLINLGKKHKT-VGV
Ngb_Latimeria_chalumnae     AGCLTSPEFIDHIHKVMMVVDAAVNSLDNLLSLEDYLINLGKKHQA-AGV
GbX_Lepisosteus_oculatus    ERLRASKELRAHGLKVMSFIEKTVARLDQLDRLDQLALELGKSHYR-YNA
GbX_Danio_rerio             ERLRTSRELRAHGLRVMSFIEKSVARLDQLERLETLALELGKSHYR-YNA
GbX_Carassius_auratus       ERLRTSRELRAHGLRVMSFIEKSVARLDQLERLETLALELGKSHYR-YNA
Mb_Anguilla_anguilla        GDLAGNAAISEHGCTVLTKLGDLLHAKGNHA---DILKPLAKTHATQHKI
Mb_Denticeps_clupeoides     GDLAGNAAVCDHGATVLKKLGELLKAKGKHA---DILKPLATRHANTHKI
Mb_Epinephe...lanceolatus   ADLASNGDLTAQGVTVLTKLGEVLKTRGNHG---EIIKNLSKSHANQHKV
Hb_beta1_Danio_rerio        AAIMGNPKVAAHGRTVMGGLERAIKNMDNVK---NTYAALSVMHSEKLHV
Hb-beta-1_C...odon_idella   AAIIGNPKVAAHGKTVMGGLERAIKNLDNIK---ATYSALSVMHSEKLHV
Hb_beta2_Mus_musculus       SAIMGNPKVKAHGKKVITAFNEGLKNLDNLK---GTFASLSELHCDKLHV
Hb-beta_Carassius_gibelio   AAIMANKKVAAHGKVVLNGLGRAVQSMDDIK---NTYAPLSVLHSEKLHV
Hb-beta-2_C...odon_idella   AAILGNPMVAAHGTVVLHGLDRAVKNMDNIK---ATYAELSVLHSEKLHV
Hb-beta-2_C..._adamanteus   SAICANPLVRAHGKKVMNSFAEAVKHMDNIK---ETFSKLSELHCEKLHV
Hb_beta1_Xenopus_laevis     DAVFHNEAVAAHGEKVVTSIGEAIKHMDDIK---GYYAQLSKYHSETLHV
Hb_beta_Lat...a_chalumnae   KAIASNPKVTEHGLKVMNKLTEAIHNLDHIK---DLFHKLSEKHFHELHV
GbE_Chelydra_serpentina     EEMEQSDQVRSHGKRVLTTINDLVQHLDSTDAFLGIVNPLGKKHAMQLKV
GbE_Chrysem...icta_bellii   EEMEQSDQVRSHGKRVLTTINDLVQHLDSTDAFLGIVNPLGKKHAMQLKV
GBE_Latimeria_chalumnae     AEMEQSAQVRTHGKKIFSALNDMIQHLDSTDALLGVVNPLGKKHATQLKV
GbE_Gallus_gallus           EEMKQSDQVRGHGKRVFTAINDMVQHLDNTEAFLGILNPLGQKHATQLKI
GbE_Patagioenas_fasciata    EEMKQSDQVRGHGKKVFTAINDMVQHLDNSEAFLGIVNPLGKKHATQLKI
Cygb-2_Amph...n_ocellaris   EEMERSSQLRHHARRVMNAINTVVENLHDPEKVSSVLALVGKAHAIKHQV
Cygb-2_Chelmon_rostratus    EEMERSSQLRHHARRVMNAINTVVENLHDPEKVSSVLALVGKAHAIKHKV
Cygb-1_Astyanax_mexicanus   EELKSSVQLKKHALRVMSALNTLVENVNDEEKTAAVLKVVAKSHAIKHSV
Cygb-2_Alosa_sapidissima    EEMEKSSQLRQHARRVMNAINSVVENLSDADKVSSVLSLVGKAHALKHKV
Cygb-1_Danio_rerio          AEMQQNAQLKKHGQRVLNALNTLVENLRDADKLNTIFNQMGKSHALRHKV
Cygb-1_Cyprinus_carpio      EEMKQSVQLKKHALRVMTALNTLVENLRDGDKLNTIFQQMGKSHALKHKV
```

```
Cygb-1_Cten...odon_idella  EEMKQSVQLKKHALRVMTALNTLVENLRDGDNLNTIFNQMGKSHALRHKV
Cygb_Neocer...us_forsteri  LEMERSIQLRKHARRIMNAINTVVENLGDPEKVSTVVALVGKAHALKHKV
Cygb_Rhinat..._bivittatum  LEMERSLQLRKHARRVMGAVNNVVENVNDMDKVSTILALVGKSHALKHKV
Cygb_Camelus_dromedarius   LEMERSPQLRKHACRVMGALNTVVENLHDPEKVSSVLALVGKAHALKHKV
Cygb_Callithrix_jacchus    LEMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKV
Cygb_Bos_taurus            LEMERSPQLRKHACRVMGALNTVVENLHDPEKVSSVLSLVGKAHALKHKV
Cygb_Homo_sapiens          LEMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKV
Cygb_Mus_musculus          LEMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKV
Cygb_Rattus_norvegicus     LEMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKV
Cygb_Budorcas_taxicolor    LEMERSPQLRKHACRVMGALNTVVENLHDPEKVSSVLALVGKAHALKHKV
Cygb_Bufo_bufo             LEMERSTQLRKHGRRVMGAINSVVENLGDPEKVATVLAIVGKSHAVKHKV
Cygb-2_Danio_rerio         EEMEKSSQLRKHARRVMNAINTVVENLHDPEKVSSVLVLVGKAHAFKYKV
Cygb_Chelydra_serpentina   LEMERTPQLRKHARRVMGAVNTVVENINDAEKVSSVLALVGKAHALKHKV
Cygb_Xenopus_laevis        LEMEGSMQLRKHARRVMGAVNSVVENLGDPEKVTTVLSIVGKSHALKHKV
Cygb_Gallus_gallus         LEMERSLQLRKHAQRVMGAINTVVENLDDPEKVSSVLALVGKAHALKHKV
GbY_Callorhinchus_milii    EEMKKSPQIKRHGKIVMSALNKLIANLDNGEELSSLLAKMAERHINVHKV
GbY_Latimeria_chalumnae    EEMQKNEQIRIHGLRVMNSLNQVIQNIDNLNEVYSILTHLAKRHQYVHRV
GbY_Xenopus_laevis         EEMQVNPGIRAHGKRVMGALNQVIQNLNDWEVVSSALTHLAQRHQDVHKV
GbY_Rhinella_marina        EEMQVNPGIRAHGKGVMGALNQVIQNLNDWEVVSSALTHLAQRHQDVHKV
GbY_Xenopus_tropicalis     QEMQDNAGIRAHGKRVMGALNHVIENLKDWDAVCSALSHLAKRHQDVHKV
Mb_Alligator_sinensis      DEMKSSEEMKKHGTNVFAALGNILKQKGNHA---EVLKPLAESHALKHKI
Mb_Mus_musculus            EDMKGSEDLKKHGCTVLTALGTILKKKGQHA---AEIQPLAQSHATKHKI
Mb_Delphinapterus_leucas   AEMKASEDLKKHGNTVLTALGGILKKKGHHD---AELKPLAQSHATKHKI
Mb_Balaenoptera_borealis   AEMKASEDLKKHGNTVLTALGGILKKKGHHE---AELKPLAQSHATKHKI
Mb_Homo_sapiens            DEMKASEDLKKHGATVLTALGGILKKKGHHE---AEIKPLAQSHATKHKI
Mb_Zootoca_vivipara        DDMKSSEELKKHGTTVLTALGKILKQKGNHE---AELAPLSQTHANTHKI
Mb_Cricetulus_griseus      DEMKASEDLKKHGCTVLTALGVILKKKGQHA---AEIQPLAQSHATKHKI
Mb_Gallus_gallus           DQMKGSEDLKKHGATVLTQLGKILKQKGNHE---SELKPLAQTHATKHKI
Mb_Crotalus_tigris         DEMKNSEELKKHGTTVLSALGKILKQKGSHE---AELSPLAQTHANRHKI
Hb_alpha_La...a_chalumnae  ---PSSQKLHAHAKVVLGALTKAVNHLDNIT---DTLHDISLVHAKKLLV
Hb-alpha_Ca...ius_gibelio  ---PGSGPVKKHGKTIMGAVGDAVSKIDDLV---GALSSLSELHAFKLRI
Hb_alpha1_Homo_sapiens     ---HGSAQVKGHGKKVADALTNAVAHVDDMP---NALSALSDLHAHKLRV
Hb-alpha_Acinonyx_jubatus  ---HGSAQVKAHGQKVADALTQAVAHMDDLP---NALSALSDLHAYKLRV

