Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

Article

Fitting together the evolutionary puzzle pieces of the Immunoglobulin T gene from Antarctic fishes

Alessia Ametrano[1,2] Marco Gerdol[3], Maria Vitale[1,4], Samuele Greco[3], Umberto Oreste[1], Maria Rosaria Coscia[1,*]

[1] Institute of Biochemistry and Cell Biology - National Research Council of Italy, 80131 Naples, Italy; alessia.ametrano@ibbc.cnr.it (A.A.); umberto.oreste@ibbc.cnr.it (U.O.); mariarosaria.coscia@ibbc.cnr.it (M.R.C.)

[2] Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, University of Campania Luigi Vanvitelli, 81100 Caserta, Italy; alessia.ametrano@unicampania.it (A.A.)

[3] Department of Life Sciences, University of Trieste, 34127 Trieste, Italy; mgerdol@units.it (M.G.); SAMUELE.GRECO@phd.units.it (S.G.)

[1,4] Department of Molecular Medicine and Medical biotechnology, University of Naples Federico II, 80131 Naples, Italy (Present address); vitalema@ceinge.unina.it (M.V.)

* Correspondence: mariarosaria.coscia@ibbc.cnr.it; Tel.: +0039 081 6132556 (M.R.C.)

Abstract: Cryonotothenioidea is the main group of fishes that thrive in the extremely cold Antarctic environment, thanks to the acquisition of peculiar morphological, physiological and molecular adaptations. We have previously disclosed that IgM, the main immunoglobulin isotype in teleosts, display typical cold-adapted features. Recently, we have analyzed the gene encoding the heavy chain constant region (CH) of the IgT isotype from the Antarctic teleost *Trematomus bernacchii* (family Nototheniidae), characterized by the near-complete deletion of the CH2 domain. Here, we aimed to track the loss of the CH2 domain along notothenioid phylogeny and to identify its ancestral origins. To this end, we obtained the *IgT* gene sequences from several species belonging to the Antarctic families Nototheniidae, Bathydraconidae and Artedidraconidae. All species display a CH2 remnant of variable size, encoded by a short *Cτ2* exon, which retains functional splicing sites and therefore is included in the mature transcript. We also considered representative species from the three non-Antarctic families: Eleginopsioidea (*Eleginops maclovinus*), Pseudaphritioidea (*Pseudaphritis urvillii*) and Bovichtidae (*Bovichtus diacanthus* and *Cottoperca gobio*). Even though only *E. maclovinus*, the sister taxa of Cryonotothenioidea, shared the partial loss of *Cτ2*, the other non-Antarctic notothenioid species displayed early molecular signatures of this event. These results shed light on the evolutionary path that underlies the origins of this remarkable gene structural modification.

Keywords: teleost fish; Notothenioidei; genome modifications; IgT; exonic remnant; immunoglobulin domain; Antarctic marine environment; molecular evolution.

1. Introduction

The peculiar architecture of the Immunoglobulin molecule (Ig) comprises two types of domains

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

with similar secondary and tertiary structures, but different amino acid sequences: the variable domains (V), which provide immunoglobulins the ability to recognize and bind foreign antigens, and the constant domains (C), which perform effector functions. Another important difference between C and V domains derives from their underlying genomic configuration. While C domains are always encoded by individual exons, the V domains of both Ig heavy (VH) and light chains (VL) are generated by a recombination activating gene -mediated rearrangement of multiple sets of *Variable (V)*, *Diversity (D)* and *Joining (J) gene* segments. Each Ig gene encodes both a membrane-bound heavy chain form, which is present on the B-cell surface as B cell receptor complex, and a secreted heavy chain, which can be mainly found in blood. Both forms are generated by an *alternative splicing* process that involves the 3′ exons but does not alter the rearrangement of *VDJ* gene segments.

Different Ig isotypes can be distinguished based on the properties of their heavy chain constant (CH) domains. Unlike mammals, which possess five Ig isotypes, teleosts only possess three different Ig isotypes, i.e. IgM, IgD and IgT. IgM, mainly found in the blood but also present in mucosal compartments, are the main players in systemic immune responses against a broad range of pathogens [1]. Although the IgD isotype emerged early in vertebrate evolution, its functional role in teleosts has remained unclear for long time. Previous data suggested that IgD participate in immune tolerance and in basophil-mediated immunity [2]. Most recently, Perdiguero et al. [3] have shown that secreted IgD interact with the gut microbiota, playing a relevant role in maintaining mucosal homeostasis in rainbow trout. In 2005 a new Ig isotype, named "IgT" (for Teleost) or IgZ (for Zebrafish), was discovered simultaneously in two teleost species, i.e. rainbow trout and zebrafish [4, 5]. Early studies had initially linked IgT to a specialized mucosal role in the gut [6]. However, further evidence have expanded this function to other tissues, which include the [7], the nasopharyngeal tract [8, 9], the gills [9] and the buccal cavity [10].

The growing body of molecular data obtained from other teleost species has progressively revealed that IgT is a particularly heterogeneous Ig isotype. First, the number of *IgT* genes per species can largely vary, from zero (e.g. channel catfish and medaka) to several paralogous copies (e.g. rainbow trout and seabass) [11, 12].

The teleost Ig heavy chain (*IgH) gene* locus organization had been originally described as similar to the mammalian "translocon" type, consisting of multiple sets of *V* gene segments located upstream of multiple *D* and *J* segments, followed by *Cμ* and *Cδ* exons, coding for the CH regions of the IgM and IgD isotypes, respectively. Later, a comparative analysis of the *IgH* locus in several teleost species overturned this concept, uncovering a novel and incredibly diversified genomic organization [13]. In most cases, another set of *Dτ -Jτ -Cτ* elements encoding the IgT heavy chain (τ) was found upstream of previously described *D-J-Cμ-Cδ* elements for IgM and IgD heavy chains. In this configuration, upstream *V* segments can be rearranged either to *DτJτCτ* for the IgT VH region, or to *DJCμ* for the same region of IgM.

The canonical structure of IgT, exemplified by zebrafish, comprises a CH region encoded by four *Cτ* exons, hereinafter referred to as *Cτ1*, *Cτ2*, *Cτ3* and *Cτ4*. However, remarkable variations have been observed in some teleosts. The pufferfish *IgT* heavy chain gene only comprises two *Cτ* exons, homologous to the zebrafish *Cτ1* and *Cτ4* exons, respectively [14]. The Nile tilapia and the common carp also display unusual IgT domain architectures: the former only possesses the first two CH domains [15], and the latter is characterized by the presence of at least three chimeric IgM/IgT

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

molecules, encoded by different gene copies [16, 17].

The first report of an *IgT* CH region gene in a teleost species living under extreme conditions dates back to 2015, when our group reported the unprecedented case of a nearly complete truncation of the CH2 domain in the Antarctic fish *Trematomus bernacchii* [18]. Two out of the three variants identified in this species, termed Long (TbeL) and Short (TbeS), respectively) were encoded by alleles characterized by a large deletion and only displayed a short remnant of the *Cτ2* exon, which was entirely skipped by alternative splicing in the third isoform, termed Shortest (TbeSts). Most Antarctic fishes belong to the Perciform suborder Notothenioidei (Cryonotothenioidea), which comprises five families (Nototheniidae, Harpagiferidae, Artedidraconidae, Bathydraconidae, Channichthyidae) and about 130 species of marine fishes found in the Southern Ocean, with a circum-Antarctic distribution, but also found in the more temperate coastal waters of the southern hemisphere [19]. Notothenioidei are among the most intensively studied lineages of marine fishes since they are a rare example of massive adaptive radiation driven by the same selective pressures (e.g. isolation and cooling) that may have led to the dramatic extinction of most fish fauna in the Southern Ocean [20-22]. The evolutionary success of Notothenioidei is marked by the acquisition of key adaptive features that enabled cold adaptation, such as the expression of antifreeze glycoproteins [23, 24]. At the same time, Notothenioidei lost other traits universally shared by non-Antarctic metazoans, such as the inducible heat shock response [25] and, in the family Channichthyidae, hemoglobin [26, 27].

Over the past 20 years, molecular and morphological studies allowed a revision of the phylogenetic relationships among notothenioid lineages, contributing to improve our knowledge about the adaptive radiation of these organisms [28-32]. Apart from the Antarctic Clade [33], three non-Antarctic lineages, distributed in proximity of the Southern Ocean, i.e. the southern regions of South America, around the Falkland Islands, Tristan da Cunha, New Zealand and south-eastern Australia, are currently recognized [20, 34]. While the first one, Bovichtidae, includes six species, the two other families are monotypic and therefore include a single species, i.e. *Pseudaphritis urvillii* (family Pseudaphritioidae) and *Eleginops maclovinus* (family Eleginopsioidea). More recently, much attention has been dedicated to these notothenioid taxa, due to their early divergence from the polar lineage, which occurred before the climatic and geographic isolation of Antarctica, the drastic reduction in water temperature, and prior to the morpho-physiological diversification of cryonotothenioid species [35]. The study of the evolutionary history of Bovichtidae, based on mitochondrial and nuclear DNA molecular phylogeny, as well as on morphological and meristic characters, has been an essential factor for clarifying the process that drove the diversification between Cryonotothenioidea and their non-Antarctic relatives. Indeed, the evolutionary radiation of the genus *Bovichtus* and of the closely related cryonotothenioid species are characterized by a similar timeline, and therefore the patterns of diversification and extinction observed in these two lineages are expected to closely match each other [35]. Moreover, the revised positioning of *E. maclovinus* as the sister lineage of Cryonotothenioidea (it was previously considered as closely related to the notothenid lineage *Dissostichus*) [30], provides another key information for the study of the evolution of these fishes. Taking into account the updated information about notothenioid phylogeny, the present work aims to extend the molecular analysis of IgT to the other notothenioid families and to solve the key question as to whether the features of the *T. bernacchii* IgT are unique to this species or shared by other Antarctic species. In order to track the evolutionary history of the *Cτ2*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

exon and to pinpoint the timing of its partial loss, we obtained and comparatively investigated the *IgT* sequences of representatives from each of the five Antarctic notothenioid families and the three non-Antarctic lineages of Bovichtidae (*Bovichtus diachantus* and *Cottoperca gobio*), Eleginopsioidea (*E. maclovinus*) and Pseudaphritioidae (*P. urvillii*). The findings reported here bring further insights into the molecular evolution of the *IgT* gene in Antarctic fishes, marking the loss of the *Cτ2* exon before the split between the Eleginopsioidea and Cryonotothenioidea lineages, and revealing early signatures of this event in the other early-branching non-Antarctic Notothenioidei.

2. Results

2.1 The IgT cDNA sequence of Eleginops maclovinus *provides new evidence about the origin of the loss of Cτ2 exon in Antarctic species*

Our investigations on the *IgT* cDNA sequences targeted several Antarctic species belonging to the Nototheniidae, Artetidraconidae, Bathydraconidae and Channichthyidae families (see the Materials and methods section for details), but also included the non-Antarctic species *E. maclovinus* due to its phylogenetic placement as a sister lineage of Cryonotothenioidea.

