PLOS GENETICS

RESEARCH ARTICLE

# Developmental constraint shaped genome evolution and erythrocyte loss in Antarctic fishes following paleoclimate change

**Jacob M. Daane**[1,2,3]*, **Juliette Auvinet**[1], **Alicia Stoebenau**[1], **Donald Yergeau**[4¤], **Matthew P. Harris**[2,3], **H. William Detrich, III**[1,4]*

**1** Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, Nahant, MA, United States of America, **2** Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA, United States of America, **3** Department of Genetics, Harvard Medical School, Boston, MA, United States of America, **4** Department of Biology, Northeastern University, Boston, MA, United States of America

¤ Current address: University at Buffalo Genomics and Bioinformatics Core, New York State Center of Excellence in Bioinformatics and Life Sciences, Buffalo, NY, United States of America
* j.daane@northeastern.edu (JMD); w.detrich@northeastern.edu (HWD)





**Citation:** Daane JM, Auvinet J, Stoebenau A, Yergeau D, Harris MP, Detrich HW, III (2020) Developmental constraint shaped genome evolution and erythrocyte loss in Antarctic fishes following paleoclimate change. PLoS Genet 16(10): e1009173. https://doi.org/10.1371/journal.pgen.1009173

**Editor:** Mary C. Mullins, University of Pennsylvania School of Medicine, UNITED STATES

**Received:** July 9, 2020

**Accepted:** October 6, 2020

**Published:** October 27, 2020



**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This work was supported by American Heart Association Postdoctoral Fellowship (No. 17POST33660801; www.heart.org) and by the Harvard Medical School Fund for Genetics of Climate Change (no URL) to J.M.D., by a John Simon Guggenheim Fellowship (www.gf.org), the US National Science Foundation (OPP- 2001584;

## Abstract

In the frigid, oxygen-rich Southern Ocean (SO), Antarctic icefishes (Channichthyidae; Notothenioidei) evolved the ability to survive without producing erythrocytes and hemoglobin, the oxygen-transport system of virtually all vertebrates. Here, we integrate paleoclimate records with an extensive phylogenomic dataset of notothenioid fishes to understand the evolution of trait loss associated with climate change. In contrast to buoyancy adaptations in this clade, we find relaxed selection on the genetic regions controlling erythropoiesis evolved only after sustained cooling in the SO. This pattern is seen not only within icefishes but also occurred independently in other high-latitude notothenioids. We show that one species of the red-blooded dragonfish clade evolved a spherocytic anemia that phenocopies human patients with this disease via orthologous mutations. The genomic imprint of SO climate change is biased toward erythrocyte-associated conserved noncoding elements (CNEs) rather than to coding regions, which are largely preserved through pleiotropy. The drift in CNEs is specifically enriched near genes that are preferentially expressed late in erythropoiesis. Furthermore, we find that the hematopoietic marrow of icefish species retained proerythroblasts, which indicates that early erythroid development remains intact. Our results provide a framework for understanding the interactions between development and the genome in shaping the response of species to climate change.

## Author summary

Our climate is rapidly changing. To better understand how species can adapt to major climate disturbance, we looked back into the past at a group of fishes that have encountered dramatic climate upheavals and thrived: Antarctic notothenioid fishes. In particular, we focus on the icefishes, which lost the ability to produce red blood cells in the frigid

www.nsf.gov) and by Milton Foundation funds (www.miltonfoundationforeducation.org) to M.P. H., and by US National Science Foundation grants (PLR-1444167 and OPP-1955368; www.nsf.gov) to H.W.D. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

environment of the Southern Ocean. By integrating past climate records with a large genetic dataset of Antarctic fishes, we show that the loss of red blood cells occurred only after sustained cooling of the Southern Ocean. As cooling continued into the modern era, we discover that even some of the "red-blooded" relatives of the icefishes show early genetic and morphological signs of erythrocyte loss. This cooling event left a non-random imprint on the genome of icefishes. With few exceptions, the genetic toolkit underlying red cell development has remained intact in icefishes because many "erythroid" genes perform important functions in other tissues. Rather, mutations have accumulated in gene regulatory regions near genes that control terminal erythroid maturation, such that icefishes continue to produce red cell progenitors but not mature erythrocytes. These results show that the genetic constraints regulating embryonic development shaped the evolutionary response of this fish group to climate change.

## Introduction

The cooling of the Southern Ocean (SO) beginning 35 million years ago (Ma) had a profound impact on the evolution of Antarctic fishes [1]. The stable, freezing temperatures, strong currents, and frequent storms created an environment in which dissolved oxygen was abundant and well mixed throughout the water column. In this unique environment, a single clade of Antarctic fishes, the icefishes (Notothenioidei: Cryonotothenioidea: Channichthyidae), lost the capacity to produce erythrocytes and the oxygen-transport protein hemoglobin (Hb)–and yet they thrive in the SO. The connection between paleoclimatic change in the SO and the origins of novel traits in notothenioid fishes provides a natural experiment for understanding the developmental and genetic mechanisms that shape phenotypic responses to environmental change.

Whereas the loss of erythrocytes among vertebrates is unique to icefishes, many closely related, but red-blooded, cryonotothenioid species cohabit the SO. Although having erythrocytes, these red-blooded species show a phylogenetic trend toward reduced hematocrit and/or mean corpuscular hemoglobin concentration, decreased hemoglobin multiplicity, and lowered hemoglobin affinity for $O_2$ as one proceeds from basal clades to the crown group Channichthyidae [1–4]. Intriguingly, several red-blooded Antarctic notothenioids survive experimentally induced anemia. Treatment of the bullhead notothen, *Notothenia coriiceps*, with the hemolytic agent phenylhydrazine reduces the percentage of erythrocytes in blood from 35% to 4% without lethality [5]. Similarly, the notothen, *Trematomus bernacchii*, survives conversion of its hemoglobin to the inactive carbonmonoxy state (95% CO-Hb)[6]; in contrast, CO-Hb exceeding 40% is lethal in humans [7]. Thus, erythrocytes and hemoglobin appear to be dispensable in red-blooded notothenioid lineages, which suggests an inherent resiliency within cryonotothenioids to accommodate extreme anemia.

Recent studies of the erythroid system in notothenioids have focused on the evolution of specific candidate genes, most notably the *alpha-* and *beta-globin* genes of the teleost *globin* clusters. These genes are almost completely deleted from the genomes of most icefishes [8–10], and globin regulatory elements are progressively deleted in the ancestral lineages leading to the icefishes [11]. Myoglobin expression is also absent from the hearts of 6 out of 16 icefish species, although mutated myoglobin genes remain in their genomes [12–15]. Furthermore, several genes encoding hemoglobin scavenging proteins, such as *haptoglobin*, have accumulated deleterious mutations and are expressed at reduced levels by icefishes [16]. Early work by Hureau *et al* [17] and by Barber *et al* [18] revealed that icefishes possess small numbers of

senescent, "erythrocyte-like" cells that are devoid of hemoglobin. These results suggest that despite the loss of hemoglobin, the block to erythropoiesis in icefishes might be constrained, or incomplete, leading to a minimally functional erythroid genetic program.

Together, these results provide insights into the evolution of the notothenioid hematological system, but a comprehensive assessment of changes to the erythroid developmental and genetic program is lacking. Part of this limitation has been the lack of the genome-wide data across cryonotothenioids necessary to establish a timeline of genomic changes supporting and potentially driving phenotypic adaptations. Recently, we published a dataset of ~250,000 loci representing protein-coding exons and conserved non-coding elements (CNEs) from 44 notothenioid species, including 10 icefishes and 6 dragonfishes (**S1 Fig**) [8]. The power of this phylogeny-wide genomic dataset lies in the ability to reconstruct the genetic steps that preceded, initiated, and follow trait evolution.

In this report, we systematically investigate the evolutionary genetic response of notothenioids to global paleoenvironmental change and explore the preconditions and consequences of erythrocyte loss on icefish genomes. Using these datasets, we discover pronounced shifts to the evolutionary rate of erythrocyte-associated CNEs following global cooling after the mid-Miocene climate transition, and we identify patterns of drift within these regions in extant high-latitude cryonotothenioids. We further demonstrate the retention of the majority of the erythroid protein coding toolkit and the presence of erythroid progenitors in icefish hematopoietic tissues, which together indicate that developmental mechanisms have been maintained through pleiotropy following trait loss.

## Results

Previous studies have focused on the end-point—the status of extant icefishes—but have lacked the genome-wide data from a sufficiently large sample of species to support high-resolution investigations into patterns of gene loss along ancestral branches and their associations with geological events.

### Drift in anemia-associated genetic regions followed the decline in global temperatures

An important assumption underlying our understanding of icefish evolution is the link between evolved character states and cooling of the SO. Our dataset permitted us to directly test these associations. To track patterns of genome evolution associated with environmental change, we integrated paleoclimate data with a time-calibrated phylogeny of notothenioids. We determined evolutionary dynamics across phyletic branches ancestral to icefishes to identify changes in selection across climatic and evolutionary history. Protein coding genes were grouped into clusters of similar function based on the Human Phenotype Ontology [19], and CNEs were assigned to adjacent genes using the 'GREAT' algorithm [20]. We detected a significant enrichment for accelerated evolutionary rates in anemia-associated genetic regions coincident with the loss of erythrocytes on the branch leading to the common ancestor of icefishes (**Fig 1A** and **1B**, **S1 Table**). Notably, this trend was found for CNEs but not for coding sequences (**S2 Fig**, **S2 Table**), revealing a bias toward drift in putative gene-regulatory regions. Relaxation of purifying selection in anemia-associated regions was not observed prior to erythrocyte loss in the phylogeny (**Fig 1B** and **1C**).

**Fig 1C** shows that relaxation of purifying selection on CNEs near anemia-associated genes in icefishes followed pronounced global cooling, decreases in sea level, and the formation of stable Antarctic ice sheets after the mid-Miocene climate transition (MMCT) 14 Ma [21–23]. Prior to and during the MMCT, regional sea surface temperature (SST) estimates and other

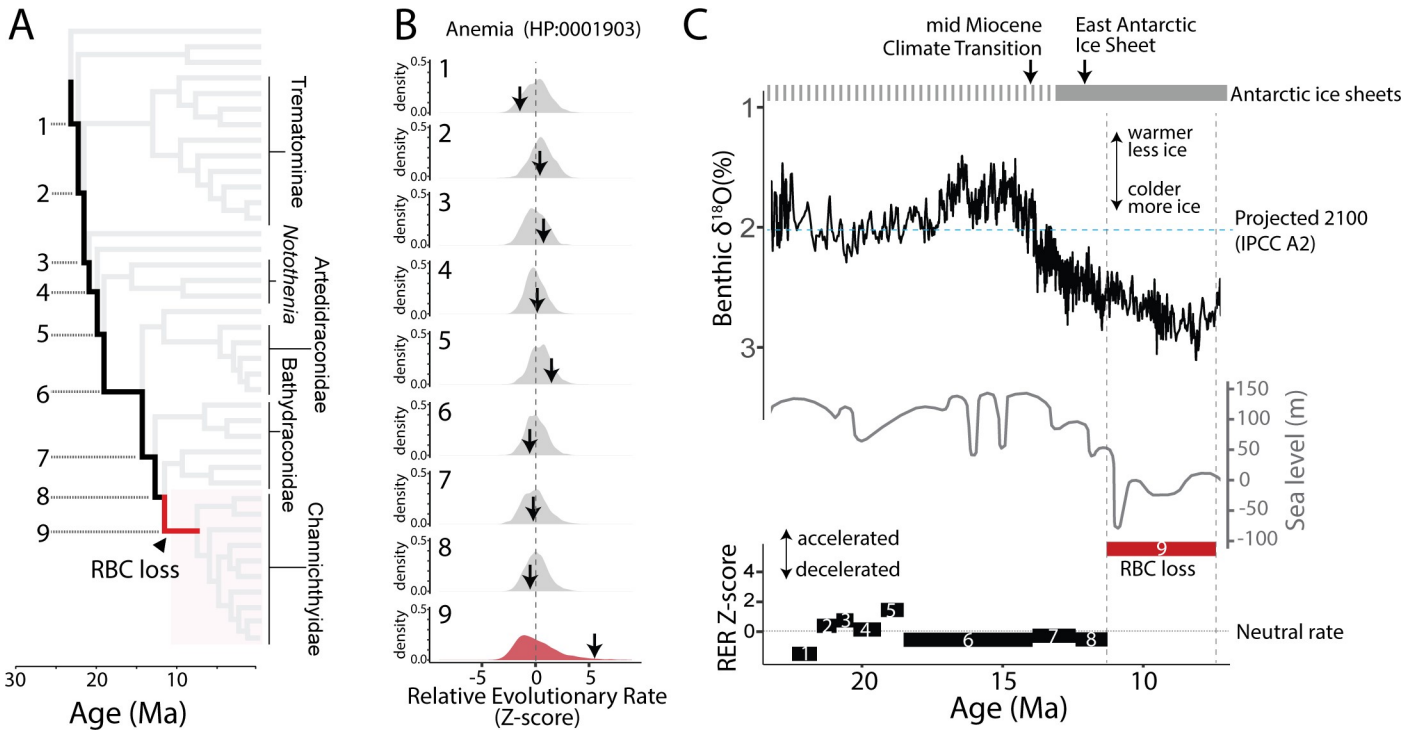



**Fig 1. Drift in anemia-associated CNEs followed erythrocyte loss and decline in global temperatures.** (**A**) Phylogeny of cryonotothenioids, highlighting the ancestral branches leading up to the loss of red blood cells (RBC) in icefishes (Channichthyidae). Numbers label branches in panel **B** and in the relative evolutionary rate (RER) plot of **C**. (**B**) Elevated RER of CNEs following loss of RBCs in icefishes. Distribution of Z-scores for average RER across groupings of conserved non-coding elements (CNEs). CNEs were linked to neighboring genes via the 'GREAT' algorithm [20] and then clustered based on the Human Phenotype Ontology (HPO) [19]. Z-scores > 0 are considered accelerated, while those < 0 have constrained evolution relative to the genome average. Arrows indicate positions in the histograms for the Anemia HPO term (HP:0001903). (**C**) RER increased in icefishes following loss of RBCs and the fall of global temperatures. The line numbers and lengths on the RER plot correspond to the branch labels and branch lengths on the time-calibrated phylogeny in **A**. The five-point moving average of global benthic $\delta^{18}O$ ratios is adapted from Zachos *et al.* [21] and sea-level estimations from Haq *et al.* [22].

https://doi.org/10.1371/journal.pgen.1009173.g001

oceanic temperature proxies [23–28] exceeded the critical thermal maxima ($CT_{max}$) of two extant icefish species [*Chaenocephalus aceratus* (13.9˚ ± 0.4˚C); *Chionodraco rastrospinosus* (13.3˚ ± 0.2˚C)] (**S3 Fig**), which are considered to be determined by the oxygen-carrying capacity of blood [29]. Therefore, evolution of the erythrocyte-null phenotype of Antarctic icefishes is tightly coupled, via physiology, to environmental cooling after the MMCT, in striking contrast to the increase in genetic diversity and positive selection for reduced skeletal density that evolved prior to the cryonotothenioid radiation [8].

## Correlation between the modern environment and relative evolutionary rate in notothenioids

Icefishes cohabit the frigid SO with several red-blooded notothenioid lineages. Although the cryonotothenioid radiation began ~22 Ma [30,31], the stably-cold temperatures and high oxygen concentrations of the SO necessary to facilitate viable reduction in hematocrit emerged well after the initial divergence of the group (**Fig 1C**, **S3 Fig**). Therefore, we propose that reduced hematocrits and tolerance of experimental anemia in several lineages of red-blooded notothenioids evolved independently of the icefish phenotypes [5,6].

To evaluate drift in anemia-associated genetic regions among extant notothenioids, we compared relative evolutionary rates both in high-latitude Antarctic (HA) and in sub-

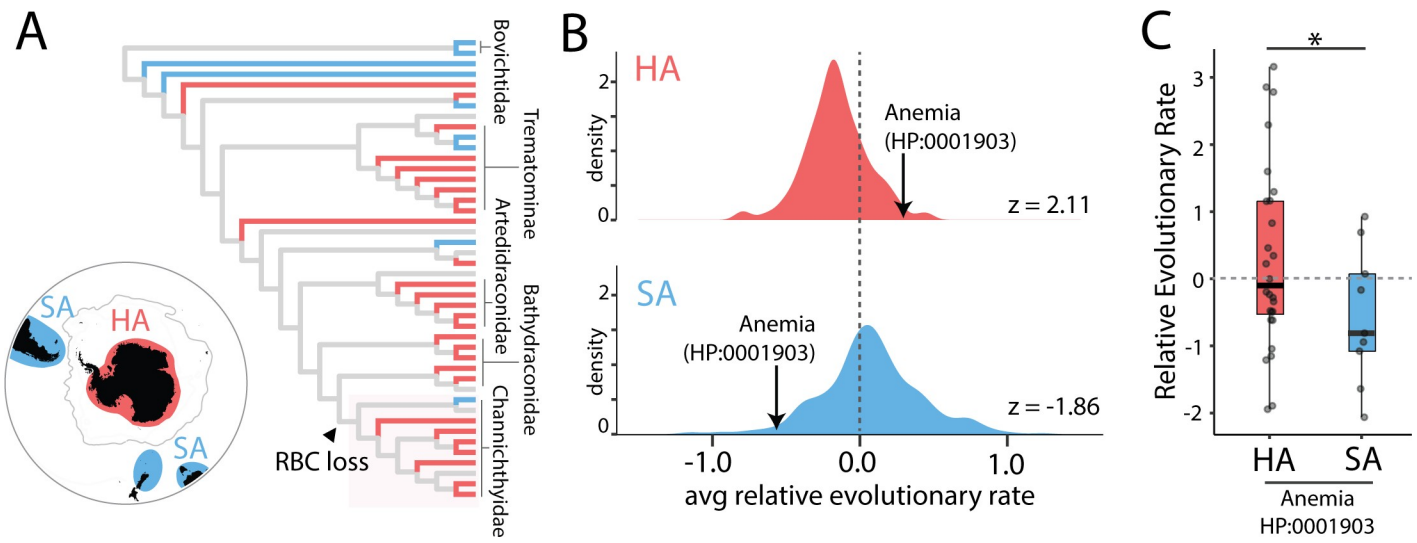



**Fig 2. Elevated evolutionary rates in anemia-associated CNEs in high latitude notothenioids.** (**A**) Notothenioid lineages designated as high-latitude Antarctic (HA) or sub-Antarctic (SA) as in Dornburg *et al*. 2017 [31]. (**B**) Distribution of average relative evolutionary rates across CNEs in all extant branches for each Human Phenotype Ontology term associated with at least 1000 CNEs. (**C**) Relative evolutionary rates of extant lineages. The asterisk indicates one-tailed t-test p-value < 0.05.

https://doi.org/10.1371/journal.pgen.1009173.g002

Antarctic (SA) notothenioids, as recently compiled by Dornburg *et al* [31] (**Fig 2A**). Consistent with the expectation that lower temperatures reduced selective pressure on erythrocyte-associated regions, we found that HA, but not SA, notothenioids showed a significant bias toward elevated evolutionary rate in CNEs (**Fig 2B and 2C**). This signal was largely driven by the icefishes, but also included several red-blooded notothenioid species (**S3 Table**). As a control, random selections of CNEs produced no deviation from neutral evolution when aggregated across these species' ensembles (**S4 Fig**). Thus, relative evolutionary rate of anemia-associated CNEs correlated with latitude in extant notothenioids, which suggests that independent weakening of purifying selection on the erythroid genetic program is ongoing.

## Independent deterioration of erythrocyte-associated genes and occurrence of spherocytic anemia in cryonotothenioids

Given the decreased hematocrits, reduced hemoglobin oxygen affinities, and apparent relaxation of purifying selection at anemia-associated genetic regions in Antarctic notothenioids, we investigated whether deleterious mutations had accumulated in erythroid genes across the notothenioid phylogeny. We searched for deleterious mutations within a set of candidate genes involved in many facets of red blood cell development and function, including genes encoding cytoskeletal proteins (7 genes; e.g., *sptb* [32,33]), membrane and solute transporters (10 genes; e.g., *slc4a1a* [34]), carbonic anhydrases (6 genes; e.g., *car2* [35]), heme and hemoglobin biosynthesis-associated proteins (16 genes; e.g., *alas2* [36]) and transcription factors that regulate erythropoiesis (32 genes; e.g., *gata1* [37,38]) (**S4 Table**). Due to ambiguity in assigning function to missense variants, we focused our analysis on truncating variants (frameshifts, premature termination codons and whole gene deletions) and on missense SNPs previously identified at orthologous sites of genetic variation in human patients.

