






## ARTICLE

<https://doi.org/10.1038/s42003-019-0685-y>

OPEN

# Draft genome assembly and transcriptome data of the icefish *Chionodraco myersi* reveal the key role of mitochondria for a life without hemoglobin at subzero temperatures

Luca Bargelloni <sup>1,2,3\*</sup>, Massimiliano Babbucci<sup>1</sup>, Serena Ferrareso<sup>1</sup>, Chiara Papetti<sup>3,4</sup>, Nicola Vitulo <sup>5</sup>, Roberta Carraro<sup>1</sup>, Marianna Pauletto <sup>1</sup>, Gianfranco Santovito<sup>3</sup>, Magnus Lucassen<sup>6</sup>, Felix Christopher Mark <sup>6</sup>, Lorenzo Zane <sup>2,3</sup> & Tomaso Patarnello<sup>1,3</sup>

Antarctic fish belonging to Notothenioidei represent an extraordinary example of radiation in the cold. In addition to the absence of hemoglobin, icefish show a number of other striking peculiarities including large-diameter blood vessels, high vascular densities, mitochondria-rich muscle cells, and unusual mitochondrial architecture. In order to investigate the bases of icefish adaptation to the extreme Southern Ocean conditions we sequenced the complete genome of the icefish *Chionodraco myersi*. Comparative analyses of the icefish genome with those of other teleost species, including two additional white-blooded and five red-blooded notothenioids, provided a new perspective on the evolutionary loss of globin genes. Muscle transcriptome comparative analyses against red-blooded notothenioids as well as temperate fish revealed the peculiar regulation of genes involved in mitochondrial function in icefish. Gene duplication and promoter sequence divergence were identified as genome-wide patterns that likely contributed to the broad transcriptional program underlying the unique features of icefish mitochondria.

<sup>1</sup> Department of Comparative Biomedicine and Food Science, University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy. <sup>2</sup> Department of Land, Environment, Agriculture, and Forestry, University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy. <sup>3</sup> Consorzio Nazionale Interuniversitario per le Scienze del Mare (CoNISMa), Piazzale Flaminio 9, 00196 Rome, Italy. <sup>4</sup> Department of Biology, University of Padova, Via G. Colombo 3, 35131 Padua, Italy. <sup>5</sup> Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy. <sup>6</sup> Section of Integrative Ecophysiology, Alfred Wegener Institute Helmholtz. Centre for Polar and Marine Research, Am Handelshafen 12, Bremerhaven 27570, Germany. \*email: [luca.bargelloni@unipd.it](mailto:luca.bargelloni@unipd.it)

Antarctic fish belonging to the family Channichthyidae are the only vertebrates that lack hemoglobin as first reported by Ruud<sup>1</sup>. They were named icefish by the British whalers sailing the Southern Ocean because of their translucent aspect, due to the absence of respiratory pigments and scaleless skin. The icefish are part of a broader taxonomic group, the Notothenioidei, which dominate the Antarctic fish fauna<sup>2</sup>. Notothenioids represent an extraordinary example of radiation in the cold, as they have filled several ecological niches emptied by the dramatic cooling of the Southern Ocean in the last 25 myr<sup>3,4</sup>. The extremely low water temperature ( $-1.8^{\circ}\text{C}$ ) is also considered the necessary condition for the evolutionary loss of hemoglobin in the icefish. According to Henry's law<sup>5</sup>, oxygen solubility is inversely correlated with water temperature, therefore in the subzero Southern Ocean the concentration of dissolved oxygen is substantially higher than in temperate waters. Such a condition, however, is not sufficient to compensate for the lack of hemoglobin. In fact, the icefish shows several cardiovascular modifications (e.g., greater blood volume, wider capillaries, and larger hearts) to offset the absence of oxygen carriers (reviewed by Sidell and O'Brien<sup>6</sup>). At the cellular level, it has long been reported that icefish show extraordinarily high mitochondrial density<sup>7,8</sup>. An increased ratio between mitochondrial and muscle cell volume appears to be a more general response to cold-acclimation in fish<sup>9–11</sup>. The Antarctic notothenioids and, in particular, the hemoglobinless icefish represent an extreme example of such response<sup>12</sup>. The role of mitochondria in cold-acclimation of ectothermic vertebrates is rather complex and remains to be fully elucidated. Going beyond the classic hypothesis that higher mitochondrial density provides compensatory increase of aerobic capacity during cold-acclimation, Sidell<sup>13</sup> suggested that denser mitochondrial membranes might favor intracellular oxygen transport. Oxygen diffusion, which is reduced in aqueous solutions at low temperatures, is greatly enhanced in lipid bilayers. In the icefish, where oxygen transport cannot be facilitated by hemoglobin, mitochondrial lipid membranes might be the key to ensure its efficient diffusion and storage within the cell. In fact, the white-blooded channichthyids have higher phospholipid mitochondrial content and smaller surface-to-volume ratio compared to red-blooded notothenioids<sup>14</sup>, which suggests a larger degree of mitochondrial fusion relative to fission<sup>15</sup>. The large icefish mitochondria, however, show sparser mitochondrial cristae, which are the mitochondrial substructures where respiratory complexes are located<sup>16</sup>. This is in keeping with the evidence that channichthyids do not show a proportionally higher aerobic metabolic capacity<sup>8,17</sup>. In recent years, there has been a resurging interest in mitochondrial biology, largely due to the appreciation of the role of mitochondria in multiple functions in addition to ATP production<sup>18</sup>. Studies on mitochondrial dynamics<sup>16</sup> and their interactions with other cellular organelles<sup>19</sup> reveal an unsuspected complexity and provided clues to explore the role of mitochondria in the adaptation to life without hemoglobin.

The evolution of the unique hemoglobinless condition of icefish has understandably attracted much attention and several studies investigated the molecular basis of hemoglobin loss and addressed the question whether it might be a form of adaptation to the extreme Antarctic environment or, conversely, it represents an example of disaptation<sup>6,20</sup>. Cocco et al.<sup>21,22</sup> characterized the remnants of the globin locus. They found that the same partial copy of one alpha-globin gene was present in several white-blooded species, differently from beta-globin genes, which were not detected. They hypothesized that a single event of deletion in the common ancestor of channichthyids might have been the driver of the evolution of the hemoglobinless condition. A subsequent study<sup>23</sup> discovered that a highly derived channichthyid species, *Neopagetopsis ionah*, shows a complete copy of both

alpha- and beta-globin genes, calling into question the hypothesis of a single event for the evolutionary loss of hemoglobin. Both studies, however, assumed a single-genomic locus for alpha- and beta-globins. Genomic analysis showed that two genomic globin clusters are present in the teleost genome, with multiple copies of globin genes within each cluster<sup>24</sup>, prompting for whole-genome investigations on globin gene evolution in the Notothenioidei. In parallel to the complete absence of hemoglobin, icefish show very few circulating erythrocytes. Comparing the transcriptome of primary erythropoietic tissues between a white-blooded species, *Chionodraco hamatus*, and two red-blooded notothenioids, Xu et al.<sup>25</sup> demonstrated the silencing of several key transcription factors involved in erythropoiesis, possibly mediated by upregulation of suppressive microRNAs. A second transcriptome study reported that microRNAs required for erythropoiesis are conserved in white-blooded notothenioids<sup>26</sup>. At variance with the nearly complete deletion of globin genes, these studies suggest that downregulation rather than gene loss underlies the limited presence of erythrocytes in icefish blood, although confirmation at the genome level is still lacking.

A decade ago, Chen et al.<sup>27</sup> discovered a general pattern in the evolution of the nototheniid genome. Using EST data and microarray-based comparative genomic hybridization they showed that massive gene duplication led to the dramatic expansion of several gene families. Such genomic evolutionary pattern was later demonstrated to provide the basis for the neofunctionalization of zona pellucida (ZP)-like proteins, which are crucial for freeze prevention of nototheniid eggs<sup>28</sup>. Similarly, the analysis of the muscle transcriptome of the icefish *C. hamatus* discovered selective duplication of genes encoding mitochondrial proteins compared to other teleost species<sup>29</sup>. The role of gene duplication in Antarctic fish evolution has been further confirmed after the recent publication of the draft genomes of two red-blooded notothenioids, *Eleginops maclovinus* and *Dissostichus mawsoni*<sup>30</sup> and one white-blooded species, *Chaenocephalus aceratus*<sup>31</sup>. Such genomic information also provided us with the unprecedented opportunity for a comparative analysis at the genome and transcriptome level between red- and white-blooded notothenioids. Such comparative analysis included *E. maclovinus*, which the closest outgroup of all Antarctic notothenioids and it belongs to a lineage that never inhabited the waters surrounding Antarctica and *P. charcoti*, which belongs to the family Bathydraconidae, which represents the closest living relatives to the hemoglobinless channichthyids<sup>32,33</sup>. Here, we report the draft genome assembly and muscle transcriptome data for a second channichthyid species, *C. myersi*, and we identify the genomic and transcriptomic patterns underlying the key role of mitochondria in the evolution of this unique group of vertebrates.

