ND6 Gene "Lost" and Found: Evolution of Mitochondrial Gene Rearrangement in Antarctic Notothenioids

Xuan Zhuang and C.-H. Christina Cheng*

Department of Animal Biology, University of Illinois at Urbana–Champaign

***Corresponding author:** E-mail: c-cheng@uiuc.edu.

Associate editor: Willie Swanson

Abstract

Evolution of Antarctic notothenioids in the frigid and oxygen-rich Southern Ocean had led to remarkable genomic changes, most notably the gain of novel antifreeze glycoproteins and the loss of oxygen-binding hemoproteins in the icefish family. Recently, the mitochondrial (mt) NADH dehydrogenase subunit 6 (*ND6*) gene and the adjacent transfer RNA^{Glu} ($tRNA^{Glu}$) were also reportedly lost. ND6 protein is crucial for the assembly and function of Complex I of the mt electron transport chain that produces adenosine triphosphate (ATP) essential for life; thus, *ND6* absence would be irreconcilable with Antarctic notothenioids being thriving species. Here we report our discovery that the *ND6* gene and $tRNA^{Glu}$ were not lost but had been translocated to the control region (CR) from their canonical location between *ND5* and cytochrome b genes. We characterized the CR and adjacent sequences of 22 notothenioid species representing all eight families of Notothenioidei to elucidate the mechanism and evolutionary history of this mtDNA rearrangement. Species of the three basal non-Antarctic families have the canonical vertebrate mt gene order, whereas species of all five Antarctic families have a rearranged CR bearing the embedded *ND6* ($ND6_{CR}$) and $tRNA^{Glu}$, with additional copies of $tRNA^{Thr}$, $tRNA^{Pro}$, and noncoding region in various lineages. We hypothesized that an initial duplication of the canonical mt region from *ND6* through CR occurred in the common ancestor to the Antarctic clade, and we deduced the succession of loss or modification of the duplicated region leading to the extant patterns of mt DNA reorganization that is consistent with notothenioid evolutionary history. We verified that the $ND6_{CR}$ gene in Antarctic notothenioids is transcribed and therefore functional. However, $ND6_{CR}$-encoded protein sequences differ substantially from basal non-Antarctic notothenioid ND6, and we detected lineage-specific positive selection on the branch leading to the Antarctic clade of $ND6_{CR}$ under the branch-site model. Collectively, the novel mt $ND6_{CR}$ genotype of the Antarctic radiation represents another major molecular change in Antarctic notothenioid evolution and may reflect an adaptive change conducive to the functioning of the protein (Complex I) machinery of mt respiration in the polar environment, driven by the advent of freezing, oxygen-rich conditions in the Southern Ocean.

Key words: H-strand duplication, novel ND6, control region reorganization, adaptive mitochondrial evolution.

Research article

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

Introduction

The teleost suborder Notothenioidei (order Perciformes) consists of eight recognized families encompassing 129 species (*sensu* Eastman [2005]). Three small basal families—Bovichtidae (11 species) and the monotypic Pseudaphritidae and Eleginopidae are non-Antarctic, having diverged before the isolation and glaciation of Antarctica. The other five families—Nototheniidae, Harpagiferidae, Artedidraconidae, Bathydraconidae, and Channichthyidae diversified within the isolated, frigid Southern Ocean, comprising an adaptive radiation. The majority of species (100) in these five families are endemic to the freezing Antarctic waters, and about 16 species have secondarily entered non-Antarctic waters over evolutionary time (Cheng et al. 2003; Eastman 2005). Notothenioidei thus provides an unparalleled evolutionary series of related species for investigations of genotypic changes that accompanied environmental change (Chen et al. 2008). Antarctic notothenioid evolution within the oceanographically isolated, frigid Southern Ocean had led to remarkable molecular changes, most notably in the gain of the novel antifreeze glycoprotein gene providing the ice-binding protein that prevents inoculative freezing by environmental ice crystals (DeVries 1971; Chen et al. 1997). Living in chronically cold and thus oxygen-rich waters had also led to evolutionary genetic loss, most remarkably in the loss of hemoglobin and red blood cells in the derived icefish family (Channichthyidae) (Cocca et al. 1995), as well as myoglobin loss in six icefish members (Sidell and O'Brien 2006).

Recently, another example of gene loss was reported—the mitochondrial (mt) genomes of Antarctic notothenioids apparently lack the genes encoding NADH dehydrogenase subunit 6 (*ND6*) and the adjacent $tRNA^{Glu}$, whereas the basal non-Antarctic notothenioid species have the canonical mt genome (Papetti et al. 2007). The gene content of vertebrate mt genome is generally fixed, containing a highly conserved set of 37 genes, encoding 2 ribosomal RNAs (*rRNAs*), 22 transfer RNAs (*tRNAs*), and 13 proteins that are essential in mt respiration and adenosine triphosphate (ATP) production (Wolstenholme 1992). Putative loss of *ND6* gene has not been reported in any animal mt genome except Antarctic notothenioids and apparently without nuclear compensation for this loss (Papetti et al. 2007). ND6 is an indispensable subunit of Complex I (NADH–quinone oxidoreductase) of

© The Author 2010. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

the mt electron transport chain, as even single amino acid mutations in ND6 can abolish the assembly of Complex I or cause disease conditions in human (Bai and Attardi 1998; Chinnery et al. 2001). Without ND6 protein and properly assembled Complex I, mt electron transport and ATP synthesis would be greatly impaired, rendering organismal survival tenuous or impossible. Thus, it is difficult to reconcile the absence of an *ND6* gene with Antarctic notothenioids being thriving species and studies that show mt respiration proceeds in these fish (Weinstein and Somero 1998; Hardewig et al. 1999; Urschel and O'Brien 2009).