Results show the truncating variants in icefish erythrocyte genes appear to have evolved independently on multiple occasions. Truncating mutations in our candidate gene set were confined/unique to the icefishes (**Fig 3A**, **S5–S7 Figs**) and absent in other notothenioid clades. Nonsense mutations or frameshifts in *alas2* (erythroid-specific isoform) that are predicted to

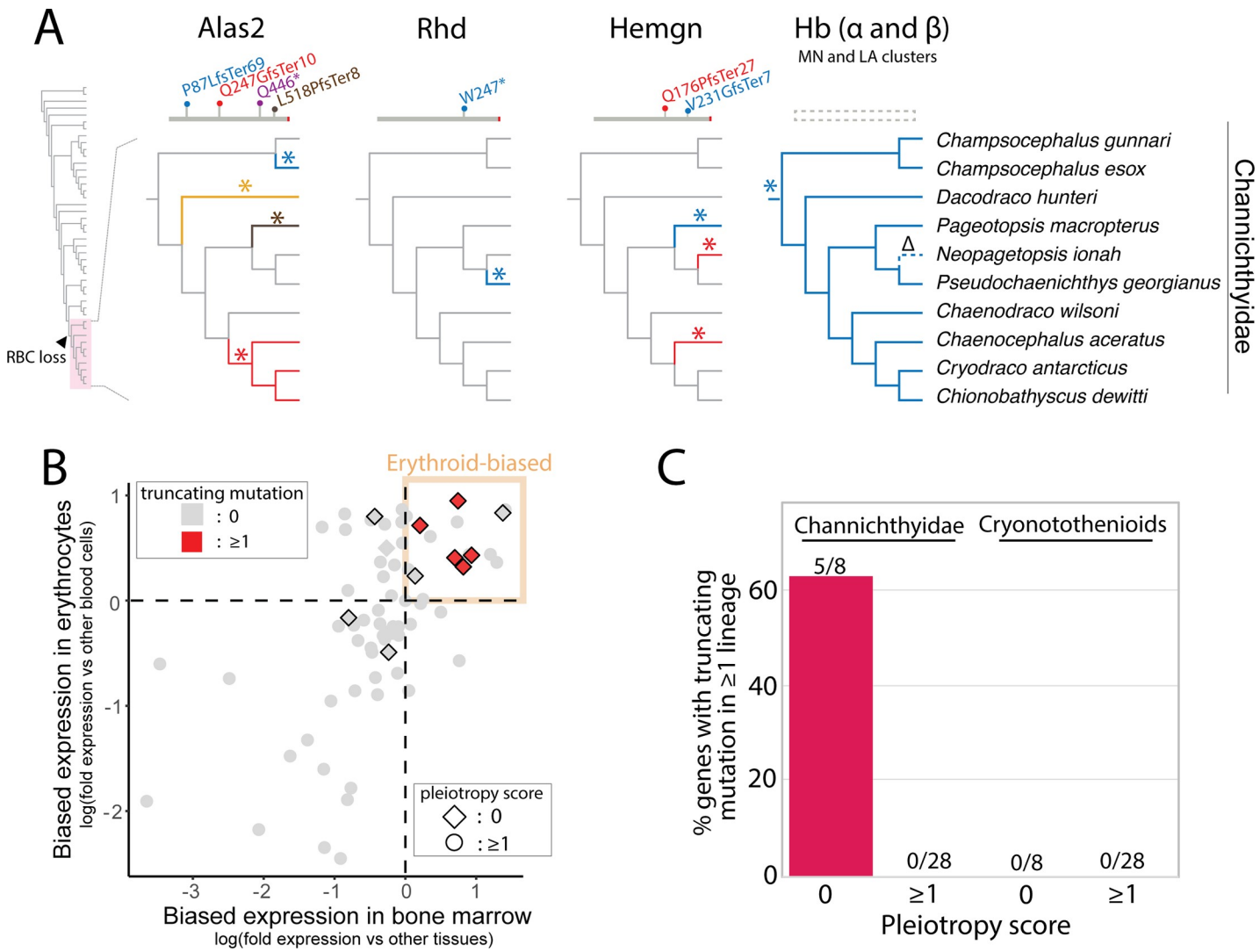



**Fig 3. Truncating mutations in erythrocyte-associated genes.** (**A**) Genes with truncating mutations or whole gene deletion in at least one icefish lineage. Asterisks indicate independent mutation events, with color corresponding to the allele above the cladogram. Δ indicates partial deletion of Hb in *N. ionah* with pseudogenization. See S5–S7 **Figs** for more detail on the Alas2, RhD and Hemgn mutations. See references 8–10 for analyses of the globin mutations. Arrow in **A** indicates position of red-blood cell (RBC) loss in icefishes (Channichthyidae) (**B**) Analysis of pleiotropy in erythroid genes (**S4 Table**). Genes were sorted by relative expression in mammalian erythrocytes vs other hematopoietic lineages and in mammalian bone marrow vs other organ systems. Genes were assigned a pleiotropy score ≥ 1 if mutations in these genes affect organ systems other than the hematopoietic system in the Mammalian Phenotype Ontology [74]. (**C**) Percentage of all erythroid-biased genes (**S5 Table**) with loss-of-function mutations in at least one lineage in icefishes (Channichthyidae) compared to other Antarctic notothenioid (Cryonotothenioids), showing enrichment for loss-of-function mutations in this gene set in species lacking red blood cells.

https://doi.org/10.1371/journal.pgen.1009173.g003

lead to premature termination were found in six of the 10 icefish species examined and arose independently in four lineages (**Fig 3A**, **S5 Fig**). Truncating mutations in *hemogen* (*hemgn*), which encodes an erythroid transcription factor [39,40], occurred independently in three icefish species, though many icefish species share a large deletion in this gene (**Fig 3A**, **S6 Fig**). Furthermore, *Rhd*, which encodes a blood group antigen, was truncated in *Pseudochaenichthys georgianus* (**Fig 3A**, **S7 Fig**). Consistent with our prior work, *globin* genes (*hba* and *hbb*) were also absent from most icefish species, with the exception of *Neopagetopsis ionah*, whose genome retained a pseudogenized version of the *globins* of the LA cluster (**Fig 3A**)[8,10].

The *sptb* (*beta-spectrin*) locus is particularly informative with respect to the timing of evolutionary decay of the erythroid program in cryonotothenioids. Sptb (erythrocytic Beta spectrin) is a cytoskeletal protein that interacts with ankyrin and other proteins to organize the erythrocyte membrane and maintain the oval shape of the red cell [32,41]. Multiple mutations in human *SPTB* disrupt the erythrocyte cytoskeleton and cause hereditary elliptocytosis or spherocytosis, which are characterized by elliptical and/or spherical erythrocytes [33]. Ten icefishes (of 10 examined) evolved variants at three highly conserved and clinically relevant amino acid positions (**Fig 4A**, **S8 Fig**). In contrast, two dragonfish species (of six examined), *Parachaenichthys charcoti* and *Gerlachea australis*, accumulated missense mutations in *sptb* at *different* sites (**Fig 4A**, **S8 Fig**) that also correspond to human *SPTB* mutations [42]. Note, because our dataset involved analysis of pools of individuals, these mutations are presumed to be fixed in the species. Given that the decrease in SO temperatures followed the divergence of the icefish and dragonfish clades (**Fig 1**, **S3 Fig**), their distinct *sptb* mutations must have arisen by independent decay (**Fig 4A**, **S8 Fig**). Nonetheless, dragonfishes are the sister taxon to the white-blooded icefishes, and the two clades may share physiological and genetic contexts that predispose the loss of red cell function and production.

To determine whether misshapen erythrocytes were present in dragonfishes with *sptb* mutations, we analyzed blood smears of *P. charcoti* with Wright/Giemsa stain. We did not examine icefish blood smears due to the absence of mature red blood cells in peripheral blood. Intriguingly, we found that *P. charcoti* has spherical erythrocytes (**Fig 4B and 4C**), the same pathology described in human patients with mutations at the same positions [42], whereas *N. coriiceps*, which does not share variation at human patient sites in *sptb* gene, possessed oval erythrocytes. Together, the genetic and morphological evidence is consistent with the independent decay of the erythroid developmental program in dragonfishes and icefishes.

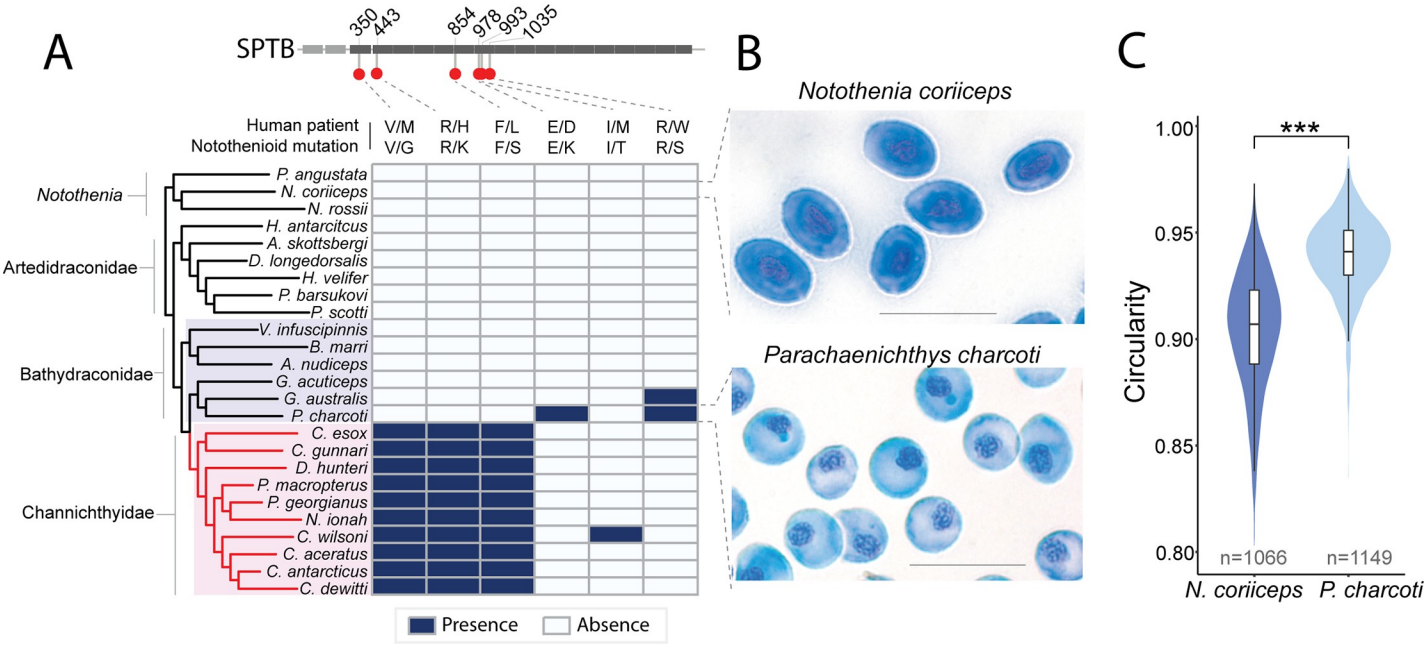



**Fig 4. Clinically relevant variation in the notothenioid *beta spectrin* gene predicts convergent phenotypes of anemia between humans and Antarctic fishes.** (**A**) Sites of mutation in human patients with hereditary spherocytosis/elliptocytosis that also have mutations in the notothenioid ortholog of *sptb*. Blue in heatmap indicates presence of the allele, white the absence. See **S8 Fig** for multiple sequence alignments. (**B**) Giemsa staining of peripheral blood revealed spherocytosis in erythrocytes of Charcot's dragonfish (*P. charcoti*) compared to the red-blooded bullhead notothen (*N. coriiceps*). Scale bar = 20 μm. C) Circularity of erythrocytes in *N. coriiceps* and *P. charcoti* (*** Wilcoxon signed-rank test p-value $< 2.2e^{-16}$). n indicates number of cells measured.

https://doi.org/10.1371/journal.pgen.1009173.g004

Outside of our candidate erythrocyte list, we found a few truncating variants in red-blooded notothenioids in non-erythrocyte oxygen-associated genes. This includes a truncating frameshift in the myoglobin of the barbled plunderfish *Artedidraco skottsbergi* and a truncation in the hemoglobin scavenging protein *haptoglobin* (*hp*) in the spiny plunderfish *Harpagifer antarcticus* (**S9 and S10 Figs**).

## Pleiotropy shaped patterns of gene evolution following loss of erythrocytes by icefishes

Patterns of drift in anemia-associated genetic regions in our erythroid dataset were found largely in CNEs rather than within coding sequences (**Figs 1** and **2**). We hypothesize that pleiotropy acts to maintain a core scaffold of erythroid genes, even in the erythrocyte-null icefishes. To test this hypothesis, we developed a pleiotropy score for cryonotothenioid genes that integrates non-hematopoietic phenotypes and gene expression patterns based on mammalian functional annotation databases (see Materials and Methods). Low pleiotropy scores correspond to genes with predominantly erythroid phenotypes and expression.

Results showed that icefish genes with truncating mutations had low pleiotropy scores and highly erythroid-biased expression (*alpha-* and *beta-globins*, *rhd*, *alas2*, *hemgn*; **Fig 3B**, **S4 Table**). Furthermore, there was a statistically significant negative association between pleiotropy and the presence of deleterious mutations among all erythroid-biased genes (**Fig 3C**, **S5 Table**; Fisher's exact test p-value = 0.0001). Thus, these classical "erythroid" genes are likely maintained in icefishes due to pleiotropic functions in other tissues.

## Patterns of mutation in CNEs predict the retention of erythroid progenitors in icefish blood

Erythrocytes develop from hematopoietic stem cells by specification of a myeloerythroid progenitor, commitment of the proerythroblast, and maturation through normoblast, reticulocyte, and terminal erythrocyte stages [43]. Given that icefishes apparently produce the full complement of myeloid and lymphoid lineages [17,18], two important questions emerge–what is (are) the stage(s) at which erythropoiesis fails, and what are the mechanism(s) underlying the failure? Because we detected specific signals of drift in icefish anemia-associated CNEs, we parsed patterns of drift in these regions across developmental specification of erythrocytes.

Using data from the murine ErythronDB to cluster genes by expression profile during erythrocyte maturation [44], we found that CNEs near genes that were highly expressed in reticulocytes had significantly elevated evolutionary rates on the ancestral branch leading to icefishes (**Fig 5A** and **5B**). This trend was observed across datasets for primitive, fetal definitive, and adult definitive erythropoiesis, and for the consensus gene set across all forms of erythropoiesis (**Fig 5A** and **5B**; **S11 Fig**). Given the conservation of CNEs near early-, but not late-, stage erythrocyte genes, we hypothesize that erythropoiesis in icefishes halts at the normoblast stages of late erythroid maturation, rather than at earlier stages, and that erythroid progenitors are present in icefish blood marrow.

To characterize hematopoietic lineages in notothenioid marrow, we examined head-kidney and spleen tissue prints (sites of leukopoiesis and erythropoiesis in fishes [43,45–47]) and peripheral blood smears from icefishes and red-blooded relatives (**Fig 5C** and **5D**; **S12 Fig**). **Fig 5C** and **5D** show that icefish marrow and peripheral blood possessed proerythroblasts and normoblasts but were devoid of reticulocytes or mature erythrocytes, whereas *N. coriiceps* marrow and blood contained the complete erythroid suite of cells. Furthermore, we also identified proerythroblasts and normoblasts in spleen prints from two other icefish species (*P. georgianus*, *C. rastrospinosus*; **S12 Fig**). In contrast, we found that the marrow and blood of icefishes

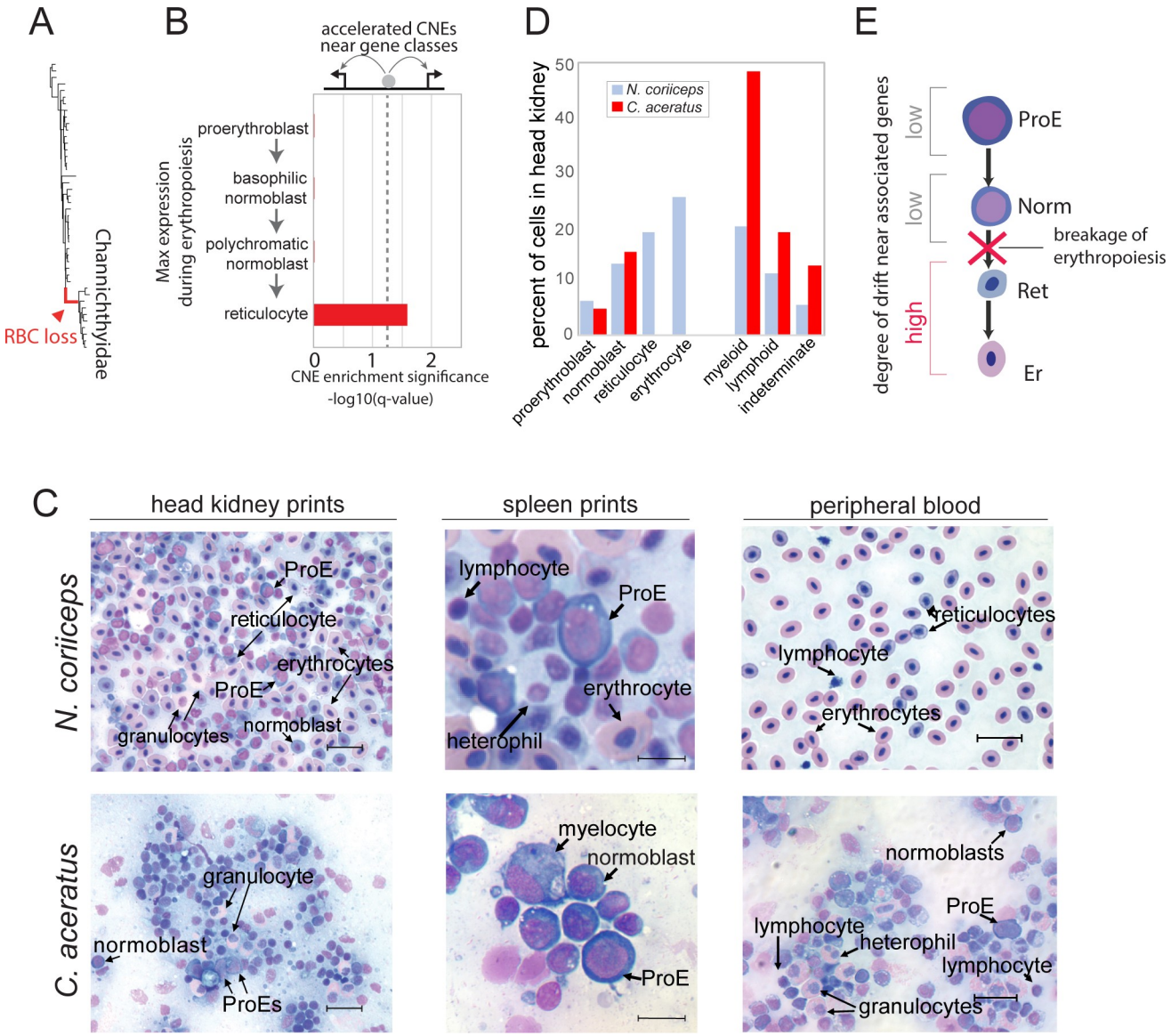



**Fig 5. Patterns of accelerated sequence evolution in CNEs predict the presence of erythroid progenitors in icefish marrow and blood.** (**A**) Branch leading to the common ancestor of icefishes from which the test for accelerated sequence evolution (phyloP) was run. (**B**) Enrichment for accelerated evolution of CNEs is biased toward genes that have maximum expression levels in reticulocytes. Data shown for the consensus gene list from primitive, fetal definitive, and adult definitive erythropoiesis in ErythronDB. See S11 Fig for individual breakdown of erythropoiesis types. (**C**) Prints of hematopoietic tissues (head kidney, spleen) and smears of peripheral blood from the "white-blooded" blackfin icefish *C. aceratus* and the red-blooded bullhead notothen *N. coriiceps* after staining with Wright/Giemsa. Both species possess erythroid progenitors (pro-erythroblasts (ProE)—circular cells with diameters ~15 μm, uncondensed nuclear chromatin, and densely blue-staining cytoplasm) and normoblasts (smaller proerythroblast derivatives with partial chromatin condensation and densely staining cytoplasm), but icefishes conspicuously lack reticulocytes (cells with an erythrocytic morphology but with a blue-staining cytoplasm due to high concentrations of globin and other mRNAs) and mature erythrocytes. Lymphoid and other myeloid lineages are present in both species. Scale bars: 25 μm in head kidney and peripheral blood, 10 μm in spleen. (**D**) Cell composition of head kidney prints from *N. coriiceps* (n = 1,420 cells) and the icefish *C. aceratus* (n = 825 cells). (**E**) Model of erythropoiesis in icefishes showing failure of differentiation/maturation (X) occurring at the normoblast (Norm) to reticulocyte (Ret) transition.

https://doi.org/10.1371/journal.pgen.1009173.g005

and red-blooded notothenioids contained apparently similar distributions of leukocytes: lymphocytes, myelocytes, and granulocytes (Fig 5C and 5D; S12 Fig). Taken together, these data indicate that erythropoiesis in icefishes fails during terminal maturation (Fig 5E).

## Discussion

Our results provide a comprehensive examination of the environmental, genetic, and developmental events involved in loss of erythrocytes in icefishes. By analyzing thousands of genes in a phylogenetic framework, we determined whether mutational events were shared among icefishes and other cryonotothenioids, and we used these patterns of mutation to predict natural phenotypes. Finally, we examined ancestral branches within the phylogeny to assess how changes to mutation rates aligned with paleoclimatic events.

As the global environment rapidly changes in response to anthropogenic impacts, understanding the mechanisms by which species have responded to past climate events is critically important. Here, we demonstrate an interaction between climate change, pleiotropy, and developmental constraint in shaping genome evolution. As many of the genes involved in erythropoiesis have pleiotropic roles in other tissues, the necessity of properly completing development appears to limit the allowable mutations that can occur in the genome–in this case most erythrocyte-associated genes remain largely intact. Thus, as species respond to climate change, this developmental constraint will play a role in shaping evolutionary trajectories.