We obtained partial cDNA sequences, coding for the CH region of the IgT secreted form in all species analyzed, with the single exception of *E. maclovinus*, where the IgT membrane-bound form was obtained. The multiple sequence alignment highlighted the high conservation of the *Cτ1* and *Cτ4* exons in all species, as expected from previous publications (Figures 1 and S1). Mirroring the previously reported case of *T. bernacchii*, all Antarctic teleosts displayed a truncated *Cτ2* exon, whose size ranged from 24 to 51 nt (Figure 1, Table 1).

```
          [Cτ1 exon                                ][Cτ2 exon
Ema1      GTGACTCATTCTGGAACCTCAAAAAGTGTGGAAATTAAAAA------TACTCAA---TGT---------CCAAAGATAGC     80
Ema2      GTGACTCATTCTGGAACCTCAAAAAGTGTGGAAATTAAAAAACAAGATACTCAA---TGT---------CCAAAGATAGC
Tb30.3    GTGACTCACCCTGGAGGACAAAAAAGTTTGAGCATTAAAAAGGTTAA------------------AGCACCTTCATGTTC
Tb30.7    GTGACTCACCCTGGAGGACAAAAAAGTTTGAGCATTAAAAAGGTTATTCAACTTTTAGGTGCTTTAGCACCTTCATGTCC
Tb30.8    GTGACTCACCCTGGAGGACAAAAAAGTTTGAGCATTAAAA----------------------------------------
Nco       GTGGCTCACCCTGGAGGACCAAAAAGTTTGAACATTAAAAAN-------------------------CCTTCATGTCC
Ggi1      GTGACTCACTCTGGAGGATCAACAAGTTTGAACATTAAAAAT-----------------------------CCTTCATGTAC
Hve1      GTGACTCACCCTGGAGGACCAACAAGTTTGAGCATTAAAAAGG------------CAATTCAACAAGCACCTTCATGTCC
Hve2      GTGACTCACCCTGGAGGACCAACAAGTTTGAGCATTAAAAAGG------------TAATTCAACAAGCACCTTCATGTCC
Gac1      GTGACTCACCCTGGAGGACCAACAAGTTTGAACATTAAAAAGG------------CAATTCAACAAGCACCTTCATGTAC
          ***.****: *****:::  ** ****:**: :*******

                  ][Cτ3 exon
Ema1      ---ATCGAAAGCTCCCAT---------------CACAATAGAACTGAAAAAATCAAGTCCCAAAGAAGTGTTCAATAACA    160
Ema2      ---ATCGAAAGCTCCCAT---------------CGCAATAGAACTGAAAAAATCAAGTCCCAAAGAAGTGTTCAATAACA
Tb30.3    AGAGACTAACGCTACCAT---------------CACGATGAAACTGAACAAATCAAGTCCCAAAGAAATGTTCAACAACA
Tb30.7    AGAGACTAACGCTCCCAT---------------CACGATGAAACTGAACAAATCAAGTCCCAAAGAAATGTTCAACAACA
Tb30.8    -----------CTACCAT---------------CACGATGAAACTGAACAAATCAAGTCCCAAAGAAATGTTCAACAACA
Nco       AAAGACTAAAGCTCCCAT---------------CACGATGAAACTGAACAAATCAAGTCCCAAAGAAATGTTCAACAACA
Ggi1      AGAGACTAAAGCTCCCAT---------------CACGATGGAACTGAACAAATCAAGTCCCAAAGAAATGTACACCAACA
Hve1      AAAGACTAAAGCTCCCGTTTGTGTAGCTCCCGTCACGATGAAACTGATCATA---AGTTCCAAAGAAATGTTCAACAACA
Hve2      AAAGACTAAAGCTCCCGT---------------CACGATGAAACTGATCAAA---AGTTCCAAAGAAATGTTCAACAACA
Gac1      AGAGACTAAAGCTCCCAT---------------CACGATGGAACTGATCAGA---AGTCCCAAAGAAATGTTCAACAGCA
                **:**:*                  *.*:**: ****** :* *   ***:********:***.**.:*.**
```

Figure 1. Multiple alignment of partial cDNA sequences spanning the 3′ end of the *Cτ1* exon, and the 5′ end of the *Cτ3* exon. Nucleotide sequences are representatives of different cDNA clones coding for different variants (wherever found), derived from the non-Antarctic species *E. maclovinus* (clones Ema1and Ema2), and from the Antarctic species *G. gibberifrons* (clone Ggi1), *H. velifer* (clones Hve1 and Hve2, both coding for the same *Cτ2* remnant but differing in *Cτ3* length), *G. acuticeps* (clone Gac1). Three cDNA transcripts (Tb30.3, Tb30.7, Tb30.8) encoding the TbeS, TbeL and TbeSts variants *in T. bernacchii* and one cDNA sequence from *N. coriiceps* (Nco), obtained in a previous work [18], have been added for comparison. *Cτ* exon boundaries are reported above the alignment. Gaps are indicated by dashes. Below the alignment, identical nucleotides are marked with an asterisk, positions where only one sequence shows a different nucleotide are marked with a dot, positions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

differing in two nucleotides are marked with a colon. The duplicated 9-nt sequence at the beginning of the *Cτ3* exon of the Hve2 transcript is underlined. Since several sequences varying in length at the 5′ and/or at 3′ were obtained from each species, only the region of each representative sequence that aligned over the same length has been shown here. Full alignments are provided in Figure S1.

The two variants cloned in *E. maclovinus* displayed a partially deleted *Cτ2* exon (30 or 36 nt long in Ema1 and Ema2, respectively; see Table 1), revealing a *Cτ2* exon structure similar to Antarctic Notothenioidei. On the other hand, we have previously shown that *B. diacanthus*, a non-Antarctic species more distantly related with Cryonotothenioidea, retains the entire *Cτ2* exon (285 nt) [35] (Table 1).

Table 1. Length of the IgT CH2 domain in non-Antarctic (in red) and Antarctic (in blue) notothenioid fish. The minimum and maximum length values observed among all species are underlined.

Species	Nucleotides	Amino acids
B. diacanthus	285	95
C. gobio	282	94
P. urvillii	282	94
E. maclovinus Ema1	30	10
E. maclovinus Ema2	36	12
T. bernacchii TbeS	33	11
T. bernacchii TbeL	51	17
G. gibberifrons Ggi1	24	8
N. coriiceps Nco1	24	8
H. velifer * Hve1	39	13
G. acuticeps Gac1	39	13

* two *Cτ3* variants

Although Antarctic species and *E. maclovinus* shared the peculiar structure of *Cτ2*, their sequences differed due to several characteristic codon indels. In detail, the *Cτ3* exon of all Antarctic species lacked one codon at positions 502, 598 and 617 (the latter was not missing in *N. coriiceps*), and two consecutive codons at position 562. On the other hand, a single codon insertion was found at position 526 in the Antarctic lineage (Figure S1). Interestingly, all Cryonotothenioidea also presented four additional codons in the highly conserved *Cτ4* exon at positions 715, 718, 778 and 784 (Figure S1).

2.2 Genomic analysis of the IgT gene in Antarctic species and Eleginops maclovinus*: putting together the first pieces of the puzzle*

The next step in the exploration of the molecular mechanisms behind the partial loss of the *Cτ2* exon in Antarctic *IgT* gene was the extension of our analyses to the neighboring genomic regions (i.e. the *Cτ1-Cτ2* and the *Cτ2-Cτ3* introns). Based on the data reported in the previous paragraph, we used *E. maclovinus* as a reference for comparative analyses (Figures 2 and S2).

The two *IgT* genomic variants found in this sub-Antarctic species (Ema1 and Ema3) were characterized by a *Cτ2* exon of variable size (30/36 nt) and matched those identified at the cDNA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

level. The *Cτ1-Cτ2* intron of the isoform with the shortest *Cτ2* exon (Ema1) displayed the insertion of a CAAACA sequence immediately before the splicing acceptor site of the *Cτ2* exon (Figure 2). On the other hand, the *Cτ2-Cτ3* intron had identical length in both isoforms.

The two introns showed a significant size variation among Antarctic species. The *Cτ1-Cτ2* intron ranged from 212 nt in *N. coriiceps* to 318 nt in *T. bernacchii*, which contained two 46-nt long repeated regions (Table 2, Figure 2). *N. coriiceps* and *G. gibberifrons* displayed an insertion of additional 15 nt at the 3′ end of the intron, which was paired with the presence of the shortest *Cτ2* exon out of all the species analyzed (24 nt, 8 aa). The size and structure of *Cτ2-Cτ3* intron was much more conserved across Antarctic species, ranging from 115 nt in *G. gibberifrons* to 128 nt in *H. velifer*.

The short length of the *G. gibberifrons* intron was due to a 9-nt long deletion, which matched the position of a 7-nt indel evidenced by the multiple sequence alignment in *E. maclovinus* (Figure 2).

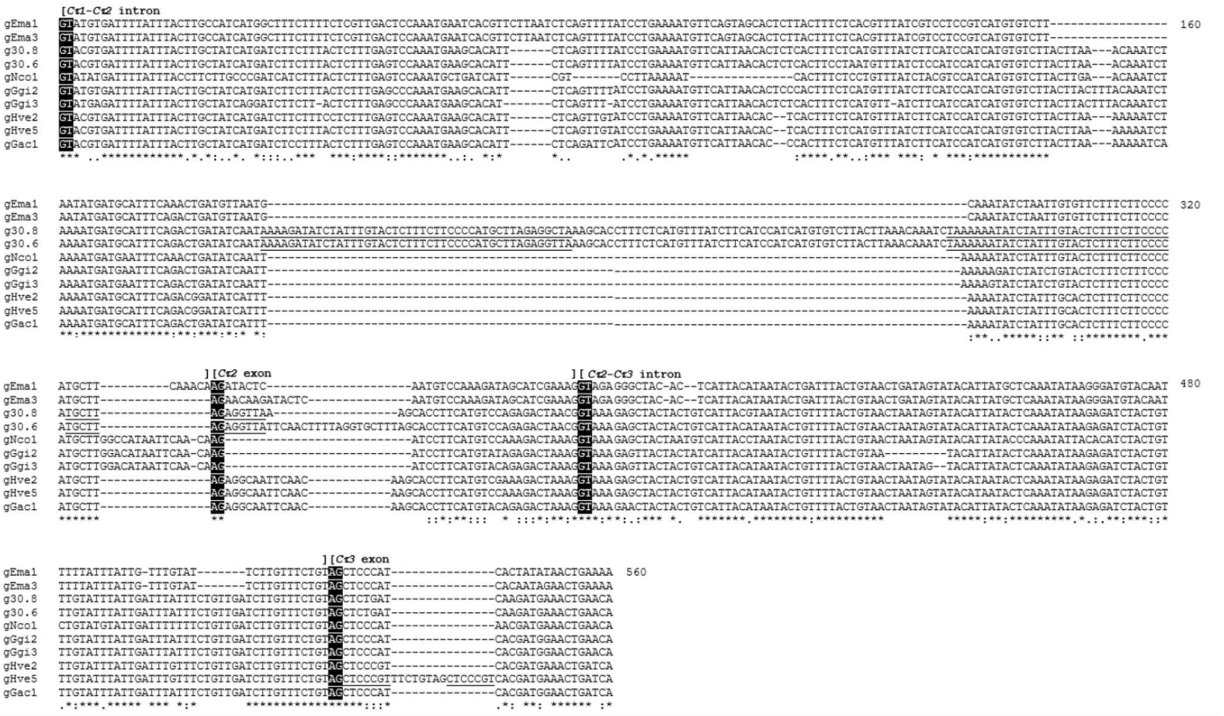

Figure 2. Multiple alignment of partial genomic sequences encoding the notothenioid IgT CH region. Nucleotide sequences are representatives of different genomic clones coding for different variants (wherever found) from the non-Antarctic species *E. maclovinus* (clones gEma1 and gEma3), and from the Antarctic species *N. coriiceps* (clone gNco1), *G. gibberifrons* (clones gGgi2 and gGgi3/pseudogene), *H. velifer* (clones gHve2 and gHve5, both coding for the same *Cτ2* remnant but differing in *Cτ3* length), and *G. acuticeps* (clone gGac1). The *T. bernacchii* clones g30.8 and g30.6, coding for the respective TbeS and TbeL variants, previously determined [18], have been added for comparison. The exon-intron boundaries are reported above the alignment. Donor and acceptor splicing sites are shaded in black. Gaps are indicated by dashes. Below the alignment, identical nucleotides are marked with an asterisk, positions where only one sequence shows a different nucleotide are marked with a dot, positions differing in two nucleotides are marked with a colon. The two duplicated 46-bp long sequences in the *T. bernacchii Cτ1-Cτ2* intron and the one duplicate 9-bp long sequence in the gHve5 *Cτ3* exon are underlined. Since several sequences varying in length at the 5′ and/or at 3′ ends were obtained for each

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

species, only the region of each sequence that aligned over the same length has been shown. Full alignments are provided in Figure S2.

2.3 The genomic analysis of the IgT gene from non-Antarctic species sheds light on the stepwise process that led to Cτ2 exon loss

To elucidate the processes by which the *Cτ2* exon was progressively lost along the evolution of Notothenioidei, we extended our analyses to the *IgT*-encoding genomic sequences of three additional key non-Antarctic notothenioid species (i.e. *B. diacanthus*, *P. urvillii*, *C. gobio*, see the Materials and Methods section).

The alignment of the partial genomic sequences (Figures 3 and S3) hinted that the evolutionary process that led to the partial loss of the *Cτ2* exon might have already started before the split between the Eleginopsioidea and Cryonotothenioidea lineages, as early as in the late Cretaceous. Indeed, the gene of *P. urvillii* (Pseudaphritioidea) was characterized by early molecular signatures of erosion shared with the Antarctic species or in their sister taxa *E. maclovinus*. In particular, a few informative small deletions, matching the position of similar gaps in the sequence of *E. maclovinus*, were found in the *Cτ1-Cτ2* intron, but not in the *Cτ2-Cτ3* intron, which only included a few indels shared by all species, regardless of their position in the phylogeny of Notothenioidei (Figure 3). Unlike *E. maclovinus*, *P. urvillii* retained a complete *Cτ2* exon, which only lacked the first three nt at its 5′ end.