## Results

**Icefish genome assembly and annotation.** The draft genome of *C. myersi* was assembled into 63,605 scaffolds for a total size of 1.12 Gbp and contained 38,127 putative protein-coding genes. The estimated genome size is very similar to that reported for another icefish (1.1 Gbp, *C. aceratus*)<sup>31</sup> and larger than those of two red-blooded nototheniid species (0.73 Gbp, *E. maclovinus*; 0.84 Gbp *D. mawsoni*)<sup>30</sup>. For all these species, genome size based on k-mer analysis is lower than estimated using DNA content and BAC library sequencing by Detrich et al.<sup>34</sup>. It is difficult to explain the reason of such a discrepancy, although a likely hypothesis is that the high content in repetitive sequences, especially for white-blooded notothenioids<sup>34</sup> might lead to underestimating genome size when using next-generation sequence data. Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis showed good completeness (86.8%) and

representation (93.5%). The same analysis was run for four more notothenioid genomes (*Notothenia coriiceps*, *C. aceratus*, *E. maclovinus*, *D. mawsoni*; Supplementary Table 1) showing that the level of completeness and representation for *C. myersi* is within the range observed for these species (80–97%). All details about *C. myersi* genome sequencing, assembly, k-mer analysis, and annotation are reported in Supplementary Methods, in Supplementary Tables 2–4 and in Supplementary Fig. 1. Putative protein-coding sequences for *C. myersi* were analyzed together with 18 other teleost genomes, including other notothenioid species, and 21,718 orthology groups (OGs) were identified (Supplementary Table 5).

**Whole-genome analysis of globin gene loss and the evolution of the erythropoiesis pathway.** Conserved flanking orthologs were used as anchors to identify the contiguous genomic regions corresponding to the two teleost globin gene clusters (MN and LA as defined in Opazo et al.<sup>24</sup>) in the hemoglobinless icefish *C. myersi* and *C. hamatus*. For this purpose, a draft genome of *C. hamatus* was assembled (see Supplementary Methods section). Kim et al.<sup>31</sup> recently compared the genomic organization of the LA and MN clusters between the icefish *C. aceratus* and temperate fish species. Here, we included in the comparison seven notothenioid species<sup>30,31,35–37</sup>, namely three white-blooded icefish and four red-blooded notothenioids, *N. coriiceps*, *P. charcoti*, *D. mawsoni*, and *E. maclovinus* (Fig. 1a). *Eleginops maclovinus* is particularly interesting as it is the closest outgroup of all Antarctic notothenioids and it belongs to a lineage that never inhabited the waters surrounding Antarctica, while *P. charcoti* belongs to the family Bathyaconidae, which represents the closest living relatives to the hemoglobinless channichthyids<sup>32,33</sup>.

In the LA cluster only one alpha-globin fragment was found in the three icefish genomes, as previously reported<sup>22</sup>, while cluster MN was precisely defined by orthologous anchors, but showed no trace of globin-like sequences as already described<sup>31</sup>. The same analysis carried out on the genomes of the four red-blooded notothenioids was extended to the sub-Antarctic species *E. maclovinus* that showed the presence of up to 10 full-length genes encoding beta- and alpha-globins (Fig. 1a).

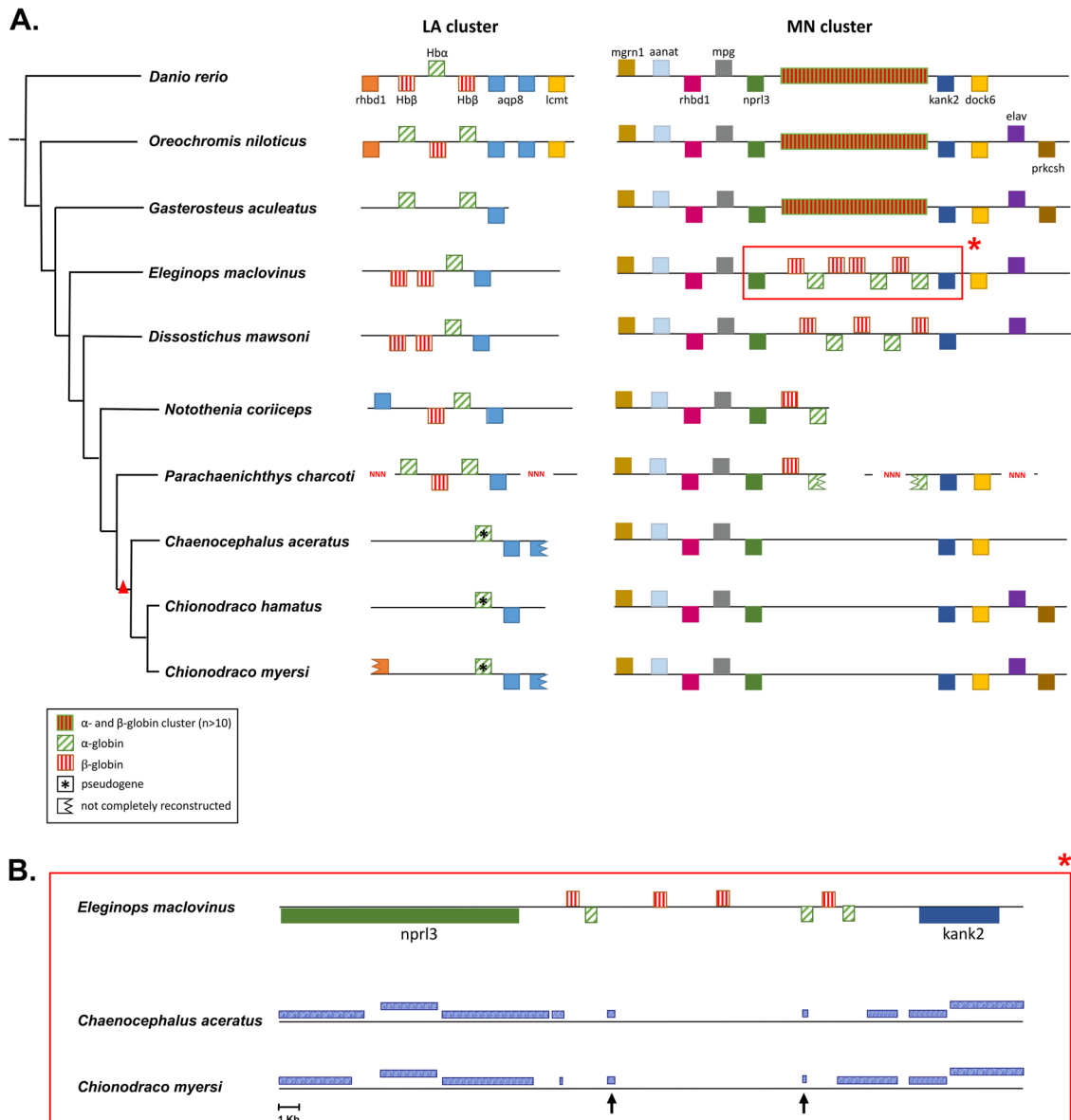
Analysis of sequence similarity between the MN globin genomic cluster in *E. maclovinus* and the homologous region in *C. myersi* and *C. aceratus* showed at least two conserved sequence fragments (indicated by arrows in Fig. 1b). Such evidence suggests that multiple events of deletion and/or rapid sequence divergence, in addition to the already reported deletion in the LA cluster, led to the loss of globin genes. Unlike the fate of the two globin clusters, full-length key erythropoietic transcription factors were identified (full list in Table S2 of Xu et al.<sup>25</sup>) in the *C. myersi* genome, but were found to be suppressed in the icefish. To further explore the role of selection on proteins involved in erythropoiesis, we used RELAX, a method specifically developed to identify relaxation of purifying selection on protein-coding genes<sup>39</sup>. Seven strictly one-to-one orthologs encoding important erythropoietic factors (GF11B, RHAG, TAL1, LMO2, ALAD, CPOX, PPOX) were analyzed across six teleost fish (*Danio rerio*, *Oreochromis niloticus*, *Gasterosteus aculeatus*, *E. maclovinus*, *D. mawsoni*, *C. myersi*) to test whether in the icefish evolutionary lineage relaxed purifying selection could be observed. Although five out of seven genes suggested relaxation of evolutionary constraints, none of these tests were statistically significant (Supplementary Table 6).

**Comparative transcriptomic analysis of icefish muscle expression profiles.** Transcriptome data of *Chionodraco myersi* skeletal muscle from five individual samples were generated and analyzed

with comparable data (adult muscle tissue, five biological replicates) that were available in public repositories for zebrafish (*D. rerio*), Nile tilapia (*O. niloticus*), and threespine stickleback (*G. aculeatus*). These three species all have high quality, well-annotated genomes and represent diverse teleost evolutionary lineages adapted to different thermal regimes. Since divergence at the transcriptome level between model fish species and icefish might be due either to the peculiar hemoglobinless condition or to the response to subzero temperatures, we included in the analysis muscle RNA-seq data from Antarctic red-blooded notothenioids, which share with icefish the long-term adaptation to the Southern Ocean conditions, but still retain functional hemoglobins. Unfortunately, a sufficient number of biological replicates was not available for a single notothenioid species in public repositories. Therefore, we decided to group together RNA-seq data from three red-blooded Antarctic species (*D. mawsoni*, *N. coriiceps*, *P. charcoti*, see methods). Although such an approach might inflate expression variance within this group and be overly conservative, we considered that it was crucial to compare at a transcriptome-wide level, the hemoglobinless icefish and the red-blood notothenioids. After quality filtering and normalization, RNA-seq expression data were compared for 9721 OGs by hierarchical clustering and principal component analyses (Fig. 2). Results showed a clear separation between *C. myersi* and all the other species, including red-blooded notothenioids, which were grouped together despite representing cross-species data. Differential expression was statistically assessed between *Chionodraco* and all three model species together and similar results were obtained when using only red-blooded notothenioid. Red- and white-blooded notothenioid expression data were analyzed separately (see methods).