### Erythrocyte loss: evolution in response to environmental change

Icefishes are the only vertebrate taxon whose species survive without red blood cells. How do species lose a cell type that is thought to be essential for viability? Through analysis of paleoclimate data and comparative genomics, we present a new perspective on the erythrocyte-null phenotype of icefishes. We find that loss of erythrocytes occurred following steep declines in global and local oceanic temperatures, which led to increased dissolved oxygen concentrations in the SO. Icefish genomes then evolved rapidly, with decay of erythrocyte-associated noncoding regions occurring only after the formation of stable Antarctic ice sheets. Thus, a proximal environmental trigger drove the dramatic changes in the erythroid genetic program of icefishes (and, independently, within the sister clade the dragonfishes). This is in striking contrast to the roles of standing genetic diversity and positive selection that underlie the reduced skeletal density that preceded the cryonotothenioid radiation [8].

### Erythrocyte loss: maladaptive or adaptive?

Loss of erythrocytes should present with strong negative selection against the resulting anemia. Montgomery and Clements [48] argue that the loss of red cells by icefishes represents a "disaptation"–"*an organismal character whose use to the organism is demonstrably inferior to that of a phylogenetically antecedent character*"–and that recovery via readaptation reduced the detrimental impact of ablation of the red cell. That the most recent common ancestor of the icefishes probably possessed the metabolic flexibility necessary to transition to an erythrocyte-null condition is demonstrated by the capacity of red-blooded notothenioids to survive poisoning of hemoglobin with CO [6] or severe anemia induced by phenylhydrazine [5]. Nevertheless, icefish blood has an oxygen carrying capacity $<10\%$ (by unit volume) that of red-blooded notothens [49]. Sidell and O'Brien [12] assert that numerous cardiovascular enhancements in icefishes, including increased vascular branching [12], enlargement of the heart [50,51], elevated mitochondrial densities in cells [52], and a four-fold increase in blood volume [53], were necessary to compensate for severe chronic anemia. As a result of these changes, an estimated 22% of resting metabolic rate in icefishes is devoted to cardiac function, compared to 0.5–5.0% in most temperate fishes [54,55]. Adding to the physiological calculus are mutations of key erythroid genes in the genomes of these fishes. We and others have identified loss-of-function alleles in the *alpha-* and *beta-globins* [8–10], *alas2*, *hemgn*, *haptoglobin* [16] and

*myoglobin* [12] that have risen to fixation in some, if not all, icefish species. These mutations are likely to constrain the adaptive landscape of icefishes and to render impossible the re-evolution of erythroid function as the SO warms. Thus, one may argue that the energetic savings achieved by abrogation of red cell production are likely to be negated by the costs of the physiological compensations to overcome anemia and the constrained ability of icefishes to adapt to environmental fluctuations.

## Underlying genetic 'scaffolding' of traits and prediction of natural phenotypes

The peripheral blood of icefishes, like red-blooded notothenioids and other fishes, contains leukocytes, lymphocytes, heterophils/granulocytes, myelocytes and thrombocytes. Although we did not detect reticulocytes or mature erythrocytes in icefish blood, we found small numbers of proerythroblasts and normoblasts. In striking contrast, we find that proerythroblasts, and to a lesser extent normoblasts, are abundant in icefish marrow prints from pronephric (head) kidney and spleen (**Fig 5**, **S12 Fig**). Progenitor accumulation might result from blockage of terminal erythroid differentiation/maturation or from a futile physiological response to hypoxia and anemia. Failure to mature beyond the normoblast stage implies that the erythroid genetic program is compromised at a step critical for terminal differentiation. The mutational event(s) causing this disruption is (are) not yet known.

We identified strong signals in CNEs that led to the prediction that erythroid progenitors do exist in icefish hematopoietic tissues. The same analyses performed on coding sequences failed to identify or predict this phenotype. Thus, pleiotropy likely constrains the types of mutations that can occur following trait loss. Although drift is a major factor shaping icefish genomes and developmental programs of erythropoiesis, pleiotropy appears to have left much of the erythroid genetic pathway intact. Dollo's Law argues that traits lost in a lineage do not re-evolve [56,57], and the nearly complete extinction of *globin* genes in icefishes makes the recovery of fully functional erythrocytes problematic. Nevertheless, the modest losses of key erythroid genes and the maintenance of erythroid progenitors might enable these species to 're-gain' some aspects of red cell function.

## Evolutionary mutant models of human disease

Traits that are adaptive to diverse organisms in various environmental contexts are often maladaptive (i.e., pathological) in humans. There has been much interest in using evolutionary diversity to gain insights into human disease, including how species evolve to compensate for the deleterious aspects of certain traits [58–60]. Comparative trait analysis can be used to identify novel disease genes, as shown by analysis of gene expression in icefishes [39,61], or as a means of filtering Genome-Wide Association Study (GWAS) hits, in which a high percentage of associated loci lack an obvious functional mechanism [62]. Cryonotothenioids have numerous traits that phenocopy human diseases, including aglomerular kidneys, lipid accumulation, low skeletal density, mitochondrial proliferation, heart enlargement with spongy myocardium, and others [1,12]. Furthermore, some cryonotothenioid traits, such as reduced skeletal density, show enrichment for selection in human-disease loci [8]. Thus, comparative genomic analyses within the cryonotothenioids have the potential to power our understanding of human diseases.

As an example, we demonstrate in this report convergence in anemic phenotypes based on mutations in the *beta-spectrin* gene of Antarctic icefishes, dragonfishes, and humans. Not only do dragonfishes and human patients share spherical erythrocytes as a result of *beta-spectrin* mutations, but the mutations introduce amino acid substitutions at the same highly conserved

positions. Furthermore, evolution of anemia appears to be ongoing in the sister taxa of the icefishes. We propose that icefishes and dragonfishes share genetic and physiological potentials to ameliorate the deleterious effects of anemia and that understanding this potential can be leveraged to treat the human disease.

## Materials and methods

### Ethics statement

The experimental use of notothenioid fishes was performed in accordance with protocol 18-0103R, which was approved by the Northeastern University Institutional Animal Care and Use Committee (IACUC).

### Notothenioid genomic datasets

We used our recently published dataset of a broad taxonomic sampling of 46 species of notothenioid fishes and close relatives, including *Percophis brasiliensis* as the sister taxon to notothenioids and *Percina caprodes* as an outgroup [8,63]. This dataset contains contigs constructed from cross-species targeted sequence enrichment for over 250,000 protein coding exons and conserved non-coding elements, with an average coverage of targeted regions >90% in all notothenioids (doi: 10.5281/zenodo.2628936).

### Multiple sequence alignment

Orthologous sequences within the dataset were mapped according to Daane *et al.* [8]. Non-coding sequences were aligned using Mafft v7.313 (parameters '*—maxiterate 1000—localpair —op 10—ep 10*')[64]. For coding sequences, the frameshift-aware program MACSE v2.03 was used (parameters '*-prog alignSequences -seq -seq_lr -fs_lr 10 -stop_lr 15*')[65]. The multiple sequence alignment was pruned using GUIDANCE v2.02 to mask residues with scores <0.6 (parameters '*—bootstrap 25—mafft—maxiterate 100,—localpair—op 10—ep 10*') [66].

### Reconstruction of gene sequences

As in Daane *et al.* [8], we reconstructed full gene sequences from the contigs that represented individual constituent coding exons. Orthologous exons were identified in the *Gasterosteus aculeatus* (three-spine stickleback) reference genome through reciprocal BLAST. We concatenated single-copy exons in the same order as they appear in the *Gasterosteus aculeatus* reference genome. Transcripts containing isoforms were merged into a non-redundant gene sequence containing all possible exons. A total of 18,600 gene sequences were reconstructed for each species.

### CNE association with genes

Because enhancers can regulate gene expression for genes many kilo- and mega-bases away from transcription start sites, prediction of regulatory targets is difficult *in silico*. To infer potential *cis*-regulatory targets of the CNEs, and thus link CNEs to putative biological function, we assigned CNEs to neighboring genes using the Genomic Regions Enrichment of Annotations Tool ('GREAT')[20]. This approach links CNEs to the transcription start site of the nearest neighboring genes within specified windows (minimum basal window is 5 kb upstream and 1 kb downstream of transcription start sites, extended up to 1 Mb or until overlap with the basal window from another gene) while allowing overlap such that multiple genes can be associated with the same CNE. This approach has much higher statistical power for detecting gene

ontology enrichment of CNEs when compared to simple distance-based approaches for associating CNEs to putative regulatory targets [20].

## Patterns of sequence evolution

Relative evolutionary rates were estimated using the program RERConverge (parameters '*transform = "sqrt"*, *weighted = T*, *scale = T*, *cutoff* = 0')[67]. As long branches exhibit higher degrees of variance compared to short branches, RERconverge includes a heteroskedasticity correction that increases comparative statistical power across the phylogeny [68].

We also assessed accelerated sequence evolution along pre-specified ancestral branches using the program phyloP, as implemented in PHAST v1.4 (parameters '*—method LRT—no-prune—features—mode ACC'*)[69,70]. The tree model for phyloP was derived using phyloFit and the species tree [8]. CNE tree models were based on 2,912 elements $\geq$ 250 bp with $\geq$ 85% coverage in all species.

## Gene cluster enrichment

We grouped notothenioid genes into specific clusters using several databases of mammalian orthologs. Since many of the evolved phenotypes in notothenioids are comparable to human pathologies, we utilized the Human Phenotype Ontology database (downloaded April 2018). We further used groupings of genes according to gene expression profiles during erythropoiesis in ErythronDB [44,71]. Gene identifiers for both databases were converted to Ensembl gene IDs followed by conversion to stickleback identifiers using Ensembl Biomart [72].

For analysis of relative evolutionary rate across a gene cluster, Z-scores were generated for each term by comparing the mean relative evolutionary rate from all genes within a gene cluster to a random distribution of 1,500 bootstrap resamples of equivalent bin sizes. Z-scores were calculated using SciPy (stats.scipy).

We also assessed patterns of cumulative polygenic enrichment within each gene cluster using the SUMSTAT approach [73]. For phyloP, we normalized the distribution of log-likelihood ratio test values (ΔlnL) by taking the fourth root (ΔlnL4). The ΔlnL4 score was then summed for all genes within an ontology and an enrichment p-value was estimated from the empirical sum(ΔlnL4) score through bootstrap resampling (1,500 replicates).

In all enrichment analyses, p-values were corrected using FDR (Python module statsmodels v0.6.1; fdrcorrection0).

## Analysis of notothenioid gene mutations in human orthologs

As our data was already converted to stickleback orthologs (see 'Reconstruction of gene sequences'), we used Ensembl Biomart to map orthologs between stickleback and human annotations [72]. To identify the site of orthologous human mutations, we performed multiple sequence alignments of each translated exon using Mafft v7.313 (parameters '*—maxiterate* 1000*—localpair—op* 10*—ep* 10*—addfragments* '). To avoid generating inferences based on non-homologous sites, we only considered amino acid positions where the ancestral notothenioid and human amino acids were identical. We then used the ClinVar database to check for variants in human patients at sites of notothenioid mutation [42].

## Assessment of pleiotropy in coding regions

We developed a pleiotropy score based on the number of recorded non-hematopoietic system phenotypes for each gene in the Mammalian Phenotype Ontology (downloaded March 2019), which is a record of phenotypes in mouse mutants organized by organ and tissue system [74].

To calculate pleiotropy scores in non-blood tissues, we removed descendent ontologies within the "Hematopoietic System Phenotype" from each gene. We excluded indirect phenotypes, such as pallor, abnormal iron or blood chemistry, body or organ size, and spleen abnormalities (**S6 Table**), because they are secondary to reduction in hematocrit. A score of 0 indicates absence of non-hematopoietic system phenotypes, whereas a score $\geq 1$ would indicate the presence of a phenotype outside of this system (e.g. craniofacial phenotype or muscle phenotype).

To complement the mouse phenotype data, we also included gene expression data to identify erythroid-biased genes. We used the Human Protein Atlas to distinguish between genes expressed throughout the body with those predominantly expressed in hematopoietic tissues (mammalian bone marrow)[75]. We further compared expression across multiple hematopoietic cell types to find erythroid-enriched genes (Array Express: E-MTAB-3079 on the Expression Atlas [76]).

Truncating mutations were indicated by the absence of read coverage across the gene and/or the presence of premature termination via frameshift or nonsense mutation. We required a minimum of three sequencing reads for any frameshift or nonsense mutation to be reported. Unless otherwise indicated, all frameshifts or nonsense mutations reported are fixed in our sequencing read data, which is pooled from populations of 5 or more individuals (see [8]).

## Tree calibration

We time-calibrated our species tree using TreePL (parameters '*smooth = 0.1*, *cv*, *randomcv*, *opt = 1 moredetail optad = 1*, *moredetailad*, *optcvad = 2*, *moredetailcvad*, *thorough*')[77]. We used date priors from two recent time-calibrated notothenioid phylogenies [30,31]. The minimum and maximum age estimate priors for the most recent common ancestor (MRCA) were: *Pseudaphritis* + Eleginopsioidea (62.5–87.1 Ma), *Harpagifer-Pogonophryne* (7.7–13.0 Ma), *Bathydraco-Chaenocephalus* (9.4–13.3 Ma), *Notothenia* (15.2–20.5 Ma), Cryonotothenioidea (18.6–23.9 Ma), Eleginopsiodea (37.2–53.2 Ma).

## Analysis of notothenioid blood

Three species of channichthyids (*Chaenocephalus aceratus*, *Pseudochaenichthys georgianus*, and *Chionodraco rastrospinosus*), two species of nototheniids (*Notothenia coriiceps* and *Gobionotothen gibberifrons*), and a single dragonfish species (*Parachaenichthys charcoti*) were collected by bottom trawling from the *R/V Polar Duke* or the *R/V Laurence M. Gould* near Low and Brabant Islands in the Palmer Archipelago. The fish were transported alive to Palmer Station, Antarctica, where they were maintained in seawater aquaria at -1.5˚C to +1.0˚C.

Whole blood (5–25 ml) was collected from live fishes via caudal venipuncture using heparinized syringes. Aliquots (~5–10 μl) from red-blooded species were directly smeared on glass microscope slides by standard techniques [78]. Because icefish blood contains ~4% cells by volume, cells were concentrated by low-speed centrifugation of 5 or 10 ml aliquots (clinical centrifuge, 1000 rpm, 5 min, room temperature), and pellets were resuspended in 0.5 ml Notothenioid Ringer's solution (260 mM NaCl, 5 mM KCl, 2.5 mM $MgCl_2$, 2.5 mM $CaCl_2$, 2 mM $NaHCO_3$, 2 mM $NaH_2PO_4$, 5 mM glucose) on ice before blood smears (~5–10 μl) were prepared. Head kidney and spleen tissues were dissected from euthanized fish, and prints were prepared by pressing each tissue gently onto microscope slides to deposit a monolayer of cells. Cells of smears and prints were then fixed in 100% methanol for 5 min.

Blood smears and tissue prints were stained with Wright's solution (0.1% w/v, pH = 6.8; Sigma-Aldrich) for 15 s, washed for 1 min in distilled water, and then stained with Giemsa solution (0.4% w/v, pH = 7.2, Sigma-Aldrich) for 1.5 min. Slides were then washed for 3 min

in distilled water and air-dried. Wright's stains the cytoplasm light blue, and Giemsa stains the nucleus a deeper blue/purple with collagen and other tissue elements staining pink to rose [79]. Micrographs were recorded using a Nikon E800 microscope equipped with differential interference contrast optics, a SPOT 7.2 Color Mosaic CCD camera (Diagnostic Instruments, Inc.), and SPOT 5.1 imaging software.

### Quantitation of erythrocyte morphology

Erythrocyte morphology was determined using Fiji [80]. Wright/Giemsa-stained peripheral blood smears from one individual of *N. coriiceps* and one of *P. charcoti* were quantified. To smooth edges and reduce background noise, a Gaussian blur (sigma = 1) and rolling ball background subtraction (rolling = 7) was applied to each image of a field of cells. Image contrast was enhanced ("saturated = 0.1 normalize"), and the image was converted to a binary through Auto-Thresholding ("method = Minimum"). Cells were further smoothed and gaps filled through opening and closing operations and the "fill holes" command. To ensure accurate measurements of cell shape, we ignored particles with unusual morphologies that may have been artifacts of automated thresholding. We also excluded cells that touched other cells or the edge of the frame by restricting particle size to an area of 70–150 $\mu m^2$ and by removing particles with circularity $< 0.80$ and solidity $< 0.93$. We analyzed 1,066 cells for *N. coriiceps* and 1,149 cells for *P. charcoti* for circularity ($C = 4\pi Area/Perimeter^2$).

## Supporting information

**S1 Fig. Phylogeny of notothenioid species included in this study.** Tree topology from Daane *et al.* [8]. Phylogenetic relationships inferred from ASTRAL using 11,627 gene trees. All nodes in the phylogeny are supported by 100% quadpartition posterior probability. Asterisk (*) indicates position of red blood cell loss in the icefishes (Channichthyidae).
(TIF)

**S2 Fig. Drift in coding sequences of anemia-associated genes did not follow erythrocyte loss or the decline in global temperatures.** (**A**) Phylogeny of cryonotothenioids, highlighting the ancestral branches leading up to the loss of red blood cells (RBC) in icefishes (Channichthyidae). Numbers label branches in panels **B** and **C**. (**B**) Elevated relative evolutionary rate (RER) following loss of RBCs in icefishes. Distribution of Z-scores for average RER across groupings of genes. These genes were then clustered based on the Human Phenotype Ontology (HPO) [19]. Arrow indicates position in histogram of the Anemia HPO term (HP:0001903). Z-scores $> 0$ are considered accelerated, while those $< 0$ have constrained evolution relative to the genome average. (**C**) Relative evolutionary rate across genes in icefishes following loss of RBCs and the fall of global temperatures remained steady. The five-point moving average of benthic δ18O ratios is adapted from Zachos et al. 2001 [21] and sea level estimations from Haq et al. 1987 [22].
(TIF)

**S3 Fig. Global and local paleo-temperature estimates and the loss of erythrocytes in icefishes.** Overlay of time-calibrated phylogeny of cryonotothenioids and paleoclimate estimates shows loss of red blood cells (*, red branch) following decreases in global and local temperatures. (**A**) Sea surface temperature (SST) reconstructions from multiple Southern Ocean drill sites. Site location, SST method and citation are indicated in the inset. Modern and paleo drill site locations adapted from Hartman *et al.*, 2018 [25], and mapped using the Ocean Drilling Stratigraphic Network Plate Tectonic Reconstruction Service (http://www.odsn.de/odsn/services/paleomap/paleomap.html). $CT_{max}$ for the blackfin icefish, *Chaenocephalus aceratus*, is

indicated by the dashed line. (**B**) The five-point moving average of global benthic $\delta^{18}O$ ratios is adapted from Zachos *et al.* 2001 [21]. Higher $\delta^{18}O$ ratios indicate colder temperatures and more ice.
(TIF)

**S4 Fig. Enrichment for elevated evolutionary rate in anemia-associated genetic regions compared to random gene sets.** Three random sets of genes equal to the number of genes in HP:0001903 (n = 360) were created and the relative evolutionary rate between species distributed in the high-Antarctic (HA) and sub-Antarctic (SA) were compared. * indicates one-tailed t-test p-value $< 0.05$; n.s. is not significant.
(TIF)

**S5 Fig. Truncating mutations identified in icefish erythroid-specific 5-aminolevulinate synthase (alas2) gene.** (**A**) Notothenioid phylogeny showing presence of truncating alleles (*) in four icefish species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D**-**G**) The icefishes show distinct frameshifts and truncations in Alas2 compared to the *N. coriiceps* reference sequence. Alignment start/stop coordinates in **D**-**G** are based on position in the *N. coriiceps* genome assembly (XP_010782407.1).
(TIF)

**S6 Fig. Truncating mutations identified in icefish hemogen gene.** (**A**) Phylogeny of the notothenioids showing the presence of truncating alleles (*) in three icefish species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D**) *Chaenocephalus aceratus* and *Neopagetopsis ionah* show identical frameshifts and truncations in Hemgn compared to the *N. coriiceps* reference. (**E**) *Pageotopsis macropterus* shows a different frameshift and truncation. Alignment start/stop coordinates in **D** and **E** are based on position in the *N. coriiceps* genome assembly (XP_010773828.1).
(TIF)

**S7 Fig. Truncating mutation identified in Pseudochaenichthys georgianus Rh blood group D antigen (rhd) gene.** (**A**) Notothenioid phylogeny showing presence of a truncating allele in *P. georgianus* (*). (**B**) The mutation encoded by the allele. (**C**) Sequencing read depth for *P. georgianus* as aligned to the *Notothenia coriiceps* reference genome. The gap in read depth corresponds to a deletion in each read relative to the reference genome. (**D**) *P. georgianus* shows a frameshift and truncation in Rhd compared to the *N. coriiceps* reference sequence. Alignment start/stop coordinates are based on position in the *N. coriiceps* genome assembly (XP_010782194.1).
(TIF)

**S8 Fig. Dragonfish and icefish mutations at highly-conserved and clinically-relevant sites in Beta-spectrin.** Variant amino acid substitutions in Beta-spectrin of the dragonfish *Parachaenichthys charcoti* and a representative icefish *Chaenodraco wilsoni* highlighted in red. Beta-spectrin sequences for three-spined stickleback (*Gasterosteus aculeatus*), spotted gar (*Lepisosteus oculatus*), elephant shark (*Gallorhinchus milii*) and human (*Homo sapiens*) are provided for comparison. The dbSNP identifier (ClinVar) for deleterious variants found in human patients with spherocytic anemia/elliptocytosis are shown above each alignment.
(TIF)

