

The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonia robalo genomes --Manuscript Draft--

Manuscript Number:	GIGA-D-18-00328R1													
Full Title:	The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonia robalo genomes													
Article Type:	Research													
Funding Information:	<table border="1"> <tr> <td>Natural Science Foundation of China (41761134050)</td> <td>Dr. liangbiao Chen</td> </tr> <tr> <td>Natural Science Foundation of China (31572611)</td> <td>Dr. liangbiao Chen</td> </tr> <tr> <td>Natural Science Foundation of China (31572598)</td> <td>Dr. Qianghua Xu</td> </tr> <tr> <td>Major Science Innovation Grant from the Shanghai Education Committee (2017-01-07-00-10-E00060)</td> <td>Dr. liangbiao Chen</td> </tr> <tr> <td>USA NSF Polar Programs grant (ANT1142158)</td> <td>Dr. Chi-Hing Christina Cheng</td> </tr> <tr> <td>Key Achievement Supporting Grant from Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology</td> <td>Dr. liangbiao Chen</td> </tr> </table>		Natural Science Foundation of China (41761134050)	Dr. liangbiao Chen	Natural Science Foundation of China (31572611)	Dr. liangbiao Chen	Natural Science Foundation of China (31572598)	Dr. Qianghua Xu	Major Science Innovation Grant from the Shanghai Education Committee (2017-01-07-00-10-E00060)	Dr. liangbiao Chen	USA NSF Polar Programs grant (ANT1142158)	Dr. Chi-Hing Christina Cheng	Key Achievement Supporting Grant from Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology	Dr. liangbiao Chen
Natural Science Foundation of China (41761134050)	Dr. liangbiao Chen													
Natural Science Foundation of China (31572611)	Dr. liangbiao Chen													
Natural Science Foundation of China (31572598)	Dr. Qianghua Xu													
Major Science Innovation Grant from the Shanghai Education Committee (2017-01-07-00-10-E00060)	Dr. liangbiao Chen													
USA NSF Polar Programs grant (ANT1142158)	Dr. Chi-Hing Christina Cheng													
Key Achievement Supporting Grant from Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology	Dr. liangbiao Chen													
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 Patagonia robalo <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:	
Response to Reviewers:	<p>Response to reviewers' comments</p> <p>Reviewer #1: Review of "Genomic bases for colonizing the freezing Southern Ocean revealed by the genomes of Antarctic toothfish and Patagonia robalo"</p> <p>This manuscript sequences the genome and transcriptome of Antarctic toothfish and its closest temperate relative, the Patagonian robalo. The authors detect >200 protein gene families that have expanded with functions in stress response and freeze resistance. The amount of work conducted in this study is impressive. I applaud the authors for estimating the genome size using both the k-mer frequency distribution and flow cytometry to compare to the final assembly lengths. I am particularly interested in the expansion of transposable elements correlate to notothenioid diversification. I find no major flaws with the study, but would like to see more detail on the assembly steps (e.g. parameters tested and used in SoapDeNovo and SSPACE).</p> <p>Minor comments: Suggest a review for spelling and English grammar throughout the paper. Thank you for your suggestion. The manuscript has been edited by two professors teach in US universities (Prof. Chi-Hing C. Cheng, University of Illinois, and Prof. George Somero, Stanford University who are native English speakers).</p> <p>L69: is the >90% catch from commercial fisheries? No, it is from many studies based on random sampling in the Southern Ocean. We added this information</p> <p>L141: do the authors know why both the scaffold and contig N50 lengths are so much lower in robalo compared to toothfish?</p> <p>For both species, collected red blood cells were embedded into agarose plugs till isolation of the genomic DNA by the same protocol. However, the Robalo DNA for some unidentified reasons is easier to degrade, which resulted in lower molecular weight of isolated genomics DNA in robalo, which is approximately 25 kb compared to 40 kb in toothfish. Accordingly, sizes of the constructed sequencing libraries ranged</p>

from 170 to 40,000 bp in toothfish, while 170 to 20,000 bp in robalo. Thus spans of the Matepair reads used in scaffolding for the robalo assemblies are smaller than those of the toothfish so that the final scaffold N50 length of the robalo genome is lower. Although we considerably increased the sequencing depth of the paired-end libraries in robalo, size of the assembled contigs were still lower.

In Fig. 4 "kidney" is spelled incorrectly on the "caudal kidney" label
Thanks. This label in Fig.4 has been corrected in the revision.

Reviewer #2: In this study, Chen and colleagues did genome sequencing of two notothenioids to understand genetic basis of Antarctic notothenioids adaptation to the Southern Ocean. I do think that the methodology of the study is sound and findings here are solid. The genomic resources are valuable for further studying genetic basis of the notothenioids adaptive radiation as the authors state. Before recommendation for its final publishing in GigaScience, I have the following concerns that authors might consider. 1) The authors might consider to incorporate the earlier sequenced notothenioid genome. I see that the authors have included the Antarctic bullhead notothen genome in their comparison analyses but do not make fair comparison among the three notothenioid genomes. For example, the authors speculated that TE might contribute to genome size increasing in derived notothenioids by comparison between Antarctic toothfish and Patagonia robalo. What about TEs in the Antarctic bullhead notothen genome? What about LINEs in Antarctic bullhead notothen genome? A thorough comparison among the three notothenioid genomes might give us more information about the topic the authors are trying to beat.

Thank you for suggestion. Comparison of the TEs among Antarctic toothfish, Patagonia robalo and Antarctic bullhead notothen genomes has been added to this revision. Accumulation of TEs are observed in both Antarctic toothfish and bullhead notothen genomes. We thus modified the previous statement regarding TE content and genome size as follows: "The doubling of TE content in the *D. mawsoni* and *N. coriiceps* genomes suggests higher activity of TEs in the Antarctic species in relative to the basal robalo, suggesting a likely contributing factor to the observed trend of increasing genome sizes in more derived Antarctic notothenioid lineages". As for as the timing of LINE insertion is concerned, we calculated their insertion time in the *N. coriiceps* genome by the same methodology as in *D. mawsoni* and *E. maclovinus*. A similar trend of expansion is observed, in *D. mawsoni* and *N. coriiceps*. The corresponding results has been added to Fig.2a in the revision.

2) As findings in their earlier works, the authors find that gene duplication plays an important role in adaptation to the freezing Southern Ocean in notothenioids. I am wondering how many gene families experienced duplication have been identified by both this study and their 2008 PNAS study.

As we stated in the manuscript, "Due to inherent inefficiency in correctly assembling highly similar DNA sequences in the shotgun sequencing strategy, there are likely many more duplicated genes that had eluded detection". From the set of duplicated genes we identified through comparative genome hybridization (Chen et al., 2008) previously, We found 23 protein coding genes are shown significantly duplicated in *D. mawsoni* genome in relative to *E. maclovinus* which was shown in Additional file 1: Fig. S4b. Among these genes included zona pellucida domain containing protein C5 (ZPC5), multiple banded antigen (previously a novel gene), serum lectin isoform 1 precursor (previously FBP32II) and hepcidin.

Many types of ZPs, such as ZPAX1, ZPC1, ZPC2 failed to detect as duplicated in this study, but are known to undergo substantial duplication through array-based genome hybridization and quantitative PCR (Cao et al., 2016), indicating the limitation of the shotgun genome sequencing strategy in finding gene duplications.

References

28. Chen, Z. et al. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. USA* 105, 12944-12949 (2008).
29. Xu, Q. et al. Adaptive evolution of hepcidin genes in antarctic notothenioid fishes.

Mol. Biol. Evol. 25, 1099-1112 (2008).

30 Cao, L. et al. Neofunctionalization of zona pellucida proteins enhances freeze-prevention in the eggs of Antarctic notothenioids. Nat. Commun. 7, 12987 (2016).

3) The authors used an RNA-seq method for their study of transcriptomic adaptation to the freezing environment. However, I do not see any details how they collected the tissues, as we all know that such analysis is very sensitive to the sampling strategy.

Thank you for your suggestion. We added the following information to the Materials and Methods section:

"To obtain tissues from the large-sized *D. mawsoni*, live specimen was anesthetized with MS222 (tricaine methanesulfonate) inside a ambient seawater filled floating sheet plastic tubing in the aquarium tank. The anesthetized specimen was then put on a V-shaped trough for dissection. Tissues were quickly removed and cut into small pieces on ice, and immediately immersed and shaken in ≥ 10 volumes of pre-chilled (-20°C) 90% ethanol (made with 100% pure ethanol and sterilized MilliQ Type 1 water). The ethanol was replaced with a fresh volume within 10 minutes, and again at 2-3 hours and 12 hours later. This preservation method serially desiccates the tissue and effectively inactivates tissue nucleases. The tissue samples were kept in -20°C freezer throughout the serial preservation process and then stored at -20°C until use. To obtain tissues from *E. maclovinus*, MS222 anesthetized specimen was quickly dissected on ice, and preserved in -20°C as described for *D. mawsoni*. The ethanol preserved tissues were shipped back to the University of Illinois on dry ice."

Reviewer #3: Reviewer report.

Title: Genomic bases for colonizing the freezing Southern Ocean revealed by the genomes of Antarctic toothfish and Patagonia robalo

General comments

The authors have sequenced and assembled the genomes of two notothenioids, and have done extensive comparisons with regards to expansions of gene families and differential expression of genes. They show that several genes in the *D. mawsoni* has undergone positive selection, highlighting the evolution of the genes of that species.

Specific comments

Abstract: An extant species is not necessary a proxy for an extinct species. Thank you for your suggestion. The sentence that mentioned the proxy is in Introduction. We corrected it according to the reviewer's comment in revision.

Introduction:

Line 89-90: You specify "whole genome sequence analysis" as the criteria for mentioning the Antarctic rockcod as the only notothenioid reported so far, but Malmstrøm et al 2016 (<https://www.nature.com/articles/ng.3645>) did publish genomic sequences and the assembly of *Chaenocephalus aceratus*. However, they did not report any genomic/biological features of that particular species, so your phrasing is entirely correct.

Thank you for your comments. We have added this citation in the section of Introduction. The corresponding sentence was corrected as: "Thus far, whole genome sequence analysis has been reported for only one notothenioid species, the Antarctic rockcod *Notothenia coriiceps* (Shin et al., 2014). A major histocompatibility complex gene loci from *Chaenocephalus aceratus* was also reported (Malmstrøm M, et al (2016)."

Line 107-8: As you no doubt are aware of, size do not necessary have any bearing on

buoyancy, only average density. It is not apparent to me that smaller size would mean easier to achieve neutral buoyancy.

Throughout the manuscript, we agree that neutral buoyancy is related to the average density of fish, not smaller size. Enhanced lipid storage and promotion of chondrogenesis while inhibiting osteogenesis in bone development play important roles for the *D. mawsoni* to achieve the neutral buoyancy.

We guess the misunderstanding of the reviewer might have resulted from our description on the evolution of smaller ZPC5 molecules in *D. mawsoni*, which is related to the enhanced capability of intracellular freezing-resistance in *D. mawsoni*, NOT related to neutral buoyancy, and nothing to do with body size of the fish.

Results:

Line 138: Why was two different genome assemblers used? Also, in the header for Table S2b it is stated that *E. maclovinus* was assembled with both SOAPdenovo and Platanus.

As we stated in answering reviewer #1's question, *E. maclovinus* DNA extracted from similarly prepared agarose plugs exhibited lower molecular weight than *D. mawsoni* for unknown reasons, which resulted in lower quality of the *E. maclovinus* assemblies. To increase the *E. maclovinus* contig length and decrease algorithm bias when a single assembler was used, we parallelly built the contigs by two assemblers SOAPdenovo and Platanus. The generated contigs were merged prior to the scaffold building.

Line 140 and other places across the manuscript: "Kb", that is, kilo base pairs, should be abbreviated "kb(p)". See: https://en.wikipedia.org/wiki/Metric_prefix

Thank you for your comment. We check the abbreviation of "Kb" and "Kbp" in several journals. "Kb", as the abbreviation of kilo base pairs, is used in most of the journals. But for the abbreviation of less than 1000 base pairs, "bp" is used. So this manuscript better to follow the universal usage, as "Kb".

Line 164: The number of common genes is a bit strange. The vast majority of genes should be common between these species. I think you have written this wrong. In the referred figure, S3, it is specified that the number 8,825 is the amount of common gene clusters, and not just genes. One cluster might contain multiple genes.

Thanks. That should be 8,825 gene clusters, not genes. We have corrected it in the revision.

Lines 182-192: You stated earlier "842 Mb for *D. mawsoni* and 727 Mb for *E. maclovinus*". You could say that quite a bit of that difference in genome size could be due to differences in repeat content, and not just percentage. 161.8 Mbp TEs in *D. mawsoni* and 74.6 Mbp in *E. maclovinus*, with a difference of 86.2 Mbp. It is not apparent that the percentages differences in repeat content actually translates to those large differences in repeats, because these repeat annotations can be quite different (many repeats are not annotated properly in different genomes).

Thank you for your suggestion. Annotation of the TEs in Antarctic toothfish, Patagonia robalo and Antarctic bullhead notothen (added in the revision) are conducted with the same pipelines and criteria. We compared TE contents (%) among the three genomes. Accumulation of TEs are observed in both Antarctic toothfish and bullhead notothen genomes, which may partially contributed the enlargement of genome size in the Antarctic notothenioids. We agree that repeat in genomes may not correctly annotated. We also correct two numbers, the TEs contents of *D. mawsoni* (21.38%) and *E. maclovinus*(10.02%), in this section, where the errors occurred in the previous version due to a mistake when citing from the results from the Additional file 1: Table S9.

Line 613: It is InterProScan, and not InterproScan.
Thanks. We have corrected it in the revision.

Additional Information:

Question	Response
<p>Are you submitting this manuscript to a special series or article collection?</p>	<p>No</p>
<p>Experimental design and statistics</p> <p>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.</p> <p>Have you included all the information requested in your manuscript?</p>	<p>Yes</p>
<p>Resources</p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough 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>Yes</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>Yes</p>

Have you have met the above requirement as detailed in our [Minimum Standards Reporting Checklist?](#)

[Click here to view linked References](#)

1 1 **The genomic basis for colonizing the freezing Southern Ocean**
2
3 2 **revealed by Antarctic toothfish and Patagonia robalo genomes**
4
5
6 3

7 4 Liangbiao Chen^{1,2,3,*}, Ying Lu^{1,2,*}, Wenhao Li^{1,2,*}, Yandong Ren⁴, Mengchao Yu^{1,2}, Shouwen
8 5 Jiang^{1,2}, Yanxia Fu^{1,2}, Jian Wang^{1,2}, Sihua Peng^{1,2}, Kevin T. Bilyk⁵, Katherine R. Murphy⁵,
9 6 Xuan Zhuang⁵, Mathias Hune⁶, Wanying Zhai^{1,2}, Wen Wang⁴, Qianghua Xu^{1,2,§}, Chi-Hing
10 7 Christina Cheng^{5,6,§}
11
12
13
14
15

16 9 ¹National Demonstration Center for Experimental Fisheries Science education at Shanghai
17 10 Ocean University, Shanghai, China, ²Key Laboratory of Exploration and Utilization of
18 11 Aquatic Genetic Resources (Ministry of Education) and International Research Center for
19 12 Marine Biosciences (Ministry of Science and Technology) at Shanghai Ocean University,
20 13 Shanghai, China, ³Laboratory for Marine Biology and Biotechnology, Qingdao National
21 14 Laboratory for Marine Science and Technology, Qingdao, China, ⁴Kunming Institute of
22 15 Zoology, Chinese Academy of Sciences, Kuming, China, ⁵Department of Animal Biology,
23 16 University of Illinois at Urbana-Champaign, IL, USA, ⁶Fundación Ictiológica, Providencia,
24 17 Santiago, Chile.
25
26
27
28
29
30
31
32
33

34 19 * These authors contributed equally to this work.
35
36
37
38

39 21 [§]Corresponding authors:

40 22 Qianghua Xu (qhXu@shou.edu.cn).

41 23 Chi-Hing Christina Cheng (c-cheng@illinois.edu).