GbX_Latimeria_chalumnae    PPKYYQYVGIEFVCAVQPILK-ERWTP-EVEEAWQTLFKYVTEVMKRGYL
GbX_Tetraod...igroviridis  PPKYYSYVGAEFICAVQPILK-ERFTS-ELEEAWKTLFQYVTGLMRKGHQ
GbX_Perca_fluviatilis      PPKYYSYVGAEFICAVQPILK-ERWTA-ELEEAWKTMFQFVTSLMKKGYE
GbX_Plectro...s_leopardus  PPKYYSYVGAEFICAVQPILK-ERWTA-ELEEAWKTMFQYLTGLMKQGYE
Ngb_Mycteria_americana     KVESFSTVGESLLYMLEKCLG-TAFSP-DMREAWSKLYGAVVKAMRRGWE
Ngb_Homo_sapiens           KLSSFSTVGESLLYMLEKCLG-PAFTP-ATRAAWSQLYGAVVQAMSRGWD
Ngb_Mus_musculus           RLSSFSTVGESLLYMLEKCLG-PDFTP-ATRTAWSRLYGAVVQAMSRGWD
```

```
Ngb_Bufo_gargarizans        KLESFKMVGEALLYALESGLG-ATFTS-DARAAWSQLYAIVVSAMSRGWW
Ngb_Xenopus_laevis          KPESFKTVGESLLFALESGLG-DAFTS-DTREAWSLLYATVVQSMSRGWH
Ngb_Xenopus_tropicalis      KLESFNTVGESLLFALESCLG-DAFTS-DTREAWSLLYANVVQSMSRGWH
GbX_Salmo_salar             PPKYYSYVGAEFICAVQPILK-ERWTP-ELEEAWKTLFLYLTRLMNRGYQ
Ngb_Danio_rerio             NTQSFALVGESLLYMLQSSLG-PAYTT-SLRQAWLTMYSIVVSAMTRGWA
Ngb_Dissostichus_mawsoni    NTQSFAEVGESLLHMLQCSLG-QAYTA-PLRQAWLNLYSIVVAAMSQGWA
Ngb_Oncorhy...tshawytscha   NTQSFAVVGESLLYMLQCSLG-HGYTG-PLRQAWLNMYTIVVAAMSRGWA
Ngb_Oncorhy...s_gorbuscha   NTQSFAVVGESLLYMLQCSLG-HGYTG-PLRQAWLNMYTIVVAAMSRGWA
Ngb_Embiotoca_jacksoni      NTQSFAVVGESLLYMLQCSLG-QAYTA-PLRQAWLNMYSIVVAAMSRGWA
Ngb_Neoceratodus_forsteri   KIDSFPAVGESLLYMLERALG-TPLCT-EIKAAWIVLYTFVTKTMIRGWD
Ngb_Latimeria_chalumnae     KMESFPVVGESLLYMLEHGLD-SAFSV-EVRHAWIALYSFVVETMARGWG
GbX_Lepisosteus_oculatus    PPKYYGYVGTEFICAVQPILK-EKWTS-EVEEAWQTLFLYVTRIMKRGYQ
GbX_Danio_rerio             PPKYYGYVGAEFICAVRPILK-DRWTP-ELEEAWKTLFQYVTSIMREGFL
GbX_Carassius_auratus       PPKYYGYVGAEFICAVRPILK-DRWTP-ELEEAWKTLFQYVTSIMREGFL
Mb_Anguilla_anguilla        KLQNFQLITEVIVKLMGEKG--VD--A-AGQEAVRKVMQAVIGDIDNFYK
Mb_Denticeps_clupeoides     GLNNFKLISEVIVKVMAEKAG-LD--G-AGQDALRRVLGIVINDIDIYYK
Mb_Epinephe...lanceolatus   PIGHFKVMAEILGKVMEEKAG-LD--G-PGQQALKRLMTAVIADIEANYK
Hb_beta1_Danio_rerio        DPDNFRLLADCITVCAAMKFGQAGFNA-DVQEAWQKFLAVVVSALCRQYH
Hb-beta-1_C...odon_idella   DPDNFRLLADCITVCAAMKFGPSGFNA-DVQEAWQKFLSVVVSALCRQYH
Hb_beta2_Mus_musculus       DPENFRLLGNAIVIVLGHHLG-KDFTP-AAQAAFQKVVAGVATALAHKYH
Hb-beta_Carassius_gibelio   DPDNFRLLGDCLTIVVASQLG-RSFTP-DIQAAWQKFLAVVVSALGKQYH
Hb-beta-2_C...odon_idella   DPDNFRLLADCLTVVVAGQLG-ASFTP-EVQAAFQKFIAVVISALGRQYH
Hb-beta-2_C..._adamanteus   DPENFRLLGNILIIVLAGHHG-KEFIP-STHAAFQKLTNVVAHALAYRYH
Hb_beta1_Xenopus_laevis     DPLNFKRFGGCLSIALARHFH-EEYTP-ELHAAYEHLFDAIADALGKGYH
Hb_beta_Lat...a_chalumnae   DPQNFKLLSKCLIIVLATKLG-KQLTP-DVQATWEKLLSVVVAALSREYH
GbE_Chelydra_serpentina     DPKNFRIICDIILQLMEEKYG-G-----DCKASFEKVTNEICTRLNNAYK
GbE_Chrysem...icta_bellii   DPKNFRIICDIILQLMEEKYG-G-----DCKASFEKVTNEICTRLNNAYK
GBE_Latimeria_chalumnae     DPKNFKIICNILLQVLDEKFG-G-----DARAGFEKVTDVLCTHLNHAYK
GbE_Gallus_gallus           DPKNFRIICDIILQLMEEKFG-G-----DCKASFEKVTNEICTHLTNIYK
GbE_Patagioenas_fasciata    DPKNFRIICDIILQLMEEKFG-G-----DCKSSFEKVTNEICSHLNNVYK
Cygb-2_Amph...n_ocellaris   EPIYFKILSGVMLEVLSEDFP-EFFTA-EVQMVWTKLMGAVYWHVTGAYT
Cygb-2_Chelmon_rostratus    EPMYFKILSGVMLEVFAEDFP-EFFTA-EVQMLWAKLMGAVYWHVTGAYT
Cygb-1_Astyanax_mexicanus   EPRYFKLLTGVILEVLVEAFP-ETFGV-EAQGAWSKLMEVVHWQVMQVYS
Cygb-2_Alosa_sapidissima    EPVYFKILSGVILEVLADDFG-ECFTT-DVARAWTKLMGALYWHVTGAYA
Cygb-1_Danio_rerio          DPVYFKILAGVILEVLVEAFP-QCFSPAEVQSSWSKLMGILYWQMNRVYA
Cygb-1_Cyprinus_carpio      DPLYFKILAGVILEVLVEAFP-QCFSPASVQSAWSKLLGVLYWQMNKVYA
Cygb-1_Cten...odon_idella   DPVYFKILAGVILEVLVEAFP-QSFSPASVQGAWSKLMGVLYWQMNKVYA
Cygb_Neocer...us_forsteri   EPMYFKILNGVILEVFPEEYP-DDFTP-EVQRAWTKLMGLFYCHVMASYK
Cygb_Rhinat..._bivittatum   EPVYFKILTGVMLEVIAEEYA-NDFTP-EVQRAWTKLKNLIYSHVTATYK
Cygb_Camelus_dromedarius    EPVYFKILSGVILEVIAEEFA-NDFPP-ETQRAWAKLRGLIYSHVTAAYK
Cygb_Callithrix_jacchus     EPVYFKILSGVILEVIAEEFA-NDFPP-ETQRAWAKLRGLIYSHVTAAYK
```