**S9 Fig. Truncating mutations identified in notothenioid myoglobin gene.** (**A**) Phylogeny of the notothenioids showing the presence of truncating alleles (*) in three species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D**) Red-blooded species *Artedidraco skottsbergi* Mb compared to the *N. coriiceps* reference. (**E**) *Champsocephalus gunnari* and *C. esox* shows identical frameshifts in Mb. Alignment start/stop coordinates in **D** and **E** are based on position in the *N. coriiceps* genome assembly (NP_001290223.1).
(TIF)

**S10 Fig. Truncating mutations identified in notothenioid haptoglobin gene.** (**A**) Phylogeny of the notothenioids showing the presence of truncating alleles (*) in three species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D**) Red-blooded species *Harpagifer antarcticus* Hp compared to the *N. coriiceps* reference. The icefish species (**E**) *Champsocephalus gunnari* and (**F**) *Dacodraco hunteri* have different frameshifts and truncations in Hp. Alignment start/stop coordinates in **D**-**F** are based on position in the *N. coriiceps* genome assembly (XP_010770321.1).
(TIF)

**S11 Fig. Enrichment for accelerated sequence evolution in conserved non-coding elements (CNEs) near genes that are maximally expressed at distinct stages of erythropoiesis.** Three waves of mammalian erythropoiesis are defined by distinct patterns of gene expression and (locations): primitive (yolk sac blood island), fetal definitive (liver) and adult definitive (bone marrow). For each erythropoietic wave, accelerated evolution of CNEs near maximally expressed genes is shown for four cellular stages of erythroid differentiation/maturation: proerythroblast, basophilic erythroblast/normoblast, polychromatic erythroblast/normoblast, reticulocyte. The Consensus is the intersection of maximally expressed genes across each the three erythropoietic waves. Dashed line corresponds to q-value of 0.05. Gene expression data from ErythronDB [44].
(TIF)

**S12 Fig. Spleen prints from three notothenioid species: Wright/Giemsa-stained.** Two "white-blooded" icefishes, *Pseudochaenichthys georgianus* and *Chionodraco rastrospinosus*, show the presence of erythroid progenitors [proerythroblasts (ProEs) and normoblasts] but lack later stages of maturation (e.g., reticulocytes, erythrocytes). By contrast, the red-blooded notothen, *Gobionotothen gibberifrons*, displays the complete erythropoietic progression: ProE → normoblast → reticulocyte → erythrocyte. Scale bar = 10 μm.
(TIF)

**S1 Table. Human phenotype ontology (HPO) enrichment of CNEs under accelerated sequence evolution.**
(XLSX)

**S2 Table. Human phenotype ontology (HPO) enrichment of coding sequences under accelerated sequence evolution.**
(XLSX)

**S3 Table. Relative evolutionary rate and notothenioid biogeography.**
(PDF)

**S4 Table. Coverage and mutations in candidate erythrocyte genes.**
(PDF)

**S5 Table. Coverage and mutations in erythroid-biased genes.**
(PDF)

**S6 Table. Excluded terms from Mammalian Phenotype Ontology (MP) in pleiotropy analysis.**
(PDF)

## Acknowledgments

We thank Dr. Thomas Desvignes (University of Oregon) for preparing the blood smears used in Fig 4C. Spherocytosis of the erythrocytes of *P. charcoti* was first described by Dr. Michael J. Peters using blood smears prepared in 1996 by H.W.D. [81]. This is contribution No. 411 from the Marine Science Center at Northeastern University.

## Author Contributions

**Conceptualization:** Jacob M. Daane, Matthew P. Harris, H. William Detrich, III.

**Data curation:** Jacob M. Daane.

**Formal analysis:** Jacob M. Daane.

**Funding acquisition:** Jacob M. Daane, Matthew P. Harris, H. William Detrich, III.

**Investigation:** Jacob M. Daane, Juliette Auvinet, Alicia Stoebenau, Donald Yergeau.

**Methodology:** Jacob M. Daane.

**Project administration:** H. William Detrich, III.

**Resources:** Matthew P. Harris, H. William Detrich, III.

**Software:** Jacob M. Daane.

**Supervision:** Matthew P. Harris, H. William Detrich, III.

**Validation:** Jacob M. Daane.

**Visualization:** Jacob M. Daane.

**Writing – original draft:** Jacob M. Daane.

**Writing – review & editing:** Jacob M. Daane, Juliette Auvinet, Donald Yergeau, Matthew P. Harris, H. William Detrich, III.

## References

1. Eastman JT. Antarctic Fish Biology: Evolution in a Unique Environment. San Diego, CA: Academic Press, Inc; 1993.
2. Egginton S. Blood rheology of Antarctic fishes: viscosity adaptations at very low temperatures. J Fish Biol. 1996; 48: 513–521. https://doi.org/10.1006/jfbi.1996.0049
3. di Prisco G, Giardina B, Avino RD, Condo SG, Bellellit A, Brunori M. Antarctic fish hemoglobin: an outline of the molecular structure and oyxgen binding properties—II. oxygen binding properties. Comp Biochem Physiol. 1988; 90: 585–591.
4. Daane JM, Giordano D, Coppola D, di Prisco G, Detrich HW, Verde C. Adaptations to environmental change: Globin superfamily evolution in Antarctic fishes. Mar Genomics. 2020; 49: 100724. https://doi.org/10.1016/j.margen.2019.100724 PMID: 31735579
5. Borley KA, Beers JM, Sidell BD. Phenylhydrazine-induced anemia causes nitric-oxide-mediated upregulation of the angiogenic pathway in Notothenia coriiceps. J Exp Biol. 2010; 213: 2865–2872. https://doi.org/10.1242/jeb.043281 PMID: 20675556
6. di Prisco G, Macdonald, J A, Brunori M. Antarctic fishes survive exposure to carbon monoxide. Experientia. 1992; 48: 473–475. https://doi.org/10.1007/BF01928166 PMID: 1601112

7. Nelson G. Effects of carbon monoxide in man: Exposure fatality studies. In: Hinschler MM, editor. Carbon Monoxide and Human Lethality: Fire and Non-fire Studies. New York: Taylor and Francis; 2006. pp. 3–62.
8. Daane JM, Dornburg A, Smits P, MacGuigan DJ, Brent Hawkins M, Near TJ, et al. Historical contingency shapes adaptive radiation in Antarctic fishes. Nat Ecol Evol. 2019; 3: 1102–1109. https://doi.org/10.1038/s41559-019-0914-2 PMID: 31182814
9. Kim B-M, Amores A, Kang S, Ahn D, Kim J-H, Kim I-C, et al. Antarctic blackfin icefish genome reveals adaptations to extreme environments. Nat Ecol Evol. 2019; 3: 469–478. https://doi.org/10.1038/s41559-019-0812-7 PMID: 30804520
10. Near TJ, Parker SK, Detrich HW III. A genomic fossil reveals key steps in hemoglobin loss by the Antarctic icefishes. Mol Biol Evol. 2006; 23: 2008–2016. https://doi.org/10.1093/molbev/msl071 PMID: 16870682
11. Lau Y, Parker SK, Near TJ, III HWD. Evolution and Function of the Globin Intergenic Regulatory Regions of the Antarctic Dragonfishes (Notothenioidei: Bathydraconidae). Mol Biol Evol. 2012; 29: 1071–1080. https://doi.org/10.1093/molbev/msr278 PMID: 22075115
12. Sidell BD, O'Brien KM. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes. J Exp Biol. 2006; 209: 1791–802. https://doi.org/10.1242/jeb.02091 PMID: 16651546
13. Grove T, Hendrickson J, Sidell B. Two species of antarctic icefishes (genus Champsocephalus) share a common genetic lesion leading to the loss of myoglobin expression. Polar Biol. 2004; 27: 579–585. https://doi.org/10.1007/s00300-004-0634-0
14. Moylan TJ, Sidell BD. Concentrations of myoglobin and myoglobin mRNA in heart ventricles from Antarctic fishes. J Exp Biol. 2000; 203: 1277–86. Available: http://www.ncbi.nlm.nih.gov/pubmed/10729277 PMID: 10729277
15. Sidell BD, Vayda ME, Small DJ, Moylan TJ, Londraville RL, Yuan ML, et al. Variable expression of myoglobin among the hemoglobinless Antarctic icefishes. Proc Natl Acad Sci U S A. 1997; 94: 3420–3424. https://doi.org/10.1073/pnas.94.7.3420 PMID: 9096409
16. Bilyk KT, Zhuang X, Murphy KR, Cheng C-HC. A tale of two genes: divergent evolutionary fate of haptoglobin and hemopexin in hemoglobinless Antarctic icefishes. J Exp Biol. 2019;222: jeb188573. https://doi.org/10.1242/jeb.188573 PMID: 30765469
17. Hureau C, Petit D, Fine J, Marneuc M. New cytological, biochemical and physiological data on the colorless blood of the Chaenichthyidae (Pisces,Teleosteans, Perciformes). In: Llano G, editor. Adaptations Within Antarctic Ecosystem—Proceedings of the 3rd SCAR Symposium on Antarctic Biology. Houston, TX: Gulf Publishing Co; 1977. pp. 459–477.
18. Barber DL, Mills Westermann JE, White MG. The blood cells of the Antarctic icefish Chaenocephalus aceratus Lönnberg: light and electron microscopic observations. J Fish Biol. 1981; 19: 11–28. https://doi.org/10.1111/j.1095-8649.1981.tb05807.x
19. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The human phenotype ontology in 2017. Nucleic Acids Res. 2017; 45: D865–D876. https://doi.org/10.1093/nar/gkw1039 PMID: 27899602
20. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28: 495–501. https://doi.org/10.1038/nbt.1630 PMID: 20436461
21. Zachos J, Pagani M, Sloan L, Thomas E, Billups K. Trends, Global Rhythms, Aberrations in Global Climate 65Ma to Present. Science. 2001; 292: 686–693. https://doi.org/10.1126/science.1059412 PMID: 11326091
22. Haq BU, HardenbolL J, Vail PR. Chronology of Fluctuating Sea Levels Since the Triassic. Science. 1987; 235: 1156–1167. https://doi.org/10.1126/science.235.4793.1156 PMID: 17818978
23. Shevenell AE, Kennett JP, Lea DW. Middle Miocene Southern Ocean cooling and antarctic cryosphere expansion. Science. 2004; 305: 1766–1770. https://doi.org/10.1126/science.1100061 PMID: 15375266
24. Shevenell AE, Ingalls AE, Domack EW, Kelly C. Holocene Southern Ocean surface temperature variability west of the Antarctic Peninsula. Nature. 2011; 470: 250–254. https://doi.org/10.1038/nature09751 PMID: 21307939
25. Hartman JD, Sangiorgi F, Salabarnada A, Peterse F, Houben AJP, Schouten S, et al. Paleoceanography and ice sheet variability offshore Wilkes Land, Antarctica–Part 3: Insights from Oligocene–Miocene TEX 86 -based sea surface temperature reconstructions. Clim Past. 2018; 14: 1275–1297. https://doi.org/10.5194/cp-14-1275-2018

26. Billups K, Schrag DP. Paleotemperatures and ice volume of the past 27 Myr revisited with paired Mg/Ca and 18 O/ 16 O measurements on benthic foraminifera. Paleoceanography. 2002; 17: 3-1-3–11. https://doi.org/10.1029/2000PA000567

27. Plancq J, Mattioli E, Pittet B, Simon L, Grossi V. Productivity and sea-surface temperature changes recorded during the late Eocene–early Oligocene at DSDP Site 511 (South Atlantic). Palaeogeogr Palaeoclimatol Palaeoecol. 2014; 407: 34–44. https://doi.org/10.1016/j.palaeo.2014.04.016

28. Petersen S V., Schrag DP. Antarctic ice growth before and after the Eocene-Oligocene transition: New estimates from clumped isotope paleothermometry. Paleoceanography. 2015; 30: 1305–1317. https://doi.org/10.1002/2014PA002769

29. Beers JM, Sidell BD. Thermal tolerance of Antarctic notothenioid fishes correlates with level of circulating hemoglobin. Physiol Biochem Zool. 2011; 84: 353–62. https://doi.org/10.1086/660191 PMID: 21743249

30. Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, et al. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc Natl Acad Sci U S A. 2012; 109: 3434–9. https://doi.org/10.1073/pnas.1115169109 PMID: 22331888

31. Dornburg A, Federman S, Lamb AD, Jones CD, Near TJ. Cradles and museums of Antarctic teleost biodiversity. Nat Ecol Evol. 2017; 1: 1379–1384. https://doi.org/10.1038/s41559-017-0239-y PMID: 29046532

32. Lux SE. Anatomy of the red cell membrane skeleton: Unanswered questions. Blood. 2016; 127: 187–199. https://doi.org/10.1182/blood-2014-12-512772 PMID: 26537302

33. Tse W. Red blood cell membrane disorders. Br J Haematol. 2000; 2–13.

34. Cabantchik ZI. Erythrocyte membrane transport. Novartis Found Symp. 1999; 226: 6–19. https://doi.org/10.1002/9780470515730.ch2 PMID: 10645535

35. Boron WF. Evaluating the role of carbonic anhydrases in the transport of HCO3−-related species. Biochim Biophys Acta—Proteins Proteomics. 2010; 1804: 410–421. https://doi.org/10.1016/j.bbapap.2009.10.021 PMID: 19879980

36. Chung J, Chen C, Paw BH. Heme metabolism and erythropoiesis. Curr Opin Hematol. 2012; 19: 156–162. https://doi.org/10.1097/MOH.0b013e328351c48b PMID: 22406824

37. Dore LC, Crispino JD. Transcription factor networks in erythroid cell and megakaryocyte development. Blood. 2011; 118: 231–239. https://doi.org/10.1182/blood-2011-04-285981 PMID: 21622645

38. Bresnick EH, Hewitt KJ, Mehta C, Keles S, Paulson RF, Johnson KD. Mechanisms of erythrocyte development and regeneration: implications for regenerative medicine and beyond. Development. 2018;145: dev151423. https://doi.org/10.1242/dev.151423 PMID: 29321181

39. Peters MJ, Parker SK, Grim J, Allard CAH, Levin J, Detrich HW III. Divergent Hemogen genes of teleosts and mammals share conserved roles in erythropoiesis: analysis using transgenic and mutant zebrafish. Biol Open. 2018; 7: bio035576. https://doi.org/10.1242/bio.035576 PMID: 30097520

40. Yang L V., Nicholson RH, Kaplan J, Galy A, Li L. Hemogen is a novel nuclear factor specifically expressed in mouse hematopoietic development and its human homologue EDAG maps to chromosome 9q22, a region containing breakpoints of hematological neoplasms. Mech Dev. 2001; 104: 105–111. https://doi.org/10.1016/s0925-4773(01)00376-8 PMID: 11404085

41. Marchesi VT, Steers E. Selective Solubilization of a Protein Component of the Red Cell Membrane. Science. 1968; 159: 203–204. https://doi.org/10.1126/science.159.3811.203 PMID: 5634911

42. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42: D980–D985. https://doi.org/10.1093/nar/gkt1113 PMID: 24234437

43. Witeska M. Erythrocytes in teleost fishes: A review. Zool Ecol. 2013; 23: 275–281. https://doi.org/10.1080/21658005.2013.846963

44. Kingsley PD, Greenfest-Allen E, Frame JM, Bushnell TP, Malik J, Mcgrath KE, et al. Ontogeny of erythroid gene expression. Blood. 2013; 121: e5–e13. https://doi.org/10.1182/blood-2012-04-422394 PMID: 23243273

45. Catton WT. Blood Cell Formation in Certain Teleost Fishes. Blood. 1951; 6: 39–60. https://doi.org/10.1182/blood.V6.1.39.39 PMID: 14791599

46. Kondera E. Haematopoiesis and haematopoietic organs in fish. Rocz Nauk Pol Tow Zootech. 2019; 15: 9–16. https://doi.org/10.5604/01.3001.0013.4535

47. Fänge R. Blood cells, haemopoiesis and lymphomyeloid tissues in fish. Fish Shellfish Immunol. 1994; 4: 405–411. https://doi.org/10.1006/fsim.1994.1036

48. Montgomery J, Clements K. Disaptation and recovery in the evolution of Antarctic fishes. Trends Ecol Evol. 2000; 15: 267–271. https://doi.org/10.1016/s0169-5347(00)01896-6 PMID: 10856946

49. Holeton GF. Oxygen uptake and circulation by a hemoglobinless antarctic fish (Chaenocephalus aceratus Lonnberg) compared with three red-blooded antartic fish. Comp Biochem Physiol. 1970; 34: 457–471. https://doi.org/10.1016/0010-406x(70)90185-4 PMID: 5426570

50. Johnston IA, Fitch N, Zummo G, Wood RE, Harrison P, Tota B. Morphometric and ultrastructural features of the ventricular myocardium of the haemoglobin-less icefish Chaenocephalus aceratus. Comp Biochem Physiol Part A Physiol. 1983; 76: 475–480. https://doi.org/10.1016/0300-9629(83)90449-8

51. Feller G, Goessens G, Gerday C, Bassleer R. Heart structure and ventricular ultrastructure of hemoglobin- and myoglobin-free icefish Channichthys rhinoceratus. Cell Tissue Res. 1985; 242: 669–676. https://doi.org/10.1007/BF00225436 PMID: 4075384

52. O'Brien KM, Sidell BD. The interplay among cardiac ultrastructure, metabolism and the expression of oxygen-binding proteins in Antarctic fishes. J Exp Biol. 2000; 203: 1287–97. Available: http://www.ncbi.nlm.nih.gov/pubmed/10729278 PMID: 10729278

53. Fitch NA, Johnson IA, Wood RE. Skeletal muscle capillary supply in a fish that lacks respiratory pigments. Respir Physiol. 1984; 57: 201–211. https://doi.org/10.1016/0034-5687(84)90093-8 PMID: 6494646

54. Farrell AP, Jones DR. The heart. In: Hoar WS, Randall DJ, Farrell AP, editors. Fish Physiology, Vol XIIA. San Diego: Academic Press; 1992. pp. 1–88.

55. Hemmingsen EA. Respiratory and Cardiovascular Adaptations in Hemoglobin-Free Fish: Resolved and Unresolved Problems. In: Prisco G di, Maresca B, Tota B, editors. Biology of Antarctic Fish. New York: Springer Verlag; 1991. pp. 191–203.

56. Dollo L. Les Lois de l'evolution. Bull La Société Belge Géologie. 1893; 7: 164–6. Available: https://paleoglot.org/files/Dollo_93.pdf

57. Gould SJ. Dollo on Dollo's law: Irreversibility and the status of evolutionary laws. J Hist Biol. 1970; 3: 189–212. https://doi.org/10.1007/BF00137351 PMID: 11609651

58. Schartl M. Beyond the zebrafish: diverse fish species for modeling human disease. Dis Model Mech. 2014; 7: 181–92. https://doi.org/10.1242/dmm.012245 PMID: 24271780

59. Albertson RC, Cresko W, Detrich HW, Postlethwait JH. Evolutionary mutant models for human disease. Trends Genet. 2009; 25: 74–81. https://doi.org/10.1016/j.tig.2008.11.006 PMID: 19108930

60. Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G. A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell Rep. 2012; 2: 817–23. https://doi.org/10.1016/j.celrep.2012.08.032 PMID: 23022484

61. Yergeau DA, Cornell CN, Parker SK, Zhou Y, Detrich HW III. bloodthirsty, an RBCC/TRIM gene required for erythropoiesis in zebrafish. Dev Biol. 2005; 283: 97–112. https://doi.org/10.1016/j.ydbio.2005.04.006 PMID: 15890331

62. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. Am J Hum Genet. 2018; 102: 717–730. https://doi.org/10.1016/j.ajhg.2018.04.002 PMID: 29727686

63. Near TJ, Dornburg A, Harrington RC, Oliveira C, Pietsch TW, Thacker CE, et al. Identification of the notothenioid sister lineage illuminates the biogeographic history of an Antarctic adaptive radiation. BMC Evol Biol. 2015; 15: 109. https://doi.org/10.1186/s12862-015-0362-9 PMID: 26062690

64. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772–80. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

65. Ranwez V, Harispe S, Delsuc F, Douzery EJP. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. Murphy WJ, editor. PLoS One. 2011; 6: e22594. https://doi.org/10.1371/journal.pone.0022594 PMID: 21949676

66. Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 2015; 43: W7–W14. https://doi.org/10.1093/nar/gkv318 PMID: 25883146

67. Kowalczyk A, Meyer WK, Partha R, Mao W, Clark NL, Chikina M. RERconverge: an R package for associating evolutionary rates with convergent traits. Valencia A, editor. Bioinformatics. 2019; 35: 4815–4817. https://doi.org/10.1093/bioinformatics/btz468 PMID: 31192356

68. Partha R, Kowalczyk A, Clark NL, Chikina M. Robust Method for Detecting Convergent Shifts in Evolutionary Rates. Mol Biol Evol. 2019; 36: 1817–1830. https://doi.org/10.1093/molbev/msz107 PMID: 31077321

69. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010; 20: 110–21. https://doi.org/10.1101/gr.097857.109 PMID: 19858363

70. Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with space/time models. Brief Bioinform. 2011; 12: 41–51. https://doi.org/10.1093/bib/bbq072 PMID: 21278375

71. Higgs DR, McGowan SJ, Soneji S, Merryweather-Clarke AT, Roberts DJ, Buckle VJ, et al. Global gene expression analysis of human erythroid progenitors. Blood. 2011; 117: e96–e108. https://doi.org/10.1182/blood-2010-07-290825 PMID: 21270440

72. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database. 2011; 2011: bar030–bar030. https://doi.org/10.1093/database/bar030 PMID: 21785142

73. Daub JT, Moretti S, Davydov II, Excoffier L. Detection of Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. Mol Biol Evol. 2017; 34: 1391–1402. https://doi.org/10.1093/molbev/msx083 PMID: 28333345

74. Smith CL, Eppig JT. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. Wiley Interdiscip Rev Syst Biol Med. 2009; 1: 390–399. https://doi.org/10.1002/wsbm.44 PMID: 20052305

75. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010; 28: 1248–1250. https://doi.org/10.1038/nbt1210-1248 PMID: 21139605

76. Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, et al. Expression Atlas update—An integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res. 2016; 44: D746–D752. https://doi.org/10.1093/nar/gkv1045 PMID: 26481351

77. Smith SA, O'Meara BC. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. 2012; 28: 2689–2690. https://doi.org/10.1093/bioinformatics/bts492 PMID: 22908216

78. Detrich HW, Yergeau DA. Comparative genomics in erythropoietic gene discovery: Synergisms between the antarctic icefishes and the zebrafish. Methods Cell Biol. 2004; 2004: 475–503. https://doi.org/10.1016/s0091-679x(04)77026-0 PMID: 15602928

79. Carson F, Matthews J, Pickett J. Preparation of tissues for laboratory examination. In: Race J, editor. Laboratory Medicine. New York: Harper and Rowe; 1980. pp. 1–57.

80. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012; 9: 676–682. https://doi.org/10.1038/nmeth.2019 PMID: 22743772

81. Peters Michael J. Modeling the evolutionary loss of erythroid genes by Antarctic icefishes: analysis of the hemogen gene using transgenic and mutant zebrafish. Northeastern University. 2018. https://doi.org/10.17760/D20293303

# Supplementary Materials for

## Developmental constraint shaped genome evolution and erythrocyte loss in Antarctic fishes following paleoclimate change

Jacob M. Daane, Juliette Auvinet, Alicia Stoebenau, Donald Yergeau, Matthew P. Harris, H. William Detrich III

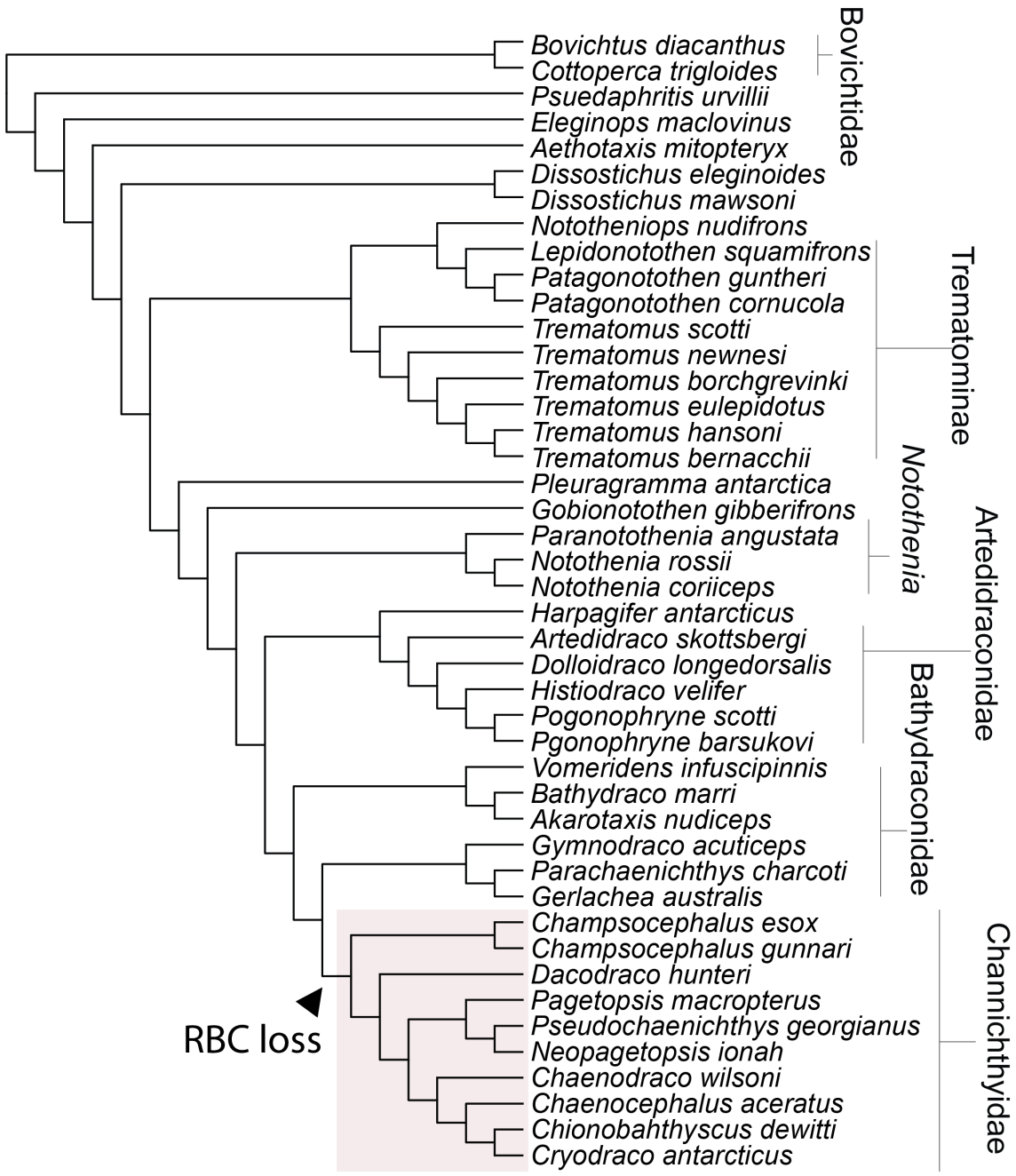
**Correspondence to:** j.daane@northeastern.edu, w.detrich@northeastern.edu

**This PDF file includes:**

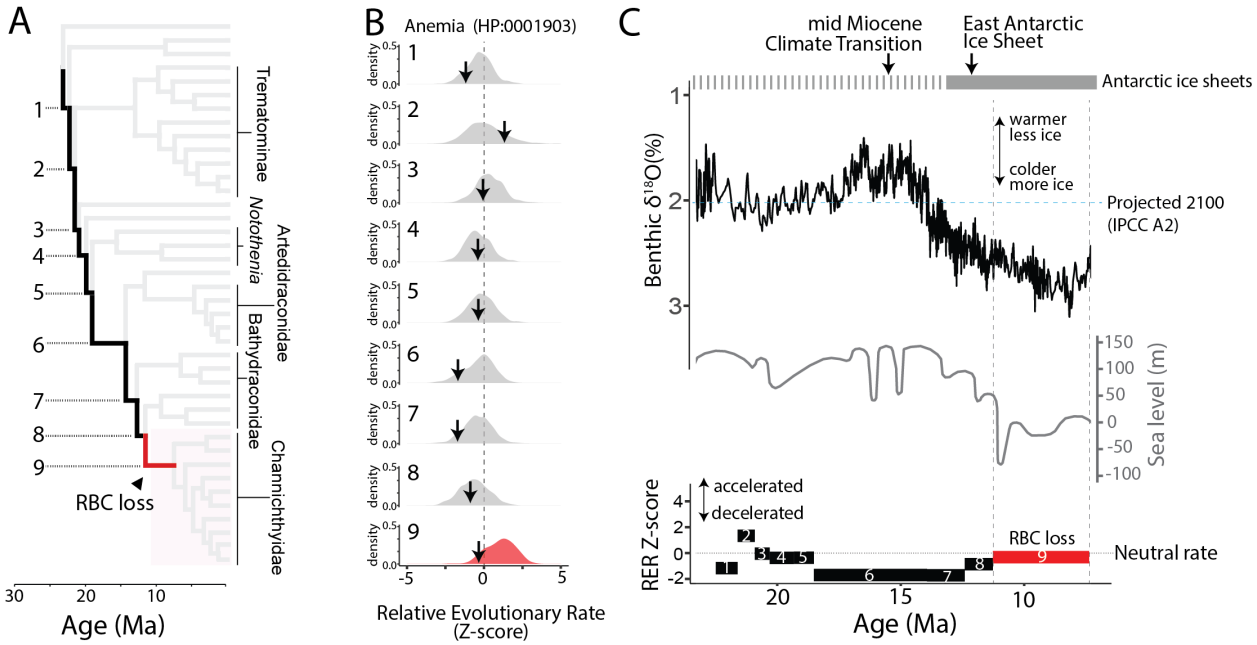
Supplementary Figures 1 to 12

Supplementary Tables 3 to 6

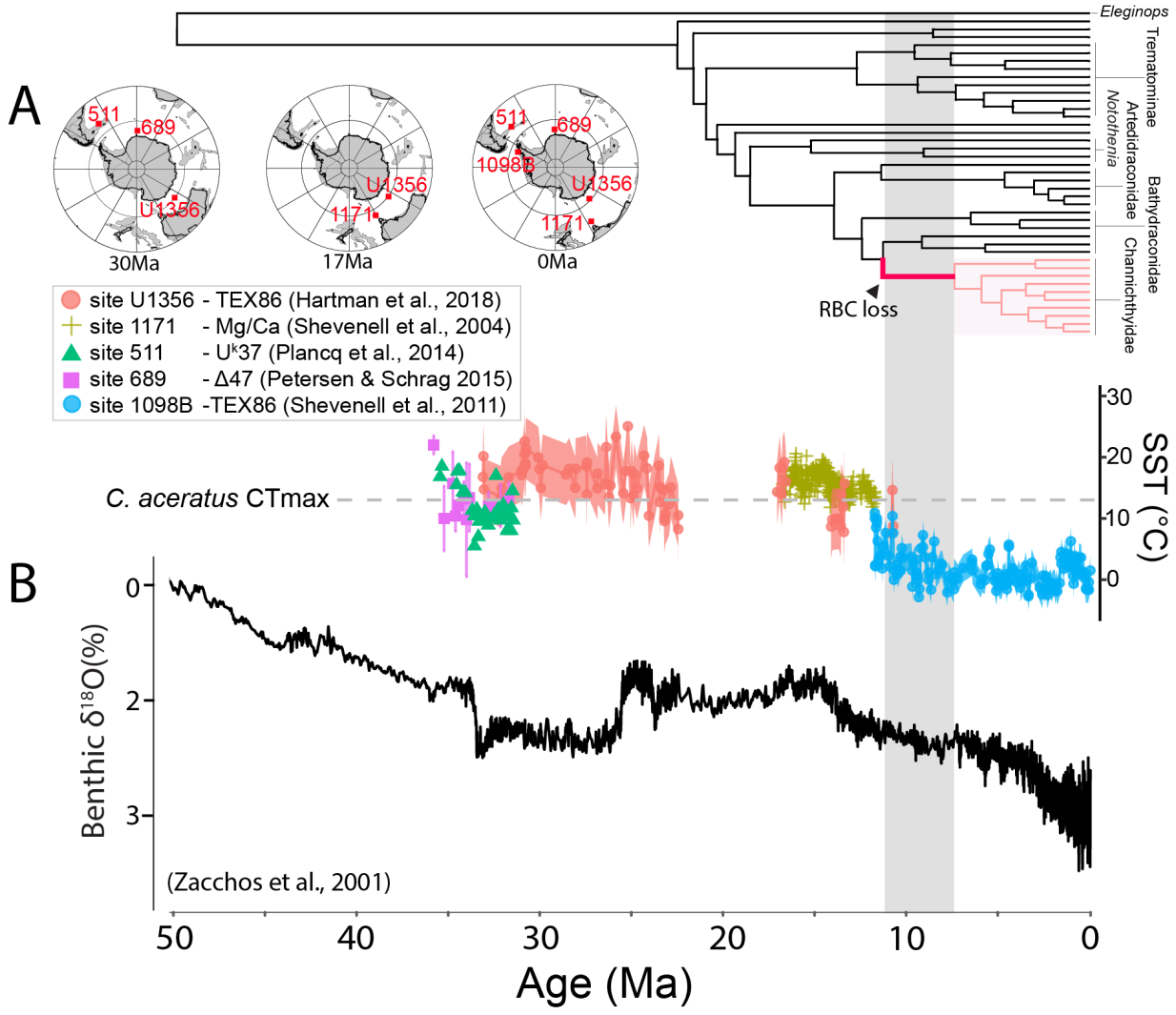
# Table of Contents



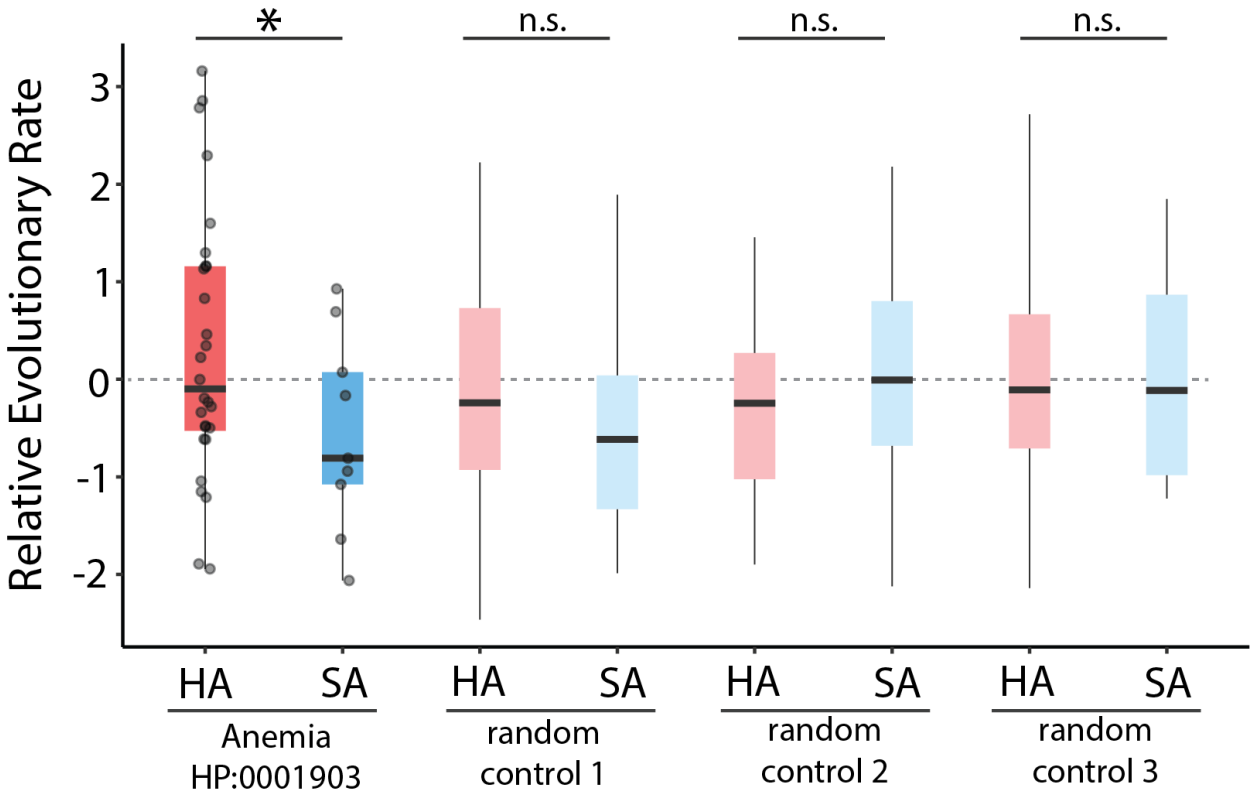

**S1 Fig. Phylogeny of notothenioid species included in this study.** Tree topology from Daane *et al*. (8). Phylogenetic relationships inferred from ASTRAL using 11,627 gene trees. All nodes in the phylogeny are supported by 100% quadpartition posterior probability. Asterisk (*) indicates position of red blood cell loss in the icefishes (Channichthyidae).



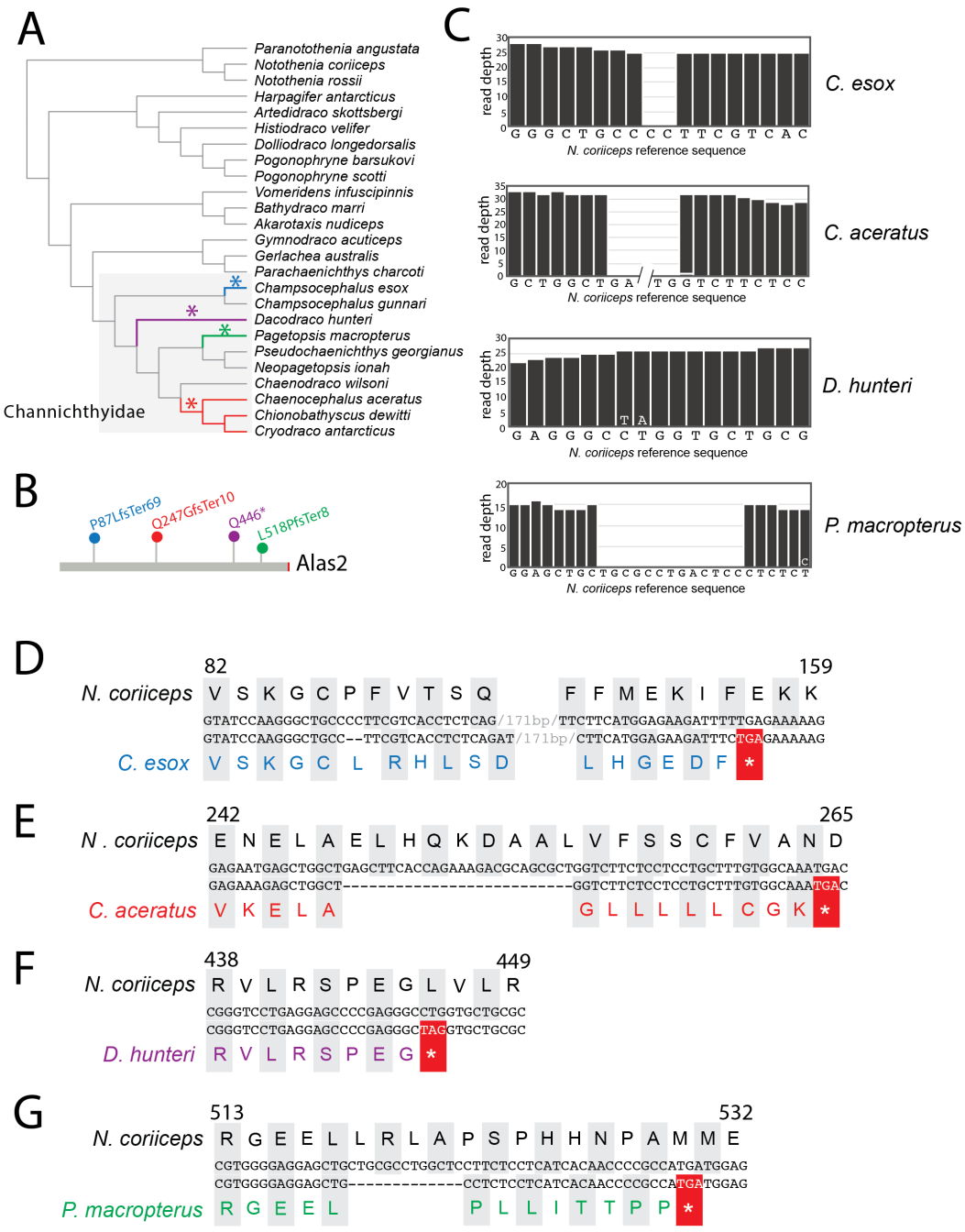

**S2 Fig. Drift in coding sequences of anemia-associated genes did not follow erythrocyte loss or the decline in global temperatures**. (**A**) Phylogeny of cryonotothenioids, highlighting the ancestral branches leading up to the loss of red blood cells (RBC) in icefishes (Channichthyidae). Numbers label branches in panels **B** and **C**. (**B**) Elevated relative evolutionary rate (RER) following loss of RBCs in icefishes. Distribution of Z-scores for average RER across groupings of genes. These genes were then clustered based on the Human Phenotype Ontology (HPO) (15). Arrow indicates position in histogram of the Anemia HPO term (HP:0001903). Z-scores > 0 are considered accelerated, while those < 0 have constrained evolution relative to the genome average. (**C**) Relative evolutionary rate across genes in icefishes following loss of RBCs and the fall of global temperatures remained steady. The five-point moving average of benthic δ18O ratios is adapted from Zachos et al. 2001 (17) and sea level estimations from Haq et al. 1987 (18).