Comparing *C. myersi* against the three model species, differential expression analysis identified 2758 and 2813 OGs over- and under-expressed, respectively ( $\log_2(\text{FC}) \geq 1$  or  $\log_2(\text{FC}) \leq -1$ , false-discovery rate (FDR)  $\leq 0.05$ ). The comparison between white- and red-blooded notothenioids yielded similar results, with 1630 up- and 2882 downregulated OGs. Gene-set enrichment analysis (GSEA) was then employed to specifically test the hypothesis that mitochondria play a key role in compensating for the lack of hemoglobin. In the comparison between icefish and non-Antarctic species, two gene sets, IMPI and MitoCarta, which include, respectively, 1626 and 1158 proteins with mitochondrial localization, were enriched with high significance (IMPI FDR  $q$ -val 0, normalized enrichment score (NES) 3.9; MitoCarta FDR  $q$ -val 0, NES 3.6). Likewise, two gene sets that contain genes predicted to be under the control of two key regulators of mitochondrial biogenesis, NRF1 and NRF2/GABPA, were significantly enriched (NRF1 FDR  $q$ -val 0.003, NES 2.3; NRF2/GABPA, FDR  $q$ -val 0.1, NES 1.7). Even stronger evidence for differential expression of mitochondria-related genes was observed in the comparison between white- and red-blooded notothenioids (IMPI FDR  $q$ -val 0, NES 3.4; MitoCarta FDR  $q$ -val 0, NES 3.2, NRF1 FDR  $q$ -val 0, NES 1.9; NRF2/GABPA, FDR  $q$ -val 0.003, NES 1.8). Two additional gene sets were significantly enriched, MITOCHONDRIAL\_INNER\_MEMBRANE (FDR  $q$ -val 0.006, NES 1.9) and BIOCARTA\_PGC1A\_PATHWAY (FDR  $q$ -val 0.03, NES 1.9). The inner mitochondrial membrane hosts the electron transport chain machinery, generates membrane potential necessary for ATP generation, and forms the signature folds of mitochondria, known as cristae<sup>40</sup>. Genes in the PCG-1a pathway are crucial for mitochondrial biogenesis<sup>41</sup>.

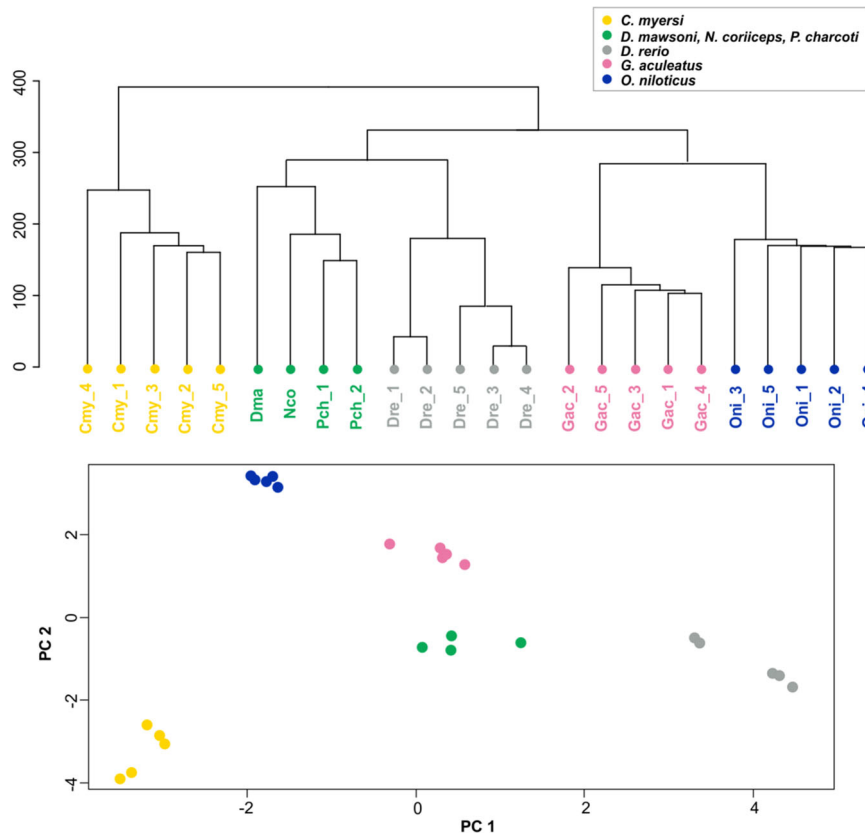
Detailed analysis of genes involved in various mitochondrial processes confirmed evidence of differential expression in the icefish as compared to model species. Among these, it is worth to note that the large majority (8–10 out of 13, Table 1) of genes encoding cofactors involved in mitochondria assembly were



**Fig. 1** LA and MN globin clusters genomic structure. **a** Genomic organization (not to scale) of the LA and MN globin gene clusters of teleost fish. The phylogenetic tree depicted in the figure is based on<sup>32,33,38</sup>. All clusters are presented in the same orientation as *D. rerio*. Genes in forward orientation are shown on top of the contig whereas genes in the reverse orientation are shown below. Red triangle depicts loss of functional hemoglobins. Hb $\beta$  = beta-globin, Hb $\alpha$  = alpha-globin, aqp8 = aquaporin 8, lcmt = leucine carboxyl methyltransferase, mgrn1 = mahogunin ring finger 1, aanat = arylalkylamine N-acetyltransferase, rhd1 = inactive rhomboid protein 1, mpg = DNA-3-methyladenine glycosylase, npr3 = nitrogen permease regulator 3-like, kank2 = KN motif and ankyrin repeat domain-containing protein 2, dock6 = dedicator of cytokinesis protein 6, prkcs = glucosidase II subunit beta. The asterisk refers to the genomic region magnified in Fig. 1b. **b** Sequence conservation of the genomic region flanking MN globin cluster either between *E. maclovinus* and *C. aceratus* or between *E. maclovinus* and *C. myersi*. Blue boxes represent Blastn best hits. The two genes, npr3 and kank2, appear to be fully functional in both icefish genomes. Arrows in Fig. 1b indicate regions with sequence similarity between *E. maclovinus* and either *C. aceratus* or *C. myersi* in the genomic region where globin genes are present in red-blooded notothenioids.

over-expressed. Assembly of the large and small ribosomal subunits is essential for mitochondrial biogenesis<sup>42</sup>. Also, proteins that play a fundamental role in mitochondrial fusion/fission and cristae remodeling<sup>16</sup> were differentially regulated at the messenger RNA (mRNA) level (Table 2). Transcripts encoding three proteins forming the mitochondrial contact site and cristae organizing system (MICOS), including IMMT (MIC60), a core MICOS component<sup>43</sup>, were significantly under-expressed (FDR < 0.05) in *C. myersi*. Mitochondrial creatine-kinase (mtCK), a relevant protein for bridging inner and outer mitochondrial membranes and promoting lipid transfer (albeit its role is still controversial)<sup>27,29–31</sup>, is largely under-expressed (Table 2).

Interestingly, mtCK knock-out in mice produces a muscle cellular phenotype similar to that of icefish, with an ultra-structural remodeling of muscle fibers, leading to higher mitochondrial density and enlarged mitochondria<sup>44</sup>. It was suggested that lack (or very low levels) of mtCK could be compensated by high levels of adenylate kinase (AK)<sup>45</sup>. This indeed seems to be the case in *C. myersi*, which exhibits significant (FDR < 0.05) upregulation of several AK genes such as AK1, AK2, AK4, and AK9 (Table 2). Also OPA1, a fundamental player in cristae remodeling, was downregulated<sup>46</sup>. Finally, the key regulator of mitochondrial fission, DNMI1/DRP1, was downregulated, whereas Mitoguarding (MIGA1), which favors mitochondrial fusion<sup>47</sup> and tethers



**Fig. 2** Hierarchical clustering and principal component analyses of muscle RNA-seq data. Clustering and spatial distribution of muscle RNA-seq data. Cmy: *C. myersi*; Dre: *D. rerio*; Gac: *G. aculeatus*; Oni: *O. niloticus*, Dma: *D. mawsoni*; Nco: *N. coriiceps*; Pch: *P. charcoti*. Biological replicates are distinguished by numbers from 1 to 5.

**Table 1 Muscle differential expression analysis of genes encoding cofactors involved in mitoribosome assembly.**

Gene symbols	OG ID	Cmy vs. Red		Cmy vs. NonAnt	
		LogFC	FDR	LogFC	FDR
<i>MRM1</i>	OG0012273	8.70	1.08E-12	2.53	0.00001
<i>MRM2</i>	OG0013246	0.46	0.68674	1.79	0.01748
<i>MRM3</i>	OG0007072	1.49	0.00828	3.56	1.77E-19
<i>MTG2</i>	OG0008966	1.51	0.00453	2.53	5.52E-11
<i>DDX28</i>	OG0008112	-0.16	0.83706	1.15	0.01680
<i>DHX30</i>	OG0012049	-0.78	0.24656	1.43	0.00214
<i>MTERF3</i>	OG0012260	1.09	0.02093	1.70	3.81E-07
<i>AFG3L2</i>	OG0001029	-0.50	0.29814	-0.50	0.16280
<i>SPG7</i>	OG0011454	0.93	0.10783	0.92	0.01888
<i>TFB1M</i>	OG0005933	2.58	0.00007	3.58	7.52E-18
<i>NSUN4</i>	OG0011353	-1.15	0.15217	-0.53	0.43104
<i>ERAL1</i>	OG0009078	0.40	0.46238	1.56	2.9E-06
<i>MPV17L2</i>	OG0006317	-0.87	0.30116	-0.68	0.32989

*Cmy* *Chionodraco myersi*, *Red* Red-Blooded Antarctic fish species, *NonAnt* non-Antarctic species

mitochondria to lipid droplets (L. Scorrano, personal communication), was upregulated (Table 2).