The sequences of the two species belonging to Bovichtidae, the most early-branch of the Notothenioidei lineage, also retained a complete *Cτ2* exon (e.g. 282 nt – 94 aa in *C. gobio* and 285 nt – 95 aa in *B. diacanthus*, see Table 1) and displayed much larger introns than *E. maclovinus* (Table 2).

Table 2. Length of the IgT *Cτ1-Cτ2* and *Cτ2-Cτ3* introns in non-Antarctic (in red) and in Antarctic (in blue) notothenioid fish. The minimum and maximum length values observed across all species are underlined.

Species	*Cτ1-Cτ2* intron (nt)	*Cτ2-Cτ3* intron (nt)
B. diacanthus	242	144
C. gobio	224	147
P. urvillii	222	142
E. maclovinus	206	108
B. diacanthus	242	144
T. bernacchii	318	122
G. gibberifrons	237	115
N. coriiceps	212	126
H. velifer	216	128
G. acuticeps	216	126
T. bernacchii	318	122

However, the *Cτ1-Cτ2* intron of *C. gobio* was also characterized by the presence of a few indels in similar positions to those observed in *P. urvillii* and *E. maclovinus* (Figure 3).

The multiple sequence alignment of the *Cτ2* exon of non-Antarctic Notothenioidei interestingly revealed that, despite the nearly complete deletion of the *Cτ2* exon, *E. maclovinus* retained the six nucleotides found at the 3′end of its remnant nearly intact (with the single synonymous substitution

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

TCG/TCT, see position 537 in Figure 3). This observation is in line with the high conservation of the canonical donor and acceptor splicing sites of the remnant *Cτ2* exon observed in all species (Figure 3).

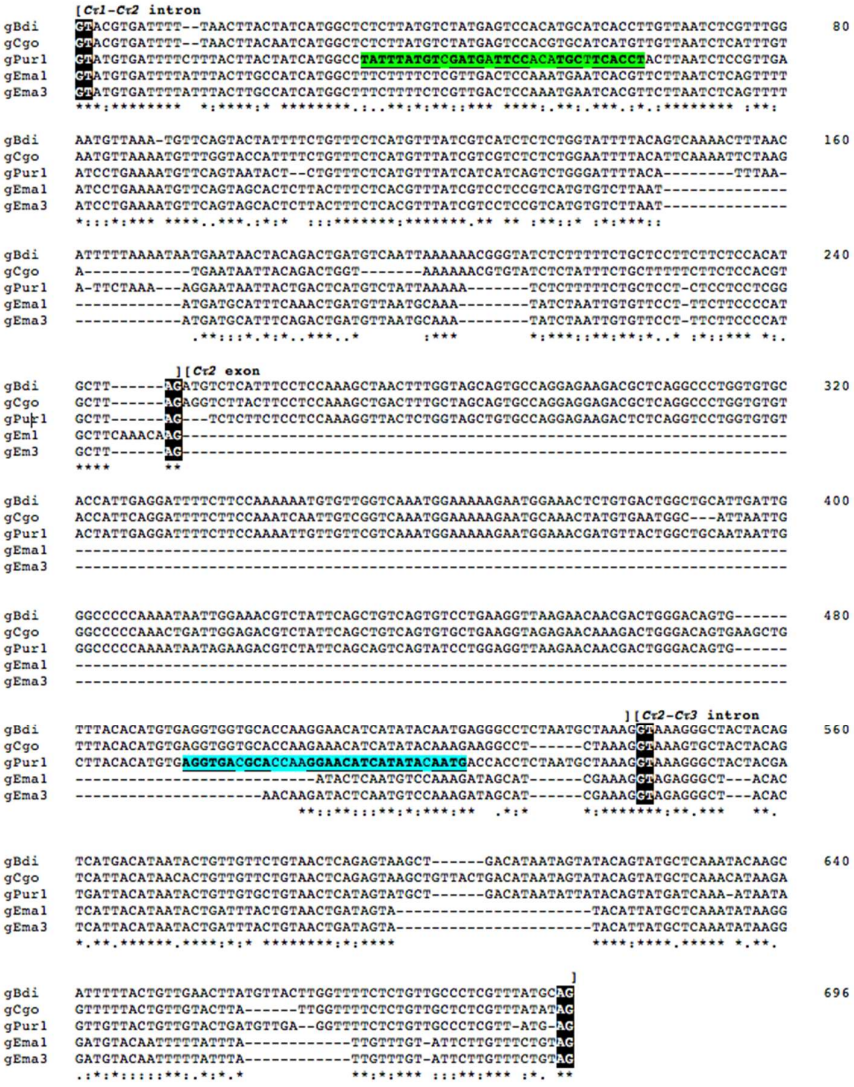

Figure 3. Multiple alignment of partial genomic sequences encoding the IgT CH region from the four non-Antarctic notothenioid species investigated. Nucleotide sequences are representatives of different genomic clones coding for different variants (wherever found) from the species *B. diacanthus* (clone gBdi), obtained in a previous work [18], *C. gobio* (gCgo), *P. urvillii* (clone gPur1) and *E. maclovinus* (clones gEma1 and gEma3). The exon-intron boundaries are reported above the alignment. Donor and acceptor splicing sites are shaded in black. Gaps are indicated by dashes. Below the alignment, identical nucleotides are marked with an asterisk, positions where only one sequence shows a different nucleotide are marked with a dot, positions differing in two nucleotides are marked with a colon. Twenty-six nucleotides (underlined, in bold) of a 32-nt long region (shaded in green) at the beginning of the *P. urvillii Cτ1-Cτ2* intron were present within the 32-nt long complementary reverse motif (shaded in cyan) in the respective *Cτ2* exon. Since several sequences varying in length at the 5′ and/or at 3′ ends were obtained for each species, only the region of each sequence that aligned over the same length has been shown. Full alignments are provided in Figure S3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

2.4 The absence of repeats in the introns of the IgT gene rules out the involvement of transposable elements in the Cτ2 exon loss

As a further step in the investigation of the molecular mechanisms that might have led to the shortening of the *Cτ2* exon, we investigated whether any repeated element could be identified in the *Cτ1-Cτ2* and *Cτ2-Cτ3* introns of the IgT gene from Antarctic notothenioids and in their sister lineage Eleginopsioidea. The activity of transposable elements (TEs) is well known to be associated with accelerated mutation, disrupting exons [36], altering splicing patterns [37], and shuffling the position of entire exons or of their parts by moving them to different genomic locations [38]. Hence, the presence of repeats could be indicative of the presence of active TEs, which may have possibly accelerated the evolution of the *Cτ2* region. Our analyses revealed that neither the *Cτ1-Cτ2* nor the *Cτ2-Cτ3* introns contained traces of repeats in any of the species analyzed in this study.

2.5 The IgT gene sequence phylogeny is consistent with the phylogenetic relationships among Notothenioidei

In line with the observations provided above, we found that the molecular evolution of the *Cτ1-Cτ2* and *Cτ2-Cτ3* introns closely followed the currently accepted phylogenetic relationship among Notothenioidei. In detail, all the sequences from Cryonotothenioidea were placed in a highly supported monophyletic clade (100% posterior probability, Figure 4).

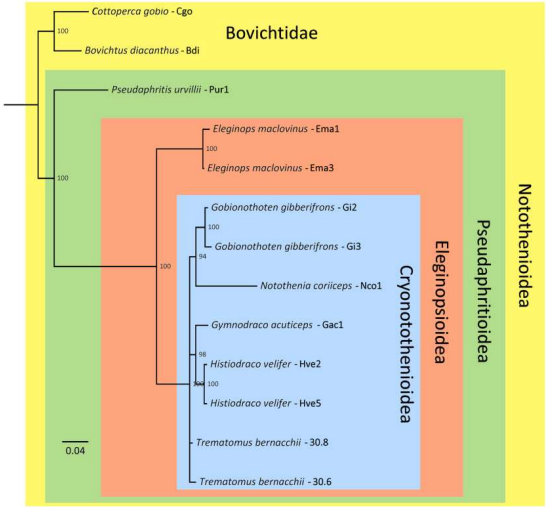

Figure 4. Bayesian phylogeny of notothenioid IgTs, based on the concatenated multiple sequence alignment of the *Cτ1-Cτ2* and *Cτ2-Cτ3* introns. Phylogeny was built with two parallel MCMC analyses, run for 1 million generations. Nodes supported by posterior probability values lower than 50% were collapsed.

As the sequence divergence between the different Antarctic species, which have been subject to a fast evolutionary radiation [39], was minimal, the topology of the Cryonotothenioidea subtree was characterized by very short branches. However, the two variants found in *G. gibberifrons* and *H. velifer* were closely related (100% posterior probability), suggesting that both have been originated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

by species-specific gene duplications (the most likely hypothesis in *G. gibberifrons*, since one of the two variants is pseudogenic) or that they represent allelic variants.

As expected, the sequences from the Bovichtidae *C. gobio* and *B. diacanthus* were placed as outgroups in a monophyletic clade at the base of the notothenioid IgT tree, whereas the sequences from *P. urvillii* and *E. maclovinus* occupied, with high statistical support (100% posterior probability in both cases), intermediate positions. These were well consistent with the recently proposed position of Pseudaphritioidea and Eleginopsioidea [35]. Curiously, the two variants from *E. maclovinus* shared closer homology with the Antarctic species than with *P. urvillii* and Bovichtidae, confirming the high relevance of this key species for the investigation of IgT evolution in Notothenioidei and further supporting the observation of shared indels, which may indicate a process of progressive loss of the *Cτ1-Cτ2* intron.

Overall, the structure of this intron can be summarized by the presence of 5 distinct conserved sequence motifs (Figure 5). Starting from the 5′ end, the first 15 nt-long motif (named motif 5 in Figure 5) was most likely involved in the recognition of the *Cτ1* donor splicing site. This was followed by the 42 nt-long motif 2, which was the least conserved among those identified and ancestrally duplicated at the 3′end of the intron, where it may be involved in the *Cτ2* splicing acceptor site. The subsequent 29 nt-long motif 3 was found in all species, except *B. diachantus* and *N. coriiceps*, whereas the highly conserved motif 1 (50 nt-long) was found in all species and corresponds to the previously mentioned duplicated region found in *T. bernacchii* (Figure 5). The last motif identified was the highly conserved 29 nt-long motif 4, which preceded the second repeat of motif 2. Curiously, the long *Cτ1-Cτ2* intron of *T. bernacchii* was characterized by the presence of an additional copy of motif 1 and motif 2 at its 3′ end.

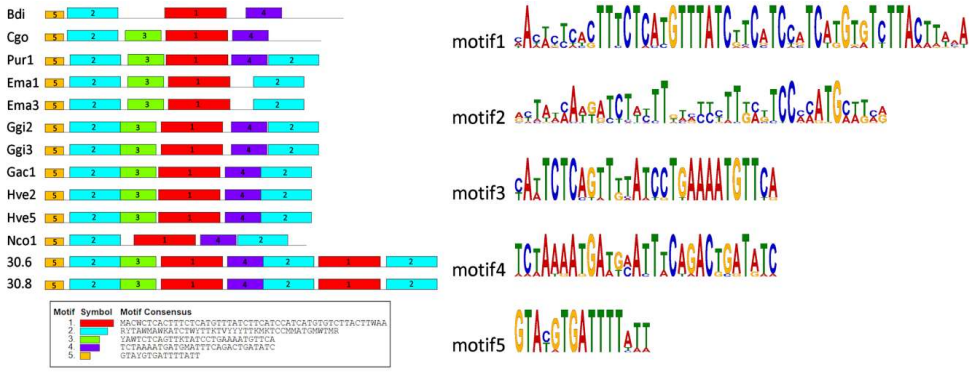

Figure 5. Significantly enriched sequence motifs identified by MEME in the *Cτ1-Cτ2* intron structure, exemplified by colored boxes. The degree of nucleotide conservation of each of the five motifs identified is reported as a sequence logo. Sequence names design different genomic clones from different species, as follows: *B. diacanthus* (gBdi), *C. gobio* (gCgo), *P. urvillii* (gPur1), *E. maclovinus* (gEma1 and gEma3), *G. gibberifrons* (gGgi2 and the pseudogene gGgi3), *G. acuticeps* (Gac1), *H. velifer* (gHve2 and gHve5), *N. coriiceps* (gNco1), and *T. bernacchii* (g30.6 and g30.8), representing the previously described TbeL and TbeS variants [18].

