

GigaScience

Genomic bases for colonizing the freezing Southern Ocean revealed by the genomes of Antarctic toothfish and *Patagonia roballo* --Manuscript Draft--

Manuscript Number:	GIGA-D-18-00328	
Full Title:	Genomic bases for colonizing the freezing Southern Ocean revealed by the genomes of Antarctic toothfish and <i>Patagonia roballo</i>	
Article Type:	Research	
Funding Information:	Natural Science Foundation of China (31572598)	Dr. Qianghua Xu
	Natural Science Foundation of China (41761134050)	Dr. liangbiao Chen
	Natural Science Foundation of China (31572611)	Dr. liangbiao Chen
	Major Science Innovation Grant from the Shanghai Education Committee (2017-01-07-00-10-E00060)	Dr. liangbiao Chen
	Key Achievement Supporting Grant from Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology	Dr. liangbiao Chen
	USA NSF Polar Programs grant (ANT1142158)	Dr. Chi-Hing Christina Cheng
Abstract:	<p>The Southern Ocean is the coldest ocean on Earth but a hotspot of evolution. The bottom-dwelling Eocene ancestor of Antarctic notothenioid fishes survived polar marine glaciation and underwent adaptive radiation forming >120 species that fill all water column niches today. Genome-wide changes enabling physiological adaptations and rapid expansion of the Antarctic Notothenioids remain poorly understood. To advance our understanding, we sequenced and compared two notothenioid genomes - the cold-adapted and neutrally buoyant Antarctic toothfish <i>Dissostichus mawsoni</i>, and the basal <i>Patagonia roballo</i> <i>Eleginops maclovinus</i> representing the temperate ancestor. We detected >200 protein gene families that had expanded and thousands of genes that had evolved faster in the toothfish, with diverse cold-relevant functions including stress response, lipid metabolism, protein homeostasis and freeze resistance. Besides AFGP, an eggshell protein had functionally diversified to aid in cellular freezing resistance. Genomic and transcriptomic comparisons revealed proliferation of Sclys-tRNA genes and broad transcriptional upregulation across anti-oxidative selenoproteins, signifying their prominent role in mitigating oxidative stress in the oxygen-rich Southern Ocean. We found expansion of transposable elements, temporally correlated to Antarctic notothenioid diversification. In addition, the toothfish exhibited remarkable shifts in genetic programs towards enhanced fat cell differentiation and lipid storage, and promotion of chondrogenesis while inhibiting osteogenesis in bone development, collectively contributing to achieving neutral buoyancy and pelagicism. Our study revealed a comprehensive landscape of evolutionary changes essential for Antarctic notothenioid cold adaptation and ecological expansion. The two genomes are valuable resources for further uncovering mechanisms underlying the spectacular notothenioids radiation driven by the coldest environment.</p>	
Corresponding Author:	liangbiao Chen Shanghai Ocean University Lingang New City, CHINA	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	Shanghai Ocean University	
Corresponding Author's Secondary Institution:		
First Author:	liangbiao Chen	

First Author Secondary Information:	
Order of Authors:	liangbiao Chen
	Ying Lu
	Wenhao Li
	Yandong Ren
	Mengchao Yu
	Shouwen Jiang
	Yanxia Fu
	Jian Wang
	Sihua Peng
	Kevin T. Bilyk
	Katherine R. Murphy
	Xuan Zhuang
	Mathias Hune
	Wanying Zhai
	Wen Wang
Qianghua Xu	
Chi-Hing Christina Cheng	
Order of Authors Secondary Information:	
Additional Information:	
Question	Response
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist . Information essential to interpreting the data presented should be made available in the figure legends. Have you included all the information requested in your manuscript?	Yes
Resources A description of all resources used, including antibodies, cell lines, animals and software tools, with enough	Yes

<p>information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	
<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?</p>	<p>Yes</p>

[Click here to view linked References](#)

1 1 **Genomic bases for colonizing the freezing Southern Ocean revealed**
2
3 2 **by the genomes of Antarctic toothfish and Patagonia robalo**

4 3
5 4 Liangbiao Chen^{1,2,3,*}, Ying Lu^{1,2,*}, Wenhao Li^{1,2,*}, Yandong Ren⁴, Mengchao Yu^{1,2}, Shouwen
6 5 Jiang^{1,2}, Yanxia Fu^{1,2}, Jian Wang^{1,2}, Sihua Peng^{1,2}, Kevin T. Bilyk⁵, Katherine R. Murphy⁵,
7 6 Xuan Zhuang⁵, Mathias Hune⁶, Wanying Zhai^{1,2}, Wen Wang⁴, Qianghua Xu^{1,2,§}, Chi-Hing
8 7 Christina Cheng^{5,6,§}

9 8
10 9 ¹National Demonstration Center for Experimental Fisheries Science education at Shanghai
11 10 Ocean University, Shanghai, China, ²Key Laboratory of Exploration and Utilization of
12 11 Aquatic Genetic Resources (Ministry of Education) and International Research Center for
13 12 Marine Biosciences (Ministry of Science and Technology) at Shanghai Ocean University,
14 13 Shanghai, China, ³Laboratory for Marine Biology and Biotechnology, Qingdao National
15 14 Laboratory for Marine Science and Technology, Qingdao, China, ⁴Kunming Institute of
16 15 Zoology, Chinese Academy of Sciences, Kuming, China, ⁵Department of Animal Biology,
17 16 University of Illinois at Urbana-Champaign, IL, USA, ⁶Fundación Ictiológica, Providencia,
18 17 Santiago, Chile.

19 18
20 19 * These authors contributed equally to this work.

21 20
22 21 [§]Corresponding authors:

23 22 Qianghua Xu (qhxu@shou.edu.cn),

24 23 Chi-Hing Christina Cheng (c-cheng@illinois.edu).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

26 **Abstract**

27 The Southern Ocean is the coldest ocean on Earth but a hotspot of evolution. The
28 bottom-dwelling Eocene ancestor of Antarctic notothenioid fishes survived polar marine
29 glaciation and underwent adaptive radiation forming >120 species that fill all water column
30 niches today. Genome-wide changes enabling physiological adaptations and rapid expansion
31 of the Antarctic Notothenioids remain poorly understood. To advance our understanding, we
32 sequenced and compared two notothenioid genomes - the cold-adapted and neutrally buoyant
33 Antarctic toothfish *Dissostichus mawsoni*, and the basal Patagonia robalo *Eleginops*
34 *maclovinus* representing the temperate ancestor. We detected >200 protein gene families that
35 had expanded and thousands of genes that had evolved faster in the toothfish, with diverse
36 cold-relevant functions including stress response, lipid metabolism, protein homeostasis and
37 freeze resistance. Besides AFGP, an eggshell protein had functionally diversified to aid in
38 cellular freezing resistance. Genomic and transcriptomic comparisons revealed proliferation
39 of Selcys-tRNA genes and broad transcriptional upregulation across anti-oxidative
40 selenoproteins, signifying their prominent role in mitigating oxidative stress in the
41 oxygen-rich Southern Ocean. We found expansion of transposable elements, temporally
42 correlated to Antarctic notothenioid diversification. In addition, the toothfish exhibited
43 remarkable shifts in genetic programs towards enhanced fat cell differentiation and lipid
44 storage, and promotion of chondrogenesis while inhibiting osteogenesis in bone development,
45 collectively contributing to achieving neutral buoyancy and pelagicism. Our study revealed a
46 comprehensive landscape of evolutionary changes essential for Antarctic notothenioid cold
47 adaptation and ecological expansion. The two genomes are valuable resources for further
48 uncovering mechanisms underlying the spectacular notothenioids radiation driven by the
49 coldest environment.

50
51
52 Key words: adaptive radiation, climate change, genome, oxidative stress, Antarctic
53 notothenioids

54 **Introduction**

55
56
57
58 The Southern Ocean (SO) surrounding Antarctica is the coldest body of water on Earth,
59 having been isolated from other world oceans by the Antarctic Circumpolar Current (ACC)

1 60 beginning in the early Oligocene ~32 million years ago (1). The formation of the ACC also
2 61 impeded species dispersal across the Antarctic Polar Front, and mass extinction of the
3
4 62 Antarctic-sequestered fish taxa occurred upon marine glaciation (2). The rich cosmopolitan
5
6 63 fish fauna prior to the isolation of Antarctica is represented today by a single predominant
7
8 64 group of related fish species - the Antarctic notothenioids. From a common temperate
9
10 65 ancestor, likely a swim-bladderless, bottom dwelling perciform species of Eocene age (3), the
11
12 66 Antarctic notothenioids have evolved to become highly adapted to life in unyielding cold,
13
14 67 spectacularly diverse in sizes and morphological innovations, and diversified into all water
15
16 68 column habitats, epitomizing an adaptive radiation and a rare marine species flock (4).
17
18 69 Abundant in fish biomass (>90% of catch) and species (≥ 128), they are vital in sustaining the
19
20 70 contemporary SO food web (2, 5).

21 71 What evolutionary processes and mechanisms propelled the Antarctic notothenioid
22
23 72 radiation replete with extraordinary trait diversification during its evolutionary history remain
24
25 73 fascinating unanswered questions. Two conspicuous trait outcomes - the evolutionary gain of
26
27 74 the novel antifreeze glycoprotein (AFGP) gene and function that averted otherwise
28
29 75 inescapable death from freezing (6, 7), and exploitation of open niches vacated by extinction
30
31 76 of fishes lacking freeze-resistance, have been recognized as major contributors to the
32
33 77 Antarctic notothenioid radiation (8, 9). However, little is known of the myriad subtler
34
35 78 adaptive changes that must also have evolved in response to challenges from freezing
36
37 79 temperatures and the associated high oxygen concentration – the two foremost modalities of
38
39 80 selection pressure from the SO environments that would pervade all levels of organismal
40
41 81 functions, from molecules to cells to system physiology. Another prominent hallmark of
42
43 82 notothenioid adaptive radiation is the secondary acquisition of pelagicism in some lineages,
44
45 83 enabling their ecological expansion from bottom habitats of their negatively buoyant ancestor
46
47 84 to upper water column niches. What evolutionary changes occurred in the cellular and
48
49 85 developmental programs that enabled neutral buoyancy and secondary pelagicism are also
50
51 86 unknown.

52 87 To address these fundamental, system-wide questions about Antarctic notothenioid
53
54 88 evolution, whole genome sequences of multiple and appropriately chosen species from the
55
56 89 diverse Antarctic notothenioids are essential. Thus far, whole genome sequence analysis has
57
58 90 been reported for only one notothenioid species, the Antarctic rockcod *Notothenia coriiceps*,
59
60 91 providing the key inference that the fast evolving hemoglobin and mitochondrial proteins are
61
62 92 adaptive in increasing efficiency of aerobic cellular respiration in the freezing environment
63
64 93 (10). *N. coriiceps* is not known to occur in the high latitude Antarctic coastal waters, but is
65

1 94 widely distributed in the lower latitude waters of the Antarctic Peninsula archipelago and the
2 95 Scotia Arc islands, reaching localities north of the Polar Front around sub-Antarctic islands in
3
4 96 the Indian Ocean sector (11), a distribution pattern that suggests a considerable degree of
5
6 97 thermal plasticity in this species. It is a heavy, bottom fish and one of the hardest boned
7
8 98 Antarctic notothenioids (12), reminiscent of the benthic ancestor. To gain insights into
9
10 99 evolutionary adaptations in the most cold-adapted and stenothermal Antarctic notothenioids,
11
12 100 as well as into the evolutionary changes leading to acquisition of neutral buoyancy that
13
14 101 enabled the transition from the ancestral benthic existence to a pelagic life history, a different
15
16 102 and more appropriate model Antarctic notothenioid species would be required.

17 103 The Antarctic toothfish *Dissostichus mawsoni* that grows to giant sizes (2.0 m in length
18 104 and 140 kg in mass) is an iconic species of the Antarctic notothenioid radiation, with wide
19 105 distributions in freezing waters of high latitude Antarctic coasts, as far south as 77.5°S
20 106 (McMurdo Sound), the southern limit of Antarctic marine life. It thus exemplifies the
21 107 stenothermal cold-adapted character state. Despite its large size, it is the only notothenioid
22 108 species that achieved complete neutral buoyancy as adults (13, 14); thus this species serves as
23 109 the best model for examining the evolutionary underpinning of secondary pelagicism in the
24 110 Antarctic clade. In addition, to discern evolutionary changes from the ancestral temperate
25 111 state to the derived polar state driven by selection in the cold, oxygen-rich Southern Ocean
26 112 environment, a closely related basal non-Antarctic notothenioid comparison species that can
27 113 serve as ancestral proxy would strongly add to the discriminating power of analyses of
28 114 genome evolution. The most appropriate species for this purpose is a South American
29 115 notothenioid, the Patagonia robalo *Eleginops maclovinus*, which is the sole species in the
30 116 basal family Elegendinopsidae. The lineage diverged prior to the isolation of Antarctica, and *E.*
31 117 *maclovinus* is phylogenetically the closest sister species to the Antarctic clade (3). Thus its
32 118 genome is the best representative of the temperate character of the most recent common
33 119 ancestor of the Antarctic notothenioids. We conducted genome sequencing and comparative
34 120 analyses of these two strategically selected species accompanied with extensive
35 121 transcriptomic characterizations to profile relevant functional outcomes of the genomic
36 122 changes. Our results provide several new key insights into evolutionary adaptation and
37 123 secondary pelagicism of the Antarctic notothenioids in the isolated and extremely cold
38 124 Southern Ocean environment.

39 125

40 126

41 127 **Results and Discussion**

128 ***D. mawsoni* and *E. maclovinus* genome sequencing and assembly**

129 The geographic distributions and sampling locations of *D. mawsoni* and *E. maclovinus* are
130 illustrated in **Fig. 1a**. The genome of one *D. mawsoni* juvenile (12 kg) of undetermined sex,
131 and one young adult male *E. maclovinus* (~100 gm) were *de novo* sequenced using Illumina
132 sequencing platforms. The genomes of both species comprise 24 pairs of chromosomes ($2n =$
133 48) (15, 16). Analyses of 17-mer frequency distribution indicated a genome size of
134 approximately 842 Mb for *D. mawsoni* and 727 Mb for *E. maclovinus* (**Additional file 1: Fig.**
135 **S1**), consistent with the mean genome sizes of 840 Mb and 780 Mb for the toothfish and
136 robalo respectively determined by flow cytometry (**Additional file 1: Table S1a**). The raw
137 sequence data after cleaning and error correction (**Additional file 1: Tables S1b, S1c**) were
138 assembled using SOAPdenovo (17) for *D. mawsoni*, and Platanus (18) for *E. maclovinus*
139 followed by scaffold building with SSPACE (19). The assembled toothfish genome had a
140 contig N50 length of 23.1 Kb and scaffold N50 length of 2.2 Mb, while those of the robalo
141 were 10.9 Kb and 0.69 Mb. The assembled toothfish and robalo genomes are approximately
142 757 Mb and 744 Mb respectively (**Table 1; Additional file 1: Tables S2a, S2b**), consistent
143 with *k*-mer and flow cytometry estimates, and achieving over 90% and 95% coverage of the
144 genome size based on flow cytometry of the two species respectively. The completeness of
145 both genomes were assessed with BUSCO (Benchmarking Universal Single-Copy Orthologs)
146 (20), referencing the lineage dataset of actinopterygii_odb9 and orthologs of zebrafish, which
147 reflected the complete BUSCOs at 97.2% for the *D. mawsoni* genome and 95.0% for the *E.*
148 *maclovinus* genome (**Additional file 1: Tables S2c**). The GC content of the *D. mawsoni*
149 genome is 0.4070, nearly identical to 0.4066 of *E. maclovinus*, and both are lower than that of
150 a model fish the stickleback *Gasterosteus aculeatus* (**Additional file 1: Fig. S2**). The
151 accuracy of the genome assembly was assessed by alignment of the scaffolds to publically
152 available unigenes of *D. mawsoni* and *E. maclovinus*, and the coverage of the initial contigs
153 was found to be approximately 98.8% and 99.1%, respectively (**Additional file 1: Table S3**),
154 suggesting an acceptable quality of the genome assemblies. Alignment of the sequence reads
155 to the assemblies estimated an overall heterozygous rate of approximately 2.58 and 2.40 per
156 Kb for *D. mawsoni* and *E. maclovinus* respectively. (**Additional file 1: Table S4**).