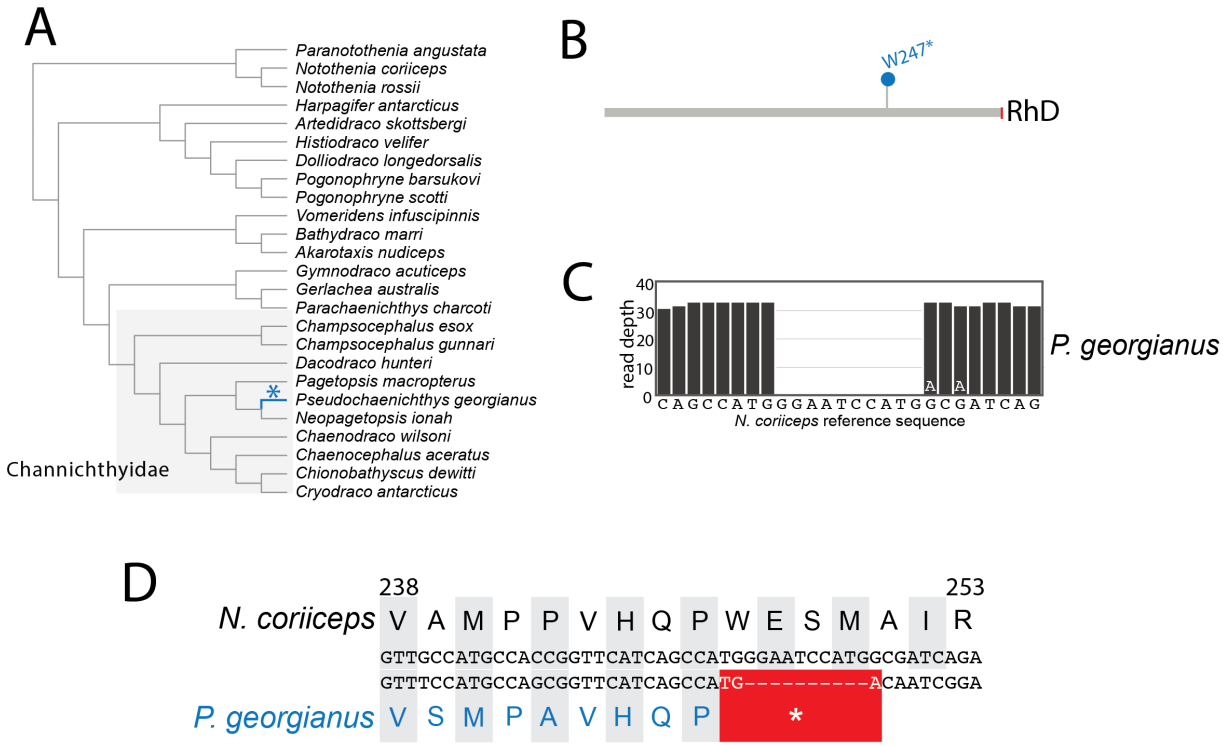

**S3 Fig. Global and local paleo-temperature estimates and the loss of erythrocytes in icefishes**. Overlay of time-calibrated phylogeny of cryonotothenioids and paleoclimate estimates shows loss of red blood cells (*, red branch) following decreases in global and local temperatures. (**A**) Sea surface temperature (SST) reconstructions from multiple Southern Ocean drill sites. Site location, SST method and citation are indicated in the inset. Modern and paleo drill site locations adapted from Hartman *et al.*, 2018 (ref), and mapped using the Ocean Drilling Stratigraphic Network Plate Tectonic Reconstruction Service (http://www.odsn.de/odsn/services/paleomap/paleomap.html). $CT_{max}$ for the blackfin icefish, *Chaenocephalus aceratus*, is indicated by the dashed line. (**B**) The five-point moving average of global benthic $\delta^{18}O$ ratios is adapted from Zachos *et al.* 2001 (17). Higher $\delta^{18}O$ ratios indicate colder temperatures and more ice.



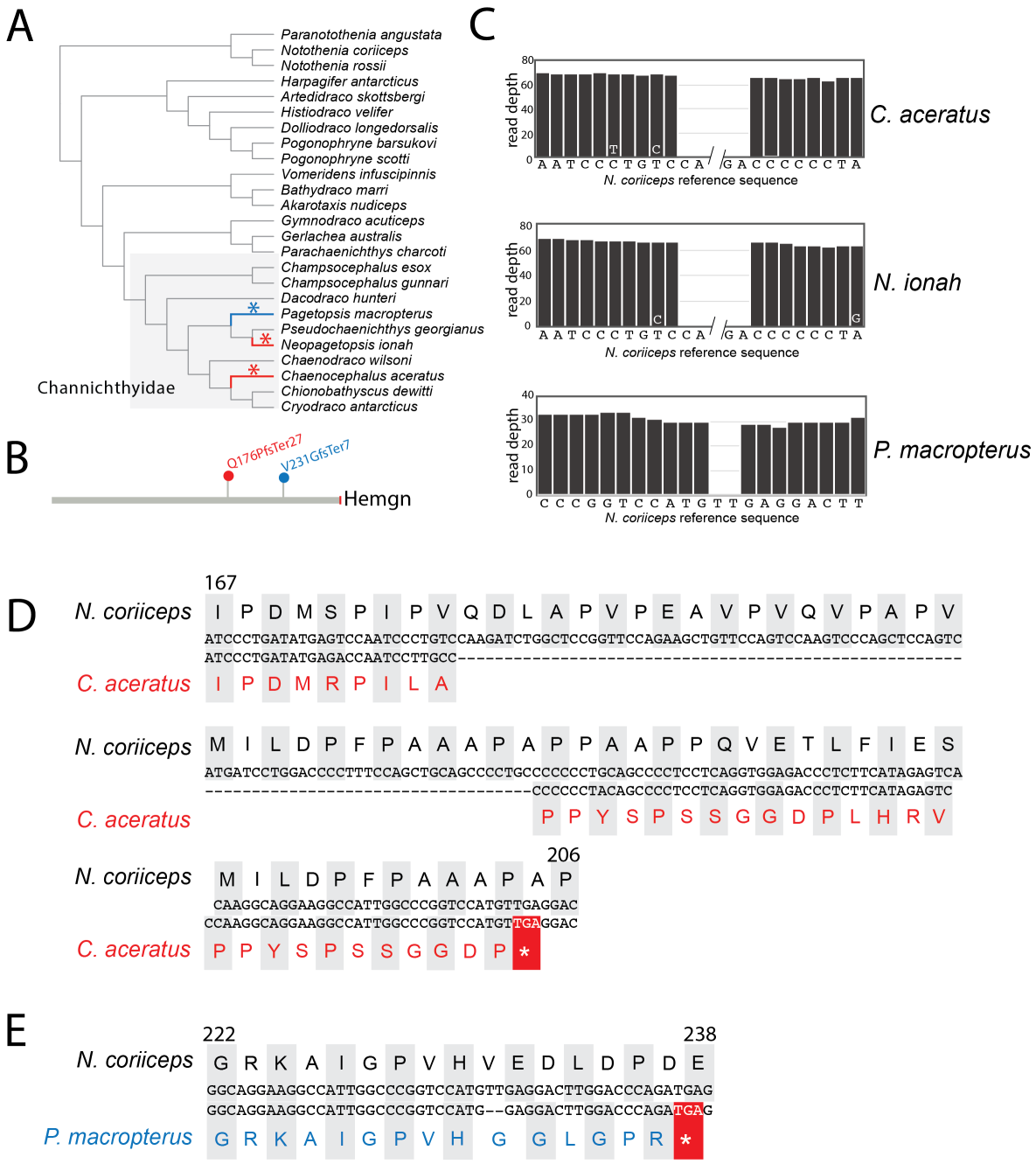

**S4 Fig. Enrichment for elevated evolutionary rate in anemia-associated genetic regions compared to random gene sets**. Three random sets of genes equal to the number of genes in HP:0001903 (n = 360) were created and the relative evolutionary rate between species distributed in the high-Antarctic (HA) and sub-Antarctic (SA) were compared. * indicates one-tailed t-test p-value < 0.05; n.s. is not significant.



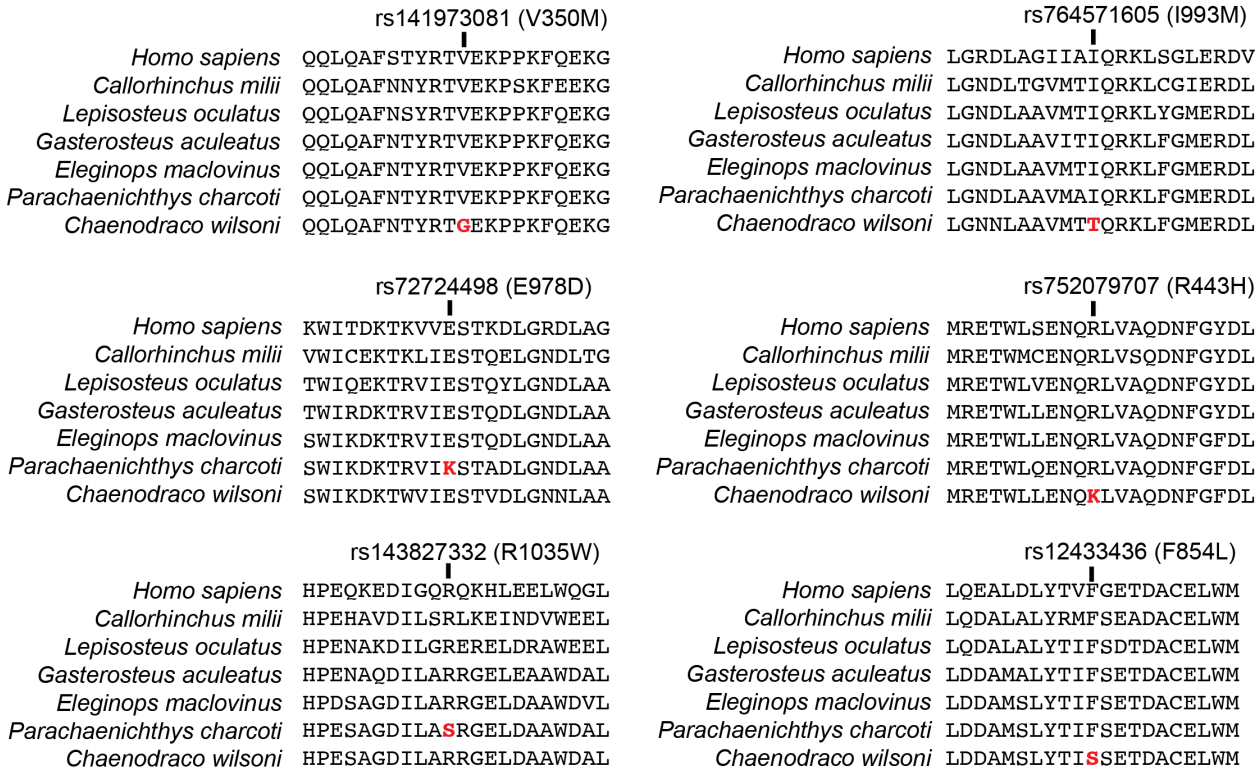

**S5 Fig. Truncating mutations identified in icefish *erythroid-specific 5-aminolevulinate synthase* (*alas2*) gene**. (**A**) Notothenioid phylogeny showing presence of truncating alleles (*) in four icefish species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D-G**) The icefishes show distinct frameshifts and truncations in Alas2 compared to the *N. coriiceps* reference sequence. Alignment start/stop coordinates in D-G are based on position in the *N. coriiceps* genome assembly (XP_010782407.1).



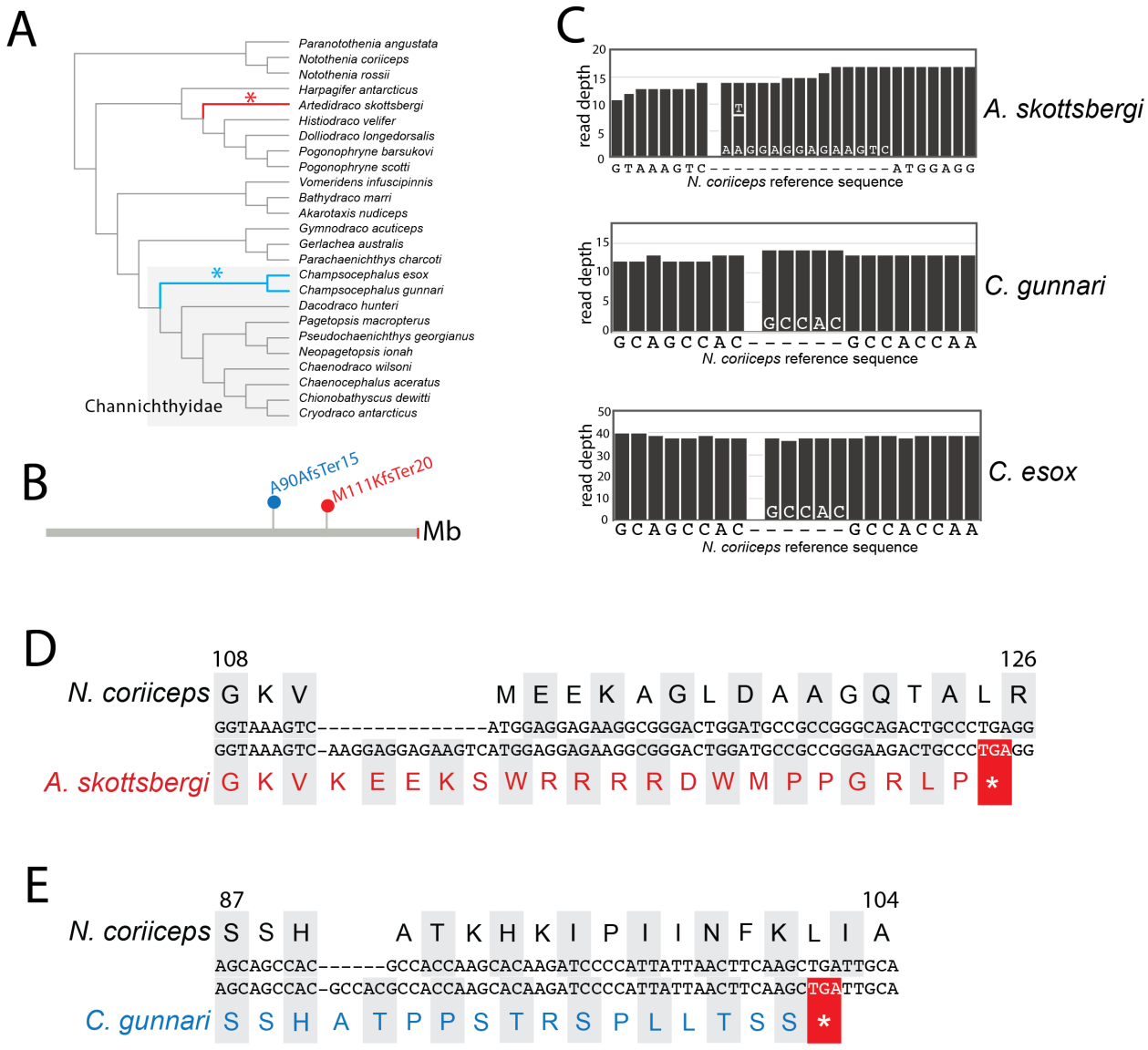

**S6 Fig. Truncating mutation identified in *Pseudochaenichthys georgianus Rh blood group D antigen (rhd)* gene**. (**A**) Notothenioid phylogeny showing presence of a truncating allele in *P. georgianus* (*). (**B**) The mutation encoded by the allele. (**C**) Sequencing read depth for *P. georgianus* as aligned to the *Notothenia coriiceps* reference genome. The gap in read depth corresponds to a deletion in each read relative to the reference genome. (**D**) *P. georgianus* shows a frameshift and truncation in Rhd compared to the *N. coriiceps* reference sequence. Alignment start/stop coordinates are based on position in the *N. coriiceps* genome assembly (XP_010782194.1).



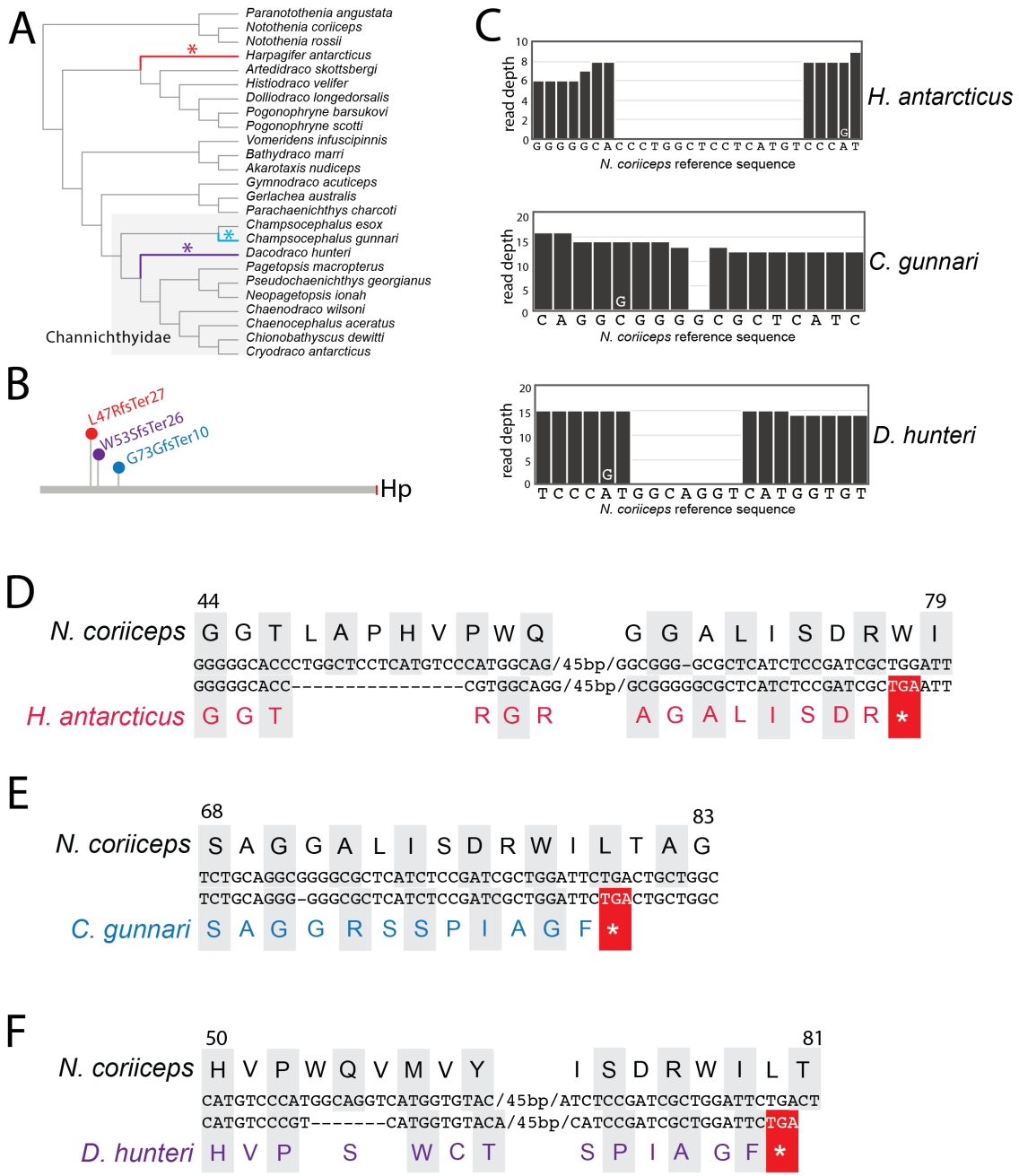

**S7 Fig. Truncating mutations identified in icefish *hemogen* gene**. (**A**) Phylogeny of the notothenioids showing the presence of truncating alleles (*) in three icefish species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D**) *Chaenocephalus aceratus* and *Neopagetopsis ionah* show identical frameshifts and truncations in Hemgn compared to the *N. coriiceps* reference. (**E**) *Pageotopsis macropterus* shows a different frameshift and truncation. Alignment start/stop coordinates in **D** and **E** are based on position in the *N. coriiceps* genome assembly (XP_010773828.1).



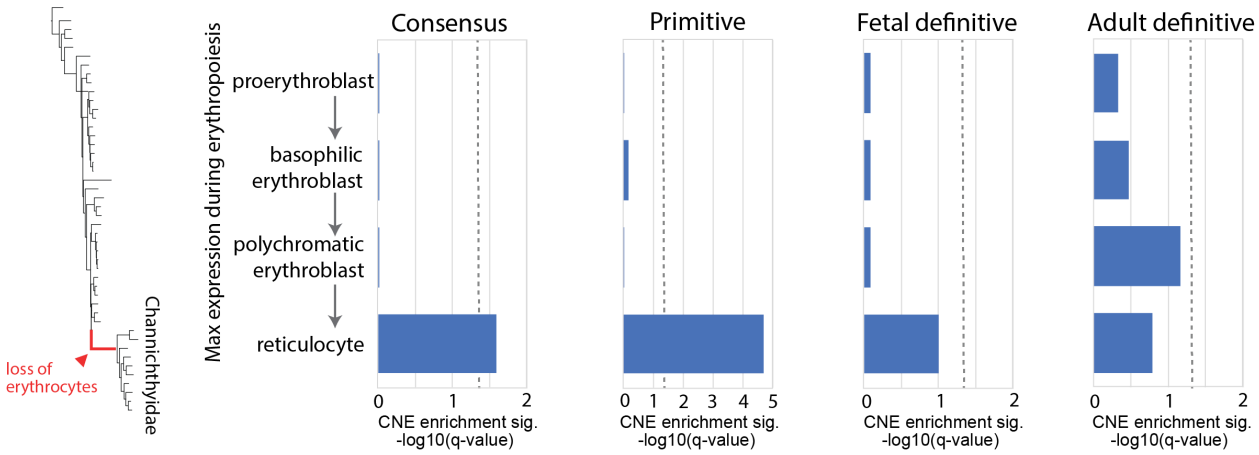

**S8 Fig. Dragonfish and icefish mutations at highly-conserved and clinically-relevant sites in Beta-spectrin.** Variant amino acid substitutions in Beta-spectrin of the dragonfish *Parachaenichthys charcoti* and a representative icefish *Chaenodraco wilsoni* highlighted in red. Beta-spectrin sequences for three-spined stickleback (*Gasterosteus aculeatus*), spotted gar (*Lepisosteus oculatus*), elephant shark (*Gallorhinchus milii*) and human (*Homo sapiens*) are provided for comparison. The dbSNP identifier (ClinVar) for deleterious variants found in human patients with spherocytic anemia/elliptocytosis are shown above each alignment.