```
Cygb_Bos_taurus             EPVYFKILSGVILEVIAEEFA-NDFPP-ETQRAWAKLRGLIYSHVTAAYK
Cygb_Homo_sapiens           EPVYFKILSGVILEVVAEEFA-SDFPP-ETQRAWAKLRGLIYSHVTAAYK
Cygb_Mus_musculus           EPMYFKILSGVILEVIAEEFA-NDFPV-ETQKAWAKLRGLIYSHVTAAYK
Cygb_Rattus_norvegicus      EPMYFKILSGVILDVIAEEFA-NDFPV-ETQKAWTKLRGLIYSHVTAAYK
Cygb_Budorcas_taxicolor     EPVYFKILSGVILEVIAEEFA-SDFPP-ETQRAWAKLRGLIYSHVTAAYK
Cygb_Bufo_bufo              D-----------------------------------------------
Cygb-2_Danio_rerio          EPIYFKILSGVILEILAEEFG-ECFTP-EVQTSWSKLMAALYWHITGAYT
Cygb_Chelydra_serpentina    EPVYFKFFTGVMLEVIAEEYA-NDFTP-EVQRAWTKMKSLIYTHVTAAYK
Cygb_Xenopus_laevis         EPVYFKILTGVMLEVIAEEYA-KDFTP-DVQLAWNKLRSLIYSQVQSAYK
Cygb_Gallus_gallus          EPVYFKKLTGVMLEVIAEAYG-NDFTP-EAHGAWTKMRTLIYTHVTAAYK
GbY_Callorhinchus_milii     DLHNFQIIFNIIIAILEETFG-NAFTP-EIRETWTKLFGVIYACLESHYK
GbY_Latimeria_chalumnae     DVHNFKLIFGVIIKILKEALG-ATFTE-EICTSWQKMLSFTYDYLVSCYH
GbY_Xenopus_laevis          GVNNFQLLFLVILTIFKEALG-ADFTP-EHCKSWEKLFSITYNFLDSCYT
GbY_Rhinella_marina         GVNNFQLLFLVILTIFKEALG-ADFTP-EHCKSWEKLFSITYNFLGSCYT
GbY_Xenopus_tropicalis      EVNNFELLFLVIISVFKEALG-SGFTP-EQSKSWEKLFSITYKYLESCYA
Mb_Alligator_sinensis       PVKYLEFISEIIVKVIAEKYP-VDFGA-DSQAAMRKALELFRNDMASKYK
Mb_Mus_musculus             PVKYLEFISEIIIEVLKKRHS-GDFGA-DAQGAMSKALELFRNDIAAKYK
Mb_Delphinapterus_leucas    PIKYLEFISEAIIHVLHSRHP-GDFGA-DAQGAMNKALELFRKDIAAKYK
Mb_Balaenoptera_borealis    PIKYLEFISDAIIHVLHSRHP-GDFGA-DAQAAMNKALELFRKDIAAKYK
Mb_Homo_sapiens             PVKYLEFISECIIQVLQSKHP-GDFGA-DAEGAMNKALELFRKDMASNYK
Mb_Zootoca_vivipara         PVKYLEFICEIIVGVIAEKHS-ADFGA-DSQAAMRKALELFRNDMASKYK
Mb_Cricetulus_griseus       PVKYLEFISESIIEVLKSRHS-GDFGD-DAQGAMSKALELFRNDIAAKYK
Mb_Gallus_gallus            PVKYLEFISEVIIKVIAEKHA-ADFGA-DSQAAMKKALELFRNDMASKYK
Mb_Crotalus_tigris          PVKYLEFISEVIIDVIAEKHS-ADFGA-DSQEAMRKALELFRNDMASKYK
Hb_alpha_La...a_chalumnae   DPVNFELLGHCLEVALAAHFA-TDFTP-EVHLAIDKFLYEVEKALFETYR
Hb-alpha_Ca...ius_gibelio   DPANFKILAHNVIVVIGMLFP-GDFTP-EVHMSVDKFFQNLALALSEKYR
Hb_alpha1_Homo_sapiens      DPVNFKLLSHCLLVTLAAHLP-AEFTP-AVHASLDKFLASVSTVLTSKYR
Hb-alpha_Acinonyx_jubatus   DPVNFKFLSHCLLVTLACHHP-AEFTP-AVHASLDKFFSAVSTVLTSKYR

GbX_Latimeria_chalumnae     Q-EEKIKRTMETVEPKQASRRANNAI
GbX_Tetraod...igroviridis   E-EGSRQRH-LALPPKDGPEKRTSAL
GbX_Perca_fluviatilis       E-ESNRQRH-LALSPKERPDKKNTAL
GbX_Plectro...s_leopardus   E-ESNRQRH-LALSPKERPDKRNTAL
Ngb_Mycteria_americana      T-LP-------------------EGD
Ngb_Homo_sapiens            ------------------------GE
Ngb_Mus_musculus            ------------------------GE
Ngb_Bufo_gargarizans        R----------------------DP
Ngb_Xenopus_laevis          R-DSQE---------------QQEGI
Ngb_Xenopus_tropicalis      R-DSQE---------------QREGI
GbX_Salmo_salar             E-RNRHQQVG--AS----PKERNIAL
Ngb_Danio_rerio             K-NGE-----------------HKSN
```