157

158 **Genome annotation and synteny alignment between *D. mawsoni* and *E. maclovinus***

159 A total of 22,516 and 22,959 protein-coding genes were annotated in the *D. mawsoni* and *E.*
160 *maclovinus* genome, respectively by combining the results from homologous and *de novo*
161 prediction methods using the gene modeler GLEAN (**Additional file 1: Table S5**). The

162 protein coding genes of the toothfish and robalo, along with the sequenced notothenioid *N.*
163 *coriiceps* and model species *G. aculeatus* and zebrafish *Danio rerio* were clustered using
164 OrthoMCL(21). We found 8,825 genes that were common to all five species. Genes shared
165 among the notothenioids are similar in number, 12,269 between toothfish and robalo, and
166 12,421 between toothfish and *N. coriiceps* (**Additional file 1: Fig. S3**). In annotations of
167 conserved non-coding RNA genes, we predicted 1,097 tRNA, 110 rRNA, 422 SnRNA, and
168 295 microRNA genes in the toothfish genome (**Additional file 1: Table S6a**), while the
169 robalo genome was annotated to carry 1,037 tRNA, 44 rRNA, 891 snRNA and 286 miRNAs
170 (**Additional file 1: Table S6b**). The much larger number of rRNA copies (2.5 fold) in *D.*
171 *mawsoni* than *E. maclovinus* is consistent with the presence of dual chromosomal loci of
172 rDNA genes detected by *in situ* fluorescent hybridization in the giant toothfish (15), as
173 opposed to the single rDNA locus in other notothenioids (22, 23). The two species showed
174 largely similar profiles in their microRNAome, with minor differences in the copy number of
175 some individual microRNA (**Additional file 1: Table S7**). Strikingly, the toothfish genome
176 contains many more selcys-tRNA genes than robalo (84 versus 1; **Additional file 1: Table**
177 **S8**). This extensive duplication of selcys-tRNA genes, accompanied with high expression of
178 selenoproteins in *D. mawsoni* (detailed in a later section) signify that mitigation of oxidative
179 stress through selenoproteins, many of which are strong antioxidants, is likely an important
180 selection force in the evolution of the Antarctic notothenioid genome in the freezing and
181 oxygen-rich waters.

182 *De novo* annotation of repeat sequences revealed >2-fold increase in overall repeat
183 content in *D. mawsoni* (28.29% of genome) over the basal *E. maclovinus* (13.40%)
184 (**Additional file 1: Table S9**). Transposable elements (TEs) including LTR-retrotransposons,
185 non-LTR-retrotransposons (long interspersed transposable elements - LINES, and short
186 interspersed transposable elements - SINEs), and DNA transposons accounted for a total of
187 21.38% of the toothfish genome, more than twice that of *E. maclovinus* (10.02%). Among
188 simple sequence repeats (SSRs), dimer repeats constituted the majority of the SSRs in both
189 genomes, and tetramers and pentamers showed highest levels of increment in *D. mawsoni*
190 (**Additional file 1: Table S10a, Table S10b**). The doubling of TE content in the toothfish
191 genome relative to the basal robalo suggests it is a likely contributing factor to the observed
192 trend of increasing genome sizes in more derived Antarctic notothenioid lineages (24).

193 For global genome alignment, we anchored the *D. mawsoni* and *E. maclovinus* scaffolds
194 to the 21 linkage groups of the well characterized three-spined stickleback genome according
195 to gene collinearity (**Fig. 1b; Additional file 2**). Most of the notothenioid scaffolds had

196 extensive collinearity to the corresponding stickleback chromosomes. The 84 duplicated
197 selcys-tRNA genes are wide-spread throughout the toothfish genome. Inversion and
198 translocation of genome segments have occurred in both *D. mawsoni* and *E. maclovinus*
199 relative to stickleback, but the *D. mawsoni* genome showed more frequent rearrangements.
200 Insertion sites for the expanded TEs of the *D. mawsoni* genome were random, thereby
201 expanding the lengths of almost all of the linkage groups. Mapping RNAseq transcript data
202 we obtained from white muscles for the two species on their respective synteny showed that
203 the heavier insertion of TEs in the toothfish genome did not appear to adversely affect the
204 expression of the neighboring protein genes as strong gene expression in regions of high TE
205 content was maintained (**Fig. 1b**).

207 **Burst of LINE expansion in the cold**

208 Involvement of gain (or loss) of mobile element copies in genome size and genome
209 re-structuring affecting species differentiation has increasingly gained empirical support (25,
210 26). TEs are normally under epigenetic regulation, but waves of TE proliferations could arise
211 from environmental changes that cause physiological stress and disrupt the epigenetic control
212 (27). We therefore examined potential linkage in timing between LINE expansion (2-fold
213 increase in the toothfish versus robalo) and onset of frigid SO marine conditions. A multiple
214 sequence alignment of several LINEs (LINE/I, LINE/L2, LINE/RTE-BovB and
215 LINE_Rex-Babar – the most abundant types in the toothfish genome) was made, which
216 clustered 349 *D. mawsoni* LINE pairs and 213 *E. maclovinus* LINE pairs, where the two
217 LINEs within each pair share the highest sequence similarity, approximating the least
218 divergence time. Calculation of the nucleotide substitution rates for each LINE pair identified
219 a burst of emergence of LINEs in the *D. mawsoni* genome with the substitution rates centered
220 at 0.04, which corresponded with a divergence time of 6.5 million years ago (**Fig. 2a; see also**
221 **Methods**) based on the average substitution rate of the LINEs. This estimated timing of LINE
222 burst expansion correlated with the radiation of the majority of the modern Antarctic
223 notothenioid clades beginning in the late Miocene when seawater temperatures steadily
224 declined (9). In contrast, no burst expansion of LINEs was detected in the *E. maclovinus*
225 genome, supporting an Antarctic/cold-specific LINE burst in the toothfish, and corroborating
226 our prior empirical evidence for an increase in retrotransposition activity of *D. mawsoni*
227 LINE-1 resulting in more copies in the transfected cells when subjected to stress from
228 non-physiologically low incubation temperature (28).

230 **Accelerated protein evolution in the cold**

231 Ectotherms are vulnerable to disruptive effects on protein structures and reaction rate
232 depression at low temperatures. Antarctic notothenioids in perennially freezing high latitude
233 waters face the extremes of these effects, as well as an oxidative environment due to high
234 oxygen concentrations resulting from increased gas solubility at low temperatures. We
235 examined evidence for adaptive evolution of proteins in response to these selection pressures.
236 The gene models of *D. mawsoni*, *E. maclovinus*, seven other teleost genomes and the mouse
237 genome as outgroup were clustered, from which 2,936 one-to-one single-copy orthologs were
238 obtained to reconstruct a phylogenetic tree (**Fig. 2b**). Stickleback is the closest species to the
239 notothenioid clade as expected, and *E. maclovinus* is sister to the two Antarctic notothenioids,
240 *D. mawsoni* and *N. coriiceps*. We found evolutionary rates of the orthologous genes based on
241 calculated *dN/dS* values (the ratio of the rate of non-synonymous substitution to the rate of
242 synonymous substitution) were elevated in the notothenioid lineage compared to the other
243 fish lineages. This faster rate is more pronounced in the two Antarctic species, about twice
244 that of the basal *E. maclovinus*, suggesting intensified selection pressures driving genome
245 evolution in the Antarctic environment. To identify Gene Ontology categories that were
246 evolving faster in the toothfish or robalo, the *dN/dS* ratios of 7,584 orthologous genes among
247 the two notothenioids and the stickleback were calculated and the average *dN/dS* value of the
248 genes associated with each GO term was calculated for each species. These orthologous genes
249 were annotated to 411 GO terms, of which 281 showed significantly higher average *dN/dS*
250 ratios in *D. mawsoni*, while only 19 demonstrated higher average rates in *E. maclovinus* (**Fig.**
251 **2c; Additional file 2**). The faster evolving GO processes in *D. mawsoni* included “gene
252 expression”, “protein folding”, “tRNA metabolic process”, “cell-redox homeostasis”,
253 “immune response”, “response to stress”, “lipid metabolic process”, “DNA repair”,
254 “vesicle-mediated transport” and others. To assess which *D. mawsoni* genes experienced
255 positive selection, we tested using the Branch-site model (in PAML) on a reconstructed
256 phylogenetic tree of six fish species with the *D. mawsoni* lineage assigned as the foreground
257 branch (**Additional file 1: Fig. S4**). A total of 526 positively selected genes (PSGs) were
258 identified (**Additional file 1: Table S11a**). The most significantly enriched KEGG pathway
259 (**Additional file 1: Table S11b**) was “Protein digestion and absorption” ($p = 5.1 \times 10^{-5}$) and
260 GO term (**Additional file 1: Table S11c**) was “protein binding” ($p = 6.0 \times 10^{-4}$), which
261 indicate that maintenance of protein homeostasis played an important role in shaping the *D.*
262 *mawsoni* genome. In addition, a complement of genes (*sorbs2a*, *acox1*, *apoA1a*, *scp2a*, *tnni3k*
263 and perilipin-like genes; **Additional file 1: Table S11a**) involved in the PPAR (peroxisome

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
264 proliferator-activated receptors) signaling pathway - the key pathway in adipocyte
265 development regulation - were found to be under positive selection, suggesting occurrence of
266 adaptive changes in lipid metabolism in *D. mawsoni*.

267 268 **Gene duplication in the freezing environment**

269 Gene duplication plays fundamental roles in emergence of adaptive features. In the list of
270 predicted protein coding genes from the toothfish and robalo genomes, we identified 202
271 families that have increased in copy number in *D. mawsoni*, compared to the other eight fish
272 genomes (**Additional file 3**). KEGG enrichment analyses of these expanded gene families
273 yielded enrichment in pathways involved in protein homeostasis and lipid and bone
274 metabolism, such as “Protein digestion and absorption”, “Regulation of lipolysis in
275 adipocyte”, “Fat digestion and absorption”, “Ether lipid metabolism” and “Osteoclast
276 differentiation” (**Fig. 2d; Additional file 2**), suggesting genomic capacity for these functional
277 pathways had increased in *D. mawsoni* during evolution in chronic cold. Corroborating this
278 cold-specificity is that the expanded gene families found in the basal *E. maclovinus* relative to
279 the other fish genomes, including *D. mawsoni* and *N. coriiceps*, yielded distinctly different
280 enriched KEGG profiles (**Fig. 2e**); this analysis indicates that the functional traits gained
281 through gene duplication in *E. maclovinus* were driven by different selective pressures,
282 consistent with our previous findings (29, 30). Due to inherent inefficiency in correctly
283 assembling highly similar DNA sequences in the shotgun sequencing strategy, there are likely
284 many more duplicated genes that had eluded detection. For example, many paralogs of Zona
285 Pellucida protein (ZPs), such as ZPAX1, ZPC1, ZPC2 in *D. mawsoni* have been shown,
286 through array-based genome hybridization and quantitative PCR, to be duplicated (29, 31).

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
287 Gene duplication had contributed prominently to the evolutionary gain of freezing
288 avoidance in Antarctic notothenioids. Generally regarded as a key innovation of the Antarctic
289 notothenioid radiation, the AFGP (antifreeze glycoprotein) gene evolved from a
290 trypsinogen-like protease ancestor followed by extensive intragenic and whole gene
291 duplications, generating a large gene family that would provide an abundance of this novel
292 life-saving protein (7). By referencing to the published AFGP haplotypes that were assembled
293 from BAC clone sequences of the same individual used in this study (Genbank accessions
294 HQ447059 and HQ447060) (32), we identified and assembled the AFGP loci shotgun reads,
295 recaptured the two AFGP haplotypes and integrated them in the draft genome, and localized
296 them to a region syntenic with a scaffold in LG 20 of the stickleback genome (**Fig.1b**).

297 Resistance to freezing extends beyond the AFGPs: we found gene duplication for a
298 protein that *a priori* would not be expected to function in freeze-resistance, zona pellucida
299 protein (or eggshell protein), additionally provided protection against cellular freezing.
300 Products of *ZPC5* from *D. mawsoni* have been shown to enhance freezing resistance of eggs
301 of recipient zebrafish both *in vivo* and *in vitro* (31). In this study, we found four copies of
302 *ZPC5* in the *D. mawsoni* genome that encoded three different sized *ZPC5* proteins in the
303 toothfish ovary - DmZPC5_1, DmZPC5_2a/DmZPC5_2b, and DmZPC_3 in decreasing order
304 of size, corresponding to gradually shortened C-termini from the conserved ZP domain due to
305 nonsense mutations in exon 9 and exon 10 (**Fig. 3a; Additional file 1: Fig. S5a**). A single
306 *ZCP5* gene in the basal *E. maclovinus* was found, which corresponded to the full-length
307 DmZPC5_1. We expressed the three DmZPC5 isoforms in Chinese Hamster Ovary (CHO)
308 cells (**Additional file 1: Fig. S5b, Fig. 3b**) and assayed cell survival rate at freezing
309 temperature (-2°C for 8 hrs). DmZPC5_3 was the most active isoform in maintaining cell
310 viability while DmZPC5_1 was the least active one (**Fig. 3c**). Further analyses showed that
311 DmZPC5_3 was more likely retained inside the cell than DmZPC5_2 and DmZPC5_1
312 (**Additional file 1: Fig. S6**), and less likely to become polymerized inside the cell compared
313 with DmZPC5_1 (**Additional file 1: Fig. S7a**), corroborating our previous finding that only
314 unpolymerized ZP proteins are active for the ice-melting promoting activity (31). We detected
315 DmZPC5 expression across many tissues besides ovary in *D. mawsoni* (**Additional file 1: Fig.**
316 **S7b**), consistent with a distribution expected of a general protective function.