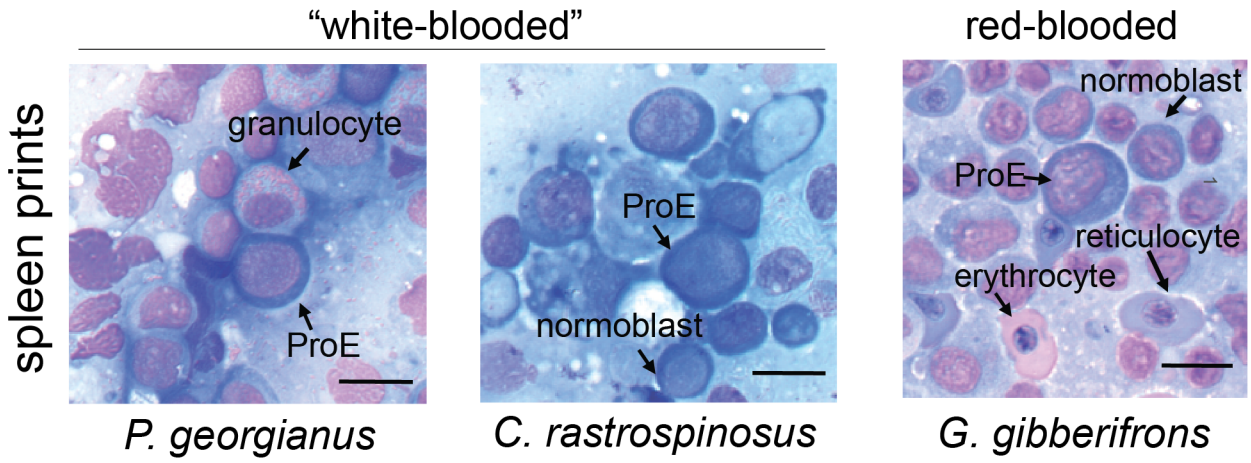

**S9 Fig. Truncating mutations identified in notothenioid *myoglobin* gene**. (**A**) Phylogeny of the notothenioids showing the presence of truncating alleles (*) in three species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D**) Red-blooded species *Artedidraco skottsbergi* Mb compared to the *N. coriiceps* reference. (**E**) *Champsocephalus gunnari* and *C. esox* shows identical frameshifts in Mb. Alignment start/stop coordinates in **D** and **E** are based on position in the *N. coriiceps* genome assembly (NP_001290223.1).




**S10 Fig. Truncating mutations identified in notothenioid *haptoglobin* gene**. (**A**) Phylogeny of the notothenioids showing the presence of truncating alleles (*) in three species. (**B**) Mutant alleles; asterisk color corresponds to branches in **A**. (**C**) Sequencing read depth for each species aligned to the *Notothenia coriiceps* reference genome. Gaps in read depth correspond to deletions in each read relative to the reference genome. (**D**) Red-blooded species *Harpagifer antarcticus* Hp compared to the *N. coriiceps* reference. The icefish species (**E**) *Champsocephalus gunnari* and (**F**) *Dacodraco hunteri* have different frameshifts and truncations in Hp. Alignment start/stop coordinates in **D**-**F** are based on position in the *N. coriiceps* genome assembly (XP_010770321.1).




**S11 Fig. Enrichment for accelerated sequence evolution in conserved non-coding elements (CNEs) near genes that are maximally expressed at distinct stages of erythropoiesis**. Three waves of mammalian erythropoiesis are defined by distinct patterns of gene expression and (locations): primitive (yolk sac blood island), fetal definitive (liver) and adult definitive (bone marrow). For each erythropoietic wave, accelerated evolution of CNEs near maximally expressed genes is shown for four cellular stages of erythroid differentiation/maturation: proerythroblast, basophilic erythroblast/normoblast, polychromatic erythroblast/normoblast, reticulocyte. The Consensus is the intersection of maximally expressed genes across each the three erythropoietic waves. Dashed line corresponds to q-value of 0.05. Gene expression data from ErythronDB [*42*].




**S12 Fig. Spleen prints from three notothenioid species: Wright/Giemsa-stained**. Two "white-blooded" icefishes, *Pseudochaenichthys georgianus* and *Chionodraco rastrospinosus*, show the presence of erythroid progenitors [proerythroblasts (ProEs) and normoblasts] but lack later stages of maturation (e.g., reticulocytes, erythrocytes). By contrast, the red-blooded notothen, *Gobionotothen gibberifrons*, displays the complete erythropoietic progression: ProE → normoblast → reticulocyte → erythrocyte. Scale bar = 10 μm.

**S3 Table. Relative evolutionary rate and notothenioid biogeography**

| Distribution† | Species | Family | RER‡ |
|---|---|---|---|
| HA | *Pogonophryne barsukovi* | Artedidraconidae | 3.160 |
| HA | *Pagetopsis macropterus* | Channichthyidae | 2.856 |
| HA | *Histiodraco velifer* | Artedidraconidae | 2.783 |
| HA | *Trematomus bernacchii* | Nototheniidae | 2.293 |
| HA | *Dacodraco hunteri* | Channichthyidae | 1.599 |
| HA | *Vomeridens infuscipinnis* | Bathydraconidae | 1.297 |
| HA | *Neopagetopsis ionah* | Channichthyidae | 1.163 |
| HA | *Trematomus newnesi* | Nototheniidae | 1.155 |
| HA | *Pleuragramma antarctica* | Nototheniidae | 1.133 |
| sub-Antarctic | *Cottoperca trigloides* | Bovichtidae | 0.927 |
| HA | *Trematomus eulepidotus* | Nototheniidae | 0.829 |
| sub-Antarctic | *Paranotothenia angustata* | Nototheniidae | 0.692 |
| HA | *Chaenodraco wilsoni* | Channichthyidae | 0.460 |
| HA | *Trematomus hansoni* | Nototheniidae | 0.344 |
| HA | *Akarotaxis nudiceps* | Nototheniidae | 0.224 |
| sub-Antarctic | *Bovichtus diacanthus* | Bovichtidae | 0.073 |
| HA | *Cryodraco antarcticus* | Channichthyidae | -0.002 |
| sub-Antarctic | *Champsocephalus esox* | Channichthyidae | -0.167 |
| HA | *Parachaenichthys charcoti* | Bathydraconidae | -0.195 |
| HA | *Pogonophryne scotti* | Artedidraconidae | -0.234 |
| HA | *Aethotaxis mitopteryx* | Nototheniidae | -0.280 |
| HA | *Trematomus borchgrevinki* | Nototheniidae | -0.339 |
| HA | *Notothenia coriiceps* | Nototheniidae | -0.477 |
| HA | *Gerlachea australis* | Bathydraconidae | -0.487 |
| HA | *Trematomus scotti* | Nototheniidae | -0.499 |
| HA | *Dissostichus mawsoni* | Nototheniidae | -0.612 |
| HA | *Lepidonotothen squamifrons* | Nototheniidae | -0.617 |
| sub-Antarctic | *Patagonotothen guntheri* | Nototheniidae | -0.810 |
| sub-Antarctic | *Patagonotothen cornucola* | Nototheniidae | -0.942 |
| HA | *Bathydraco marri* | Bathydraconidae | -1.044 |
| sub-Antarctic | *Dissostichus eleginoides* | Nototheniidae | -1.078 |
| HA | *Chionobathyscus dewitti* | Channichthyidae | -1.152 |

| Distribution† | Species | Family | RER‡ |
|---|---|---|---|
| HA | *Gymnodraco acuticeps* | Bathydraconidae | -1.210 |
| sub-Antarctic | *Eleginops maclovinus* | Eleginopsidae | -1.639 |
| HA | *Dolloidraco longedorsalis* | Artedidraconidae | -1.893 |
| HA | *Artedidraco skottsbergi* | Artedidraconidae | -1.943 |
| sub-Antarctic | *Pseudaphritis urvillii* | Pseudaphritidae | -2.063 |

† HA - high latitude Antarctic

‡ Relative evolutionary rate across all conserved non-coding elements flanking human anemia-associated genes (HP:0001903)

# S4 Table. Coverage and mutations in candidate erythrocyte genes

| Ensembl ID | Gene Name | Type | Avg Coverage† | Avg Coverage Icefish† | FE marrow | FE erythrocytes | Pleiotropy Score | Truncating variant (s) ‡ |
|---|---|---|---|---|---|---|---|---|
| ENSGACG00000006807 | Alas2 | heme and hemoglobin biosynthesis | 97.8% | 100.0% | 6.58 | 2.09 | 0 | ✓ |
| ENSGACG00000001091 | Dmtn | cytoskeleton | 91.2% | 94.1% | 0.37 | 6.30 | 0 | |
| ENSGACG00000017373 | Hemgn | development/transcription factors | 95.7% | 100.0% | 1.61 | 5.18 | 0 | ✓ |
| ENSGACG00000013918 | Hbb | heme and hemoglobin biosynthesis | 78.9% | 13.0% | 8.59 | 2.70 | 0 | ✓ |
| ENSGACG00000014492 | Hba | heme and hemoglobin biosynthesis | 81.0% | 37.2% | 4.99 | 2.56 | 0 | ✓ |
| ENSGACG00000009865 | Rhag | membrane proteins and solute transporters | 71.9% | 73.7% | 23.77 | 6.82 | 0 | |
| ENSGACG00000015628 | Gypc | cytoskeleton | 81.6% | 100.0% | 1.37 | 1.72 | 0 | |
| ENSGACG00000007369 | Rhd | membrane proteins and solute transporters | 94.3% | 98.4% | 5.54 | 8.88 | 0 | ✓ |
| ENSGACG00000007574 | Slc25a28 | membrane proteins and solute transporters | 96.8% | 99.8% | 0.58 | 0.32 | 0 | |
| ENSGACG00000001549 | Tfcp2 | development/transcription factors | 96.1% | 99.5% | 0.16 | 0.69 | 0 | |
| ENSGACG00000009512 | Eif2ak1 | heme and hemoglobin biosynthesis | 73.6% | 74.5% | 0.49 | 1.69 | 1 | |
| ENSGACG00000007530 | Slc25a38 | membrane proteins and solute transporters | 96.1% | 100.0% | 0.54 | 3.16 | 1 | |
| ENSGACG00000007468 | Car6 | carbonic anhydrases | 97.8% | 100.0% | 0.00 | 0.25 | 1 | |
| ENSGACG00000007890 | Car3 | carbonic anhydrases | 74.8% | 80.9% | 0.00 | 0.18 | 1 | |
| ENSGACG00000016482 | Hmgb2 | development/transcription factors | 95.5% | 100.0% | 3.16 | 0.78 | 1 | |
| ENSGACG00000002257 | Kdm1b | development/transcription factors | 97.8% | 100.0% | 0.40 | 0.13 | 1 | |
| ENSGACG00000008971 | Isg15 | development/transcription factors | 62.2% | 63.1% | 0.67 | 0.57 | 1 | |
| ENSGACG00000012154 | Aqp9 | membrane proteins and solute transporters | 97.5% | 100.0% | 0.11 | 0.57 | 2 | |
| ENSGACG00000010765 | Slc25a37 | membrane proteins and solute transporters | 91.2% | 93.4% | 1.64 | 0.94 | 2 | |
| ENSGACG00000014377 | Urod | heme and hemoglobin biosynthesis | 97.7% | 100.0% | 0.91 | 3.53 | 2 | |
| ENSGACG00000003213 | Car5b | carbonic anhydrases | 95.5% | 99.9% | 0.07 | 0.03 | 2 | |
| ENSGACG00000015396 | Car7 | carbonic anhydrases | 97.3% | 100.0% | 0.00 | 1.00 | 2 | |
| ENSGACG00000018021 | Alad | heme and hemoglobin biosynthesis | 97.7% | 100.0% | 0.33 | 5.84 | 3 | |
| ENSGACG00000001433 | Epb41 | cytoskeleton | 95.0% | 98.2% | 0.80 | 0.56 | 3 | |
| ENSGACG00000001495 | Trf | heme and hemoglobin biosynthesis | 96.2% | 99.7% | 0.00 | 0.01 | 3 | |
| ENSGACG00000004180 | Ppox | heme and hemoglobin biosynthesis | 83.2% | 87.7% | 0.44 | 2.33 | 3 | |
| ENSGACG00000018597 | Tmod1 | cytoskeleton | 91.5% | 94.5% | 0.15 | 1.25 | 3 | |
| ENSGACG00000012574 | Slc2a1 | membrane proteins and solute transporters | 96.0% | 99.9% | 0.17 | 0.02 | 3 | |
| ENSGACG00000011803 | Car14 | carbonic anhydrases | 96.9% | 99.9% | 0.02 | 0.03 | 3 | |
| ENSGACG00000018705 | Chd4 | development/transcription factors | 96.5% | 99.5% | 0.33 | 0.35 | 3 | |
| ENSGACG00000020462 | Crebrf | development/transcription factors | 96.8% | 100.0% | 1.12 | 0.14 | 4 | |
| ENSGACG00000018134 | Gfi1b | development/transcription factors | 97.1% | 100.0% | 15.87 | 2.75 | 4 | |
| ENSGACG00000009442 | Nfe2 | development/transcription factors | 97.2% | 100.0% | 5.76 | 0.27 | 4 | |
| ENSGACG00000013189 | Add2 | cytoskeleton | 90.4% | 91.2% | 0.90 | 7.34 | 4 | |
| ENSGACG00000008037 | Ftl1 | heme and hemoglobin biosynthesis | 97.5% | 100.0% | 0.34 | 0.32 | 4 | |
| ENSGACG00000003384 | Xk | membrane proteins and solute transporters | 97.8% | 99.9% | 0.70 | 2.18 | 4 | |
| ENSGACG00000019222 | Bcl11a | development/transcription factors | 97.5% | 100.0% | 0.26 | 0.65 | 4 | |

| Ensembl ID | Gene Name | Type | Avg Coverage† | Avg Coverage Icefish† | FE marrow | FE erythrocytes | Pleiotropy Score | Truncating variant (s) ‡ |
|---|---|---|---|---|---|---|---|---|
| ENSGACG00000019143 | Klf1 | development/transcription factors | 96.0% | 100.0% | 25.70 | 7.32 | 5 | |
| ENSGACG00000009874 | Lmo2 | development/transcription factors | 97.7% | 100.0% | 0.66 | 0.49 | 5 | |
| ENSGACG00000016373 | Tfrc | heme and hemoglobin biosynthesis | 94.8% | 100.0% | 1.10 | 1.98 | 6 | |
| ENSGACG00000010082 | Cpox | heme and hemoglobin biosynthesis | 97.6% | 100.0% | 1.04 | 6.33 | 6 | |
| ENSGACG00000011054 | Myb | development/transcription factors | 96.3% | 99.9% | 0.77 | 0.20 | 6 | |
| ENSGACG00000002554 | Zfpm1 | development/transcription factors | 96.5% | 98.7% | 0.42 | 0.81 | 6 | |
| ENSGACG00000005173 | Hmbs | heme and hemoglobin biosynthesis | 94.7% | 97.3% | 5.39 | 5.59 | 6 | |
| ENSGACG00000012004 | Flvcr1 | heme and hemoglobin biosynthesis | 55.7% | 63.1% | 0.19 | 0.58 | 6 | |
| ENSGACG00000013741 | Dntm1 | development/transcription factors | 97.2% | 99.8% | 0.54 | 0.50 | 6 | |
| ENSGACG00000006591 | Ets1 | development/transcription factors | 97.5% | 100.0% | 0.07 | 0.00 | 6 | |
| ENSGACG00000013704 | Jak2 | development/transcription factors | 97.5% | 100.0% | 0.51 | 0.43 | 6 | |
| ENSGACG00000017365 | Jak2 | development/transcription factors | 97.5% | 100.0% | 0.51 | 0.43 | 6 | |
| ENSGACG00000009622 | Slc4a1 | membrane proteins and solute transporters | 92.1% | 94.8% | 7.93 | 2.74 | 7 | |
| ENSGACG00000015484 | Fth1 | heme and hemoglobin biosynthesis | 96.5% | 100.0% | 1.18 | 0.60 | 7 | |
| ENSGACG00000000651 | Fech | heme and hemoglobin biosynthesis | 94.1% | 96.8% | 2.25 | 4.07 | 8 | |
| ENSGACG00000018336 | Ldb1 | development/transcription factors | 96.8% | 99.8% | 0.44 | 0.60 | 8 | |
| ENSGACG00000015635 | Sox6 | development/transcription factors | 97.8% | 100.0% | 0.64 | 1.11 | 8 | |
| ENSGACG00000010218 | Gata1 | development/transcription factors | 96.7% | 100.0% | 19.53 | 2.31 | 8 | |
| ENSGACG00000002608 | Uros | heme and hemoglobin biosynthesis | 95.4% | 100.0% | 0.51 | 5.32 | 8 | |
| ENSGACG00000004999 | Car2 | carbonic anhydrases | 96.9% | 99.0% | 0.14 | 4.73 | 8 | |
| ENSGACG00000009013 | Acvr1ba | development/transcription factors | 97.7% | 100.0% | 0.15 | 0.01 | 8 | |
| ENSGACG00000000719 | Acvr1bb | development/transcription factors | 97.8% | 100.0% | 0.15 | 0.01 | 8 | |
| ENSGACG00000017383 | Aqp1 | membrane proteins and solute transporters | 97.8% | 100.0% | 0.07 | 5.00 | 9 | |
| ENSGACG00000009608 | Gata2 | development/transcription factors | 96.9% | 99.7% | 0.04 | 0.05 | 9 | |
| ENSGACG00000013846 | Tal1 | development/transcription factors | 97.6% | 100.0% | 1.77 | 1.03 | 9 | |
| ENSGACG00000011100 | Sptb | cytoskeleton | 95.1% | 97.5% | 0.89 | 5.58 | 10 | |
| ENSGACG00000009679 | Kdm1a | development/transcription factors | 97.7% | 100.0% | 0.21 | 0.42 | 10 | |
| ENSGACG00000008634 | Stat5a | development/transcription factors | 97.6% | 100.0% | 0.49 | 0.46 | 10 | |
| ENSGACG00000015405 | Stat5b | development/transcription factors | 97.1% | 99.8% | 0.38 | 0.19 | 11 | |
| ENSGACG00000009373 | Kitl | development/transcription factors | 95.7% | 100.0% | 0.01 | 0.01 | 12 | |
| ENSGACG00000017699 | Ank1 | cytoskeleton | 97.1% | 100.0% | 0.14 | 6.67 | 12 | |
| ENSGACG00000015083 | Acvr2a | development/transcription factors | 97.7% | 99.6% | 0.09 | 0.11 | 13 | |
| ENSGACG00000015589 | Stat1 | development/transcription factors | 96.2% | 99.6% | 0.19 | 0.14 | 13 | |
| ENSGACG00000008910 | Foxo3 | development/transcription factors | 98.1% | 100.0% | 0.81 | 0.47 | 16 | |

† Average coverage across dataset at a minimum depth of 4x reads

‡ Whole gene deletion or truncating variant (nonsense, frameshift) in at least one icefish lineage