In consideration of the potential role of mitochondrial lipids in the adaptation to low temperature and loss of hemoglobin, we then examined genes involved in lipid synthesis. Several genes encoding key enzymes in glycerolipids biosynthetic pathways (GPAT1/2, GPAT3/4, LPAAT, AGPAT2) were significantly upregulated (FDR < 0.05) in *Chionodraco* compared to non-Antarctic species. Increased expression of GPAT3/4 and LPAAT might favor phosphatidic acid (PA) synthesis. PCYT1A, which is important for regulating phosphatidylcholine (PC) biosynthesis,

was also over-expressed. On the opposite, mRNA levels for enzymes (TAMM41) involved in cardiolipin (CL) synthesis were not significantly changed. Evidence from the comparison between red- and white-blooded notothenioids is less strong (Table 3). Lipid metabolism of mitochondria requires transfer of phospholipids from the endoplasmic reticulum through specialized contact sites<sup>19</sup>. A family of highly conserved proteins (EMC1-10) appears to be crucial to facilitate such transfer<sup>48</sup>. The majority of transcripts encoding EMC proteins were found to be over-expressed in the icefish muscle transcriptome (Table 3) compared to model species, while the comparison with red-blooded



**Table 2 Muscle differential expression analysis of genes encoding proteins that play a fundamental role in mitochondrial fusion/fission and cristae remodeling.**

Gene symbols	OG ID	Cmy vs. Red		Cmy vs. NonAnt	
		LogFC	FDR	LogFC	FDR
CHCHD3	OG0000439	-0.56	0.34327	-1.06	0.01315
CHCHD6	OG0000439	-0.56	0.34327	-1.06	0.01315
IMMT	OG0009745	-1.01	0.03134	-1.82	1.17E-06
DNM1L	OG0000826	-2.18	1.55E-06	-1.63	0.00001
OPA1	OG0010429	-0.88	0.11476	-1.47	0.00063
MIGA1	OG0012804	1.27	0.08039	3.08	3.6E-12
AK1	OG0001654	2.37	0.01166	1.75	0.00294
AK2	OG0009063	0.56	0.21309	2.31	2.42E-16
AK4	OG0009643	1.83	0.00065	2.21	7.65E-10
AK9	OG0004716	5.10	2.96E-06	3.33	0.0001
CKMT1A mtCK	OG0000821	-15.43	1.68E-24	-16.04	1.41E-27

Cmy *Chionodraco myersi*, Red Red-Blooded Antarctic fish species, NonAnt non-Antarctic species

**Table 3 Muscle differential expression analysis of genes encoding endoplasmic reticulum membrane protein complex (EMC).**

Gene symbols	OG ID	Cmy vs. Red		Cmy vs. NonAnt	
		LogFC	FDR	LogFC	FDR
EMC1	OG0006378	-0.25	0.65393	1.87	1.63E-08
EMC2	OG0010325	0.70	0.07284	1.56	4.97E-10
EMC3	OG0002273	-1.49	0.00033	-0.64	0.05495
EMC7	OG0008614	-0.40	0.35842	0.36	0.23869
EMC8	OG0008125	1.98	0.00022	2.87	1.38E-15
EMC9	OG0008569	-0.19	0.75571	0.52	0.21265
EMC10	OG0010959	0.08	0.89155	1.09	0.00044
GPAT3/4	OG0000399	0.65	0.24223	1.46	0.00003
GPAT1/2	OG0007344	0.72	0.65832	0.64	0.53293
AGPAT2	OG0012024	1.22	0.03150	1.31	0.00104
MBOAT2/LPAAT	OG0004179	0.65	0.21398	1.09	0.00388
PCYT1A	OG0007558	0.56	0.23464	1.21	0.00007
TAMM41	OG0010111	-0.35	0.63748	0.53	0.29101

Cmy *Chionodraco myersi*, Red Red-Blooded Antarctic species, NonAnt non-Antarctic species

notothenioids showed no evidence of general upregulation of ECM-encoding transcripts (Table 3).

Several nuclear-encoded genes are potentially involved in the transcriptional control of mitochondrial biogenesis. Tfam (mitochondrial transcription factor A) plays an important role in replication of mtDNA and transcription of mitochondrial genes. In mammals two other regulative proteins, TFB1M, TFB2M (mitochondrial transcription factors B1 and B2) are relevant in mitochondrial genes transcriptional activation<sup>49</sup>. Of these, Tfam and TFB1M appeared to be over-expressed in the muscle of *C. myersi* as compared to the three model species, whilst genes known to activate the mitochondrial machinery like NRF1, NRF2/GABPA, and PGC1 $\alpha$ <sup>50</sup>, do not appear to be consistently upregulated in the icefish muscle transcriptome (Table 4). It was suggested, however, that the PGC1 $\alpha$  - NRF1 pathway may be disrupted in fish due to insertions in the gene sequence of the teleost PGC1 $\alpha$  ortholog within the NRF1-binding domain<sup>51</sup>. This is also the case in the *C. myersi* PGC1 $\alpha$  that shows a serine-rich insertion in the putative NRF1-binding region. Icefish PGC1 $\alpha$  is also mutated at the site responsive to AMPK (AMP-activated protein kinase) which, in mammals, requires a critical threonine

at position 177. AMPK is a major energy sensor implicated in energy balance that in mammals activates metabolic pathways via the PGC1 $\alpha$ -NRF1 cascade to respond to energetic challenges when the ATP:AMP ratio is altered<sup>52</sup>. In the icefish, as in other teleosts, threonine is replaced by a proline residue precluding PGC1 $\alpha$  phosphorylation by AMPK. Although in fish the AMPK-PGC1 $\alpha$ -NRF1 axis seems disrupted, the role of AMPK in responding to muscle metabolic stress was shown to remain central in trout<sup>53</sup>. This could explain the observed remarkable overexpression of AMPK (PRKAA1) in *C. myersi* (Table 4). Similar to AMPK, other master genes of muscle mitochondrial biogenesis, Myocyte enhancer factor 2A (MEF2A), are largely upregulated in the icefish. Also, MEF2A interacts with PGC1 $\alpha$  in mammals. MEF2A knock-out mice show disorganized mitochondria with major structural alteration pointing out the central role of MEF2A in mitochondrial biogenesis<sup>54</sup>. Similar evidence was observed for other members of the MEF family (MEF2B, MEF2C, MEF2D; Table 4).

**Comparative analysis of gene duplication.** Massive gene duplication is reported to have affected the evolution of the notothenioid genome<sup>27,29-31</sup>. We compared the genomes of two Antarctic red-blooded (*N. coriiceps* and *D. mawsoni*) and two white-blooded notothenioids (*C. myersi* and *C. aceratus*). The number of annotated coding genes in the genome of *C. myersi* is 38,140, which is considerably higher than what reported for the closely related icefish species *C. aceratus* (30,773). Likewise, the number of predicted coding genes in *N. coriiceps* (32,331) is significantly higher than what described for *D. mawsoni* (22,516), another red-blooded notothenioid species with a comparable genome size. Fragmentation of genome assemblies is likely the main reason for such discrepancies as a highly fragmented assembly might inflate the estimated number of copies within OGs, because partial gene fragments are counted as two or more copies. Therefore, to control for such phenomenon, we compared all protein-coding genes from all four notothenioid species against the phylogenetically closest species with a high-quality genome, the threespine stickleback, *G. aculeatus*. Only notothenioid transcripts covering at least 60% of the orthologous sequence in stickleback were retained for duplication analysis (see methods). After such correction, the total number of coding genes was much more similar between species with comparable genome size (*C. myersi*, 26,805; *C. aceratus*, 24,900; *N. coriiceps*, 21,612; *D. mawsoni*, 19,585). The remaining difference in gene copy number between *C. myersi* and *C. aceratus* was apparently attributable to few gene families (<15), encoding immune-related proteins that

**Table 4 Muscle differential expression analysis of key genes involved in mitochondrial biogenesis.**

Gene symbols	OG ID	Cmy vs. Red		Cmy vs. NonAnt	
		logFC	FDR	logFC	FDR
<i>TFB1M</i>	OG0005933	2.58	0.00007	3.58	7.52E-18
<i>TFB2M</i>	OG0007876	-0.23	0.72366	0.11	0.84489
<i>TFAM</i>	OG0007700	0.01	0.99426	0.77	0.02033
<i>NRF1</i>	OG0007192	-1.52	0.00821	-1.36	0.00395
<i>GABPA</i>	OG0007460	-0.86	0.04306	-0.33	0.36865
<i>PRKAA1/AMPK</i>	OG0001835	0.50	0.26308	1.79	6.93E-11
<i>MEF2A</i>	OG0001443	3.26	7.36E-08	2.43	2.65E-11
<i>MEF2B</i>	OG0005631	1.94	0.06032	2.37	0.00056
<i>MEF2C</i>	OG0002939	1.94	0.00104	0.29	0.53004
<i>MEF2D</i>	OG0003723	1.88	0.00150	1.31	0.00076
<i>PPARGC1A</i>	OG0004227	1.79	0.05546	-0.17	0.85610