318 **Transcriptomic adaptation to the cold environment**

319 To assess the functional relevance of the detected genomic outcomes to life in freezing
320 condition, we characterized and compared transcriptomes of 12 tissues including brain, liver,
321 red muscle, white muscle, gill, skin, intestine, stomach, spleen, head kidney, caudal kidney
322 and ovary between native specimens of *D. mawsoni* and *E. maclovinus*. We found over ten
323 thousand genes were differentially expressed in pairwise comparisons between the two
324 species, with the toothfish showing substantially more up-regulated genes than the robalo.
325 Enrichment test on KEGG pathways yielded many signaling pathways in the tissues of
326 toothfish being significantly enriched in differentially expressed genes (DEGs), including the
327 TGF-beta, AMPK and PPAR pathways, known to play essential roles in development,
328 metabolism and stress responses (**Additional file 1: Fig. S8a**). GO enrichment analysis of the
329 DEGs demonstrated up-regulation of hundreds of GO biological processes, including
330 translation, transferrin transport, cell redox homeostasis, cellular response to unfolded protein,

1 331 ubiquitin-dependent protein catabolic process, regulation of innate immunity, MAPK cascade,
2 332 positive regulation of apoptosis pathways, many of the pathways involved in lipid metabolism,
3 333 and anti-ROS pathways represented by “selenium compound metabolic process” and
4 334 “selenocysteine metabolic process” (**Additional file 1: Fig. S8b**). These results corroborated
5 335 the expression profiles in a previous study with the lower depth of sequencing available at
6 336 that time (29), and additionally revealed further details of transcriptomic cold adaptation from
7 337 the much deeper sequencing across a comprehensive set of tissue transcriptomes. A striking
8 338 finding was the greatly increased transcriptional activities across many selenium-containing
9 339 protein genes in the toothfish tissues compared with robalo (**Fig. 4; Additional file 1: Fig.**
10 340 **S9**). Correspondingly, the genes involved in the translation of selenocysteine-containing
11 341 proteins were also significantly up-regulated in the toothfish (**Fig. 4**). Selenoproteins have
12 342 well-known functions in coping with cellular oxidative stress, thus the great expansion of
13 343 selcys-tRNA genes (**Fig. 1b, Additional file 1: Table S8**), and the significantly upregulated
14 344 expression of many kinds of selenoprotein mRNAs indicate augmented anti-ROS capacity has
15 345 evolved as an important adaptation to the constantly freezing environment, where saturated
16 346 levels of O₂ and cold-depressed metabolic rates would make oxidative stress a formidable
17 347 challenge for cellular life. Interestingly, the expression of glutathione peroxidase 4b (gpx4b)
18 348 (**Fig. 4**), a selenoprotein uniquely able to reduce lipid hydroperoxides (33, 34), was lower in
19 349 *D. mawsoni*, suggesting alternative lipid metabolic programs may exist in the toothfish.
20 350 Accordingly, we found all isoforms of the major players in lipid droplet assembly (PLN2,
21 351 PLN5, fitm, seipin) important for lipid storage in adipose tissue, were upregulated in all
22 352 toothfish tissues examined relative to *E. maclovinus*, signifying a shift of lipid distribution
23 353 towards storage in *D. mawsoni*, as described in detail below.

24 354

25 355 **Altered lipid metabolism in *D. mawsoni* for neutral buoyancy**

26 356 In a striking evolutionary departure from the heavy, bottom-water ancestral character
27 357 (exemplified by *E. maclovinus*), a handful of Antarctic notothenioids have secondarily
28 358 acquired neutral or near neutral buoyancy, enabling ecological diversification into and filling
29 359 of mid-water niches – a distinctive hall mark of the Antarctic notothenioid adaptive radiation.
30 360 The giant toothfish *D. mawsoni*, despite growing to massive sizes, being robustly muscled,
31 361 and lacking a swim bladder, is the only notothenioid that has attained complete neutrally
32 362 buoyancy (13, 14). Known morphological specializations include extensive lipid (mostly
33 363 triglycerides) deposits under skin and in the musculature, and a light skeleton of mostly
34 364 cartilage and little mineralized bone, adaptations that reduce overall density and provide static

1 365 lift (13). To understand the genetic basis of the large accumulation of lipids and reduced
2 366 mineralization in *D. mawsoni*, we carried out transcriptome comparisons between *D. mawsoni*
3
4 367 and several other notothenioids in which neutral buoyancy is not developed.

5 368 We compared gene expression profiles of muscles of *D. mawsoni*, and of the negatively
6
7 369 buoyant Antarctic *N. coriiceps* and the basal *E. maclovinus* to elucidate evolutionary
8
9 370 differences in the mechanisms of intermuscular lipid deposit. Compared with *E. maclovinus*,
10
11 371 genes involved in triacylglycerol synthesis in the toothfish muscle were overrepresented in
12
13 372 DEGs ($p < 0.05$) and markedly upregulated in transcription, including the key enzymes
14
15 373 acylglycerol-3-phosphate O-acyltransferase (AGPAT) isoforms and CDP-diacylglycerol
16
17 374 synthase (CDS) (**Additional file 1: Fig. S10**). An important regulator of this process, lipin1,
18
19 375 was downregulated in *D. mawsoni*. Lipin1 is known to exert dual effects on lipid metabolism
20
21 376 - it acts as a phosphatidate phosphatase enzyme to form diacylglycerol required for lipid
22
23 377 synthesis, but also serves as a transcriptional co-activator to promote fatty acid oxidation (35).
24
25 378 The down-regulation of lipin1 was consistent with down-regulation of fatty acid oxidation in
26
27 379 *D. mawsoni*, as about half of the genes involved in fatty acid oxidation were also
28
29 380 down-regulated relative to the robalo (**Additional file 1: Table S12**). At the same time, genes
30
31 381 involved in regulation of lipid storage were overrepresented in the DEGs ($p < 0.05$) and all
32
33 382 but one gene (MEST) were upregulated in the toothfish (**Additional file 1: Table S13**). These
34
35 383 data strongly suggest a shift of metabolic pathways from lipid breakdown to lipid biosynthesis
36
37 384 and lipid storage in *D. mawsoni* muscle relative to *E. maclovinus*, favoring deposit of lipids,
38
39 385 thus contributing to neutral buoyancy. Compared with *N. coriiceps*, *D. mawsoni* muscle
40
41 386 showed an overall trend of upregulation of genes involved in glycerolipid biosynthesis and
42
43 387 lipid storage, but the differences are not as dramatic as in the *D. mawsoni/E. maclovinus*
44
45 388 comparison. Expression levels of many genes relevant to lipid oxidation were fairly similar
46
47 389 between the two species, suggesting common downregulation in lipid oxidation in the
48
49 390 Antarctic species compared to the temperate *E. maclovinus* (**Additional file 1: Fig. S10**;
50
51 391 **Additional file 1: Tables S12, S13**). Consistent with downregulation of the lipid oxidation in
52
53 392 the muscles of the two Antarctic fishes were their lower expression levels of lipid oxidation
54
55 393 mitigating selenoenzyme gpx4 compared to the robalo. In total, the transcriptome
56
57 394 comparisons revealed substantial genetic reprogramming in *D. mawsoni* muscle that would
58
59 395 favor the large lipid deposition in this species.

60
61 396 GO enrichment tests on the muscle DEGs also indicated regulatory change ($p = 0.073$) in
62
63 397 fat cell differentiation between *D. mawsoni* and *E. maclovinus*. Adipogenesis in almost all
64
65 398 animals is predominately regulated by peroxisome proliferator activated receptor gamma

399 (PPAR γ)(36). Expression of PPAR γ in *D. mawsoni* muscle was upregulated by more than
400 5.6-fold compared to *E. maclovinus* (**Additional file 1: Table S14**). In addition, as many as
401 16 known pro-adipogenic factors were upregulated from 1 to 8 fold. Several factors in the
402 TGF-beta (TGFB1, smad3), Wnt (Sirt1, sirt2, frizzled-related protein (FRZB)) and Notch
403 (jag1b, Hes1) pathways and Jun dimerization protein 2 (JDP2), reportedly negative regulators
404 of adipogenesis were also upregulated. When compared with *N. coriiceps*, about half of the
405 DEGs in the *D. mawsoni/E. maclovinus* comparison under this GO term were statistically
406 insignificant, but remarkably the comparison yielded significant upregulation of 10
407 pro-adipogenic factors in *D. mawsoni* muscle, including the most important regulator,
408 PPAR γ (**Fig. 5a**), and only one negative regulating factor (TGFB1) (37, 38). These results
409 strongly suggest that regulatory promotion of adipogenesis in *D. mawsoni* muscle is a key
410 contributing factor to fat deposition and attainment of neutral buoyancy.

411

412 **Reduction of ossification in *D. mawsoni***

413 To reveal the genes involved in the reduced bone ossification in *D. mawsoni*, we compared
414 the transcriptomes of pelvic girdle bones between *D. mawsoni* (0% body weight in seawater)
415 and the negatively buoyant Antarctic notothenioids *Trematomus bernacchii* and *Pagothenia*
416 *borchgrevinki*, (3.52% and 2.75% of body weight, respectively, in seawater) (13). A total of
417 1,733 DEGs showing the same direction of change in the *D. mawsoni/P. borchgrevinki* and
418 the *D. mawsoni/T. bernacchii* comparisons were identified and used for further analysis
419 (**Additional file 3**). We found that 48 genes encoding various ribosomal proteins were
420 significantly reduced in expression in the *D. mawsoni* bone, suggesting either a lower protein
421 translation activity or fewer metabolically active cells in the pectoral girdle than in the other
422 two nototheniids. The DEGs were enriched with hundreds of GO biological processes,
423 including “extracellular matrix”, “ossification”, “response to hypoxia”, “angiogenesis” and
424 “lipid storage”, indicating multiple genetic programs were distinctly regulated in the toothfish
425 bone (**Additional file 1: Fig. S11**). In terms of ossification, it was noteworthy that
426 expressions of the two major regulators of vertebrate bone development, *sox9* and *runx2* (39)
427 were not significantly altered among the three notothenioids. However, expression of many
428 genes of the BMP pathways, wnt pathways and many regulatory factors known to be involved
429 in the process were specifically altered in *D. mawsoni*, which likely shifted the developmental
430 balance between chondrogenesis and osteogenesis (**Fig. 5b; Additional file 1: Table S15**).
431 Among these highly upregulated genes (depicted in **Fig. 5b**), CTGF (Connective Tissue
432 Growth Factor) has been implicated in early events of osteogenic differentiation including

1 433 proliferation and recruitment of osteoprogenitors, however, when expressed constitutively,
2 434 CTGF would inhibit both Wnt-3A and BMP-9 induced osteoblast differentiation (40). HSPG2
3
4 435 (prostaglandin-endoperoxide synthase 2) is required for the chondrogenic and adipogenic
5
6 436 differentiation from synovial mesenchymal cells via its regulation of *sox9* and PPAR γ , but not
7
8 437 for osteogenic differentiation via *runx2* (41), and ECM1 (extracellular matrix protein 1)
9
10 438 interacts with HSPG2 to regulate chondrogenesis (42). MEF2C, a transcription factor that
11
12 439 regulates muscle and cardiovascular development, controls bone development by activating
13
14 440 the genetic program for chondrocyte hypertrophy (43). Some of the upregulated genes are
15
16 441 known to inhibit osteoblastogenesis, such as *Tob2* (44), *CTNNBIP1* (45), secreted
17
18 442 frizzled-related protein 1 (*SFRP1*) (46) and *ZBTB16* (47). Some members of the TGF-beta
19
20 443 superfamily (*BMPR1a*, TGF-beta1, *SMAD1*), which were upregulated in *D. mawsoni* are
21
22 444 known to promote both chondrogenesis and osteoblastogenesis (48). *CYR61* and *PTN*
23
24 445 specifically promote osteoblastogenesis (49, 50), but expression of *PTN* is drastically reduced,
25
26 446 consistent with reduced hard bone formation. A few genes influence ossification via
27
28 447 regulating osteoclastogenesis, for example, *Sbno2*-promotes osteoclast fusion (51) and
29
30 448 activation of the EphA2 signaling on osteoblasts led to bone reabsorption(52). We found
31
32 449 genes associated with osteoclast differentiation are significantly enriched in the DEGs (p
33
34 450 <0.05) and all were upregulated (**Additional file 1: Tables S16, S17**). Overall, the gene
35
36 451 expression patterns in the toothfish bone demonstrated a genetic shift to chondrogenesis over
37
38 452 osteoblastogenesis in bone development, which would reduce bone density and contribute to
39
40 453 achieving neutral buoyancy.

41
42 454 Studies have indicated that the majority of clinical conditions associated with human
43
44 455 bone loss are accompanied by increased marrow adiposity possibly due to shifting of the
45
46 456 balance between osteoblast and adipocyte differentiation in bone marrow stromal (skeletal)
47
48 457 stem cells (53). A few signaling pathways such as the TGF-beta/BMP pathways and the Wnt
49
50 458 pathway (represented by *CTNNBIP* and *SFRP1* in this case) are known to participate in
51
52 459 regulation of both bone and adipocyte development in animals. In the toothfish, we found
53
54 460 enriched GO terms relevant to regulation of “response to lipid” and “lipid storage”
55
56 461 (**Additional file 1: Table S12**) indicating possible linkage in the regulatory network that
57
58 462 orchestrates the loss of ossification and gain of lipids in *D. mawsoni* bones.

59
60 463 To verify whether the elevated transcription of the regulatory factors indeed resulted in
61
62 464 more abundant protein, we selected the factor CTGF for immunohistochemical staining in the
63
64 465 bone and surrounding tissues of pelvic fins of *D. mawsoni* and *E. maclovinus* since it is the
65
66 466 only factor for which an effective monoclonal antibody is currently available. Much stronger

1 467 signal was detected in the *D. mawsoni* fin tissue (Fig. 5c), supporting a correlation between
2 468 protein abundance and mRNA transcription in the case of CTGF. This result further supports
3
4 469 the involvement of CTGF in the reduced ossification in *D. mawsoni*.

5 470

7 471 **Conclusions**

9 472 We sequenced and compared the genomes and transcriptomes of the cold-adapted
10 473 high-latitude Antarctic toothfish *D. mawsoni* and the basal temperate relative *E. maclovinus*
11 474 representing the ancestral character state to deduce Antarctic-specific evolutionary and
12 475 adaptive changes supporting physiological activities of notothenioid fishes in freezing and
13 476 oxygen rich Southern Ocean waters, as well as the gain of secondary pelagicism fundamental
14 477 to Antarctic notothenioid niche expansion and adaptive radiation. The assembled genomes
15 478 achieved 90% (*D. mawsoni*) and 95% (*E. maclovinus*) coverage of the respective genome size
16 479 estimated by cell flow cytometry, and with greater scaffold N50 than the currently available
17 480 sole Antarctic notothenioid (*N. coriiceps*) genome, greatly enhancing comprehensive,
18 481 genome-wide discovery of evolutionary processes.

27 482 We found two-fold expansion of TEs in the Antarctic toothfish over the temperate robalo
28 483 *E. maclovinus* and deduced the timing of a burst of one major class of TEs (LINEs) to about
29 484 6.5 mya, temporally correlating with the late Miocene onset of steady cooling trend of the
30 485 Southern Ocean (SO) and diversification of the modern Antarctic notothenioid clade,
31 486 suggesting a role of cold-induced TE expansion in notothenioid speciation. We found many
32 487 of the protein coding genes in the toothfish evolved rapidly and experienced positive selection,
33 488 among which genes relevant to preservation of protein homeostasis were particularly
34 489 prominent. Multiple gene families have undergone duplication during evolution in the cold,
35 490 as exemplified by genes that confer resistance to freezing in the cold SO waters: the AFGP
36 491 family that evolved *de novo* and confers extracellular freeze avoidance, and duplicated zona
37 492 pellucida ZPC5 genes that functionally diversified to aid in cellular freezing resistance.
38 493 Through transcriptome comparisons, we found functional output of the cellular apparatus for
39 494 selenoprotein production in the Antarctic toothfish was greatly elevated compared to the basal
40 495 temperate robalo, suggesting evolutionary mobilization of antioxidant selenoproteins in
41 496 mitigating intensified oxidative stresses arising from the O₂-rich SO environment.

54 497 The evolutionary transition from the negatively buoyant ancestral character to complete
55 498 neutral buoyancy in the Antarctic toothfish entailed remarkable genetic reprogramming of fat
56 499 deposition and bone development. We found upregulation of processes of adipogenesis in
57 500 skeletal muscle, and triacylglycerol synthesis and fat storage were favored over fatty acid
58 501 oxidation. In bone development, a regulatory cascade favoring chondrogenesis over

1 502 osteoblastogenesis was especially evident. The shift in fat synthesis and storage, together with
2 503 reduction of ossification are therefore key in evolutionary gain of neutral buoyancy and
3 504 secondary pelagicism in *D. mawsoni*, and likely in the handful of other pelagic notothenioids,
4 505 allowing them to diversify into mid-water niches, a distinctive hallmark of the Antarctic
5 506 notothenioid adaptive radiation.