## S5 Table. Coverage and mutations in erythroid-biased genes

| Ensembl ID | Gene Name | Avg Coverage† | Avg Coverage Icefish† | FE marrow | FE erythrocytes | Pleiotropy Score | Truncating variant (s) ‡ |
|---|---|---|---|---|---|---|---|
| ENSGACG00000000027 | Mcm2 | 93.3% | 96.4% | 1.27 | 1.18 | 11 | |
| ENSGACG00000004199 | Hdgf | 97.9% | 100.0% | 1.06 | 1.46 | 1 | |
| ENSGACG00000006776 | Mcm5 | 72.1% | 69.9% | 1.05 | 1.41 | 1 | |
| ENSGACG00000009622 | Slc4a1 | 92.1% | 94.8% | 7.93 | 2.74 | 7 | |
| ENSGACG00000007018 | Slc4a1 | 96.7% | 98.4% | 7.93 | 2.74 | 7 | |
| ENSGACG00000002407 | Cdt1 | 94.1% | 98.8% | 4.34 | 1.45 | 0 | |
| ENSGACG00000003179 | Timm23 | 97.8% | 100.0% | 1.34 | 1.02 | 4 | |
| ENSGACG00000018996 | Usp15 | 97.5% | 99.9% | 1.81 | 1.83 | 5 | |
| ENSGACG00000004862 | Josd1 | 83.2% | 86.6% | 1.04 | 1.62 | 1 | |
| ENSGACG00000010082 | Cpox | 97.6% | 100.0% | 1.04 | 6.33 | 6 | |
| ENSGACG00000016373 | Tfrc | 94.8% | 100.0% | 1.10 | 1.98 | 6 | |
| ENSGACG00000005398 | Tfrc | 95.4% | 100.0% | 1.10 | 1.98 | 6 | |
| ENSGACG00000009865 | Rhag | 71.9% | 73.7% | 23.77 | 6.82 | 0 | |
| ENSGACG00000000651 | Fech | 94.1% | 96.8% | 2.25 | 4.07 | 8 | |
| ENSGACG00000006807 | Alas2 | 97.8% | 100.0% | 6.58 | 2.09 | 0 | ✓ |
| ENSGACG00000020793 | Rabgef1 | 68.1% | 71.0% | 1.14 | 2.15 | 4 | |
| ENSGACG00000013350 | Pecam1 | 95.5% | 99.9% | 6.11 | 5.67 | 0 | |
| ENSGACG00000019062 | Tk1 | 97.0% | 100.0% | 1.19 | 1.70 | 10 | |
| ENSGACG00000014938 | Pigq | 96.9% | 99.8% | 1.37 | 3.27 | 1 | |
| ENSGACG00000019155 | Mcm10 | 96.3% | 100.0% | 1.02 | 1.33 | 2 | |
| ENSGACG00000018134 | Gfi1b | 97.1% | 100.0% | 15.87 | 2.75 | 4 | |
| ENSGACG00000017832 | Clp1 | 97.9% | 100.0% | 1.17 | 1.21 | 8 | |
| ENSGACG00000005726 | Pcna | 97.8% | 100.0% | 1.20 | 1.06 | 5 | |
| ENSGACG00000017373 | Hemgn | 95.7% | 100.0% | 1.61 | 5.18 | 0 | ✓ |
| ENSGACG00000005437 | Orc1 | 81.5% | 86.2% | 2.27 | 2.43 | 1 | |
| ENSGACG00000013846 | Tal1 | 97.6% | 100.0% | 1.77 | 1.03 | 9 | |
| ENSGACG00000007369 | Rhd | 94.3% | 98.4% | 5.54 | 8.88 | 0 | ✓ |
| ENSGACG00000010218 | Gata1 | 96.7% | 100.0% | 19.53 | 2.31 | 8 | |
| ENSGACG00000016176 | Abcb10 | 89.8% | 92.4% | 1.97 | 3.94 | 3 | |
| ENSGACG00000005173 | Hmbs | 94.7% | 97.3% | 5.39 | 5.59 | 6 | |
| ENSGACG00000012552 | Blvrb | 97.1% | 100.0% | 2.17 | 2.85 | 2 | |
| ENSGACG00000004430 | Rpia | 97.3% | 100.0% | 1.14 | 1.63 | 2 | |
| ENSGACG00000019143 | Klf1 | 96.0% | 100.0% | 25.70 | 7.32 | 5 | |
| ENSGACG00000014492 | Hba-a1 | 79.9% | 37.2% | 4.99 | 2.56 | 0 | ✓ |
| ENSGACG00000004078 | Fastkd5 | 97.4% | 100.0% | 1.20 | 1.27 | 4 | |
| ENSGACG00000015628 | Gypc | 81.6% | 100.0% | 1.37 | 1.72 | 0 | |
| ENSGACG00000013918 | Hbb | 79.9% | 13.6% | 8.59 | 2.70 | 0 | ✓ |

† Average coverage across dataset at a minimum depth of 4x reads

‡ Whole gene deletion or truncating variant (nonsense, frameshift) in at least one icefish lineage

**S6 Table. Excluded terms from Mammalian Phenotype Ontology (MP) in pleiotropy analysis**

| MP Term | Name |
|---|---|
| MP:0000202 | abnormal circulating alkaline phosphatase level |
| MP:0000208 | decreased hematocrit |
| MP:0000215 | absent erythrocytes |
| MP:0000226 | abnormal mean corpuscular volume |
| MP:0000233 | abnormal blood flow velocity |
| MP:0000237 | obsolete decreased blood cell number |
| MP:0000245 | abnormal erythropoiesis |
| MP:0000248 | macrocytosis |
| MP:0000256 | echinocytosis |
| MP:0000314 | schistocytosis |
| MP:0000315 | hemoglobinuria |
| MP:0000332 | hemoglobinemia |
| MP:0000348 | abnormal aerobic fitness |
| MP:0000603 | pale liver |
| MP:0000689 | abnormal spleen morphology |
| MP:0000734 | muscle hypoplasia |
| MP:0000748 | progressive muscle weakness |
| MP:0000752 | dystrophic muscle |
| MP:0000759 | abnormal skeletal muscle morphology |
| MP:0001189 | absent skin pigmentation |
| MP:0001190 | reddish skin |
| MP:0001191 | abnormal skin condition |
| MP:0001201 | translucent skin |
| MP:0001264 | increased body size |
| MP:0001265 | decreased body size |
| MP:0001569 | abnormal circulating bilirubin level |
| MP:0001574 | abnormal oxygen level |
| MP:0001577 | anemia |
| MP:0001585 | hemolytic anemia |
| MP:0001586 | abnormal erythrocyte cell number |
| MP:0001588 | abnormal hemoglobin |
| MP:0001589 | abnormal mean corpuscular hemoglobin |
| MP:0001598 | abnormal blood viscosity |
| MP:0001599 | abnormal blood volume |
| MP:0001697 | abnormal embryo size |
| MP:0001698 | decreased embryo size |

MP:0001699 increased embryo size
MP:0001721 absent visceral yolk sac blood islands
MP:0001722 pale yolk sac
MP:0001730 embryonic growth arrest
MP:0001731 abnormal postnatal growth
MP:0001732 postnatal growth retardation
MP:0001770 abnormal iron level
MP:0001786 skin edema
MP:0001933 abnormal litter size
MP:0001934 increased litter size
MP:0001935 decreased litter size
MP:0002088 abnormal embryonic growth/weight/body size
MP:0002089 abnormal postnatal growth/weight/body size
MP:0002095 abnormal skin pigmentation
MP:0002106 abnormal muscle physiology
MP:0002108 abnormal muscle morphology
MP:0002224 abnormal spleen size
MP:0002225 obsolete abnormal spleen cellularity
MP:0002227 abnormal spleen capsule morphology
MP:0002228 abnormal spleen trabecular vein morphology
MP:0002288 obsolete litter size
MP:0002319 hyperoxia
MP:0002329 abnormal blood gas level
MP:0002354 abnormal spleen trabecular artery morphology
MP:0002355 obsolete abnormal spleen venous sinus
MP:0002356 abnormal spleen red pulp morphology
MP:0002357 abnormal spleen white pulp morphology
MP:0002358 abnormal spleen periarteriolar lymphoid sheath morphology
MP:0002359 abnormal spleen germinal center morphology
MP:0002361 abnormal spleen central arteriole morphology
MP:0002362 abnormal spleen marginal zone morphology
MP:0002363 abnormal spleen marginal sinus morphology
MP:0002424 abnormal reticulocyte morphology
MP:0002447 abnormal erythrocyte morphology
MP:0002591 decreased mean corpuscular volume
MP:0002592 obsolete mean erythrocyte count traits
MP:0002593 high mean erythrocyte cell number
MP:0002594 low mean erythrocyte cell number
MP:0002596 abnormal hematocrit
MP:0002608 increased hematocrit

MP:0002640 reticulocytosis
MP:0002641 anisopoikilocytosis
MP:0002642 anisocytosis
MP:0002643 poikilocytosis
MP:0002810 microcytic anemia
MP:0002811 macrocytic anemia
MP:0002812 spherocytosis
MP:0002813 microcytosis
MP:0002814 hyperchromasia
MP:0002874 decreased hemoglobin content
MP:0002875 decreased erythrocyte cell number
MP:0002897 blotchy skin
MP:0002954 obsolete abnormal aerobic energy metabolism
MP:0002966 decreased circulating alkaline phosphatase level
MP:0002968 increased circulating alkaline phosphatase level
MP:0003015 abnormal circulating bicarbonate level
MP:0003016 increased circulating bicarbonate level
MP:0003017 decreased circulating bicarbonate level
MP:0003060 increased aerobic running capacity
MP:0003131 increased erythrocyte cell number
MP:0003342 accessory spleen
MP:0003396 abnormal embryonic hematopoiesis
MP:0003656 abnormal erythrocyte physiology
MP:0003657 abnormal erythrocyte osmotic lysis
MP:0003717 pallor
MP:0003852 skeletal muscle necrosis
MP:0003956 abnormal body size
MP:0003984 embryonic growth retardation
MP:0004142 abnormal muscle tone
MP:0004143 muscle hypertonia
MP:0004151 decreased circulating iron level
MP:0004152 abnormal circulating iron level
MP:0004196 abnormal prenatal growth/weight/body size
MP:0004197 abnormal fetal growth/weight/body size
MP:0004198 abnormal fetal size
MP:0004199 increased fetal size
MP:0004200 decreased fetal size
MP:0004201 fetal growth retardation
MP:0004229 abnormal embryonic erythropoiesis
MP:0004230 abnormal embryonic erythrocyte morphology

MP:0004232 decreased muscle weight
MP:0004233 abnormal muscle weight
MP:0004797 increased anti-erythrocyte antigen antibody level
MP:0004817 abnormal skeletal muscle mass
MP:0004818 increased skeletal muscle mass
MP:0004819 decreased skeletal muscle mass
MP:0004827 increased susceptibility to autoimmune hemolytic anemia
MP:0004828 decreased susceptibility to autoimmune hemolytic anemia
MP:0004846 absent skeletal muscle
MP:0004951 abnormal spleen weight
MP:0004952 increased spleen weight
MP:0004953 decreased spleen weight
MP:0004969 pale kidney
MP:0005028 abnormal trophectoderm morphology
MP:0005097 polychromatophilia
MP:0005152 pancytopenia
MP:0005288 abnormal oxygen consumption
MP:0005289 increased oxygen consumption
MP:0005290 decreased oxygen consumption
MP:0005344 increased circulating bilirubin level
MP:0005369 muscle phenotype
MP:0005406 abnormal heart size
MP:0005505 thrombocytosis
MP:0005561 increased mean corpuscular hemoglobin
MP:0005562 decreased mean corpuscular hemoglobin
MP:0005563 abnormal hemoglobin content
MP:0005564 increased hemoglobin content
MP:0005635 decreased circulating bilirubin level
MP:0005637 abnormal iron homeostasis
MP:0005640 abnormal mean corpuscular hemoglobin concentration
MP:0005641 increased mean corpuscular hemoglobin concentration
MP:0005642 decreased mean corpuscular hemoglobin concentration
MP:0005649 increased spleen neoplasm incidence
MP:0006034 myoglobinuria
MP:0006208 lethality throughout fetal growth and development
MP:0006351 abnormal glycosylated hemoglobin level
MP:0006352 decreased glycosylated hemoglobin level
MP:0006353 increased glycosylated hemoglobin level
MP:0008234 absent spleen marginal zone
MP:0008387 hypochromic anemia

MP:0008388 hypochromic microcytic anemia
MP:0008389 hypochromic macrocytic anemia
MP:0008473 abnormal spleen follicular dendritic cell network
MP:0008474 absent spleen germinal center
MP:0008475 intermingled spleen red and white pulp
MP:0008476 increased spleen red pulp amount
MP:0008477 decreased spleen red pulp amount
MP:0008478 increased spleen white pulp amount
MP:0008479 decreased spleen white pulp amount
MP:0008481 increased spleen germinal center number
MP:0008482 decreased spleen germinal center number
MP:0008483 increased spleen germinal center size
MP:0008484 decreased spleen germinal center size
MP:0008737 abnormal spleen physiology
MP:0008738 abnormal liver iron level
MP:0008739 abnormal spleen iron level
MP:0008740 abnormal intestinal iron level
MP:0008741 abnormal heart iron level
MP:0008742 abnormal kidney iron level
MP:0008743 decreased liver iron level
MP:0008772 increased heart ventricle size
MP:0008807 increased liver iron level
MP:0008808 decreased spleen iron level
MP:0008809 increased spleen iron level
MP:0008810 increased circulating iron level
MP:0008849 abnormal hemoglobin concentration distribution width
MP:0008850 increased hemoglobin concentration distribution width
MP:0008851 decreased hemoglobin concentration distribution width
MP:0008941 reticulocytopenia
MP:0008945 hyperchromic macrocytic anemia
MP:0008954 abnormal cellular hemoglobin content
MP:0008955 increased cellular hemoglobin content
MP:0008956 decreased cellular hemoglobin content
MP:0008962 abnormal carbon dioxide production
MP:0008963 increased carbon dioxide production
MP:0008964 decreased carbon dioxide production
MP:0009246 pale spleen
MP:0009323 abnormal spleen development
MP:0009395 increased nucleated erythrocyte cell number
MP:0009398 abnormal skeletal muscle fiber size

MP:0009399 increased skeletal muscle fiber size
MP:0009403 increased variability of skeletal muscle fiber size
MP:0009405 increased skeletal muscle fiber number
MP:0009406 decreased skeletal muscle fiber number
MP:0009408 decreased skeletal muscle fiber density
MP:0009409 abnormal skeletal muscle fiber type ratio
MP:0009410 abnormal skeletal muscle satellite cell proliferation
MP:0009411 abnormal skeletal muscle fiber triad morphology
MP:0009412 skeletal muscle fiber degeneration
MP:0009413 skeletal muscle fiber atrophy
MP:0009414 skeletal muscle fiber necrosis
MP:0009415 skeletal muscle degeneration
MP:0009416 cardiac muscle degeneration
MP:0009417 skeletal muscle atrophy
MP:0009418 cardiac muscle atrophy
MP:0009458 abnormal skeletal muscle size
MP:0009459 skeletal muscle hyperplasia
MP:0009460 skeletal muscle hypoplasia
MP:0009461 skeletal muscle hypertrophy
MP:0009462 skeletal muscle hypotrophy
MP:0009547 elliptocytosis
MP:0009568 abnormal red blood cell deformability
MP:0009642 abnormal blood homeostasis
MP:0009701 abnormal birth body size
MP:0009702 increased birth body size
MP:0009703 decreased birth body size
MP:0009841 foam cell reticulosis
MP:0009931 abnormal skin appearance
MP:0010020 spleen vascular congestion
MP:0010034 abnormal erythrocyte clearance
MP:0010035 increased erythrocyte clearance
MP:0010036 decreased erythrocyte clearance
MP:0010067 increased red blood cell distribution width
MP:0010068 decreased red blood cell distribution width
MP:0010074 stomatocytosis
MP:0010175 leptocytosis
MP:0010176 dacryocytosis
MP:0010177 acanthocytosis
MP:0010178 increased number of Howell-Jolly bodies
MP:0010237 abnormal skeletal muscle weight

| | |
|---|---|
| MP:0010238 | increased skeletal muscle weight |
| MP:0010239 | decreased skeletal muscle weight |
| MP:0010240 | decreased skeletal muscle size |
| MP:0010245 | abnormal spleen perifollicular zone morphology |
| MP:0010375 | increased kidney iron level |
| MP:0010376 | decreased kidney iron level |
| MP:0010399 | decreased skeletal muscle glycogen level |
| MP:0010401 | increased skeletal muscle glycogen level |
| MP:0010563 | increased heart right ventricle size |
| MP:0010577 | abnormal heart right ventricle size |
| MP:0010579 | increased heart left ventricle size |
| MP:0010580 | decreased heart left ventricle size |
| MP:0010630 | abnormal cardiac muscle tissue morphology |
| MP:0010632 | cardiac muscle necrosis |
| MP:0010696 | increased siderocyte number |
| MP:0010832 | lethality during fetal growth through weaning |
| MP:0010865 | prenatal growth retardation |
| MP:0010866 | abnormal prenatal body size |
| MP:0010957 | abnormal aerobic respiration |
| MP:0011089 | perinatal lethality, complete penetrance |
| MP:0011091 | prenatal lethality |
| MP:0011098 | embryonic lethality during organogenesis, complete penetrance |
| MP:0011099 | lethality throughout fetal growth and development, complete penetrance |
| MP:0011101 | prenatal lethality, incomplete penetrance |
| MP:0011109 | lethality throughout fetal growth and development, incomplete penetrance |
| MP:0011111 | lethality during fetal growth through weaning, complete penetrance |
| MP:0011112 | lethality during fetal growth through weaning, incomplete penetrance |
| MP:0011171 | increased number of Heinz bodies |
| MP:0011188 | increased erythrocyte protoporphyrin level |
| MP:0011204 | abnormal visceral yolk sac blood island morphology |
| MP:0011235 | abnormal blood oxygen capacity |
| MP:0011236 | increased blood oxygen capacity |
| MP:0011237 | decreased blood oxygen capacity |
| MP:0011239 | abnormal skin coloration |
| MP:0011240 | abnormal fetal derived definitive erythrocyte morphology |
| MP:0011241 | abnormal fetal derived definitive erythrocyte cell number |
| MP:0011242 | increased fetal derived definitive erythrocyte cell number |
| MP:0011243 | decreased fetal derived definitive erythrocyte cell number |

MP:0011244 absent fetal derived definitive erythrocytes
MP:0011245 abnormal fetal derived definitive erythrocyte physiology
MP:0011263 abnormal spleen mesenchyme morphology
MP:0011514 skin hemorrhage
MP:0011519 abnormal placenta labyrinth size
MP:0011520 increased placental labyrinth size
MP:0011521 decreased placental labyrinth size
MP:0011526 abnormal placenta fetal blood space morphology
MP:0011630 increased mitochondria size
MP:0011631 decreased mitochondria size
MP:0011890 increased circulating ferritin level
MP:0011891 decreased circulating ferritin level
MP:0011892 abnormal circulating transferrin level
MP:0011893 increased circulating transferrin level
MP:0011894 decreased circulating transferrin level
MP:0011895 abnormal circulating unsaturated transferrin level
MP:0011896 increased circulating unsaturated transferrin level
MP:0011897 decreased circulating unsaturated transferrin level
MP:0011913 abnormal reticulocyte cell number
MP:0011992 increased erythrocyte catalase activity
MP:0012056 abnormal polar trophectoderm morphology
MP:0012057 abnormal mural trophectoderm morphology
MP:0012102 absent trophectoderm
MP:0012115 abnormal trophectoderm cell proliferation
MP:0012116 increased trophectoderm cell proliferation
MP:0012117 decreased trophectoderm cell proliferation
MP:0012118 absent trophectoderm cell proliferation
MP:0012119 increased trophectoderm apoptosis
MP:0012120 trophectoderm cell degeneration
MP:0012363 abnormal erythrocyte sodium level
MP:0012364 decreased erythrocyte sodium level
MP:0012365 increased erythrocyte sodium level
MP:0012366 abnormal erythrocyte magnesium level
MP:0012367 decreased erythrocyte magnesium level
MP:0012368 increased erythrocyte magnesium level
MP:0012369 abnormal erythrocyte potassium level
MP:0012370 decreased erythrocyte potassium level
MP:0012371 increased erythrocyte potassium level
MP:0012372 abnormal erythrocyte ion content
MP:0012373 abnormal erythrocyte magnesium ion content

MP:0012374 decreased erythrocyte magnesium ion content
MP:0012375 increased erythrocyte magnesium ion content
MP:0012376 abnormal erythrocyte potassium ion content
MP:0012377 decreased erythrocyte potassium ion content
MP:0012378 increased erythrocyte potassium ion content
MP:0012379 abnormal erythrocyte sodium ion content
MP:0012380 decreased erythrocyte sodium ion content
MP:0012381 increased erythrocyte sodium ion content
MP:0012384 abnormal erythrocyte ion transport
MP:0012385 abnormal erythrocyte potassium:chloride symporter activity
MP:0012386 decreased erythrocyte potassium:chloride symporter activity
MP:0012387 increased erythrocyte potassium:chloride symporter activity
MP:0012388 abnormal erythrocyte sodium:hydrogen antiporter activity
MP:0012389 decreased erythrocyte sodium:hydrogen antiporter activity
MP:0012390 increased erythrocyte sodium:hydrogen antiporter activity
MP:0012391 abnormal erythrocyte sodium:potassium-exchanging ATPase activity
MP:0012392 decreased erythrocyte sodium:potassium-exchanging ATPase activity
MP:0012393 increased erythrocyte sodium:potassium-exchanging ATPase activity
MP:0012394 abnormal erythrocyte calcium-activated potassium channel activity
MP:0012395 decreased erythrocyte calcium-activated potassium channel activity
MP:0012396 increased erythrocyte calcium-activated potassium channel activity
MP:0012397 abnormal nucleated erythrocyte cell number
MP:0012398 decreased nucleated erythrocyte cell number
MP:0012650 abnormal erythrocyte catalase level
MP:0012653 decreased erythrocyte catalase level
MP:0012656 increased erythrocyte catalase level
MP:0012663 decreased haptoglobin level
MP:0012664 decreased circulating haptoglobin level
MP:0012665 increased haptoglobin level
MP:0012666 increased circulating haptoglobin level
MP:0013215 abnormal haptoglobin level
MP:0013301 abnormal pancreas iron level
MP:0013302 increased pancreas iron level
MP:0013303 decreased pancreas iron level
MP:0013403 abnormal circulating lactate level
MP:0013404 decreased circulating lactate level
MP:0013405 increased circulating lactate level
MP:0013657 abnormal blood cell morphology
MP:0020240 increased skeletal muscle cell apoptosis
MP:0020241 decreased skeletal muscle cell apoptosis

| | |
|---|---|
| MP:0020323 | abnormal heart apex size |
| MP:0020365 | increased brain iron level |
| MP:0020366 | decreased brain iron level |
| MP:0020367 | increased heart iron level |
| MP:0020368 | decreased heart iron level |
| MP:0020369 | increased intestinal iron level |
| MP:0020453 | abnormal erythrocyte aggregation |
| MP:0020454 | decreased erythrocyte aggregation |
| MP:0020455 | increased erythrocyte aggregation |
| MP:0020825 | ectopic spleen |