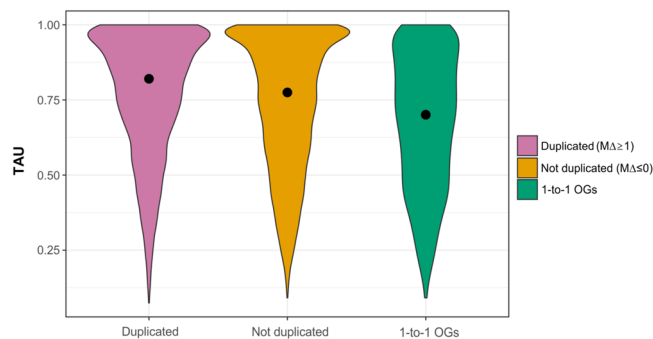
*Cmy* *Chionodraco myersi*, *Red* Red-Blooded Antarctic fish species, *NonAnt* non-Antarctic species

are reported to have undergone large and rapid lineage-specific expansions in other teleost genomes and are notoriously difficult to correctly annotate. Considering the limited number of gene families accounting for such residual discrepancy, we proceeded by comparing the mean gene copy number for each Orthology Group (OG) between red- and white-blooded notothenioid species. A list of putative duplicated OGs and the corresponding annotations in *C. myersi* is reported in Supplementary Data 1. We then tested whether mitochondrial proteins might be specifically involved in gene duplication events. We observed that genes encoding proteins included in Mitocarta and IMPI were significantly enriched in duplicated OGs between red- and white-blooded species (Fisher-exact test Mitocarta  $p = 0.02$ , IMPI  $p = 0.02$ ).

Gene duplications in notothenioids have been proposed to increase expression through a gene dosage effect<sup>27,28</sup>, although experimental evidence is limited. Here, we compared OGs with  $MD \geq 1$ , i.e., at least one additional gene copy in white-blooded notothenioids, against OGs with  $MD \leq 0$ , i.e., genes not duplicated in the icefish genome (*C. myersi* and *C. aceratus*). We observed that duplicated genes were also upregulated ( $\logFC \geq 1$ ) more frequently than not duplicated genes (Fisher-exact test  $p = 0.01$ ).

Gene duplication also provides the opportunity for functional specialization of duplicated copies. Tissue-specific expression is considered an indication of such specialization. We thus tested whether genes that underwent duplication specifically in the icefish lineage showed higher tissue specificity compared to not duplicated genes. After calculating the tissue-specificity index  $\tau$  (TAU<sup>55</sup>) for all genes across five tissues, genes that belong to those OGs (3601 genes) with a higher number of co-orthologs in *C. myersi* ( $MD \geq 1$ , i.e., at least one additional copy in the icefish lineage compared to red-blooded notothenioids) showed significantly greater tissue specificity (median  $\tau = 0.82$ ) compared to either genes included in OGs (14,948 genes) in which the number of co-orthologs was equal or lower than other species (median  $\tau = 0.77$ , Wilcoxon-rank test;  $p < 0.00001$ ) or strictly one-to-one orthologs (1019 genes) (median  $\tau = 0.70$ , Wilcoxon-rank test;  $p < 0.00001$ ). The distributions of  $\tau$ -values within the three gene categories are shown in Fig. 3.

**Comparative analysis of promoter regions.** Results of GSEA showed significant enrichment in genes that are putatively regulated by NRF1 and NRF2/GABPA in model vertebrate species. These two transcription factors control the expression of a large set of proteins involved in mitochondrial functioning. Using in silico prediction of putative-binding motifs, we



**Fig. 3** Violin plots describing the distribution of  $\tau$  (TAU) values within three gene categories.

identified the number of potential binding sites for either NRF1 or NRF2/GABPA in the upstream region of unambiguous 1-to-1 orthologs in zebrafish, tilapia, stickleback, one red-blooded notothenioid (*D. mawsoni*), and one icefish (*C. myersi*). We then compared the number of NRF1 and NRF2/GABPA putative-binding sites in orthologous genes between either *C. myersi* and all three non-Antarctic species or *C. myersi* and *D. mawsoni*. Finally, we tested whether a higher number of putative-binding sites for either NRF1 or NRF2/GABPA in *C. myersi* was associated with significantly higher gene expression ( $\logFC > 0$ ,  $p \leq 0.05$ ) in the icefish muscle. While no significant association was observed for NRF1, for NRF2/GABPA, we found that *C. myersi* genes with at least one additional NRF2/GABPA-binding site in their promoter compared to model fish species also were more frequently upregulated (Fisher-exact test,  $p = 0.003$ ). The same test between *C. myersi* and *D. mawsoni* was not significant. However, when considering only genes with at least two additional NRF2/GABPA-binding sites, we observed a significant association with higher expression in the comparison between *C. myersi* and non-Antarctic species (Fisher-exact test  $p = 0.01$ ), as well as between *C. myersi* and *D. mawsoni* (Fisher-exact test  $p = 0.03$ ). Among the genes showing at least two additional predicted NRF2/GABPA motifs in the white-blooded *C. myersi* compared either to non-Antarctic species or to the red-blooded Antarctic species, we found those encoding for WFS1, ZC3H10, and MTG2. The first one of these is a protein involved in mitochondrial-endoplasmic reticulum contacts<sup>56</sup>, ZC3H10 is a new mitochondrial regulator<sup>57</sup>, and MTG2 is involved in mitoribosome assembly, which is crucial for mitochondrial biogenesis<sup>42</sup>.

## Discussion

In this work, we built upon a recent comparison of two genomic regions in the icefish *C. aceratus* that corresponded to the MN and LA globin clusters in other teleost species<sup>31</sup>. Here, we expanded the reconstruction of these two globin gene clusters, including four red-blooded notothenioids and three white-blooded species, defining a new perspective on globin gene loss in the icefish. It has been suggested that the notothenioids only harbor one globin locus—compared to other teleosts containing two—and that this locus was almost completely lost in a single event in the common ancestor of channichthyids, possibly driving the evolution of the hemoglobinless condition<sup>22</sup>. This scenario was partially called into question by Near et al.<sup>23</sup>, who however limited their analysis to a single-genomic region. The presence of at least two genomic clusters containing up to ten globin genes in red-blooded notothenioids, reported here for the first time to our knowledge, suggests that the hemoglobin loss was achieved through a series of deletion events and/or rapid sequence divergence due to relaxed purifying selection occurring at separate genomic regions, likely in parallel with progressive reduction of functional importance for oxygen carriers rather than as a single driving mutational event. The irreversible loss of all globin genes is not mirrored by the corresponding deletion of genes involved in the differentiation of the cells that normally carry hemoglobin, the erythrocytes. These genes are downregulated<sup>25</sup>, but conserved at the genomic level in white-blooded notothenioids. Likewise, we found no clear evidence for relaxed selection at seven genes encoding proteins with a relevant role in erythropoiesis. A possible explanation is that these genes have a role in other biological processes beside erythropoiesis. It should be noted, however, that orthologous sequences were available only for a very limited number of notothenioid species, which might have reduced the power of the test. As soon as genomes from other red- and white-blooded notothenioids become available, a more powerful analysis of the selective pressures acting on specific icefish genes should be possible.

Loss of hemoglobin is accompanied by unique modifications to favor within-cell oxygen diffusion, through increased mitochondrial density, higher phospholipid content, and altered mitochondrial dynamics<sup>6</sup>. While such adaptations have been investigated at the biochemical and cellular level, the underlying genomic basis has remained elusive so far. Here, we show that a vast transcriptional program significantly involves several pathways controlling key mitochondrial features, with potential effects toward higher mitochondrial biogenesis, increased mitochondrial fusion, modified (reduced) cristae, enhanced transfer between endoplasmic reticulum and mitochondria, and a peculiar lipid profile, with higher phosphatidylcholine and phosphatidic acid (PA), but not cardiolipin (CL). Since PA is considered fusogenic<sup>58</sup> while CL seems to favor mitochondrial fission, the observed expression profile of enzymes involved in lipid biosynthesis is also concordant with marked downregulation of DNMI1/DRP1, which has a key role in mitochondrial fission. The divergent transcriptional profile in the muscle of *Chionodraco* emerging from the comparison with non-Antarctic fish species nicely recapitulates all the cellular adaptations that have been proposed to allow for a life without hemoglobin. The analogous comparison with muscle transcriptome data from red-blooded Antarctic notothenioids provided fully consistent results, with even stronger evidence for a key role of mitochondria in the adaptation to the hemoglobinless condition. In fact, the results obtained from the comparison between *C. myersi* and non-Antarctic species might be explained, at least in part, as the response to the freezing conditions of the Southern Ocean rather than the absence of oxygen carriers. On the opposite, Antarctic red- and white-blooded notothenioids share the same environmental conditions,

therefore, the obtained evidence is the likely consequence of the unique icefish biology. A possible exception is lipid metabolism where divergence at the transcriptome level was less pronounced, which might imply that modifications of lipid metabolism are mostly due to the adaptation to low temperatures. In fact, Chen et al.<sup>30</sup> reported altered lipid metabolism in the Antarctic red-blooded *D. mawsoni* compared to the non-Antarctic notothenioid *E. maclovinus*, although this observation appeared to be mostly linked to increased neutral buoyancy in *D. mawsoni*.

