Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish

Zuozhou Chen*†, C.-H. Christina Cheng†‡, Junfang Zhang*†§, Lixue Cao*§, Lei Chen*, Longhai Zhou*, Yudong Jin*, Hua Ye*§, Cheng Deng*§, Zhonghua Dai*§, Qianghua Xu*, Peng Hu*§, Shouhong Sun*, Yu Shen*, and Liangbiao Chen*¶

*Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, §Graduate School of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100101, China; and ‡Department of Animal Biology, University of Illinois, 515 Morrill Hall, Urbana, IL 61801

Edited by Sean B. Carroll, University of Wisconsin, Madison, WI, and approved July 14, 2008 (received for review March 15, 2008)

The antifreeze glycoprotein-fortified Antarctic notothenioid fishes comprise the predominant fish suborder in the isolated frigid Southern Ocean. Their ecological success undoubtedly entailed evolutionary acquisition of a full suite of cold-stable functions besides antifreeze protection. Prior studies of adaptive changes in these teleost fishes generally examined a single genotype or phenotype. We report here the genome-wide investigations of transcriptional and genomic changes associated with Antarctic notothenioid cold adaptation. We sequenced and characterized 33,560 ESTs from four tissues of the Antarctic notothenioid *Dissostichus mawsoni* and derived 3,114 nonredundant protein gene families and their expression profiles. Through comparative analyses of same-tissue transcriptome profiles of *D. mawsoni* and temperate/tropical teleost fishes, we identified 177 notothenioid protein families that were expressed many fold over the latter, indicating cold-related up-regulation. These up-regulated gene families operate in protein biosynthesis, protein folding and degradation, lipid metabolism, antioxidation, antiapoptosis, innate immunity, choriongenesis, and others, all of recognizable functional importance in mitigating stresses in freezing temperatures during notothenioid life histories. We further examined the genomic and evolutionary bases for this expressional up-regulation by comparative genomic hybridization of DNA from four pairs of Antarctic and basal non-Antarctic notothenioids to 10,700 *D. mawsoni* cDNA probes and discovered significant to astounding (3- to >300-fold, $P < 0.05$) Antarctic-specific duplications of 118 protein-coding genes, many of which correspond to the up-regulated gene families. Results of our integrative tripartite study strongly suggest that evolution under constant cold has resulted in dramatic genomic expansions of specific protein gene families, augmenting gene expression and gene functions contributing to physiological fitness of Antarctic notothenioids in freezing polar conditions.

cold adaptation | comparative genomics | gene duplication | genome evolution | retrotransposon

The detachment of Antarctica from Gondwana and the establishment of the Antarctic Circumpolar Current commenced the geographic and thermal isolation of the Southern Ocean during early Cenozoic (1, 2). The circumpolar barriers and the ensuing climatic changes leading to sea-level glaciation and freezing water temperatures had driven major episodes of vicariance and extinction of the Antarctic marine fish fauna (3). The ancestral Antarctic notothenioid fish evolved antifreeze glycoproteins (AFGP) and was able to colonize freezing habitats (4). With little niche competition, it underwent an adaptive radiation to become the predominant Antarctic fish group today (5), providing an excellent system for exploring the relationship between evolutionary genomic change and environmental adaptation (6).

Adaptive changes in Antarctic notothenioids essential for cold survival have been reported in various studies. Besides the key evolutionary innovation, AFGP (4, 7), examples include adaptive modification of enzyme protein structures (8), cold-efficient microtubule assembly (9), cold-adapted protein translocation (10), and elevated mitochondrial densities (11). Evolution in chronic cold was also accompanied by striking gene loss, most notably the loss of hemoproteins in the icefish family *Channichthyidae* (12). These studies revealed important cold-related changes but were limited to a single genotype or phenotype. We report here integrated genome-wide investigations into the full range of changes in the genome and gene expression patterns relevant to evolutionary cold adaptation in Antarctic notothenioids.

We first performed large-scale EST sequencing of the brain, liver, head kidney, and ovary of the Antarctic notothenioid *Dissostichus mawsoni* and profiled the expression patterns of these transcriptomes. We then carried out multitissue comparative transcriptome analyses between *D. mawsoni* and non-notothenioid warm-water teleost fishes to identify cold-specific expressed genes. Third, using *D. mawsoni* ESTs as probes, we performed array-based comparative genomic hybridizations (13) for pairs of related Antarctic and non-Antarctic notothenioids to elucidate evolutionary cold-specific genomic changes. This integrated tripartite study identified >200 protein gene families with diverse putative stress-mitigating functions that were significantly up-regulated in expression and/or expanded through gene duplications in Antarctic notothenioids, providing a comprehensive view into the genome-wide responses in these teleost fish to evolutionary selection pressures in the freezing Southern Ocean.

Results and Discussion

Gene Expression Profile of *D. mawsoni* Transcriptomes. We obtained 33,560 high-quality ESTs averaging >460 reliable nucleotide reads from sequencing randomly selected clones of unnormalized long-insert cDNA libraries of *D. mawsoni* brain, liver, ovary, and head kidney. The ESTs were assembled into 13,000 unigenes, of which 6,208 (48%) were identified as protein coding genes [Table S1 in supporting information (SI) *Appendix*]. The protein genes were further condensed into 3,114 nonredundant gene families using an in-house pipeline under rigorous accuracy checks (*SI Text, File s1* in *SI Appendix*) to ensure each gene family represents a unique protein or subunits of the same protein. The

Author contributions: Liangbiao Chen designed research; C.-H.C.C., J.Z., L. Cao, Lei Chen, L.Z., Y.J., H.Y., C.D., Z.D., Q.X., P.H., S.S., and Y.S. performed research; Z.C. and Liangbiao Chen analyzed data; and C.-H.C.C. and Liangbiao Chen wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the National Center for Biotechnology Information Expressed Sequence Tags (EST) database (accession nos. FE193727–FE230830).