42 24
43 25 ORCID IDs:

44 26 Qianghua Xu, ORCID: 0000-0003-0351-1765

45 27 Chi-Hing Christina Cheng, ORCID: 0000-0001-6308-8685

46 28 Katherine R. Murphy, ORCID: 0000-0001-8898-8719

47 29 Kevin T. Bilyk, ORCID: 0000-0002-2262-2703
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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 **Abstract**

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

55
56 **Keywords:** adaptive radiation, climate change, genome, oxidative stress, Antarctic
57 notothenioids

58
59 **Introduction**

60
61 The Southern Ocean (SO) surrounding Antarctica is the coldest body of water on Earth,
62 having been isolated from other world oceans by the Antarctic Circumpolar Current (ACC)
63 beginning in the early Oligocene ~32 million years ago (1). The formation of the ACC also
64 impeded species dispersal across the Antarctic Polar Front, and mass extinction of the

1 65 Antarctic-sequestered fish taxa occurred upon marine glaciation (2). The rich cosmopolitan
2 66 fish fauna prior to the isolation of Antarctica is represented today by a single predominant
3
4 67 group of related fish species - the Antarctic notothenioids. From a common temperate
5
6 68 ancestor, likely a swim-bladderless, bottom dwelling perciform species of the Eocene age (3),
7
8 69 the Antarctic notothenioids have evolved to become highly adapted to life in unyielding cold,
9
10 70 spectacularly diverse in sizes and morphological innovations, and diversified into all water
11
12 71 column habitats, epitomizing an adaptive radiation and a rare marine species flock (4).
13
14 72 Abundant in fish biomass (>90% of random catch) and species (≥ 128), they are vital in
15
16 73 sustaining the contemporary SO food web (2, 5).

17 74 What evolutionary processes and mechanisms propelled the Antarctic notothenioid
18
19 75 radiation replete with extraordinary trait diversification during its evolutionary history remain
20
21 76 fascinating unanswered questions. Two conspicuous trait outcomes - the evolutionary gain of
22
23 77 the novel antifreeze glycoprotein (AFGP) gene and function that averted otherwise
24
25 78 inescapable death from freezing (6, 7), and exploitation of open niches vacated by extinction
26
27 79 of fishes lacking freeze-resistance, have been recognized as major contributors to the
28
29 80 Antarctic notothenioid radiation (8, 9). However, little is known of the myriad subtler
30
31 81 adaptive changes that must also have evolved in response to challenges from freezing
32
33 82 temperatures and the associated high oxygen concentration – the two foremost modalities of
34
35 83 selection pressure from the SO environments that would pervade all levels of organismal
36
37 84 functions, from molecules to cells to system physiology. Another prominent hallmark of
38
39 85 notothenioid adaptive radiation is the secondary acquisition of pelagicism in some lineages,
40
41 86 enabling their ecological expansion from bottom habitats of their negatively buoyant ancestor
42
43 87 to upper water column niches. What evolutionary changes occurred in the cellular and
44
45 88 developmental programs that enabled neutral buoyancy and secondary pelagicism are also
46
47 89 unknown.

48 90 To address these fundamental, system-wide questions about Antarctic notothenioid
49
50 91 evolution, whole genome sequences of multiple and appropriately chosen species from the
51
52 92 diverse Antarctic notothenioids are essential. Thus far, whole genome sequence analysis has
53
54 93 been reported for only one notothenioid species, the Antarctic rock cod *Notothenia coriiceps*
55
56 94 (10). A major histocompatibility complex gene loci from *Chaenocephalus aceratus* was also
57
58 95 reported (11). The *N. coriiceps* genome provided the key inference that the fast evolving
59
60 96 hemoglobin and mitochondrial proteins are adaptive in increasing efficiency of aerobic
61
62 97 cellular respiration in the freezing environment. *N. coriiceps* is not known to occur in the high
63
64 98 latitude Antarctic coastal waters. Instead, it is widely distributed in the lower latitude waters
65

1 99 of the Antarctic Peninsula archipelago and the Scotia Arc islands, reaching localities north of
2 100 the Polar Front around sub-Antarctic islands in the Indian Ocean sector (12), a distribution
3
4 101 pattern that suggests a considerable degree of thermal plasticity in this species. It is a heavy,
5
6 102 bottom fish and one of the hardest boned Antarctic notothenioids (13), reminiscent of the
7
8 103 benthic ancestor. To gain insights into evolutionary adaptations in the most cold-adapted and
9
10 104 stenothermal Antarctic notothenioids, as well as into the evolutionary changes leading to
11
12 105 acquisition of neutral buoyancy that enabled the transition from the ancestral benthic
13
14 106 existence to a pelagic life history, a different and more appropriate model Antarctic
15
16 107 notothenioid species would be required.

17 108 The Antarctic toothfish *Dissostichus mawsoni* (NCBI: txid6530, Fishbase ID:7039) that
18
19 109 grows to giant sizes (2.0 m in length and 140 kg in mass) is an iconic species of the Antarctic
20
21 110 notothenioid radiation, with wide distributions in freezing waters of high latitude Antarctic
22
23 111 coasts, as far south as 77.5°S (McMurdo Sound), the southern limit of Antarctic marine life. It
24
25 112 thus exemplifies the stenothermal cold-adapted character state. Despite its large size, it is the
26
27 113 only notothenioid species that achieved complete neutral buoyancy as adults (14, 15); thus
28
29 114 this species serves as the best model for examining the evolutionary underpinning of
30
31 115 secondary pelagicism in the Antarctic clade. In addition, to discern evolutionary changes from
32
33 116 the ancestral temperate state to the derived polar state driven by selection in the cold,
34
35 117 oxygen-rich Southern Ocean environment, a closely related basal non-Antarctic notothenioid
36
37 118 comparison species would strongly add to the discriminating power of analyses of genome
38
39 119 evolution. The most appropriate species for this purpose is a South American notothenioid,
40
41 120 the Patagonian robalo *Eleginops maclovinus* (NCBI: txid56733, Fishbase ID:466), which is
42
43 121 the sole species in the basal family Eleginopsidae. The lineage diverged prior to the isolation
44
45 122 of Antarctica, and *E. maclovinus* is phylogenetically the closest sister species to the modern
46
47 123 Antarctic clade (3). Thus its genome is the best representative of the temperate character of
48
49 124 the most recent common ancestor of the Antarctic notothenioids. We conducted genome
50
51 125 sequencing and comparative analyses of these two strategically selected species accompanied
52
53 126 with extensive transcriptomic characterizations to profile relevant functional outcomes of the
54
55 127 genomic changes. Our results provide several new key insights into evolutionary adaptation
56
57 128 and secondary pelagicism of the Antarctic notothenioids in the isolated and extremely cold
58
59 129 Southern Ocean environment.

60 130

61 131

62 132 **Results and Discussion**

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

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

162
163 **Genome annotation and synteny alignment between *D. mawsoni* and *E. maclovinus***

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

167 protein coding genes of the toothfish and robalo, along with the sequenced notothenioid *N.*
168 *coriiceps* and model species *G. aculeatus* and zebrafish *Danio rerio* were clustered using
169 OrthoMCL(22). We found 8,825 gene clusters that were common to all five species. Genes
170 shared among the notothenioids are similar in number, 12,269 between toothfish and robalo,
171 and 12,421 between toothfish and *N. coriiceps* (**Additional file 1: Fig. S3**). In annotations of
172 conserved non-coding RNA genes, we predicted 1,097 tRNA, 110 rRNA, 422 SnRNA, and
173 295 microRNA genes in the toothfish genome (**Additional file 1: Table S6a**), while the
174 robalo genome was annotated to carry 1,037 tRNA, 44 rRNA, 891 snRNA and 286 miRNAs
175 (**Additional file 1: Table S6b**). The much larger number of rRNA copies (2.5 fold) in *D.*
176 *mawsoni* than *E. maclovinus* is consistent with the presence of dual chromosomal loci of
177 rDNA genes detected by *in situ* fluorescent hybridization in the giant toothfish (16), as
178 opposed to the single rDNA locus in other notothenioids (23). The two species showed
179 largely similar profiles in their microRNAome, with minor differences in the copy number of
180 some individual microRNA (**Additional file 1: Table S7**). Strikingly, the toothfish genome
181 contains many more selcys-tRNA genes than robalo (84 versus 1; **Additional file 1: Table**
182 **S8**). This extensive duplication of selcys-tRNA genes, accompanied with high expression of
183 selenoproteins in *D. mawsoni* (detailed in a later section) signify that mitigation of oxidative
184 stress through selenoproteins, many of which are strong antioxidants, is likely an important
185 selection force in the evolution of the Antarctic notothenioid genome in the freezing and
186 oxygen-rich waters.

187 *De novo* annotation of repeat sequences revealed >2-fold increase in overall repeat
188 content in *D. mawsoni* (21.38% of genome) over the basal *E. maclovinus* (10.02%)
189 (**Additional file 1: Table S9**). Transposable elements (TEs) of the toothfish genome,
190 including LTR-retrotransposons, non-LTR-retrotransposons (long interspersed transposable
191 elements - LINEs, and short interspersed transposable elements - SINEs), and DNA
192 transposons, was more than twice that of *E. maclovinus*. Examination of the *N. coriiceps*
193 genome exhibited similar accumulation of the TEs (23.5%). Among simple sequence repeats
194 (SSRs), dimer repeats constituted the majority of the SSRs in both genomes, and tetramers
195 and pentamers showed highest levels of increment in *D. mawsoni* (**Additional file 1: Table**
196 **S10a, Table S10b**). The doubling of TE content in the *D. mawsoni* and *N. coriiceps* genomes
197 suggests higher activity of TEs in the Antarctic species in relative to the basal robalo,
198 suggesting a likely contributing factor to the observed trend of increasing genome sizes in
199 more derived Antarctic notothenioid lineages (24).

1 200 For global genome alignment, we anchored the *D. mawsoni* and *E. maclovinus* scaffolds
2 201 to the 21 linkage groups of the well characterized three-spined stickleback genome according
3 202 to gene collinearity (**Fig. 1b; Additional file 2**). Most of the notothenioid scaffolds had
4 203 extensive collinearity to the corresponding stickleback chromosomes. The 84 duplicated
5 204 selcys-tRNA genes are wide-spread throughout the toothfish genome. Inversion and
6 205 translocation of genome segments have occurred in both *D. mawsoni* and *E. maclovinus*
7 206 relative to stickleback, but the *D. mawsoni* genome showed more frequent rearrangements.
8 207 Insertion sites for the expanded TEs of the *D. mawsoni* genome were random, thereby
9 208 expanding the lengths of almost all of the linkage groups. Mapping RNAseq transcript data
10 209 we obtained from white muscles for the two species on their respective synteny showed that
11 210 the heavier insertion of TEs in the toothfish genome did not appear to adversely affect the
12 211 expression of the neighboring protein genes as strong gene expression in regions of high TE
13 212 content was maintained (**Fig. 1b**).
14 213

15 214 **Burst of LINE expansion in the cold**

16 215 Involvement of gain (or loss) of mobile element copies in genome size and genome
17 216 re-structuring affecting species differentiation has increasingly gained empirical support (25,
18 217 26). TEs are normally under epigenetic regulation, but waves of TE proliferations could arise
19 218 from environmental changes that cause physiological stress and disrupt the epigenetic control
20 219 (27). We therefore examined potential linkage in timing between LINE expansion (2-fold
21 220 increase in the toothfish versus robalo) and onset of frigid SO marine conditions. A multiple
22 221 sequence alignment of several LINES (LINE/I, LINE/L2, LINE/RTE-BovB and
23 222 LINE_Rex-Babar – the most abundant types in the toothfish genome) was made, which
24 223 clustered 349 *D. mawsoni* LINE pairs and 213 *E. maclovinus* LINE pairs, where the two
25 224 LINES within each pair share the highest sequence similarity, approximating the least
26 225 divergence time. Calculation of the nucleotide substitution rates for each LINE pair identified
27 226 a burst of emergence of LINES in the *D. mawsoni* genome with the substitution rates centered
28 227 at 0.04, which corresponded with a divergence time of 6.5 million years ago (**Fig. 2a; see also**
29 228 **Methods**) based on the average substitution rate of the LINES. A similar trend of LINE
30 229 expansion was also observed in the *N. coriiceps* genome. This estimated timing of LINE burst
31 230 expansion correlated with the radiation of the majority of the modern Antarctic notothenioid
32 231 clades beginning in the late Miocene when seawater temperatures steadily declined (9). In
33 232 contrast, no burst expansion of LINES was detected in the *E. maclovinus* genome, supporting
34 233 an Antarctic/cold-specific LINE burst in the toothfish, and corroborating our prior empirical
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

1 234 evidence for an increase in retrotransposition activity of *D. mawsoni* LINE-1 resulting in
2 235 more copies in the transfected cells when subjected to stress from non-physiologically low
3
4 236 incubation temperature (28).

5 237

6 238 **Accelerated protein evolution in the cold**

7
8
9 239 Ectotherms are vulnerable to disruptive effects on protein structures and reaction rate
10 240 depression at low temperatures. Antarctic notothenioids in perennially freezing high latitude
11 241 waters face the extremes of these effects, as well as an oxidative environment due to high
12 242 oxygen concentrations resulting from increased gas solubility at low temperatures. We
13 243 examined evidence for adaptive evolution of proteins in response to these selection pressures.
14 244 The gene models of *D. mawsoni*, *E. maclovinus*, seven other teleost genomes and the mouse
15 245 genome as outgroup were clustered, from which 2,936 one-to-one single-copy orthologs were
16 246 obtained to reconstruct a phylogenetic tree (**Fig. 2b**). Stickleback is the closest species to the
17 247 notothenioid clade as expected, and *E. maclovinus* is sister to the two Antarctic notothenioids,
18 248 *D. mawsoni* and *N. coriiceps*. We found evolutionary rates of the orthologous genes based on
19 249 calculated *dN/dS* values (the ratio of the rate of non-synonymous substitution to the rate of
20 250 synonymous substitution) were elevated in the notothenioid lineage compared to the other
21 251 fish lineages. This faster rate is more pronounced in the two Antarctic species, about twice
22 252 that of the basal *E. maclovinus*, suggesting intensified selection pressures driving genome
23 253 evolution in the Antarctic environment. To identify Gene Ontology categories that were
24 254 evolving faster in the toothfish or robalo, the *dN/dS* ratios of 7,584 orthologous genes among
25 255 the two notothenioids and the stickleback were calculated and the average *dN/dS* value of the
26 256 genes associated with each GO term was calculated for each species. These orthologous genes
27 257 were annotated to 411 GO terms, of which 281 showed significantly higher average *dN/dS*
28 258 ratios in *D. mawsoni*, while only 19 demonstrated higher average rates in *E. maclovinus* (**Fig.**
29 259 **2c; Additional file 2**). The faster evolving GO processes in *D. mawsoni* included “gene
30 260 expression”, “protein folding”, “tRNA metabolic process”, “cell-redox homeostasis”,
31 261 “immune response”, “response to stress”, “lipid metabolic process”, “DNA repair”,
32 262 “vesicle-mediated transport” and others. To assess which *D. mawsoni* genes experienced
33 263 positive selection, we tested using the Branch-site model (in PAML) on a reconstructed
34 264 phylogenetic tree of six fish species with the *D. mawsoni* lineage assigned as the foreground
35 265 branch (**Additional file 1: Fig. S4a**). A total of 526 positively selected genes (PSGs) were
36 266 identified (**Additional file 1: Table S11a**). The most significantly enriched KEGG pathway
37 267 (**Additional file 1: Table S11b**) was “Protein digestion and absorption” ($p = 5.1 \times 10^{-5}$) and

1 268 GO term (**Additional file 1: Table S11c**) was “protein binding” ($p = 6.0 \times 10^{-4}$), which
2 269 indicate that maintenance of protein homeostasis played an important role in shaping the *D.*
3 270 *mawsoni* genome. In addition, a complement of genes (*sorbs2a*, *acox1*, *apoa1a*, *scp2a*, *tnni3k*
4 271 and perilipin-like genes; **Additional file 1: Table S11a**) involved in the PPAR (peroxisome
5 272 proliferator-activated receptors) signaling pathway - the key pathway in adipocyte
6 273 development regulation - were found to be under positive selection, suggesting occurrence of
7 274 adaptive changes in lipid metabolism in *D. mawsoni*.

8 275

9 276 **Gene duplication in the freezing environment**

10 277 Gene duplication plays fundamental roles in emergence of adaptive features. In the list of
11 278 predicted protein coding genes from the toothfish and robalo genomes, we identified 202
12 279 families that have increased in copy number in *D. mawsoni*, compared to the other eight fish
13 280 genomes (**Additional file 3**). KEGG enrichment analyses of these expanded gene families
14 281 yielded enrichment in pathways involved in protein homeostasis and lipid and bone
15 282 metabolism, such as “Protein digestion and absorption”, “Regulation of lipolysis in
16 283 adipocyte”, “Fat digestion and absorption”, “Ether lipid metabolism” and “Osteoclast
17 284 differentiation” (**Fig. 2d; Additional file 2**), suggesting genomic capacity for these functional
18 285 pathways had increased in *D. mawsoni* during evolution in chronic cold. Corroborating this
19 286 cold-specificity is that the expanded gene families found in the basal *E. maclovinus* relative to
20 287 the other fish genomes, including *D. mawsoni* and *N. coriiceps*, yielded distinctly different
21 288 enriched KEGG profiles (**Fig. 2e**); this analysis indicates that the functional traits gained
22 289 through gene duplication in *E. maclovinus* were driven by different selective pressures,
23 290 consistent with our previous findings (29, 30). Due to inherent inefficiency in correctly
24 291 assembling highly similar DNA sequences in the shotgun sequencing strategy, there are likely
25 292 many more duplicated genes that had eluded detection. For example, many paralogs of Zona
26 293 Pellucida protein (ZPs), such as ZPAX1, ZPC1, ZPC2 in *D. mawsoni* have been shown,
27 294 through array-based genome hybridization and quantitative PCR, to be duplicated (29, 31).
28 295 Among the set of genes found duplicated in our previous report (28), 23 are identified in this
29 296 study (Additional file 1: Fig. S4b).

