BMC Genomics

RESEARCH Open Access

Gene loss in Antarctic icefish: evolutionary adaptations mimicking Fanconi Anemia?

Seung Chul Shin[1*], Sanghee Kim[1], Han-Woo Kim[1,2], Jun Hyuck Lee[1,2] and Jin-Hyoung Kim[1,2*]

Abstract

Background The white-blooded Antarctic icefishes is a representative organism that survive under the stenothermal conditions of the Southern Ocean without the hemoglobin genes. To compensate for inefficient oxygen transport, distinct features such as increased heart size, greater blood volume, and reduced hematocrit density enhance the amount of dissolved oxygen and the velocity of blood flow.

Results Here, we investigated these unique characteristics by comparing high-quality genomic data between white-blooded and red-blooded fishes and identified the loss of *FAAP20*, which is implicated in anemia. Although the gene region containing *FAAP20* is conserved in notothenioids as shown through collinear analysis, only remnants of *FAAP20* persist in several icefish species. Additionally, we observed the loss of *SOAT1*, which plays a pivotal role in cholesterol metabolism, providing a clue for further investigations into the unique mitochondrial form of the icefish.

Conclusions The loss of *FAAP20*, which is known to reduce erythrocyte counts under stress conditions in mice and humans, may provide a clue to understanding the genomic characteristics related to oxygen supply, such as low hematocrit, in Antarctic icefishes.

Keywords Antarctic icefish, Gene loss, Fanconi anemia, *FAAP20*

Background

Antarctic icefishes, belonging to the Channichthyidae family, have evolved to survive in the extremely cold and oxygen-rich waters of the Southern Ocean surrounding Antarctica, where temperatures range from −2 to 4 °C [1–5]. These unique conditions have enabled the evolution of distinct physiological characteristics. Antarctic icefish exhibit increased mitochondrial density in cardiac myocytes and oxidative skeletal muscle, along with elevated levels of mitochondrial phospholipids—phosphatidylethanolamine and phosphatidylcholine [3]. Additionally, their substantially reduced bone density may facilitate buoyancy control [6]. Notably, white-blooded icefishes lack hemoglobin, and some species also lack myoglobin [7]. They carry oxygen in their blood plasma at approximately 10% of the levels found in their red-blooded relatives [8, 9]. These extreme phenotypes have led to their classification as the only known vertebrates without hemoglobin [10]. To compensate, they have developed larger hearts, increased blood volume—two to four times that of red-blooded teleosts—and lower hematocrit density [7, 11]. Such characteristics likely enhance dissolved oxygen levels and blood flow velocity, which appears to enable survival in the cold environments of the Southern Ocean.

*Correspondence:
Seung Chul Shin
biotech21@gmail.com
Jin-Hyoung Kim
kimjh@kopri.re.kr
[1]Division of Life Sciences, Korea Polar Research Institute (KOPRI), Incheon 21990, Republic of Korea
[2]Polar Sciences, University of Science and Technology, Incheon 21990, Republic of Korea

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Recent advancements in high-quality genomic data have facilitated studies on the evolution of Antarctic icefish [1, 2, 5, 10, 12–14] and reports continue to be published regarding chromosome-level genome sequences of notothenioids including *Cottoperca gobio* from the Bovichidae family and *Eleginops maclovinus* from the Eleginopidae family [1, 5, 12, 13, 15] which diverged before the Drake Passage opened between 49 and 17 million years ago [16] and evolved in a non-freezing environment unlike other notothenioids. To date, seven chromosome-scale genome sequences of notothenioids have been published, with gene annotations available for all except *Notothenia rossii* [1, 5, 12]. An analysis of 36 notothenioid genome assemblies revealed detrimental alterations in hemoglobin genes affecting oxygen affinity and multiplicity [2], which likely contributed to the hemoglobin gene loss [2]. A genomic study of 46 notothenioid species indicated that icefish retain proerythroblasts, with the evolution of erythrocyte-associated conserved non-coding elements rather than coding regions affecting the transition from normoblast to erythrocyte, resulting in anemia in white-blooded icefish [10]. However, further genomic studies examining distinct characteristics such as blood volume, low hematocrit, and larger heart size are warranted. Therefore, by comparing high-quality genomes of red-blooded and white-blooded Antarctic fish, we investigated orthologous gene groups to find genetic factors crucial for these modifications to extreme cold and oxygen-rich marine environments and identified gene loss associated with anemia to elucidate the mechanisms underlying erythrocyte loss in Antarctic icefish.

Results

Chromosome-level genome assembly and annotation of *Champsocephalus gunnari*

We sequenced the genome of *Champsocephalus gunnari* as a chromosome-level genome assembly by integrating nanopore reads, Illumina short-reads, and Hi-C methods, achieving a BUSCO (Benchmarking Universal Single-Copy Orthologs) [17] genome completeness of 97.9% (Additional file 1: Table S1). Using the BRAKER3 tool [18] with RNA-seq and protein evidences from nine teleost fish, including five notothenioids, we predicted 45,002 protein-coding genes and the completeness of the gene annotation was determined to be 95.9% against the Actinopterygii database obd10 using BUSCO (Additional file 1: Table S2). The proportion of BUSCO duplicates and fragmented BUSCOs were 1.1% and 1%, respectively. Recently, a chromosome-level assembly for this species was independently reported [5]; however, we used the results of the present study for a representative species.

Gene annotation of *Notothenia rossii*

To compare gene sets of white-blooded and red-blooded fishes and explore distinct characteristics to life without hemoglobin, comprehensive gene annotation was essential for all seven fishes. The gene annotation for *N. rossii* was unavailable; therefore, we annotated the genome of *N. rossii* using RNA-seq data (ERX4524565-ERX4524570: gonad, muscle, and brain) from the Sequence Read Archive (SRA) and utilized the BRAKER3 tool, incorporating the same protein evidence used in the annotation process of *C. gunnari*. Consequently, we obtained high-quality gene annotations with gene set completeness exceeding 90% (Additional file 1: Table S3), with a predicted gene count of 44,502. The complete BUSCO score was 91.9% [19] for *N. rossii*, while the proportion of duplicated and fragmented BUSCOs were 1.1% and 1.6%, respectively.

Before further analyses, we confirmed the BUSCO completeness for the gene set, and verified that the complete BUSCOs of the predicted coding genes for all 10 fishes exceeded 90% against Actinopterygii database obd10 (Additional file 1: Table S4).