The vertebrate mt gene order was generally considered conservative, but with increasing number of mt genomes sequenced and characterized, deviations from the canonical order have been identified in many chordate and vertebrate lineages, including amphioxus (Boore et al. 1999), lampreys (Lee and Kocher 1995), bony fishes (Inoue et al. 2001), amphibians (Macey, Larson, Ananjeva, Fang, et al. 1997; Mueller and Boore 2005; Kurabayashi et al. 2008), reptiles (Amer and Kumazawa 2007), birds (Mindell et al.1998), and mammals (Janke et al. 1994). Hence, the absence of an mt gene at its normal position in the typical vertebrate mt gene order does not necessarily indicate gene loss. In characterizing mt control region (CR) of Antarctic notothenioids, we discovered that the "missing" *ND6* gene and the adjacent *tRNA^{Glu}* are not lost but have become translocated within the mt CR from their canonical location between NADH dehydrogenase subunit 5 (*ND5*) and cytochrome b (*Cytb*) gene. We verified that the translocated *ND6* gene is transcribed and thus likely produces a functional protein. We analyzed the structural organization of the CR and rearranged CR from a large number of species representing all eight notothenioid families and deduced the molecular mechanism leading to the translocation of *ND6*/*tRNA^{Glu}* to the CR and the evolutionary process of the observed CR rearrangements in the Antarctic families. Additionally, we tested for presence of lineage-specific positive selection on the CR-embedded *ND6* to assess if the evolution of the genotype was of an adaptive nature.

Materials and Methods

Specimen and Tissue Collection

Notothenioid species representing all eight recognized families were collected using various methods in Southern Ocean and non-Antarctic habitats. Species from the three basal, non-Antarctic families include *Bovichtus variegatus* (Bovichtidae) from Otago Harbor, New Zealand, *Pseudaphritis urvillii* (Pseudaphritidae) from Onkaparinga River, South Australia, and *Eleginops maclovinus* (Eleginopidae) from Puerto Natales, Chile. Nineteen species from the five Antarctic families include the following: Nototheniidae—*Trematomus bernacchii*, *T. newnesi*, *Pagothenia borchgrevinki*, and *Pleuragramma antarcticum* from McMurdo Sound, *Notothenia coriiceps*, *N. rossii*, *T. eulepidotus*, and *Lepidonotothen squamifrons* from Antarctic Peninsula waters, and two secondarily cool-temperate nototheniid species *N. angustata* and *N. microlepidota* from Otago Harbor, New Zealand; Harpagiferidae—*Harpagifer antarcticus* from South Georgia; Artedidraconidae—*Histiodraco velifer* from McMurdo Sound and *Pogonophryne cerebropogon* and *P. scotti* from the Ross Sea; Bathydraconidae—*Racovitzia glacialis* from the Ross Sea; and Channichthyidae—*Chaenocephalus aceratus* and *Chionodraco rastrospinosus* from Antarctic Peninsula, *Chionodraco myersi* from the Ross Sea, and the secondarily cool-temperate channichthyid *Champsocephalus esox* from the Falklands. Tissues from fish specimens were flash frozen in liquid nitrogen and stored at −80 °C until use.

DNA Extraction, mtDNA Amplification, Subcloning, and Sequencing

Total genomic DNA was isolated from tissues (mostly liver or spleen) using standard phenol–chloroform extraction and ethanol precipitation methods. Thirteen primers were designed (sequences and amplicons detailed in supplementary table S1.I [Supplementary Material online]) and used in various combinations to amplify by polymerase chain reaction (PCR) the complete or partial mt CR from 22 notothenioid species representing all eight families of Notothenioidei. The primer pair Noto_Cytb_F and Pa_12S_R (supplementary table S1.I, Supplementary Material online), designed based on conserved sites in notothenioid *Cytb* and *12S rRNA* gene (flanking the CR) sequences available in the database, was first used to PCR amplify the entire CR from the Antarctic species. Only one species, the nototheniid *P. antarcticum*, yielded a clear single product, confirmed to contain the CR region upon sequencing. After obtaining the *P. antarcticum* CR sequence and discovering *ND6* and *tRNA^{Glu}* genes to be embedded within the CR (see Results and fig. 1), we designed a primer Noto_Glu_R and its complement Noto_Glu_F (supplementary table S1.I, Supplementary Material online) to conserved sequence in the *tRNA^{Glu}* gene, to PCR amplify the entire CR in two reactions from eight select species representing all five Antarctic families. Noto_Cytb_F was paired with Noto_Glu_R, and Noto_Glu_F with Noto_12S_R designed to another *12S rRNA* sequence site (supplementary table S1.I, Supplementary Material online) were used to amplify the 5′ and 3′ segment of the CR, respectively, and the PCR products were sequenced. To obtain the actual sequence that spans the Noto_Glu_R/F primer site, a species-specific primer 3′ to the site (supplementary table S1.I, Supplementary Material online) was designed for each species after its CR sequence was obtained and paired with Noto_Cytb_F to amplify the 5′ segment of the CR inclusive of the *tRNA^{Glu}* gene. The primer pair Noto_Cytb_F and Noto_Glu_R was also used to successfully amplify the 5′ segment of the CR that contains the embedded *ND6* from an additional 10 species across Notothenioidei (supplementary table S1.I, Supplementary Material online).