30 297 Gene duplication had contributed prominently to the evolutionary gain of freezing
31 298 avoidance in Antarctic notothenioids. Generally regarded as a key innovation of the Antarctic
32 299 notothenioid radiation, the AFGP (antifreeze glycoprotein) gene evolved from a
33 300 trypsinogen-like protease ancestor followed by extensive intragenic and whole gene
34 301 duplications, generating a large gene family that would provide an abundance of this novel

1
2
3
4
5
6
7
8
302 life-saving protein (7). By referencing to the published AFGP haplotypes that were assembled
303 from BAC clone sequences of the same individual used in this study (Genbank accessions
304 HQ447059 and HQ447060) (32), we identified and assembled the AFGP loci shotgun reads,
305 recaptured the two AFGP haplotypes and integrated them in the draft genome, and localized
306 them to a region syntenic with a scaffold in LG 20 of the stickleback genome (**Fig.1b**).

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

46 47 328 **Transcriptomic adaptation to the cold environment**

48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
329 To assess the functional relevance of the detected genomic outcomes to life in freezing
330 condition, we characterized and compared transcriptomes of 12 tissues including brain, liver,
331 red muscle, white muscle, gill, skin, intestine, stomach, spleen, head kidney, caudal kidney
332 and ovary between native specimens of *D. mawsoni* and *E. maclovinus*. We found over ten
333 thousand genes were differentially expressed in pairwise comparisons between the two
334 species, with the toothfish showing substantially more up-regulated genes than the robalo.
335 Enrichment test on KEGG pathways yielded many signaling pathways in the tissues of

1 336 toothfish being significantly enriched in differentially expressed genes (DEGs), including the
2 337 TGF-beta, AMPK and PPAR pathways, known to play essential roles in development,
3
4 338 metabolism and stress responses (**Additional file 1: Fig. S8a**). GO enrichment analysis of the
5
6 339 DEGs demonstrated up-regulation of hundreds of GO biological processes, including
7
8 340 translation, transferrin transport, cell redox homeostasis, cellular response to unfolded protein,
9
10 341 ubiquitin-dependent protein catabolic process, regulation of innate immunity, MAPK cascade,
11
12 342 positive regulation of apoptosis pathways, many of the pathways involved in lipid metabolism,
13
14 343 and anti-ROS pathways represented by “selenium compound metabolic process” and
15
16 344 “selenocysteine metabolic process” (**Additional file 1: Fig. S8b**). These results corroborated
17
18 345 the expression profiles in a previous study with the lower depth of sequencing available at
19
20 346 that time (29), and additionally revealed further details of transcriptomic cold adaptation from
21
22 347 the much deeper sequencing across a comprehensive set of tissue transcriptomes. A striking
23
24 348 finding was the greatly increased transcriptional activities across many selenium-containing
25
26 349 protein genes in the toothfish tissues compared with robalo (**Fig. 4; Additional file 1: Fig.**
27
28 350 **S9**). Correspondingly, the genes involved in the translation of selenocysteine-containing
29
30 351 proteins were also significantly up-regulated in the toothfish (**Fig. 4**). Selenoproteins have
31
32 352 well-known functions in coping with cellular oxidative stress, thus the great expansion of
33
34 353 selcys-tRNA genes (**Fig. 1b, Additional file 1: Table S8**), and the significantly upregulated
35
36 354 expression of many kinds of selenoprotein mRNAs indicate augmented anti-ROS capacity has
37
38 355 evolved as an important adaptation to the constantly freezing environment, where saturated
39
40 356 levels of O₂ and cold-depressed metabolic rates would make oxidative stress a formidable
41
42 357 challenge for cellular life. Interestingly, the expression of glutathione peroxidase 4b (gpx4b)
43
44 358 (**Fig. 4**), a selenoprotein uniquely able to reduce lipid hydroperoxides (33, 34), was lower in
45
46 359 *D. mawsoni*, suggesting alternative lipid metabolic programs may exist in the toothfish.
47
48 360 Accordingly, we found all isoforms of the major players in lipid droplet assembly (PLN2,
49
50 361 PLN5, fitm, seipin) important for lipid storage in adipose tissue, were upregulated in all
51
52 362 toothfish tissues examined relative to *E. maclovinus*, signifying a shift of lipid distribution
53
54 363 towards storage in *D. mawsoni*, as described in detail below.

50 364

52 365 **Altered lipid metabolism in *D. mawsoni* for neutral buoyancy**

54 366 In a striking evolutionary departure from the heavy, bottom-water ancestral character
55
56 367 (exemplified by *E. maclovinus*), a handful of Antarctic notothenioids have secondarily
57
58 368 acquired neutral or near neutral buoyancy, enabling ecological diversification into and filling
59
60 369 of mid-water niches – a distinctive hall mark of the Antarctic notothenioid adaptive radiation.

1 370 The giant toothfish *D. mawsoni*, despite growing to massive sizes, being robustly muscled,
2 371 and lacking a swim bladder, is the only notothenioid that has attained complete neutrally
3 372 buoyancy (14, 15). Known morphological specializations include extensive lipid (mostly
4 373 triglycerides) deposits under skin and in the musculature, and a light skeleton of mostly
5 374 cartilage and little mineralized bone, adaptations that reduce overall density and provide static
6 375 lift (14). To understand the genetic basis of the large accumulation of lipids and reduced
7 376 mineralization in *D. mawsoni*, we carried out transcriptome comparisons between *D. mawsoni*
8 377 and several other notothenioids in which neutral buoyancy is not developed.

9 378 We compared gene expression profiles of muscles of *D. mawsoni*, and of the negatively
10 379 buoyant Antarctic *N. coriiceps* and the basal *E. maclovinus* to elucidate evolutionary
11 380 differences in the mechanisms of intermuscular lipid deposit. Compared with *E. maclovinus*,
12 381 genes involved in triacylglycerol synthesis in the toothfish muscle were overrepresented in
13 382 DEGs ($p < 0.05$) and markedly upregulated in transcription, including the key enzymes
14 383 acylglycerol-3-phosphate O-acyltransferase (AGPAT) isoforms and CDP-diacylglycerol
15 384 synthase (CDS) (**Additional file 1: Fig. S10**). An important regulator of this process, lipin1,
16 385 was downregulated in *D. mawsoni*. Lipin1 is known to exert dual effects on lipid metabolism
17 386 - it acts as a phosphatidate phosphatase enzyme to form diacylglycerol required for lipid
18 387 synthesis, but also serves as a transcriptional co-activator to promote fatty acid oxidation (35).
19 388 The down-regulation of lipin1 was consistent with down-regulation of fatty acid oxidation in
20 389 *D. mawsoni*, as about half of the genes involved in fatty acid oxidation were also
21 390 down-regulated relative to the robalo (**Additional file 1: Table S12**). At the same time, genes
22 391 involved in regulation of lipid storage were overrepresented in the DEGs ($p < 0.05$) and all
23 392 but one gene (MEST) were upregulated in the toothfish (**Additional file 1: Table S13**). These
24 393 data strongly suggest a shift of metabolic pathways from lipid breakdown to lipid biosynthesis
25 394 and lipid storage in *D. mawsoni* muscle relative to *E. maclovinus*, favoring deposit of lipids,
26 395 thus contributing to neutral buoyancy. Compared with *N. coriiceps*, *D. mawsoni* muscle
27 396 showed an overall trend of upregulation of genes involved in glycerolipid biosynthesis and
28 397 lipid storage, but the differences are not as dramatic as in the *D. mawsoni*/*E. maclovinus*
29 398 comparison. Expression levels of many genes relevant to lipid oxidation were fairly similar
30 399 between the two species, suggesting common downregulation in lipid oxidation in the
31 400 Antarctic species compared to the temperate *E. maclovinus* (**Additional file 1: Fig. S10**;
32 401 **Additional file 1: Tables S12, S13**). Consistent with downregulation of the lipid oxidation in
33 402 the muscles of the two Antarctic fishes were their lower expression levels of lipid oxidation
34 403 mitigating selenoenzyme gpx4 compared to the robalo. In total, the transcriptome

1 404 comparisons revealed substantial genetic reprogramming in *D. mawsoni* muscle that would
2 405 favor the large lipid deposition in this species.

3
4 406 GO enrichment tests on the muscle DEGs also indicated regulatory change ($p = 0.073$) in
5 407 fat cell differentiation between *D. mawsoni* and *E. maclovinus*. Adipogenesis in almost all
6 408 animals is predominately regulated by peroxisome proliferator activated receptor gamma
7 409 (PPAR γ)(36). Expression of PPAR γ in *D. mawsoni* muscle was upregulated by more than
8
9 410 5.6-fold compared to *E. maclovinus* (**Additional file 1: Table S14**). In addition, as many as
10 411 16 known pro-adipogenetic factors were upregulated from 1 to 8 fold. Several factors in the
11 412 TGF-beta (TGFB1, smad3), Wnt (Sirt1, sirt2, frizzled-related protein (FRZB)) and Notch
12 413 (jag1b, Hes1) pathways and Jun dimerization protein 2 (JDP2), reportedly negative regulators
13 414 of adipogenesis were also upregulated. When compared with *N. coriiceps*, about half of the
14 415 DEGs in the *D. mawsoni/E. maclovinus* comparison under this GO term were statistically
15 416 insignificant, but remarkably the comparison yielded significant upregulation of 10
16 417 pro-adipogenetic factors in *D. mawsoni* muscle, including the most important regulator,
17 418 PPAR γ (**Fig. 5a**), and only one negative regulating factor (TGFB1) (37, 38). These results
18 419 strongly suggest that regulatory promotion of adipogenesis in *D. mawsoni* muscle is a key
19 420 contributing factor to fat deposition and attainment of neutral buoyancy.
20
21 421

22 422 **Reduction of ossification in *D. mawsoni***

23 423 To reveal the genes involved in the reduced bone ossification in *D. mawsoni*, we compared
24 424 the transcriptomes of pelvic girdle bones between *D. mawsoni* (0% body weight in seawater)
25 425 and the negatively buoyant Antarctic notothenioids *Trematomus bernacchii* and *Pagothenia*
26 426 *borchgrevinki*, (3.52% and 2.75% of body weight, respectively, in seawater) (14). A total of
27 427 1,733 DEGs showing the same direction of change in the *D. mawsoni/P. borchgrevinki* and
28 428 the *D. mawsoni/T. bernacchii* comparisons were identified and used for further analysis
29 429 (**Additional file 3**). We found that 48 genes encoding various ribosomal proteins were
30 430 significantly reduced in expression in the *D. mawsoni* bone, suggesting either a lower protein
31 431 translation activity or fewer metabolically active cells in the pectoral girdle than in the other
32 432 two nototheniids. The DEGs were enriched with hundreds of GO biological processes,
33 433 including “extracellular matrix”, “ossification”, “response to hypoxia”, “angiogenesis” and
34 434 “lipid storage”, indicating multiple genetic programs were distinctly regulated in the toothfish
35 435 bone (**Additional file 1: Fig. S11**). In terms of ossification, it was noteworthy that
36 436 expressions of the two major regulators of vertebrate bone development, *sox9* and *runx2* (39)
37 437 were not significantly altered among the three notothenioids. However, expression of many
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

1 438 genes of the BMP pathways, wnt pathways and many regulatory factors known to be involved
2 439 in the process were specifically altered in *D. mawsoni*, which likely shifted the developmental
3
4 440 balance between chondrogenesis and osteogenesis (**Fig. 5b; Additional file 1: Table S15**).
5 441 Among these highly upregulated genes (depicted in **Fig. 5b**), CTGF (Connective Tissue
6 442 Growth Factor) has been implicated in early events of osteogenic differentiation including
7 443 proliferation and recruitment of osteoprogenitors, however, when expressed constitutively,
8 444 CTGF would inhibit both Wnt-3A and BMP-9 induced osteoblast differentiation (40). HSPG2
9 445 (prostaglandin-endoperoxide synthase 2) is required for the chondrogenic and adipogenic
10 446 differentiation from synovial mesenchymal cells via its regulation of sox9 and PPAR γ , but not
11 447 for osteogenic differentiation via *runx2* (41), and ECM1 (extracellular matrix protein 1)
12 448 interacts with HSPG2 to regulate chondrogenesis (42). MEF2C, a transcription factor that
13 449 regulates muscle and cardiovascular development, controls bone development by activating
14 450 the genetic program for chondrocyte hypertrophy (43). Some of the upregulated genes are
15 451 known to inhibit osteoblastogenesis, such as Tob2 (44), CTNNBIP1 (45), secreted
16 452 frizzled-related protein 1 (SRFP1) (46) and ZBTB16 (47). Some members of the TGF-beta
17 453 superfamily (BMPR1a, TGF-beta1, SMAD1), which were upregulated in *D. mawsoni* are
18 454 known to promote both chondrogenesis and osteoblastogenesis (48). CYR61 and PTN
19 455 specifically promote osteoblastogenesis (49, 50), but expression of PTN is drastically reduced,
20 456 consistent with reduced hard bone formation. A few genes influence ossification via
21 457 regulating osteoclastogenesis, for example, *Sbno2*-promotes osteoclast fusion (51) and
22 458 activation of the EphA2 signaling on osteoblasts led to bone reabsorption(52). We found
23 459 genes associated with osteoclast differentiation are significantly enriched in the DEGs (p
24 460 <0.05) and all were upregulated (**Additional file 1: Tables S16, S17**). Overall, the gene
25 461 expression patterns in the toothfish bone demonstrated a genetic shift to chondrogenesis over
26 462 osteoblastogenesis in bone development, which would reduce bone density and contribute to
27 463 achieving neutral buoyancy.

28 464 Studies have indicated that the majority of clinical conditions associated with human
29 465 bone loss are accompanied by increased marrow adiposity possibly due to shifting of the
30 466 balance between osteoblast and adipocyte differentiation in bone marrow stromal (skeletal)
31 467 stem cells (53). A few signaling pathways such as the TGF-beta/BMP pathways and the Wnt
32 468 pathway (represented by CTNNBIP and SFRP1 in this case) are known to participate in
33 469 regulation of both bone and adipocyte development in animals. In the toothfish, we found
34 470 enriched GO terms relevant to regulation of “response to lipid” and “lipid storage”

1 471 (Additional file 1: Table S12) indicating possible linkage in the regulatory network that
2 472 orchestrates the loss of ossification and gain of lipids in *D. mawsoni* bones.

3
4 473 To verify whether the elevated transcription of the regulatory factors indeed resulted in
5 474 more abundant protein, we selected the factor CTGF for immunohistochemical staining in the
6
7 475 bone and surrounding tissues of pelvic fins of *D. mawsoni* and *E. maclovinus* since it is the
8
9 476 only factor for which an effective monoclonal antibody is currently available. Much stronger
10
11 477 signal was detected in the *D. mawsoni* fin tissue (Fig. 5c), supporting a correlation between
12
13 478 protein abundance and mRNA transcription in the case of CTGF. This result further supports
14
15 479 the involvement of CTGF in the reduced ossification in *D. mawsoni*.

16 480

18 481 **Conclusions**

19
20 482 We sequenced and compared the genomes and transcriptomes of the cold-adapted
21
22 483 high-latitude Antarctic toothfish *D. mawsoni* and the basal temperate relative *E. maclovinus*
23
24 484 representing the ancestral character state to deduce Antarctic-specific evolutionary and
25
26 485 adaptive changes supporting physiological activities of notothenioid fishes in freezing and
27
28 486 oxygen rich Southern Ocean waters, as well as the gain of secondary pelagicism fundamental
29
30 487 to Antarctic notothenioid niche expansion and adaptive radiation. The assembled genomes
31
32 488 achieved 90% (*D. mawsoni*) and 95% (*E. maclovinus*) coverage of the respective genome size
33
34 489 estimated by cell flow cytometry, and with greater scaffold N50 than the currently available
35
36 490 sole Antarctic notothenioid (*N. coriiceps*) genome, greatly enhancing comprehensive,
37
38 491 genome-wide discovery of evolutionary processes.