†Z.C., C.-H.C.C., and J.Z. contributed equally to this work.

¶To whom correspondence should be addressed. E-mail: lbchen@genetics.ac.cn.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802432105/DCSupplemental.

B

Gene	Percentage
Brain (9,979)	
Ependymin-1	0.88%
MMP nas-17	0.67%
S100 β	0.65%
Liver (8,718)	
Apolipoprotein A-I	6.06%
Fibrinogen β	1.38%
14 kDa apolipoprotein	1.06%
Ovary (9,754)	
ZPC5	4.48%
ZPC1	4.10%
ZPA	1.83%
Head Kidney (5,109)	
Hepcidin	5.44%
α1-globin	2.64%
Ig λ-5	1.35%
All Tissue (33,560)	
Tubulin β2	0.33%
Ubiquitin	0.28%
EF-1α	0.26%
Actin	0.24%
Tubulin α-ubiquitous	0.23%
Calmodulin	0.23%
Creatine kinase	0.22%
RBP S18	0.21%
HSP 90 β	0.21%
HSC(P) 70	0.21%
PPIase	0.20%
Glycine-rich RNA binding, GRP1A	0.19%

Fig. 1. Transcriptome analysis of an Antarctic notothenioid fish. (*A*) The expression profile of 3,114 protein gene families in brain, liver, ovary, and head kidney of *D. mawsoni*. The profile was generated with data in Table S2 in *SI Appendix* and arranged in the same order. It was clustered with Cluster 3.0 (14) using $\log_2$ transformation with the Pearson correlation coefficient metric and rendered with JAVA TreeView (15). Blue and white indicate the presence or absence, respectively, of the specific transcripts. Transcript abundance is indicated by the intensity of the blue color. (*B*) The top tissue-specific and all-tissue-expressed genes and their transcript percentages in the transcriptome(s). Total number of ESTs in the transcriptomes is shown in parentheses.

transcript frequencies, i.e., the total number of ESTs within a gene family, were tabulated for each of the four *D. mawsoni* transcriptomes (Table S2 in *SI Appendix*), clustered (14), and shown as TreeView plots (15) in Fig. 1*A*. The most abundant transcripts within each tissue and across all four tissues are listed in Fig. 1*B*.

The transcriptome analyses revealed several interesting features of *D. mawsoni* gene expression. Most strikingly, each transcriptome is dominated by transcripts from a small number of genes. Transcripts of apolipoprotein A-I (a high-density lipid transporter), zona pellucida (egg shell) proteins, and hepcidin (dual function as antibacterial agent and iron regulation hormone) constitute 6%, >10%, and 5% of total liver, ovary and head kidney ESTs respectively (Fig. 1*B*). Although this highly biased transcription may reflect up-regulation for major tissue-specific functions, some of the dominant transcripts are known to be associated with stress or cold response, for instance, the top three transcripts in the brain (Fig. 1*B*). S100β, an astrocyte produced molecule in brain trauma, suppresses oxidative cell damage and performs neurotrophic and neuroprotective functions (16), matrix metalloproteinases are involved in extracellular matrix remodeling during damaged neuronal tissue recovery (17), and the cell matrix adhesion protein ependymin exhibits cold-induced up-regulation in the brain of two teleost fishes (18). The top 12 all-tissue transcripts (Fig. 1*B*) included HSP90β, HSP(C)70s, PPIase, glycine-rich RNA-binding protein, creatine kinase and calmodulin, all major players in cellular defenses against thermal, oxidative and other stressors (19). Thus, the four *D. mawsoni* transcriptomes collectively appear to convey a transcriptional shift toward a functional theme of elevated stress response in this cold-adapted notothenioid.

Transcriptome Comparisons Between *D. mawsoni* and Warm-Water Teleost Fishes. The highly biased transcription patterns and the notable stress-responding nature of the transcriptomes of the cold-adapted *D. mawsoni* may represent manifestations of cold-adaptive gene expression. To investigate this possibility and uncover potential cold-adaptive genes in *D. mawsoni*, we performed same-tissue comparative analyses of *D. mawsoni* transcriptomes and available transcriptomes of model warm-water teleost fishes (because transcriptome data from related non-Antarctic notothenioids were lacking). We downloaded the sequences of 11 (four brain, three liver, three ovary, and one head kidney) unnormalized high-volume EST datasets of five temperate/tropical teleosts, *Danio rerio*, *Salmo salar*, *Gasterosteus aculeatus*, *Fundulus heteroclitus*, and *Oryzias latipes*, and processed them with the same bioinformatics pipeline used for *D. mawsoni* EST data. The ratio of the expression level (EST frequency) of each of the 3,114 *D. mawsoni* protein gene families and its homolog in the same tissue in the warm-water reference teleost was calculated (detailed methods in *SI Text*, *File s1*, in *SI Appendix*). Through rigorous statistical evaluation and stringent consistency checks for all comparison pairs (false discovery Rate at 10%; corrected P value of comparisons at <0.05 in Fisher's exact test; *SI Text*, *File s1*, in *SI Appendix*), we identified 189 gene families as differentially expressed (Table S3 in *SI Appendix*) in *D. mawsoni* relative to the warm-water species. Of the 189, 177 (94%) were up-regulated in *D. mawsoni* (Fig. 2*A*), remarkably consistent with the observation in carp *Cyprinus carpio* that 97% (252 of 260) of differentially expressed genes during cold acclimation were up-regulated (20). Furthermore, 85 (48%) of the up-regulated genes in *D. mawsoni* were homologs of the cold-induced *C. carpio* transcripts (20), attesting to the feasibility of our comparative transcriptomics approach in detecting cold-related gene expression in *D. mawsoni* despite phylogenetic distances of the reference species.

