# SCIENTIFIC REPORTS

**OPEN**

# Evolution of Hemoglobin Genes in Codfishes Influenced by Ocean Depth

Helle Tessand Baalsrud[1], Kjetil Lysne Voje[1], Ole Kristian Tørresen[1], Monica Hongrø Solbakken[1], Michael Matschiner[1,2], Martin Malmstrøm[1], Reinhold Hanel[3], Walter Salzburger[1,2], Kjetill S. Jakobsen[1] & Sissel Jentoft[1,4]

Understanding the genetic basis of adaptation is one of the main enigmas of evolutionary biology. Among vertebrates, hemoglobin has been well documented as a key trait for adaptation to different environments. Here, we investigate the role of hemoglobins in adaptation to ocean depth in the diverse teleost order Gadiformes, with species distributed at a wide range of depths varying in temperature, hydrostatic pressure and oxygen levels. Using genomic data we characterized the full hemoglobin (*Hb*) gene repertoire for subset of species within this lineage. We discovered a correlation between expanded numbers of Hb genes and ocean depth, with the highest numbers in species occupying shallower, epipelagic regions. Moreover, we demonstrate that the *Hb* genes have functionally diverged through diversifying selection. Our results suggest that the more variable environment in shallower water has led to selection for a larger *Hb* gene repertoire and that *Hb*s have a key role in adaptive processes in marine environments.

Understanding the genetic basis for how organisms adapt to specific environments is a fundamental challenge within evolutionary biology. The use of model systems allowing genetic manipulation has in many cases shown to be a powerful approach[1–3]. However, for non-model species - including deep-sea and cold-adapted teleost fish - with limited options for experimental manipulations and where little pre-existing knowledge is available, comparative genomics is a powerful route to apply[4, 5]. The teleost fish lineage Gadiformes – the codfishes – is a large group of 610 species with substantial ecological and economic impact distributed across the world in a wide variety of marine and freshwater environments. Most Gadiformes species are characterized as benthopelagic, i.e. neutrally buoyant in close association with the sea floor, and they inhabit almost every section of the continental slope, from shallow waters to the deep sea[6]. Fossil and phylogenetic evidence suggest that the ancestor of cod-fishes was a deep-sea fish, which later diversified into species inhabiting shallower waters[7–9]. Elucidation of biological traits crucial for such evolutionary transitions is not trivial, particularly because deep-sea adapted species are not well studied. The large latitudinal and vertical ranges in which Gadiformes reside represent a wide span of environmental conditions, including temperature, dissolved oxygen and hydrostatic pressure[10]. The hemoglobin (*Hb*) gene-family is a prime candidate for investigating molecular adaptation to depth because the relationship between the structure and function and how this links to environmental factors is well characterized.

The Hb protein is a key component of respiration, and consists of two $\alpha$- and $\beta$-globins subunits assembled as a tetramer. Teleost fish have evolved a more diverse *Hb* repertoire compared to other vertebrates due to being ectotherms and water-breathers, and thus exposed to a wider span in temperatures and oxygen availability[11]. The foundation for this diverse repertoire is partially associated with the teleost whole-genome duplication (TGD) 320–400 Ma[12], which resulted in two *Hb* clusters located on different chromosomes; the LA cluster and the MN cluster, respectively[11, 13]. Such duplication events provide raw evolutionary material allowing species to acquire new biological functions when additional copies are relieved from the functional constraints associated with the original gene[14, 15]. The genome sequencing of Atlantic cod (*Gadus morhua*)[16] confirmed that it contains four $\alpha$ genes and five $\beta$ genes[17], and this high multiplicity of *Hb* genes might increase its ability to respond to different

[1]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Oslo, Norway. [2]Institute of Fisheries Ecology, Johann Heinrich von Thünen-Institute, Federal Research Institute for Rural Areas, Forestry and Fisheries, Hamburg, Germany. [3]Zoological Institute, University of Basel, Basel, Switzerland. [4]Department of Natural Sciences, Centre for Coastal Research, University of Agder, Kristiansand, Norway. Correspondence and requests for materials should be addressed to S.J. (email: sissel.jentoft@ibv.uio.no)
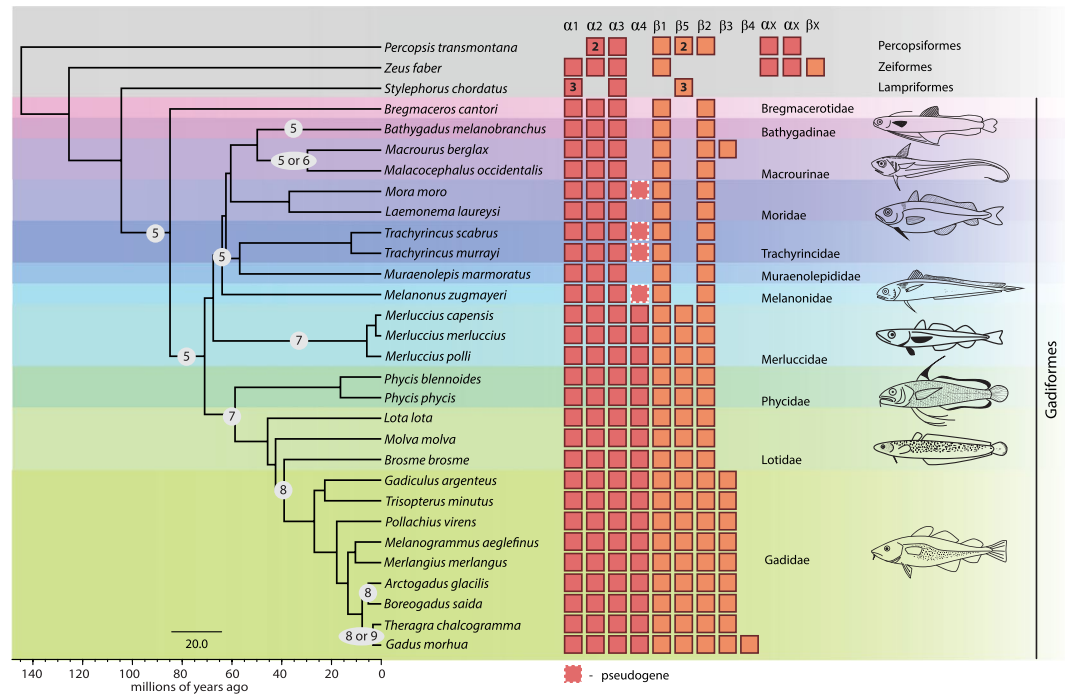
**Figure 1.** The repertoire of hemoglobin genes in the Gadiformes. The number of *Hb*s in 27 species of Gadiformes, as well as three outgroup species from Stylephoriformes, Zeiformes and Percopsiformes, are mapped onto a time-calibrated molecular phylogeny. This phylogeny is part of a larger teleost phylogeny presented in ref. 9. $\alpha$- and $\beta$-globin genes are indicated by boxes. Some species have more than one copy of a gene, which is indicated by a number. $\alpha$x and $\beta$x refer to $\alpha$- and $\beta$ genes that are not 1:1 orthologs to the gadiform *Hb* genes. The ancestral *Hb* copy number with the highest likelihood is indicated at nodes where there has been an evolutionary change, as well as any ambiguity (Supplementary Fig. 1). Time is given in million years. Fish illustrations drawn by Geir Holm are reprinted with permission from ref. 9.

environmental conditions[17, 18]. Furthermore, the *β1* gene displays a clear latitudinal gradient between two haplotypes, which may be linked to temperature adaptation[19, 20]. Hb could consequently be important for adaptation to the variety of environments occupied by the other gadiform species.