Three other primers (supplementary table S1.II, Supplementary Material online) were designed to amplify the mtDNA region between *ND5* and *Cytb* genes (the typical location of *ND6* gene in vertebrate animals) from four of the nine Antarctic species whose complete CRs were

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

sequenced in this study. Sequences of this region in the remaining species were reported by Papetti et al. (2007) and available in GenBank (accession numbers DQ526430, DQ526431, DQ526437, EF538671, and EF538675). For *P. antarcticum*, complete Cytb gene was also amplified using the primer pair Pa_Cytb_F and Pa_Cytb_R (supplementary table S1.II, Supplementary Material online); thus, the mtDNA sequence spanning 3′ end of *ND5* through 5′ end of *12S rRNA* was obtained for this species.

For the basal non-Antarctic species *B. variegatus* (Bovichtidae), *P. urvillii* (Pseudaphritidae), and *E. maclovinus* (Eleginopidae), their complete CR and canonical *ND6* gene were also amplified (except *B. variegatus ND6*, which is available in GenBank). *Bovichtus variegatus* CR was amplified with primer pair Bv_ Cytb_F/Noto_12S_R. *P. urvillii* CR and *ND6* were amplified with Noto_Glu_F/Noto_12S_R and PaNcPu_3tRNA_F/Noto_Glu_R, respectively. *Eleginops maclovinus* CR and *ND6* were amplified with Em_Cytb_F/Em_12S_R and Em_ND6_F/Em_ND6_R, respectively. Primers are shown in supplementary table S1.I–III (Supplementary Material online).

PCR amplifications were carried out using PTC-200 thermocycler (MJ Research), in reaction volumes of 50 μl containing 1 μg of genomic DNA, 0.2 mM dNTPs, 0.2 μM each primer, 2.0 mM $MgCl_2$, 5.0 μl 10× reaction buffer, and 2 U *Taq* polymerase, using the following cycling parameters: 94 °C initial denaturation for 3 min, 35 cycles of 94 °C denaturation for 55 s, 54 °C annealing for 55 s and 72 °C elongation for 1–4 min, and a final extension at 72 °C for 7 min. Purified PCR products were either directly sequenced with the PCR primers or sequenced after cloning into the pGem-T_{easy} vector (Promega). Sequencing reactions were performed using BigDye v.3.1 (Applied Biosystems) and read on an ABI3730*xl* automated sequencer at University of Illinois Keck Center for Comparative and Functional Genomics. Sequences were edited and assembled using ChromasPro v.1.42 (Technelysium). Alignments of CR, *ND6*, and other nucleotide sequences and *in silico* translated ND6 amino acid sequences were made with ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (gap parameters set to default: open = 10, extend = 0.05) and minor manual improvement. Sequence similarity scores (%) were calculated by ClustalW2 with the above setting.

RNA Extraction and Amplification of *ND6* Transcripts

To assess if the CR-embedded *ND6* gene (hereon called $ND6_{CR}$) is functional, seven select species representing all five Antarctic notothenioid families were tested for expression of $ND6_{CR}$ transcript by reverse transcription (RT)–PCR amplification. The basal non-Antarctic species *E. maclovinus* was included as a positive control. Total RNA was isolated from tissues (mostly liver, spleen, or gill) using the Ultraspec RNA isolation reagent (Biotecx) and treated with RNase-free DNase (Promega) to remove potential mtDNA contamination. For first-strand cDNA synthesis, the RNA from each species was primed with a species-specific reverse primer (supplementary table S1.III, Supplementary Material online) designed to the 3′ terminus beginning with the stop codon of the respective *ND6* coding sequence obtained in this study. This species specificity was necessary because we found considerable sequence divergence among *ND6* genes (see Results) precluding a single reliable degenerate 3′ primer applicable to all species. For the same reason, species-specific 5′ primer was designed for four species, and a degenerate primer Noto_ND6_R (supplementary table S1.III, Supplementary Material online) was found to be applicable to the remaining four Antarctic species. The 5′ primers were designed to begin at the translation start. About 2–5 μg of total RNA from each species was reverse transcribed using SuperScript III (Invitrogen) in a 20-μl volume following manufacturer instructions. One-tenth (2 μl) of the RT reaction was then amplified by PCR with the appropriate primer pair to obtain the full-length *ND6* cDNA. PCR amplification conditions were as described above except for shortening the elongation step to 45 s per cycle. Two negative controls, PCR reagents without first-strand cDNA, and with one-tenth the amount of RNA used in the RT reaction (to validate absence of mtDNA contamination), were performed in parallel for each species. The RT-PCR product of each species was sequenced to verify that they are indeed *ND6* cDNA.

Tests for Positive Selection on Notothenioid *ND6*

A total of 22 full-length *ND6* and $ND6_{CR}$ gene sequences were aligned with codon constraint and used in Bayesian (MrBayes 3.1.2) and maximum likelihood (ML; PHYML 2.4.4) phylogenetic analyses, implementing the evolutionary models GTR + I + G, and HKY + I + G and GTR + I + G (closest best-fit models selected by Modeltest version 3.8 (Posada and Buckley 2004) under bayesian information criterion and corrected akaike infromation criterion that are accepted by MrBayes and PHYML), respectively. The trees from all analyses have identical topology (representative ML tree shown in fig. 4), which was used as input to test for presence of positive selection on the $ND6_{CR}$ lineage. The nonsynonymous/synonymous substitution rate ratio ($\omega = d_N/d_S$) tests for positive selection were carried out using modified branch-site Model A (Zhang et al. 2005) using CODEML in the PAML package 4.3. The branch leading to the Antarctic clade of $ND6_{CR}$ was marked as foreground and the three non-Antarctic lineages as background. Sites within the foreground lineage were tested for positive selection. Branch-site Model A assumes four-site classes: class 0, conserved sites throughout the tree ($0 < \omega < 1$); class 1, neutral sites throughout the tree ($\omega = 1$); and class 2a and 2b, sites experiencing positive selection in the foreground, but are conserved or neutral in the background (Zhang et al. 2005). Likelihood ratio test (LRT) was carried out to test for the statistical significance of the log-likelihood difference between Model A and the null model where ω is fixed as 1. Sites under positive selection in the foreground lineage