Conserved synteny between temperate and Antarctic icefish genomes

We conducted genomic synteny and protein-based collinearity analyses among seven fish species. The genomic synteny analysis was performed using SyMAP [20] between *C. gunnari* and four other Antarctic fish species with chromosome-scale assemblies (Additional file 1: Fig. S1-4). Twenty-four orthologous chromosomes showed one-to-one correspondence, indicating conservation in large-scale genome organization between *C. gunnari* in the present study and the genomes of three other Antarctic icefishes (Additional file 1: Fig. S1-3). Each chromosome of *N. rossii* exhibited one-to-two correspondence, also showing conservation in large-scale genome organization with the *C. gunnari* genome (Additional file 1: Fig. S4). When compared with *C. gunnari* from a previous study [5], although a higher degree of conserved synteny along whole chromosomes was observed (Additional file 1: Fig. S1), several chromosomal rearrangements, including inversions, were also identified. Additionally, when compared with the *C. esox* genome, two large chromosomal inversions specific to *C. esox* in chromosome 21 were identified as well (Additional file 1: Fig. S2) [5]. Protein-based collinearity analysis, performed using OrthoVenn3 [21], further confirmed the highly conserved gene order on chromosomes among the seven fish species with genomic synteny (Fig. 1a).

Orthologous gene cluster analysis

We investigated orthologous gene groups using OrthoVenn3 with OrthoMCL algorithm analyzing

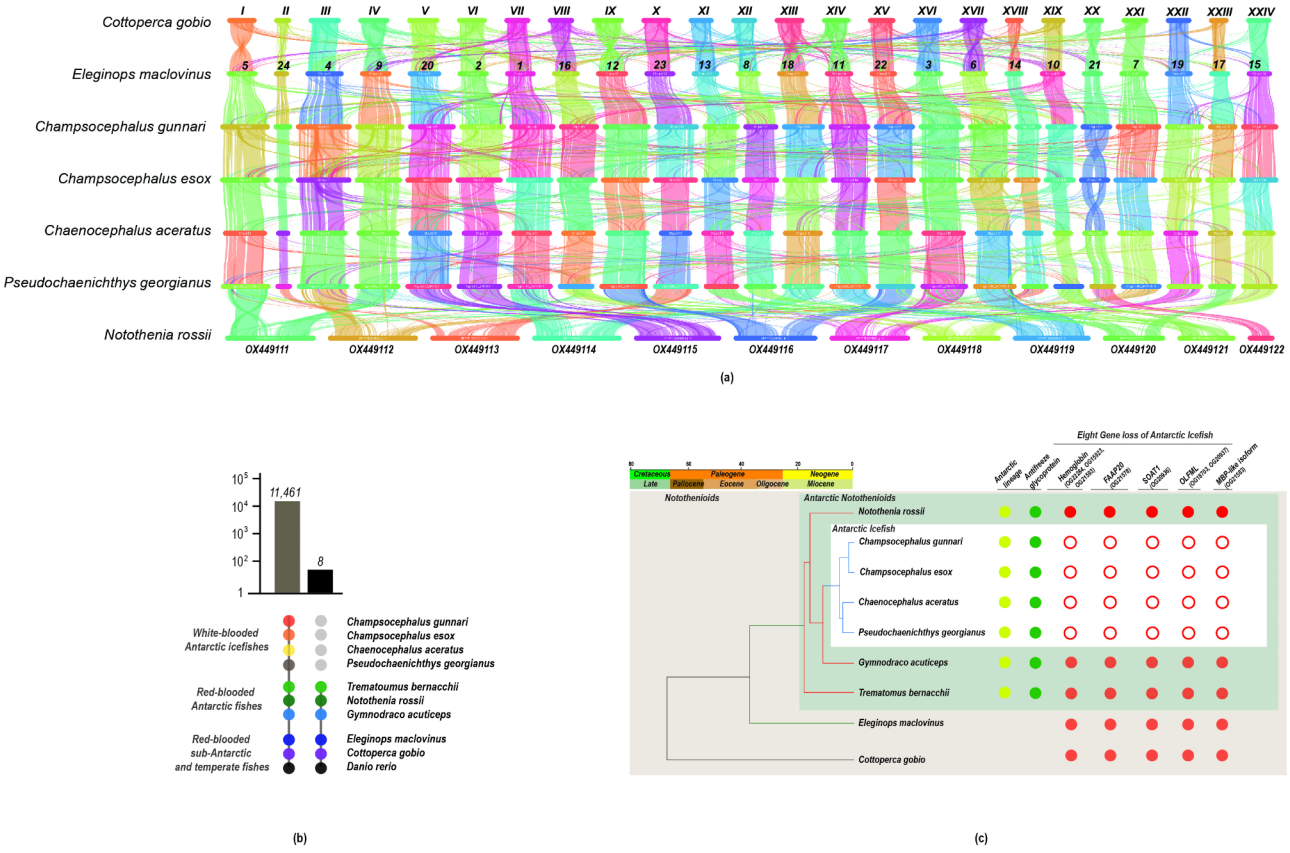
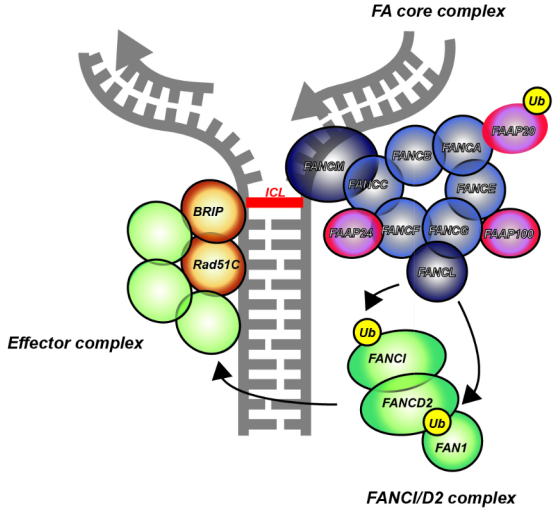