Here, we examine the evolutionary history of *Hb* in a wider phylogenetic perspective using low coverage (approx. 30x) genome sequencing to characterize the full *Hb* repertoire across the Gadiformes lineage. We selected 27 gadiform species, as well as 3 closely related outgroup species[9]: *Stylephorus chordatus* (closest living relative of Gadiformes[9]), *Zeus faber* and *Percopsis transmontanta*. Combining the extracted *Hb* gene repertoire with data on depth and latitude of occurrence we show that the expanded number of *Hb* genes found within this lineage is negatively correlated to depth, with the highest numbers in species living in shallower or epipelagic regions. Furthermore, *in silico* modeling of the Hb tetramer combined with comprehensive tests of natural selection revealed strong signs of diversifying selection on the surface of the protein structure indicating that the different variants have functionally diverged. Taken together, our findings suggest that the evolution of the hemoglobin gene repertoire has played a fundamental role in the Gadiformes' adaptation to the wide range of depths it occupies today.

## Results/Discussion

### Expansion of Hb Genes in Codfishes.
A full characterization of the *Hb* genes was obtained from the genome sequence data of the 30 selected species[9] (Supplementary Table 1). The number of *Hb* genes varied from five to nine in extant Gadiformes (Fig. 1). The ancestral state at the root of the Gadiformes lineage was estimated to five *Hb* genes (Supplementary Fig. 1), which indicates an overall expansion of the *Hb* repertoire within Merlucciidae, Phycidae, Lotidae and Gadidae (Fig. 1 and Supplementary Fig. 1). The observed variation in *Hb* gene number is indicative of gene duplications and/or gene deletions in different lineages. Construction of phylogenetic trees for all $\alpha$- (Fig. 2) and $\beta$-globin (Fig. 3) sequences, including sequences from *Danio rerio, Oreochromis niloticus, Gasterosteus aculeatus, Oryzias latipes, Salmo salar* and *Xenopus tropicalis*, the latter as an outgroup species, enabled proper gene annotation of the $\alpha$- and $\beta$-sequences. In concordance with previous phylogenies of teleost Hbs[11, 17] our gene trees reveal the dynamic nature of *Hb* family gene evolution. For the gadiform $\alpha$-globins, $\alpha$1, $\alpha$2, $\alpha$3 and $\alpha$4 form monophyletic groups (Fig. 2) with $\alpha$1, $\alpha$2, and $\alpha$4 showing high sequence similarity, which indicates recent gene duplications. However, due to high degree of sequence similarity in $\alpha$-globins across all teleosts, either as a result of homoplasy, gene conversion or sequence conservation, the phylogenetic relationship between the clades containing $\alpha$1, $\alpha$2, and $\alpha$4 could not be resolved with sufficient statistical support (i.e. bootstrap or posterior probabilities). Furthermore, it seems like $\alpha$4 has been lost, or pseudogenized,
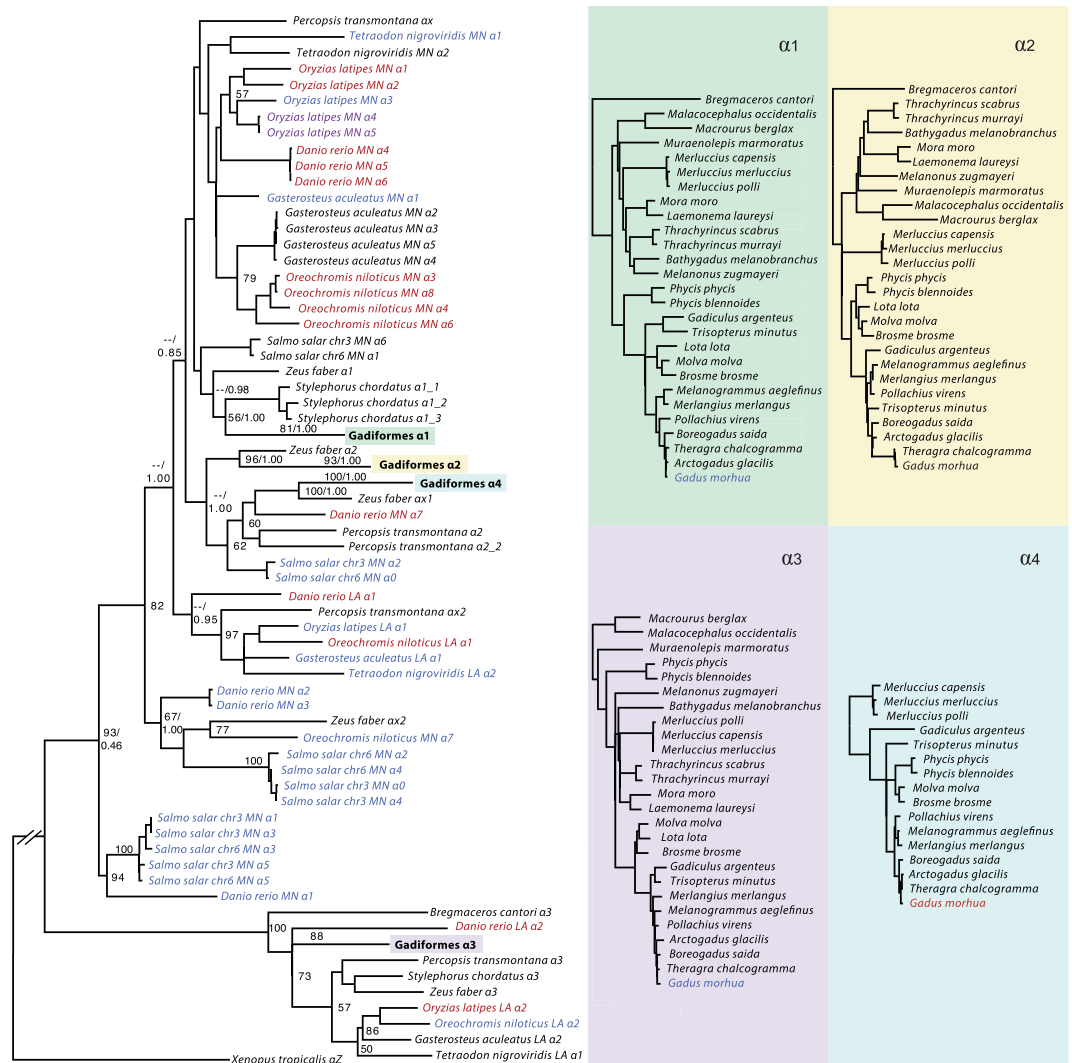
**Figure 2.** Phylogenetic relationships of $\alpha$-globin genes. ML phylogeny of $\alpha$-globin genes from 36 species of teleosts, and western clawed frog (*Xenopus tropicalis*) as the outgroup species. Numbers on nodes show bootstrap values and Bayesian posterior probabilities where topology is concordant, -- denotes support lower than 50/0.50. Sequences are colored according to timing of expression[11, 45]; embryonic (red), adult (blue), embryonic and adult (purple) and unknown (black). For each gadiform $\alpha$-globin gene the phylogenetic tree is shown separately. Some branches are shortened for convenience, which is indicated by gaps.