```
Ngb_Dissostichus_mawsoni   K-NGE-----------------DKAD
Ngb_Oncorhy...tshawytscha  K-KGE-----------------HKTD
Ngb_Oncorhy...s_gorbuscha  K-NGE-----------------HKTD
Ngb_Embiotoca_jacksoni     K-NGE-----------------DKAD
Ngb_Neoceratodus_forsteri  K-NGE-----------------NKIE
Ngb_Latimeria_chalumnae    A-NGE-----------------NKLN
GbX_Lepisosteus_oculatus   E-EEKSKRNNVVIASKERPEKKGTAI
GbX_Danio_rerio            E-EERNKRSNTQTSSRERPDKRSTAI
GbX_Carassius_auratus      E-EERNKRSNTQTSSRERPDKRSTAI
Mb_Anguilla_anguilla       E-FGF-------------------QG
Mb_Denticeps_clupeoides    E-LGF-------------------AG
Mb_Epinephe...lanceolatus  E-LGF-------------------SG
Hb_beta1_Danio_rerio       --------------------------
Hb-beta-1_C...odon_idella  --------------------------
Hb_beta2_Mus_musculus      --------------------------
Hb-beta_Carassius_gibelio  --------------------------
Hb-beta-2_C...odon_idella  --------------------------
Hb-beta-2_C..._adamanteus  --------------------------
Hb_beta1_Xenopus_laevis    --------------------------
Hb_beta_Lat...a_chalumnae  --------------------------
GbE_Chelydra_serpentina    E-AGW--------------------
GbE_Chrysem...icta_bellii  E-AGW--------------------
GBE_Latimeria_chalumnae    E-AGW--------------------
GbE_Gallus_gallus          E-AGW--------------------
GbE_Patagioenas_fasciata   E-AGW--------------------
Cygb-2_Amph...n_ocellaris  E-VGWLQV-------------SSSAV
Cygb-2_Chelmon_rostratus   E-VGWLQV-------------SSSAV
Cygb-1_Astyanax_mexicanus  E-IGWTSA-------------NSATE
Cygb-2_Alosa_sapidissima   E-VGWVKV-------------SSSAV
Cygb-1_Danio_rerio         E-VGWE----------------NSKK
Cygb-1_Cyprinus_carpio     E-VGWDSI-------------KNSAK
Cygb-1_Cten...odon_idella  E-VGWESI-------------KNSAK
Cygb_Neocer...us_forsteri  E-VGWVQY-------------PNSTM
Cygb_Rhinat..._bivittatum  E-VGWA-------------------Q
Cygb_Camelus_dromedarius   E-VGWVQQVPNA------------TT
Cygb_Callithrix_jacchus    E-VGWVQQVPNATTP--PATLPSSGP
Cygb_Bos_taurus            E-VGWVQQVPNATTP--PATLPSSGP
Cygb_Homo_sapiens          E-VGWVQQVPNATTP--PATLPSSGP
Cygb_Mus_musculus          E-VGWVQQVPNTTTP--PATLPSSGP
Cygb_Rattus_norvegicus     E-VGWVQQVPNTTTL--PATLPSSGP
Cygb_Budorcas_taxicolor    E-VGWVQQVPNATTP--PATLPSSGP
```

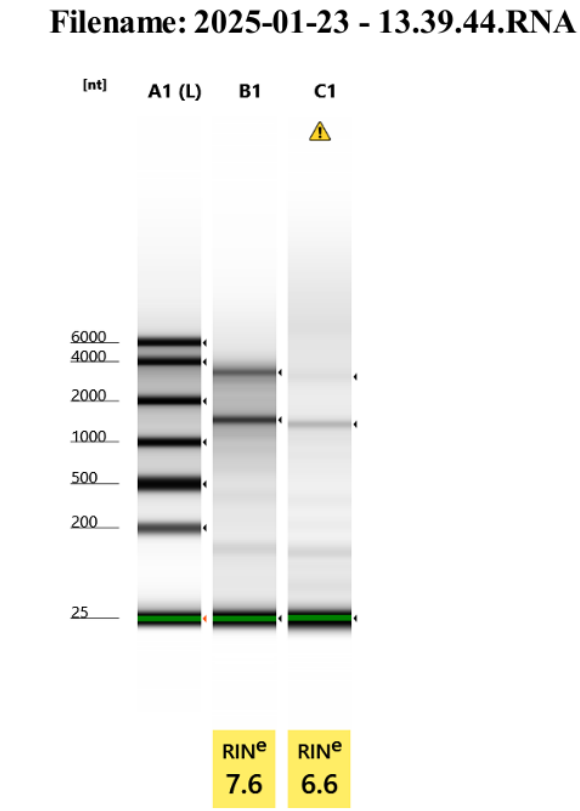
```
Cygb_Bufo_bufo              -------------------------
Cygb-2_Danio_rerio          E-VGWVKL------------SSSAV
Cygb_Chelydra_serpentina    E-VGWVQY------------PNSTM
Cygb_Xenopus_laevis         E-VGWTQY------------PSNSV
Cygb_Gallus_gallus          E-AGWVSY------------PSATL
GbY_Callorhinchus_milii     D-AGF-------------------YP
GbY_Latimeria_chalumnae     KSS---------------------GP