Many of the differentially expressed gene families are tissue-specific, with brain and liver showing greater numbers (Fig. 2*A*). Approximately 20% of the differentially up-regulated genes are found in two or more tissues, suggesting the presence of a common underlying set of enhanced cellular responses. The lack of qualified head kidney transcriptomes in the database limited us to a single comparison for this tissue; nevertheless, a trend toward up-regulation was visible (Fig. 2*A*), most notably in the coelevation of 29 ribosomal protein genes (Table S3 in *SI Appendix*). The increase in ribosomal biogenesis suggests enhanced protein synthesis capacity in this tissue, which was reported for the gill and white muscle of another cold-adapted Antarctic teleost, the zoarcid *Pachycara brachycephalum* (21).

To assist in functional interpretations and evaluate potential cold-adaptive response, we performed Gene Ontology (GO) annotation and literature searches by which most of the differentially expressed genes were assigned to 15 functional groups (Table S3 in *SI Appendix*). The main subsets of genes with related biological functions in 11 of the 15 GO groups are presented in Fig. 2*B i-xi*. These are: (*i*) molecular chaperones including the conserved heat-shock proteins and protein-specific chaperones (22); (*ii* and *iii*) genes that operate in ubiquitin-dependent and -independent protein degradation machineries; (*iv*) lipid transport and membrane metabolism-related genes; (*v*) reactive oxygen species (ROS) scavengers; (*vi*) apoptosis regulators, especially those of antiapoptotic functions; (*vii*) genes involved in metal ion and solute homeostasis; (*viii*) plasma proteins of blood coagulation and innate immunity; (*ix*) diverse eggshell (ZP) proteins; (*x*) components of the Ras/MAPK signal transduction pathway; and (*xi*) factors involved in messenger RNA and ribosomal RNA transcription. The concomitant differential