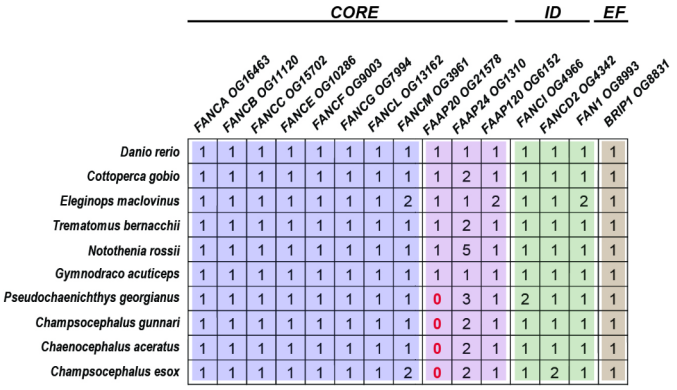
Fig. 1 Protein-based collinearity analysis and OGs in Antarctic notothenioids. **(a)** Collinearity analysis of seven chromosome-scale genomes from Antarctic and sub-Antarctic fish. Each chromosome is numbered with Roman numerals, except for those of *Eleginops maclovinus* and *Champsocephalus esox*, which are numbered with Arabic numerals. For *Notothenia rossii*, 12 chromosomes, organized by their NCBI accession numbers, are presented for comparison. **(b)** A total of 11,461 orthologous groups were identified across all fish, while 8 orthologous groups were found exclusively in red-blooded fish and not in white-blooded icefishes. **(c)** Phylogenetic tree (left) depicting the evolutionary relationships among nine notothenioid species, highlighting the divergence between Antarctic and sub-Antarctic lineages. Branches are color-coded by lineage: Antarctic icefish (blue), red-blooded notothenioids (red), and their related species (green). The ultrametric tree includes divergence times and marks each species with closed circle for specific evolutionary traits, highlighting the presence of antifreeze glycoprotein (AFGP), and the number of orthologous groups for each lineage. For the 8 genes not found in white-blooded icefish, the presence or absence of the genes was indicated by closed circles and open circles. All species presented are part of the Antarctic notothenioid clade, except for *C. gobio* and *E. maclovinus* which are members of two out of the three non-Antarctic notothenioid families

high-quality gene sets from four white-blooded icefishes, three red-blooded Antarctic fishes, and three red-blooded temperate fishes, including *Danio rerio*, *C. gobio*, and *E. maclovinus*. The genome sequences of *T. bernacchii* [1] and *G. acuticeps* [1] were assembled with long reads but not at the chromosome scale. Except for zebrafish (*D. rerio*), which serves as our non-notothenioid teleost outgroup, all species were notothenioids. A total of 29,436 orthologous gene groups were identified, with 11,461 gene groups shared among ten species (Additional file 1: Fig. S5). In total, 86.4% of protein coding genes in *C. gunnari* and 92.1% of those in *N. rossii* were clustered into orthologous gene groups.

We focused on orthologous groups (OGs) that were present in red-blooded but absent in white-blooded fishes to determine the gene losses impacting their unique physiological traits, such as enlarged hearts, increased blood volume, and atypical mitochondrial functions. We observed only 8 OGs that were absent specifically in the four icefishes, and present in the genomes of all the red-blooded species studied (Fig. 1b). We identified several genes and their associated functions. Hemoglobin genes (OG22284, OG15293, and OG21593), whose absence is well-documented in Antarctic icefishes [2], are essential for oxygen transport; olfactomedin-like genes (OLFML) (OG18703 and OG20937) contribute to early development, nervous system organization, and hematopoiesis [22–24]; the myelin basic protein-like isoform (MBP) (OG21583) supports the sheathing of myelin, enhancing neural signal transmission [25, 26]; sterol O-acyltransferase 1 (SOAT1) (OG20936) plays a role in cholesterol metabolism with implications for Alzheimer's disease [27, 28] (Additional file 1: Supplementary Note) ; and Fanconi anemia core complex-associated protein 20 (FAAP20) (OG21578) is linked directly to Fanconi anemia (Fig. 1c) [4, 29–32] (Additional file 1: Table S5).

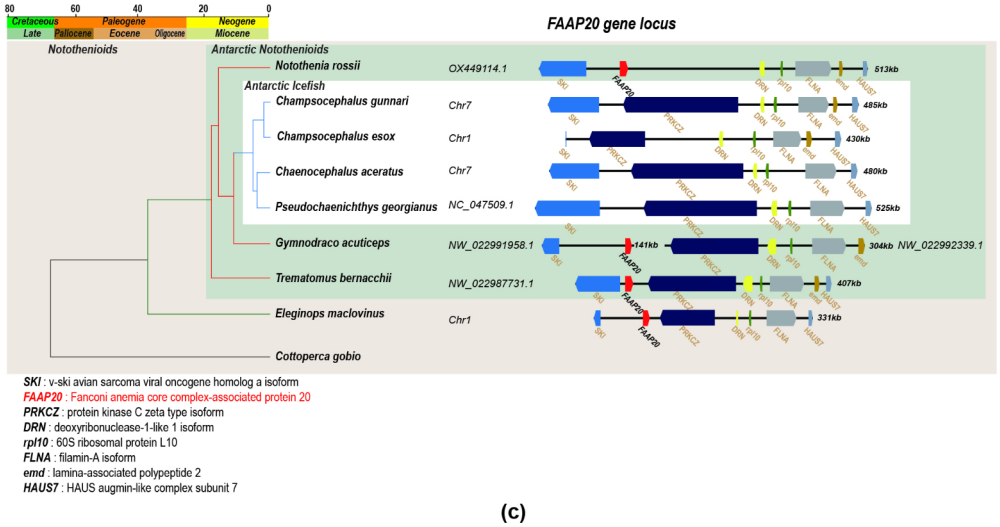


(a)

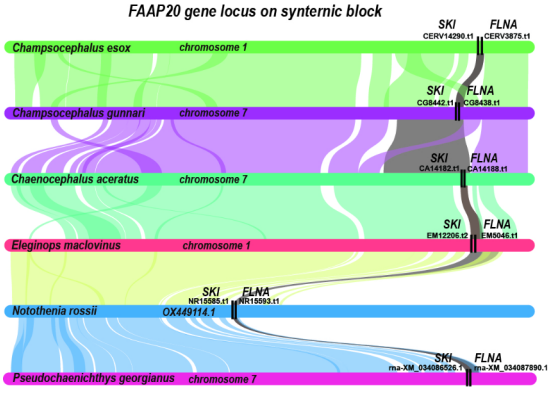


	CORE										ID		EF	
	FANCA OG16463	FANCB OG11120	FANCC OG15702	FANCE OG10286	FANCF OG9003	FANCG OG7994	FANCL OG13162	FANCM OG3961	FAAP20 OG21578	FAAP24 OG1310	FAAP100 OG6152	FANCI OG4966	FANCD2 OG4342	FAN1 OG8993	BRIP1 OG8831
Danio rerio	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cottoperca gobio	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
Eleginops maclovinus	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1
Trematomus bernacchii	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
Notothenia rossii	1	1	1	1	1	1	1	1	1	5	1	1	1	1	1
Gymnodraco acuticeps	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pseudochaenichthys georgianus	1	1	1	1	1	1	1	1	0	3	1	2	1	1	1
Champsocephalus gunnari	1	1	1	1	1	1	1	1	0	2	1	1	1	1	1
Chaenocephalus aceratus	1	1	1	1	1	1	1	1	0	2	1	1	1	1	1
Champsocephalus esox	1	1	1	1	1	1	1	2	0	2	1	1	2	1	1