The analysis of the nucleotide composition of the *Cτ1-Cτ2* intron did not reveal any significant bias in Antarctic species (Figures 6 and S6), but at the same time it revealed an interesting trend in the *Cτ2-Cτ3* intron. Indeed, in line with the placement of *E. maclovinus* as a sister group

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

Cryonotothenioidea [30], the intron of this species had an AT content similar to the Antarctic species, which all showed a similar AT content (>70%) regardless of their length, and significantly higher than the other three non-Antarctic species (see Figure 6).

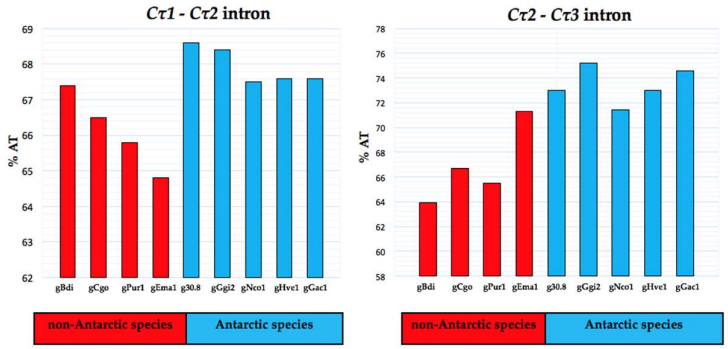

Figure 6. Distribution of AT content in the *Cτ1-Cτ2* and *Cτ2-Cτ3* introns. The representative genomic clones encoding the *Cτ* region are from the non-Antarctic species *B. diacanthus* (clone gBdi), *C. gobio* (gCgo), *P. urvillii* (clone gPur1), *E. maclovinus* (clone gEma1), and from the Antarctic species *T. bernacchii* (clone g30.8), *G. gibberifrons* (clone gGgi2), *N. coriiceps* (clone gNco1), *H. velifer* (clone gHve2), and *G. acuticeps* (clone gGac1).

2.6 Amino acid multiple sequence alignment sums up the main features of the notothenioid IgT CH domains

Consistent with the previously reported analysis of the *IgT* cDNA and genomic sequences (Figures 1 and 2, Table 1), the multiple sequence alignment of the deduced amino acid sequences highlighted the high conservation of the first, third and fourth CH domains (Figure 7). On the contrary, the size of the CH2 domain varied significantly from species to species, ranging from 8 to 95 aa, depending on its partial deletion (in all Antarctic species plus *E. maclovinus*) or full retention (in the most basal notothenioid lineages) (Table 1).

All the complete constant domains were characterized by the presence of two constitutive cysteines and one tryptophan, labeled as Cys 23, Cys 104, and Trp 41, according to the IMGT unique numbering (http://www.imgt.org). These residues are required to allow the correct folding of immunoglobulin domains, together with the hydrophobic residue (in this case a valine) typically found at the conserved position 89. The remnant portion of the CH2 domain just maintained the second of the two aforementioned canonical cysteines, whereas an additional conserved cysteine residue, known to be involved in the formation of an interchain disulfide bridge with the IgL chain, was found in CH1. While the presence of other cysteine residues in addition to those reported above is quite uncommon in temperate species, we observed several such cases in both Antarctic and non-Antarctic Notothenioidei (Figure 7). In detail, one extra cysteine was found in the N-terminal end of the CH3 domain of one of the two isoforms of *H. velifer*, precisely between the two APV repeats, and a second one was found at the C-terminal end of CH3 of *G. acuticeps.* The two non-Antarctic species *B. diacanthus* and *E. maclovinus* also showed extra cysteines, in CH2 and in the extracellular membrane proximal region, respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

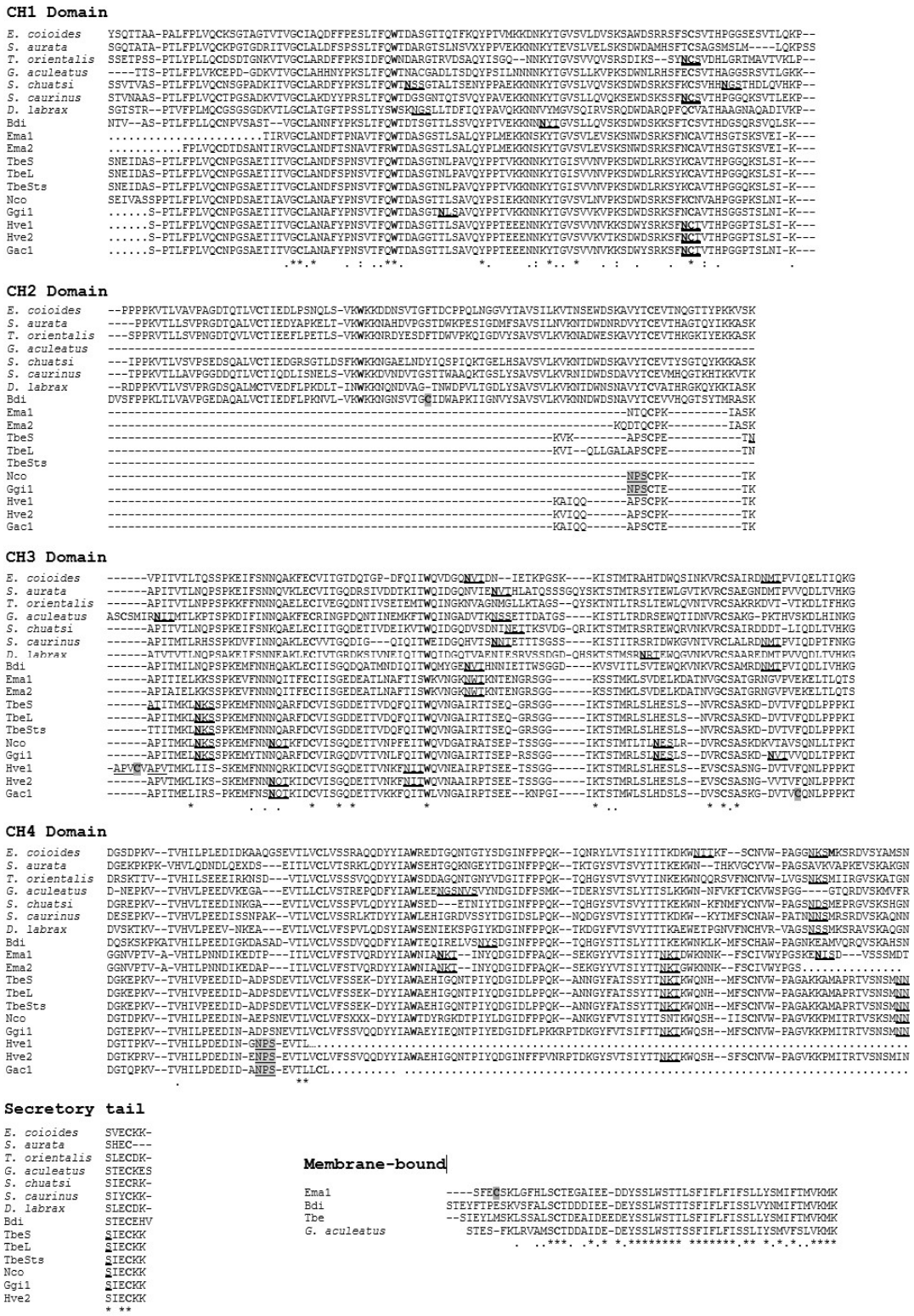

Figure 7. Multiple alignment of deduced amino acid sequences of the IgT CH domains from temperate, Antarctic and non-Antarctic species. Temperate Perciform species: *Epinephelus coioides*, orange-spotted grouper (GU182366), *Sparus aurata*, seabream (KX599200), *Thunnus orientalis*, Pacific bluefin tuna (KF713336), *Gasterosteus aculeatus*, three-spined stickleback [40], *Siniperca chuatsi*, mandarin fish (DQ016660), *Sebastes caurinus*, copper rockfish (GE798008), *Dicentrarchus labrax*, European seabass (KM410929); non-Antarctic

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

species: *B. diacanthus* (Bdi, KP876590), *E. maclovinus* (Ema1and Ema); Antarctic species: *T. bernacchii* (TbeS, TbeL and TbeSts variants), *N. coriiceps* (Nco, KP876589), *G. gibberifrons* (Ggi1), *H. velifer* (Hve1 and Hve2, both coding for same CH2 but differing in the CH3 domain length) and *G. acuticeps* (Gac1). Gaps are indicated by dashes. Identical amino acids are marked with an asterisk, conservative substitutions with colon, and semiconservative substitutions with a dot. Canonical cysteine and tryptophan residues are marked in bold. Extra cysteines are marked in bold and shaded in dark gray. The putative *N*-glycosylation sites are underlined. Asparagines predicted to be glycosylated are shown in red. Sequons containing a proline are in bold, underlined and shaded in gray. The duplicated 3-aa sequence APV at the beginning of the CH3 domain of the Hve1 variant is underlined. The membrane-bound form obtained only for *E. maclovinus* (Ema1) has been aligned with that of *B. diacanthus* (Bdi) and *T. bernacchii* (Tbe), previously obtained [18], and with that of *G. aculeatus* [40], the closest temperate species to the notothenioid species.

We investigated whether the IgT sequences of Antactic species were associated with the presence of conserved motifs that could be considered as possible "cold hallmarks", due to their absence in non-Antarctic Notothenioidei and temperate Perciformes, combined with signatures of purifying selection. Although a few short motifs conserved in Antarctic species and progressively fading out in their sister lineages were found (e.g. VV[NK]V in CH1, [IF]DCVI[SR] and TSTMXL[ST] in CH3) (Figure 7), none of these contained sites subject to purifying selection (Figure 8).

We could detect, however, a few sites subject to significant purifying selection in the two exons flanking the *Cτ2* exonic remnant. In detail, four and one negatively selected sites were detected in the *Cτ1* and *Cτ3* exons, respectively. The former exon included codons encoding highly conserved Leu, Asn and a Val residues (plus a Glu/Val residue found at the N-terminus of the domain, but just observed in three sequences), whereas the latter included a single codon encoding a Phe/Ile residue (Figure 8). Interestingly, in spite of its short length, two of the five positively selected sites detected by our analysis were located within CH2. The first of such sites was placed in a region subject to indel in a few species (e.g. *G. gibberifons* and *N. coriiceps*), close to the N-terminal end of the CH2 remnant. The second hypervariable site was found buried at the center of the domain remnant, adjacent to the aforementioned conserved cysteine. The three other sites evolving under positive selection were found at >20 residues of distance from the boundaries of the CH2 remnant. In detail, the single positively selected codon found in CH1 (encoding either Pro, Thr or Lys) was found near the negatively selected Val-encoding codon mentioned above (Figure 8). The two hypervariable sites of CH3 were close to each other and located approximately at the center of the domain (Figure 8).

We evaluated the amino acid composition of the CH2 domain in both Antarctic species and non-Antarctic species, comparing it with temperate teleosts. The most represented amino acid residues were Ala and Lys in all species, with Val, Thr, Ser and Leu being more abundant in temperate species and Ala and Pro being more frequently found in Antarctic species.

Since the attached glycans can influence solubility, structural stability, and biological function of Ig molecules, we assessed whether Asn-X-Ser/Thr sequons, i.e. possible glycosylation sites, were present in IgT sequences. Antarctic species showed a high number of sequons (Figures 7 and S4), potentially subject to a higher degree of glycosylation (65% in all Cryonotothenioidea, 100% in the family Nototheniidae) compared with both non-Antarctic (40%) and temperate species (47%), suggesting a significant role of glycosylation for cold-adapted proteins.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

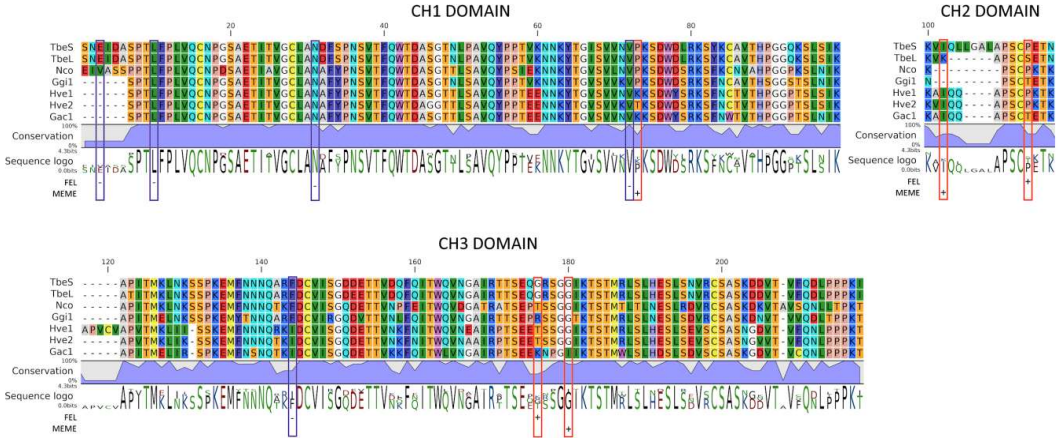

Figure 8. Sites predicted to evolve under purifying and diversifying selection in the regions surrounding the remnant CH2 domain in Cryonotothenioidea. Sites with statistically significant ω values are indicated by blue (negative selection) and red (positive selection) boxes, as detected by FEL and MEME.