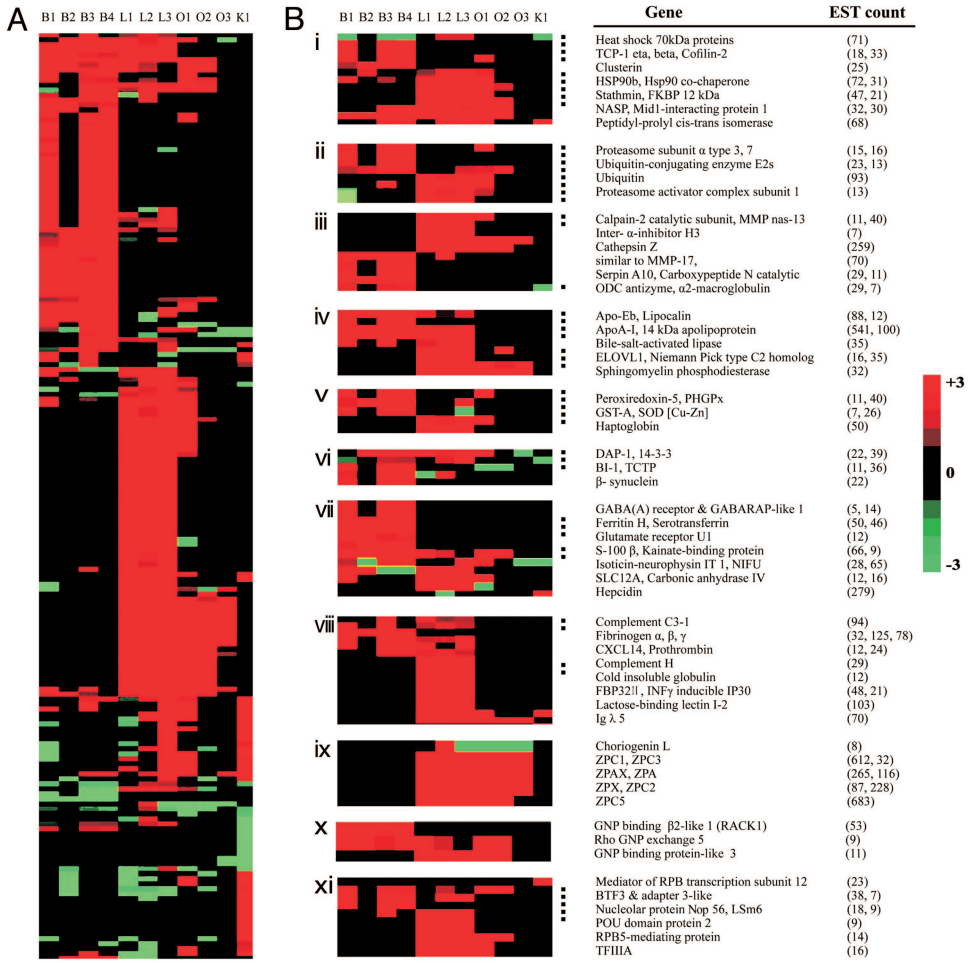

Fig. 2. Transcriptome comparison between *D. mawsoni* and temperate/tropical teleosts revealed co-up-regulation of multiple gene groups in the Antarctic notothenioid fish. (*A*) Tissue distribution and expression ratio of 189 differentially expressed genes between *D. mawsoni* and five temperate/tropical teleost fishes identified from same-tissue transcriptome comparisons. Lanes B1-B4, L1-L3, O1-O3, and K1 showed comparison results of *D. mawsoni* brain vs. *D. rerio*-1522, -14409 and *G. aculeatus*-17209, -18921 brain libraries; *D. mawsoni* liver versus *D. rerio*-14410, *F. heteroclitus*-15870 and *O. latipes*-17414 liver libraries; *D. mawsoni* ovary versus *S. salar*-15459 and *D. rerio*-9767, -15519 ovary libraries; and *D. mawsoni* head kidney versus *S. salar*-15454 head kidney library, respectively. The graph was derived from data in Table S3 in *SI Appendix*, clustered and rendered by the same programs described for Fig. 1*A*. (*B*) Eleven functional groups of differentially up-regulated genes in *D. mawsoni*. (*Left*) TreeView plot. Red and green indicate statistically significant up- or down-regulation, respectively, of transcript abundance in *D. mawsoni* relative to the compared species ($P < 0.05$ in Fisher's exact test). Black indicates no difference. The color scale is derived from $\log_2$ transformation of differential expression ratios in Table S3 in *SI Appendix*. Rows dotted with a small black square indicate the homologs of cold-responding genes identified in *C. carpio* (20). (*Right*) Names and EST counts of the highly up-regulated genes in each functional group. Sequentially, the 11 groups represent: molecular chaperones, ubiquitin-dependent protein degradation, proteolytic enzymes and inhibitors, lipid metabolism, antioxidant, apoptosis regulation, ion and solute homeostasis, immunity, zona pellucida proteins, Ras/MAPK pathway, and transcription factors.

up-regulation of multiple genes within these functional categories in *D. mawsoni* compared with warm-water teleosts suggests greater demands for these functions in the cold Antarctic environment. The cold-adaptive relevance of their expression is supported by the fact that the majority (75%) of the top expressed genes in *D. mawsoni* tissues (Fig. 1*B*) are found in this differentially up-regulated gene set (Table S3 in *SI Appendix*, Fig. 2*B*).

The co-up-regulated putative cold-adaptive genes identified lend support to reported evidence and provide grounds for testable hypotheses regarding mechanisms of cold adaptation at cellular and systems levels. One example is the co-up-regulation of three groups of genes operating in protein homeostasis, i.e., maintenance of functioning protein pools (Figs. 1*B*, 2*Bi*, 2*Bii* and 2*Biii*). The concerted up-regulation of HSPs and protein-specific chaperones (Figs. 1*B* and 2*Bi*) in *D. mawsoni* strongly indicates a constitutive system-wide need for mitigating cold denaturation of proteins (23) in the face of chronic cold. It supports the reported constitutive HSP70 expression and loss of HSP70 inducibility in some notothenioids (24). The augmented need for cellular protein homeostasis is also indicated in the up-regulation of genes in ubiquitin-dependent protein degradation pathways (Fig. 2*Bii*) and support the reported high levels of ubiquitin-conjugated proteins in Antarctic notothenioid tissues (25). Besides genes in intracellular protein catabolism, transcripts of several extracellular proteases such as Cathepsin Z and MMP-17 homolog (Fig. 2*Biii*) were ≥11-fold greater in *D. mawsoni* than in warm-water teleosts (Table S3 in *SI Appendix*). Collectively, these up-regulated expressions argue for the hypothesis that adaptive enhancements in protein homeostasis capabilities, both intracellular and extracellular, are important for life in chronic cold.

A second example pertains to antioxidant function. The subzero Antarctic waters are also oxygen-rich due to increased O_2 solubility at low temperatures. Living in constant high-oxygen tension, Antarctic notothenioids conceivably would face in-

creased levels of ROS in their tissues and thus system-wide oxidative stress (26). This hypothesis finds support in three GO categories of up-regulated genes in *D. mawsoni* involved in ROS and oxidative stress mitigations. Group v (Fig. 2*Bv*) contains genes that act in ROS scavenging, including PHGPx, peroxiredoxin-5, and superoxide dismutase, and the potent antioxidant haptoglobin that alleviates hemoglobin-driven oxidative stress in blood cells (27). Group vi (Fig. 2*Bvi*) contains antiapoptotic genes, including BI-1 and TCTP with known roles in mitigating ROS-mediated apoptosis (28) and cellular oxidative stress (29). Group vii (Fig. 2*Bvii*) contains genes involved in iron transport and storage, including ferritin and serotransferrin that bind free iron and the small peptide hormone hepcidin that reduces iron absorption (30). The up-regulation of these genes would be consistent with the need for controlling free Fe^{2+}-catalyzed ROS generation (31). Simultaneous up-regulation of these three groups of genes suggests augmented defenses against oxidative stress likely constituted an important aspect of evolutionary adaptation of the Antarctic notothenioids in its oxygen-rich environment. The identification of cold-specific adaptive evolution of novel hepcidin isoforms in the Antarctic notothenioids (32) lent support to this hypothesis.

Collectively, the results of transcriptome expression patterns in *D. mawsoni* and comparative transcriptome analyses provided us with an overview of the nature of transcriptomic changes potentially important in cold adaptation. The spectrum of up-regulated gene groups we identified will enable further in-depth investigations into mechanisms of notothenioid evolutionary cold adaptation in much wider physiological contexts than currently examined in the field.