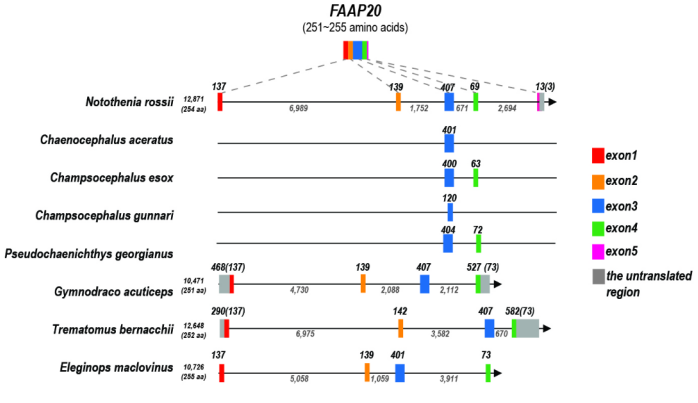
(b)



(c)



(d)



(e)

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Analysis of *FAAP20* loss in Antarctic icefish. **(a)** Fanconi anemia (FA) repair pathway depicting the FA core complex, including FANCA, B, C, D, E, F, G, L, M, and accessory components (FAAP20, FAAP24, FAAP100). The ID complex comprises FANCI, FANCD2, and FAN1, whereas BRIP1 functions in the effector complex. The FA core complex, through FANCM, recognizes interstrand crosslink (ICL) lesions and monoubiquitinates the ID complex (FANCI and FANCD2), which subsequently recruits the effector complex to break the crosslink. **(b)** Fifteen orthologous groups linked to Fanconi anemia. **(c)** Gene regions between the SKI and FLNA loci on the orthologous chromosomes or contigs of the eight fish species. Each horizontal line depicts the chromosome span, arranged from the 5′ to 3′ orientation. Colored arrows depict various genes, with each color representing a different gene within the locus. *FAAP20* region is split into two contigs in *G. acuticeps*: NW_022991958.1 and NW_022992339.1. **(d)** Gene region containing *FAAP20* identified in a syntenic block based on the orthologous protein group. The *FAAP20* region is located on the orthologous chromosomes in five Antarctic fish (*N. rossii*, *C. gunnari*, *C. esox*, *C. aceratus*, and *P. georgianus*) and on OX449114.1 in *N. rossii*. Collinearity is visualized to focus on the chromosome length proportion within each species, with the gray region depicting the syntenic block containing the *FAAP20* locus. **(e)** *FAAP20* spanning approximately 10–12 kb on the chromosome, with the FAAP20 protein (ranging from 251–255 amino acids in length) translated from 4 or 5 exons

Loss of *FAAP20* in Antarctic white-blooded icefish

The absence of *FAAP20* represents the first documented gene loss associated with anemia in Antarctic icefishes. The Fanconi anemia core complex in fish comprises 11 genes (Fig. 2a) [30, 31]. We confirmed the absence of FAAP20, while the presence of components such as *FANCA*, *FANCB*, *FANCC*, *FANCE*, *FANCF*, *FANCG*, *FANCL*, *FANCM*, *BRIP1*, *FAAP100*, and *FAAP24* was observed in the OGs across 10 fish species (Fig. 2b). Additionally, components, such as FANCI, FAN1, and FANCD2, necessary for the FANCI/FANCD2 (ID) complex, and BRIP1, involved in the downstream effector complex, were present across the 10 species (Fig. 2b).

We identified regions of *FAAP20* locus within the genome sequence (Fig. 2c). In all assayed genomes, the *FAAP20* locus lies in a conserved synteny block between the SKI and FLNA genes. As revealed by the collinear analysis (Fig. 2d), this locus is located in orthologous chromosomes in all species, named chromosome 7 in *E. maclovinus*, *C. gunnari*, and *C. esox*, chromosome OX449114.1 in *N. rossii*, and chromosome NC_047509.1 in *P. georgianus*. The genome *T. bernacchii* is only assembled at a scaffold-level; however, we could identify a syntenic and complete *FAAP20* locus on scaffold NW_022987731.1. In the *G. acuticeps* genome, also assembled at a scaffold level, the *FAAP20* locus was identified in two separate scaffolds (NW_022991958.1 and NW_022992339.1). *FAAP20* was located 5′ upstream of SKI, similar to the location in the other notothenioid genomes (Fig. 2c). Although the specific mechanism of gene loss remains uncertain, exons 1 and 2 of *FAAP20* were lost in all four Antarctic icefishes, while the fourth exon was present only in *P. georgianus* and *C. esox*. The third exon of *FAAP20* remained in the locus of four Antarctic icefishes (Fig. 2e). *FAAP20* might be lost in the icefish ancestor following their divergence from dragonfishes.

Discussion

The loss of hemoglobin and myoglobin genes, along with the emergence of the AFGP gene, are well-known genetic traits specific to Antarctic icefish [1, 7, 11, 33]. However, the genes associated with unique phenotypes such as increased blood volume, low hematocrit, larger heart size, and increased mitochondrial density remain largely unexplored [8, 9]. In the present study, we identified that Antarctic icefishes have lost genes related to Fanconi anemia, which are known in humans and mice [30, 32].