GbY_Xenopus_laevis          K-S---------------------DS
GbY_Rhinella_marina         K-S---------------------DS
GbY_Xenopus_tropicalis      N-T---------------------DS
Mb_Alligator_sinensis       E-FGY-------------------QG
Mb_Mus_musculus             E-LGF-------------------QG
Mb_Delphinapterus_leucas    E-LGF-------------------HG
Mb_Balaenoptera_borealis    E-LGF-------------------QG
Mb_Homo_sapiens             E-LGF-------------------QG
Mb_Zootoca_vivipara         E-HGF-------------------QG
Mb_Cricetulus_griseus       E-LGF-------------------QG
Mb_Gallus_gallus            E-FGF-------------------QG
Mb_Crotalus_tigris          E-LGF-------------------QG
Hb_alpha_La...a_chalumnae   -------------------------
Hb-alpha_Ca...ius_gibelio   -------------------------
Hb_alpha1_Homo_sapiens      -------------------------
Hb-alpha_Acinonyx_jubatus   -------------------------
//
```

cDNA primer list

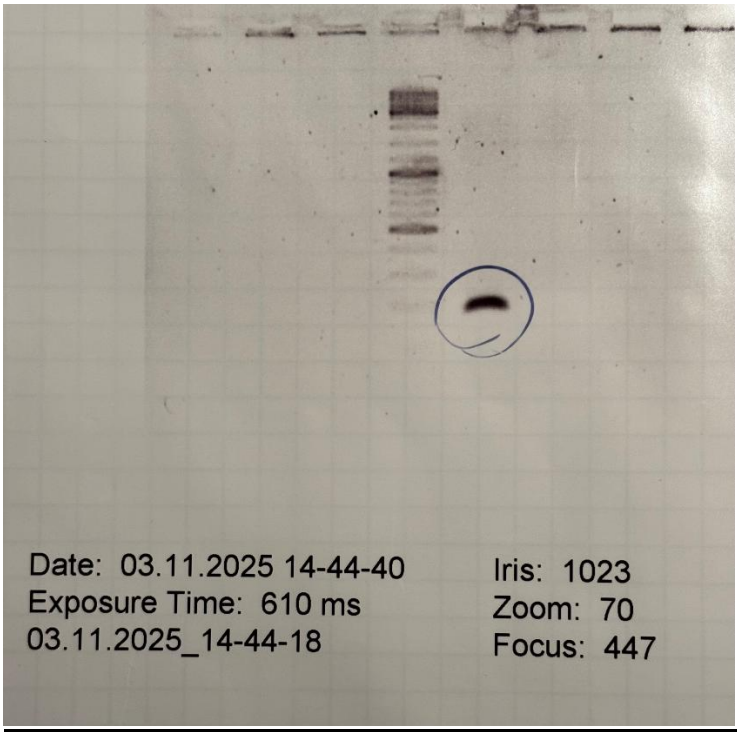
Name	Primer	Sequenz 5‘- 3‘	Length (bp)	Tm	GC %	Tm (Salt adjusted)
Nfo_G1	NfoG1_cDNA_fw	GTG TTG AGT AAA TGA ATT TTG ATG TC	26	51,5	30,8	60,1
	NfoG1_cDNA_rv	CAGCCATTTAAACAAAGGTGTTG	21	50,8	38,1	55,4
Nfo_G2	NfoG2_cDNA_fw	CAT TTA CGC TGT ATT AAA TGC CTT TAG	27	53,1	33,3	62,1
	NfoG2_cDNA_rv	CATATACCCATTTATCCAACGCTAC	25	53,5	40	62,5
Nfo_G3	NfoG3_cDNA_fw	CTGTATGAATGTGTAATGATTGGG	24	51,9	37,5	60,3
	NfoG3_cDNA_rv	GGCATCTAACGATTTATAATGCTTAC	26	52,3	34,6	61,7
Nfo_G4	NfoG4_cDNA_fw	GCTTGccttaaataaATGATTATCTGC	27	53,4	33,3	62,1
	NfoG4_cDNA_rv	CATTCATGCTTTGTGTTCTATTTGC	25	53,7	36	60,9
Nfo_G5	NfoG5_cDNA_fwd	GTTGAATGTTTCTGAATTACAAAACAAGC	29	54,5	31	63,3
	NfoG5_cDNA_rv	GTA GTA ACA GCT AAT CAG CAT CTA ATG	27	53,5	37	63,7
Nfo_G6	Nfo_G6_cDNA_fwd	GTCATAGATGATTGAGTTGGTTTTCC	26	54,1	38,5	62,9
	Nfo_G6_cDNA_rv	CAG CCA ATC AGC ATC TAA TGA TTT ATG	27	54,6	37	63,7
Nfo_G7	Nfo_G7_cDNA_fwd	GCAAGTAGATGCCTTTAGGG	20	53,1	50	58,4
	Nfo_G7_cDNA_rv	GGA AGA CAT TTA TCC AAA ATT CCT TC	26	52,7	34,6	61,7
Nfo_G10	Nfo_G10_cDNA_fwd	CTAGATGAGGTGTGCTTTATCAG	23	52,8	43,5	60,9
	Nfo_G10_cDNA_rv	CTT TCA TTT ACT GCT CAC AAT GTG	24	52,8	37,5	60,3
Nfo_G11	Nfo_G11E1_cDNA_fwd	GGGTTTCACTGCACTGATAAAC	22	53	45,5	60,1
	Nfo_G11E2_cDNA_rv	CTTTCATTGAAAAGAGTGTTGCCA	24	52,3	37,5	60,3
	Nfo_G11E2_cDNA_fwd	AGTTCAAAGGCTATCTGTTCAAGC	24	54	41,7	62
	Nfo_G11EX_cDNA_fwd	ctcgtTTCATCTAGGGGATTTAAG	24	54	41,7	62
	Nfo_G11E3_cDNA_rv	CCTGGCAATATCCTGGTGAAG	21	54,4	52,4	61,2
	Nfo_G11E3_cDNA_fwd	CTTTATCAGGACCTGGAATAACAC	24	54	41,7	62
	Nfo_G11E4_cDNA_rv	GAAGACATCTTTGCACTCAGGG	22	54,8	50	62,1
Nfo_G1	Nfo_G1E1_cDNA_fwd	GAATACCCGAGCCTGTCTATG	21	54,4	52,4	61,2
	Nfo_G1E2_cDNA_rv	GATATGGTCCAGGAACTCCG	20	53,8	55	60,5