It remains to be assessed whether such articulated transcriptome modifications are the consequence of either genomic divergence or phenotypic plasticity, a long-standing question in evolutionary biology. Looking for evidence of genomic divergence, we tested whether gene duplication might be involved in differential expression of genes encoding mitochondrial proteins. Genes that are duplicated in the *C. myersi* genome compared to other Antarctic red-blooded species showed evidence for enrichment for genetic loci encoding mitochondrial proteins (IMPI or Mitocarta). Gene duplication was significantly associated with higher expression at the mRNA level, as well as to functional specialization in terms of higher tissue specificity. Massive gene duplication has already been shown to characterize the notothenioid genome<sup>27,29–31</sup>. Here, we present compelling evidence that such an evolutionary pattern is particularly important in relation to the peculiar mitochondrial biology of the hemoglobinless icefish and provide a significant link between gene duplication and functional divergence at the whole-transcriptome level.

The transcription factor NRF2/GABPA is known to regulate genes involved in several processes, including mitochondrial biogenesis<sup>59,60</sup>. We found that in the *Chionodraco* genome the number of NRF2/GABPA-binding sites was significantly (Fisher-exact test,  $p = 0.003$ ) higher in the upstream genomic region of genes that are over-expressed in muscle compared to the orthologous region in non-Antarctic, as well as Antarctic red-blooded fish genomes. It has been shown that multiplicity of transcription factor-binding sites in the promoter region generally increases gene expression, with a clear positive relationship between the number of sites and average expression<sup>61,62</sup>. This represents additional strong evidence that the icefish genome has evolved structural modifications that are associated with major changes in the muscle transcriptome profile. Therefore, genomic divergence seems at least partially responsible for the unique cellular adaptations that allow the hemoglobinless icefish to survive. While working on Antarctic fish biology remains difficult for obvious reasons, the popularization of genomic tools, such as comparative transcriptome analyses and functional genomic assays, holds the promise to help uncover the molecular mechanisms underlying the evolution of this unique group of vertebrates.

## Methods

**Ethics.** No specific permits were required for the work described here. Animals included in the present study were not subjected to any experimental manipulation. The study was performed in accordance with the EU directive 2010/63/EU and Italian DL 2014/26. Experiments and killing procedures were monitored and carried out by authorized staff to minimize animals' suffering.

**Ortholog analysis.** Homology relationships between *C. myersi*, and model teleost genomes were reconstructed with the software OrthoFinder<sup>63</sup> using a total of 19 annotated genomes. OrthoFinder uses the BLAST tool<sup>64</sup> to compute sequence similarity scores between sequences in multiple species and then uses the MCL clustering algorithm<sup>65</sup> to identify groups of highly similar sequences within this dataset. Protein datasets from Ensembl or NCBI databases were downloaded for *Astyanax mexicanus* (Ensembl GCA\_000372685.2), *Danio rerio* (Ensembl GRCz11 GCA\_000002035.4), *Gadus morhua* (Ensembl gadMor1), *Gasterosteus aculeatus* (Ensembl BROAD S1), *Oreochromis niloticus* (Ensembl GCA\_000188235.1), *Oryzias latipes* (Ensembl GCA\_002234715.1), *Poecilia formosa* (Ensembl Poecilia\_formosa-5.1.2), *Takifugu rubripes* (Ensembl GCA\_000180615.2), *Tetraodon nigroviridis*



(Ensembl TETRAODON 8.0), *Xiphophorus maculatus* (Ensembl GCA\_002775205.2), *Lepisosteus oculatus* (Ensembl GCA\_000242695.1), and *Larimichthys crocea* (NCBI Release 102). The proteomes of *Sparus aurata*<sup>66</sup> along with the closely related European seabass (*Dicentrarchus labrax*, downloaded from <http://seabass.mpipz.mpg.de/cgi-bin/hgGateway>), five Antarctic fish *Parachaenichthys charcoti* (Bathydraconidae)<sup>35</sup>, *Notothenia coriiceps* (Nototheniidae)<sup>37</sup>, *Dissostichus mawsoni* (Nototheniidae)<sup>30</sup>, *Eleginops maclovinus* (Eleginopsidae)<sup>30</sup>, and *Chaenocephalus aceratus* (Channichthyidae)<sup>31</sup> were added in the analysis. The orthology groups (OGs) were calculated for all species using OrthoFinder default parameters.

In order to reduce the effect of the fragmentation of Antarctic genome assemblies on the estimated number of duplicated genes, all protein-coding genes of *C. myersi*, *D. mawsoni*, *N. coriiceps*, *C. aceratus*, and *E. maclovinus* were compared to the proteome of *G. aculeatus* (Ensembl BROAD S1) by BLASTP analysis. Only proteins covering at least 60% of the orthologous sequence in stickleback were retained in the final datasets (Supplementary Table 7).

After a first round with OrthoFinder and all the 20 species, the *P. charcoti* dataset was excluded due to the low-quality of sequence data resulting in unreliable protein alignments.

Protein-coding gene gain and loss between icefish and red-blooded Antarctic species were then assessed by employing Mean  $\Delta$  ( $M\Delta$ ) metrics: mean number of *C. myersi* and *C. aceratus* genes—mean number of *N. coriiceps* and *D. mawsoni* genes.

**Comparative analysis of muscle transcriptomes.** In order to compare muscle gene expression across different species, RNA sequencing (RNA-seq) datasets were retrieved from the NCBI Sequence Read Archive. RNA-seq libraries of muscle from *D. rerio*, *O. niloticus*, and *G. aculeatus* were downloaded (Supplementary Table 8). For each species, five biological replicates were retrieved. RNA-seq data from three Antarctic species (*D. mawsoni*, *N. coriiceps*, *P. charcoti*) were also used for RNA-seq analysis, for a total of four replicates representing the red-blooded Antarctic sample group (Supplementary Table 8).

The reads of each species were mapped against the corresponding assembled genome by means of STAR aligner<sup>67</sup> and following the *two-pass* mapping mode. The maximum number of mismatches allowed was fixed to 6% and a threshold for multiple matches was set to 200 (–outFilterMultimapNmax 200). Read counts for each sample, at the gene level, were extracted by GeneCounts quantification while running STAR.

Analysis of differential gene expression was conducted in EdgeR<sup>68</sup>. For each sample, the sum of raw cpm (count per million) of all genes belonging to the same OG was calculated by means of *aggregate* function in R v. 3.5.3. OGs showing a cpm value <1 in more than half of the samples for each species were filtered out. Remaining OGs ( $n = 9721$ ) were normalized with the Trimmed Mean of  $M$ -values (TMM) method and, after estimating common and tagwise dispersions, likelihood-ratio test (lrt) as implemented in EdgeR was employed to assess differentially expressed genes (DEGs) between *C. myersi* and the other three model species, as well as between *C. myersi* and the red-blooded Antarctic samples.

OGs with  $\log_2$  fold-change (logFC) < –1 or >1 and false-discovery rate (FDR)  $\leq 0.05$  were considered significant.

**Gene-set enrichment analysis.** GSEA<sup>69</sup> was used to identify enriched functional categories within OGs differentially expressed in *C. myersi* compared to other teleost species.

The GSEAPreranked v1<sup>69</sup> tool was applied to RNA-seq data used in the comparative transcriptomic analysis (9721 OGs). Ranking values were assigned to OGs based on EdgeR lrt  $p$ -value as follows: (i)  $\logFC > 0$ , score =  $1 - p$ -val; (ii)  $\logFC < 0$ , score =  $-(1 - p$ -val). The enrichment analysis was carried out by setting Enrichment statistic = Classic, Normalization mode = meandiv, and Number of permutations = 1000. A panel of 25 Gene sets was interrogated comprising BIOARTA, KEGG, and GO sets related to mitochondrial biogenesis and functioning as downloaded from the Molecular Signatures Database v6.1 (MSigDB, <http://software.broadinstitute.org/gsea/msigdb>), as well as two gene sets, IMPI (Integrated Mitochondrial Protein Index) and MitoCarta 2.0<sup>70</sup> genes, which were downloaded from MitoMiner 4.0 database<sup>71</sup>.

**Assessment of duplication enrichment on gene sets.** All OGs were used in order to investigate whether the above mentioned gene sets were significantly enriched in OGs that duplicated in the icefish compared to red-blooded notothenioids. To this end, the  $M\Delta$  metric was used. All OGs with  $M\Delta \geq 1$  were considered duplicated. All genes that composed a single OG were retrieved and, whenever possible, multiple Gene Symbols were retained. In case of multiple Gene Symbols for the same OG, they were assigned the same  $M\Delta$  value. A Fisher's exact test was then used to assess whether gene-set showed a significant difference in the frequency of duplicated OGs between icefish and red-blooded notothenioids.