as indicated by premature stop codons or frame-shifting indels in the clade flanked by *Bregmaceros cantori* and *Melanonus zugmayeri* (Fig. 1). The phylogeny of the $\beta$-sequences suggests that $\beta1$, the ancestor of $\beta2/3/4$ and $\beta5$, represents the ancestral $\beta$ genes in Gadiformes (Fig. 3). Additionally, the gene tree indicates that $\beta2$ has been duplicated and giving rise to $\beta3$ in Gadidae – and again independently duplicated in *Macrourus berglax* – implying that $\beta3$ in *Macrourus berglax* and Gadidae are not strict 1:1 orthologs (Fig. 3). Within Gadidae, $\beta2$ seems to have been duplicated independently to form $\beta3$ in *Gadiculus argentus* and *Trisopterus minutus*, respectively. This may call for renaming of $\beta3$ in these species. In *Gadus morhua* $\beta3$ has been duplicated giving rise to $\beta4$. Overall, the observed expansion of *Hb* genes in Gadiformes relative to the ancestral state is indicative of an increased repertoire of combinatory Hb tetramers, likely contributing to a high respiratory plasticity.

According to the birth-death model of gene family evolution, new genes are often redundant or deleterious, and are consequently pseudogenized or lost[14, 21]. If the new gene confers some advantage to the organism either by increasing gene dosage or acquiring a new function/sub-function, it will be maintained by natural selection. Thus, duplicated *Hb* genes could have slightly different oxygen-carrying properties due to diversifying or positive selection, i.e. evolutionarily optimized for functionality at different environmental conditions. To address this, we tested whether the identified *Hb* genes within the Gadiformes have diversified faster than expected by chance with multiple site-specific tests (SLAC, FEL and REL) for natural selection based on the nonsynonymous (dN)/synonymous (dS) substitution rate ratio within the different *Hb* genes. For $\alpha1$, $\alpha2$, $\alpha3$, $\alpha4$, $\beta1$ and $\beta5$ many sites were under diversifying selection (dN/dS > 1) (Fig. 4, Supplementary Table 3). For the three tests of natural selection (REL, FEL and SLAC, respectively) that were carried out, not all tests reported the same sites (Fig. 4,
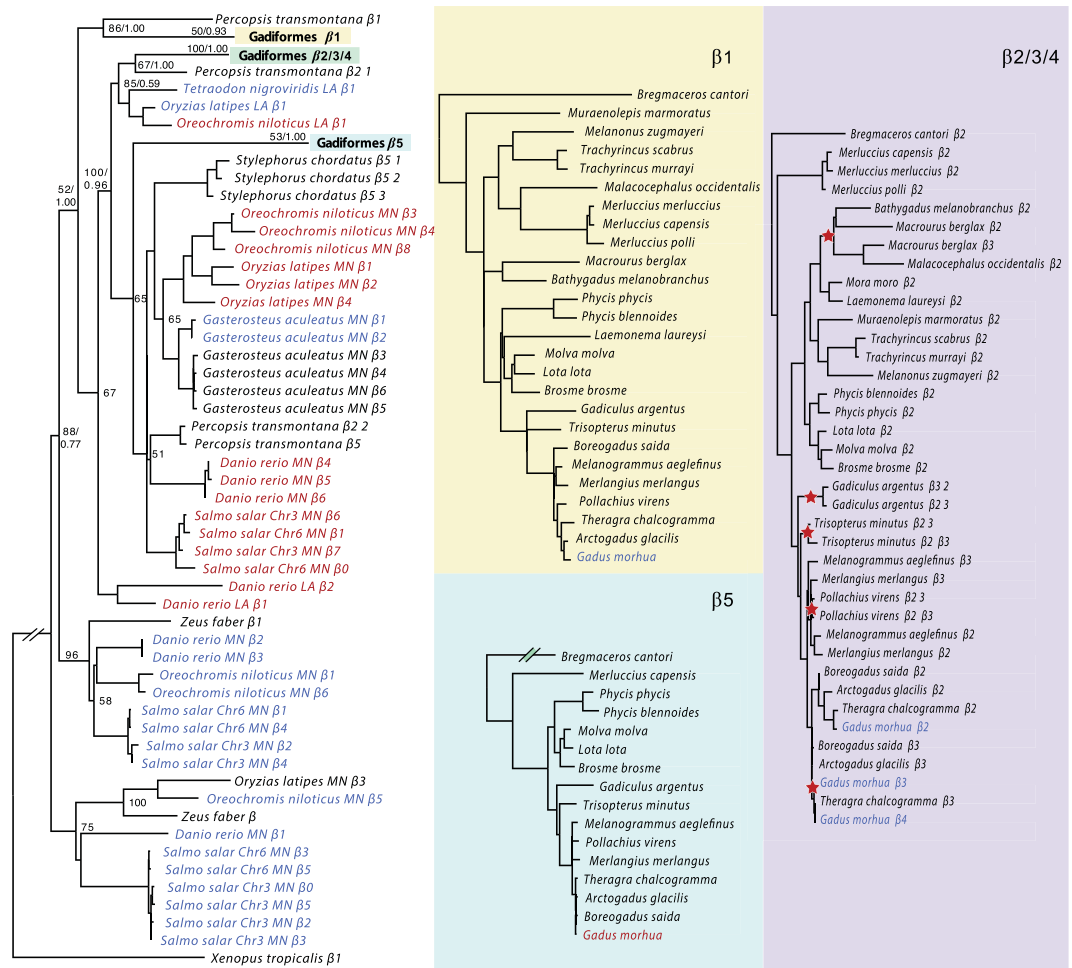
**Figure 3.** Phylogenetic relationships among $\beta$-globin genes. ML phylogeny of $\beta$-globin genes from 36 species of teleosts, and western clawed frog (*Xenopus tropicalis*) as the outgroup species. Numbers on nodes show bootstrap values and Bayesian posterior probabilities where topology is concordant, -- denotes support lower than 50/0.50. Sequences are colored according to timing of expression[11, 45]; embryonic (red), adult (blue) and unknown (black). For each gadiform $\beta$-globin gene the phylogenetic tree is shown separately. Lineage specific duplications of $\beta2$ are indicated by red stars. Some branches are shortened for convenience, indicated by gaps.