39
40 492 We found two-fold expansion of TEs in the Antarctic toothfish over the temperate robalo
41
42 493 *E. maclovinus* and deduced the timing of a burst of one major class of TEs (LINEs) to about
43
44 494 6.5 mya, temporally correlating with the late Miocene onset of steady cooling trend of the
45
46 495 Southern Ocean (SO) and diversification of the modern Antarctic notothenioid clade,
47
48 496 suggesting a role of cold-induced TE expansion in notothenioid speciation. We found many
49
50 497 of the protein coding genes in the toothfish evolved rapidly and experienced positive selection,
51
52 498 among which genes relevant to preservation of protein homeostasis were particularly
53
54 499 prominent. Multiple gene families have undergone duplication during evolution in the cold,
55
56 500 as exemplified by genes that confer resistance to freezing in the cold SO waters: the AFGP
57
58 501 family that evolved *de novo* and confers extracellular freeze avoidance, and duplicated zona
59
60 502 pellucida ZPC5 genes that functionally diversified to aid in cellular freezing resistance.
61
62 503 Through transcriptome comparisons, we found functional output of the cellular apparatus for
63
64 504 selenoprotein production in the Antarctic toothfish was greatly elevated compared to the basal
65

1 505 temperate robalo, suggesting evolutionary mobilization of antioxidant selenoproteins in
2 506 mitigating intensified oxidative stresses arising from the O₂-rich SO environment.

3
4 507 The evolutionary transition from the negatively buoyant ancestral character to complete
5 508 neutral buoyancy in the Antarctic toothfish entailed remarkable genetic reprogramming of fat
6 509 deposition and bone development. We found upregulation of processes of adipogenesis in
7 510 skeletal muscle, and triacylglycerol synthesis and fat storage were favored over fatty acid
8 511 oxidation. In bone development, a regulatory cascade favoring chondrogenesis over
9 512 osteoblastogenesis was especially evident. The shift in fat synthesis and storage, together with
10 513 reduction of ossification are therefore key in evolutionary gain of neutral buoyancy and
11 514 secondary pelagicism in *D. mawsoni*, and likely in the handful of other pelagic notothenioids,
12 515 allowing them to diversify into mid-water niches, a distinctive hallmark of the Antarctic
13 516 notothenioid adaptive radiation.

14
15
16
17
18
19 517 The remarkable diversification of Antarctic notothenioids (and several other polar fish
20 518 lineages) is integral to the conclusion from a recent analysis of latitudinal diversity gradient of
21 519 marine fishes that high-latitude cold water lineages exhibit exceptionally high rates of
22 520 speciation compared to tropical lineages, counter to expectation based on latitudinal species
23 521 richness (54). Rates of molecular evolution based on phylogenetic tree branch lengths are not
24 522 found to be slower at high latitudes (54). We have shown more definitively in this study that
25 523 in the cold adapted Antarctic notothenioid fish, evolutionary rates in fact accelerated in
26 524 thousands of protein coding genes, extensive cold-specific gene duplication and functional
27 525 diversification had occurred, such as the ZP protein gene families, and TE mobility was
28 526 remarkably elevated which likely contributing to the observed higher frequency of
29 527 chromosomal rearrangements. In mammals, ZP3 is known to function in sperm-egg
30 528 recognition (55) and TE activity is positively related to rate of speciation (56). How these
31 529 genomic and functional changes elicited by selective pressures from the cold SO temperatures
32 530 might have acted as intrinsic factors affecting notothenioid speciation are rich questions for
33 531 further investigation.

34
35
36
37
38
39
40
41
42
43
44
45
46 532 In summary, the results of this study provided robust new insights into genomic and
47 533 transcriptomic alterations enabling cold adaptation and niche expansion of the predominant
48 534 and ecologically vital Antarctic fish group in the SO. The genomes also serve as valuable
49 535 resources for future investigations of genomic and evolutionary changes in the diverse
50 536 Antarctic notothenioid families driven by paleoclimate changes in the SO, studies that may
51 537 shed light on questions of why the coldest ocean has been a hotspot of species formation.

52 538
53 539
54 540
55
56
57
58
59
60
61
62
63
64
65

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

541 MATERIALS AND METHODS

542

543 **Specimens, sampling, and DNA and RNA isolation.**

544 Antarctic toothfish *D. mawsoni* was collected using vertical setline through drilled hole in sea ice
545 of McMurdo Sound, Antarctica (77° 53'S, 166° 34.4'E and vicinity) during austral summer field
546 seasons (Oct.-Dec). Specimens were transported to the aquarium facility in the US National
547 Science Foundation Crary Lab at McMurdo Station and kept in ambient (-1.6°C) flow through
548 seawater tanks, and sacrificed at 2-4 weeks post capture for blood and tissue sampling. The
549 temperate basal notothenioid *E. maclovinus* was collected by rod and reel in the Patagonia waters
550 of southern Chile during austral winter (June) and transported to the NSF R/V Laurence Gould at
551 Punta Arenas in a large, aerated Styrofoam cooler of ambient water (~8°C), where they were
552 sacrificed and sampled within a few days prior to southbound transit for winter field season. To
553 obtain tissues from the large-sized *D. mawsoni*, live specimen was anesthetized with MS222
554 (tricaine methanesulfonate) inside a ambient seawater filled floating sheet plastic tubing in the
555 aquarium tank. The anesthetized specimen was then put on a V-shaped trough for dissection.
556 Tissues were quickly removed and cut into small pieces on ice, and immediately immersed
557 and shaken in ≥ 10 volumes of pre-chilled (-20°C) 90% ethanol (made with 100% pure ethanol
558 and sterilized MilliQ Type 1 water). The ethanol was replaced with a fresh volume within 10
559 minutes, and again at 2-3 hours and 12 hours later. This preservation method serially
560 desiccates the tissue and effectively inactivates tissue nucleases. The tissue samples were kept
561 in -20°C freezer throughout the serial preservation process and then stored at -20°C until use.
562 To obtain tissues from *E. maclovinus*, MS222 anesthetized specimen was quickly dissected on
563 ice, and preserved in -20°C as described for *D. mawsoni*. The ethanol preserved tissues were
564 shipped back to the University of Illinois on dry ice.

565

566 RNA for transcriptome sequencing was isolated from -20°C ethanol preserved tissues using
567 Trizol (Invitrogen) and quality verified by visualization on gel electrophoresis and an Agilent
568 BioAnalyzer. Collection, handling and sampling of the Antarctic toothfish and S. American
569 *E. maclovinus* in this study were carried out in compliance with protocol # 12123 approved
570 by the University of Illinois Institutional Animal Care and Use Committee (IACUC).
571 Additional juvenile specimens of *D. mawsoni* were collected by trawl from the waters of Antarctic
572 Peninsula during the same winter season, and sampled on ship board shortly after capture. The
573 dissected carcasses of *E. maclovinus* and juvenile *D. mawsoni* were kept frozen at -80°C, which
574 provided the pelvic bone samples for immunohistochemical detection for expression of candidate

1 575 genes in bone development. To preserve high molecular weight (HMW) DNA for genome
2 576 sequencing, red blood cells of each species were washed with notothernioid saline (0.1M sodium
3 577 phosphate buffer, pH8.0, adjusted to 420 mOsm with NaCl for *E. maclovinus*, and 540 mOsm
4 578 for *D. mawsoni*), and then embedded in 1% melt agarose to provide about 20ug DNA per 90uL
5 579 block using BioRad plug molds (CHEF Mammalian Genomic DNA Plug Kit #1703591,
6 580 Bio-rad, USA). The embedded cells were lysed *in situ* and the DNA in the agarose blocks was
7 581 preserved following Amemiya *et al* (1996) (57). To recover HMW DNA, the agarose plugs
8 582 were digested with β -agarase (NEB, USA) followed by phenol extraction and dialysis, and
9 583 quality verified using pulsed field electrophoresis.
10 584

18 585 **Sequencing and genome assembly**

19 586 The sequencing libraries with insert sizes of 170, 250, 500 bp were prepared for sequence of
20 587 the paired-end reads, following a modified version of the manufacturer's protocol (Illumina).
21 588 An integrated protocol from the Mate-Pair Library v2 Sample Preparation Guide (Illumina)
22 589 and the Paired-End Library Preparation Method Manual (Roche) was used to prepare
23 590 mate-pair libraries with insert sizes of 3, 6, 10, 15 and 20 kb (**Additional file 1: Tables S1b,**
24 591 **S1c**). For the transcriptome sequencing, Poly(A)+ mRNA was purified using the DynaBeads
25 592 mRNA Purification kit (Life Technologies). Paired-end cDNA libraries were constructed
26 593 using the RNA-Seq NGS Library Preparation Kit for Whole-Transcriptome Discovery
27 594 (Gnomegen). All of the libraries are sequenced on an Illumina HiSeq 1500 sequencer. The *D.*
28 595 *mawsoni* genome was assembled using SOAPdenovo (SOAPdenovo, RRID:SCR_010752)
29 596 (18) to build the contigs and SSPACE (SSPACE, RRID:SCR_005056) (20) to scaffold the
30 597 contigs. The *E. maclovinus* genome was assembled using Platanus¹⁶ (Platanus,
31 598 RRID:SCR_015531) to build the contigs, and SSPACE to scaffold the contigs.
32 599

45 600 **Annotation of the genomes**

46 601 We identified repeats, protein-coding genes and non-coding RNA in the genome assemblies
47 602 of the two species. First, a *de novo* repeat annotation of *D. mawsoni* and *E. maclovinus*
48 603 genomes was carried out by successively using RepeatModeler (RepeatModeler,
49 604 RRID:SCR_015027) (version 1.0.8) and RepeatMasker (RepeatMasker, RRID:SCR_012954)
50 605 (version 4.0.5). *De novo* repeat libraries of the two species were constructed with two
51 606 complementary programs, RECON (RECON, RRID:SCR_006345) (58) and RepeatScout
52 607 (RepeatScout, RRID:SCR_014653) (59) implemented in the RepeatModeler package. The
53 608 generated consensus sequences were manually checked by aligning to the Repbase
54
55
56
57
58
59
60
61
62
63
64
65

1 609 transposable element library (<http://www.girinst.org/replib/>) and genes from the NCBI
2 610 database (nt and nr). The *D. mawsoni* and *E. maclovinus* repeat library consisted of 975 and
3 611 676 consensus sequences with classification information respectively, which were used to run
4 612 RepeatMasker on the assembled scaffolds. Secondly, protein-coding genes were predicted
5 613 using a combination of homology-based and *de novo* approaches. GLEAN was used to create
6 614 consensus gene set by integrating evidence from each prediction. Then RNA-Seq data were
7 615 used to rectify gene models. Generated coding genes were aligned to known protein databases,
8 616 including InterPro (60), KEGG (61) and Uniprot (62), and functional assignment was based
9 617 on that of the best database match. Thirdly, the tRNA genes were predicted with
10 618 tRNAscan-SE (tRNAscan-SE, RRID:SCR_010835) (63). Aligning the rRNA template
11 619 sequences from fishes using BlastN (BlastN, RRID:SCR_001598) with E-value 1e-5
12 620 identified the rRNA fragments. The miRNA and snRNA genes were predicted with
13 621 INFERNAL (INFERNAL, RRID:SCR_011809) (64) software against the Rfam database
14 622 (Release 12) (65).

15 623

16 624 **Phylogenetic reconstruction of 10 vertebrate genomes**

17 625 Protein coding genes of Atlantic cod (*Gadus morhua*), tetraodon (*Tetraodon nigroviridis*),
18 626 Antarctic notothenioid *N. coriiceps*, stickleback (*Gasterosteus aculeatus*), tilapia
19 627 (*Oreochromis niloticus*), medaka (*Oryzias latipes*), zebrafish (*Danio rerio*) and mouse (*Mus*
20 628 *musculus*) genomes were collected from Ensembl release 84 or NCBI, and *D. mawsoni* and *E.*
21 629 *maclovinus* genes from this study, were used to build orthologous clusters with OrthoMCL
22 630 Ver. 2.0.9 (OrthoMCL, RRID:SCR_007839)(22) with default parameters and options. A total
23 631 of 2,936 one-to-one single-copy genes were identified among the ten species. Protein-coding
24 632 sequences of the orthologs were aligned using PRANK (Ver.140603) (66) under a protein
25 633 model with default parameters. The coding sequences of the genes were concatenated to a
26 634 supergene for each species. The supergene sequence dataset was subjected to phylogenetic
27 635 analysis using MrBayes (MrBayes, RRID:SCR_012067) (67), implementing best-fit
28 636 substitution model (GTR+gamma+I) as determined by Modeltest (68). The analysis was run
29 637 800,000 generations, sampling every 100 generations, with the first 2,000 sample set as
30 638 burn-in. Branch-specific *dN* and *dS* were estimated with codeml of the PAML package (69).
31 639 The analysis of changes in gene family size were computed with CAFÉ (CAFE,
32 640 RRID:SCR_005983) (70).

33 641

34 642 **GO annotation and identification of positive selection genes**

1 643 GO terms of the *D. mawsoni*, *E. maclovinus* and stickleback orthologs were built with
2 644 InterProScan (InterProScan, RRID:SCR_005829) (71). The orthologs of each GO terms were
3
4 645 concatenated to estimate branch-specific dN and dS using codeml of PAML (PAML,
5
6 646 RRID:SCR_014932). A binomial test was used to identify the excess of nonsynonymous
7
8 647 changes of GO categories in either *D. mawsoni* or *E. maclovinus* lineages referenced to the
9
10 648 stickleback. Only the GO terms carrying more than 30 orthologs were put into this calculation.
11
12 649 To detect genes evolving under positive selection in *D. mawsoni*, we used the branch-site
13
14 650 model in which likelihood ratio test (LRT) p values were computed. Fisher's exact tests were
15
16 651 used to test for over-represented functional categories among the positive select genes. GO
17
18 652 enrichment analyses of the genes under positive selection were performed using a
19
20 653 hypergeometric method.

21 654

22 655 **Calling of heterozygous SNPs**

23 656 All of the paired-end reads were mapped to the assembled scaffolds with the aligner SMALT
24
25 657 (SMALT, RRID:SCR_005498) to detect the heterozygous sequence polymorphism in the
26
27 658 genomes. The heterozygous SNPs were called with SSAHA_Pileup (version 0.8;
28
29 659 ftp://ftp.sanger.ac.uk/pub/zn1/ssaha_pileup/). Five thresholds were used to post-filter
30
31 660 unreliable SNPs: (1) SSAHA_Pileup SNP score ≥ 20 ; (2) ratio of two alleles between 3:17
32
33 661 to 17:3; (3) the lowest sequencing depth for each allele ≥ 5 ; (4) the minimum distance for
34
35 662 adjacent SNPs ≥ 5 bp; (5) only one polymorphism detected at each SNP position.

36 663

37 664 **Transcriptome analyses**

38
39 665 RNA-seq data derived from liver, gill, stomach, white muscle, red muscle, skin, small
40
41 666 intestine, brain, head kidney, caudal kidney, spleen, and ovary were analyzed for variations in
42
43 667 gene expression of *D. mawsoni* and *E. maclovinus*. RNA-seq reads were trimmed using
44
45 668 Trimmomatic (Trimmomatic, RRID:SCR_011848) (Ver. 0.33) (72) with the parameter set to
46
47 669 AVGQUAL at 20, TRAILING at 20 and MINLEN at 50. The cleaned Illumina paired-end
48
49 670 reads of each tissue were mapped to the annotated scaffolds of *D. mawsoni* and *E. maclovinus*
50
51 671 genome using HISAT2 (HISAT2, RRID:SCR_015530) aligner (Ver. 2.0.4)(73). Cufflinks
52
53 672 (Cufflinks, RRID:SCR_014597) (Ver. 2.2.1) (74) normalized gene expressions to the
54
55 673 quantified transcription levels (FPKM). Differentially expressions of the genes were assessed
56
57 674 using DEGseq (DEGseq, RRID:SCR_008480) (Ver. 1.28.0) (75) with cutoff at $q < 0.001$ (76).
58
59 675 GO and KEGG enrichment analyses for the identified differentially expressed genes were
60
61 676 performed using clusterProfiler packages (77) with the cutoff at $p < 0.05$.

677

678 **Construction of gene collinearity among *D. mawsoni*, *E. maclovinus* and stickleback**
679 **genomes.**

680 The genes of *D. mawsoni* and *E. maclovinus* were aligned to the gene model set of
681 stickleback by Blastp (BLASTP, RRID:SCR_001010) with E-value at 1e-20. Two criteria
682 were used to call syntenic gene blocks in the *D. mawsoni* or *E. maclovinus* scaffolds: (1)
683 Number of the gene on the syntenic block ≥ 3 ; (2) number of non-syntenic genes between
684 two adjacent syntenic genes ≤ 10 . Each syntenic block was anchored on the stickleback
685 genomes according to the orders of the reference gene.