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

```
     →ND5                                                                              non-coding spacer
   1 CAGGAACTGTACCCCGAAGTGATTAAC----/438/---CTTGTCTCATCAATGGTTATATTTTAACGGGCAACCCCACAACACTAACAAGCTCCTGACAGACCCTGTCC
                      ├→ Cytb
 837 ACGAGCCCCCCCCCTAATGGCAAGCCTTCGAAAGACGCACCCCC----/1052/----ATTTTCCTCCTCTTCTTCCCCCTGGCGGGATGACTGGAAAATAAGGTCCTGG
           ├→ tRNA-Thr1
1685 ATTTGGCATGCGCTAGTAGCTCAGTCCTAGAGCGCCGGTCTTGTAAACCGGAGGCCGGAAGTTAGATCCCTCCCTACCGCTCAAAGAGAGGAGACTCTAACTCCCGCCCC
                             tRNA-Pro1←┤├→ CR1                       TAS   cTAS
1795 CAACTCCCAAAGCTGGAGTTCTAAGCTAAACTACTCTGAGGCGCCCCCCAAAAATGTACTTTTAAGTACATTATGTATTATCAACACTCATTTATATTAACCACTTAATG

1905 GGCATTCGTGGACAGGATTTGATTTTAGGACAAAATTTGACTCGACACACAAATATATCAGAATTAACGGGTGATATACGCAAAGCATTTGAGAAGCCTCACATAATCCA
                                                                                                               CSB-F
2015 ATTTAATGACAGGCGAAATTTAAGACCGAACTCTAACCTCATTGGTTAAGTTATACCTTTATCCAACTTCCTTGCAGTTTACAGATTCTTAATGTAGTAAGAGCCGACCA
                          CSB-E                                                            CSB-D
2125 ACAAGCTCATATCTTAAGGCTCACGGTTATTGAGGGTGAGGGACAAGTAATGTGGGGGTTTCACATGGTGAACTATTCCTGGCATTTGGTTGCTACTTCAGGGCCAATGA

2235 TTGGTGTCATCCCCCGCACTTTTACCGACGCTTACATAAGTTAATGTTTGTATTACATACTCCTCGTTACCCAGCAAGCCGGGCGTTCACTCCAGCGAGCCAGGGGTTCC

2345 TTTTTTTTTTTCCTTTCCCCTTGCATTTCAGAGTGCGCGCGGTTTTAACTAACAAGCGTGAGCACTTTTCTTGCTTGAGAGAGAATAGTCTGATCTATATTAAGACTCCG
                              CSB1                                                                          CSB2
2455 CTTGCGGTTTTAATTTTTTTTTTTTCTAGGGCATAAAGTACCTAACCTCTCACCTGAGTATTATAAGAGATATTTTAACGCCTCTTAGTCGAGAACCCCCCCCCACCCCC

2565 TTACTTCCCTGAAATCCATAAAGACTACAAGAGACACTACGAACAAGAACGCCCTTACAACCCCGTTTGGAGCATACTTCTTAACAAGTGAACAAGTGACTGGTGACATT

2675 TACAAAACCTCTCACCAGTTTTACAAACCCTTTGCACTACCACAACTCTTGTGTAACTTACACCCCATTCTCCCAGGTTTGTAGGTCTGAGCACCTGCCCGCAATCTAAA

2785 GAATCCGCTCTATACATGTACTAATTTAGACCCATCCTTAAACACTTAAATTTCACAGCTTCCTAACACCCGCTTGTTGTTTAAGAAAGCCCCGACCTCTTAAACACGTC

2895 CCTCCCGAGCGCCTGGGCCCCTATCCCCGCCTGTTTAATAACAAACCCCAGCATAATCTCCTTTTGAATGGGCGTACCGGCCCCCGGCTAGGACCCCGAGACACCTCCAG
                                                                  *  K  F  P  R  V  P  G  R  S  P  G  R  S  V  E  L

3005 CGCCACATAAAGGGCAATCAACAGGGCCCATCCACACGTCACTACTAGGAACCCCCCGCCCCCGTAGGCTTCTGAGACCCCTGCGGCCTCCCCCTGTATTACAGCTATCT
       A  V  Y  L  A  I  L  L  A  W  G  C  T  V  V  L  F  G  G  G  G  Y  A  E  S  V  G  A  A  E  G  Q  M  V  A  M

3115 CCCACTCTCCCCCTGCAAACCACCAGAAAGCGGGGCCTTCTGAACTTCCTTGTAAAAAGCCTACAACACCAGTCACCCCTAGATACCCGGCTATACACTTCAACACCGAC
     E  W  E  G  G  A  G  W  W  F  A  P  G  E  S  S  G  Q  L  F  G  V  V  G  T  V  G  L  Y  G  A  M  C  K  L  V  S

3225 TCCTCCCCCAAACCTGTTGAAGGAGCACCAGCACACAATGCTACTGAGTAAGCAAATACAACCAACATCCCTCCAAGATAGATTAAGAACAGTACTAGACATGAAAAAGT