Fanconi anemia is a rare genetic disorder that affects red blood cells, platelets, and leukocytes, often resulting in inherited bone marrow failure and pancytopenia [30]. Unlike typical genetic disorders that impact blood cell differentiation, Fanconi anemia results from a compromised response to DNA damage, specifically to interstrand crosslinks within the FA/BRCA pathway. This pathway comprises the FA core, the ID, and the downstream effector complexes [30] (Fig. 2a). FAAP20, an integral component of the FA core complex, directly stabilizes FANCA and is essential for DNA repair and genome maintenance. Deletion of Faap20 in mice results in a mild FA-like phenotype with defects in the reproductive and hematopoietic systems [32]. Notably, hematopoietic stem and progenitor cells (HSPCs) from these mice showed defects in long-term multilineage reconstitution in lethally irradiated recipient mice. Additionally, these mice are prone to mitomycin C (MMC)-induced pancytopenia, a condition primarily associated with interstrand cross-link formation, where MMC mainly alkylates the N-2 position of guanine [34]. Acute MMC stress also leads to substantial progenitor loss, especially of the megakaryocyte-erythroid progenitors and the common myeloid progenitor in Faap20$^{-/-}$ mice [32], resulting in severe pancytopenia with a considerably reduced erythroid lineage when compared with that in the myeloid/lymphoid lineage (Fig. 3) [32].

Zebrafish, a key model organism, possesses a complete complement of FA family members found in humans [31, 35]. Although the loss of FAAP20 has not been studied in fish [36], studies on rad51$^{-/-}$, a component of the effector complex in zebrafish, have shown reduced HSPC proliferation when compared with that in their wild-type counterparts, suggesting that zebrafish can serve as a model for Fanconi anemia in fish. In our study of orthologous proteins, Antarctic icefish exhibited a complete complement of FA core complex genes associated with Fanconi anemia, except for FAAP20. We also identified FAN1,

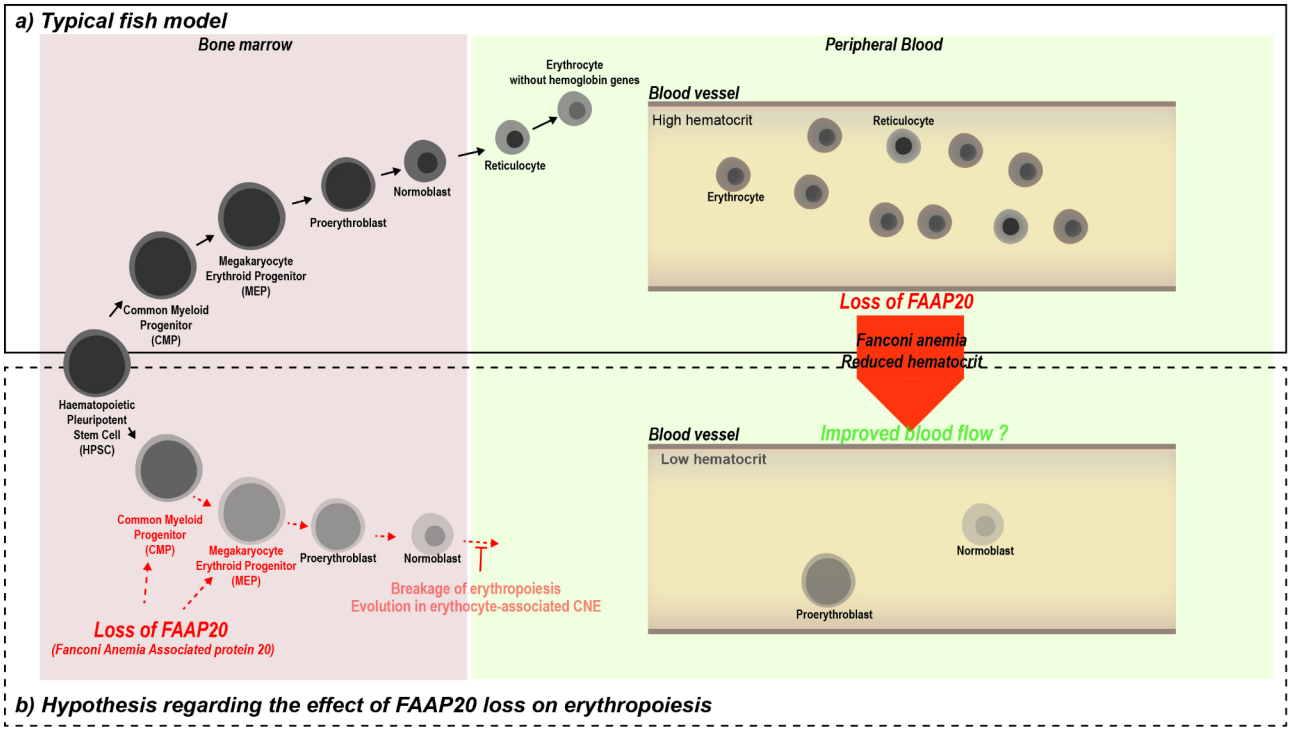


Fig. 3 Hypothesis regarding the effect of FAAP20 loss on erythropoiesis. **(a)** The development of the hematopoietic lineage in a typical fish model (upper panel). In fish, red blood cells are nucleated. The presence of FAAP20 results in the normal differentiation of erythrocyte from hematopoietic stem cells (HPSCs) in Antarctic fish. **(b)** Contrastingly, the diagram (lower panel) illustrates the hypothesis regarding the effect of FAAP20 loss on the predicted development of the hematopoietic lineage in Antarctic icefish. The absence of FAAP20 may reduce the common myeloid progenitor and GMP in the erythroid progenitors, similar to that in the FAAP$^{-/-}$mouse model. Additionally, the evolution of the erythrocyte-associated conserved non-coding elements affects the differentiation of the HPSC into erythrocytes, resulting in the breakdown of the erythropoietic process and anemia in Antarctic icefish. FAAP20 loss in Antarctic fish might be a factor contributing to pancytopenia in the erythroid progenitors

FANCI, FANCD (ID complex), and BRCA1 (effector complex). The phenotype associated with FAAP20 loss in Antarctic icefish may mirror that observed in Faap20 knockout (Faap20$^{-/-}$) mice, which exhibit pancytopenia in erythroid progenitors (Fig. 3). Consequently, the loss of FAAP20 in Antarctic icefish might be one of the factors responsible for their failure to differentiate from HSPCs into erythrocytes, a process supported by evolutionary changes in erythrocyte-associated conserved non-coding elements (CNEs) [10].