**Expression breadth.** Tissue specificity (i.e., expression breadth) of OGs was estimated using the  $\tau$  index<sup>55</sup>. The index  $\tau$  ranges from 0 to 1 and it is defined as  $\tau = \sum N_i = 1(1 - x_i)/N - 1$ , where  $N$  is the number of tissues and  $x_i$  is the expression profile component normalized by the maximal component value

( $\tau = 1$  single tissue expression,  $\tau = 0$  ubiquitous expression). Tissue specificity was calculated on the normalized expression (TMM normalized  $\log_2$  cpm) evaluated in brain, liver, skeletal muscle, kidney, and spleen from RNA-seq data produced in the present study (see Supplementary Table 8). Median values of  $\tau$  were calculated on OGs found duplicated ( $M\Delta \geq 1$ ) in the icefish genome compared to not duplicated OGs ( $M\Delta \leq 0$ ) and 1-to-1 OGs. In order to assess the significance of inter-group  $\tau$  differences, a pairwise Wilcoxon rank-sum test was implemented in R.

**Transcription factor-binding sites distribution.** Position frequency matrices (PFM) of transcription factor (TF)-binding profiles for NRF2/GABPA were downloaded from the JASPAR database (<http://jaspar.genereg.net/>). The presence and number of TF-binding sites (TFBS) were investigated in 1-to-1 orthologues ( $n = 3,895$ ) between *C. myersi*, *D. mawsoni*, and the three model species *D. rerio*, *O. niloticus*, and *G. aculeatus*. A region of length of 5 kb upstream the putative transcription starting site of these orthologous genes was then used as input for FIMO<sup>72</sup>, a MEME<sup>73</sup> suite's tool (Motif-based sequence analysis tools). For each gene,  $M\Delta$  was calculated between the number of TFBS found in *C. myersi* and the median TFBS number across non-Antarctic species. The same comparison was carried out between *C. myersi* and *D. mawsoni* (number of *C. myersi* TFBS—number of *D. mawsoni* TFBS). A Fisher's exact test was performed in order to test for significant differences in the  $M\Delta$  distribution between over-expressed 1-to-1 orthologous genes and not differentially expressed in *C. myersi*.

**Annotation of hemoglobin clusters in Antarctic fish genomes.** The LA and MN hemoglobin clusters were identified in the genome of *C. myersi*, *C. hamatus*, *N. coriiceps*, *P. charcoti*, *C. aceratus*, *D. mawsoni*, and *E. maclovinus* based on the genomic organization described by Opazo et al.<sup>24</sup>. There is a high-degree conservation for five genes to the left of the MN cluster: *aanat*, *mgn1*, *rhbdf1a*, *mvg*, and *nprl3*. The two genes to the right, *kank2* and *dock6*, are also well conserved. The LA cluster is less conserved with the gene *rhbdf1* to the left, *aqp8* and *lcmt1* on the right side<sup>24</sup>. For both clusters, *N. coriiceps* protein sequences of conserved flanking genes, together with  $\alpha$ - and  $\beta$ -globin genes, were downloaded from the NCBI database and employed as query in BLASTp searches ( $10^{-5}$   $e$ -value) against the draft genomes of the six species. Contigs containing target genes were then further employed for manual annotation, using BLASTp output in the software UGENE<sup>74</sup>, to draw the putative LA and MN clusters.

One-to-one orthologs for seven genes coding for key erythropoiesis factors from six teleost species (*D. rerio*, *G. aculeatus*, *O. niloticus*, *E. maclovinus*, *D. mawsoni*, *C. myersi*) were aligned using a codon-guided alignment software (TranslatorX)<sup>75</sup> available at <http://translatorx.co.uk/>. The algorithm MAFFT with default settings was selected for alignment in TranslatorX and codon-based alignment were refined using GBLOCKS. Aligned coding sequences were then processed with the software RELAX<sup>39</sup>, available at <https://www.datamonkey.org/relax>. In RELAX, the tree branch leading the icefish species (*C. myersi*) was compared to the remaining branches to test whether relaxed selection was detected as a consequence of hemoglobin loss.

**Statistics and reproducibility.** Protein-coding gene gain and loss analysis was assessed by employing Mean  $\Delta$  ( $M\Delta$ ) metrics: mean number of icefish genes – mean number of red-blooded Antarctic species genes.

Gene expression analysis was conducted using five biological replicates for non-Antarctic species and icefish while four replicates were employed for red-blooded Antarctic species. Analysis of differential gene expression was conducted in EdgeR. For each sample, the sum of raw cpm (count per million) of all genes belonging to the same OG was calculated by means of *aggregate* function on R. OGs showing a cpm value <1 in more than half of the samples for each species were filtered out. A likelihood-ratio test (lrt) as implemented in EdgeR was employed to assess differentially expressed genes considering significant a  $\log_2$  fold-change (logFC) < –1 or >1 and false-discovery rate (FDR)  $\leq 0.05$ .

In order to assess duplication enrichment on specific gene sets a Fisher's exact test (FET) was conducted considering as duplicated all OGs having  $M\Delta$  values >0. Gene sets with FET  $p$ -value  $\leq 0.05$  were considered significantly enriched.

FET was also employed in order to assess if duplicated OGs were more frequently over-expressed than OGs not duplicated. In this case, OGs with  $M\Delta \geq 1$  were compared to OGs with  $M\Delta \leq 0$  and overexpression cutoff was set to  $\logFC \geq 1$ .

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

*Chionodraco myersi* and *Chionodraco hamatus* genome assemblies were deposited in DDBJ/ENA/GenBank under the accession RQJG0000000 (version RQJG0100000) and RRCA00000000 (version RRCA01000000), respectively. All raw sequence data produced in this study were deposited in NCBI Short Reads Archive (SRA) under accession numbers from SRR8197047 to SRR8197058. Details are reported in Supplementary Table 8.

Received: 4 February 2019; Accepted: 8 November 2019;