      E  E  G  L  G  T  S  P  A  G  A  C  L  A  V  S  Y  A  F  V  V  L  M  G  G  L  Y  I  L  F  L  V  L  C  S  F  T

3335 ACCCCCCAGAGACAAGATCAACCCACACCCCATTGCTGCGACTACAACCAACCCAAGAGCCGCATAAAAAGGCGGAGGATTAGAGGCGACCGCCGCTATTCCTACTACTA
       G  G  L  S  L  I  L  G  C  G  M  A  A  V  V  V  L  G  L  A  A  Y  F  P  P  P  N  S  A  V  A  A  M  G  V  V
                        ND6CR←┤                                                                        tRNA-Glu←┤
3445 TACCCAATATAAATAAATATCCTCAATACAACACAATTCTTGCCAGGATTTTAACCAGGACTAACGGCATGAAAAGCCATCGTTGTTATTCAACTACAAGAACCCTTGTA
     M  G  L  M  F  L  Y  G  W  Y  L  M
                                                                  ├→ tRNA-Thr2
3555 AAAAGCCTCTTCTTCCCCCTGGCGGGATGACTGGAAAATAAGGTCCTGGATTTGGCATGCGCTAGTAGCTCAGTCCTAGAGCGCCGGTCTTGTAAACCGGAGGCCGGAAG
                ├→ tRNA-Pro2                                                                        ├→ CR2
3665 TTAGATCCCTCCCTACCGCTCAAAGAGAGGAGACTCTAACTCCCGCCCCCAACTCCCAAAGCTGGAGTTCTAAGCTAAACTACTCTGAGGCGCCCCCCAAAAATGTACTT
          TAS   cTAS
3775 TTAAGTACATTATGTATTATCAACACTCATTTATATTAACCACTTAATGGGCATTCGTGGACGGGATTTGATTTTAGGACAAAATTTGACTCGACACACAAATATATCAG

3885 AATTAACGGGTGATATACGCAAAGCATTTGAGAAGCCTCACATAATCCAATTTAATGACAGGCGAAATTTAAGACCGAACTCTAACCTCATTGGTTAAGTTATACCTTTA
                                           CSB-F                                    CSB-E
3995 TCCAACTTCCTTGCAGTTTACAGATTCTTAATGTAGTAAGAGCCGACCAACAAGCTCATATCTTAAGGCTCACGGTTATTGAGGGTGAGGGACAAGTAATGCGGGGGTTT
                      CSB-D
4105 CACATGGTGAACTATTCCTGGCATTTGGTTCCTACTTCAGGGCCAATGATTGGTGTCATCCCCCGCACTTTTACCGACGCTTACATAAGTTAATGTTTGTATTACATACT

4215 CCTCGTTACCCAGCAGGCCGGGCGTTCACTCCAGCGAGCCAGGGGCTCTCTTTTTTTTTTTCCTTTCTCACTTGCATTTACAGAGTGCGCGCGGTTTTAACTAACAAGCG
                                                                                                    CSB1
4325 TGAGCACTTTTCTTGGCTTGAGAGAGAATAGTCTGATCTATATTAAGACTCCGACTTGCGGTTTTAATTTTTTTTTTTTTTCTGGGGCATAAAGTACCTAACCTCTCACC
                                                      CSB2
4435 TGAGTATTATAAGAGATATTTTAACGCCCTTTAGTCGAGAACCCCCCCCCACCCCCTTACTTCCCTGAAATCCATAAAGACTACAAGAGACACTACGAACAAGAACGCCC

4545 TTACAACCCCGTTTGGAGCATACTTCTTAACAAGTGAACAAGTGACTGGTGACATTTACAAAACCTCTCACCAGTTTTACAAACCCTTGTGTTGCAGGGATTAGACGTCC

4655 ATTAAACCAGAGCCTACTCTAATCACCACTTGACAAGTATTCCAAACTGCTTAAGCCACAAGTACACATTAAAAACTTGTGTAGCTCTTGAACACTCCCTTGGAGCCATG

4765 CCCGAAGTAATAAAGGTTGCTCTATACATCTATTAATTTGTACACTTCCCCAAACACCTCAAATTTCCACTTCTTTGCAACCCTTCCCTGTTACCCCCGTAAGGGGCCTC

4875 CCGGGGGTAAACTAGTTGACGTCACCGTATACATATACACACTTTGATACCTTCCACACTCAGGTTTTAAACCACCTCCCCCACCTCTATAAGAGGCCCCAACCCTTCTT
                                                                                                      ├→ tRNA-Phe
4985 GTTCAGGGGTCCCCCGAAATTTGTAGATATATACAGCTGTTATGTTGTACACACCCTTGAACACCTCATCCTTATTGGCCCCTTAAATGCCGTAGCTAGCGTGGTTTATT
                                                           ├→ 12SrRNA
5095 TAAAACGTAACACTGAAAATGTTGAAATGGGCCCTAGAAAGCTCCGCAAGCACAAAGGCTTGGTCCTGACTTTAT---/486/---CCCAAAGGACTTGGCGGTGCTTTA
```


Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

(Antarctic notothenioid clade) with significant LRTs were identified under Bayes empirical Bayes.

Results

Antarctic Notothenioid *ND6* and *tRNA^{Glu}* Genes "Lost" and Found

We obtained the partial mt genome sequence (5,955 nt) from the Antarctic nototheniid *P. antarcticum*, spanning from the 3′ end of *ND5* through the 5′ end of *12S rRNA* and discovered that the previously reported "missing" *ND6* gene and the adjacent *tRNA^{Glu}* (Papetti et al. 2007) were not missing but have become embedded in between two copies of mt CR sequences (fig. 1). The CR-embedded *ND6* and *tRNA^{Glu}* (hereon called *ND6$_{CR}$/tRNA^{Glu}*) appears to be intact genes. *ND6$_{CR}$* has a 525-nt open-reading frame with correct translation start and stop and encodes a 174-residue protein (fig. 1) of high BLAST similarity (*e* value = 2×10^{-26}) to teleost ND6 sequences. The *ND6$_{CR}$/tRNA^{Glu}* genes are centrally located in the modified CR, flanked on each side by an apparently complete copy of CR sequence, as each contains a full set of control sequences and conserved sequence blocks (termination-associated sequences [TASs], complementary termination-associated sequence [cTAS], extended termination-associated sequence [ETAS], and the conserved sequence blocks [CSBs]) (fig. 1). Also present are two copies each of the adjacent *tRNA^{Thr}* and *tRNA^{Pro}*, one at their canonical position right next to *Cytb* and the other immediately downstream from *ND6$_{CR}$/tRNA^{Glu}* (fig. 1). The residual sequence in the canonical location of *ND6/tRNA^{Glu}*, between *ND5* and *Cytb* genes, is a short noncoding spacer (60 nt for *P. antarcticum*) (gray block, fig. 1), confirming the same feature reported for various Antarctic notothenioid species by Papetti et al. (2007). To determine if these mt DNA rearrangements are unique to *P. antarcticum* or a shared character among Antarctic notothenioids, we obtained the complete CR sequences for eight other species representing all five Antarctic families, as well as the noncoding spacer sequences between *ND5* and *Cytb* for four of these species which were not examined by Papetti et al. (2007). For more comprehensive taxon representation, we additionally amplified and sequenced partial CRs from 10 other species. In all 18 species, we found that a complete *ND6* sequence and the adjacent *tRNA^{Glu}* gene to be present in their respective mt CR, similar to *P. antarcticum* (CR and *ND6$_{CR}$* sequence alignments in supplementary figs. S1and S2, Supplementary Material online, respectively; details in subsequent Results sections). For comparison, we also obtained the complete CR and *ND6* sequences of *B. variegatus* (Bovichtidae), *P. urvillii* (Pseudaphritidae), and *E. maclovinus* (Eleginopidae) representing all three basal non-Antarctic notothenioid families (fig. 2; supplementary figs. S1 and S2, Supplementary Material online) and confirmed that they have the canonical vertebrate gene order (*ND5_ND6_Cytb_CR_12S*) for this mt region (fig. 2) and that their respective CR is entirely noncoding sequence. The *E. maclovinus ND6* sequence we obtained is 100% identical to that reported by Papetti et al. (2007), whereas the CR sequence is 99% identical to the partial CR they obtained. The schematic representations of the canonical or rearranged *ND6* plus flanking regions of all 22 species along with a notothenioid phylogeny are shown in figure 2.

CR Gene Organization in Antarctic Notothenioids

Geneic and structural variations are present in the respective rearranged CR of the nine Antarctic notothenioids (*P. antarcticum*, *N. angustata*, *N. coriiceps*, *H. antarcticus*, *P. scotti*, *R. glacialis*, *C. aceratus*, *C. myersi*, and *C. rastrospinosus*) for which complete CR sequences were obtained (fig. 2). The total length of the region (3′ end of *Cytb* to 5′ end of *12S rRNA*) varies substantially, from 2,064 to 3,453 bp. All contain a single copy of *ND6$_{CR}$* and *tRNA^{Glu}* gene, but vary in the copy number of CRs and other *tRNA* genes, and in the lengths of CRs and other small noncoding regions (fig. 2). Four geneic/structural patterns are observed from these variations: 1) in the nototheniids *P. antarcticum* and *N. angustata*—two apparently complete CRs and two copies of *tRNA^{Thr}* and *tRNA^{Pro}*; 2) in the nototheniid *N. coriiceps* and the harpagiferid *H. antarcticus*—similar to pattern I except only one copy each of *tRNA^{Thr}* and *tRNA^{Pro}*. *Harpagifer antarcticus* also differs slightly in that its 3′ copy of CR appears incomplete as it lacks TASs; 3) in the artedidraconid *P. scotti* and the three channichthyids *C. aceratus*, *C. myersi*, and *C. rastrospinosus*—similar to pattern II but only a single complete CR, whereas the region corresponding to the 5′ CR in pattern II is reduced to a short noncoding sequence; and 4) in the bathydraconid *R. glacialis*—similar to pattern I except only one copy of *tRNA^{Thr}*, and a sizable noncoding region (367 nt) intercalates between the 5′ *tRNA^{Thr}* and *tRNA^{Pro}*. Nucleotide sequence similarities of the duplicated copies of CR and *tRNA* genes where present are very high, from 95% to 99%. The annotated alignment of

←
Fig. 1. L-strand nucleotide sequence of *Pleuragramma antarcticum* partial mt genome (5955 nt), from the 3′ end of *ND5* through the 5′ end of *12S rRNA*. Genes and noncoding regions occur in the following order: partial *ND5*, an intergenic noncoding spacer, *Cytb*, two tandem copies of CR_*tRNA^{Thr}*_*tRNA^{Pro}* with the translocated *ND6$_{CR}$* and adjacent *tRNA^{Glu}* in between, *tRNA^{Phe}*, and partial *12S rRNA*. Numbers within slash marks indicate nucleotides not shown. The start and end of each gene are indicated by vertical lines. Arrow indicates direction of transcription of each gene. Multiple gene copies resulting from duplication are named numerically. The noncoding spacer that replaced the canonical *ND6* first identified by Papetti et al. (2007) is boxed in gray. Unannotated sequence segments are noncoding regions of unknown function. *In silico* translation of the CR-embedded *ND6$_{CR}$* gene shown in one-letter abbreviations uses vertebrate mt genetic code. The sense sequence of *ND6$_{CR}$* gene is on the H-strand; thus, the amino acid sequence is derived from the complementary sequence of the L-strand shown. Asterisk indicates stop codon of *ND6$_{CR}$*. Within the two copies of CR, TAS and CSB motifs are boxed. The two underlined segments share high (99%) identity, indicative of a tandem duplication underlying the rearrangement in this mt region.