6 507 The remarkable diversification of Antarctic notothenioids (and several other polar fish
7 508 lineages) is integral to the conclusion from a recent analysis of latitudinal diversity gradient of
8 509 marine fishes that high-latitude cold water lineages exhibit exceptionally high rates of
9 510 speciation compared to tropical lineages, counter to expectation based on latitudinal species
10 511 richness (54). Rates of molecular evolution based on phylogenetic tree branch lengths are not
11 512 found to be slower at high latitudes (53). We have shown more definitively in this study that
12 513 in the cold adapted Antarctic notothenioid fish, evolutionary rates in fact accelerated in
13 514 thousands of protein coding genes, extensive cold-specific gene duplication and functional
14 515 diversification had occurred, such as the ZP protein gene families, and TE mobility was
15 516 remarkably elevated which likely contributing to the observed higher frequency of
16 517 chromosomal rearrangements. In mammals, ZP3 is known to function in sperm-egg
17 518 recognition (55) and TE activity is positively related to rate of speciation (56). How these
18 519 genomic and functional changes elicited by selective pressures from the cold SO temperatures
19 520 might have acted as intrinsic factors affecting notothenioid speciation are rich questions for
20 521 further investigation.

21 522 In summary, the results of this study provided robust new insights into genomic and
22 523 transcriptomic alterations enabling cold adaptation and niche expansion of the predominant
23 524 and ecologically vital Antarctic fish group in the SO. The genomes also serve as valuable
24 525 resources for future investigations of genomic and evolutionary changes in the diverse
25 526 Antarctic notothenioid families driven by paleoclimate changes in the SO, studies that may
26 527 shed light on questions about why the coldest ocean has been a hotspot of species formation.

27 528

28 529

29 530

30 531 MATERIALS AND METHODS

31 532

32 533 **Specimens, sampling, and DNA and RNA isolation.**

33 534 Antarctic toothfish *D. mawsoni* was collected using vertical setline through drilled hole in sea ice
34 535 of McMurdo Sound, Antarctica (77° 53'S, 166° 34.4'E and vicinity) during austral summer field
35 536 seasons (Oct.-Dec). Specimens were transported to the aquarium facility in the US National
36 537 Science Foundation Crary Lab at McMurdo Station and kept in ambient (-1.6°C) flow through

1 538 seawater tanks, and sacrificed at 2-4 weeks post capture for blood and tissue sampling. The
2 539 temperate basal notothenioid *E. maclovinus* was collected by rod and reel in the Patagonia waters
3
4 540 of southern Chile during austral winter (June) and transported to the NSF R/V Laurence Gould at
5
6 541 Punta Arenas in a large Styrofoam cooler of ambient water (~8°C), where they were sacrificed and
7
8 542 sampled within a few days prior to southbound transit for winter field season. Additional juvenile
9
10 543 specimens of *D. mawsoni* were collected by trawl from the waters of Antarctic Peninsula during
11
12 544 the same winter season, and sampled on ship board shortly after capture. The dissected carcasses
13
14 545 of *E. maclovinus* and juvenile *D. mawsoni* were kept frozen at -80°C, which provided the pelvic
15
16 546 bone samples for immunohistochemical detection for expression of candidate genes in bone
17
18 547 development. To preserve high molecular weight (HMW) DNA for genome sequencing, red blood
19
20 548 cells of each species were embedded in 1% melt agarose to provide about 20ug DNA per 90uL
21
22 549 block using BioRad plug molds (CHEF Mammalian Genomic DNA Plug Kit #1703591,
23
24 550 Bio-rad, USA), lysed *in situ* and preserved following Amemiya *et al* (1996) (57). To recover
25
26 551 HMW DNA, the agarose plugs were digested with β -agarase (NEB, USA) followed by phenol
27
28 552 extraction and dialysis, and quality verified using pulsed field electrophoresis. RNA for
29
30 553 transcriptome sequencing was isolated from -20°C ethanol preserved tissues using Trizol
31
32 554 (Invitrogen) and quality verified using an Agilent BioAnalyzer. Collection, handling and
33
34 555 sampling of the Antarctic toothfish and S. American robalo in this study were carried out in
35
36 556 compliance with protocol # 12123 approved by the University of Illinois Institutional Animal
37
38 557 Care and Use Committee (IACUC).

39 558

40 559 **Sequencing and genome assembly**

41 560 The sequencing libraries with insert sizes of 170, 250, 500 bp were prepared for sequence of
42
43 561 the paired-end reads, following a modified version of the manufacturer's protocol (Illumina).
44
45 562 An integrated protocol from the Mate-Pair Library v2 Sample Preparation Guide (Illumina)
46
47 563 and the Paired-End Library Preparation Method Manual (Roche) was used to prepare
48
49 564 mate-pair libraries with insert sizes of 3, 6, 10, 15 and 20 kb (**Additional file 1: Tables S1b,**
50
51 565 **S1c**). For the transcriptome sequencing, Poly(A)+ mRNA was purified using the DynaBeads
52
53 566 mRNA Purification kit (Life Technologies). Paired-end cDNA libraries were constructed
54
55 567 using the RNA-Seq NGS Library Preparation Kit for Whole-Transcriptome Discovery
56
57 568 (Gnomegen). All of the libraries are sequenced on an Illumina HiSeq 1500 sequencer. The *D.*
58
59 569 *mawsoni* genome was assembled using SOAPdenovo (17) to build the contigs and SSPACE
60
61 570 (19) to scaffold the contigs. The *E. maclovinus* genome was assembled using Platanus¹⁶ to
62
63 571 build the contigs, and SSPACE to scaffold the contigs.

572

573 **Annotation of the genomes**

574 We identified repeats, protein-coding genes and non-coding RNA in the genome assemblies
575 of the two species. First, a *de novo* repeat annotation of *D. mawsoni* and *E. maclovinus*
576 genomes was carried out by successively using RepeatModeler (version 1.0.8) and
577 RepeatMasker (version 4.0.5). *De novo* repeat libraries of the two species were constructed
578 with two complementary programs, RECON (58) and RepeatScout (59) implemented in the
579 RepeatModeler package. The generated consensus sequences were manually checked by
580 aligning to the Repbase transposable element library (<http://www.girinst.org/replib/>) and
581 genes from the NCBI database (nt and nr). The *D. mawsoni* and *E. maclovinus* repeat library
582 consisted of 975 and 676 consensus sequences with classification information respectively,
583 which were used to run RepeatMasker on the assembled scaffolds. Secondly, protein-coding
584 genes were predicted using a combination of homology-based and *de novo* approaches.
585 GLEAN was used to create consensus gene set by integrating evidence from each prediction.
586 Then RNA-Seq data were used to rectify gene models. Generated coding genes were aligned
587 to known protein databases, including InterPro (60), KEGG (61) and Uniprot (62), and
588 functional assignment was based on that of the best database match. Thirdly, the tRNA genes
589 were predicted with tRNAscan-SE (63). Aligning the rRNA template sequences from fishes
590 using BlastN with E-value 1e-5 identified the rRNA fragments. The miRNA and snRNA
591 genes were predicted with INFERNAL (64) software against the Rfam database (Release 12)
592 (65).

593

594 **Phylogenetic reconstruction of 10 vertebrate genomes**

595 Protein coding genes of Atlantic cod (*Gadus morhua*), tetraodon (*Tetraodon nigroviridis*),
596 Antarctic notothenioid *N. coriiceps*, stickleback (*Gasterosteus aculeatus*), tilapia
597 (*Oreochromis niloticus*), medaka (*Oryzias latipes*), zebrafish (*Danio rerio*) and mouse (*Mus*
598 *musculus*) genomes were collected from Ensembl release 84 or NCBI, and *D. mawsoni* and *E.*
599 *maclovinus* genes from this study, were used to build orthologous clusters with OrthoMCL
600 (Ver. 2.0.9) (21) with default parameters and options. A total of 2,936 one-to-one single-copy
601 genes were identified among the ten species. Protein-coding sequences of the orthologs were
602 aligned using PRANK (Ver.140603) (66) under a protein model with default parameters. The
603 coding sequences of the genes were concatenated to a supergene for each species. The
604 supergene sequence dataset was subjected to phylogenetic analysis using MrBayes (67)],
605 implementing best-fit substitution model (GTR+gamma+I) as determined by Modeltest (68).

1 606 The analysis was run 800,000 generations, sampling every 100 generations, with the first
2 607 2,000 sample set as burn-in. Branch-specific *dN* and *dS* were estimated with codeml of the
3
4 608 PAML package (69). The analysis of changes in gene family size were computed with CAFÉ
5
6 609 (70).
7
8 610

9 611 **GO annotation and identification of positive selection genes**

10 612 GO terms of the *D. mawsoni*, *E. maclovinus* and stickleback orthologs were built with
11 613 InterproScan (71). The orthologs of each GO terms were concatenated to estimate
12 614 branch-specific *dN* and *dS* using codeml of PAML. A binomial test was used to identify the
13 615 excess of nonsynonymous changes of GO categories in either *D. mawsoni* or *E. maclovinus*
14 616 lineages referenced to the stickleback. Only the GO terms carrying more than 30 orthologs
15 617 were put into this calculation. To detect genes evolving under positive selection in *D.*
16 618 *mawsoni*, we used the branch-site model in which likelihood ratio test (LRT) *p* values were
17 619 computed. Fisher's exact tests were used to test for over-represented functional categories
18 620 among the positive select genes. GO enrichment analyses of the genes under positive
19 621 selection were performed using a hypergeometric method.
20
21 622

22 623 **Calling of heterozygous SNPs**

23 624 All of the paired-end reads were mapped to the assembled scaffolds with the aligner SMALT
24 625 to detect the heterozygous sequence polymorphism in the genomes. The heterozygous SNPs
25 626 were called with SSAHA_Pileup (version 0.8; ftp://ftp.sanger.ac.uk/pub/zn1/ssaha_pileup/).
26 627 Five thresholds were used to post-filter unreliable SNPs: (1) SSAHA_Pileup SNP score ≥ 20 ;
27 628 (2) ratio of two alleles between 3:17 to 17:3; (3) the lowest sequencing depth for each
28 629 allele ≥ 5 ; (4) the minimum distance for adjacent SNPs ≥ 5 bp; (5) only one polymorphism
29 630 detected at each SNP position.
30
31 631

32 632 **Transcriptome analyses**

33 633 RNA-seq data derived from liver, gill, stomach, white muscle, red muscle, skin, small
34 634 intestine, brain, head kidney, caudal kidney, spleen, and ovary were analyzed for variations in
35 635 gene expression of *D. mawsoni* and *E. maclovinus*. RNA-seq reads were trimmed using
36 636 Trimmomatic (Ver. 0.33) (72) with the parameter set to AVGQUAL at 20, TRAILING at 20
37 637 and MINLEN at 50. The cleaned Illumina paired-end reads of each tissue were mapped to the
38 638 annotated scaffolds of *D. mawsoni* and *E. maclovinus* genome using HISAT2 aligner (Ver.
39 639 2.0.4)(73). Cufflinks (Ver. 2.2.1) (74) normalized gene expressions to the quantified
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 640 transcription levels (FPKM). Differentially expressions of the genes were assessed using
2 641 DEGseq (Ver. 1.28.0) (75)with cutoff at $q < 0.001$ (76). GO and KEGG enrichment analyses
3
4 642 for the identified differentially expressed genes were performed using cluserProfiler packages
5
6 643 (77) with the cutoff at $p < 0.05$.

7 644

9 645 **Construction of gene collinearity among *D. mawsoni*, *E. maclovinus* and stickleback**
10 **genomes.**

11 646
12 647 The genes of *D. mawsoni* and *E. maclovinus* were aligned to the gene model set of
13 648 stickleback by Blastp with E-value at $1e-20$. Two criteria were used to call syntenic gene
14 649 blocks in the *D. mawsoni* or *E. maclovinus* scaffolds: (1) Number of the gene on the syntenic
15 650 block ≥ 3 ; (2) number of non-syntenic genes between two adjacent syntenic genes ≤ 10 .
16 649
17
18 650
19 651 Each syntenic block was anchored on the stickleback genomes according to the orders of the
20 651
21 652 reference gene.

22 653

23 654
24 654 **Western-blot analysis of ZP proteins**

25 655
26 655 The proteins were separated on 10% SDS–PAGE at 100V for 90 min in 193 mM glycine and
27 656 25 mM Tris (pH 8.8). The resolved proteins were electrophoretically transferred to a
28 656
29 657 nitrocellulose membrane (Millipore) using a Mini-Protean Tetra Cell (BioRad) in a buffer
30 657
31 658 containing 193mM glycine, 25mM Tris (pH 8.3) and 20% methanol. The membrane was
32 658
33 659 treated with blocking agent (5% nonfat milk in 1x TBST) for 2 h at room temperature on a
34 659
35 660 shaker. FLAG antibody or β -actin antibody (Hua An Biotechnology Co. Ltd, Hangzhou,
36 660
37 661 China) was added, and the membranes were incubated at room temperature for 1 h. The
38 661
39 662 membrane was then washed with 1x TBST three times for 15 min each. The secondary
40 662
41 663 antibody (1:2,000 in 1x TBST, Boston Biomedical Inc.) was then added and incubated for 1 h
42 663
43 664 at room temperature. The membrane was washed with 1x TBST twice and 1x TBS once for
44 664
45 665 15 min each. Color was developed using SuperSignal West Pico Chemiluminescent Substrate
46 665
47 666 (Thermo Scientific) according to the manufacturer's instructions. Images were acquired using
48 666
49 667 a ChemiDoc MP Imaging System (BioRad).

50 668

51 669
52 669 **Assay of CHO cell survival rate at freezing temperature.**

53 670
54 670 DmZPC5 exons were serially deleted using PCR amplification of the DmZPC5 expression
55 670
56 671 vector with primers designed to eliminate desired coding sequences (Fig.3b). The full-length
57 671
58 672 sequences of three DmZPC5 isoforms (DmZPC5-1, DmZPC5-2 and DmZPC5-3) were
59 672
60 673 engineered to contain a FLAG octapeptide and cloned into the expression vector

1 674 pIRES2-EGFP (**Additional file 1: Fig. S5**). The three constructed vectors and blank control
2 675 (vector pIRES2-EGFP) were transferred into the CHO cells (American Type Culture
3
4 676 Collection CCL-61) obtained from American Type Culture Collection. The CHO cells were
5
6 677 cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum (Gibco).
7
8 678 CHO cells were incubated at 37°C for 2 days and then kept at -2 °C for 8 hours. The treated
9
10 679 CHO cells were collected and washed by DPBS (Dulbecco's Phosphate-Buffered Saline)
11
12 680 twice. The cells were stained with 10 µg/mL propidium iodide at room temperature for 5
13
14 681 minutes, and numbers of PI-stained cells (dead cells) were determined by flow cytometry.
15
16 682 The survival rate is calculated with the equation: survival rate = S / (S + D), where S is the
17
18 683 number of surviving cells and D is the number the dead cells.
19

20 684

21 685 **Identification of the LINES and estimation of their divergence time.**

22 686 The seed sequences of the LINES (28) were aligned against the *D. mawsoni* and *E.*
23
24 687 *maclovinus* draft genomes, respectively, using BlastN at E-value of 1e-10. According to loci
25
26 688 of the alignments, the sequences were extracted from the genomes, which were considered as
27
28 689 the candidates of the LINES. If distance of two adjacent candidates was 200 bp or less, these
29
30 690 two candidates were connected by the sequence between them. All of the candidates and their
31
32 691 corresponding seed sequences were mutually aligned by BlastN. Those candidates with over
33
34 692 60% identity and over 100 bp of alignments to the seed sequences were collected as the
35
36 693 LINES of *D. mawsoni* and *E. maclovinus*, respectively.