Unlike temperate and non-Antarctic species, where sequons are evenly distributed along CH1 and CH3, in Cryonotothenioidea most glycosylation sites were found in CH3 (Figure S5).

Although *N. coriiceps* and *G. gibberifrons* presented an Asn-Pro-Ser sequon in the CH2 remnant (also found in the CH4 domain of Artedidraconidae and Bathydraconidae), this is unlikely to be a real glycosylation site due to the proximity between a Pro and an Asn residue, which is expected to make the Asn residue inaccessible [41].

3. Discussion

The immunoglobulins of Antarctic fishes have been fascinating us since the early discovery of unforeseen features of IgM from Cryonotothenioidea [42-44]. For several years, our studies have been mostly focused on IgM, an ancient Ig isotype that first appeared in jawed fish along with the emergence of an adaptive immune system [45]. However, we recently moved our attention to the study of the heavy chain gene of IgT, a fish-specific Ig isotype, whose discovery revealed the origins of the most ancient Ig specialized in mucosal immune response. We disclosed that the gene of Antarctic species *T. bernacchii*, unlike the early-branching non-Antarctic notothenioid species *B. diacanthus* and most species living in temperate environments, displayed an unusual truncated *Cτ2* exon, which only encoded a short remnant of the CH2 domain [18]. This finding was the starting point of the present work, which extends our molecular investigations to other Antarctic species and to the early-branching non-Antartic notothenioid lineages, in the attempt to pinpoint the origins of this partial exon loss event.

The groundwork for placing down the first piece of this evolutionary puzzle was provided by the observation that the *Cτ2* exon was nearly completely missing also in *E. maclovinus*, a non-Antarctic species which belongs to Eleginopsioidea, the sister group of Cryonotothenioidea. This allowed us to move backwards through the phylogeny of Notothenioidei, characterizing the IgT sequence of *C. gobio*, belonging to the most basal group of non-Antactic Notothenioidei, i.e.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

Bovichtidae. This species showed a nearly complete *Cτ2* exon, except for a small deletion located at its 3'end of the exon, which matched a similar indel carried by *E. maclovinus.* Moreover, *C. gobio* also displayed a shorter *Cτ1-Cτ2* intron than its close relative *B. diacanthus*, with a few deletions shared with *E. maclovinus*.

The cornerpiece of the puzzle was provided by the analysis of the IgT sequence of *P. urvillii*, belonging to the monotypic family Pseudaphritioidea, which covers an intermediate position in notothenioid evolution, between Bovichtidae and the Eleginopsioidea + Cryonotothenioidea lineage. The *IgT* heavy chain gene of this species showed evident signatures of the evolutionary process that led to the *Cτ2* exon loss in Cryonotothenioidea, which included a small deletion of three nucleotides at the 5' end of the *Cτ2* exon and several short deletions within the *Cτ1-Cτ2* intron, shared with *E. maclovinus*.

Based on these observations, we propose a timeline for the loss of the *Cτ2* exon, which might have occurred after the divergence of Pseudaphritioidea from the main notothenioid lineage, but before the split between Eleginopsioidea and Cryonotothenioidea. This temporal placement would be consistent with the recently revised phylogeny of Notothenioidei [35]. Several observations suggest that this process may have been initiated by the modifications of the two introns flanking the *Cτ2* exon. First, intron length followed a clear trend towards reduction in the Antarctic species, consistent in particular with the progressive loss of the upstream *Cτ1-Cτ2* intron. Second, the phylogenetic analysis, the organization of the motifs and shared indels found in each of the two introns neighboring *Cτ2* were fully consistent with notothenioid phylogeny and marked a close similarity between the introns of *E. maclovinus* and those of Antarctic species, with a progressive disappearance of the Antarctic features in the most early-branching notothenioid species. Third, a significant bias towards a rich AT content in the *Cτ2-Cτ3* intron was observed in Antarctic species and their sister taxa *E. maclovinus*, but not in the early-branching non-Antarctic lineages. The latter observation raises the interesting question of whether the high AT content may be interpreted as an adaptive feature to improve replication, transcription or other molecular processes in a cooling environment. AT-rich introns are a well-known peculiar feature of most teleost genomes, which may be viewed as an ancestral characteristic of ectothermic vertebrates [46]. An integrated revision of the genomes readily available at present for many notothenioid species may provide a useful framework for assessing this issue. These observations are in line with key role that introns cover in the dynamic process of genome evolution [47]. The role of introns in the adaptation to varying environmental pressure may have been particularly relevant in teleosts, where these elements are found in higher number and usually have a shorter length than other vertebrates, as a consequence of the teleost-specific whole-genome duplication event [48, 49]. We have previously revealed, as a key example of this potential for adaptation, the peculiar rearrangement of the exon/intron architecture of the region encoding the C-terminal Extracellular Membrane-Proximal Domain in the IgM heavy chain gene locus of Antarctic fish [50].

Although Transposable Elements (TEs), found in 20–60% of introns in vertebrate genomes, can also provide a significant contribution to the modification of gene architecture [51, 52], we could not find any active TEs in the two introns flanking the *Cτ2* exon, neither in Antarctic, nor in non-Antarctic species. However, we cannot exclude the possibility that TEs that were lost along evolution and that are not detectable anymore have played a role in the process that led to the loss of the *Cτ2* exon.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

Most certainly, the reconstruction of this scenario suffers from missing data and significant "evolutionary gaps" due to the high phylogenetic distance between the genera we took into account [35]. In any case, the *IgT* sequence features outlined above for non-Antarctic species are "molecular fossils" that can provide useful information to infer the stepwise *Cτ2* exon loss that occurred in the Antarctic lineage.

The evolutionary dynamics of exon–intron architecture are another significant aspect of genome adaptation [53, 54], as they accommodate a high variation of intron and exon size but require an extreme conservation of the splicing site motifs located at their boundaries [55]. We show that the *Cτ2* exon remnant of all Antarctic species, despite its short length, retains both functional acceptor (5') and donor (3') splice sites. While the conservation of the donor splice site can be explained by the preservation of the 6 nucleotides at the 3'end of this exon, the entire 5'end of *Cτ2* was deleted in Antarctic species, along with the acceptor splicing site. We speculate that the correct splicing might still occur in this region thanks to the presence of an AGGCAA motif, which nearly matches one of the five predicted splicing regulatory hexamers of mammals [54].

The recurrent reorganization of *Ig* gene loci during vertebrate evolution has often led to the generation of multiple functional gene copies and pseudogenes [56]. A similar evolutionary process might have targeted Antarctic *IgT* genes, thereby generating novel sequence variants that could have granted the acquisition of new structural arrangements and functions, favorable for the adaptation to the polar environment, as in the previously reported case of *T. bernacchii*.

Most of the variants reported in this work had a length similar to the variant S of *T. bernacchii* (24-39 nt - 8-13 aa, vs 33 nt - 11aa). However, due to the limited number of specimens available for each species, we cannot exclude that other Notothenioid species share a similar repertoire of transcript variants. Interestingly, we could not identify any variant characterized, like the Sts isoform of *T. bernacchii*, by the skipping of the *Cτ2* exon remnant. Therefore, Sts may be considered a species-specific isoform, possibly linked with the significant genomic reorganization of Trematominae [57]. Most likely, the availability of additional fully sequenced notothenioid genomes will clarify the orthology and paralogy relationships of the different sequence variants identified in this work.

A question that remains to be solved is why Antarctic species retained a very short remnant of the IgT CH2 domain, instead of completely losing it. A possible explanation might reside in the structural utility of this region as a linker, with the function to keep the first and second CH domains apart from each other, and to provide the Ig molecule with greater flexibility of the Fab arms, as suggested by (i) its high solvent exposure predicted by 3D molecular modeling (Figure 9); (ii) the abundance of amino acid residues typically found in hinge regions (i.e. proline, glycine and cysteine), recalling in particular the human IgA1 hinge region, which is also of similar length.

The finding that the *Cτ2* exon remnant, despite its short length, contained two sites subject to diversifying selection was intriguing. While the first one was found in a region characterized by indels in some species, the second one, encoding Pro/Ser/Thr, was adjacent to a highly conserved cysteine residue. This observation will undoubtedly lead to further investigations aimed at clarifying the functional role of these residues in the context of cold adaptation, which presently remains unknown.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

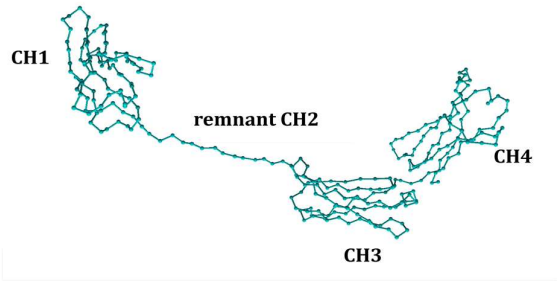

Figure 9. Predicted 3D molecular model of the IgT CH1, remnant CH2, CH3 and CH4 domains from *E. maclovinus* (sequence Ema2), obtained with Phyre 2 (http://www.sbg.bio.ic.ac.uk/phyre2).

Since the IgT heavy chain does not possess the additional cysteine residue engaged in the formation of the covalent bond between the two heavy chains in the monomer, the presence of a conserved cysteine residue in all Antarctic species is another peculiarity of the CH2 remnant. Cysteine pairs involved in the formation of disulfide bonds are highly conserved relative to unpaired cysteines and to other amino acids [58]. Consequently, whenever disulfide bonds are permanently consolidated in proteins, the cysteine residues involved in the formation of such bonds rarely vary. Since the remnant portion of the CH2 domain just maintained the second of the two canonical cysteines, we speculate that the high conservation of this residue might be linked with a function in: (i) bridging the two heavy chains in the monomer; (ii) keeping the monomer units covalently linked in the multimeric form, creating the so-called "redox forms" observed for other Ig isotypes.

In general, variations in disulfide connectivity allow a higher degree of polymerization, influencing the structure, stability and effector functions of Ig molecules [59], as it typically happens in the case of IgM [60]. In light of these observations, it is tempting to speculate that the remnant CH2 domain may provide an advantage for the activity of this Ig isotype under the thermodynamically unfavorable cold conditions of the Antarctica.

Although little information is available concerning the aggregation status of fish IgT, biochemical studies carried out in the rainbow trout have revealed that they are present in the serum as a monomer and that, unlike IgM, the multimeric complexes found in the mucus are non-covalently linked [61]. Considering that teleosts lack the J chain, essential for the formation of Ig polymers, the principles that drive their association in multimers are presently unknown.

Teleost mucosal Igs can associate with the polymeric Ig receptor (pIgR), which is also present in *T. bernacchii* (unpublished data) and possibly in other Antarctic fish, and enables their secretion into the gut lumen, similar to mammalian IgA and IgM [61, 62]. Antarctic IgT contain, at the carboxy terminus of the secretory tail, a sequence motif similar to the one found in other teleosts, which may be involved in polymerization.

A third factor that might allow IgT polymerization relies on glycosylation, even though the role of glycan moieties in teleost mucosa-associated Igs still needs to be clarified. The IgT of Antarctic species contain a high number of sequons consistent with glycosylation sites. In particular, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

presence of two such sites in the C-terminal secretory tail of the Antarctic IgT may suggest a role in the assembly of Ig complexes, as they match the position of conserved glycosylation sites found in mammalian IgM and involved in the polymerization of this Ig isotype [63].