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

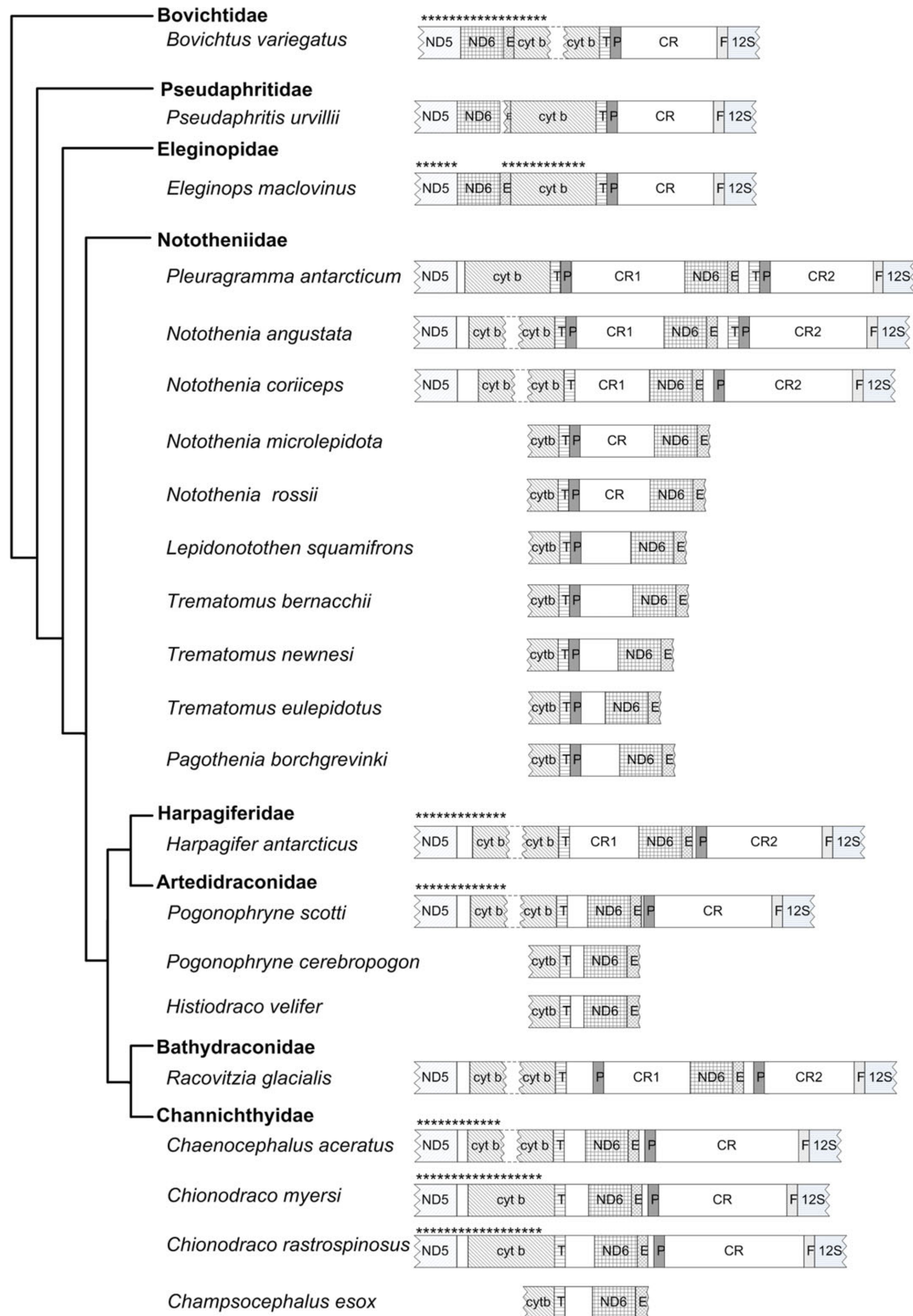

FIG. 2. Organization of notothenioid canonical and rearranged mt regions in relation to known notothenioid familial phylogeny. The lengths of the boxed genes approximately scale to actual gene sequence lengths. Transfer RNA genes are shown as single-letter amino acid codes. Boxes representing coding genes are shaded with different textured patterns and boxes representing noncoding sequences including CR and intergenic spacers have blank background. Sequences of regions indicated with asterisks above are curated from GenBank database (accession numbers [from top to bottom] are EF538674, DQ526429, EF538675, EF538671, EF538679, DQ526430, and DQ526431). Phylogenetic relationships of notothenioid families follow published phylogeny (Near et al. 2004).

complete CR sequences from all examined species is shown in supplementary figure S1 (Supplementary Material online). Whether single copy or in duplicates, the complete Antarctic notothenioid CRs are deemed complete because they contain a full set of conserved regulatory sequence motifs (TAS, cTAS, ETAS, and the CSBs), similar to the CR of the three basal non-Antarctic notothenioids (supplementary fig. S1, Supplementary Material online).