Faap20-null mice do not spontaneously develop pancytopenia without MMC. In Antarctic icefish, inducers that heighten Faap20-null symptoms appear to be present. Icefish have a reduced capacity to detoxify reactive oxygen species (ROS) compared with that of red-blooded species [37, 38], and their mitochondria potentially produce higher levels of ROS than those of red-blooded notothenioids, whereas polyunsaturated fatty acid (PUFA) levels are higher in Antarctic fish than in temperate fish [39]. Therefore, low antioxidant levels combined with high PUFA levels may render icefish particularly susceptible to oxidative stress. DNA interstrand crosslinks that occur under oxidative conditions [34] could be predicted to form in a manner similarly to those induced by MMC, and the effects of FAAP20 loss without MMC might also occur in Antarctic icefish. Recent reports have highlighted the pivotal role of FAAP20 in homologous recombination at DNA double-strand breaks not associated with interstrand crosslinks [4]. Thus, the FA occurrence may readily result from oxidative and replication stress in the absence of FAAP20 [40].

Conclusions

In conclusion, our findings offer insights into potential genomic features in Antarctic icefish, suggesting gene losses that may contribute to their specialized physiological traits. The observed loss of FAAP20 raises the possibility of its role in anemia observed in Antarctic icefishes, hinting at a genetic alteration that could facilitate improved blood flow in these hemoglobin-free organisms and potentially enhance oxygen transport in extreme environments. While the causal nature of the FAAP20 loss in icefish remains unknown, our findings set the stage for future studies exploring the biological and clinical implications of these gene loses through experimental validation.

Materials and methods

Sample collection and sequencing

The mackerel icefish, *Champsocephalus gunnari*, was obtained from Jeong Il Corporation (Seoul, KOREA), a krill fishery. Wild *C. gunnari* was caught as a by-catch species in Subareas 48.1 and 48.2 of the CCAMLR (Convention on the Conservation of Antarctic Marine Living Resources) Convention area from November 2020 to February 2021. Samples were frozen in a freezer (−20 °C) immediately after collection and were transported to the Korea Polar Research Institute within 6 months. One individual was selected for this study. A DNeasy Tissue Kit (Qiagen, Valencia, CA, USA) was used to extract the total DNA. The DNA quality was assessed using a NanoDrop Spectrophotometer (Thermo Fisher, USA), and quantified using a Qubit Fluorometer (Invitrogen, USA). Phyzen Co. Ltd. (Seongnam, Korea) performed all library construction and sequencing procedures using Oxford Nanopore technology (Additional file 1: Table S9). The paired-end library (2×150 bp) was constructed and sequenced on the Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA) for genome polishing and genome size prediction (Additional file 1: Table S9).

Twelve tissues comprising the brain, egg, esophagus, eye, fin, heart, kidney, muscle, skin, spinal, spleen, and stomach were dissected and subsequently stored in RNAlater (Ambion, Inc., Austin, TX, USA) solution for RNA extraction. The total RNA was extracted from each tissue using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. For constructing the library, total RNA was used and sequencing was performed on the Illumina HiSeq 2000 platform (Additional file 1: Table S9). The paired-end reads (2×151 bp) obtained for total RNAs were trimmed using FASTX-Toolkit (v.0.0.11) [41] with the parameters "-t 30," "-l 80," and "-Q 33" and subsequently used for further annotation.

Chromatin conformation capture data was generated by Phase Genomics (Seattle, WA, USA) using the Proximo Hi-C 2.0 Kit, which is a commercially available version of the Hi-C protocol. The muscle tissue was frozen in liquid nitrogen, ground, crosslinked using a formaldehyde solution, and sent to Phase Genomics for library preparation following the manufacturer's protocol. Hi-C library sequencing was performed on an Illumina HiSeq 4000.

Genome assembly and Hi-C scaffolding

The Oxford Nanopore long reads obtained from DNA sequencing were assembled using NextDenovo (v.2.4.0) (task=all; input type=raw; read type=ont; read_cutoff=1k, seed_cutoff=1k) [42] and polished using NextPolish (v.1.3.1) with default options [43], using trimmed Illumina short reads generated from the muscle tissue. Read trimming was performed using Trimmomatic (v.0.38) [44], with the following command line parameters: "ILLUMINACLIP: adaptor.file:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36". Haploid assemblies were scaffolded using Hi-C after processing with Purge Haplotigs [45] [17]. For Hi-C scaffolding, BWA-MEM (v.0.7.17-r1198-dirty) [46] was used to align the reads to the draft assembly. Phase Genomics Proximo Hi-C genome scaffolding platform was used to create chromosome-scale scaffolds from the polished assembly. We evaluated the completeness of the assembled genome using BUSCO (v.5.1.3) against Actinopterygii_odb10 database [17].

Gene annotation of *C. gunnari*

The BRAKER3 [18] annotation pipeline (v.3.0.8) was used to predict the protein-coding genes in the assembled genome. Protein sequence of *Danio rerio* (GCA_000002035.4), *Takifugu rubripes* (GCA_901000725.2), *Xiphophorus maculatus* (GCA_002775205.2), *Oryzias latipes* (GCA_002234675.1), *Eleginops maclovinus* (GCA_036324505.1), *Trematomus bernacchii* (GCA_902827165.1), *Gymnodraco acuticeps* (GCA_902827175.1), *Pseudochaenichthys georgianus* (GCA_902827115.1), and *Cottoperca gobio* (GCA_900634415.1) were used as protein hints, and the RNA evidence included trimmed paired-end reads for total RNA of 10 tissues. The reference sequence included masked genome sequences using RepeatMasker with the option '-species Actinopterygii' against the RepBase library (RepBaseRepeatMaskerEdition-20181026) (v.4.0.7) [47]. BRAKER3 was performed in protein and transcript modes with the default option. We verified the protein-coding gene annotation using BUSCO in protein mode against the Actinopterygii_odb10 database. Functional annotation of the protein-coding genes was performed using EggNOG mapper (v.2.1.12) [48] against the EggNOG database (v.5.0) [49] and BLASTp (v.2.0.8) against the nr database.

Gene annotation of *N. rossii*

Protein-coding genes of *N. rossii* (GCA_949606895.1) were predicted using the BRAKER3 annotation pipeline using SRA (ERX4524565-ERX4524570: brain, gonads, muscle, and brain) as RNA-Seq evidence. Genome sequence (GCA_949606895.1) was downloaded from NCBI and masked using RepeatMasker with the option '-species Actinopterygii' against the RepBase library (RepBaseRepeatMaskerEdition-20181026) (v.4.0.7). The same protein evidence used for the annotation of *C. gunnari* was applied to the BRAKER3 annotation pipeline. The predicted protein-coding genes were annotated using EggNOG mapper (v.2.1.12) [48] against the

EggNOG database (v.5.0) [49] and using BLASTp (v.2.0.8) against the nr database.