Table S3) due to different underlying models and assumptions affecting evolutionary inference. The SLAC test is the most conservative with a low false positive rate, however, SLAC sometimes misses sites that are under selection. The less stringent tests (FEL, followed by REL) usually identify more sites, with the cost of a higher type I error-rate[22]. These tests can either be used on species level phylogenies or on phylogenetic gene trees. Because species and gene phylogenies might not have the exact same topology, we performed the tests at both levels. Notably, many sites were consistently reported using either a species tree or a gene tree by two or all of the three tests. These sites are therefore likely to be the most important for the evolution of divergent functions (Fig. 4, Table S3). Moreover, most sites reported under natural selection were found to be under purifying selection (dN/dS < 1) for all genes, as many domains are conserved to uphold the function of $O_2$ transport (Table S3). Since $\beta2$, $\beta3$ and $\beta4$ are more similar due to a recent duplication (Fig. 3), uncertainties regarding true 1:1 orthology did not allow for dN/dS tests.

Sites showing diversifying selection are indicative of sub-functionalization of the different *Hb* genes. To further evaluate whether the sites identified to be under diversifying selection have any potential impact on the actual function of the *Hb* tetramer, we constructed a structural protein model based on *G. morhua* hemoglobin sequences and plotted these sites onto the structure (Fig. 4). Most of the sites in gadiform *Hb* genes appear to be on the surface of the tetramer (Fig. 4), which again supports sub-functionalization, as substitutions on the surface of the protein affect Hb subunit interactions and affinities, and are thus likely to influence its $O_2$ binding capacity under different environmental conditions (reviewed in ref. 23).

**The Hb Copy Number is Inversely Correlated with Depth.** To investigate the relationship between the number of *Hb* genes observed and environmental factors – e.g. the geographical and vertical distribution of the Gadiformes (Table 1) – we used a phylogenetic comparative approach implemented in the R program SLOUCH[24]. For vertical distribution, we chose to use maximum depth (Table 1) as an indicator of depth of usual occurrence
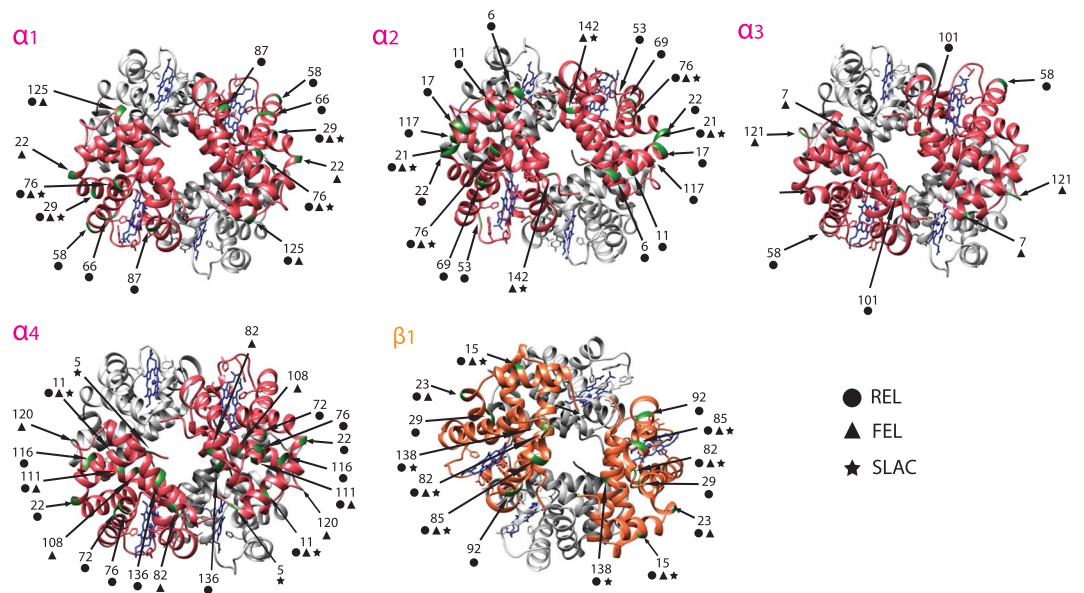
**Figure 4.** Sites under diversifying selection at the surface of hemoglobin tetramers. *In silico* models of the hemoglobin tetramers, based on sequences from Atlantic cod (*Gadus morhua*). $\alpha$ sequences are highlighted in pink, and $\beta$ sequences in orange, with respective gene names shown. Three different tests (indicated by symbols according to the key) were used to test for diversifying selection; REL, FEL and SLAC, respectively. Arrows point to sites under diversifying selection, which are also highlighted in green.

| Species | Latitude (°N) | | | Depth (m) | | |
|---|---|---|---|---|---|---|
| | South | North | Range | Min | Max | Range |
| *Gadus morhua* | 35 | 83 | 48 | 0 | 600 | 600 |
| *Arctogadus glacilis* | 69 | 87 | 18 | 0 | 1000 | 1000 |
| *Boreogadus saida* | 52 | 87 | 35 | 0 | 400 | 400 |
| *Trisopterus minutus* | 28 | 68 | 40 | 1 | 440 | 439 |
| *Pollachius virens* | 33 | 77 | 44 | 37 | 364 | 327 |
| *Melanogrammus aeglefinus* | 35 | 79 | 44 | 10 | 450 | 440 |
| *Merlangius merlangus* | 35 | 72 | 37 | 10 | 200 | 190 |
| *Theragra chalcogramma* | 34 | 68 | 34 | 0 | 1280 | 1280 |
| *Gadiculus argentus* | 24 | 74 | 50 | 100 | 1000 | 900 |
| *Phycis phycis* | 13 | 45 | 32 | 13 | 614 | 601 |
| *Molva molva* | 35 | 75 | 40 | 100 | 1000 | 900 |
| *Lota lota* | 40 | 78 | 38 | 1 | 700 | 699 |
| *Brosme brosme* | 37 | 83 | 46 | 18 | 1000 | 982 |
| *Merluccius merluccius* | 18 | 76 | 58 | 70 | 1000 | 930 |
| *Merluccius capensis* | 11 | 37 | 26 | 50 | 1000 | 950 |
| *Merluccius polli* | −19 | 29 | 48 | 50 | 910 | 860 |
| *Melanonus zugmayeri* | −49 | 60 | 109 | 0 | 3000 | 3000 |
| *Macrourus berglax* | 37 | 82 | 45 | 100 | 1000 | 900 |
| *Malacocephalus occidentalis* | −37 | 43 | 80 | 140 | 1945 | 1805 |
| *Bathygadus melanobranchus* | −34 | 53 | 87 | 400 | 2600 | 2200 |
| *Muraenolepis marmoratus* | −56 | −44 | 12 | 30 | 1600 | 1570 |
| *Bregmaceros cantori* | NA | NA | NA | 450 | 475 | 25 |
| *Mora moro* | −51 | 64 | 115 | 450 | 2000 | 1550 |
| *Laemonema laureysi* | −8 | 8 | 16 | 200 | 618 | 418 |
| *Trachyrincus murrayi* | NA | NA | NA | 0 | 1630 | 1630 |
| *Phycis blennoides* | 20 | 71 | 51 | 10 | 1047 | 1037 |
| *Trachyrincus scabrus* | −27 | 55 | 82 | 395 | 1700 | 1305 |

**Table 1.** The latitudinal distribution and depth of occurrence for the species of Gadiformes included in this study. Latitude is given in degrees north (°N) for the most southern and most northern observation for each species, respectively, as well as the range. Depth of occurrence is given in meters (m), with minimum and maximum depth recorded for each species, including range.

since benthopelagic species usually swim near the bottom[25]. In the best model (according to its AIC$_c$-score) *Hb* copy numbers have evolved towards optima that are functions of maximum depth. This model explained about 28% of the variance in *Hb* copy number (Supplementary Table 2) and returned an optimal regression slope of −0.018, which implies a reduction of two copies with every 100 meters added to maximum depth. However, species are far from realizing this optimal relationship between depth and copy number since evolution of copy number towards the maximum depth optimum is extremely slow (Supplementary Table 2). Additionally, the southernmost latitudinal distribution of the species (Table 1) explained about 25% of the variance in *Hb* copy number, but that model has a substantially higher (i.e. worse) AIC$_c$ score compared to the model using only maximum depth as predictor (Supplementary Materials and Methods).