Gene Duplication in Antarctic Notothenioid Fish Genomes. To assess genomic contribution to the transcriptional up-regulations in *D. mawsoni*, we performed array-based Comparative Genomic Hybridization (aCGH) of genomic DNA from selected pairs of related Antarctic and non-Antarctic notothenioids, using 10,700 *D. mawsoni* ESTs as probes. Three Antarctic species, *D. mawsoni*, *Pagothenia borchgrevinki*, and *Chaenocephalus aceratus*, were referenced to the phylogenetically basal New Zealand species *Bovichtus variegatus* or the South American species *Eleginops maclovinus* (phylogenetic relationship in Fig. S1 in *SI Appendix*) in four aCGHs (Fig. 3*A*). Differential hybridization signal intensity of a gene across all four aCGH pairs would indicate a gain or loss of gene copy number during the evolutionary radiation of Antarctic notothenioids.

Using a stringent selection criterion (Antarctic:non-Antarctic signal ratio of 2.4:1) (detailed methods in *SI Text*, *File s2*, and Fig. S2 in *SI Appendix*), we identified 101 protein-coding genes distinct from the well known AFGP gene and 17 long interspersed nuclear elements (LINEs) that were duplicated or potentially newly acquired in the three Antarctic species, and only 12 genes that were contracted (signal ratio <0.43, i.e., 1/2.4) (Fig. 3*A*; gene details and signal ratios in Table S4 in *SI Appendix*). Thus, duplicated genes greatly outnumbered contracted ones, by 10:1. Also, mean duplication ranged from ≈3- to >300-fold ($P < 0.05$), far greater than reduction, which are 0.25 and 0.3 ($P < 0.01$), respectively, for the two most contracted genes, a LINE and Na^+/K^+ ATPase α subunit (Table S4 in *SI Appendix*). The majority of duplicated genes are homologs of known or hypothetical proteins, indicating duplication of preexisting genes to augment gene function. The dominance of gene duplication over contraction held true when the aCGH selection ratio was lowered and non-protein coding genes were included (Table S5 in *SI Appendix*). Southern blot hybridizations using probes derived from four of the putative duplicated genes (FBP32II, ZPC5, a 10-aa repeat protein, LINE, with β-actin as control) showed many intense hybridizing bands in the Antarctic species vs. few and weakly hybridizing bands in *B. variegatus* and

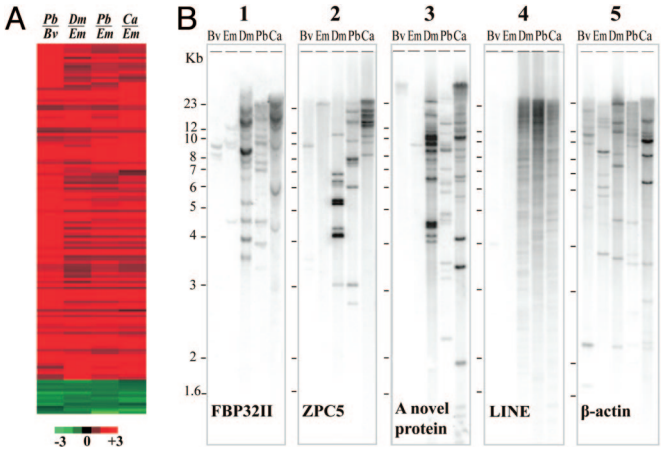

Fig. 3. Extensive gene duplications occurred in Antarctic notothenioid fish genomes. (*A*) Antarctic-specific gene duplication and contraction revealed by aCGH pairs of Antarctic (*Pb*, *P. borchgrevinki*; *Dm*, *D. mawsoni*; *Ca*, *C. aceratus*) and non-Antarctic notothenioids (*Bv*, *B. variegatus*; *Em*, *E. maclovinus*). The plot is derived from data in Table S4 in *SI Appendix*, clustered and rendered by the same programs described for Fig. 1*A*. Gene duplications (red) greatly outnumbered gene contractions (green). The red color may be saturated for highly duplicated genes (see Table S4 in *SI Appendix* for the detail of genes, duplication folds, and annotations). (*B*) Southern blot hybridization of genomic DNA from the five species used in aCGH to verify gene duplication. Approximately 20 μg of EcoRI digested genomic DNA was applied in each lane except lanes *Dm*, *Pb*, and *Ca* in *4*, where 5 μg was applied. Four genes that show duplications in aCGH-FBP32II (*L129B07*), ZPC5 (*O116C07*), a 10-aa repeat protein (*O120F01*), and LINE (*L042D05*) (Table S4 in *SI Appendix*) were probed. β-actin (*B383D08*) is nonduplicated control gene. The test genes, including the fast-evolving ZPC5 and LINE, were detectable in the non-Antarctic notothenioids, indicating that sequence divergence was not a factor in the differential aCGH signals between the Antarctic/non-Antarctic species pairs.

E. maclovinus (Fig. 3*B*), validating the occurrence of extensive gene duplications in the Antarctic lineages inferred from aCGHs.