Ortholog, phylogenetic, and gene family analyses

OrthoVenn3 [21]was used to identify orthologous genes across the *C. gunnari* assembled genome and nine fish genomes *Eleginops maclovinus* (GCA_036324505.1), *Trematomus bernacchii* (GCA_902827165.1), *Danio rerio* (GCA_000002035.4), *Gymnodraco acuticeps* (GCA_902827175.1), *Pseudochaenichthys georgianus* (GCA_902827115.1), and *Cottoperca gobio* (GCA_900634415.1), *Notothenia rossii* (GCA_949606895.1) and *Chaenocephalus aceratus* (GCA_023974075.1), *Chaenocephalus esox* (GCA_036324585.1), which were used for protein homology analysis for gene prediction. All are chromosome-scale genome sequences except for *Trematomus bernacchii* and *Gymnodraco acuticeps*. Protein sequences from the 10 fish genomes were clustered with an E-value of 1e-5 and an inflation value of 1.5 with the OrthoFinder algorithm [50]. TimeTree [51] in OrthoVenn3 was used to infer the divergence time between species.

Syntenic alignment and collinearity analysis

SyMAP (v.4.0) was used to analyze the syntenic blocks among the fish genomes [20], and a dot plot and circle map were generated. The conserved and variable genomic structures were detected using collinearity analysis in OrthoVenn3. The input of OrthoVenn3 included a GFF file downloaded from NCBI and generated from BRAKER3 [21].

Validation of gene loss in assembly and sequencing reads

To confirm the loss of thegenes encoding FAAP20, SOAT1, OLFM, OLFM4, and MBP-like isoform, we searched for the gene sequences in both the assembly and sequencing reads. For the assembly search, we used a TBLASTN search in SequenceServer (v.2.0.0) [52]. First, we established a database using the genome sequences of four white-blooded icefishes and performed a TBLASTN [53] search for gene sequences from red-blooded Antarctic fish corresponding to genes lost in white-blooded icefishes. We used the protein sequences of *G. acuticeps* for FAAP20 (XP_034056600.1), SOAT1(XP_034094487.1), OLFM4 (XP_034051626.1 and XP_034051602.1), and MBP-like isoform (XP_034091857.1). Additionally, we searched for the gene sequences in the sequencing reads of *C. gunnari*. We established a database using Nanopore and Illumina reads and performed a TBLASTN search for gene sequences from red-blooded Antarctic fish corresponding to genes lost in white-blooded icefishes again. We confirmed the complete loss of the SOAT1, OLFM, OLFM4, and MBP-like isoform genes, excluding FAAP20. In the TBLASTN search of FAAP20, the loss of exons 1, 2, and 4 was identified.

Abbreviations

BUSCO	Benchmarking Universal Single-Copy Orthologs
CCAMLR	Convention on the Conservation of Antarctic Marine Living Resources
FA	Fanconi anemia
HSPC	Hematopoietic stem and progenitor cells
KOPRI	Korea Polar Research Institute
ROS	Reactive oxygen species
SRA	Sequence Read Archive

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11028-0.

Supplementary Material 1

Acknowledgements

The authors would like to thank the Jeong Il Corporation for supplying the mackerel icefish (*Champsocephalus gunnari*) samples.

Author contributions

S.C.S. and J.-H.K. conceived and designed the experiments. S.C.S. and J.-H.K. collected the samples and performed the experiments. S.C.S., H.S.K., H.W.K., J.H.L., and J.-H.K. analyzed the data. All authors participated in the writing of the manuscript.

Funding

The Korea Polar Research Institute (Grant No. PE24160) supported this work.

Data availability

The raw sequencing data were deposited in the NCBI SRA under BioProject PRJNA1113419 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1113419) with accessions SRX24633213, SRX24633214, SRX24633215, SRX24633216, SRX24633217, SRX24633218, SRX24633219, SRX24633220, SRX24633221, SRX24633222, SRX24633223, SRX24633224, SRX24633225, SRX24633226 and SRX24633227 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1113419). The assembled genome sequence, protein sequence, annotation information, and orthologous group were deposited in Figshare (https://doi.org/10.6084/m9.figshare.25837762.v1) [19].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 October 2024 / Accepted: 11 November 2024
Published online: 18 November 2024

References

1. Bista I, Wood JMD, Desvignes T, McCarthy SA, Matschiner M, Ning Z, et al. Genomics of cold adaptations in the Antarctic notothenioid fish radiation. Nat Commun. 2023;14:3412.
2. Desvignes T, Bista I, Herrera K, Landes A, Postlethwait JH. Cold-driven hemoglobin evolution in Antarctic notothenioid fishes prior to hemoglobin gene loss in white-blooded icefishes. Mol Biol Evol. 2023;40:msad236.