The significant decline in number of gadiform *Hb* genes with depth of occurrence (Supplementary Table 2) suggests that the diversification from ancestral deep-sea habitats into more shallow zones has been facilitated by an expansion of *Hb* genes within the Gadiformes. Our analyses therefore suggest that the number of *Hb* genes is a result of adaptive evolution, albeit slow, and that the optimal number of *Hb* genes in different branches of the phylogeny is related to maximum depth, with a broader *Hb* repertoire in species living in more shallow waters. However, given that 72% of the variation remains unexplained in our best model, other factors than maximum depth may be important for the evolution of *Hb* copy number. Furthermore, an Ornstein–Uhlenbeck process is the simplest stochastic model that allows evolution toward a specific state, and we acknowledge that the evolution of *Hb* copy number in the Gadiformes may not be perfectly described by this model. Nevertheless, 28% is a high number given the complexity of the interaction between ecological factors and evolution and diversification of codfishes over millions of years.

It should be emphasized that it might not be depth *per se* that is the environmental factor affecting the evolution of *Hb* copy number evolution. Suggestively, other abiotic factors that correlate with depth might be the causal drivers, such as hydrostatic pressure and temperature, which are both known to influence the function of the Hb tetramer[26]. Hydrostatic pressure increases with roughly 1 atm every 10 meters and temperature generally drops with depth[10]. Deep-sea fishes are thus exposed to enormously high pressure and low temperatures (0–6 °C) which require special adaptations physiologically and at the molecular level[27], including the Hb repertoire[26]. However, deep-sea adaptation may not require a broad repertoire of *Hb* genes since it represents an extreme, yet stable environment. Furthermore, decreasing temperature increases Hbs O$_2$ affinity, while metabolic rate is negatively correlated with depth; this implies that oxygen demands are readily met[28] with a relatively narrow Hb repertoire. In contrast, species exposed to a wider range of depths are dependent on a repertoire of Hbs with O$_2$ affinities optimized for different temperatures and pressures, especially in shallow zones with highly fluctuating temperatures.

The adaptive significance of *Hb* multiplicity in teleost fishes in general is not clear-cut, and the link between multiplicity and environmental variation has been questioned by some reports[29, 30]. There are, however, few experimental studies comparing closely related species in a phylogenetic context. One such study conducted on triplefin fishes (family Tripterygiidae) demonstrated that species in shallow waters have a greater diversity of Hbs and a higher O$_2$ affinity and reduced pH sensitivity than species living in deeper and more temperature stable habitats[31]. In concordance with these findings, goldfish (*Carassius auratus*) express a larger Hb repertoire when acclimated to a more variable temperature regime[32]. Antarctic Notothenioid fishes are cold-water specialists – adapted to a thermally stable climate – displaying an extremely narrow Hb repertoire, even a complete loss in the Channichthyidae family (reviewed in ref. 33). On the other hand, in mullets no connection was found between Hb multiplicity and temperature variability[34]. Thus, the adaptive role of the size of the Hb repertoire in different codfishes may not be obvious. Demonstration of diversifying selection acting on several *Hb* genes, however, suggests that they are evolutionarily and functionally fine-tuned for different environmental conditions, also supported by the finding that most of these sites are on the surface of the Hb tetramer (Fig. 4). Numerous studies have demonstrated that *Hb* evolution is driven by temperature adaptation, which is not surprising given the fact that O$_2$ affinity is negatively correlated with temperature[35]. Notably, Campbell *et al.*[36] found several surface substitutions on the chimeric β/δ-globin subunit of mammoth hemoglobin that have a large phenotypic effect and contribute to O$_2$ offloading at cold temperatures. In Atlantic cod two linked substitutions on the β1 gene with potentially the same effect has been identified in populations that are more cold-tolerant [20], although a recent report questions whether these substitutions are linked to temperature adaptation[37]. Further, equivalent substitutions have been linked to adaptation to altitudinal ranges in North American pikas[38] and deer mice[39].

Temperature and hydrostatic pressure are obvious candidates underlying depth adaptation, however, there could be other selective drivers correlated with depth shaping *Hb* copy number. Expressing different Hb tetramers could increase Hb solubility and thus cellular Hb concentration resulting in a higher O$_2$ carrying capacity of the blood[30]. This could be of importance to species living in more shallow waters, which generally have higher metabolic rates[28]. Moreover, Hb heterogeneity is shown to boost cell longevity as well as affect cellular metabolism[30]. Biotic factors such as pathogen load may also play a role in depth adaptation, as the microbial community varies at different depths [8, 40]. Hbs are known to have immune-related functions, such as Hb-derived antimicrobial peptides found in the channel catfish[41] and humans[42], or killing microbes directly by creating reactive oxygen species[43]. Further, many fish species are also exposed to diverse environments at different life stages. This likely promotes subfunctionalization due to various paralogs being differentially expressed during embryonic and larval development. Hemoglobin genes in teleosts do not cluster together phylogenetically according to the developmental timing of their expression (Figs 2 and 3)[11], thus the evolution of *Hb* regulation seems to be less constrained in teleosts compared to birds and mammals[44]. In Atlantic cod all nine variants are expressed throughout their lifespan, however, α1, β1, α2 and β2 are the most common adult globins, whereas β5 and α4 are more highly expressed in juvenile fish[45, 17, 45]. In addition to timing of expression, neo-/subfunctionalization could involve expression in novel tissues, or different allosteric regulation[35]. Perhaps the most famous example of Hb subfunctionalization is the pH-dependent reduction in O$_2$ carrying capacity known as the Root effect – involved in retinal oxygenation and swim bladder inflation (reviewed in ref. 46), but also generally enhances O$_2$ delivery during stress[47].