4. Materials and Methods

4.1 Biological samples

The biological material was collected from a set of species representing the Antarctic families Nototheniidae (*Gobionotothen gibberifrons*); Artedidraconidae (*Histiodraco veliver*); Bathydraconidae (*Gymnodraco acuticeps*); Channichthyidae (*Chionodraco hamatus*, *Chionodraco rastrospinosus*, *Chaenocephalus aceratus*, and *Pageotopsis macropterus*); and the three non-Antarctic lineages Bovichtidae (*Bovichtus diachantus* and *Cottoperca gobio*), Eleginopsioidea (*Eleginops maclovinus*) and Pseudaphritioidae (*Pseudaphritis urvillii*). *H. velifer* specimens, as well as the Channichthyidae *C. hamatus* and *P. macropterus* specimens, were caught in the Ross Sea, in the proximity of the Italian "Mario Zucchelli" Station at 74°42′ S, 164°07′ E, during the XXV Italian Antarctic Expedition (2009-2010). Specimens of the other two species of the family Channichthyidae, *C. rastrospinosus* and *C. aceratus*, were collected by bottom trawling from the research vessel L.M. Gould near Low and Braband Islands in the Palmer Archipelago during the XIX Italian Antarctic expedition (2003-2004). The activity permit, released by the Italian National Program for Antarctic Research (PNRA), was in agreement with the "Protocol on environmental protection to the Antarctic Treaty" Annex V. *P. urvillii*, *E. maclovinus*, *G. acuticeps*, and *G. gibberifrons* specimens were collected during the ICEFISH cruise 2004 (International Collaborative Expedition to collect and study Fish Indigenous to Sub-Antarctic Habitats). Spleen and head kidney samples were collected and immediately frozen in liquid nitrogen. All samples, including the blood from *P. urvillii* and *G. gibberifrons*, were kept at -80 °C until use. The tissue sample of each species and classification of the non-Antarctic and Antarctic sampled specimens are summarized in Table S1.

4.2 RNA extraction, PCR amplification, and cloning of cDNA encoding the notothenioid IgT CH region

Considering the variable number of specimens and tissues available for each species, we selected tissue samples of two individuals per species for RNA extraction. The one exception was *E. maclovinus*, as only a single specimen was available for this species. Following the instructions of the SV Total RNA Isolation System kit (Promega), RNA extractions were carried out from 150-200 mg tissue samples, homogenized by Potter-Elvehjem glass-Teflon in TriPure Isolation reagent (Roche). Initial quality checks revealed the successful isolation of RNA suitable for downstream analyses in all samples, with the only exception of *P. urvillii*, where heavy degradation of RNA (obtained from blood cells), incompatible with reverse transcription and PCR applications, was observed.
RNA quality and concentration were assessed by the presence of sharp rRNA bands on 1% agarose gel and by calculating the ratio of absorbance at 260 nm and 280 nm measured by a NanoDrop 1000 Spectrophotometer (Thermo Scientific). cDNA was obtained from 5 ◎g of total RNA using Maxima H Minus Reverse Transcriptase (Thermo Scientific). The target sequence was amplified in a final volume of 25 µl using 2 µl cDNA (20 ng), 1,25 µM of specific primers (1,0µM), 0,5uL dNTP Mix (0,2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

µM), 2,5 µl 10X DreamTaq Buffer, 0,5 µl (1 U) of DreamTaq DNA polymerase (Thermo Scientific), up to volume with H20 as follows: 95 °C for 3 min, 35 cycles of 95 °C (30 s), 60 °C (30 s), and 72 °C (1 min) with a final extension at 72 °C for 10 min. In the case of *E. maclovinus*, a second amplification was carried out following the same conditions as the first PCR in order to increase the amount of the specific product. Primers used in all the PCR experiments are reported in Table S2, which also shows the targeted constant domain of the IgT heavy chain for each primer. PCR products were analyzed on 1% agarose gel, purified by NucleoSpin ® Gel and PCR Clean-up (Macherey-Nagel), and cloned into pGEM®-T Easy Vector (Promega). Positive clones were identified by the blue/white screening method and sequenced on both strands on an ABI PRISM 3100 automated sequencer at Eurofins Genomics Europe Sequencing GmbH (Jakob-Stadler-Platz 7, 78467 Konstanz, Germany). The 5' end of nucleotide sequences of the positive cDNA clones matched the forward primer except for *E. maclovinus* (Figure S1). In this case, the obtained sequences were missing several nucleotides at the 5' end due to a high background noise in the chromatogram. The number of positive cDNA clones encoding the Ig *Cτ2* variants identified both in non-Antarctic and Antarctic species are shown in Table S3.

4.3 DNA extraction, PCR amplification, and cloning of the notothenioid IgT CH region gene

DNA was isolated from 100 mg of head kidney sample of the single *E. maclovinus* specimen available, using the PureLink® Genomic DNA Mini Kit (Thermo Scientific), following the manufacturer's instructions. The same experimental procedure was applied for *G. gibberifrons*, *H. velifer* and *G. acuticeps* spleen and for *N. coriiceps* and *P. urvillii* blood. PCR amplification of genomic DNA was performed using 150 ng of template gDNA and the same Taq polymerase described in the previous section. Primers used are reported in supplementary table 1. PCR products were analysed on 1% agarose gel, purified by NucleoSpin ® Gel and PCR Clean-up (Macherey-Nagel), and cloned into pGEM®-T Easy Vector (Promega). Positive clones were identified by the blue/white screening method and sequenced on both strands on an ABI PRISM 3100 automated sequencer Eurofins Genomics Europe Sequencing GmbH (Jakob-Stadler-Platz 7, 78467 Konstanz, Germany). Not all 5' ends of nucleotide sequences of the positive genomic clones matched the forward primer (Figures S2 and S3) due to the high background noise in the chromatogram. The number of positive genomic clones and the corresponding encoded Ig *Cτ2* variants identified both in non-Antarctic and Antarctic species are summarized in Table S4.

2.4 Additional sequences obtained from public repositories

Sequence data concerning the *IgT* gene and cDNA sequences from *T. bernacchii* (family Nototheniidae) and from *B. diacanthus* (family Bovichtidae) were retrieved from a previous study carried out by our research team [18]. The genomic DNA and predicted cDNA sequences of the *IgT* genes of *Notothenia coriiceps* (family Nototheniidae) and *C. gobio* were retrieved from publicly available annotated genome assemblies. In detail, the *N. coriiceps* genome refers to the study by Shin et al. [28] whereas the genome of *C. gobio* (v. fCotCob3.1) has been recently released within the frame of the Vertebrate Genome Project [64].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

2.5 Computational analysis

Sequencing chromatograms were visualized using the program FinchTV (version 1.3.0). The nucleotide sequences obtained were verified by sequence similarity searches against the GenBank database, using the BLAST program [65]. Amino acid sequences were deduced from nucleotide sequences using the ExPASy Translate Tool tool. The amino acid composition was analysed using the ExPASy ProtParam and Pep-Calc (www.pepcalc.com) tools. The GC-content of introns was calculated with the GC Content Calculator (Biologics International Corp, Indianapolis, USA). Multiple sequence alignments were performed with ClustalW [66]. Sequons and putative *N*-Glycosylation sites were identified using the NetNGlyc 4.0 Server [67] (at https://www.expasy.org/proteomics). A 3D molecular model was built for the IgT CH domains of *E. maclovinus* (sequence Ema2) using the Phyre2 tool [68] available at http://www.sbg.bio.ic.ac.uk/phyre2. The structure of the human secretory IgA1 was used as template (PDB entry: 3CHN).

2.6 Association between IgT genes and repeated elements

The reference genomes assembly of *C. gobio* [64] was analyzed with RepeatScout v.1.05 [69] to generate a species-specific repeat library. The IgT genomic DNA sequences obtained in this study, as well as those identified in the genome of *C. gobio*, were analyzed with RepeatMasker v.4.09 [70], with particular attention to the *Cτ1-Cτ2* and *Cτ2-Cτ3* intronic regions. All sequences were screened for the presence of repeats against the Dfam v.3.1 [71] library of known repeats found in the genomes of Actinopterygii. The *IgT* gene of *C. gobio* was subjected to an additional round of screening against the custom species-specific repeat libraries generated as described above.

2.7 Molecular evolution of the Cτ1-Cτ2 and Cτ2-Cτ3 introns

We investigated the evolutionary history of the *IgT* genes identified in Notothenioidei, with particular focus on the genomic region subjected to the highest molecular diversity, i.e. the region spanning the *Cτ1-Cτ2* and *Cτ2-Cτ3* introns. Due to the truncation of the *Cτ2* exon in Cryonotothenioidea (see the results section), this region was disregarded. The nucleotide sequences of the two introns were separately aligned with MUSCLE v.3.8.31 [72] The multiple sequence alignments were manually refined and processed with GBLOCKS v.0.91b [73] to remove phylogenetically uninformative positions. The two sequence blocks were concatenated and tested with ModelTest-NG v.0.1.3 [74] to identify the best-fitting model of molecular evolution. This was found to be the GTR+I model (Generalized time-reversible, with a proportion of invariable sites) [75], based on the corrected Akaike information criterion [76]. The multiple sequence alignment file was used as an input for Bayesian phylogenetic inference, carried out with MrBayes v.3.2.7a [77], run for one million generations and two parallel Markov Chain Monte Carlo (MCMC) analyses. The convergence of the two independent analyses was checked with Tracer [78], by evaluating that all the estimated parameters reached an ESS value higher than 200.

2.8 Selection analyses

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

The Ig *Cτ* cDNA sequences of the available species from Cryonotothenioidea, either determined by cloning or inferred from genomic DNA, were aligned with a strategy aimed at preserving the integrity of codon triplets. This was achieved by aligning the translated amino acid sequences with MUSCLE v.3.8.31 [72] within the MEGAX environment [79] and back-translating the aligned sequences to the original gapped coding nucleotide sequence.

The multiple sequence analysis was subjected to tests aimed at detecting signatures of selection with the DataMonkey adaptive evolution platform [80]. Sites evolving under pervasive positive and negative selection were detected with FEL (Fixed Effect Likelihood) [81] and those with evidence of episodic positive selection were identified with MEME (Mixed EffectsModel of Evolution) [82] using default p-value thresholds.

2.9 Identification of short conserved sequence motifs in Antarctic species

The unaligned IgT protein sequences of Cryonotothenioidea were analyzed to detect the presence of short ungapped conserved amino acid motifs of 3-8 aa length with MEME (Multiple Em for Motif Elicitation) [83] within the MEME suite v.5.0.1 environment [84]. The specific association of the detected motifs with Cryonotothenioidea was subsequently tested by assessing their absence in the IgT sequences of non-Antarctic Notothenioidei and temperate Perciformes. Finally, the relevance of the identified motifs in the context of evolution in the Antarctic environment was evaluated by inspecting their overlap with negatively selected sites, identified as explained in the previous section.

The *Cτ1-Cτ2* and *Cτ2-Cτ3* introns of all the available Notothenioidei sequences were similarly screened, looking for conserved nucleotide motifs with size comprised between 6 and 50 base pairs, allowing any number of motif repeats for each sequence.

2.10 Data availability

Genomic DNA and/or cDNA sequences from *P. urvillii*, *E. maclovinus*, *G. gibberifrons*, *N. coriiceps*, *H. velifer*, and *G. acuticeps* have been deposited in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) are available with the following accession numbers: MN583561 (gDNA clone gPur1); MN583559 (gDNA clone gEma1); MN583560 (gDNA clone gEma3); MN583563 (gDNA clone gGgi2); MN583562 (gDNA clone gNco1); MN602803 (gDNA clone gGac1); MN602804 (clone gGgi3, pseudogene); MN602807 (gDNA clone gHve2); MN602808 (gDNA clone gHve5); MN583555 (cDNA clone Ema1); MN583556 (cDNA clone Ema2); MN583557 (cDNA clone Ggi1); MN602805 (cDNA clone Hve1); MN602806 (cDNA clone Hve2); MN583558 (cDNA clone Gac1).

5. Conclusions

The many challenges notothenioid fishes faced during their evolutionary history triggered a wave of genomic changes that led to a number of peculiar features that might have been preserved as adaptive traits. The development of an IgT molecule nearly completely devoid of a CH domain

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

might be just another addition to the list of the highly successful strategies these organisms have used to adapt to this challenging environment. The evolutionary scenario we propose involves a gradual process, which has shaped the *IgH* gene locus over a long period of time, through the action of contrasting evolutionary forces, contributing to the generation of an Ig molecule with a unique architecture among vertebrates.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, A.A., M.R.C. and M.G.; Validation, A.A., S.G. and M.G.; Formal Analysis, A.A., M.R.C., S.G. and M.G.; Investigation, A.A., M.G., S.G. and M.V.; Resources, M.R.C. and M.G.; Writing - Original Draft, M.R.C. and M.G.; Writing - Review & Editing, A.A., M.R.C., M.G., S.G., U.O., and M.V.; Visualization, A.A., U.O., M.G.; Supervision, M.R.C. and M.G.; Project Administration, M.R.C. and M.G.; Funding Acquisition, M.R.C. and M.G.