3. O'Brien KM, Mueller IA. The unique mitochondrial form and function of Antarctic channichthyid icefishes. Integr Comp Biol. 2010;50:993–1008.
4. Palovcak A, Yuan F, Verdun R, Luo L, Zhang Y. Fanconi anemia associated protein 20 (FAAP20) plays an essential role in homology-directed repair of DNA double-strand breaks. Commun Biol. 2023;6:873.
5. Rivera-Colón AG, Rayamajhi N, Minhas BF, Madrigal G, Bilyk KT, Yoon V, et al. Genomics of secondarily temperate adaptation in the only non-antarctic icefish. Mol Biol Evol. 2023;40:msad029.
6. Eastman JT, Witmer LM, Ridgely RC, Kuhn KL. Divergence in skeletal mass and bone morphology in Antarctic notothenioid fishes. J Morphol. 2014;275:841–61.
7. 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.
8. Ruud JT. Vertebrates without erythrocytes and blood pigment. Nature. 1954;173:848–50.
9. 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;34:457–71.
10. Daane JM, Auvinet J, Stoebenau A, Yergeau D, Harris MP, Detrich HW III. Developmental constraint shaped genome evolution and erythrocyte loss in Antarctic fishes following paleoclimate change. PLoS Genet. 2020;16:e1009173.
11. Hemmingsen EA, Douglas EL. Respiratory characteristics of the hemoglobin-free fish *Chaenocephalus aceratus*. Comp Biochem Physiol. 1970;33:733–44.
12. Lee SJ, Kim J, Choi EK, Jo E, Cho M, Kim J-H, et al. A chromosome-level reference genome of the Antarctic blackfin icefish *Chaenocephalus aceratus*. Sci Data. 2023;10:657.
13. Bista I, McCarthy SA, Wood J, Ning Z, Detrich iii HW, Desvignes T, et al. The genome sequence of the channel bull blenny, *Cottoperca gobio* (Günther, 1861). Wellcome Open Res. 2020;5:148.
14. Ansaloni F, Gerdol M, Torboli V, Fornaini NR, Greco S, Giulianini PG, et al. Cold adaptation in Antarctic notothenioids: comparative transcriptomics reveals novel insights in the peculiar role of gills and highlights signatures of cobalamin deficiency. Int J Mol Sci. 2021;22:1812.
15. Auvinet J, Graça P, Dettai A, Amores A, Postlethwait JH, Detrich HW, et al. Multiple independent chromosomal fusions accompanied the radiation of the Antarctic teleost genus Trematomus (Notothenioidei:Nototheniidae). BMC Evol Biol. 2020;20:39.
16. 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.
17. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Gene prediction. Methods Protoc. 2019:227 – 45.
18. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. Genome Res. 2024;34:769–77.
19. Shin SC. Chromosome-level genome assembly and annotation of *Champsocephalus gunnari*. Figshare. 2024. https://doi.org/10.6084/m9.figshare.25837762.v1.
20. Soderlund C, Nelson W, Shoemaker A, Paterson A. SyMAP: A system for discovering and viewing syntenic regions of FPC maps. Genome Res. 2006;16:1159–68.
21. Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res. 2023;51:W397–403.
22. Anholt RRH. Olfactomedin proteins: central players in development and disease. Front Cell Dev Biol. 2014;2:6.
23. Chin KL, Aerbajinai W, Zhu J, Drew L, Chen L, Liu W, et al. The regulation of OLFM4 expression in myeloid precursor cells relies on NF-κB transcription factor. Br J Haematol. 2008;143:421–32.
24. Liu W, Liu Y, Wang R, Li C, Deng C, Rodgers GP. Olfactomedin 4 is essential for superoxide production and sensitizes oxidative stress-induced apoptosis in neutrophils. Blood. 2009;114:1356.
25. Martenson RE. In: Myelin, editor. Myelin basic protein isoforms: structural and evolutionary implications of alternative exon splicing. Routledge; 2023. pp. 387–411.
26. Boggs JM. Myelin basic protein: a multifunctional protein. Cell Mol Life Sci. 2006;63:1945–61.
27. Wu N, Li R-Q, Li L. SOAT1 deficiency attenuates atherosclerosis by regulating inflammation and cholesterol transportation via HO-1 pathway. Biochem Biophys Res Commun. 2018;501:343–50.
28. Guan C, Niu Y, Chen S-C, Kang Y, Wu J-X, Nishi K, et al. Structural insights into the inhibition mechanism of human sterol O-acyltransferase 1 by a competitive inhibitor. Nat Commun. 2020;11:2478.
29. Leung JWC, Wang Y, Fong KW, Huen MSY, Li L, Chen J. Fanconi anemia (FA) binding protein FAAP20 stabilizes FA complementation group A (FANCA) and participates in interstrand cross-link repair. Proc Natl Acad Sci U S A. 2012;109:4491–6.
30. Jones MJK, Huang TT. The fanconi anemia pathway in replication stress and DNA crosslink repair. Cell Mol Life Sci. 2012;69:3963–74.
31. Titus TA, Yan Y-L, Wilson C, Starks AM, Frohnmayer JD, Bremiller RA, et al. The fanconi anemia/BRCA gene network in zebrafish: embryonic expression and comparative genomics. Mutat Res. 2009;668:117–32.
32. Zhang T, Wilson AF, Mahmood Ali A, Namekawa SH, Andreassen PR, Ruhikanta Meetei A, et al. Loss of Faap20 causes hematopoietic stem and progenitor cell depletion in mice under genotoxic stress. Stem Cells. 2015;33:2320–30.
33. Harding MM, Anderberg PI, Haymet AD. Antifreeze' glycoproteins from polar fish. Eur J Biochem. 2003;270:1381–92.
34. Rozelle AL, Cheun Y, Vilas CK, Koag M-C, Lee S. DNA interstrand cross-links induced by the major oxidative adenine lesion 7,8-dihydro-8-oxoadenine. Nat Commun. 2021;12:1897.
35. Oyarbide U, Topczewski J, Corey SJ. Peering through zebrafish to understand inherited bone marrow failure syndromes. Haematologica. 2019;104:13–24.
36. Botthof JG, Bielczyk-Maczyńska E, Ferreira L, Cvejic A. Loss of the homologous recombination gene rad51 leads to fanconi anemia-like symptoms in zebrafish. Proc Natl Acad Sci U S A. 2017;114:E4452–61.
37. Cassini A, Favero M, Albergoni V. Comparative studies of antioxidant enzymes in red-blooded and white-blooded Antarctic teleost fish. Pagothenia Bernacchii and Chionodraco hamatus. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol. 1993;106:333–6.
38. Witas H, Gabryelak T, Matkovics B. Comparative studies on superoxide dismutase and catalase activities in livers of fish and other Antarctic vertebrates. Comp Biochem Physiol C Comp Pharmacol Toxicol. 1984;77:409–11.
39. Mueller IA, Grim JM, Beers JM, Crockett EL, O'Brien KM. Inter-relationship between mitochondrial function and susceptibility to oxidative stress in red- and white-blooded Antarctic notothenioid fishes. J Exp Biol. 2011;214:3732–41.
40. Li N, Chen H, Wang JDNA. Acta Biochim Biophys Sin (Shanghai). 2022;54:847–57.
41. Gordon A. Fastx-toolkit. FASTQ/A Short-Reads Preprocessing Tool. 2010;5.
42. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. 2024;25:107.
43. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36:2253–5.
44. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
45. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19:460.
46. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Preprint arXiv:13033997; 2013.
47. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinf. 2004;5:11–4.
48. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:5825–9.
49. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14.
50. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
51. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;34:1812–9.

52. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: a modern graphical user interface for custom BLAST databases. Mol Biol Evol. 2019;36:2922–4.
53. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32(suppl2):W20–5.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.