**The Evolutionary Transition from Obligate Deep-Sea Habitats to Various Depths and Shallow Zones Promoted a Large Hb Repertoire and Ecological Speciation.** To our knowledge, no experiments have so far been carried out on deep-sea organisms to address the function of Hb in relation to hydrostatic pressure. However, Noble *et al.*[48] investigated the functional properties of Hbs in five gadiform species (belonging to the Moridae and Macrouridae families) – occupying various depths – using carbon monoxide (CO) affinity as a proxy for $O_2$ affinity under high hydrostatic pressure. They found that deep-sea species have heme groups with a much lower CO affinity than more shallow-water species, most likely an adaptation to allow pumping of $O_2$ into the swim bladder under high-pressure conditions. On a broader scale, studies of other proteins show that functional properties in deep-sea species are basically unaffected by high pressure compared to species not occupying high-pressure environments[27, 49]. This suggests that the functionality of proteins in deep-sea organisms has evolved to withstand hydrostatic pressure. Accordingly, we postulate that during the evolutionary transition from obligate deep-sea habitats to more shallow zones, gadiform species have undergone an adaptive selection towards a broader Hb repertoire functionally optimized to low pressure as well as coping with fluctuating temperatures. Moreover, the specialized adaptation to an array of different ocean depths has most likely promoted ecological speciation within the Gadiformes. Ecological speciation along environmental gradients is seen as the major mode of speciation where strict geographic isolation is generally non-existent, such as in marine environments[50, 51]. In deep-sea populations, differentiation and speciation can occur along environmental gradients, attenuating with depth. This has been shown in bivalves[52], corals[53] and rockfish[54]. The weak relationship between number of *Hb* genes and geographical distribution observed in this dataset is in line with other studies that support isolation by depth as a stronger driver of population differentiation and ultimately speciation than isolation by distance in deep-sea taxa[52, 54].

**Hb Gene Duplications and the High Diversification Rates in Codfishes.** Investigating the evolutionary and ecological changes accompanying the transition from deep sea to more shallow waters is challenging as little is known about the ecology of deep-sea teleosts[10]. Here, we have shown that comparative genomic studies can provide vital insight into the evolutionary history of lineages where data is otherwise scarce. To conclude, we demonstrate that the evolution of the *Hb* gene repertoire in Gadiformes is characterized by gene duplications as well as losses, accompanied with high degree of gene diversification indicative of subfunctionalization. This is influenced by ocean depth and putatively linked to adaptation to temperature and hydrostatic pressure. Ultimately, our data suggest that hemoglobin plays an important role in the evolutionary puzzle explaining the diversification of Gadiformes, which has one of the highest diversification rates of teleosts[9] and is one of the most species-rich teleost clades, with species displaying a wide distribution with respect to geographic and vertical range.

## Materials and Methods

**Specimen collection.** Working with animals we always aim to limit the effect our research afflict populations and individuals. Whenever possible we collaborate with other sources, such as commercial fisheries or museums. This way, no animals need to be euthanized to serve our scientific purpose alone. The tissue samples used in this study are either from museum specimen or commercially fished individuals intended for human consumption. The commercially caught fish were immediately stunned by bleeding, following standard procedures by a local fisherman. Sampling in this manner does not fall under any specific legislation in Norway, but it is in accordance with the guidelines set by the 'Norwegian consensus platform for replacement, reduction and refinement of animal experiments' (www.norecopa.no). For more information regarding the samples see[9].

**Whole-genome sequencing.** We selected 27 species, which represent most of the lineages in the Gadiformes order, in addition to its closest living relatives, *Stylephorus chordatus*, *Zeus faber* and *Percopsis transmontanta*[9, 55]. We sequenced paired end libraries with an average insert size of 350 bp ($2 \times 150$ bp reads on Illumina HiSeq. 2000) with coverage ranging from 18 to 40x (average coverage 28x). This sequence strategy gives contigs spanning the average median gene-length[56], making it ideal for finding and identifying genes, but without substantial gene-order information (synteny). The Celera assembler[57] was used to assemble the genomes, with contig N50 ranging from 3.1 to 8.1 kb with an average of 4.1 kb. CEGMA[58] and BUSCO[59] were used to evaluate gene completeness; CEGMA gave, on average, complete or partial hits for 69% of the conserved eukaryotic genes included in the CEGMA analysis and BUSCO gave, on average, 68% of the conserved genes belonging to the Actinopterygii clade in the BUSCO analysis. A list of species with relevant genome statistics is given in Supplementary Table 1. For further information regarding the sequencing see[9, 55].

**Gene mining and annotation.** *Hb* genes were annotated by tBLASTn[60] searches with known *Hb* sequences from *Gadus morhua*, *Oryzias latipes*, *Tetraodon nigroviridis*, *Oreochromis niloticus*, *Gasterosteus aculeatus, Salmo salar* and *Danio rerio* (annotation and nomenclature following[11]). For paralogous genes that have recently been duplicated or are similar due to gene conversion, gene copies can collapse in the assembly process. In contrast, with polymorphic genes alleles could be misjudged as copies. However, by manually inspecting alignments of intronic sequences it was possible to distinguish paralogous gene copies from alleles.

**Phylogenetic tree construction.** To identify orthologous *Hb* sequences phylogenetic gene trees were constructed, α and β sequences were analyzed separately. Amino acid sequences were aligned using ClustalW[61] as implemented in MEGA7[62] with default settings for all species (alignments of α and β sequences are in Supplementary Data 1). Using the model selection tool in MEGA7 we determined that the best model (i.e. having the lowest AIC score) for molecular evolution was TN93 + G + I for α-sequences and GTR + G + I for β-sequences. Phylogenetic trees were constructed based on codon triplets using maximum likelihood (ML)

implemented in MEGA7 and a Bayesian method in MrBayes 3.2.2[63]. A ML tree was constructed based on the models of molecular evolution stated above, with 1000 bootstrap replicates. Bayesian trees were run using standard priors, with four chains of simulations for $1 \times 10^7$ generations sampling every $1 \times 10^3$ generation. The GTR + G + I model was used for both α and β as the TN93 + G + I is not available in this program. A given run was considered to have reached convergence when the likelihood scores leveled off asymptotically. All trees sampled prior to convergence were discarded and support (posterior probability) was calculated based on a consensus of the last 2250 trees. Previous work on teleost *Hb*s shows that *Hb* from the frog *Xenopus tropicalis* is clearly outside the clade constituting teleost *Hb*s[11], therefore it was chosen as an outgroup species.

The identified α and β genes were then mapped on a phylogenomic species tree based on 567 exons of 111 genes, selected after stringent filtering for single-copy orthologous markers. Branching times were estimated in BEAST v.2.2[64] using a relaxed clock model and 17 fossil constraints. This phylogeny is a modified version from[9], which describes the procedures in more detail.

### Ancestral reconstruction of the number of Hbs.
The ancestral reconstruction of number of *Hbs* in gadiformes was estimated using the function ace implemented in the R package APE[65]. *Percopsis transmontana*, *Zeus faber* and *Stylephorus chordatus* were not included as many of the *Hbs* found in these species are not 1:1 orthologs with gadiform *Hbs*. We used maximum likelihood estimation of the ancestral state for discrete characters with three different models, an equal rates model (ER), an all rates different model (ARD) and a symmetrical model (SYM), goodness of fit was estimated using a Chi-square test. All statistics was carried out in R v3.1.3[66].