	Nfo_G1E2_cDNA_fwd	CGGATCTGGTGCATCTCTTC	20	53,8	55	60,5
	Nfo_G1E3_cDNA_rv	CTCCAACGGTCTTGTGCTTC	20	53,8	55	60,5
	Nfo_G1E3_cDNA_fwd	GAACAACCTAGAGAACCTGCC	21	54,4	52,4	61,2
	Nfo_G1E4_cDNA_rv	CCCATTCTTGTCCCAGCCT	19	53,2	57,9	59,2
Nfo_G57	Nfo_G57E1_cDNA_fwd	GAGTAACTTTGATACCCATGTTCC	24	54	41,7	62
	Nfo_G57E2_cDNA_rv	CCTCATTCTCTTCAAGGGTGC	21	54,4	52,4	61,2
	Nfo_G57E2_cDNA_fwd	CAGGGATCTTGTTGTAATGCTG	22	53	45,5	60,1
	Nfo_G57E3_cDNA_rv	CCATCCAATGTAAAGAATCTGAGGA	25	54,4	40	62,5
Nfo_G2	Nfo_G2E1_cDNA_rv	GGCATTCTGGACCTCAACATC	21	54,4	52,4	61,2
	Nfo_G2E1_cDNA_fwd	GATGTTGAGGTCCAGAATGCC	21	54,4	52,4	61,2
	Nfo_G2E2_cDNA_rv	GTGAACATTCTGCAAGGCAAAAG	23	53,5	43,5	60,9

	Nfo_G2E2_cDNA_fwd	CCCACTTCACCCATTTCAAGG	21	54,4	52,4	61,2
Nfo_G2	NfoG2E2a_cDNA_rv	GGTCCACTTTGAGCTGTGTAG	21	54,4	52,4	61,2
Nfo_G57	NfoG57E3a_cDNA_rv	CCATCGACCAGTAttttagtctagat	26	54,8	38,5	62,9
	NfoG57E1a_cDNA_fwd	CAACGTGGAACCTCATACCTTC	22	54,8	50	62,1
Nfo_G11	Nfo_G11E3_cDNA_rv	GATTTGACTTCAATATGCCTGATCC	25	54,4	40	62,5
Nfo_G1	GSP1_G1E4_5-RACE	CCAGCCTCGGATCATGGTTTTGT	23	57,1	52	64,6
	GSP1_G1E2_3-RACE	GGATCTGGTGCATCTCTTCCATTAC	25	57,7	48	65,8
Nfo_G57	GSP1_G57E3_3-RACE	CCATCCAATGTAAAGAATCTGAGGAAG	27	56,7	41	65,3
	GSP1_G57E2_5-RACE	GCTTGCACCCTTGAAGAGAATGAG	24	57,4	50	65,2
	GSP2_G57E2_3-RACE	GAGTGGCCCAAGTGTGTTTAGG	22		54,5	64,9
	GSP3_G57E2_3-RACE	CCAGCTTTTATCCCATAATTGCATTAGATG	30		36,7	64,8
Nfo_G1	GSP2_G1E3_3-RACE	GTGATTGATGCTGCCGTGAACA	22		50	65,5
	GSP3_G1E4_3-RACE	GTGGCTGCTCTATATGTTGGGAG	23		52,2	64,3
Nfo_G57	GSP2_G57E2_5-RACE	CCTAAACACACTTGGGCCACTC	22	62	54,5	62
	GSP3_G57E2_5-RACE	CATCTAATGCAATTATGGGATAAAAGCTGG	30	60	36,7	60
NfoG_00	NfoG_00E1_fwd	GCAGAGGACAATGGAGCAAT	20	56,8	50	
	NfoG_00E3_rv	TGTAGGCGGCGTTCATTTG	19	57,8	52,6	
NfoG1	NfoG1_E1_fwd	GCACTGTGTTGTTCTCCAGG	20	58,3	55	
NfoG_00	NfoG_00E2a_fwd	CCTGAACAACATGGATGCCT	20	57		
	NfoG_00E2b_fwd	CATTTCAAGGGGATGGGGTC	20	59		
	NfoG_00E3b_rv	GGCATCCCCACCAAATTTCT	20	58		
	NfoG_00E3a_rv	GCGTTCATTTGGGCACACAC	20	59		

RNA Integrity Check



PCR amplicon Ngb



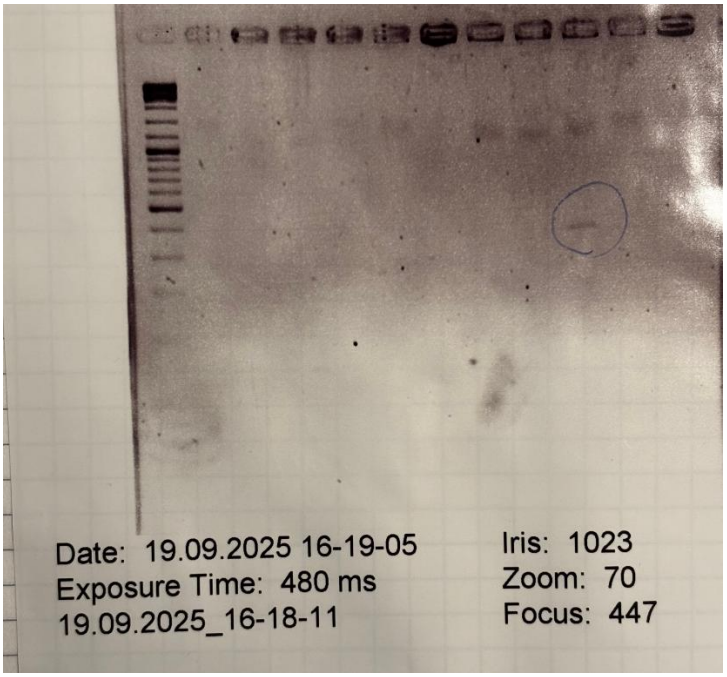
Primer combination:
E1_fwd/1_5GSP

Expected fragment size:
228 bp

cDNA tissue: brain

Annealing temperature: 59 °C

PCR amplicon Cygb



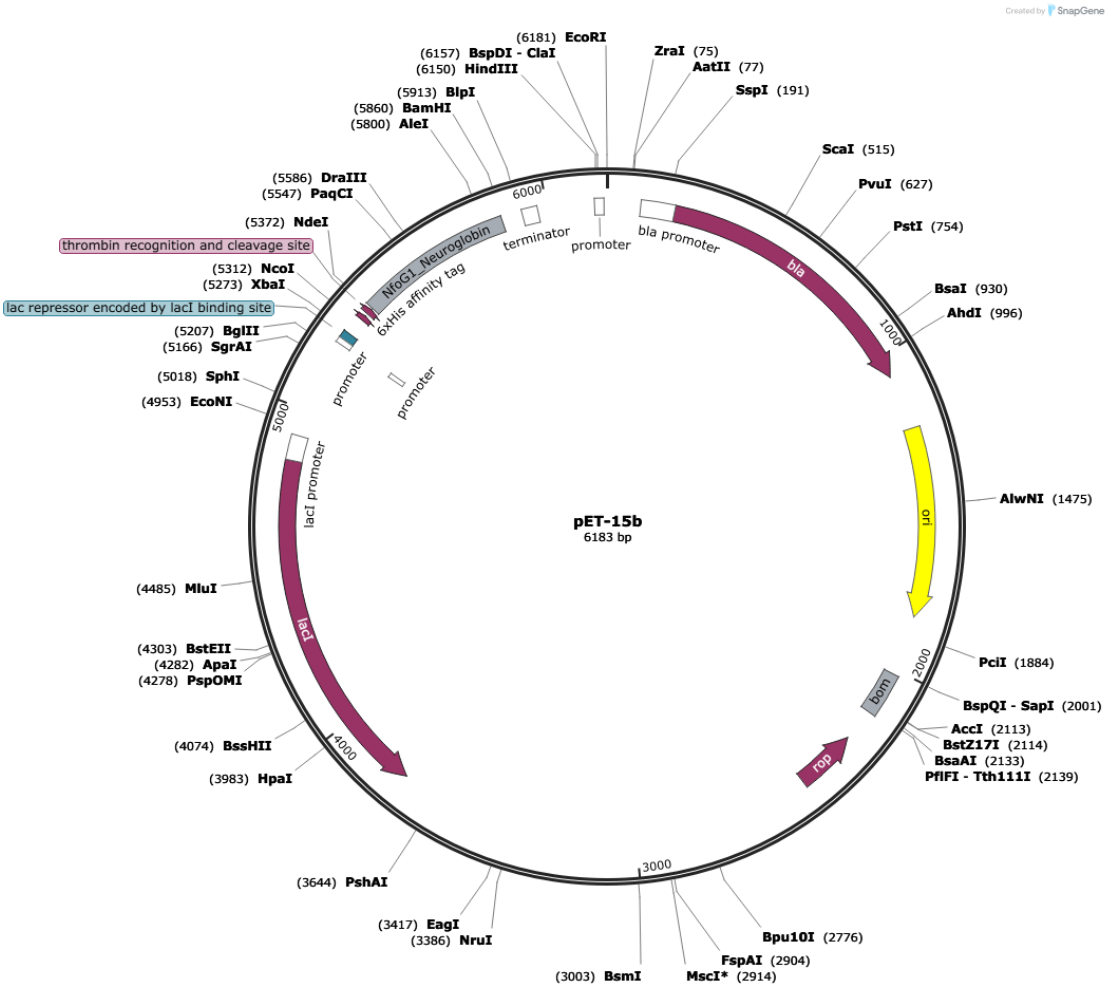
Primer combination:
E1_fwd/E3rv

Expected fragment size:
349 bp

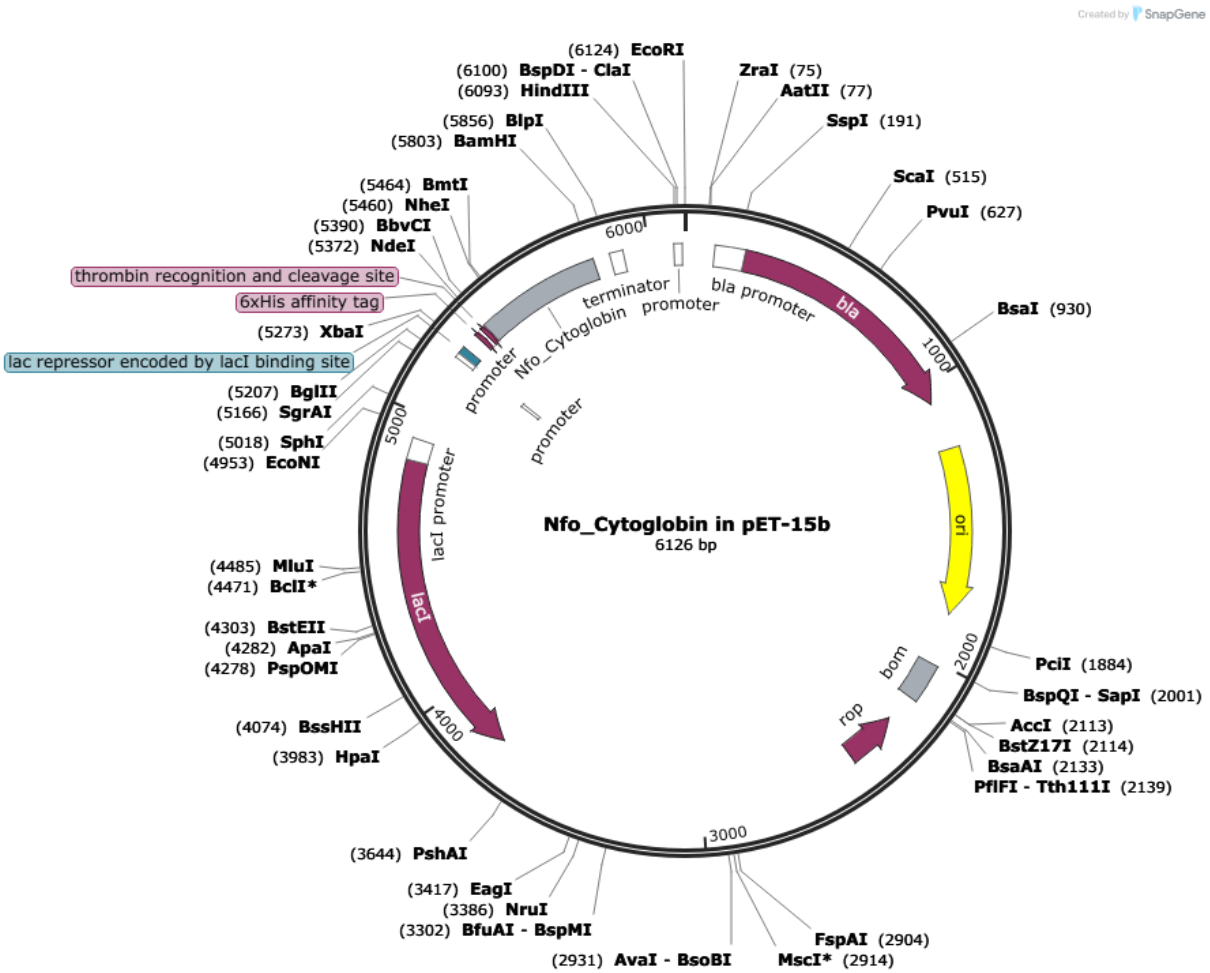
cDNA tissue: heart

Annealing temperature: 58 °C

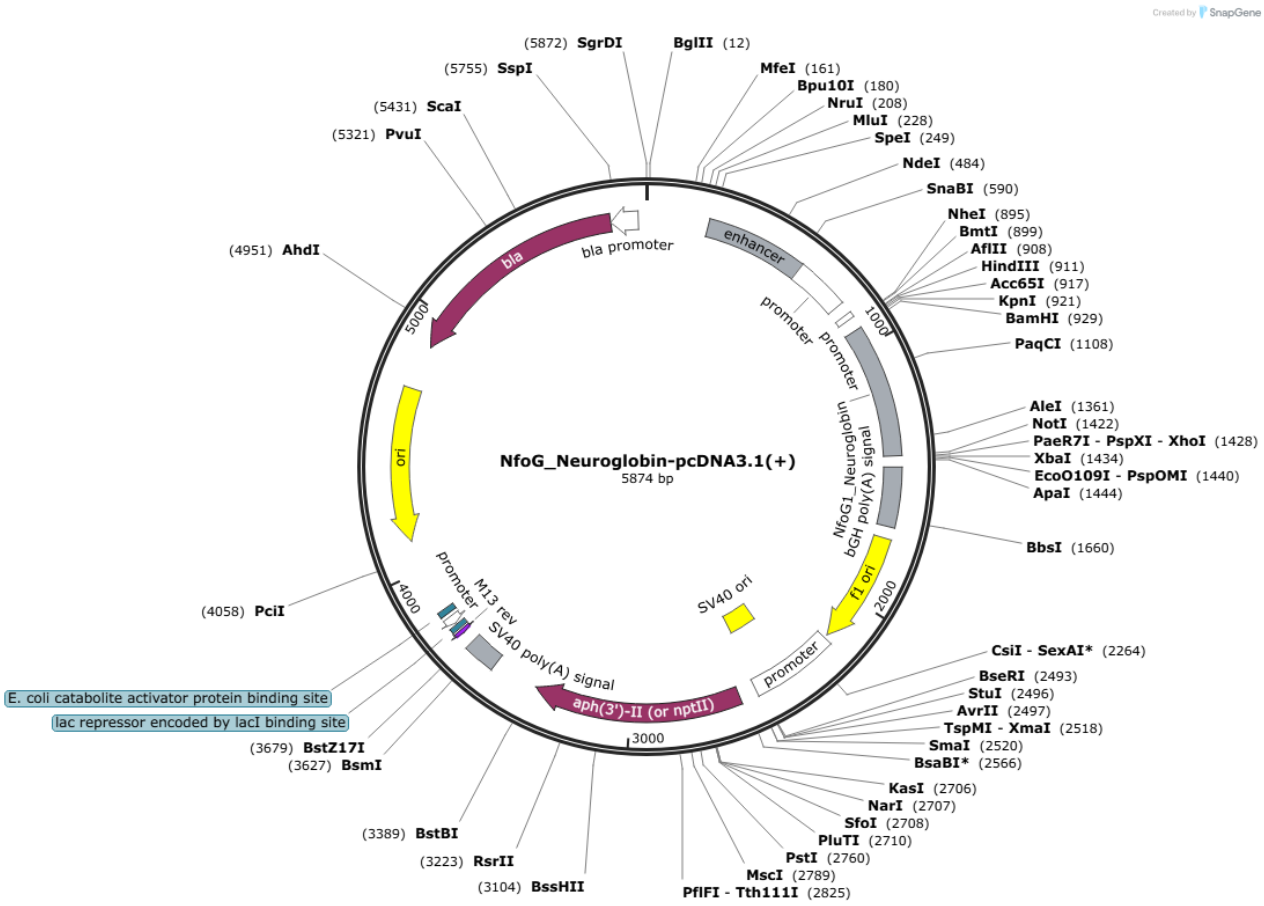
Plasmid map Nfo Neuroglobin in pET-15b



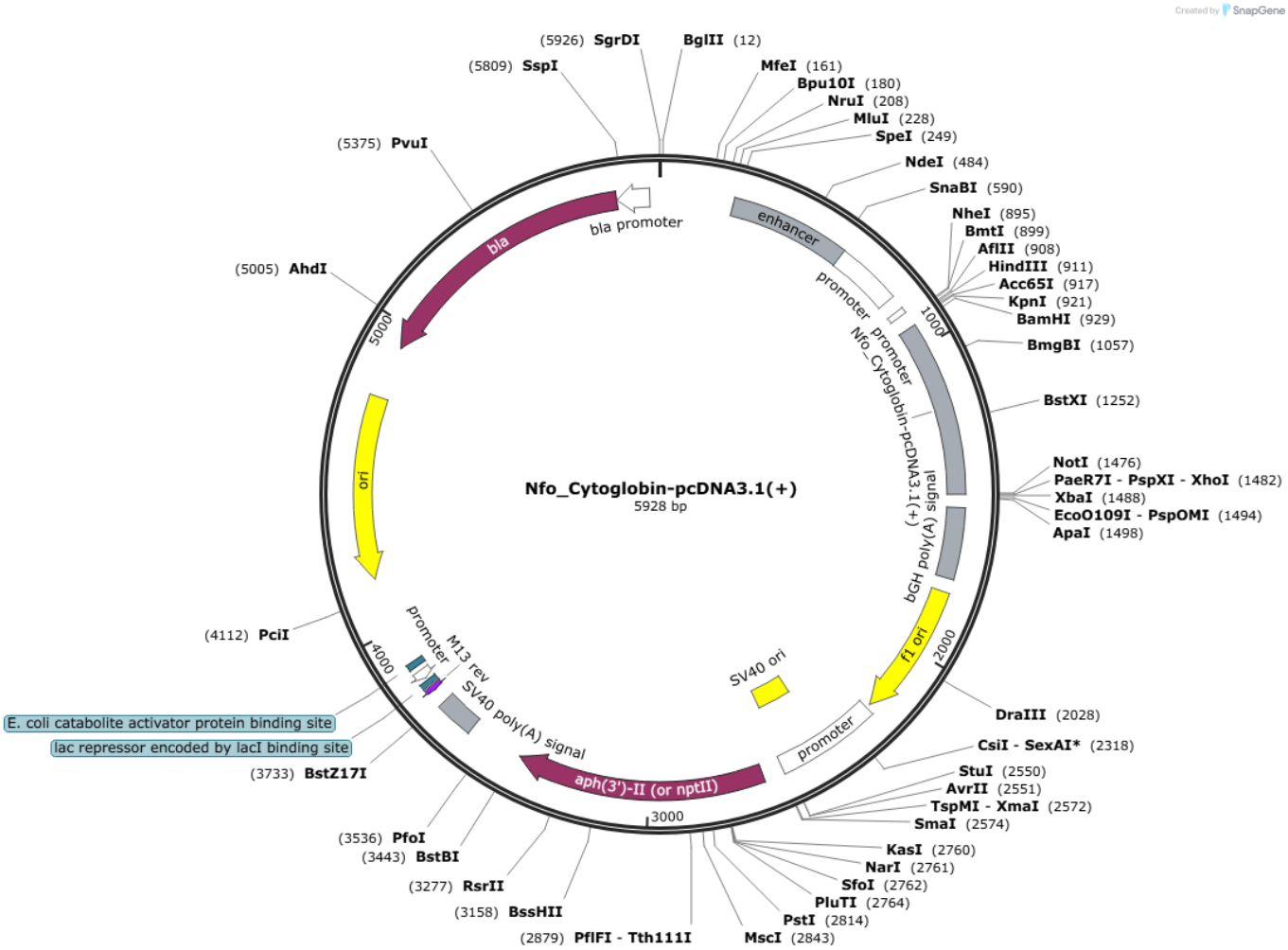
Plasmid map Nfo Cytoglobin in pET-15b



Plasmid map Nfo Neuroglobin in pcDNA3.1(+)



Plasmid map Nfo Cytoglobin in pcDNA3.1(+)



H_2O_2 stress experiment statistics