686

687 **Western-blot analysis of ZP proteins**

688 The proteins were separated on 10% SDS-PAGE at 100V for 90 min in 193 mM glycine and
689 25 mM Tris (pH 8.8). The resolved proteins were electrophoretically transferred to a
690 nitrocellulose membrane (Millipore) using a Mini-Protean Tetra Cell (BioRad) in a buffer
691 containing 193mM glycine, 25mM Tris (pH 8.3) and 20% methanol. The membrane was
692 treated with blocking agent (5% nonfat milk in 1x TBST) for 2 h at room temperature on a
693 shaker. FLAG antibody or β -actin antibody (Hua An Biotechnology Co. Ltd, Hangzhou,
694 China) was added, and the membranes were incubated at room temperature for 1 h. The
695 membrane was then washed with 1x TBST three times for 15 min each. The secondary
696 antibody (1:2,000 in 1x TBST, Boston Biomedical Inc.) was then added and incubated for 1 h
697 at room temperature. The membrane was washed with 1x TBST twice and 1x TBS once for
698 15 min each. Color was developed using SuperSignal West Pico Chemiluminescent Substrate
699 (Thermo Scientific) according to the manufacturer's instructions. Images were acquired using
700 a ChemiDoc MP Imaging System (BioRad).

701

702 **Assay of CHO cell survival rate at freezing temperature.**

703 DmZPC5 exons were serially deleted using PCR amplification of the DmZPC5 expression
704 vector with primers designed to eliminate desired coding sequences (Fig.3b). The full-length
705 sequences of three DmZPC5 isoforms (DmZPC5-1, DmZPC5-2 and DmZPC5-3) were
706 engineered to contain a FLAG octapeptide and cloned into the expression vector
707 pIRES2-EGFP (**Additional file 1: Fig. S5**). The three constructed vectors and blank control
708 (vector pIRES2-EGFP) were transferred into the CHO cells (American Type Culture
709 Collection CCL-61) obtained from American Type Culture Collection. The CHO cells were
710 cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum (Gibco).

1 711 CHO cells were incubated at 37°C for 2 days and then kept at -2 °C for 8 hours. The treated
2 712 CHO cells were collected and washed by DPBS (Dulbecco's Phosphate-Buffered Saline)
3
4 713 twice. The cells were stained with 10 µg/mL propidium iodide at room temperature for 5
5 714 minutes, and numbers of PI-stained cells (dead cells) were determined by flow cytometry.
6
7 715 The survival rate is calculated with the equation: survival rate = S / (S + D), where S is the
8
9 716 number of surviving cells and D is the number the dead cells.

10 717

11 718 **Identification of the LINEs and estimation of their divergence time.**

12
13 719 The seed sequences of the LINEs (28) were aligned against the *D. mawsoni* and *E.*
14 720 *maclovinus* draft genomes, respectively, using BlastN at E-value of 1e-10. According to loci
15
16 721 of the alignments, the sequences were extracted from the genomes, which were considered as
17
18 722 the candidates of the LINEs. If distance of two adjacent candidates was 200 bp or less, these
19
20 723 two candidates were connected by the sequence between them. All of the candidates and their
21
22 724 corresponding seed sequences were mutually aligned by BlastN. Those candidates with over
23
24 725 60% identity and over 100 bp of alignments to the seed sequences were collected as the
25
26 726 LINEs of *D. mawsoni* and *E. maclovinus*, respectively.

27
28
29 727 Alignment of the LINEs was conducted by a multiple sequence aligner ClustalW (ClustalW,
30
31 728 RRID:SCR_002909) (78) to cluster any two LINEs with highest sequence similarity into a
32
33 729 LINE pair in *D. mawsoni* or *E. maclovinus*. The evolutionary distance of two LINES for each
34
35 730 LINE pair was calculated by the Kimura two-parameter method (EMBOSS distmat, version
36
37 731 6.6.0.0), which reflected the substitution rate per site between the two LINEs. According to
38
39 732 the calculated synonymous substitution rates for 7,958 *D. mawsoni*-*E. maclovinus*
40
41 733 orthologous pairs, the mean synonymous substitution rate is around 0.227. The peak of
42
43 734 substitution rate is at 0.04, which estimated the LINE burst to be about 6.5 million years ago
44
45 735 when the species divergence time between *D. mawsoni* and *E. maclovinus* is around 37
46
47 736 million years ago (79).

48 737

49 738 **Tissue fixation and immunohistochemistry**

50 739 Pieces of pelvic bone (with muscle) and the attached fins (no larger than 20 × 20 × 5 mm)
51
52 740 were dissected from frozen specimens of young *D. mawsoni* and *E. maclovinus* and immersed
53
54 741 in a fixation solution, KINFix which contains (62.5% (v/v) ethanol, 6.71% (v/v) acetic acid,
55
56 742 and 6% (w/v) trehalose (80) for over 24hrs. Tissue are decalcified in EDTA solution (cat.
57
58 743 no.041-22031, WAKO) for about 2 weeks. Then the specimen was dehydrated in graded
59
60
61
62
63
64
65

1 744 ethanol (70%, 80%, 90%,95%, 1×1h each), 100% ethanol 2 ×1h at room temperature, xylene
2 745 for 2×1h, and embedded in low-melting paraffin for 2 × 1h, and kept overnight at 56°C, then
3
4 746 embedded in paraffin. For each tissue, 5 µm thick serial sections were cut with a microtome
5
6 747 (RM2245, Leica). Immunostaining was performed using the EnVision detection system (cat.
7
8 748 no.K5007, Dako). Slides were deparafinized in xylene and rehydrated in a descending series
9
10 749 of ethanol (100%, 95%, 90% and 70%), and washed in phosphate buffered saline (PBS).
11
12 750 Endogenous peroxidases were blocked with 3% H₂O₂ for 10 min, after which the sections
13
14 751 were incubated with 5% BSA for 35min. Then, the slides are incubated overnight with the
15
16 752 primary CTGF antibody (1:400 dilution) (cat. no.ab6992, Abcam) at 4°C. Next, the sections
17
18 753 were washed four times with PBS for 15 min followed by incubation with a goat anti-rabbit
19
20 754 secondary antibody for 35min at 37°C. After four washes with PBS, 3, 3'-diaminobenzidine
21
22 755 (DAB) was added to visualize the immunoreactivity. All slides were counterstained with
23
24 756 haematoxylin. The sections were dehydrated in a mounting series of alcohol (70%, 90%, 100%
25
26 757 and 100%) and in xylene. Finally, slides are mounted using neutral balsam mounting medium,
27
28 758 and analyzed under a bright field microscope (AXIO imager. M2, ZEISS).
29
30
31 759

32 760 **Availability of data and material**

33 761 All of the Illumina short read sequencing data of this project have been deposited at NCBI
34
35 762 under the accession no. BioProject PRJNA401363 (<http://www.ncbi.nlm.nih.gov/sra/>). The
36
37 763 assembled draft genomes and their annotations have been released at the official website of
38
39 764 the Shanghai Ocean University (<http://202.121.66.128/>). The current version of the data set is
40
41 765 the first version (v1). Antarctic toothfish genome and transcriptome (86), Patagonian robalo
42
43 766 genome and transcriptome (87) and other supporting data are also available via the
44
45 767 *GigaScience* GigaDB repository(88).
46
47 768

48 769 **Abbreviations**

49 770 ACC: Antarctic Circumpolar Current; AFGP: antifreeze glycoprotein; AGPAT:
50
51 771 acylglycerol-3-phosphate O-acyltransferase; BUSCO: Benchmarking Universal Single-Copy
52
53 772 Orthologs; CDS: CDP-diacylglycerol synthase; CHO: Chinese Hamster Ovary; CTGF:
54
55 773 Connective Tissue Growth Factor; DEG: differentially expressed genes; DPBS: Dulbecco's
56
57 774 Phosphate-Buffered Saline; ECM1: extracellular matrix protein 1; GO: Gene Ontology;
58
59 775 HMW: high molecular weight; HSPG2: prostaglandin-endoperoxide synthase 2; IACUC:

1 776 Institutional Animal Care and Use Committee; JDP2: Jun dimerization protein 2; LINE: long
2 777 interspersed transposable elements; LRT: likelihood ratio test; PBS: phosphate buffered saline;
3
4 778 PPAR:peroxisome proliferator-activated receptors; PPAR γ : peroxisome proliferator activated
5 779 receptor gamma; PSG: positively selected genes; SINE: short interspersed transposable
6
7 780 elements; SO: Southern Ocean; SRF1: secreted frizzled-related protein 1; SSR: simple
8
9 781 sequence repeats; TE: Transposable elements; ZP: Zona Pellucida protein.
10

11 782

12 783 **Additional files**

13 784 Additional file 1 : Figs. S1 to S11 and Tables S1 to S17.

14 785 Additional file 2 : supporting data for Fig. 1b, Fig.2c, Fig. 2d, and Fig. 4.

15 786 Additional file 3 : list of duplicated protein gene families of *D. mawsoni* and *E. maclovinus*.

16 787 Additional file 4: list of DEGs between *D.mawsoni* and two negatively buoyant
17 788 notothenioids.

18 789

19 790 **Acknowledgements**

20 791 We thank BGI-Shenzhen for their technical support in sequencing of *D. mawsoni* genome, Mr.
21 792 Liping Shu of Wuhan Ice-harbor Biological Technological Co. Ltd. for his help in the
22 793 annotation of the *E. maclovinus* genome.
23

24 794

25 795 **Funds**

26 796 The work was supported by grants from the Natural Science Foundation of China (No.
27 797 41761134050, 31572611, 31572598) and the Major Science Innovation Grant
28 798 (2017-01-07-00-10-E00060) from the Shanghai Education Committee, the Key Achievement
29 799 Supporting Grant from Laboratory for Marine Biology and Biotechnology, Qingdao National
30 800 Laboratory for Marine Science and Technology to LC, and the USA NSF Polar Programs
31 801 grant ANT1142158 to CHCC.
32

33 802

34 803 **Authors' contribution**

35 804 LC and CHCC conceived and managed the project and its components. CHCC and MH
36 805 performed fish and tissue collections and sample preparations. KRM and XZ contributed to
37 806 DNA and RNA preparations and genome size determination. YL, MY, and WL performed
38 807 genome annotation and RNA-seq data analysis. YR and SP conducted *de novo* genome
39 808 assembly. KTB confirmed the AFGP loci. QX, YF and LC designed and performed the
40 809 biological experiments. Sample preparation and genome sequencing were carried out by SJ,
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

1 810 WZ and JW. LC, QX, YL and CHCC analyzed the data as a whole and wrote the manuscript,
2 811 and CHCC and WW contributed interpretation of data and edits to the manuscript.
3
4 812

5 813 **Competing interests**

6
7 814 The authors declare they have no competing financial interests.
8
9 815

10
11 816 **URLs** KEGG, <http://www.genome.jp/kegg/>; KAAS,
12 <http://www.genome.jp/tools/kaas/>;SSPACE,
13 <https://www.baseclear.com/genomics/bioinformatics/basetools/SSPACE>; Platanus,
14 <http://platanus.bio.titech.ac.jp/>; SMALT, <http://www.sanger.ac.uk/resources/software/smalt/>;
15 819 SOAPdenovo, <http://soap.genomics.org.cn>; RepeatModeler,
16 820 <http://www.repeatmasker.org/RepeatModeler.html>; RepeatMasker,
17 821 <http://www.repeatmasker.org>; Repbase, <http://www.girinst.org/repbase/>; Timetree,
18 822 <http://www.timetree.org/>;Ensembl, <ftp://ftp.ensembl.org/pub/>;
19 823 <http://evolution.genetics.washington.edu/phylip.html>; PAML,
20 824 <http://abacus.gene.ucl.ac.uk/software/paml.html>; GLEAN, <https://github.com/glean/glean>;
21 825 Interpro, <http://www.ebi.ac.uk/interpro/>; Infernal, <http://eddylab.org/infernal/>; Rfam,
22 826 <http://rfam.xfam.org/>; OrthoMCL, <http://orthomcl.org/orthomcl/>; Mrbayes,
23 827 <http://mrbayes.sourceforge.net/>; SSAHA_Pileup, ftp://ftp.sanger.ac.uk/pub/zn1/ssaha_pileup/;
24 828 HISAT2, <http://ccb.jhu.edu/software/hisat2/index.shtml>.
25 829
26 830
27 831
28 832
29 833

30 834 **Legends**

31 835
32 836 **Fig. 1** Sequenced species and genome synteny. (a) Sampling location and geographic distribution.
33 837 The red and blue filled circles are the geographic distributions of *D. mawsoni* and *E. maclovinus*,
34 838 respectively (www.fishbase.org, version 02/2018 (81) and Hanchet et al.(82). The red and blue
35 839 stars show the respective locations where the sequenced individuals were collected. The Antarctic
36 840 Polar Front, an approximation of the mean position of the Antarctic Circumpolar Current, is
37 841 adopted from Barker and Thomas(83). The image of *D. mawsoni* is a courtesy from Elliot
38 842 DeVries and that of *E. maclovinus* is from Dirk Schories. (b) Gene collinearity among *D.*
39 843 *mawsoni*, *E. maclovinus* and *G. aculeatus* (Stickleback). The scaffolds of *D. mawsoni* (the
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

844 circularized red blocks labelled with “D”) and *E. maclovinus* (the circularized green blocks
845 labelled with “E”) are anchored on the twenty-one Stickleback chromosomes (the circularized
846 light blue blocks labelled with “G”, 1 to 21), according to the gene collinearity (the
847 connecting yellow lines). The black vertical lines within the *D. mawsoni* and *E. maclovinus*
848 scaffolds indicate occurrence of LINEs greater than 500 bps in these positions. The sequence
849 length is indicated by the 5-Mb tick marks on the reference Stickleback chromosomes. The
850 outermost circle of red vertical lines and the innermost circle of green vertical lines indicated
851 the quantified expression levels (FPKM) of the genes located on the corresponding *D.*
852 *mawsoni* and *E. maclovinus* scaffolds, respectively. The expression profiles are derived from
853 the transcriptome data of white muscles (see the transcriptome section). The small white
854 squares and rectangles scattered in the scaffolds show the locations of the Selcys-tRNA genes
855 of *D. mawsoni* and *E. maclovinus*. The single yellow square shows the location of AFGP
856 genes in the *D. mawsoni* genome.

Fig. 2 Evolution of the genomes and genes. **(a)** Timing and frequency of LINE insertion in *D.*
mawsoni, *E. maclovinus* and *N. coriiceps* showing correlation between onset of late Miocene
deep cooling and burst LINE insertions in the Antarctic toothfish and bullhead notothen. The
black trace indicates global temperature trends during Oligocene, Miocene, Pliocene (Pli) and
Pleistocene (Ple) from 30 to 0 million years ago (mya), modified from Zachos *et al.*
(2008)(84), Near *et al.* (2012) (9) and Favre *et al.* (2015) (85). The red and blue line indicate
the insertion frequency of LINEs (the percentage of the calculated LINE pairs) in the *D.*
mawsoni and *E. maclovinus* genomes respectively during these periods. **(b)** Reconstructed
phylogeny of nine teleost fish lineages using 2,936 orthologous genes (mouse serving as
outgroup) and the calculated *dN/dS* ratio for each branch, showing a 2-fold faster evolutionary
rate in the Antarctic notothenioids. **(c)** Comparison of adaptive evolution between *D. mawsoni*
and *E. maclovinus* genomes. Data points represent average *dN/dS* value of each GO term,
each of which consists of at least 30 genes. The red and blue circles show the GO terms with
significantly higher *dN/dS* ratios ($p < 0.05$, binomial test) in *D. mawsoni* and *E. maclovinus*,
respectively. The grey circles are those showing no significant difference. GO terms falling on
the dashed line of linearity have the same *dN/dS* ratios in the two species. **(d)** Gene
duplication in *D. mawsoni*. A subset (26) of the 202 gene families detected to contain higher
gene copy numbers in the *D. mawsoni* genome relative to other species are listed on the left,
with their respective KEGG pathway listed on the right. The gene copy numbers are measured
by color difference. The pathways highlighted in red are especially abundant in *D. mawsoni*

1 878 and might be relevant to physiological adaptation of *D. mawsoni* in the freezing environment.
2 879 (e) A subset of duplicated gene families in *E. maclovinus*, showing different KEGG pathways
3
4 880 between *D. mawsoni* and *E. maclovinus* in terms of gene duplication. The red highlighted
5
6 881 pathway (ether lipid metabolism) indicates a common duplication occurred in the three
7
8 882 Notothenioids.