All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian National Program for Research in Antarctica (PNRA), grant number PNRA 16_00099–A1 Project.

Acknowledgments

The authors wish to thank Dr Ennio Cocca (IBBR, CNR, Naples, Italy), for providing tissue samples of *E. maclovinus, C. rastrospinosus,* and *C. aceratus* specimens; Drs Daniela Giordano, and Cinzia Verde (IBBR, CNR, Naples, Italy) for providing *P. urvillii* and *G. gibberifrons* blood samples. The authors are grateful to Prof. Alberto Pallavicini (University of Trieste, Italy) for his critical comments on the manuscript.

In memory of Guido di Prisco, for his scientific career entirely devoted to unveil the mystery of evolutionary adaptations for life in Antarctica.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

Ig	Immunoglobulin
Ig VH	Immunoglobulin heavy chain Variable region
Ig CH	Immunoglobulin heavy chain Constant domain
$C\tau$	IgT heavy chain constant exon
TbeL	*Trematomus bernacchii* IgT heavy chain Long variant
TbeS	*Trematomus bernacchii* IgT heavy chain Short variant
TbeSts	*Trematomus bernacchii* IgT heavy chain Shortest variant

References

1. Salinas, I.; Zhang, Y.-A.; Sunyer, J. O. Mucosal Immunoglobulins and B Cells of Teleost Fish. *Dev Comp Immunol* **2011**, 35, 1346–1365, doi:10.1016/j.dci.2011.11.009.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

2. Chen, K.; Cerutti, A. New Insights into the Enigma of Immunoglobulin D. Immunological Reviews **2010**, 237, 160–179, doi:10.1111/j.1600-065X.2010.00929.x.

3. Perdiguero, P.; Martín-Martín, A.; Benedicenti, O.; Díaz-Rosales, P.; Morel, E.; Muñoz-Atienza, E.; García-Flores, M.; Simón, R.; Soleto, I.; Cerutti, A. et al. Teleost IgD+IgM− B Cells Mount Clonally Expanded and Mildly Mutated Intestinal IgD Responses in the Absence of Lymphoid Follicles. *Cell Reports* **2019**, 29, 4223-4235.e5, doi: 10.1016/j.celrep.2019.11.101.

4. Danilova, N.; Bussmann, J.; Jekosch, K.; Steiner, L. A. The Immunoglobulin Heavy-Chain Locus in Zebrafish: Identification and Expression of a Previously Unknown Isotype, Immunoglobulin Z. *Nat Immunol* **2005**, 6, 295–302, doi:10.1038/ni1166.

5. Hansen, J. D.; Landis, E. D.; Phillips, R. B. Discovery of a Unique Ig Heavy-Chain Isotype (IgT) in Rainbow Trout: Implications for a Distinctive B Cell Developmental Pathway in Teleost Fish. *Proc Natl Acad Sci U S A* **2005**, 102, 6919–6924, doi:10.1073/pnas.0500027102.

6. Zhang, Y.-A.; Salinas, I.; Li, J.; Parra, D.; Bjork, S.; Xu, Z.; LaPatra, S. E.; Bartholomew, J.; Sunyer, J. O. IgT, a Primitive Immunoglobulin Class Specialized in Mucosal Immunity. *Nat Immunol* **2010**, 11, 827–835, doi:10.1038/ni.1913.

7. Xu, Z.; Parra, D.; Gómez, D.; Salinas, I.; Zhang, Y.-A.; von Gersdorff Jørgensen, L.; Heinecke, R. D.; Buchmann, K.; LaPatra, S.; Sunyer, J. O. Teleost Skin, an Ancient Mucosal Surface That Elicits Gut-like Immune Responses. *Proc Natl Acad Sci U S A* **2013**, 110, 13097–13102, doi:10.1073/pnas.1304319110.

8. Tacchi, L.; Musharrafieh, R.; Larragoite, E. T.; Crossey, K.; Erhardt, E. B.; Martin, S. A. M.; LaPatra, S. E.; Salinas, I. Nasal Immunity Is an Ancient Arm of the Mucosal Immune System of Vertebrates. *Nat Commun* **2014**, 5, 5205, doi:10.1038/ncomms6205.

9. Xu, Z.; Takizawa, F.; Parra, D.; Gómez, D.; von Gersdorff Jørgensen, L.; LaPatra, S. E.; Sunyer, J. O. Mucosal Immunoglobulins at Respiratory Surfaces Mark an Ancient Association That Predates the Emergence of Tetrapods. *Nat Commun* **2016**, 7, 10728, doi:10.1038/ncomms10728.

10. Yu, Y.-Y.; Kong, W.-G.; Xu, H.-Y.; Huang, Z.-Y.; Zhang, X.-T.; Ding, L.-G.; Dong, S.; Yin, G.-M.; Dong, F.; Yu, W.; et al. Convergent Evolution of Mucosal Immune Responses at the Buccal Cavity of Teleost Fish. *iScience* **2019**, 19, 821–835, doi:10.1016/j.isci.2019.08.034.

11. Zhang, N.; Zhang, X.-J.; Chen, D.-D.; Sunyer, J. O.; Zhang, Y.-A. Molecular Characterization and Expression Analysis of Three Subclasses of IgT in Rainbow Trout (Oncorhynchus Mykiss). *Dev Comp Immunol* **2017**, 70, 94–105, doi:10.1016/j.dci.2017.01.001.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

12. Buonocore, F.; Stocchi, V.; Nunez-Ortiz, N.; Randelli, E.; Gerdol, M.; Pallavicini, A.; Facchiano, A.; Bernini, C.; Guerra, L.; Scapigliati, G.; Picchietti, S. Immunoglobulin T from Sea Bass (Dicentrarchus Labrax L.): Molecular Characterization, Tissue Localization and Expression after Nodavirus Infection. *BMC Mol Biol* **2017**, 18, 8, doi:10.1186/s12867-017-0085-0.

13. Fillatreau, S.; Six, A.; Magadan, S.; Castro, R.; Sunyer, J. O.; Boudinot, P. The Astonishing Diversity of Ig Classes and B Cell Repertoires in Teleost Fish. *Front Immunol* **2013**, 4, doi:10.3389/fimmu.2013.00028.

14. Savan, R.; Aman, A.; Sato, K.; Yamaguchi, R.; Sakai, M. Discovery of a New Class of Immunoglobulin Heavy Chain from Fugu. *Eur J Immunol* **2005**, 35, 3320–3331, doi:10.1002/eji.200535248.

15. Velázquez, J.; Acosta, J.; Lugo, J. M.; Reyes, E.; Herrera, F.; González, O.; Morales, A.; Carpio, Y.; Estrada, M. P. Discovery of Immunoglobulin T in Nile Tilapia (Oreochromis Niloticus): A Potential Molecular Marker to Understand Mucosal Immunity in This Species. *Dev Comp Immunol* **2018**, 88, 124–136, doi:10.1016/j.dci.2018.07.013.

16. Savan, R.; Aman, A.; Nakao, M.; Watanuki, H.; Sakai, M. Discovery of a Novel Immunoglobulin Heavy Chain Gene Chimera from Common Carp (Cyprinus Carpio L.). *Immunogenetics* **2005**, 57, 458–463, doi:10.1007/s00251-005-0015-z.

17. Ryo, S.; Wijdeven, R. H. M.; Tyagi, A.; Hermsen, T.; Kono, T.; Karunasagar, I.; Rombout, J. H. W. M.; Sakai, M.; Verburg-van Kemenade, B. M. L.; Savan, R. Common Carp Have Two Subclasses of Bonyfish Specific Antibody IgZ Showing Differential Expression in Response to Infection. *Dev Comp Immunol* **2010**, 34, 1183–1190, doi:10.1016/j.dci.2010.06.012.

18. Giacomelli, S.; Buonocore, F.; Albanese, F.; Scapigliati, G.; Gerdol, M.; Oreste, U.; Coscia, M. R. New Insights into Evolution of IgT Genes Coming from Antarctic Teleosts. *Mar Genomics* **2015**, 24, 55–68, doi:10.1016/j.margen.2015.06.009.

19. Eastman, J. T. The Nature of the Diversity of Antarctic Fishes. *Polar Biol* **2005**, 28, 93–107, doi:10.1007/s00300-004-0667-4.

20. Hubold, G. Antarctic Fish Biology: Evolution in a Unique Environment. Joseph T. Eastman. 1993. San Diego: Academic Press, Xiii + 322 p, Illustrated, Hard Cover. ISBN 0-12-228140-3. US$74.95. Polar Record 1994, 30, 59–60, doi:10.1017/S0032247400021100.

21. Clarke, A.; Johnston, I. A. Evolution and Adaptive Radiation of Antarctic Fishes. Trends in Ecology & Evolution **1996**, 11, 212–218, doi:0.1016/0169-5347(96)10029-X.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

22. Near, T. J.; Eytan, R. I.; Dornburg, A.; Kuhn, K. L.; Moore, J. A.; Davis, M. P.; Wainwright, P. C.; Friedman, M.; Smith, W. L. Resolution of Ray-Finned Fish Phylogeny and Timing of Diversification. *Proc Natl Acad Sci U S A* **2012**, 109, 13698–13703, doi:10.1073/pnas.1206625109.

23. DeVries, A. L.; Wohlschlag, D. E. Freezing Resistance in Some Antarctic Fishes. *Science* **1969**, 163, 1073–1075, doi:10.1126/science.163.3871.1073.

24. Chen, L.; DeVries, A. L.; Cheng, C.-H. C. Evolution of Antifreeze Glycoprotein Gene from a Trypsinogen Gene in Antarctic Notothenioid Fish. *Proc Natl Acad Sci U S A* **1997**, 94, 3811–3816, doi:10.1073/pnas.94.8.3811.

25. Hofmann, G. E.; Buckley, B. A.; Airaksinen, S.; Keen, J. E.; Somero, G. N. Heat-Shock Protein Expression Is Absent in the Antarctic Fish Trematomus Bernacchii (Family Nototheniidae). *J Exp Biol* **2000**, 203, 2331–2339.

26. Hureau, J.C.; Petit, D.; Fine, J.M.; Marneux, M. Adaptations within Antarctic Ecosystems. **1977**, pp 459-477, Smithsonian Institution, Washington, D.C.

27. Barber, D. L.; Westermann, J. E. M.; White, M. G. The Blood Cells of the Antarctic Icefish Chaenocephalus Aceratus Lönnberg: Light and Electron Microscopic Observations. *J Fish Biol* **1981**, 19, 11–28. doi:10.1111/j.1095-8649.1981.tb05807.x.

28. Shin, S. C.; Kim, S. J.; Lee, J. K.; Ahn, D. H.; Kim, M. G.; Lee, H.; Lee, J.; Kim, B.-K.; Park, H. Transcriptomics and Comparative Analysis of Three Antarctic Notothenioid Fishes. *PLoS One* **2012**, 7, e43762, doi:10.1371/journal.pone.0043762.

29. Near, T. J.; Dornburg, A.; Harrington, R. C.; Oliveira, C.; Pietsch, T. W.; Thacker, C. E.; Satoh, T. P.; Katayama, E.; Wainwright, P. C.; Eastman, J. T.; et al. Identification of the Notothenioid Sister Lineage Illuminates the Biogeographic History of an Antarctic Adaptive Radiation. *BMC Evol Biol* **2015**, 15, doi:10.1186/s12862-015-0362-9.

30. Near, T. J.; MacGuigan, D. J.; Parker, E.; Struthers, C. D.; Jones, C. D.; Dornburg, A. Phylogenetic Analysis of Antarctic Notothenioids Illuminates the Utility of RADseq for Resolving Cenozoic Adaptive Radiations. *Mol Phylogenet Evol* **2018**, 129, 268–279, doi:10.1016/j.ympev.2018.09.001.

31. Bargelloni, L.; Babbucci, M.; Ferraresso, S.; Papetti, C.; Vitulo, N.; Carraro, R.; Pauletto, M.; Santovito, G.; Lucassen, M.; Mark, F. C.; et al. 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. *Commun Biol* **2019**, 2, 443, doi:10.1038/s42003-019-0685-y.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

32. Kim, B.-M.; Amores, A.; Kang, S.; Ahn, D.-H.; Kim, J.-H.; Kim, I.-C.; Lee, J. H.; Lee, S. G.; Lee, H.; Lee, J.; et al. Antarctic Blackfin Icefish Genome Reveals Adaptations to Extreme Environments. *Nat Ecol Evol* **2019**, 3, 469–478, doi:10.1038/s41559-019-0812-7.