```
Call:
glm(formula = cbind(live, dead) ~ Construct, family = binomial,
    data = df)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-81.42  -39.78    0.00   33.87  102.27

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.763584   0.007167  246.08   <2e-16 ***
ConstructMock      -0.925258   0.008655 -106.90   <2e-16 ***
ConstructNgb ctrl  -0.377273   0.011557  -32.64   <2e-16 ***
ConstructNgb       -1.724356   0.008485 -203.22   <2e-16 ***
ConstructCygb ctrl  1.637351   0.017978   91.07   <2e-16 ***
ConstructCygb      -1.111255   0.008208 -135.39   <2e-16 ***
ConstructeGFP ctrl  0.140663   0.009723   14.47   <2e-16 ***
ConstructeGFP      -0.739073   0.008769  -84.28   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 177532  on 15  degrees of freedom
Residual deviance:  40065  on  8  degrees of freedom
AIC: 40262

Number of Fisher Scoring iterations: 5

> anova(model, test = "Chisq")
Analysis of Deviance Table

Model: binomial, link: logit
```

```
Response: cbind(live, dead)

Terms added sequentially (first to last)


          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                         15     177532
Construct  7   137467         8      40065 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Post-hoc Vergleiche
> emm <- emmeans(model, ~ Construct, type = "response")
>
> emm
 Construct   prob       SE  df asymp.LCL asymp.UCL
 Mock ctrl 0.8537 0.000895 Inf    0.8519    0.8554
 Mock      0.6981 0.001020 Inf    0.6961    0.7001
 Ngb ctrl  0.8000 0.001450 Inf    0.7971    0.8028
 Ngb       0.5098 0.001140 Inf    0.5076    0.5120
 Cygb ctrl 0.9677 0.000515 Inf    0.9667    0.9687
 Cygb      0.6575 0.000901 Inf    0.6558    0.6593
 eGFP ctrl 0.8704 0.000741 Inf    0.8689    0.8718
 eGFP      0.7359 0.000982 Inf    0.7339    0.7378

Confidence level used: 0.95
Intervals are back-transformed from the logit scale
> pairs(emm, adjust = "holm")
 contrast              odds.ratio       SE  df null  z.ratio p.value
 Mock ctrl / Mock          2.5225 0.021800 Inf    1  106.902 <0.0001
 Mock ctrl / Ngb ctrl      1.4583 0.016900 Inf    1   32.644 <0.0001
 Mock ctrl / Ngb           5.6089 0.047600 Inf    1  203.216 <0.0001
 Mock ctrl / Cygb ctrl     0.1945 0.003500 Inf    1  -91.073 <0.0001
```

Mock ctrl / Cygb	3.0382	0.024900	Inf	1	135.394	<0.0001
Mock ctrl / eGFP ctrl	0.8688	0.008450	Inf	1	-14.467	<0.0001
Mock ctrl / eGFP	2.0940	0.018400	Inf	1	84.282	<0.0001
Mock / Ngb ctrl	0.5781	0.005950	Inf	1	-53.285	<0.0001
Mock / Ngb	2.2235	0.014800	Inf	1	120.206	<0.0001
Mock / Cygb ctrl	0.0771	0.001330	Inf	1	-149.097	<0.0001
Mock / Cygb	1.2044	0.007570	Inf	1	29.574	<0.0001
Mock / eGFP ctrl	0.3444	0.002810	Inf	1	-130.486	<0.0001
Mock / eGFP	0.8301	0.005820	Inf	1	-26.574	<0.0001
Ngb ctrl / Ngb	3.8462	0.039000	Inf	1	132.829	<0.0001
Ngb ctrl / Cygb ctrl	0.1334	0.002510	Inf	1	-107.065	<0.0001
Ngb ctrl / Cygb	2.0834	0.020600	Inf	1	74.063	<0.0001
Ngb ctrl / eGFP ctrl	0.5957	0.006670	Inf	1	-46.254	<0.0001
Ngb ctrl / eGFP	1.4359	0.014900	Inf	1	34.855	<0.0001
Ngb / Cygb ctrl	0.0347	0.000593	Inf	1	-196.560	<0.0001
Ngb / Cygb	0.5417	0.003280	Inf	1	-101.282	<0.0001
Ngb / eGFP ctrl	0.1549	0.001240	Inf	1	-233.458	<0.0001
Ngb / eGFP	0.3733	0.002540	Inf	1	-144.994	<0.0001
Cygb ctrl / Cygb	15.6208	0.265000	Inf	1	162.001	<0.0001
Cygb ctrl / eGFP ctrl	4.4669	0.079300	Inf	1	84.324	<0.0001
Cygb ctrl / eGFP	10.7663	0.186000	Inf	1	137.802	<0.0001
Cygb / eGFP ctrl	0.2860	0.002200	Inf	1	-162.735	<0.0001
Cygb / eGFP	0.6892	0.004440	Inf	1	-57.746	<0.0001
eGFP ctrl / eGFP	2.4103	0.020000	Inf	1	106.128	<0.0001

P value adjustment: holm method for 28 tests

Tests are performed on the log odds ratio scale

Temperature stress experiment statistics