9 883
10 884

11 **Fig. 3** Evolutionary and functional analyses of the DmZPC5 genes involved in cellular
12 freezing resistance. (a) Duplication of ZPC5 gene (DmZPC5) in *D. mawsoni*. Phylogenetic
13 neighbor-joining tree of ZPC5 genes among *D. mawsoni*, *E. maclovinus*, *Larimichthys crocea*
14 (Lc), *T. rubripes* (Tr) and *O. latipes* (Ol). The gene structures are illustrated on the right. The
15 different colored blocks indicate the exons encoding signal peptides (red), zp domains (blue)
16 and the remaining exons (incarnadine). The jagged blocks contain the nonsense mutations in
17 DmZPC5-2a/b and DmZPC5-3 genes that cause premature termination of coding sequences.
18 (b) Western-blot analysis of the DmZPC5 isoforms indicated their sizes and temperature
19 sensitive accumulation. Purified proteins encoded by the three DmZPC5 isoforms were
20 detected by an anti-FLAG antibody on the SDS-PAGE gels. All of these three DmZPC5
21 protein had higher expression levels at 0°C than at 37 °C. (c) Assays of cell survival rate
22 under recombinant expression of different DmZPC5 isoforms in CHO cells at a freezing
23 temperature (-2°C for 8 hrs). The bars represent the mean \pm s.d ($n = 3$, biological replicates).
24 The sample pIRES2-EGFP is the expression vector as control. Significances of different
25 survival rate are indicated by * (unpaired Student's t-test, $p < 0.05$) and ** ($p < 0.01$).

26 890
27 891
28 892
29 893
30 894
31 895
32 896
33 897
34 898
35 899
36 900
37 901
38 902

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

43 903
44 904
45 905
46 906
47 907
48 908

49 **Fig. 5** Schematic diagram showing changed regulation of buoyancy related developmental
50 pathways. (a). Enhanced adipogenetic pathways in *D. mawsoni* muscle. The genes shadowed
51 with the dark red color were upregulated in *D. mawsoni* while those shadowed in light grey
52 were unchanged. (b). Changed osteogenetic regulation in *D. mawsoni* bone. Genes shadowed
53 with the dark grey color were upregulated in *D. mawsoni* while those in light grey were not
54
55
56
57
58
59
60
61
62
63
64
65

1 914 changed. The arrows (in dark red or dark grey) indicate a positive effect on the process while
2 915 blocked (in blue) lines indicate inhibitory effect. MSC: Mesenchymal Stem Cell. (c).
3
4 916 Immunohistochemical (IHC) staining to detect the abundance of Connective Tissue Growth
5 917 Factor (CTGF) in cross sections of pelvic fin of *D. mawsoni* and *E. maclovinus*. The left
6 918 panels of each fish are IHC staining without the first antibody as negative control. The
7 919 presence of CTGF is indicated by the brown signals in the right tissues. Scale bar, 50µm.
8
9
10
11 920
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

921 **References**

- 1
2 922 1. Livermore R, Nankivell A, Eagles G, & Morris P (2005) Paleogene opening of Drake Passage.
3 923 *Earth Planet Sc Lett* 236(1-2):459-470.
4 924 2. Eastman JT (2005) The nature of the diversity of Antarctic fishes. *Polar Biol* 28(2):93-107.
5
6 925 3. Near TJ, *et al.* (2015) Identification of the notothenioid sister lineage illuminates the
7 926 biogeographic history of an Antarctic adaptive radiation. *Bmc Evol Biol* 15:109.
8
9 927 4. Eastman JT (2000) Antarctic notothenioid fishes as subjects for research in evolutionary
10 928 biology. *Antarct Sci* 12(3):276-287.
11 929 5. La Mesa M, Eastman JT, & Vacchi M (2004) The role of notothenioid fish in the food web of
12 930 the Ross Sea shelf waters: a review. *Polar Biol* 27(6):321-338.
13 931 6. Devries AL & Eastman JT (1978) Lipid sacs as a buoyancy adaptation in an Antarctic fish.
14 932 *Nature* 271(5643):352-353.
15 933 7. Chen LB, DeVries AL, & Cheng CHC (1997) Evolution of antifreeze glycoprotein gene from a
16 934 trypsinogen gene in Antarctic notothenioid fish. *P Natl Acad Sci USA* 94(8):3811-3816.
17 935 8. Matschiner M, Hanel R, & Salzburger W (2011) On the Origin and Trigger of the Notothenioid
18 936 Adaptive Radiation. *Plos One* 6(4).
19 937 9. Near TJ, *et al.* (2012) Ancient climate change, antifreeze, and the evolutionary diversification
20 938 of Antarctic fishes. *P Natl Acad Sci USA* 109(9):3434-3439.
21 939 10. Shin SC, *et al.* (2014) The genome sequence of the Antarctic bullhead notothen reveals
22 940 evolutionary adaptations to a cold environment. *Genome Biol* 15(9).
23 941 11. Malmstrom M, *et al.* (2016) Evolution of the immune system influences speciation rates in
24 942 teleost fishes. *Nat Genet* 48(10):1204-1210.
25 943 12. Gon O & Heemstra P (1990) *Fishes of the Southern Ocean* (J.L.B. Smith Institute of
26 944 Ichthyology) p 305.
27 945 13. Eastman JT, Witmer LM, Ridgely RC, & Kuhn KL (2014) Divergence in Skeletal Mass and Bone
28 946 Morphology in Antarctic Notothenioid Fishes. *J Morphol* 275(8):841-861.
29 947 14. Eastman JT & Devries AL (2010) Buoyancy adaptations in a swim - bladderless Antarctic fish.
30 948 *J Morphol* 167(1):91-102.
31 949 15. Eastman JT & Devries AL (1982) Buoyancy Studies of Notothenioid Fishes in McMurdo Sound,
32 950 Antarctica. *Copeia* 1982(2):385-393.
33 951 16. Ghigliotti L, *et al.* (2007) The two giant sister species of the Southern Ocean, *Dissostichus*
34 952 *eleginoides* and *Dissostichus mawsoni*, differ in karyotype and chromosomal pattern of
35 953 ribosomal RNA genes. *Polar Biol* 30(5):625-634.
36 954 17. Mazzei F, *et al.* (2008) Chromosomal characteristics of the temperate notothenioid fish
37 955 *Eleginops maclovinus* (Cuvier). *Polar Biol* 31(5):629-634.
38 956 18. Li RQ, Li YR, Kristiansen K, & Wang J (2008) SOAP: short oligonucleotide alignment program.
39 957 *Bioinformatics* 24(5):713-714.
40 958 19. Kajitani R, *et al.* (2014) Efficient de novo assembly of highly heterozygous genomes from
41 959 whole-genome shotgun short reads. *Genome Res* 24(8):1384-1395.
42 960 20. Boetzer M, Henkel CV, Jansen HJ, Butler D, & Pirovano W (2011) Scaffolding pre-assembled
43 961 contigs using SSPACE. *Bioinformatics* 27(4):578-579.
44 962 21. SimãEO FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015) BUSCO:
45 963 assessing genome assembly and annotation completeness with single-copy orthologs.
46 964 *Bioinformatics* 31(19):3210-3212.

- 1 965 22. Li L, Stoeckert CJ, & Roos DS (2003) OrthoMCL: Identification of ortholog groups for
2 966 eukaryotic genomes. *Genome Res* 13(9):2178-2189.
- 3 967 23. Pisano E & Ozouf-Costaz C (2003) Cytogenetics and evolution in extreme environment: the
4 968 case of Antarctic fishes. *Science Publishers Inc, Enfield (NH) USA*:309-330.
- 5 969 24. Detrich HW, *et al.* (2010) Genome Enablement of the Notothenioidei: Genome Size Estimates
6 970 from 11 Species and BAC Libraries from 2 Representative Taxa. *J Exp Zool Part B*
7 971 314b(5):369-381.
- 8 972 25. Rebollo R, Horard B, Hubert B, & Vieira C (2010) Jumping genes and epigenetics: Towards new
9 973 species. *Gene* 454(1-2):1-7.
- 10 974 26. Auvinet J, *et al.* (2018) Mobilization of retrotransposons as a cause of chromosomal
11 975 diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus*.
12 976 *Bmc Genomics* 19.
- 13 977 27. Chenais B, Caruso A, Hiard S, & Casse N (2012) The impact of transposable elements on
14 978 eukaryotic genomes: From genome size increase to genetic adaptation to stressful
15 979 environments. *Gene* 509(1):7-15.
- 16 980 28. Chen S, *et al.* (2017) Cold-induced retrotransposition of fish LINEs. *Journal of genetics and*
17 981 *genomics = Yi chuan xue bao* 44(8):385-394.
- 18 982 29. Chen ZZ, *et al.* (2008) Transcritomic and genomic evolution under constant cold in Antarctic
19 983 notothenioid fish. *P Natl Acad Sci USA* 105(35):12944-12949.
- 20 984 30. Xu QH, *et al.* (2008) Adaptive evolution of hepcidin genes in antarctic notothenioid fishes.
21 985 *Mol Biol Evol* 25(6):1099-1112.
- 22 986 31. Cao LX, *et al.* (2016) Neofunctionalization of zona pellucida proteins enhances
23 987 freeze-prevention in the eggs of Antarctic notothenioids. *Nat Commun* 7.
- 24 988 32. Nicodemus-Johnson J, Silic S, Ghigliotti L, Pisano E, & Cheng CHC (2011) Assembly of the
25 989 antifreeze glycoprotein/trypsinogen-like protease genomic locus in the Antarctic toothfish
26 990 *Dissostichus mawsoni* (Norman). *Genomics* 98(3):194-201.
- 27 991 33. Brigelius-Flohe R & Maiorino M (2013) Glutathione peroxidases. *Bba-Gen Subjects*
28 992 1830(5):3289-3303.
- 29 993 34. Angeli JPF, *et al.* (2014) Inactivation of the ferroptosis regulator Gpx4 triggers acute renal
30 994 failure in mice. *Nat Cell Biol* 16(12):1180-U1120.
- 31 995 35. Chen Y, Rui BB, Tang LY, & Hu CM (2015) Lipin Family Proteins - Key Regulators in Lipid
32 996 Metabolism. *Ann Nutr Metab* 66(1):10-18.
- 33 997 36. Farmer SR (2006) Transcriptional control of adipocyte formation. *Cell Metab* 4(4):263-273.
- 34 998 37. Ignatz RA & Massagué J (1985) Type beta transforming growth factor controls the adipogenic
35 999 differentiation of 3T3 fibroblasts. *P Natl Acad Sci USA* 82(24):8530-8534.
- 36 1000 38. Petruschke T, Röhrig K, & Hauner H (1994) Transforming growth factor beta (TGF-beta)
37 1001 inhibits the differentiation of human adipocyte precursor cells in primary culture. *Int J Obes*
38 1002 *Relat Metab Disord* 18(8):532-536.
- 39 1003 39. Long FX & Ornitz DM (2013) Development of the Endochondral Skeleton. *Csh Perspect Biol*
40 1004 5(1).
- 41 1005 40. Luo Q, *et al.* (2004) Connective tissue growth factor (CTGF) is regulated by Wnt and bone
42 1006 morphogenetic proteins signaling in osteoblast differentiation of mesenchymal stem cells. *J*
43 1007 *Biol Chem* 279(53):55958-55968.
- 44 1008 41. Sadatsuki R, *et al.* (2017) Perlecan is required for the chondrogenic differentiation of synovial

1009 mesenchymal cells through regulation of Sox9 gene expression. *J Orthop Res* 35(4):837-846.

1 1010 42. Mongiat M, *et al.* (2003) Perlecan protein core interacts with extracellular matrix protein 1
2 1011 (ECM1), a glycoprotein involved in bone formation and angiogenesis. *J Biol Chem*
3 1012 278(19):17491-17499.

4 1013 43. Arnold MA, *et al.* (2007) MEF2C transcription factor controls chondrocyte hypertrophy and
5 1014 bone development. *Dev Cell* 12(3):377-389.

6 1015 44. Gamez B, Rodriguez-Carballo E, Bartrons R, Rosa JL, & Ventura F (2013) MicroRNA-322
7 1016 (miR-322) and Its Target Protein Tob2 Modulate Osterix (Osx) mRNA Stability. *J Biol Chem*
8 1017 288(20):14264-14275.

9 1018 45. Tago K, *et al.* (2000) Inhibition of Wnt signaling by ICAT, a novel beta-catenin-interacting
10 1019 protein. *Gene Dev* 14(14):1741-1749.

11 1020 46. Yao W, *et al.* (2010) Overexpression of Secreted Frizzled-Related Protein 1 Inhibits Bone
12 1021 Formation and Attenuates Parathyroid Hormone Bone Anabolic Effects. *J Bone Miner Res*
13 1022 25(2):190-199.

14 1023 47. Onizuka S, *et al.* (2016) ZBTB16 as a Downstream Target Gene of Osterix Regulates
15 1024 Osteoblastogenesis of Human Multipotent Mesenchymal Stromal Cells. *J Cell Biochem*
16 1025 117(10):2423-2434.

17 1026 48. Dexheimer V, *et al.* (2016) Differential expression of TGF-beta superfamily members and role
18 1027 of Smad1/5/9-signalling in chondral versus endochondral chondrocyte differentiation. *Sci*
19 1028 *Rep-Uk* 6.

20 1029 49. Tare RS, Oreffo ROC, Clarke NMP, & Roach HI (2002) Pleiotrophin/osteoblast-stimulating
21 1030 factor 1: Dissecting its diverse functions in bone formation. *J Bone Miner Res*
22 1031 17(11):2009-2020.

23 1032 50. Su JL, *et al.* (2010) CYR61 Regulates BMP-2-dependent Osteoblast Differentiation through the
24 1033 alpha(v)beta(3) Integrin/Integrin-linked Kinase/ERK Pathway. *J Biol Chem*
25 1034 285(41):31325-31336.

26 1035 51. Maruyama K, *et al.* (2013) Strawberry notch homologue 2 regulates osteoclast fusion by
27 1036 enhancing the expression of DC-STAMP. *J Exp Med* 210(10):1947-1960.

28 1037 52. Diercke K, Sen S, Kohl A, Lux CJ, & Erber R (2011) Compression-dependent Up-regulation of
29 1038 Ephrin-A2 in PDL Fibroblasts Attenuates Osteogenesis. *J Dent Res* 90(9):1108-1115.

30 1039 53. Abdallah BM & Kassem M (2012) New factors controlling the balance between
31 1040 osteoblastogenesis and adipogenesis. *Bone* 50(2):540-545.

32 1041 54. Rabosky DL, *et al.* (2018) An inverse latitudinal gradient in speciation rate for marine fishes.
33 1042 *Nature* 559(7714):392-395.

34 1043 55. Litscher ES, Williams Z, & Wassarman PM (2009) Zona Pellucida Glycoprotein ZP3 and
35 1044 Fertilization in Mammals. *Mol Reprod Dev* 76(10):933-941.

36 1045 56. Ricci M, Peona V, Guichard E, Taccioli C, & Boattini A (2018) Transposable Elements Activity is
37 1046 Positively Related to Rate of Speciation in Mammals. *Journal of Molecular Evolution*
38 1047 86(5):303-310.

39 1048 57. Amemiya CT, Ota T, & Litman GW (1996) Construction of P1 Artificial Chromosome (PAC)
40 1049 Libraries from Lower Vertebrates. *Nonmammalian Genomic Analysis*, (San Diego: Academic
41 1050 Press), pp 223-256.

42 1051 58. Bao ZR & Eddy SR (2002) Automated de novo identification of repeat sequence families in
43 1052 sequenced genomes. *Genome Res* 12(8):1269-1276.

1 1053 59. Price AL, Jones NC, & Pevzner PA (2005) De novo identification of repeat families in large
2 1054 genomes. *Bioinformatics* 21:1351-1358.

3 1055 60. Apweiler R, *et al.* (2001) The InterPro database, an integrated documentation resource for
4 1056 protein families, domains and functional sites. *Nucleic Acids Res* 29(1):37-40.

5 1057 61. Kanehisa M & Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids*
6 1058 *Res* 28(1):27-30.

7 1059 62. Bairoch A, *et al.* (2005) The universal protein resource (UniProt). *Nucleic Acids Res*
8 1060 33:D154-D159.

9 1061 63. Lowe TM & Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA
10 1062 genes in genomic sequence. *Nucleic Acids Res* 25(5):955-964.

11 1063 64. Nawrocki EP, Kolbe DL, & Eddy SR (2009) Infernal 1.0: inference of RNA alignments.
12 1064 *Bioinformatics* 25(10):1335-1337.

13 1065 65. Nawrocki EP, *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res*
14 1066 43(D1):D130-D137.

15 1067 66. Loytynoja A & Goldman N (2010) webPRANK: a phylogeny-aware multiple sequence aligner
16 1068 with interactive alignment browser. *Bmc Bioinformatics* 11.

17 1069 67. Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees.
18 1070 *Bioinformatics* 17(8):754-755.

19 1071 68. Posada D & Crandall KA (1998) MODELTEST: testing the model of DNA substitution.
20 1072 *Bioinformatics* 14(9):817-818.

21 1073 69. Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*
22 1074 24(8):1586-1591.

23 1075 70. De Bie T, Cristianini N, Demuth JP, & Hahn MW (2006) CAFE: a computational tool for the
24 1076 study of gene family evolution. *Bioinformatics* 22(10):1269-1271.

25 1077 71. Zdobnov EM & Apweiler R (2001) InterProScan - an integration platform for the
26 1078 signature-recognition methods in InterPro. *Bioinformatics* 17(9):847-848.