37 694 Alignment of the LINES was conducted by a multiple sequence aligner ClustalW (78) to
38
39 695 cluster any two LINES with highest sequence similarity into a LINE pair in *D. mawsoni* or *E.*
40
41 696 *maclovinus*. The evolutionary distance of two LINES for each LINE pair was calculated by
42
43 697 the Kimura two-parameter method (EMBOSS distmat, version 6.6.0.0), which reflected the
44
45 698 substitution rate per site between the two LINES. According to the calculated synonymous
46
47 699 substitution rates for 7,958 *D. mawsoni*-*E. maclovinus* orthologous pairs, the mean
48
49 700 synonymous substitution rate is around 0.227. The peak of substitution rate is at 0.04, which
50
51 701 estimated the LINE burst to be about 6.5 million years ago when the species divergence time
52
53 702 between *D. mawsoni* and *E. maclovinus* is around 37 million years ago (79).

54 703

55 704 **Tissue fixation and immunohistochemistry**

56 705 Pieces of pelvic bone (with muscle) and the attached fins (no larger than 20 × 20 × 5 mm)
57
58 706 were dissected from frozen specimens of young *D. mawsoni* and *E. maclovinus* and immersed
59
60 707 in a fixation solution, KINFix which contains (62.5% (v/v) ethanol, 6.71% (v/v) acetic acid,
61
62
63
64
65

1 708 and 6% (w/v) trehalose (80) for over 24hrs. Tissue are decalcified in EDTA solution (cat.
2 709 no.041-22031, WAKO) for about 2 weeks. Then the specimen was dehydrated in graded
3
4 710 ethanol (70%, 80%, 90%,95%, 1×1h each), 100% ethanol 2 ×1h at room temperature, xylene
5
6 711 for 2×1h, and embedded in low-melting paraffin for 2 × 1h, and kept overnight at 56°C, then
7
8 712 embedded in paraffin. For each tissue, 5 µm thick serial sections were cut with a microtome
9
10 713 (RM2245, Leica). Immunostaining was performed using the EnVision detection system (cat.
11 714 no.K5007, Dako). Slides were deparaffinized in xylene and rehydrated in a descending series
12 715 of ethanol (100%, 95%, 90% and 70%), and washed in phosphate buffered saline (PBS).
13 716 Endogenous peroxidases were blocked with 3% H₂O₂ for 10 min, after which the sections
14 717 were incubated with 5% BSA for 35min. Then, the slides are incubated overnight with the
15 718 primary CTGF antibody (1:400 dilution) (cat. no.ab6992, Abcam) at 4°C. Next, the sections
16 719 were washed four times with PBS for 15 min followed by incubation with a goat anti-rabbit
17 720 secondary antibody for 35min at 37°C. After four washes with PBS, 3, 3'-diaminobenzidine
18 721 (DAB) was added to visualize the immunoreactivity. All slides were counterstained with
19 722 haematoxylin. The sections were dehydrated in a mounting series of alcohol (70%, 90%, 100%
20 723 and 100%) and in xylene. Finally, slides are mounted using neutral balsam mounting medium,
21 724 and analyzed under a bright field microscope (AXIO imager. M2, ZEISS).
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

726 **Availability of data and material**

727 All of the Illumina short read sequencing data of this project have been deposited at NCBI
728 under the accession no. BioProject PRJNA401363 (<http://www.ncbi.nlm.nih.gov/sra/>). The
729 assembled draft genomes and their annotations have been released at the official website of
730 the Shanghai Ocean University (<http://202.121.66.128/>). The current version of the data set is
731 the first version (v1).

733 **Additional files**

734 Additional file 1 : Figs. S1 to S11 and Tables S1 to S17.

735 Additional file 2 : supporting data for Fig. 1b, Fig.2c, Fig. 2d, and Fig. 4.

736 Additional file 3 : list of duplicated protein gene families of *D. mawsoni* and *E. maclovinus*.

737 Additional file 4: list of DEGs between *D.mawsoni* and two negatively buoyant
738 notothenioids.

739

740 **Acknowledgements**

741 We thank BGI-Shen Zhen for their technical support in sequencing of *D. mawsoni* genome,
742 Mr. Liping Shu of Wuhan Ice-harbor Biological Technological Co. Ltd. for his help in the
743 annotation of the *E. maclovinus* genome.

745 **Funds**

746 The work was supported by grants from the Natural Science Foundation of China (No.
747 41761134050, 31572611, 31572598) and the Major Science Innovation Grant
748 (2017-01-07-00-10-E00060) from the Shanghai Education Committee, the Key Achievement
749 Supporting Grant from Laboratory for Marine Biology and Biotechnology, Qingdao National
750 Laboratory for Marine Science and Technology to LC, and the USA NSF Polar Programs
751 grant ANT1142158 to CHCC.

753 **Authors' contribution**

754 LC and CHCC conceived and managed the project and its components. CHCC and MH
755 performed fish and tissue collections and sample preparations. KRM and XZ contributed to
756 DNA and RNA preparations and genome size determination. YL, MY, and WL performed
757 genome annotation and RNA-seq data analysis. YR and SP conducted *de novo* genome
758 assembly. KTB confirmed the AFGP loci. QX, YF and LC designed and performed the
759 biological experiments. Sample preparation and genome sequencing were carried out by SJ,
760 WZ and JW. LC, QX, YL and CHCC analyzed the data as a whole and wrote the manuscript,
761 and CHCC and WW contributed interpretation of data and edits to the manuscript.

763 **Competing interests**

764 The authors declare to no competing financial interests.

766 **URLs** KEGG, <http://www.genome.jp/kegg/>; KAAS,
767 <http://www.genome.jp/tools/kaas/>;SSPACE,
768 <https://www.baseclear.com/genomics/bioinformatics/basetools/SSPACE>; Platanus,
769 <http://platanus.bio.titech.ac.jp/>; SMALT, <http://www.sanger.ac.uk/resources/software/smalt/>;
770 SOAPdenovo, <http://soap.genomics.org.cn>; RepeatModeler,
771 <http://www.repeatmasker.org/RepeatModeler.html>; RepeatMasker,
772 <http://www.repeatmasker.org>; Rebase, <http://www.girinst.org/rebase/>; Timetree,

1 773 <http://www.timetree.org/>;Ensembl, <ftp://ftp.ensembl.org/pub/>;
2 774 <http://evolution.genetics.washington.edu/phylip.html>; [PAML](#),
3
4 775 <http://abacus.gene.ucl.ac.uk/software/paml.html>; [GLEAN](#), <https://github.com/glean/glean>;
5
6 776 Interpro, <http://www.ebi.ac.uk/interpro/>; Infernal, <http://eddylab.org/infernal/>; Rfam,
7 777 <http://rfam.xfam.org/>; OrthoMCL, <http://orthomcl.org/orthomcl/>; Mrbayes,
8 778 <http://mrbayes.sourceforge.net/>; SSAHA_Pileup, ftp://ftp.sanger.ac.uk/pub/zn1/ssaha_pileup/;
9 779 HISAT2, <http://ccb.jhu.edu/software/hisat2/index.shtml>.

10
11
12
13
14
15
16
17
18
19

20 784 **Legends**

21
22 785

23 786 **Fig. 1** Sequenced species and genome synteny. (a) Sampling location and geographic distribution.
24 787 The red and blue filled circles are the geographic distributions of *D. mawsoni* and *E. maclovinus*,
25 788 respectively (www.fishbase.org, version 02/2018 (81) and Hanchet et al.(82)). The red and blue
26 789 stars show the respective locations where the sequenced individuals were collected. The Antarctic
27 790 Polar Front, an approximation of the mean position of the Antarctic Circumpolar Current, is
28 791 adopted from Barker and Thomas(83). The image of *D. mawsoni* is a courtesy from Elliot
29 792 DeVries and that of *E. maclovinus* is from Dirk Schories. (b) Gene collinearity among *D.*
30 793 *mawsoni*, *E. maclovinus* and *G. aculeatus* (Stickleback). The scaffolds of *D. mawsoni* (the
31 794 circularized red blocks labelled with “D”) and *E. maclovinus* (the circularized green blocks
32 795 labelled with “E”) are anchored on the twenty-one Stickleback chromosomes (the circularized
33 796 light blue blocks labelled with “G”, 1 to 21), according to the gene collinearity (the
34 797 connecting yellow lines). The black vertical lines within the *D. mawsoni* and *E. maclovinus*
35 798 scaffolds indicate occurrence of LINEs greater than 500 bps in these positions. The sequence
36 799 length is indicated by the 5-Mb tick marks on the reference Stickleback chromosomes. The
37 800 outermost circle of red vertical lines and the innermost circle of green vertical lines indicated
38 801 the quantified expression levels (FPKM) of the genes located on the corresponding *D.*
39 802 *mawsoni* and *E. maclovinus* scaffolds, respectively. The expression profiles are derived from
40 803 the transcriptome data of white muscles (see the transcriptome section). The small white
41 804 squares and rectangles scattered in the scaffolds show the locations of the Selcys-tRNA genes
42 805 of *D. mawsoni* and *E. maclovinus*. The single yellow square shows the location of AFGP
43 806 genes in the *D. mawsoni* genome.

807

808 **Fig. 2** Evolution of the genomes and genes. **(a)** Timing and frequency of LINE insertion in *D.*
809 *mawsoni* and *E. maclovinus* showing correlation between onset of late Miocene deep cooling
810 and burst LINE insertions in the Antarctic toothfish. The black trace indicates global
811 temperature trends during Oligocene, Miocene, Pliocene (Pli) and Pleistocene (Ple) from 30
812 to 0 million years ago (mya), modified from Zachos *et al.* (2008)(84), Near *et al.* (2012)
813 (9) and Favre *et al.* (2015) (85). The red and blue line indicate the insertion frequency of
814 LINEs (the percentage of the calculated LINE pairs) in the *D. mawsoni* and *E. maclovinus*
815 genomes respectively during these periods. **(b)** Reconstructed phylogeny of nine teleost fish
816 lineages using 2,936 orthologous genes (mouse serving as outgroup) and the calculated dN/dS
817 ratio for each branch, showing a 2-fold faster evolutionary rate in the Antarctic notothenioids.
818 **(c)** Comparison of adaptive evolution between *D. mawsoni* and *E. maclovinus* genomes. Data
819 points represent average dN/dS value of each GO term, each of which consists of at least 30
820 genes. The red and blue circles show the GO terms with significantly higher dN/dS ratios ($p <$
821 0.05, binomial test) in *D. mawsoni* and *E. maclovinus*, respectively. The grey circles are those
822 showing no significant difference. GO terms falling on the dashed line of linearity have the
823 same dN/dS ratios in the two species. **(d)** Gene duplication in *D. mawsoni*. A subset (26) of
824 the 202 gene families detected to contain higher gene copy numbers in the *D. mawsoni*
825 genome relative to other species are listed on the left, with their respective KEGG pathway
826 listed on the right. The gene copy numbers are measured by color difference. The pathways
827 highlighted in red are especially abundant in *D. mawsoni* and might be relevant to
828 physiological adaptation of *D. mawsoni* in the freezing environment. **(e)** A subset of
829 duplicated gene families in *E. maclovinus*, showing different KEGG pathways between *D.*
830 *mawsoni* and *E. maclovinus* in terms of gene duplication. The red highlighted pathway (Ether
831 lipid metabolism) indicates a common duplication occurred in the three Notothenioids.

832

833

834 **Fig. 3** Evolutionary and functional analyses of the DmZPC5 genes involved in cellular
835 freezing resistance. **(a)** Duplication of ZPC5 gene (DmZPC5) in *D. mawsoni*. Phylogenetic
836 neighbor-joining tree of ZPC5 genes among *D. mawsoni*, *E. maclovinus*, *Larimichthys crocea*
837 (Lc), *T. rubripes* (Tr) and *O. latipes* (Ol). The gene structures are illustrated on the right. The
838 different colored blocks indicate the exons encoding signal peptides (red), zp domains (blue)
839 and the remaining exons (incarnadine). The jagged blocks contain the nonsense mutations in
840 DmZPC5-2a/b and DmZPC5-3 genes that cause premature termination of coding sequences.
841 **(b)** Western-blot analysis of the DmZPC5 isoforms indicated their sizes and temperature

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

842 sensitive accumulation. Purified proteins encoded by the three DmZPC5 isoforms were
843 detected by an anti-FLAG antibody on the SDS-PAGE gels. All of these three DmZPC5
844 protein had higher expression levels at 0°C than at 37 °C. (c) Assays of cell survival rate
845 under recombinant expression of different DmZPC5 isoforms in CHO cells at a freezing
846 temperature (-2°C for 8 hrs). The bars represent the mean ± s.d ($n = 3$, biological replicates).
847 The sample pIRES2-EGFP is the expression vector as control. Significances of different
848 survival rate are indicated by * (unpaired Student's t-test, $p < 0.05$) and ** ($p < 0.01$).

849
850
851
852 **Fig. 4.** Comparison of gene expression between *D. mawsoni* and *E. maclovinus* tissues. The
853 squares/triangles, circles and diamonds filled in different colors represented the genes
854 involved in three metabolic processes (listed on the right). The genes with significantly higher
855 expression in *D. mawsoni* or *E. maclovinus* are labeled on the corresponding organs.

856
857
858 **Fig. 5** Schematic diagram showing changed regulation of buoyancy related developmental
859 pathways. (a). Enhanced adipogenetic pathways in *D. mawsoni* muscle. The genes shadowed
860 with the dark red color were upregulated in *D. mawsoni* while those shadowed in light grey
861 were unchanged. (b). Changed osteogenetic regulation in *D. mawsoni* bone. Genes shadowed
862 with the dark grey color were upregulated in *D. mawsoni* while those in light grey were not
863 changed. The arrows (in dark red or dark grey) indicate a positive effect on the process while
864 blocked (in blue) lines indicate inhibitory effect. MSC: Mesenchymal Stem Cell. (c).
865 Immunohistochemical (IHC) staining to detect the abundance of Connective Tissue Growth
866 Factor (CTGF) in cross sections of pelvic fin of *D. mawsoni* and *E. maclovinus*. The left
867 panels of each fish are IHC staining without the first antibody as negative control. The
868 presence of CTGF is indicated by the brown signals in the right tissues. Scale bar, 50µm.