Published online: 29 November 2019

## References

- Ruud, J. T. Vertebrates without erythrocytes and blood pigment. *Nature* **173**, 848–850 (1954).
- Eastman, J. T. *Antarctic Fish Biology: Evolution in A Unique Environment*. (Academic Press, San Diego, 1993).
- Bargelloni, L., Marcato, S., Zane, L. & Patarnello, T. Mitochondrial phylogeny of Notothenioids: a molecular approach to antarctic fish evolution and biogeography. *Syst. Biol.* **49**, 114–129 (2000).
- Near, T. J. et al. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proc. Natl Acad. Sci.* **109**, 3434–3439 (2012).
- Henry, W. Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures. *Philos. Trans. R. Soc.* **93**, 274–276 (1803).
- Sidell, B. D. & O'Brien, K. M. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes. *J. Exp. Biol.* **209**, 1791–1802 (2006).
- Archer, S. D. & Johnston, I. A. Density of cristae and distribution of mitochondria in the slow muscle fibers of antarctic fish. *Physiological. Zool.* **64**, 242–258 (1991).
- Johnston, I. A. & Harrison, P. Morphometrics and ultrastructure of myocardial tissue in Notothenioid fishes. *Fish. Physiol. Biochem.* **3**, 1–6 (1987).
- Egginton, S. & Sidell, B. D. Thermal acclimation induces adaptive changes in subcellular structure of fish skeletal muscle. *Am. J. Physiol.* **256**, R1–9 (1989).
- Johnston, I. A. & Maitland, B. Temperature acclimation in crucian carp, *Carassius carassius* L., morphometric analyses of muscle fibre ultrastructure. *J. Fish. Biol.* **17**, 113–125 (1980).
- O'Brien, K. M. Mitochondrial biogenesis in cold-bodied fishes. *J. Exp. Biol.* **214**, 275–285 (2011).
- Johnston, I. A., Calvo, J., Guderley, H., Fernandez, D. & Palmer, L. Latitudinal variation in the abundance and oxidative capacities of muscle mitochondria in perciform fishes. *J. Exp. Biol.* **201**, 1119–1129 (1998).
- Sidell, B. D. Intracellular oxygen diffusion: the roles of myoglobin and lipid at cold body temperature. *J. Exp. Biol.* **201**, 1119–1128 (1998).
- O'Brien, K. M. & Mueller, I. A. The unique mitochondrial form and function of Antarctic channichthyid icefishes. *Integr. Comp. Biol.* **50**, 993–1008 (2010).
- Keenan, K. A., Grove, T. J., Oldham, C. A. & O'Brien, K. M. Characterization of mitochondrial glycerol-3-phosphate acyltransferase in notothenioid fishes. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **204**, 9–26 (2017).
- Pernas, L. & Scorrano, L. Mito-morphosis: mitochondrial fusion, fission, and cristae remodeling as key mediators of cellular function. *Annu. Rev. Physiol.* **78**, 505–531 (2016).
- O'Brien, K. M. & Sidell, B. D. The interplay among cardiac ultrastructure, metabolism and the expression of oxygen-binding proteins in Antarctic fishes. *J. Exp. Biol.* **203**, 1287–1297 (2000).
- Van der Bliek, A. M., Sedensky, M. M. & Morgan, P. G. Cell biology of the mitochondrion. *Genetics* **207**, 843–871 (2017).
- Csordás, G., Weaver, D. & Hajnóczky, G. Endoplasmic reticulum-mitochondrial contactology: structure and signaling functions. *Trends Cell Biol.* **28**, 523–540 (2018).
- Montgomery, J. & Clements, K. Disaptation and recovery in the evolution of Antarctic fishes. *Trends Ecol. Evol. (Amst.)* **15**, 267–271 (2000).
- Cocca, E., Detrich, H. W., Parker, S. K. & Prisco, G. A cluster of four globin genes from the Antarctic fish *Notothenia coriiceps*. *J. Fish. Biol.* **57**, 33–50 (2005).
- Cocca, E. et al. Genomic remnants of alpha-globin genes in the hemoglobinless antarctic icefishes. *Proc. Natl. Acad. Sci. USA* **92**, 1817–1821 (1995).
- Near, T. J., Parker, S. K. & Detrich, H. W. A genomic fossil reveals key steps in hemoglobin loss by the antarctic icefishes. *Mol. Biol. Evol.* **23**, 2008–2016 (2006).
- Opazo, J. C., Butts, G. T., Nery, M. F., Storz, J. F. & Hoffmann, F. G. Whole-genome duplication and the functional diversification of teleost fish hemoglobins. *Mol. Biol. Evol.* **30**, 140–153 (2013).
- Xu, Q. et al. Evolutionary suppression of erythropoiesis via the modulation of TGF- $\beta$  signalling in an Antarctic icefish. *Mol. Ecol.* **24**, 4664–4678 (2015).
- Desvignes, T., Detrich, H. W. & Postlethwait, J. H. Genomic conservation of erythropoietic microRNAs (erythromiRs) in white-blooded Antarctic icefish. *Mar. Genomics* **30**, 27–34 (2016).
- Chen, Z. et al. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc. Natl Acad. Sci. U.S.A.* **105**, 12944–12949 (2008).
- Cao, L. et al. Neofunctionalization of zona pellucida proteins enhances freeze-prevention in the eggs of Antarctic notothenioids. *Nat. Commun.* **7**, 12987 (2016).
- Coppe, A. et al. Genome evolution in the cold: Antarctic icefish muscle transcriptome reveals selective duplications increasing mitochondrial function. *Genome Biol. Evol.* **5**, 45–60 (2013).
- Chen, L. et al. The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonian robalo genomes. *Gigascience*. **8**, giz016 (2019).
- Kim, B. -M. et al. Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nat. Ecol. Evolution* **3**, 469 (2019).
- Near, T. J. et al. Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. *Mol. Phylogenet. Evol.* **129**, 268–279 (2018).
- Near, T. J. et al. Identification of the notothenioid sister lineage illuminates the biogeographic history of an Antarctic adaptive radiation. *BMC Evol. Biol.* **15**, 109 (2015).
- Detrich, H. W. et al. Genome enablement of the Notothenioidei: genome size estimates from 11 species and BAC libraries from 2 representative taxa. *J. Exp. Zool. B Mol. Dev. Evol.* **314**, 369–381 (2010).
- Ahn, D. H. et al. Draft genome of the Antarctic dragonfish, *Parachaenichthys charcoti*. *Gigascience* **6**, 1–6 (2017).
- Baalsrud, H. T. et al. De novo gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msx311> (2017).
- Shin, S. C. et al. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biol.* **15**, 468 (2014).
- Near, T. J. et al. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl Acad. Sci.* **109**, 13698–13703 (2012).
- Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky, P., S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).
- Feng, Y., Madungwe, N. B. & Bopassa, J. C. Mitochondrial inner membrane protein, Mic60/mitofilin in mammalian organ protection. *J. Cell. Physiol.* **234**, 3383–3393 (2019).
- Fernandez-Marcos, P. J. & Auwerx, J. Regulation of PGC-1 $\alpha$ , a nodal regulator of mitochondrial biogenesis. *Am. J. Clin. Nutr.* **93**, 884S–90 (2011).
- De Silva, D., Tu, Y. -T., Amunts, A., Fontanesi, F. & Barrientos, A. Mitochondrial ribosome assembly in health and disease. *Cell Cycle* **14**, 2226–2250 (2015).
- Hessenberger, M. et al. Regulated membrane remodeling by Mic60 controls formation of mitochondrial crista junctions. *Nat. Commun.* **8**, 15258 (2017).
- Novotová, M., Tarabová, B., Tylková, L., Ventura-Clapier, R. & Zahradník, I. Ultrastructural remodelling of slow skeletal muscle fibres in creatine kinase deficient mice: a quantitative study. *Gen. Physiol. Biophysics* **35**, 477–486 (2016).
- O'Brien, K. M., Mueller, I. A., Orczewska, J. I., Dullen, K. R. & Ortego, M. Hearts of some Antarctic fishes lack mitochondrial creatine kinase. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* **178**, 30–36 (2014).
- Cogliati, S. et al. Mitochondrial cristae shape determines respiratory chain supercomplexes assembly and respiratory efficiency. *Cell* **155**, 160–171 (2013).
- Zhang, Y. et al. Mitoguardin regulates mitochondrial fusion through MitoPLD and is required for neuronal homeostasis. *Mol. Cell* **61**, 111–124 (2016).
- Lahiri, S. et al. A conserved endoplasmic reticulum membrane protein complex (EMC) facilitates phospholipid transfer from the ER to mitochondria. *PLoS Biol.* **12**, e1001969 (2014).
- Falkenberg, M. et al. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat. Genet.* **31**, 289–294 (2002).
- Ventura-Clapier, R., Garnier, A. & Veksler, V. Transcriptional control of mitochondrial biogenesis: the central role of PGC-1 $\alpha$ . *Cardiovasc Res.* **79**, 208–217 (2008).
- Bremer, K., Kocha, K. M., Snider, T. & Moyes, C. D. Sensing and responding to energetic stress: the role of the AMPK-PGC1 $\alpha$ -NRF1 axis in control of mitochondrial biogenesis in fish. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **199**, 4–12 (2016).
- Magnoni, L. J., Vraskou, Y., Palstra, A. P. & Planas, J. V. AMP-activated protein kinase plays an important evolutionary conserved role in the regulation of glucose metabolism in fish skeletal muscle cells. *PLoS ONE* **7**, e31219 (2012).
- Magnoni, L. J., Palstra, A. P. & Planas, J. V. Fueling the engine: induction of AMP-activated protein kinase in trout skeletal muscle by swimming. *J. Exp. Biol.* **217**, 1649–1652 (2014).
- Naya, F. J. et al. Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat. Med.* **8**, 1303–1309 (2002).
- Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).

56. Cagalinec, M. et al. Role of mitochondrial dynamics in neuronal development: mechanism for Wolfram syndrome. *PLoS Biol.* **14**, e1002511 (2016).
57. Audano, M. et al. Zc3h10 is a novel mitochondrial regulator. *EMBO Rep.* **19**, pii: e45531 (2018).
58. Adachi, Y. et al. Coincident phosphatidic acid interaction restrains Drp1 in mitochondrial division. *Mol. Cell* **63**, 1034–1043 (2016).
59. Bruni, F., Polosa, P. L., Gadaleta, M. N., Cantatore, P. & Roberti, M. Nuclear respiratory factor 2 induces the expression of many but not all human proteins acting in mitochondrial DNA transcription and replication. *J. Biol. Chem.* **285**, 3939–3948 (2010).
60. Yang, Z.-F., Drumea, K., Mott, S., Wang, J. & Rosmarin, A. G. GABP transcription factor (nuclear respiratory factor 2) is required for mitochondrial biogenesis. *Mol. Cell. Biol.* **34**, 3194–3201 (2014).
61. Sharon, E. et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30**, 521–530 (2012).
62. Smith, R. P. et al. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat. Genet.* **45**, 1021–1028 (2013).
63. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
64. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
65. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
66. Pauletto, M. et al. Genomic analysis of *Sparus aurata* reveals the evolutionary dynamics of sex-biased genes in a sequential hermaphrodite fish. *Communications. Biology* **1**, 119 (2018).
67. Dobin, A. & Gingeras, T. R. Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinforma.* **51**, 11.14.1–19 (2015).
68. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
69. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
70. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucl. Acids Res* **44**, D1251–D1257 (2016).
71. Smith, A. C. & Robinson, A. J. MitoMiner v3.1, an update on the mitochondrial proteomics database. *Nucl. Acids Res.* **44**, D1258–1261 (2016).
72. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
73. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucl. Acids Res.* **37**, W202–208 (2009).
74. Okonechnikov, K., Golosova, O. & Fursov, M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).
75. Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucl. Acids Res.* **38**, W7–13 (2010).

## Acknowledgements

This paper is dedicated to the memory of our colleague Dr. Guido di Prisco, a pioneer in Antarctic science, who recently passed away. This research was supported by Italian PNRA 2013/A21.10, PNRA 16\_00226, PNRA 16\_00099, and PNRA 2016\_00307. We would like to thank Nils Koschnick (AWI) and Emilio Riginella (Stazione Zoologica Anton Dohrn, Napoli) for collecting the samples during “Polarstern” ANT-PS82 and ANT-PS86 cruises. We are indebted to the Polarstern crew for their indispensable help during fishery activities. C.P. acknowledges financial support from the University of Padua (BIRD164793/16) and from the European Marie Curie project “Polarexpress” Grant No. 622320.

## Author contributions

T.P. and L.B. conceived the study. M.L., F.C.M. and C.P. provided the fish. R.C. isolated the *C. myersi* and *C. hamatus* genomic DNA and prepared genomic and RNA-seq libraries. M.B. and N.V. performed the genome assembly and the annotation of the two *Chionodraco* genomes. M.B. and S.F. performed the homology and gene family expansions/contractions analyses. M.B. and S.F. performed the molecular evolution and comparative transcriptomic analyses of icefish data. L.B. oversaw molecular evolution data analysis. C.P., G.S., M.P., M.L., F.C.M., N.V. and L.Z. participated in data interpretation and in the manuscript writing. M.B., S.F., L.B. and T.P. drafted the paper. All authors read, amended, and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42003-019-0685-y>.

**Correspondence** and requests for materials should be addressed to L.B.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019