Among the 118 duplicated genes (Table S4 in *SI Appendix*), 15 encode different enzymes potentially adaptive for cold survival. For example, matrix metalloproteinases MMP9 and nephrosin are involved in extracellular matrix remodeling, a known process of adaptive importance in cold acclimation (20). Seven genes are involved in protein folding and ubiquitin-dependent protein degradation, and nine genes function in stress-response, anti-ROS, and antiapoptotic processes. Duplication also occurred in nine genes participating in Ras/MAPK and TGF-β signal transduction pathways and in 10 genes encoding proteins of cellular structural components, such as the nuclear pore and cytoskeleton. Also duplicated are five genes involved in RNA binding and processing known to be important for mRNA function in the cold, and five zinc finger proteins potentially involved in gene transcription (Table S4 in *SI Appendix*). Many of these duplicated genes correspond to genes showing elevated transcription, either being the same genes or genes in the same biological processes (Figs. 1*B* and 2*B*), similarly for genes involved in innate immunity, such as hepcidins that act as antibacterial agents besides iron regulator, and FBP32II that encodes F-type lectins. FBP32II showed a 14-fold average increase in gene copies in the three Antarctic species compared with *B. variegatus* and *E. maclovinus* (Table S4 in *SI Appendix*, Fig. 3*B1*), and >53-fold greater transcription in *D. mawsoni* liver relative to warm-water teleosts (Fig. 2*Bviii*, Table S3 in *SI Appendix*). Lectins recognize specific carbohydrate domains on bacterial cell surfaces and elicit immune responses. Interestingly, fish type II antifreeze proteins evolved from C-type lectins (33), suggesting lectin gene

EVOLUTION

duplications may be a prevalent response in coldwater teleosts, preceding the cooption of one copy to become an antifreeze.

The highly expressed eggshell protein genes (Figs. 1*B* and 2*Bix*) are also substantially duplicated. ZPC1, -2, -3, -5; ZPAX; ZPX; and ZPB genes expanded by 2.7- to 7.7-fold ($P < 0.05$) in the Antarctic notothenioids (Table S4 in *SI Appendix*). Extensive duplication of ZPC5 was verified in Southern hybridization of genomic DNA (Fig. 3*B2*). The enlarged ZP gene family undoubtedly contributed to the abundant ZP transcripts seen in *D. mawsoni*. Antarctic notothenioids synthesize a dense and thick chorion (34). Embryos of *Gymnodraco acuticeps* and *Pleuragramma antarcticum* develop in ambient icy freezing seawater but have drastically inadequate amounts of AFGPs to avoid freezing; thus, the dense chorion was considered to act as a crucial physical barrier against ice penetration (35). The need for a durable protective eggshell could have been a strong selection for the extensive duplication of ZP genes.

We also identified >20 duplicated genes without known biological functions (Table S4 in *SI Appendix*). Two of these, which encode proteins of 10- (EKLNGTMSDE) and 11-residue (RHDGVNETNDV) repeats, respectively, have no homologs in other organisms and may be new acquisitions in Antarctic notothenioids. The extensive duplication of the repetitive 10-residue protein is shown in Fig. 3*B3*. Both genes are highly transcribed in *D. mawsoni* ovary (Fig. S3 in *SI Appendix*), suggesting a role in ovarian function that remains to be elucidated.

The most astounding Antarctic-specific gene duplication revealed by aCGH occurred in the LINEs, genetic elements prevalent in eukaryotic genomes capable of retrotransposition through self-coded enzymes. Seventeen LINE genes encoding reverse transcriptase and endonuclease showed Antarctic-specific duplication by 8- to >300-fold ($P < 0.05$), whereas only one underwent contraction (Table S4 in *SI Appendix*, Fig. 3*B4*). The LINE expansions correlate with increasing genome sizes from the phylogenetically basal notothenioids to the derived species. The C values of *B. variegatus*, *D. mawsoni*, *P. borchgrevinki*, *G. acuticeps*, and *C. aceratus* are 0.84, 0.97, 1.28, 1.83, and 1.86 pg, respectively (C.-H.C.C., unpublished data), suggesting the gain of LINEs contributed to genome size enlargement in notothenioid evolution. The LINE expansion might have in turn facilitated the observed protein gene duplications in Antarctic notothenioids through retrotransposition-mediated sequence transduction (36). Although these putative roles of LINEs in Antarctic notothenioids remain to be explored, the remarkable Antarctic-specific LINE expansion and diversification represent a direct link between LINE evolution and cold-selection pressure.

Conclusion

Through integrated genome-wide investigations, we conducted a comprehensive elucidation of the genotypic changes accompanying evolutionary adaptation of Antarctic notothenioid fishes in the freezing Southern Ocean. Evolution in constant cold had apparently resulted in remarkable transcriptomic shifts and gene expansion in Antarctic notothenioids with respect to related and unrelated temperate species. From 6,208 unique protein-coding genes expressed in four tissues of *D. mawsoni*, we identified 177 gene families that are transcriptionally elevated, and 118 protein genes that are substantially duplicated, bringing about a nonredundant set of >200 genes that are specifically augmented in the Antarctic notothenioids. These genes in concert would enhance an array of biological processes including protein synthesis, protein folding and degradation, lipid metabolism, antioxidation, antiapoptosis, innate immunity, and choriongenesis, forging a collective stress-responding or mitigating phenotype to overcome various physiological challenges posed by the freezing and oxygen-rich Antarctic environment during notothenioid life histories. Because many of the duplicated protein genes correspond or are functionally similar to the dominant transcripts, gene family expansions likely contributed significantly to augmentation of cold-adaptive functions. The extensive duplications of the eggshell protein genes, LINEs, and two putative newly acquired proteins of repetitive sequences in the Antarctic species are suggestive of new cellular or genomic mechanisms of cold adaptation that will stimulate further research. The results of this study provided a comprehensive view into the making of notothenioid cold fitness and, in addition, a framework for exploring the genomic adaptability of the ecologically vital Antarctic notothenioids in face of global climate change.