### Phylogenetic comparative analyses.
We used a phylogenetic comparative method called SLOUCH (Stochastic Linear Ornstein-Uhlenbeck models for Comparative Hypotheses)[24, 67–70] implemented in R v3.1.3[66], to investigate whether the number of *Hb* genes has evolved as a response to changes in maximum depth and latitude, respectively (data was obtained for the different species in the global information system FishBase[6]). The assumed model of trait evolution (trait is here the number of *Hb* genes) is an Ornstein-Uhlenbeck (OU) process, where the trait evolves towards an optimum that is assumed to be a linear function of a predictor *x*, as $\theta = a + b_a x$, the regression parameters are informative of the relationship between the optimum and the trait. The deterministic pull of the trait towards the optimum is can be quantified with the phylogenetic half-life, $t_{1/2} = \frac{\ln 2}{\alpha}$, the average time it takes for a species to move half the way from an ancestral state to a new optimum i.e. a half-life above zero indicates adaptation is not immediate. SLOUCH returns an "optimal regression", which represents the best fit of the estimated primary optimum[67] on *Hb* copy number. In other words, this optimal regression describes the expected relationship between the number of *Hb* genes and the predictor in the model if adaptation was instantaneous (i.e. there are no constraints on the evolution of number of *Hb* genes towards the optimal state). A model that includes a predictor variable can be contrasted with an intercept-only model where no predictor variables are included. Phylogenetic effect is a measure of how well the phylogeny alone explains the distribution of the trait (number of *Hb* genes). Model comparisons are done using the small sample-size corrected version of Akaike information criterion (AICc).

### Analyses of natural selection.
For each Hb gene translated amino acid sequences from all species available for that gene in the dataset were aligned following same procedure as described above (alignments presented in Supplementary Data 1). To test for diversifying and purifying selection we used the SLAC, FEL and REL analyses[22] as implemented in the Hyphy software package on the Datamonkey server (www.datamonkey.org) and using the phylogenies in Fig. 1 (referred to as the species tree), and Figs 2 and 3 (referred to as the gene trees).

### Homology model building.
A 3D protein model was created using the SWISS-MODEL Workspace and the DeepView software[71] for *Gadus morhua* Hb-I (α1 and β1) based on homology. A template search was carried out in the SWISS-MODEL Workspace, identifying hemoglobin from *Trematomus bernacchii* (Protein Data Bank (PDB) code 1HBH) as the best template. *Gadus morhua* α1, α2, α3, α4 and β1, respectively were aligned to the template in DeepView, the alignment was then submitted to the SWISS-MODEL Workspace under project mode. The automated modeler procedure gave one model with high quality (QMEAN4 = 1.34) of a Hb tetramer with two β1 units, and two alpha units of either α1, α2, α3 or α4. This gave four different Hb tetramers in total, which are all shown in Supplementary Data 2.

## References

1. Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nat Rev Genet* **14**, 168–178 (2013).
2. Mackay, T. F. C. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nat Rev Genet* **15**, 22–33 (2013).
3. Hedges, S. B. The origin and evolution of model organisms. *Nat Rev Genet* **3**, 838–849 (2002).
4. Ekblom, R. & Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1–15 (2010).
5. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29**, 51–63 (2014).
6. Eschemeyer, W. N. & Fricke, R. Catalog of fishes. http://research.calacademy.org/researchichthyology/catalog/fishcatmain.asp. Available at: (Accessed: 4 November 2014)
7. Kriwet, J. & Hecht, T. A review of early gadiform evolution and diversification: first record of a rattail fish skull (Gadiformes, Macrouridae) from the Eocene of Antarctica, with otoliths preserved *in situ*. *Naturwissenschaften* **95**, 899–907 (2008).
8. Star, B. & Jentoft, S. Why does the immune system of Atlantic cod lack MHC II? *BioEssays* **34**, 648–651 (2012).
9. Malmstrøm, M. *et al.* Evolution of the immune system influences speciation rates in teleost fishes. *Nat Genet* **48**, 1204–1212 (2016).
10. Robison, B. H. Deep pelagic biology. *Journal of Experimental Marine Biology and Ecology* **300**, 253–272 (2004).
11. Opazo, J. C., Butts, G. T., Nery, M. F., Storz, J. F. & Hoffmann, F. G. Whole-genome duplication and the functional diversification of teleost fish hemoglobins. *Molecular Biology and Evolution* **30**, 140–153 (2013).