870 **References**

1
2 871
3
4 872 1. Livermore R, Nankivell A, Eagles G, & Morris P (2005) Paleogene opening of Drake Passage.
5 873 *Earth Planet Sc Lett* 236(1-2):459-470.
6 874 2. Eastman JT (2005) The nature of the diversity of Antarctic fishes. *Polar Biol* 28(2):93-107.
7 875 3. Near TJ, *et al.* (2015) Identification of the notothenioid sister lineage illuminates the
8 876 biogeographic history of an Antarctic adaptive radiation. *Bmc Evol Biol* 15:109.
9 877 4. Eastman JT (2000) Antarctic notothenioid fishes as subjects for research in evolutionary
10 878 biology. *Antarct Sci* 12(3):276-287.
11 879 5. La Mesa M, Eastman JT, & Vacchi M (2004) The role of notothenioid fish in the food web of
12 880 the Ross Sea shelf waters: a review. *Polar Biol* 27(6):321-338.
13 881 6. Devries AL & Eastman JT (1978) Lipid sacs as a buoyancy adaptation in an Antarctic fish.
14 882 *Nature* 271(5643):352-353.
15 883 7. Chen LB, DeVries AL, & Cheng CHC (1997) Evolution of antifreeze glycoprotein gene from a
16 884 trypsinogen gene in Antarctic notothenioid fish. *P Natl Acad Sci USA* 94(8):3811-3816.
17 885 8. Matschiner M, Hanel R, & Salzburger W (2011) On the Origin and Trigger of the Notothenioid
18 886 Adaptive Radiation. *Plos One* 6(4).
19 887 9. Near TJ, *et al.* (2012) Ancient climate change, antifreeze, and the evolutionary diversification
20 888 of Antarctic fishes. *P Natl Acad Sci USA* 109(9):3434-3439.
21 889 10. Shin SC, *et al.* (2014) The genome sequence of the Antarctic bullhead notothen reveals
22 890 evolutionary adaptations to a cold environment. *Genome Biol* 15(9).
23 891 11. Gon O & Heemstra P (1990) *Fishes of the Southern Ocean* (J.L.B. Smith Institute of
24 892 Ichthyology) p 305.
25 893 12. Eastman JT, Witmer LM, Ridgely RC, & Kuhn KL (2014) Divergence in Skeletal Mass and Bone
26 894 Morphology in Antarctic Notothenioid Fishes. *J Morphol* 275(8):841-861.
27 895 13. Eastman JT & Devries AL (2010) Buoyancy adaptations in a swim - bladderless Antarctic fish.
28 896 *J Morphol* 167(1):91-102.
29 897 14. Eastman JT & Devries AL (1982) Buoyancy Studies of Notothenioid Fishes in McMurdo Sound,
30 898 Antarctica. *Copeia* 1982(2):385-393.
31 899 15. Ghigliotti L, *et al.* (2007) The two giant sister species of the Southern Ocean, *Dissostichus*
32 900 *eleginoides* and *Dissostichus mawsoni*, differ in karyotype and chromosomal pattern of
33 901 ribosomal RNA genes. *Polar Biol* 30(5):625-634.
34 902 16. Mazzei F, *et al.* (2008) Chromosomal characteristics of the temperate notothenioid fish
35 903 *Eleginops maclovinus* (Cuvier). *Polar Biol* 31(5):629-634.
36 904 17. Li RQ, Li YR, Kristiansen K, & Wang J (2008) SOAP: short oligonucleotide alignment program.
37 905 *Bioinformatics* 24(5):713-714.
38 906 18. Kajitani R, *et al.* (2014) Efficient de novo assembly of highly heterozygous genomes from
39 907 whole-genome shotgun short reads. *Genome Res* 24(8):1384-1395.
40 908 19. Boetzer M, Henkel CV, Jansen HJ, Butler D, & Pirovano W (2011) Scaffolding pre-assembled
41 909 contigs using SSPACE. *Bioinformatics* 27(4):578-579.
42 910 20. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015) BUSCO:
43 911 assessing genome assembly and annotation completeness with single-copy orthologs.
44 912 *Bioinformatics* 31(19):3210-3212.
45 913 21. Li L, Stoeckert CJ, & Roos DS (2003) OrthoMCL: Identification of ortholog groups for

914 eukaryotic genomes. *Genome Res* 13(9):2178-2189.

1 915 22. Pisano E & Ozouf-Costaz C (2003) Cytogenetics and evolution in extreme environment: the
2 916 case of Antarctic fishes. *Science Publishers Inc, Enfield (NH) USA*:309-330.

3 917 23. Mazzei F, *et al.* (2004) Chromosomal patterns of major and 5S ribosomal DNA in six icefish
4 918 species (Perciformes, Notothenioidei, Channichthyidae). *Polar Biol* 28(1):47-55.

5 919 24. Detrich HW, *et al.* (2010) Genome Enablement of the Notothenioidei: Genome Size Estimates
6 920 from 11 Species and BAC Libraries from 2 Representative Taxa. *J Exp Zool Part B*
7 921 314b(5):369-381.

8 922 25. Rebollo R, Horard B, Hubert B, & Vieira C (2010) Jumping genes and epigenetics: Towards new
9 923 species. *Gene* 454(1-2):1-7.

10 924 26. Auvinet J, *et al.* (2018) Mobilization of retrotransposons as a cause of chromosomal
11 925 diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus*.
12 926 *Bmc Genomics* 19.

13 927 27. Chenais B, Caruso A, Hiard S, & Casse N (2012) The impact of transposable elements on
14 928 eukaryotic genomes: From genome size increase to genetic adaptation to stressful
15 929 environments. *Gene* 509(1):7-15.

16 930 28. Chen S, *et al.* (2017) Cold-induced retrotransposition of fish LINEs. *J Genet Genomics*
17 931 44(8):385-394.

18 932 29. Chen ZZ, *et al.* (2008) Transcrintomic and genomic evolution under constant cold in Antarctic
19 933 notothenioid fish. *P Natl Acad Sci USA* 105(35):12944-12949.

20 934 30. Xu QH, *et al.* (2008) Adaptive evolution of hepcidin genes in antarctic notothenioid fishes.
21 935 *Mol Biol Evol* 25(6):1099-1112.

22 936 31. Cao LX, *et al.* (2016) Neofunctionalization of zona pellucida proteins enhances
23 937 freeze-prevention in the eggs of Antarctic notothenioids. *Nat Commun* 7.

24 938 32. Nicodemus-Johnson J, Silic S, Ghigliotti L, Pisano E, & Cheng CHC (2011) Assembly of the
25 939 antifreeze glycoprotein/trypsinogen-like protease genomic locus in the Antarctic toothfish
26 940 *Dissostichus mawsoni* (Norman). *Genomics* 98(3):194-201.

27 941 33. Brigelius-Flohe R & Maiorino M (2013) Glutathione peroxidases. *Bba-Gen Subjects*
28 942 1830(5):3289-3303.

29 943 34. Angeli JPF, *et al.* (2014) Inactivation of the ferroptosis regulator Gpx4 triggers acute renal
30 944 failure in mice. *Nat Cell Biol* 16(12):1180-U1120.

31 945 35. Chen Y, Rui BB, Tang LY, & Hu CM (2015) Lipin Family Proteins - Key Regulators in Lipid
32 946 Metabolism. *Ann Nutr Metab* 66(1):10-18.

33 947 36. Farmer SR (2006) Transcriptional control of adipocyte formation. *Cell Metab* 4(4):263-273.

34 948 37. Iqnotz RA & Massagué J (1985) Type beta transforming growth factor controls the adipogenic
35 949 differentiation of 3T3 fibroblasts. *P Natl Acad Sci USA* 82(24):8530-8534.

36 950 38. Petruschke T, Röhrig K, & Hauner H (1994) Transforming growth factor beta (TGF-beta)
37 951 inhibits the differentiation of human adipocyte precursor cells in primary culture. *Int J Obes*
38 952 *Relat Metab Disord* 18(8):532-536.

39 953 39. Long FX & Ornitz DM (2013) Development of the Endochondral Skeleton. *Csh Perspect Biol*
40 954 5(1).

41 955 40. Luo Q, *et al.* (2004) Connective tissue growth factor (CTGF) is regulated by Wnt and bone
42 956 morphogenetic proteins signaling in osteoblast differentiation of mesenchymal stem cells. *J*
43 957 *Biol Chem* 279(53):55958-55968.

1 958 41. Sadatsuki R, *et al.* (2017) Perlecan is required for the chondrogenic differentiation of synovial
2 959 mesenchymal cells through regulation of Sox9 gene expression. *J Orthop Res* 35(4):837-846.
3 960 42. Mongiat M, *et al.* (2003) Perlecan protein core interacts with extracellular matrix protein 1
4 961 (ECM1), a glycoprotein involved in bone formation and angiogenesis. *J Biol Chem*
5 962 278(19):17491-17499.
6
7 963 43. Arnold MA, *et al.* (2007) MEF2C transcription factor controls chondrocyte hypertrophy and
8 964 bone development. *Dev Cell* 12(3):377-389.
9
10 965 44. Gamez B, Rodriguez-Carballo E, Bartrons R, Rosa JL, & Ventura F (2013) MicroRNA-322
11 966 (miR-322) and Its Target Protein Tob2 Modulate Osterix (Osx) mRNA Stability. *J Biol Chem*
12 967 288(20):14264-14275.
13
14 968 45. Tago K, *et al.* (2000) Inhibition of Wnt signaling by ICAT, a novel beta-catenin-interacting
15 969 protein. *Gene Dev* 14(14):1741-1749.
16
17 970 46. Yao W, *et al.* (2010) Overexpression of Secreted Frizzled-Related Protein 1 Inhibits Bone
18 971 Formation and Attenuates Parathyroid Hormone Bone Anabolic Effects. *J Bone Miner Res*
19 972 25(2):190-199.
20
21 973 47. Onizuka S, *et al.* (2016) ZBTB16 as a Downstream Target Gene of Osterix Regulates
22 974 Osteoblastogenesis of Human Multipotent Mesenchymal Stromal Cells. *J Cell Biochem*
23 975 117(10):2423-2434.
24
25 976 48. Dexheimer V, *et al.* (2016) Differential expression of TGF-beta superfamily members and role
26 977 of Smad1/5/9-signalling in chondral versus endochondral chondrocyte differentiation. *Sci*
27 978 *Rep-Uk* 6.
28
29 979 49. Tare RS, Oreffo ROC, Clarke NMP, & Roach HI (2002) Pleiotrophin/osteoblast-stimulating
30 980 factor 1: Dissecting its diverse functions in bone formation. *J Bone Miner Res*
31 981 17(11):2009-2020.
32
33 982 50. Su JL, *et al.* (2010) CYR61 Regulates BMP-2-dependent Osteoblast Differentiation through the
34 983 alpha(v)beta(3) Integrin/Integrin-linked Kinase/ERK Pathway. *J Biol Chem*
35 984 285(41):31325-31336.
36
37 985 51. Maruyama K, *et al.* (2013) Strawberry notch homologue 2 regulates osteoclast fusion by
38 986 enhancing the expression of DC-STAMP. *J Exp Med* 210(10):1947-1960.
39
40 987 52. Diercke K, Sen S, Kohl A, Lux CJ, & Erber R (2011) Compression-dependent Up-regulation of
41 988 Ephrin-A2 in PDL Fibroblasts Attenuates Osteogenesis. *J Dent Res* 90(9):1108-1115.
42
43 989 53. Abdallah BM & Kassem M (2012) New factors controlling the balance between
44 990 osteoblastogenesis and adipogenesis. *Bone* 50(2):540-545.
45
46 991 54. Rabosky DL, *et al.* (2018) An inverse latitudinal gradient in speciation rate for marine fishes.
47 992 *Nature* 559(7714):392-395.
48
49 993 55. Litscher ES, Williams Z, & Wassarman PM (2009) Zona Pellucida Glycoprotein ZP3 and
50 994 Fertilization in Mammals. *Mol Reprod Dev* 76(10):933-941.
51
52 995 56. Ricci M, Peona V, Guichard E, Taccioli C, & Boattini A (2018) Transposable Elements Activity is
53 996 Positively Related to Rate of Speciation in Mammals. *Journal of Molecular Evolution*
54 997 86(5):303-310.
55
56 998 57. Amemiya CT, Ota T, & Litman GW (1996) Construction of P1 Artificial Chromosome (PAC)
57 999 Libraries from Lower Vertebrates. *Nonmammalian Genomic Analysis*, (San Diego: Academic
58 1000 Press), pp 223-256.
59 1001 58. Bao ZR & Eddy SR (2002) Automated de novo identification of repeat sequence families in
60
61
62
63
64
65

1002 sequenced genomes. *Genome Res* 12(8):1269-1276.

1 1003 59. Price AL, Jones NC, & Pevzner PA (2005) De novo identification of repeat families in large
2 1004 genomes. *Bioinformatics* 21:I351-I358.

4 1005 60. Apweiler R, *et al.* (2001) The InterPro database, an integrated documentation resource for
5 1006 protein families, domains and functional sites. *Nucleic Acids Res* 29(1):37-40.

7 1007 61. Kanehisa M & Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids
8 1008 Res* 28(1):27-30.

9 1009 62. Bairoch A, *et al.* (2005) The universal protein resource (UniProt). *Nucleic Acids Res*
10 1010 33:D154-D159.

12 1011 63. Lowe TM & Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA
13 1012 genes in genomic sequence. *Nucleic Acids Res* 25(5):955-964.

15 1013 64. Nawrocki EP, Kolbe DL, & Eddy SR (2009) Infernal 1.0: inference of RNA alignments.
16 1014 *Bioinformatics* 25(10):1335-1337.

18 1015 65. Nawrocki EP, *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res*
19 1016 43(D1):D130-D137.

21 1017 66. Loytynoja A & Goldman N (2010) webPRANK: a phylogeny-aware multiple sequence aligner
22 1018 with interactive alignment browser. *Bmc Bioinformatics* 11.

23 1019 67. Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees.
24 1020 *Bioinformatics* 17(8):754-755.

26 1021 68. Posada D & Crandall KA (1998) MODELTEST: testing the model of DNA substitution.
27 1022 *Bioinformatics* 14(9):817-818.

29 1023 69. Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*
30 1024 24(8):1586-1591.

31 1025 70. De Bie T, Cristianini N, Demuth JP, & Hahn MW (2006) CAFE: a computational tool for the
32 1026 study of gene family evolution. *Bioinformatics* 22(10):1269-1271.

34 1027 71. Zdobnov EM & Apweiler R (2001) InterProScan - an integration platform for the
35 1028 signature-recognition methods in InterPro. *Bioinformatics* 17(9):847-848.

37 1029 72. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina
38 1030 sequence data. *Bioinformatics* 30(15):2114-2120.

40 1031 73. Kim D, Landmead B, & Salzberg SL (2015) HISAT: a fast spliced aligner with low memory
41 1032 requirements. *Nat Methods* 12(4):357-U121.

42 1033 74. Trapnell C, *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq
43 1034 experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562-578.

45 1035 75. Wang LK, Feng ZX, Wang X, Wang XW, & Zhang XG (2010) DEGseq: an R package for
46 1036 identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26(1):136-138.

48 1037 76. Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and
49 1038 Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57(1):289-300.

51 1039 77. Yu GC, Wang LG, Han YY, & He QY (2012) clusterProfiler: an R Package for Comparing
52 1040 Biological Themes Among Gene Clusters. *Omics* 16(5):284-287.

54 1041 78. Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of
55 1042 progressive multiple sequence alignment through sequence weighting, position-specific gap
56 1043 penalties and weight matrix choice. *Nucleic Acids Res*:4673-4680.