```
Response: survival
               Df Sum Sq Mean Sq F value Pr(>F)
Construct       3  63.28  21.094  0.3039 0.8221
Temp            1   6.05   6.050  0.0872 0.7729
Construct:Temp  3 400.62 133.539  1.9237 0.1796
Residuals      12 833.00  69.417

> emmeans(fit, pairwise ~ Construct | Temp, adjust = "tukey")
$emmeans
Temp = 20:
 Construct emmean   SE df lower.CL upper.CL
 Mock        84.0 5.89 12     71.2     96.8
 eGFP        86.5 5.89 12     73.7     99.3
 Ngb         75.3 4.81 12     64.9     85.8
 Cygb        79.3 4.81 12     68.9     89.8

Temp = 23:
 Construct emmean   SE df lower.CL upper.CL
 Mock        80.5 5.89 12     67.7     93.3
 eGFP        72.0 5.89 12     59.2     84.8
 Ngb         83.3 4.81 12     72.9     93.8
 Cygb        87.0 4.81 12     76.5     97.5

Confidence level used: 0.95

$contrasts
Temp = 20:
 contrast    estimate   SE df t.ratio p.value
 Mock - eGFP    -2.50 8.33 12  -0.300  0.9901
 Mock - Ngb      8.67 7.61 12   1.139  0.6735
 Mock - Cygb     4.67 7.61 12   0.614  0.9258
 eGFP - Ngb     11.17 7.61 12   1.468  0.4845
 eGFP - Cygb     7.17 7.61 12   0.942  0.7833
```

```
 Ngb - Cygb     -4.00 6.80 12  -0.588  0.9338

Temp = 23:
 contrast    estimate   SE df t.ratio p.value
 Mock - eGFP     8.50 8.33 12   1.020  0.7412
 Mock - Ngb     -2.83 7.61 12  -0.373  0.9815
 Mock - Cygb    -6.50 7.61 12  -0.855  0.8276
 eGFP - Ngb    -11.33 7.61 12  -1.490  0.4725
 eGFP - Cygb   -15.00 7.61 12  -1.972  0.2511
 Ngb - Cygb     -3.67 6.80 12  -0.539  0.9477

> pairs(emm, adjust = "tukey")
 contrast                  estimate   SE df t.ratio p.value
 Mock Temp20 - eGFP Temp20   -2.500 8.33 12  -0.300  1.0000
 Mock Temp20 - Ngb Temp20     8.667 7.61 12   1.139  0.9348
 Mock Temp20 - Cygb Temp20    4.667 7.61 12   0.614  0.9979
 Mock Temp20 - Mock Temp23    3.500 8.33 12   0.420  0.9998
 Mock Temp20 - eGFP Temp23   12.000 8.33 12   1.440  0.8230
 Mock Temp20 - Ngb Temp23     0.667 7.61 12   0.088  1.0000
 Mock Temp20 - Cygb Temp23   -3.000 7.61 12  -0.394  0.9999
 eGFP Temp20 - Ngb Temp20    11.167 7.61 12   1.468  0.8098
 eGFP Temp20 - Cygb Temp20    7.167 7.61 12   0.942  0.9749
 eGFP Temp20 - Mock Temp23    6.000 8.33 12   0.720  0.9945
 eGFP Temp20 - eGFP Temp23   14.500 8.33 12   1.740  0.6653
 eGFP Temp20 - Ngb Temp23     3.167 7.61 12   0.416  0.9998
 eGFP Temp20 - Cygb Temp23   -0.500 7.61 12  -0.066  1.0000
 Ngb Temp20 - Cygb Temp20    -4.000 6.80 12  -0.588  0.9984
 Ngb Temp20 - Mock Temp23    -5.167 7.61 12  -0.679  0.9961
 Ngb Temp20 - eGFP Temp23     3.333 7.61 12   0.438  0.9998
 Ngb Temp20 - Ngb Temp23     -8.000 6.80 12  -1.176  0.9245
 Ngb Temp20 - Cygb Temp23   -11.667 6.80 12  -1.715  0.6796
 Cygb Temp20 - Mock Temp23   -1.167 7.61 12  -0.153  1.0000
 Cygb Temp20 - eGFP Temp23    7.333 7.61 12   0.964  0.9716
 Cygb Temp20 - Ngb Temp23    -4.000 6.80 12  -0.588  0.9984
```

```
 Cygb Temp20 - Cygb Temp23   -7.667 6.80 12  -1.127  0.9381
 Mock Temp23 - eGFP Temp23    8.500 8.33 12   1.020  0.9620
 Mock Temp23 - Ngb Temp23    -2.833 7.61 12  -0.373  0.9999
 Mock Temp23 - Cygb Temp23   -6.500 7.61 12  -0.855  0.9852
 eGFP Temp23 - Ngb Temp23   -11.333 7.61 12  -1.490  0.7992
 eGFP Temp23 - Cygb Temp23  -15.000 7.61 12  -1.972  0.5339
 Ngb Temp23 - Cygb Temp23    -3.667 6.80 12  -0.539  0.9991


P value adjustment: tukey method for comparing a family of 8 estimates
> View(summary_dat)
> View(emm)
> emmeans(fit, pairwise ~ Temp | Construct, adjust = "tukey")
$emmeans
Construct = Mock:
 Temp emmean   SE df lower.CL upper.CL
 20     84.0 5.89 12     71.2     96.8
 23     80.5 5.89 12     67.7     93.3

Construct = eGFP:
 Temp emmean   SE df lower.CL upper.CL
 20     86.5 5.89 12     73.7     99.3
 23     72.0 5.89 12     59.2     84.8

Construct = Ngb:
 Temp emmean   SE df lower.CL upper.CL
 20     75.3 4.81 12     64.9     85.8
 23     83.3 4.81 12     72.9     93.8

Construct = Cygb:
 Temp emmean   SE df lower.CL upper.CL
 20     79.3 4.81 12     68.9     89.8
 23     87.0 4.81 12     76.5     97.5
```

```
Confidence level used: 0.95

$contrasts
Construct = Mock:
 contrast        estimate   SE df t.ratio p.value
 Temp20 - Temp23     3.50 8.33 12   0.420  0.6818

Construct = eGFP:
 contrast        estimate   SE df t.ratio p.value
 Temp20 - Temp23    14.50 8.33 12   1.740  0.1074

Construct = Ngb:
 contrast        estimate   SE df t.ratio p.value
 Temp20 - Temp23    -8.00 6.80 12  -1.176  0.2624

Construct = Cygb:
 contrast        estimate   SE df t.ratio p.value
 Temp20 - Temp23    -7.67 6.80 12  -1.127  0.2818


Shapiro-Wilk normality test

data:  residuals(fit)
W = 0.97762, p-value = 0.8998
```

8 Acknowledgments

First of all, I would like to thank my supervisor, Dr. Andrej Fabrizius, for believing in me and giving me plenty of freedom to shape my project into what it is today. Thank you for always taking the time right away to answer my questions, correct and fix things, and for your strong optimism!

A big thank you also goes to my other supervisor, Prof. Dr. Kathrin Otte, for immediately agreeing to supervise me even though you don't work with globins. Nevertheless, our meetings helped me a lot, and I really appreciate it!

Next, I'd like to give a special thank you to my unofficial supervisor, Dr. Cäcilia Plate. Our great collaboration really helped move my project forward, and I've learned so much from you! Thank you also for your emotional support - it's something that can't be underestimated!

I'd also like to extend a special thank you to my former colleague, Dr. Raphael Koll. You introduced me to the world of HPC right from the start and were a lifesaver! Thank you for sharing your scripts and helping me with troubleshooting!

A huge thank you goes out to my colleagues Eda Dönmez and Juliane Schulte. You've always given me a reason to look forward to coming to the office. We've laughed a lot together, gotten upset together, and, above all, supported one another.

Of course, I'd also like to thank my great master's student, Lasse Roggenkamp, for the help with the project. I'd also like to thank my intern, Carina Lehmann, for the beautiful Circos plots.

I would like to thank my entire working group, especially the former Molecular Animal Physiology people!

Of course, a huge thank you goes to my partner, Dr. Jan Patrick Formanski - I probably wouldn't have been able to pull this off without you.

I would also like to thank the Department of Chemistry for providing good drinking water.

9 Affidavit

Affidavit:

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated. If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Eidesstattliche Versicherung:

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, 20th of July 2026



Soraya Safavi