12. Amores, A. *et al*. Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**, 1711–1714 (1998).
13. Hardison, R. C. Globin genes on the move. *J. Biol.* **7**, 35 (2008).
14. Ohno, S. *Evolution by gene duplication*. (Springer-Verlag, 1970).
15. Hughes, A. L. Gene duplication and the origin of novel proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8791–8792 (2005).
16. Star, B. *et al*. The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207–210 (2011).
17. Borza, T., Stone, C., Gamperl, A. K. & Bowman, S. Atlantic cod (*Gadus morhua*) hemoglobin genes: multiplicity and polymorphism. *BMC Genetics* **10**, 51 (2009).
18. Brix, O., Thorkildsen, S. & Colosimo, A. Temperature acclimation modulates the oxygen binding properties of the Atlantic cod (Gadus morhua L.) genotypes—HbI*1/1, HbI*1/2, and HbI*2/2—by changing the concentrations of their major hemoglobin components (results from growth studies at different temperatures. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* **138**, 241–251 (2004).
19. Karpov, A. K. & Novikov, G. G. Hemoglobin alloforms in cod, *Gadhus morhua* (Gadiformes, Gadidae), their functional characteristics and occurrence in populations. *Journal of Ichthyology* **20**, 45–50 (1980).
20. Andersen, Ø. *et al*. Haemoglobin polymorphisms affect the oxygen-binding properties in Atlantic cod populations. *Proc. Biol. Sci.* **276**, 833–841 (2009).
21. Demuth, J. P. & Hahn, M. W. The life and death of gene families. *BioEssays* **31**, 29–39 (2009).
22. Poon, A. F. Y., Frost, S. D. W. & Pond, S. L. K. Detecting signatures of selection from DNA sequences using Datamonkey. *Methods in molecular biology (Clifton, N.J.)* **537**, 163–183 (2009).
23. Storz, J. F. & Moriyama, H. Mechanisms of Hemoglobin Adaptation to High Altitude Hypoxia. *High Altitude Medicine & Biology* **9**, 148–157 (2008).
24. Hansen, T. F., Pienaar, J. & Orzack, S. H. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* **62**, 1965–1977 (2008).
25. Koslow, J. A. Energetic and life-history patterns of deep-sea benthic, benthopelagic and seamount-associated fish. *Journal of Fish Biology* **49**, 54–74 (1996).
26. Hourdez, S. & Weber, R. Molecular and functional adaptations in deep-sea hemoglobins. *Journal of Inorganic Biochemistry* **99**, 130–141 (2005).
27. Somero, G. N. Adaptations to High Hydrostatic-Pressure. *Annual Review of Physiology* **54**, 557–577 (1992).
28. Graham, M. S., Haedrich, R. L. & Fletcher, G. L. Hematology of three deep-sea fishes: a reflection of low metabolic rates. *Comp Biochem Physiol A Comp Physiol* **80**, 79–84 (1985).
29. Wells, R. M. G. *Chapter 6 - Blood-Gas Transport and Hemoglobin Function: Adaptations for Functional and Environmental Hypoxia*. *Hypoxia* **27**, 255–299 (Elsevier Inc., 2009).
30. Ingermann, R. L. Vertebrate Hemoglobins in *Handbook of Physiology* (John Wiley & Sons, Inc., 1997).
31. Brix, O., Clements, K. D. & Wells, R. M. G. Haemoglobin components and oxygen transport in relation to habitat distribution in triplein fishes (Tripterygiidae). *J Comp Physiol B* **169**, 329–334 (1999).
32. Houston, A. H. & Gingrasbedard, J. H. Variable Versus Constant-Temperature Acclimation Regimes - Effects on Hemoglobin Isomorph Profile in Goldfish, Carassius-Auratus. *Fish Physiol. Biochem.* **13**, 445–450 (1994).
33. di Prisco, G., Eastman, J. T., Giordano, D., Parisi, E. & Verde, C. Biogeography and adaptation of Notothenioid fish: Hemoglobin function and globin–gene evolution. *Gene* **398**, 143–155 (2007).
34. Olianas, A. *et al*. Striped mullet (Mugil cephalus) hemoglobin system: multiplicity and functional properties. *J Comp Physiol B* **181**, 187–197 (2010).
35. Weber, R. E. & Campbell, K. L. Temperature dependence of haemoglobin-oxygen affinity in heterothermic vertebrates: mechanisms and biological significance. *Acta Physiologica* **202**, 549–562 (2010).
36. Campbell, K. L. *et al*. Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nat Genet* **42**, 536–540 (2010).
37. Barlow, S. L., Metcalfe, J., Righton, D. A. & Berenbrink, M. Life on the edge: O2 binding in Atlantic cod red blood cells near their southern distribution limit is not sensitive to temperature or haemoglobin genotype. *Journal of Experimental Biology* **220**, 414–424 (2017).
38. Tufts, D. M. *et al*. Epistasis Constrains Mutational Pathways of Hemoglobin Adaptation in High-Altitude Pikas. *Molecular Biology and Evolution* **32**, 287–298 (2015).
39. Natarajan, C. *et al*. Epistasis among adaptive mutations in deer mouse hemoglobin. *Science* **340**, 1324–1327 (2013).
40. Orcutt, B. N., Sylvan, J. B., Knab, N. J. & Edwards, K. J. Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol. Mol. Biol. Rev.* **75**, 361–422 (2011).
41. Ullal, A. J., Wayne Litaker, R. & Noga, E. J. Antimicrobial peptides derived from hemoglobin are expressed in epithelium of channel catfish (Ictalurus punctatus, Rafinesque). *Developmental and Comparative Immunology* **32**, 1301–1312 (2008).
42. Liepke, C. *et al*. Human hemoglobin-derived peptides exhibit antimicrobial activity: a class of host defense peptides. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **791**, 345–356 (2003).
43. Jiang, N., Tan, N. S., Ho, B. & Ding, J. L. Respiratory protein–generated reactive oxygen species as an antimicrobial strategy. *Nat Immunol* **8**, 1114–1122 (2007).
44. Storz, J. F. Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet* **17**, 239–250 (2016).
45. Wetten, O. F. *et al*. Genomic organization and gene expression of the multiple globins in Atlantic cod: conservation of globin-flanking genes in chordates infers the origin of the vertebrate globin clusters. *BMC Evolutionary Biology* **10**, 315 (2010).
46. Brittain, T. Root effect hemoglobins. *Journal of Inorganic Biochemistry* **99**, 120–129 (2005).
47. Rummer, J. L., McKenzie, D. J., Innocenti, A., Supuran, C. T. & Brauner, C. J. Root effect hemoglobin may have evolved to enhance general tissue oxygen delivery. *Science* **340**, 1327–1329 (2013).
48. Noble, R. W. *et al*. Functional-properties of hemoglobins from deep-dea fish - correlations with depth distribution and presence of a swimbladder. *Biochim. Biophys. Acta* **870**, 552–563 (1986).
49. Somero, G. N. Biochemical ecology of deep-sea animals. *Experientia* **48**, 537–543 (1992).
50. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421**, 259–264 (2003).
51. Puebla, O. Ecological speciation in marine v.freshwater fishes. *Journal of Fish Biology* **75**, 960–996 (2009).
52. Jennings, R. M., Etter, R. J. & Ficarra, L. Population differentiation and species formation in the deep sea: the potential role of environmental gradients and depth. *PLoS ONE* **8**, e77594 (2013).
53. Prada, C. & Hellberg, M. E. Long prereproductive selection and divergence by depth in a Caribbean candelabrum coral. *Proceedings of the National Academy of Sciences* **110**, 3961–3966 (2013).
54. Ingram, T. Speciation along a depth gradient in a marine adaptive radiation. *Proc. Biol. Sci.* **278**, 613–618 (2011).
55. Malmstrøm, M., Matschiner, M., Tørresen, O. K., Jakobsen, K. S. & Jentoft, S. Whole genome sequencing data and *de novo* draft assemblies for 66 teleost species. *Sci. Data* **4**, 1–13 (2017).
56. Howe, K. *et al*. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
57. Miller, J. R. *et al*. Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* **24**, 2818–2824 (2008).
58. Parra, G., Bradnam, K., Ning, Z., Keane, T. & Korf, I. Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**, 289–297 (2009).

59. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
60. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
61. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882 (1997).
62. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870–1874 (2016).
63. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
64. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol* **10**, e1003537 (2014).
65. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
66. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013 (2014).
67. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351 (1997).
68. Labra, A., Pienaar, J. & Hansen, T. F. Evolution of Thermal Physiology in Liolaemus Lizards: Adaptation, Phylogenetic Inertia, and Niche Tracking. *The American Naturalist* **174**, 204–220 (2009).
69. Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S. & Hansen, T. F. A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology* **314**, 204–215 (2012).
70. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist* **164**, 683–695 (2004).
71. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**, S162–S173 (2009).

## Acknowledgements

## Author Contributions

S.J. and H.T.B. initially conceived and designed the study, with input from K.S.J. and W.S.; R.H. provided some of the tissue samples; Ma.M. conducted DNA extraction and library construction; O.K.T. set up and carried out the genome assemblies, with assistance from Ma.M. and H.T.B.; the phylogenetic species tree was created by Mi.M.; annotation of *Hb* genes, construction of gene-trees, selection analyses, ancestral reconstruction and *in silico* modeling of Hb was carried out by H.T.B. with assistance from M.H.S.; K.L.V. performed the comparative analyses with help from H.T.B.; H.T.B. wrote the manuscript and Supplementary Information together with S.J. and K.S.J. with input from all other authors.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-08286-2

**Competing Interests:** The authors declare that they have no competing interests.

**Accession codes**: All reads generated for this project have been deposited in the European Nucleotide Archive (ENA) under study accession PRJEB12469 (sample identifiers ERS1199874–ERS1199939). All new assemblies (unitigs and scaffolds) reported on here have been deposited in the Dryad repository under doi:10.5061/dryad.326r8.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.