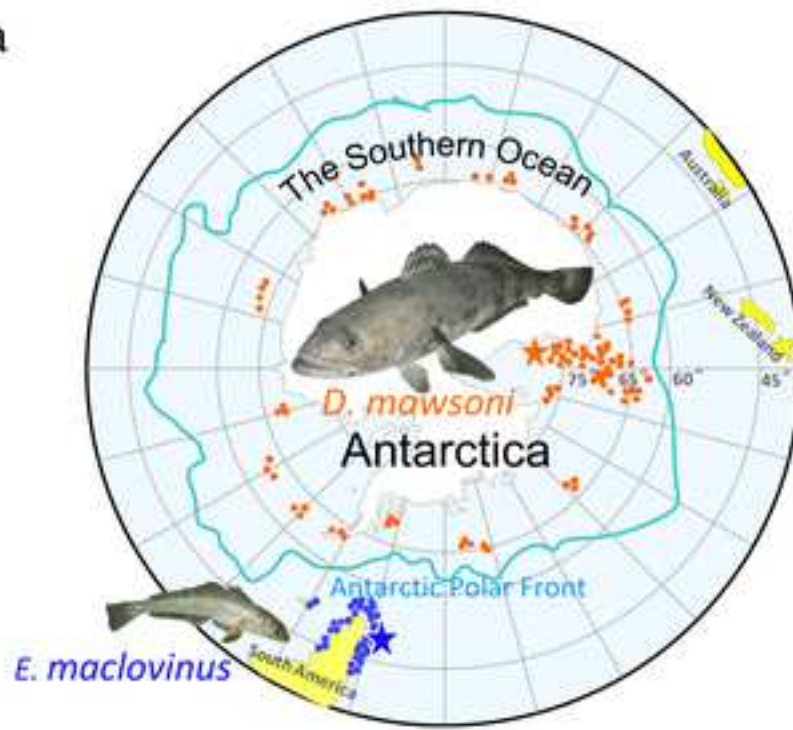
58 1044 79. Hedges SB, Dudley J, & Kumar S (2006) TimeTree: a public knowledge-base of divergence
59 1045 times among organisms. *Bioinformatics* 22(23):2971-2972.

1 1046 80. Stefanits H, *et al.* (2016) KINFix - A formalin-free non-commercial fixative optimized for
2 1047 histological, immunohistochemical and molecular analyses of neurosurgical tissue specimens.
3 1048 *Clin Neuropathol* 35(1):3-12.
4 1049 81. Pauly FRaD (2018) FishBase. in *World Wide Web electronic publication*.
5 1050 82. Hanchet S, *et al.* (2015) The Antarctic toothfish (*Dissostichus mawsoni*): biology, ecology, and
6 1051 life history in the Ross Sea region. *Hydrobiologia* 761(1):397-414.
7 1052 83. Barker PF & Thomas E (2004) Origin, signature and palaeoclimatic influence of the Antarctic
8 1053 Circumpolar Current. *Earth-Sci Rev* 66(1-2):143-162.
9 1054 84. Zachos JC, Dickens GR, & Zeebe RE (2008) An early Cenozoic perspective on greenhouse
10 1055 warming and carbon-cycle dynamics. *Nature* 451(7176):279-283.
11 1056 85. Favre A, *et al.* (2015) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of
12 1057 Tibetan biotas. *Biol Rev* 90(1):236-253.
13 1058
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

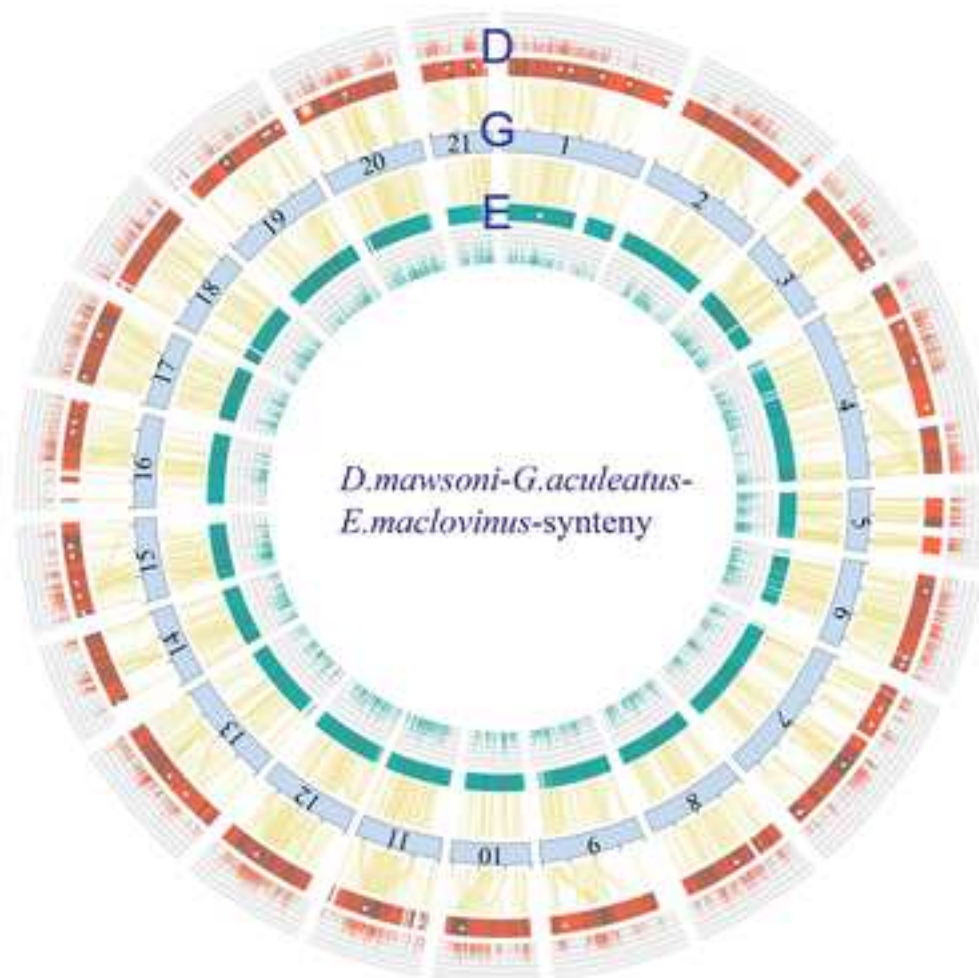
Table 1 Overview of assembly and annotation

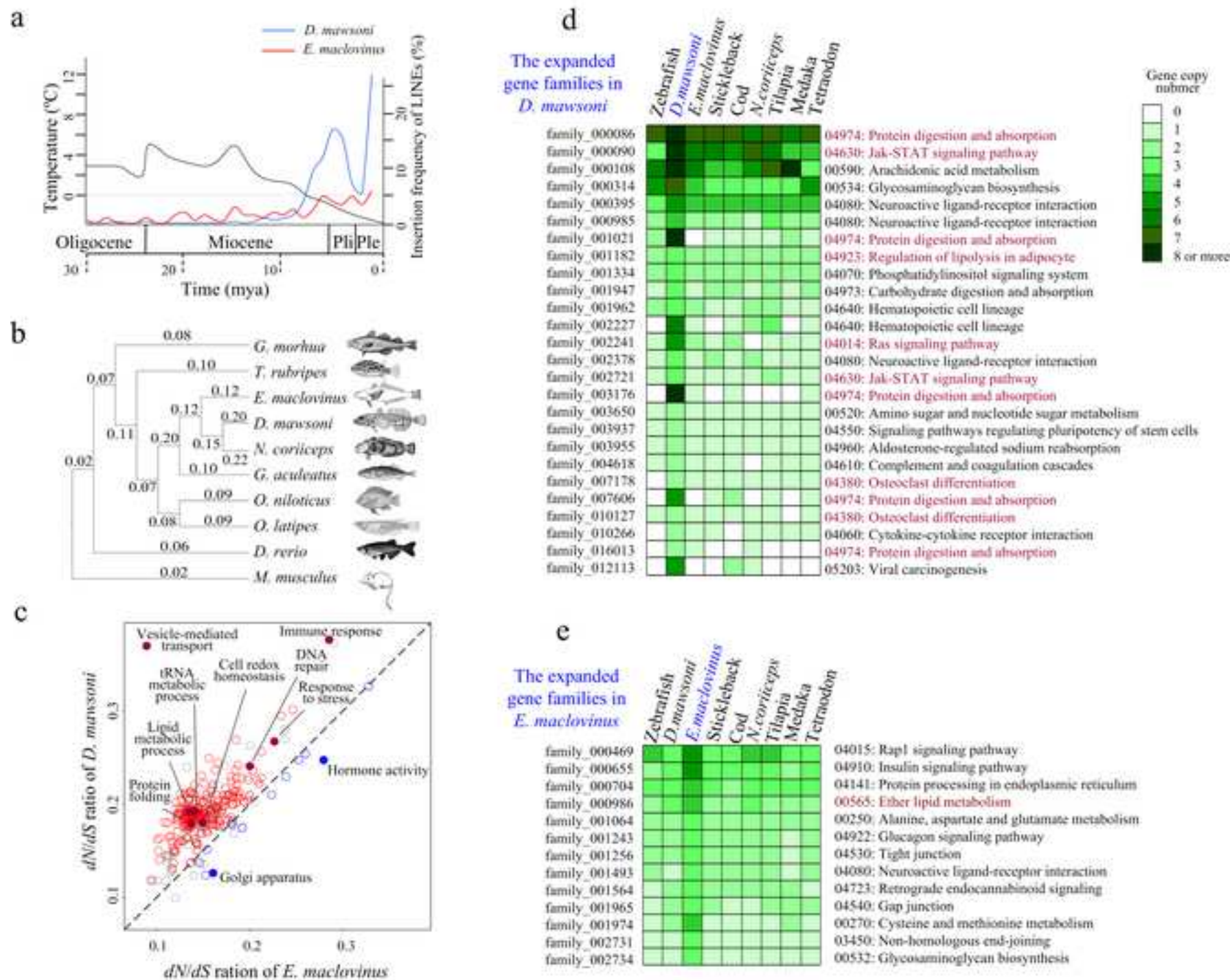
	<i>D. mawsoni</i>	<i>E. maclovinus</i>
Assembly		
Total length (Mb)	756.8	744.4
Contig N50 length (Kb)	23.1	10.9
Scaffold N50 length (Kb)	2,216.2	694.7
Scaffold N90 length (Kb)	202.7	167.2
Largest scaffold (Mb)	13.8	4.9
Quantity of scaffolds (>N90 length)	536	1,185
Annotation		
Quantity of predicted protein-coding genes	22,516	22,959
Quantity of predicted non-coding RNA genes	2,434	2,185
Content of transposable elements	21.38%	10.02%
Heterozygous SNP rate (SNPs per kb)	2.58	2.40

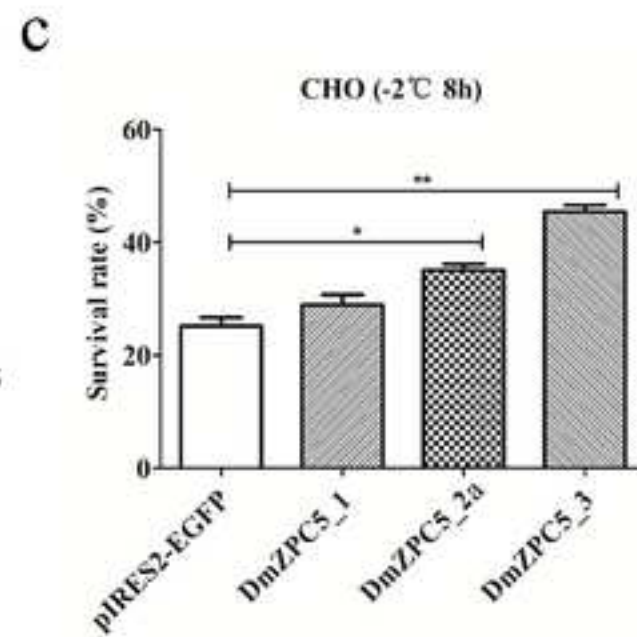
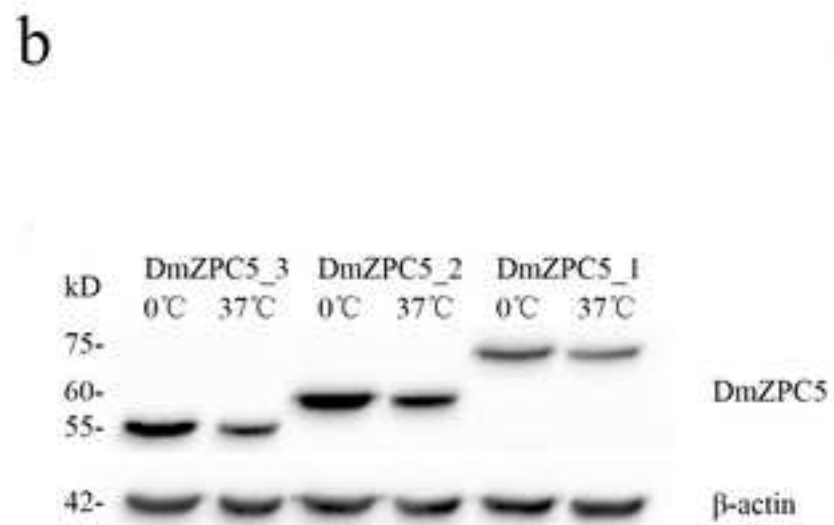
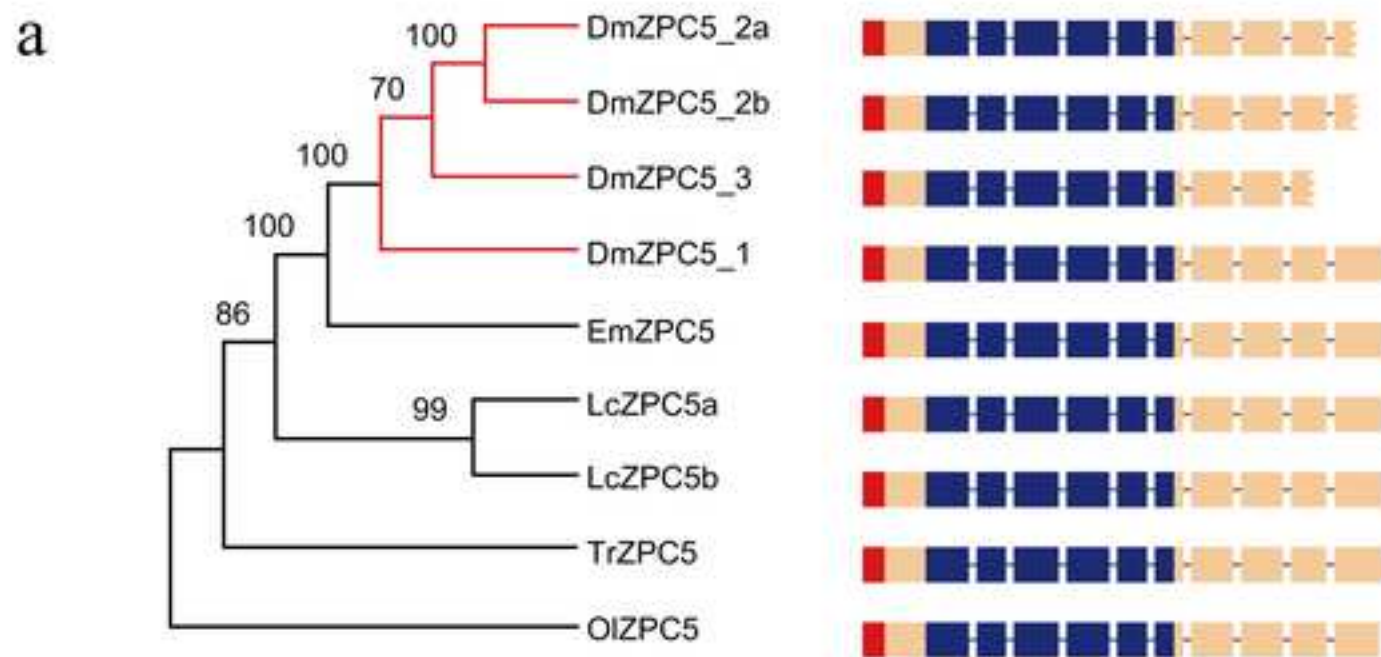
a