Materials and Methods

Sample Collection. Antarctic notothenioids *D. mawsoni* and *P. borchgrevinki* were captured from McMurdo Sound, Antarctica, and *C. aceratus* from the Antarctic Peninsula. The two basal non-Antarctic notothenioids, *B. variegatus* and *E. maclovinus*, were obtained from Otago Harbor, New Zealand, and Ushuaia, Argentina, respectively. Tissues were dissected from anesthetized specimens and kept frozen at −80°C until use.

Construction of cDNA Libraries and Sequencing. Complementary DNA libraries of *D. mawsoni* tissues were constructed by using the pCMV-Script XR cDNA Library Construction Kit (Stratagene). Seven micrograms of polyA$^+$ RNA was reverse-transcribed with the addition of 0.6 M D-trehalose to improve first-strand cDNA length, and cDNAs with sizes >0.7 kb were recovered and directionally cloned. Analysis of random recombinant clones showed an average insert size of ≈1.8 kb (37). Plasmid DNA was isolated using AxyPrep Easy-96 isolation kit (Axygen Biosciences) and sequenced from the 5′ end of the insert using BigDye v.3.1 (Applied Biosystems) and resolved on an ABI 3730 sequencer.

Sequence Annotation and Transcriptome Comparison. Raw sequence data were edited with PHRED (38) by using a stringent error rate of 0.1%. The ESTs were assembled with CAP3 (39) and annotated with BLAST against National Center for Biotechnology Information databases. A bioinformatics pipeline was designed to generate accurate and nonredundant *D. mawsoni* gene families. The assembled genes have ≈50% homology to known proteins and are at least 70 aa in length (details in *SI Text*, *File s1*, in *SI Appendix*). Associated ESTs of each gene family were summed to obtain the expression level in each tissue. Eleven unnormalized EST libraries of warm/tropical teleost fishes were downloaded from the UniGene database. EST frequencies of a homologous gene family in *D. mawsoni* and the comparison species were divided to obtain the expression ratio. Gene families showing at least 1.5-fold of expression difference and passing Fisher's exact tests at FDR = 0.1 were retained and further screened for multiple-comparison consistency. Those consistently up- or down-regulated in at least three of the four brain comparisons, all three liver comparisons, all three ovary comparisons, or the head kidney comparison were regarded as differentially expressed genes in *D. mawsoni* relative to the reference teleost fish (details in *SI Text*, *File s1*, in *SI Appendix*).

Construction of cDNA Microarray and aCGH. Microarray slides containing 10,700 *D. mawsoni* ESTs and positive (AFGP coding sequence) and negative (DMSO) controls were constructed (details in *SI Text*, *File s2*, in *SI Appendix*). Genomic DNA from livers of *D. mawsoni*, *P. borchgrevinki*, *C. aceratus*, *B. variegatus*, and *E. maclovinus* were purified, sheared, and quantified. Three micrograms of sheared DNA from each species of a aCGH pair were labeled with Cy5- or Cy3-dCTP. Dye-reversed controls were performed for all aCGH pairs. Hybridizations followed the procedure developed by CapitalBio (*SI Text*, *File s2*, *SI Appendix*).

Data Normalization, Selection of Duplicated Genes. and Verification. Hybridized microarrays were scored and the fluorescence intensities extracted using GenePix Pro 4.0 (Axon Instruments). The Cy5 and Cy3 signals were normalized by setting the $\log_2$ average Cy5/Cy3 ratio of all of the elements equal to 0. Consistency of the replicated spots was evaluated by Student's *t* test, and the average Cy5/Cy3 ratios of spots that passed the test were calculated from the replicas, whereas failed spots were eliminated.

In interhominoid aCGHs, a $\log_2$ signal ratio of 0.5 between the test and reference species was used to infer gene duplication (13). To reflect the longer divergence times among the notothenioids, we adopted the signal ratio of a known duplicated gene, hepcidin between *D. mawsoni* and *E. maclovinus* as the selection criterion (Fig. S2 in *SI Appendix*), which was 2.4 ($\log_2$ 2.4 = 1.26). EST spots with signal ratios ≥2.4 or ≤0.43 in all four CGHs were selected as duplicated or contracted genes and annotated (detail see *SI Text*, *File s2*, in *SI Appendix*). A geometric mean duplication/contraction fold of an individual EST was calculated

from the four aCGH signal ratios (*Pb*/*Bv*, *Dm*/*Em*, *Pb*/*Em*, *Ca*/*Em*) to refer the extent of duplication or contraction for this gene between Antarctic and non-Antarctic notothenioids. To validate aCGH detected gene duplication, genomic DNA of the five notothenioids used in aCGHs was digested with EcoRI and probed with five selected ESTs by Southern blot hybridizations.

ACKNOWLEDGMENTS. This work was supported by Natural Science Foundation of China Grant 30330080, the 973 Program (Grant 2004CB117404), and the 863 Program (Grant 2006AA02Z331), to Liangbiao Chen, and National Science Foundation Grants OPP 0231006 and OPP 0636696, to C.-H.C.C.