27 1079 72. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina
28 1080 sequence data. *Bioinformatics* 30(15):2114-2120.

29 1081 73. Kim D, Landmead B, & Salzberg SL (2015) HISAT: a fast spliced aligner with low memory
30 1082 requirements. *Nat Methods* 12(4):357-U121.

31 1083 74. Trapnell C, *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq
32 1084 experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562-578.

33 1085 75. Wang LK, Feng ZX, Wang X, Wang XW, & Zhang XG (2010) DEGseq: an R package for
34 1086 identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26(1):136-138.

35 1087 76. Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and
36 1088 Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57(1):289-300.

37 1089 77. Yu GC, Wang LG, Han YY, & He QY (2012) clusterProfiler: an R Package for Comparing
38 1090 Biological Themes Among Gene Clusters. *Omic* 16(5):284-287.

39 1091 78. Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of
40 1092 progressive multiple sequence alignment through sequence weighting, position-specific gap
41 1093 penalties and weight matrix choice. *Nucleic Acids Res*:4673-4680.

42 1094 79. Hedges SB, Dudley J, & Kumar S (2006) TimeTree: a public knowledge-base of divergence
43 1095 times among organisms. *Bioinformatics* 22(23):2971-2972.

44 1096 80. Stefanits H, *et al.* (2016) KINFix - A formalin-free non-commercial fixative optimized for

1097 histological, immunohistochemical and molecular analyses of neurosurgical tissue specimens.
1098 *Clin Neuropathol* 35(1):3-12.

1099 81. Pauly FRaD (2018) FishBase. in *World Wide Web electronic publication*.

1100 82. Hanchet S, *et al.* (2015) The Antarctic toothfish (*Dissostichus mawsoni*): biology, ecology, and
1101 life history in the Ross Sea region. *Hydrobiologia* 761(1):397-414.

1102 83. Barker PF & Thomas E (2004) Origin, signature and palaeoclimatic influence of the Antarctic
1103 Circumpolar Current. *Earth-Sci Rev* 66(1-2):143-162.

1104 84. Zachos JC, Dickens GR, & Zeebe RE (2008) An early Cenozoic perspective on greenhouse
1105 warming and carbon-cycle dynamics. *Nature* 451(7176):279-283.

1106 85. Favre A, *et al.* (2015) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of
1107 Tibetan biotas. *Biol Rev* 90(1):236-253.

1108 86. Chen L; Lu Y; Li W; Ren Y; Yu M; Jiang S; Fu Y; Wang J; Peng S; Bilyk KT; Murphy KR; Zhuang X;
1109 Hune M; Zhai W; Wang W; Xu Q; Cheng CC (2019): The genome and transcriptome of the
1110 Antarctic toothfish *Dissostichus mawsoni* GigaScience Database.
1111 <http://dx.doi.org/10.5524/102162>

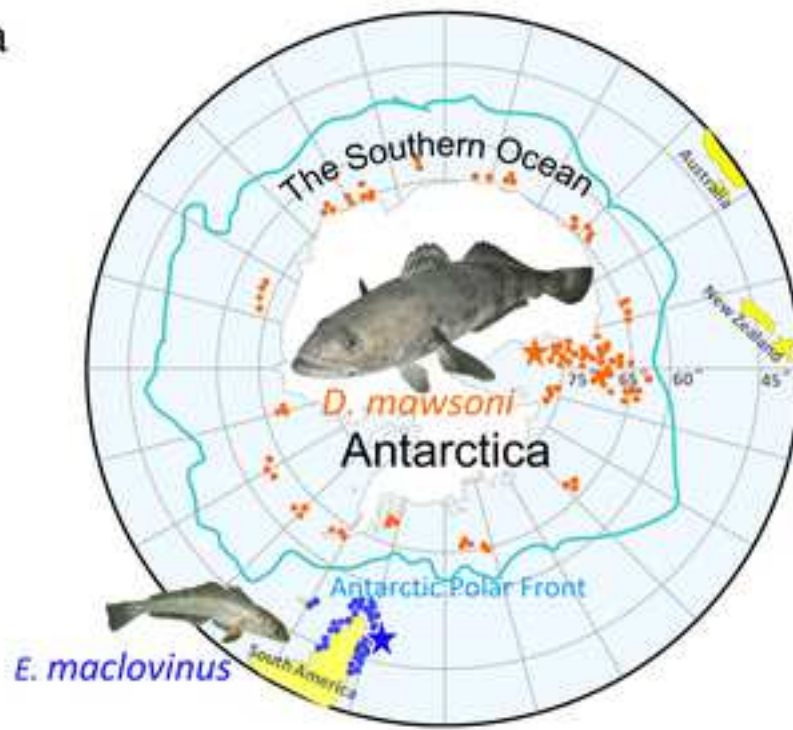
1112 87. Chen L; Lu Y; Li W; Ren Y; Yu M; Jiang S; Fu Y; Wang J; Peng S; Bilyk KT; Murphy KR; Zhuang X;
1113 Hune M; Zhai W; Wang W; Xu Q; Cheng CC (2019): The genome and transcriptome of the
1114 Patagonia robalo *Eleginops maclovinus* GigaScience Database.
1115 <http://dx.doi.org/10.5524/102163>

1116 88. Chen L; Lu Y; Li W; Ren Y; Yu M; Jiang S; Fu Y; Wang J; Peng S; Bilyk KT; Murphy KR; Zhuang X;
1117 Hune M; Zhai W; Wang W; Xu Q; Cheng CC (2019): Supporting data for "Genomic bases for
1118 colonizing the freezing Southern Ocean revealed by the genomes of Antarctic toothfish and
1119 Patagonia robalo" GigaScience Database. <http://dx.doi.org/10.5524/100545>

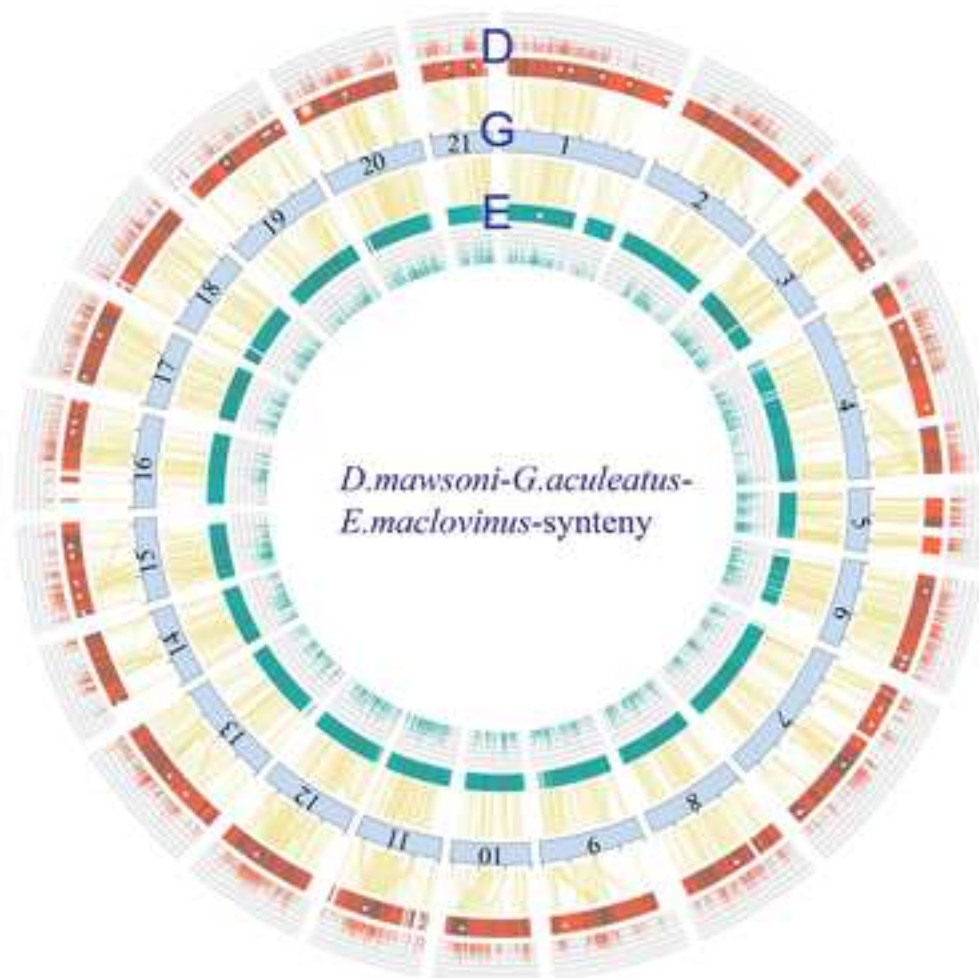
Table 1 Overview of assembly and annotation

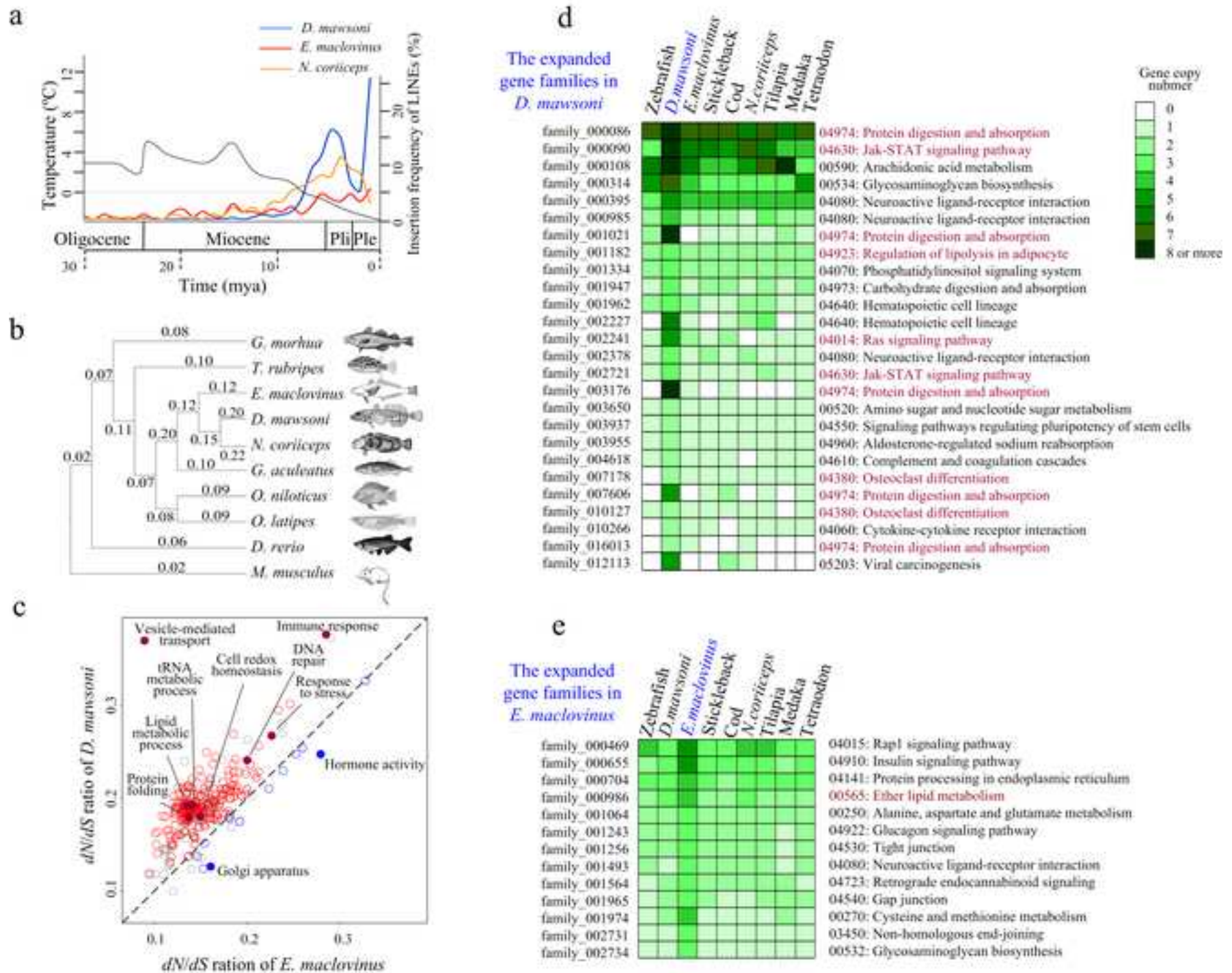
	<i>D. mawsoni</i>	<i>E. maclovinus</i>
Assembly		
Total length (Mb)	756.8	744.4
Contig N50 length (Kb)	23.1	10.9
Scaffold N50 length (Kb)	2,216.2	694.7
Scaffold N90 length (Kb)	202.7	167.2
Largest scaffold (Mb)	13.8	4.9
Quantity of scaffolds (>N90 length)	536	1,185
Annotation		
Quantity of predicted 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

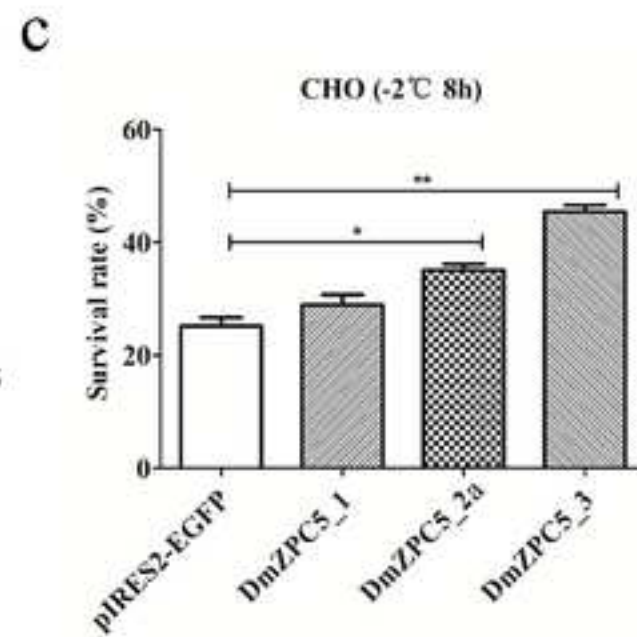
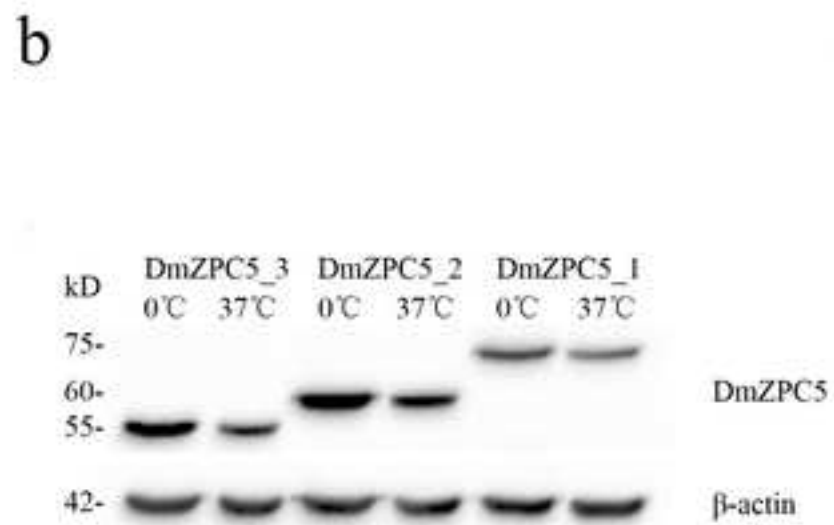
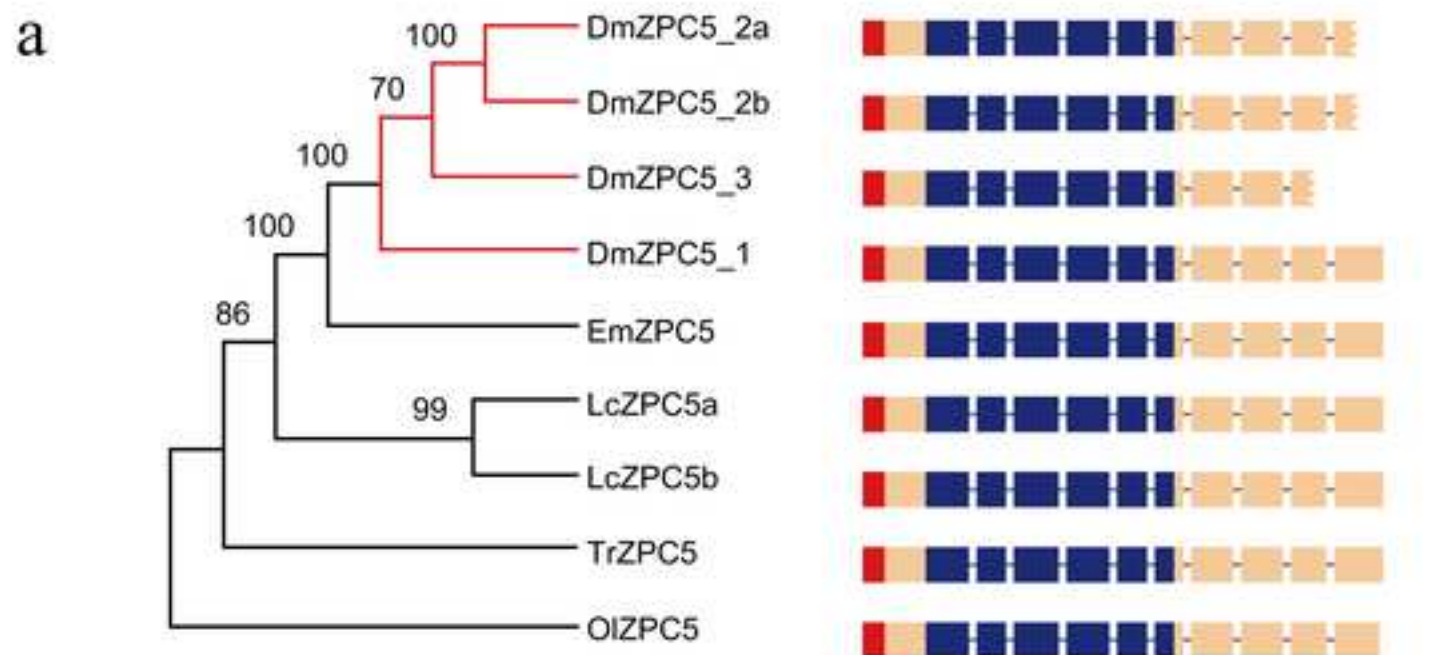
a