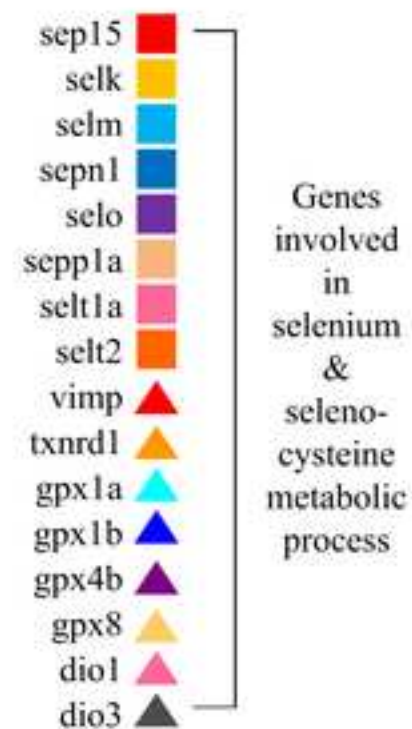
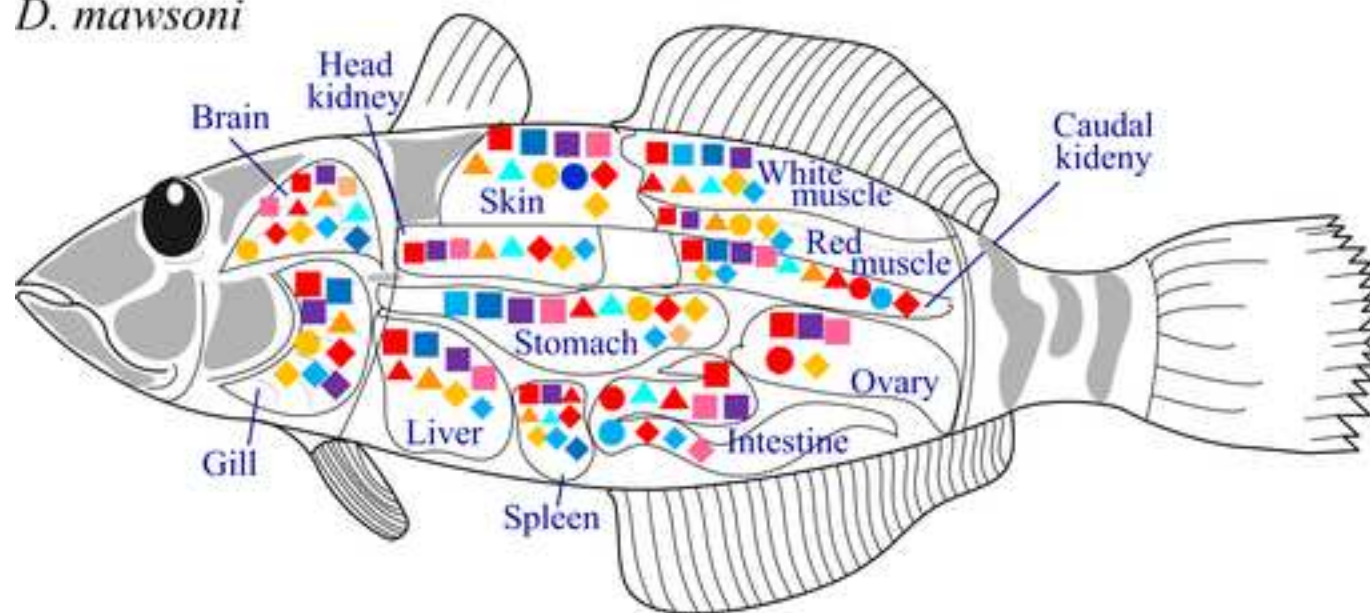
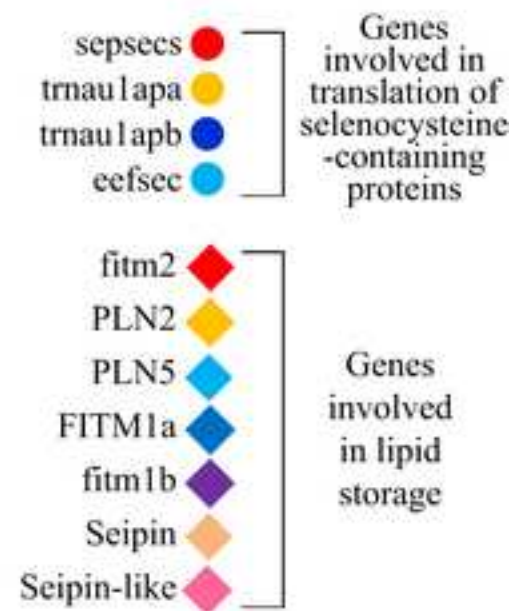
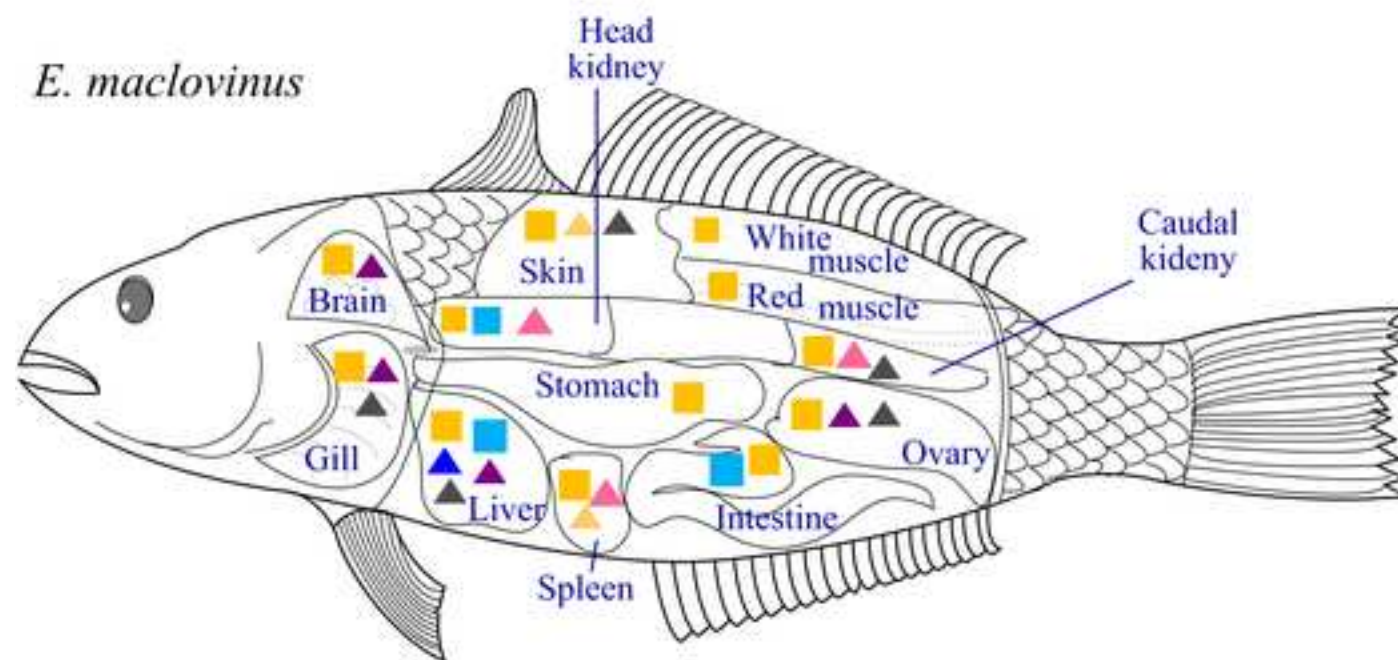


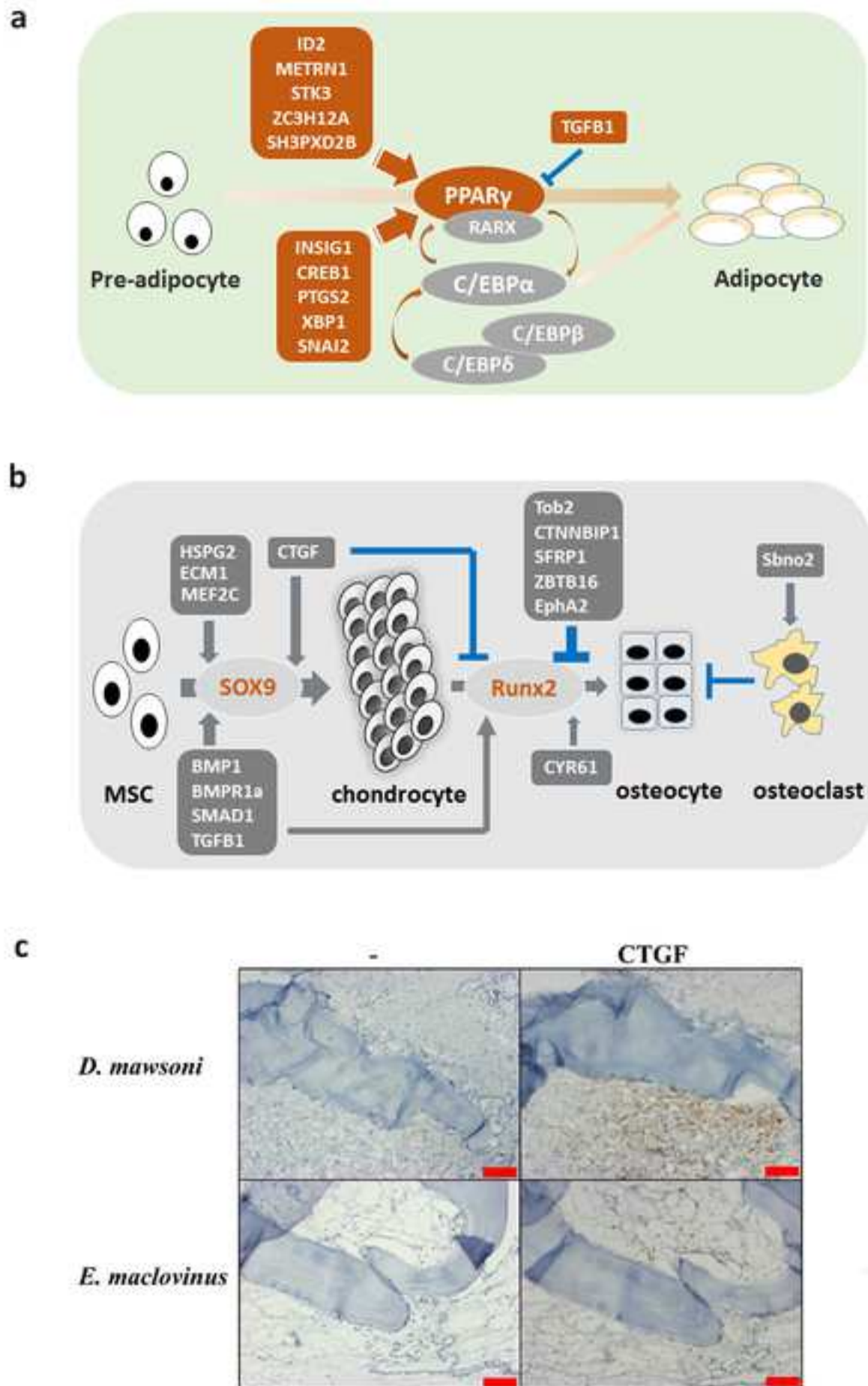
b

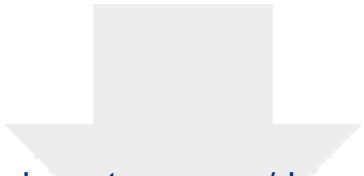







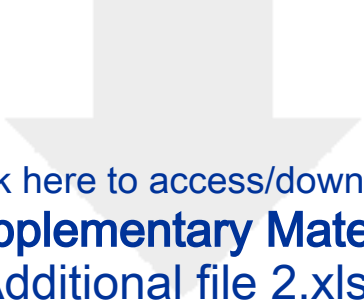
D. mawsoni*E. maclovinus*



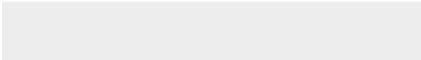




Click here to access/download
Supplementary Material
Additional file 1.docx



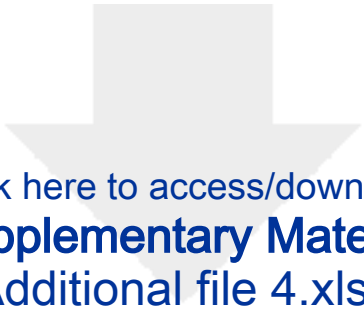


Click here to access/download
Supplementary Material
Additional file 2.xlsx






Click here to access/download
Supplementary Material
Additional file 3.xlsx



Click here to access/download
Supplementary Material
Additional file 4.xlsx



August 31, 2018

Dear Editor,

We wish to submit our manuscript entitled “**Genomic bases for colonizing the freezing Southern Ocean revealed by the genomes of Antarctic toothfish and Patagonia robalo**” for consideration for publication in *Gigascience*.

As a brief background to this work and its importance, the endemic Antarctic notothenioid fishes, which comprise the bulk of fish biomass in the Southern Ocean, are derived from a temperate-water ancestor that was a heavy, bottom-dwelling fish that would not have been able to survive in the freezing temperatures found in the current Southern Ocean. Thus, widespread adaptive genetic change underlies the success of contemporary notothenioids, which resist freezing and have come to occupy all depths of the marine water column.

In most case, owing to the difficulties in sample collection, storage and transportation from distant Antarctica, tissues of the icefish are hard to isolate high quality genomic DNA, which limited the use of new sequencing technologies (such as the technology of high-through chromosome conformation capture (Hi-C)). This is why, till now, only tow icefish genomes (Shin SC, et al. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biol* 15(9), 2014; Ahn et al. Draft genome of the Antarctic dragonfish, *Parachaenichthys charcoti*. *Gigascience*, 6(8):1-6, 2017) was published. In present study, we isolated the HMW genomic DNA and applied the Illumina sequencing technology to perform the whole-genome shotgun sequencing of the two fish genomes through the sequencing libraries with the insert sizes from 170 to 40000 bp. The quality of the assemblies is significantly higher than the two published icefish genomes.

Using the generated genome and annotation data, we compare the genomes of a proxy for the benthic ancestral species, the South American temperate notothenioid *Eleginops maclovinus*, and a neutrally buoyant, highly freeze-resistant Antarctic pelagic species (*Dissostichus mawsoni*) to elucidate the mechanisms underlying the latter species' tolerance of subzero temperatures, neutral buoyancy, and ability to cope with the high dissolved oxygen levels of the Southern Ocean. We reveal heretofore unknown aspects of genome evolution that have led to adaptive radiation of notothenioids. Thus, we show that rapid expansion of transposable elements, adaptive evolution of protein-coding genes, reorganization of development programs for lipid synthesis and bone development, exploitation of new genes (egg chorion proteins) for enhancing freeze-resistance, and expansion of genes encoding proteins that provide anti-oxidant function accompanied the decrease in Southern Ocean water temperatures. Our new data also provide a molecular-mechanistic basis for interpreting a recent trend reported by Rabosky *et al.* (An inverse latitudinal gradient in speciation rate for marine fishes, *Nature* (2018) 559: 3920395) that cold, high-latitude waters are a ‘hotspot’ for fish evolution. In the context, the close temporal link between rapid expansion of transposable elements, which support chromosomal/genetic rearrangements and falling water temperatures is an especially noteworthy and novel discovery. Our discoveries concerning modification of the developmental programs essential for bone generation and lipid deposition also have biomedical implications.

Because of the novelty and broad implications of our discoveries for evolutionary biology, biogeography, developmental physiology, and biomedicine, we believe that the *Gigascience* would be an excellent site for our work to be published.

Sincerely

Liangbiao Chen