1. Livermore R, Nankivell A, Eagles G, Morris P (2005) Paleogene opening of Drake Passage. *Earth Planet Sci Lett* 236:459–470.
2. Shevenell AE, Kennett JP, Lea DW (2004) Middle Miocene Southern Ocean cooling and Antarctic cryosphere expansion. *Science* 305:1766–1770.
3. Eastman JT (2005) The nature of the diversity of Antarctic fishes. *Polar Biol* 28:93–107.
4. Chen L, DeVries AL, Cheng C-HC (1997) Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc Natl Acad Sci USA* 94:3811–3816.
5. Eastman JT, McCune AR (2000) Fishes on the Antarctic continental shelf: Evolution of a marine species flock? *J Fish Biol* 57:84–102.
6. National Research Council ed (2003) *Frontiers in Polar Biology in the Genomic Era* (Natl Academy of Sciences, Washington, DC).
7. DeVries AL (1971) Glycoproteins as biological antifreeze agents in antarctic fishes. *Science* 172:1152–1155.
8. Fields PA, Somero GN (1998) Hot spots in cold adaptation: Localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc Natl Acad Sci USA* 95:11476–11481.
9. Detrich HW, 3rd, Parker SK, Williams RC, Jr, Nogales E, Downing KH (2000) Cold adaptation of microtubule assembly and dynamics. Structural interpretation of primary sequence changes present in the alpha- and beta-tubulins of Antarctic fishes. *J Biol Chem* 275:37038–37047.
10. Romisch K, *et al*. (2003) Protein translocation across the endoplasmic reticulum membrane in cold-adapted organisms. *J Cell Sci* 116:2875–2883.
11. O'Brien KM, Sidell BD (2000) The interplay among cardiac ultrastructure, metabolism and the expression of oxygen-binding proteins in Antarctic fishes. *J Exp Biol* 203:1287–1297.
12. Sidell BD, O'Brien KM (2006) When bad things happen to good fish: The loss of hemoglobin and myoglobin expression in Antarctic icefishes. *J Exp Biol* 209:1791–1802.
13. Fortna A, *et al.* (2004) Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol* 2:E207.
14. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20:1453–1454.
15. Saldanha AJ (2004) Java Treeview-extensible visualization of microarray data. *Bioinformatics* 20:3246–3248.
16. Nishikawa T, *et al.* (1997) Identification of S100b protein as copper-binding protein and its suppression of copper-induced cell damage. *J Biol Chem* 272:23037–23041.
17. Gasche Y, Soccal PM, Kanemitsu M, Copin JC (2006) Matrix metalloproteinases and diseases of the central nervous system with a special emphasis on ischemic brain. *Front Biosci* 11:1289–1301.
18. Tang SJ, *et al.* (1999) Cold-induced ependymin expression in zebrafish and carp brain: Implications for cold acclimation. *FEBS Lett* 459:95–99.
19. Kultz D (2005) Molecular and evolutionary basis of the cellular stress response. *Annu Rev Physiol* 67:225–257.
20. Gracey AY, *et al*. (2004) Coping with cold: An integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proc Natl Acad Sci USA* 101:16970–16975.
21. Storch D, Lannig G, Portner HO (2005) Temperature-dependent protein synthesis capacities in Antarctic and temperate (North Sea) fish (Zoarcidae). *J Exp Biol* 208:2409–2420.
22. Lopez-Fanarraga M, Avila J, Guasch A, Coll M, Zabala JC (2001) Review: Postchaperonin tubulin folding cofactors and their role in microtubule dynamics. *J Struct Biol* 135:219–229.
23. Franks F (1995) Protein destabilization at low temperatures. *Adv Protein Chem* 46:105–139.
24. Buckley BA, Place SP, Hofmann GE (2004) Regulation of heat shock genes in isolated hepatocytes from an Antarctic fish, *Trematomus bernacchii*. *J Exp Biol* 207:3649–3656.
25. Todgham AE, Hoaglund EA, Hofmann GE (2007) Is cold the new hot? Elevated ubiquitin-conjugated protein levels in tissues of Antarctic fish as evidence for cold-denaturation of proteins *in vivo*. *J Comp Physiol B* 177:857–866.
26. Abele D, Puntarulo S (2004) Formation of reactive species and induction of antioxidant defense systems in polar and temperate marine invertebrates and fish. *Comp Biochem Physiol A* 138:405–415.
27. Gueye PM, Glasser N, Ferard G, Lessinger JM (2006) Influence of human haptoglobin polymorphism on oxidative stress induced by free hemoglobin on red blood cells. *Clin Chem Lab Med* 44:542–547.
28. Lee GH, *et al*. (2007) Bax inhibitor-1 regulates endoplasmic reticulum stress-associated reactive oxygen species and heme oxygenase-1 expression. *J Biol Chem* 282:21618–21628.
29. Gnanasekar M, Ramaswamy K (2007) Translationally controlled tumor protein of. *Brugia malayi* functions as an antioxidant protein. *Parasitol Res* 101:1533–1540.
30. Ganz T, Nemeth E (2006) Regulation of iron acquisition and iron distribution in mammals. *Biochim Biophys Acta* 1763:690–699.
31. Kruszewski M (2003) Labile iron pool: The main determinant of cellular response to oxidative stress. *Mutat Res* 531:81–92.
32. Xu Q, *et al.* (2008) Adaptive evolution of hepcidin genes in antarctic notothenioid fishes. *Mol Biol Evol* 25:1099–1112.
33. Liu Y, *et al*. (2007) Structure and evolutionary origin of Ca^{2+}-dependent herring type II antifreeze protein. *PLoS ONE* 2:e548.
34. Motta CM, *et al*. (2005) Oogenesis at subzero temperatures: A comparative study of the oocyte morphology in nine species of Notothenioids. *Tissue Cell* 37:233–240.
35. Cziko PA, Evans CW, Cheng CH-C, DeVries AL (2006) Freezing resistance of antifreeze-deficient larval Antarctic fish. *J Exp Biol* 209:407–420.
36. Xing J, *et al*. (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci USA* 103:17608–17613.
37. Chen L, *et al*. (2007) Trehalose as a good candidate for enriching full-length cDNAs in cDNA library construction. *J Biotechnol* 127:402–407.
38. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
39. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:868–877.