33. Near, T. J.; Eytan, R. I.; Dornburg, A.; Kuhn, K. L.; Moore, J. A.; Davis, M. P.; Wainwright, P. C.; Friedman, M.; Smith, W. L. Resolution of Ray-Finned Fish Phylogeny and Timing of Diversification. *Proc Natl Acad Sci U S A* **2012**, 109, 13698–13703, doi:10.1073/pnas.1206625109.

34. Ceballos, S. G.; Lessa, E. P.; Victorio, M. F.; Fernández, D. A. Phylogeography of the Sub-Antarctic Notothenioid Fish Eleginops Maclovinus: Evidence of Population Expansion. *Mar Biol* **2012**, 159, 499–505, doi:10.1007/s00227-011-1830-4.

35. Near, T. J.; Ghezelayagh, A.; Ojeda, F. P.; Dornburg, A. Recent Diversification in an Ancient Lineage of Notothenioid Fishes (Bovichtus: Notothenioidei). *Polar Biol* **2019**, 42, 943–952, doi:10.1007/s00300-019-02489-1.

36. Hancks, D. C.; Kazazian, H. H. Roles for Retrotransposon Insertions in Human Disease. *Mobile DNA* **2016**, 7, 65, doi:10.1186/s13100-016-0065-9.

37. Davis, M. B.; Dietz, J.; Standiford, D. M.; Emerson, C. P. Transposable Element Insertions Respecify Alternative Exon Splicing in Three Drosophila Myosin Heavy Chain Mutants. *Genetics* **1998**, 150, 1105–1114.

38. Ejima, Y.; Yang, L. Trans Mobilization of Genomic DNA as a Mechanism for Retrotransposon-Mediated Exon Shuffling. *Hum Mol Genet* **2003**, 12, 1321–1328, doi:10.1093/hmg/ddg138.

39. Colombo, M.; Damerau, M.; Hanel, R.; Salzburger, W.; Matschiner, M. Diversity and Disparity through Time in the Adaptive Radiation of Antarctic Notothenioid Fishes. *J Evol Biol* **2015**, 28, 376–394, doi:10.1111/jeb.12570.

40. Gambòn-Deza, F.; Sànchez-Espinel, C.; Magadàn-Mompò, S. Presence of an unique IgT on the IGH locus in three-spined stickleback fish (*Gasterosteus aculeatus*) and the very recent generation of a repertoire of VH genes. *Dev Comp Immunol* **2010**, 34, 114-122, doi: 10.1016/j.dci.2009.08.011

41. Bause, E. Structural Requirements of N-Glycosylation of Proteins. Studies with Proline Peptides as Conformational Probes. *Biochem J* **1983**, 209, 331–336.

42. Coscia, M. R.; Morea, V.; Tramontano, A.; Oreste, U. Analysis of a CDNA Sequence Encoding the Immunoglobulin Heavy Chain of the Antarctic Teleost *Trematomus bernacchii*. *Fish Shellfish Immunol* **2000**, 10, 343–357, doi:10.1006/fsim.1999.0244.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

43. Coscia, M. R.; Varriale, S.; Giacomelli, S.; Oreste, U. Antarctic Teleost Immunoglobulins: More Extreme, More Interesting. *Fish Shellfish Immunol* **2011**, 31, 688–696,doi:10.1016/j.fsi.2010.10.018.

44. Coscia, M. R.; Giacomelli, S.; Oreste, U. Allelic Polymorphism of Immunoglobulin Heavy Chain Genes in the Antarctic Teleost Trematomus Bernacchii. *Mar Genomics* **2012**, 8, 43–48, doi:10.1016/j.margen.2012.04.002.

45. Flajnik, M. F. A Cold-Blooded View of Adaptive Immunity. *Nat Rev Immunol* **2018**, 18, 438–453, doi:10.1038/s41577-018-0003-9.

46. Bernardi, G.; Bernardi, G. Compositional Constraints and Genome Evolution. *J Mol Evol* **1986**, 24, 1–11, doi:10.1007/BF02099946.

47. Jeffares, D. C.; Mourier, T.; Penny, D. The Biology of Intron Gain and Loss. *Trends Genet* **2006**, 22, 16–22, doi:10.1016/j.tig.2005.10.006.

48. Meyer, A.; Van de Peer, Y. From 2R to 3R: Evidence for a Fish-Specific Genome Duplication (FSGD). *Bioessays* **2005**, 27, 937–945; doi:10.1002/bies.20293.

49. Glasauer, S. M. K.; Neuhauss, S. C. F. Whole-Genome Duplication in Teleost Fishes and Its Evolutionary Consequences. *Mol Genet Genomics* **2014**, 289, 1045–1060, doi:10.1007/s00438-014-0889-2.

50. Coscia, M. R.; Varriale, S.; De Santi, C.; Giacomelli, S.; Oreste, U. Evolution of the Antarctic Teleost Immunoglobulin Heavy Chain Gene. *Mol Phylogenet Evol* **2010**, 55, 226–233, doi:10.1016/j.ympev.2009.09.033.

51. Mills, R. E.; Bennett, E. A.; Iskow, R. C.; Devine, S. E. Which Transposable Elements Are Active in the Human Genome? *Trends Genet* **2007**, 23, 183–191, doi:10.1016/j.tig.2007.02.006.

52. Sela, N.; Mersch, B.; Hotz-Wagenblatt, A.; Ast, G. Characteristics of Transposable Element Exonization within Human and Mouse. *PLoS One* **2010**, 5, doi:10.1371/journal.pone.0010907.

53. Roy, S. W.; Gilbert, W. The Evolution of Spliceosomal Introns: Patterns, Puzzles and Progress. *Nat Rev Genet* **2006**, 7, 211–221, doi:10.1038/nrg1807.

54. Gelfman, S.; Burstein, D.; Penn, O.; Savchenko, A.; Amit, M.; Schwartz, S.; Pupko, T.; Ast, G. Changes in Exon–Intron Structure during Vertebrate Evolution Affect the Splicing Pattern of Exons. *Genome Res* **2012**, 22, 35–50, doi:10.1101/gr.119834.110.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

55. Schwartz, S.; Silva, J.; Burstein, D.; Pupko, T.; Eyras, E.; Ast, G. Large-Scale Comparative Analysis of Splicing Signals and Their Corresponding Splicing Factors in Eukaryotes. *Genome Res* **2008**, 18, 88–103,doi:10.1101/gr.6818908.

56. Das, S.; Hirano, M.; Tako, R.; McCallister, C.; Nikolaidis, N. Evolutionary Genomics of Immunoglobulin-Encoding Loci in Vertebrates. *Curr Genomics* **2012**, 13, 95–102, doi:10.2174/138920212799860652.

57. Pisano, E.; Coscia, M. R.; Mazzei, F.; Ghigliotti, L.; Coutanceau, J.-P.; Ozouf-Costaz, C.; Oreste, U. Cytogenetic Mapping of Immunoglobulin Heavy Chain Genes in Antarctic Fish. *Genetica* **2007**, 130, 9–17, doi:10.1007/s10709-006-0015-4.

58. Wong, J. W. H.; Ho, S. Y. W.; Hogg, P. J. Disulfide Bond Acquisition through Eukaryotic Protein Evolution. *Mol Biol Evol* **2011**, 28, 327–334, doi:10.1093/molbev/msq194.

59. Liu, H.; May, K. Disulfide Bond Structures of IgG Molecules: Structural Variations, Chemical Modifications and Possible Impacts to Stability and Biological Function. *MAbs* **2012**, 4, 17–23, doi: 10.4161/mabs.4.1.18347.

60. Kaattari, S.; Evans, D.; Klemer, J. Varied Redox Forms of Teleost IgM: An Alternative to Isotypic Diversity? *Immunol Rev* **1998**, 166, 133–142, doi:10.1111/j.1600-065x.1998.tb01258.x.

61. Zhang, Y.-A.; Salinas, I.; Sunyer, J. O. Recent Findings on the Structure and Function of Teleost IgT. *Fish Shellfish Immunol* **2011**, 31, 627–634, doi:10.1016/j.fsi.2011.03.021.

62. Kaetzel, C. S. Coevolution of Mucosal Immunoglobulins and the Polymeric Immunoglobulin Receptor: Evidence That the Commensal Microbiota Provided the Driving Force. *Int Sch Res Notices.* Volume **2014**, Article ID 541537, doi: 10.1155/2014/541537.

63. Su, Y.-L.; Wang, B.; Hu, M.-D.; Cui, Z.-W.; Wan, J.; Bai, H.; Yang, Q.; Cui, Y.-F.; Wan, C.-H.; Xiong, L.; Zhang, Y.-A.; Geng, H. Site-Specific N-Glycan Characterization of Grass Carp Serum IgM. *Front Immunol* **2018**, 9, doi:10.3389/fimmu.2018.02645.

64. Koepfli, K.-P.; Paten, B.; Genome 10K Community of Scientists; O'Brien, S. J. The Genome 10K Project: A Way Forward. *Annu Rev Anim Biosci* **2015**, 3, 57–111, doi:10.1146/annurev-animal-090414-014900.

65. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J Mol Biol* **1990**, 215, 403–410, doi:10.1016/S0022-2836(05)80360-2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

66. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res* **1994**, 22, 4673–4680, doi:10.1093/nar/22.22.4673.

67. Gupta, R.; Jung, E.; Brunak, S. Prediction of N-glycosylation sites in human proteins. **2004**

68. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat Protoc* **2015**, 10, 845–858, doi:10.1038/nprot.2015.053.

69. Price, A. L.; Jones, N. C.; Pevzner, P. A. De Novo Identification of Repeat Families in Large Genomes. *Bioinformatics* **2005**, 21 Suppl 1, i351-358, doi:10.1093/bioinformatics/bti1018.

70. Smit, A.F.A; Hubley, R.; Green, P. **1996**. *RepeatMaster Open-3.0*

71. Hubley, R.; Finn, R. D.; Clements, J.; Eddy, S. R.; Jones, T. A.; Bao, W.; Smit, A. F. A.; Wheeler, T. J. The Dfam Database of Repetitive DNA Families. *Nucleic Acids Res* **2016**, 44, D81-89, doi:10.1093/nar/gkv1272.

72. Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res* **2004**, 32, 1792–1797, doi:10.1093/nar/gkh340.

73. Talavera, G.; Castresana, J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst Biol* **2007**, 56, 564–577, doi:10.1080/10635150701472164.

74. Darriba, D.; Posada, D.; Kozlov, A. M.; Stamatakis, A.; Morel, B.; Flouri, T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol Biol Evol* **2020**, 37, 291–294, doi:10.1093/molbev/msz189.

75. Tavare, S. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Some mathematical questions in biology / DNA sequence analysis edited by Robert M. Miura **1986**.

76. Cavanaugh, J. E. Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria. *Statistics & Probability Letters* **1997**, 33, 201–208,doi:10.1016/S0167-7152(96)00128-9.

77. Huelsenbeck, J. P.; Ronquist, F. MRBAYES: Bayesian Inference of Phylogenetic Trees. *Bioinformatics* **2001**, 17, 754–755, doi:10.1093/bioinformatics/17.8.754.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 November 2020 doi:10.20944/preprints202011.0685.v1

78. Rambaut, A.; Drummond, A. J.; Xie, D.; Baele, G.; Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* **2018**, 67, 901–904, doi:10.1093/sysbio/syy032.

79. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **2018**, 35, 1547–1549, doi:10.1093/molbev/msy096.

80. Weaver, S.; Shank, S. D.; Spielman, S. J.; Li, M.; Muse, S. V.; Kosakovsky Pond, S. L. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol Biol Evol* **2018**, 35, 773–777, doi:10.1093/molbev/msx335.

81. Kosakovsky Pond, S. L.; Frost, S. D. W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol* **2005**, 22, 1208–1222, doi:10.1093/molbev/msi105.

82. Murrell, B.; Wertheim, J. O.; Moola, S.; Weighill, T.; Scheffler, K.; Pond, S. L. K. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLOS Genet* **2012**, 8, e1002764, doi:10.1371/journal.pgen.1002764.

83. Bailey, T. L.; Elkan, C. Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Biopolymers. *Proc Int Conf Intell Syst Mol Biol* **1994**, 2, 28–36.

84. Bailey, T. L.; Boden, M.; Buske, F. A.; Frith, M.; Grant, C. E.; Clementi, L.; Ren, J.; Li, W. W.; Noble, W. S. MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Res* **2009**, 37 (Web Server issue), W202-208, doi:10.1093/nar/gkp335.