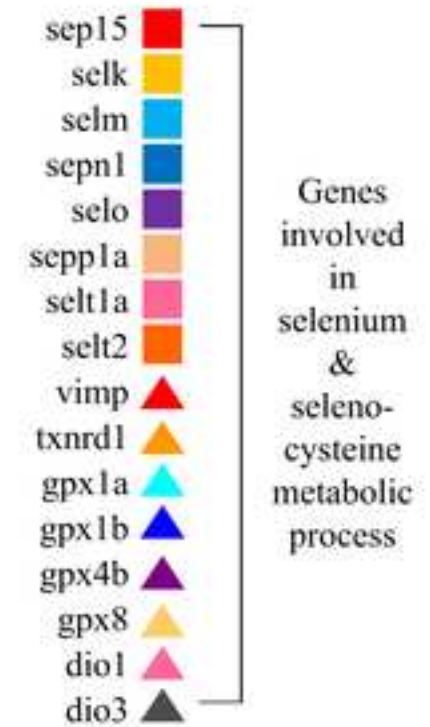
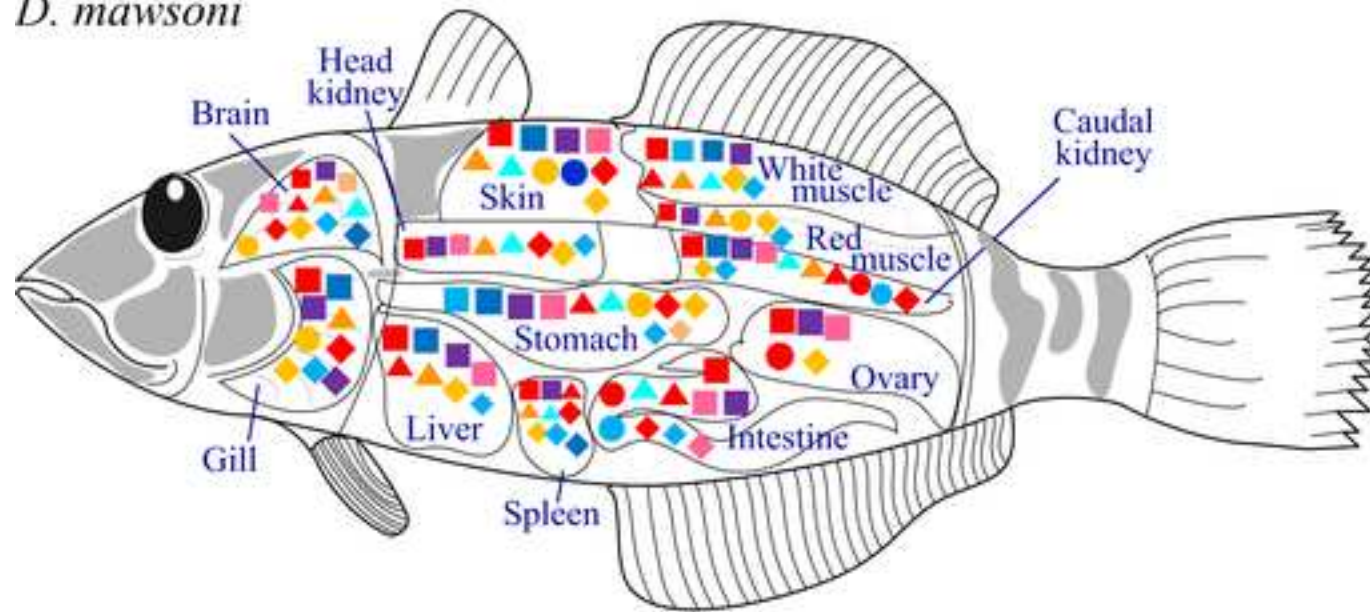
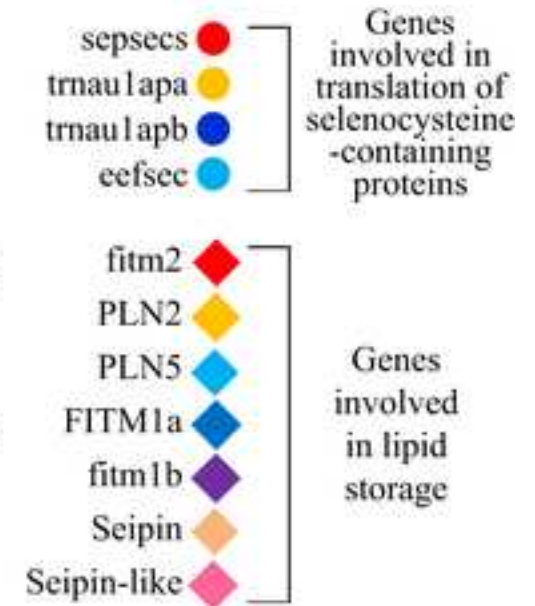
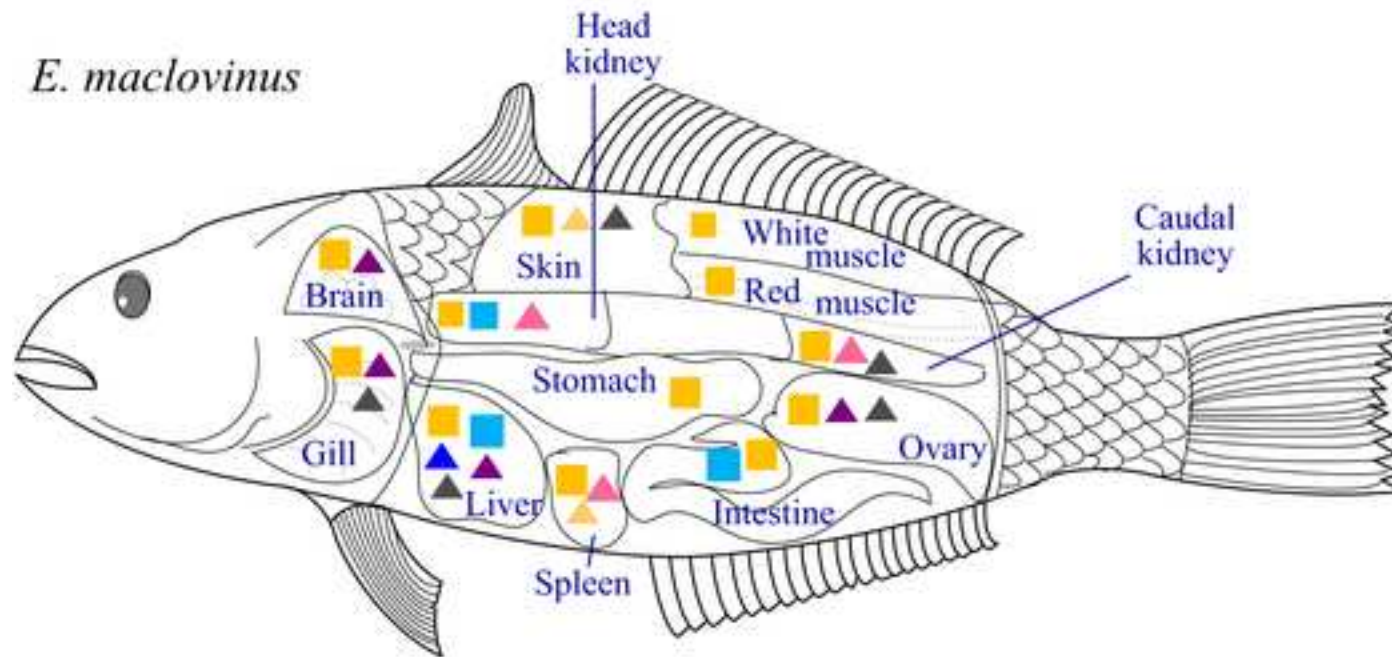


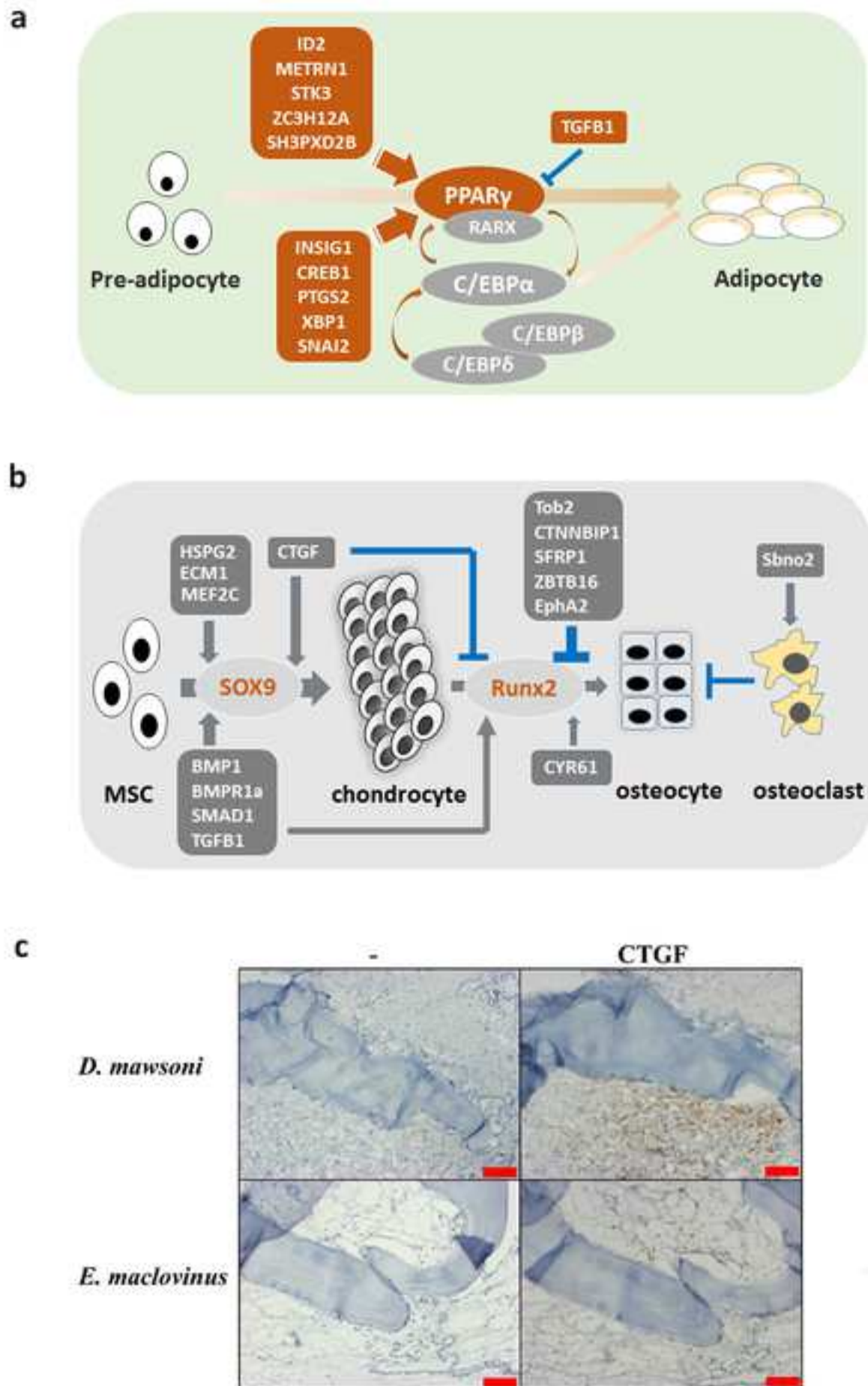
b







D. mawsoni*E. maclovinus*



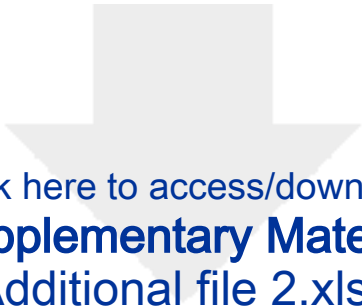


[Click here to access/download](#)

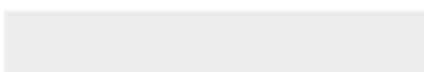
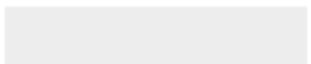
Supplementary Material

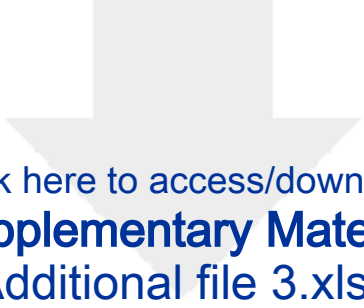
Additional file 1 revised26Nov2018.docx



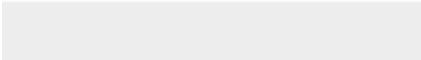



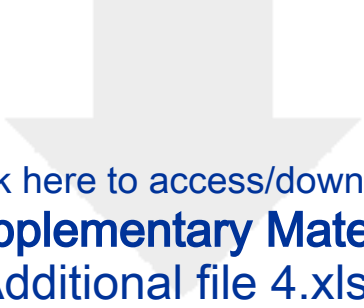
Click here to access/download
Supplementary Material
Additional file 2.xlsx



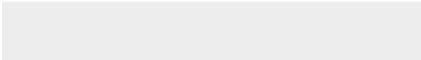



Click here to access/download
Supplementary Material
Additional file 3.xlsx





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



Dear Gigascience editors,

We responded to the reviewers' comments point-to-point and made corresponding revisions to the manuscript entitled "Genomic bases for colonizing the freezing Southern Ocean revealed by the genomes of Antarctic toothfish and Patagonia robalo". The important revisions are outline as below:

1. Paragraph 3 of Introduction: we added a citation of the paper that reports the MHC loci of the Antarctic icefish *C. aceratus* to be more accurately reflecting the status of the genomic studies in the Antarctic fishes.
2. Paragraph 2 of Section "Genome annotation and synteny alignment between *D. mawsoni* and *E. maclovinus*". We added the information of transposable elements of the published *Notothenia coriiceps* genome to comparison, and found that both *D. mawsoni* and *N. coriiceps* contained higher TE contents in relative to *E. maclovinus*. We also corrected two numbers: the TEs content of *D. mawsoni* (21.38%) and *E. maclovinus*(10.02%), in this section. The errors were resulted from mistakes made during citing the Additional file 1: Table S9. The statistics of *N. coriiceps* repeats is also added to **Additional file 1: Table S9**.
3. Paragraph 1 of Section "Burst of LINE expansion in the cold": information of LINE expansion in *N. coriiceps* is added. The estimated timing of *N. coriiceps* LINE expansion is also added to **Fig. 2a**.
4. Paragraph 1 of Section " Gene duplication in the freezing environment": the identities of the duplicated genes detected in this study which are consistent with those previously reported in (Chen et al., 2008) were added and shown in **Additional file 1: Fig. S4b**.
5. Section "Specimens, sampling, and DNA and RNA isolation": Sample collection and DNA/RNA preparation are described in more detail.

We believe these revisions have thoroughly addressed the reviewer's concerns. We also edited the manuscript to meet the standard for publication in Gigascience. We look forward to a favorable decision from you.

Sincerely

Liangbiao Chen and Chi-Hing C. Cheng