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

For the 10 species for which the CR was partially characterized, the embedded $ND6_{CR}/tRNA^{Glu}$ is an invariant feature (fig. 2). A complete 5′ copy of CR is found in two other *Notothenia* species, *N. microlepidota* and *N. rossii*, similar to their congeners and the confamilial *P. antarcticum*. In all other species, the 5′ CR region is reduced to shorter noncoding sequences, suggesting that a complete CR would be found in the 3′ copy downstream from $ND6_{CR}/tRNA^{Glu}$ but this awaits verification. One copy each of $tRNA^{Thr}$ and $tRNA^{Pro}$ are found in their canonical position next to *Cytb* in the nototheniids, apparently a predominant arrangement (9 of 10 species) in Nototheniidae (fig. 2). In the artedidraconids *P. scotti* and *P. cerebropogon* and the icefish *C. esox*, $tRNA^{Pro}$ is missing at its canonical position immediately adjacent to $tRNA^{Thr}$, apparently the predominant arrangement (all species) in the remaining four Antarctic families based on the species we examined (fig. 2).

Comparison of $ND6_{CR}$ and Canonical *ND6* in Notothenioids

The alignment of the nucleotide sequences and *in silico* protein translations of the $ND6_{CR}$ gene of the 19 representative species from the five Antarctic families and of the canonical *ND6* genes from the three basal non-Antarctic species are shown in supplementary figure S2 (Supplementary Material online). All 19 $ND6_{CR}$ gene sequences have an intact opening reading frame with correct start and stop codons. The start codons (ATG, CTG, and GTG) and stop codons (TAG, AGG, and AGA) vary among species (supplementary fig. S2A, Supplementary Material online), but all follow the standard vertebrate mt genetic code. The length of the *ND6* genes is 525 nucleotides encoding a protein of 174 amino acids in 14 of the examined Antarctic notothenioids and the basal non-Antarctic *P. urvillii* and *E. maclovinus*, 522 nucleotides (173 amino acids) in Antarctic *L. squamifrons* and basal non-Antarctic *B. variegatus*, and 519 nucleotides (172 amino acids) in the remaining four Antarctic species, *P. borchgrevinki*, *T. bernacchii*, *T. eulepidotus*, and *T. newnesi* belonging to the nototheniid subfamily Trematominae (supplementary fig. S2A and B, Supplementary Material online). The gene and protein lengths of Antarctic notothenioid $ND6_{CR}$ are consistent with those of canonical *ND6* in most teleost fishes (522 nucleotides, 173 amino acids) we identified in GenBank entries, the rearranged *ND6* in the conger eel *Conger myriaster* (171 amino acids) (Inoue et al. 2001), and a number of Perciformes fishes as well as mammals (172 or 174 amino acids) in the database (Miya et al. 2003; Yamanoue et al. 2007). Antarctic notothenioid $ND6_{CR}$ sequences share high similarities (69%–96% in nucleotides, 67%–98% in amino acids) but differ substantially from the canonical *ND6* sequences of the basal non-Antarctic notothenioids *E. maclovinus*, *P. urvillii*, and *B. variegatus*, sharing much lower sequence similarities (33%–66% in nucleotides and 45%–57% in amino acids) (supplementary table S2, Supplementary Material online). Sequence divergence is most pronounced at the midsection of notothenioid $ND6_{CR}$ and *ND6* genes, whereas the 5′ and 3′ sections are relatively more conserved, with the 5′ section having longer stretches of nucleotide identities than the 3′ end (supplementary fig. S2, Supplementary Material online).

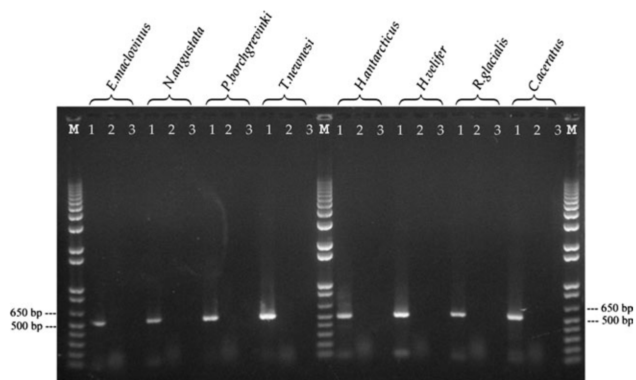

Fig. 3. RT-PCR detection of the $ND6_{CR}$ gene transcript in seven Antarctic notothenioid species and of *ND6* transcript in the non-Antarctic positive control *Eleginops maclovinus*. Amplified cDNA products were electrophoresis on 1% agarose gel. Lanes 1—all species show cDNA product of expected size for *ND6* (500+ bp). Lanes 2 and 3—parallel negative control PCR reactions; RNA as template (lane 2) and no template (lane 3). M = 1 kb plus DNA ladder (Invitrogen).

Antarctic Notothenioid $ND6_{CR}$ Genes Are Transcribed

Using species-specific and other appropriately designed primers (supplementary table S1.III, Supplementary Material online) for first-strand cDNA synthesis followed by PCR amplification, the seven select notothenioid species representing all five Antarctic families and the non-Antarctic notothenioid *E. maclovinus* as the basal positive control produced cDNA products of the expected size for full-length ND6 coding sequence (500-plus bp) (fig. 3). No amplified products were detected in the two parallel negative PCR controls for each species (lanes 2—RNA as template instead of first-strand cDNA; lanes 3—no template) (fig. 3); thus, these RT-PCR products are cDNAs derived from *ND6* and $ND6_{CR}$ mRNA and not from contaminating DNA. The RT-PCR products were sequenced, and the sequences were identical to the *ND6* or $ND6_{CR}$ coding sequence of the corresponding species we obtained from sequencing amplified mtDNA products in this study.

Positive Selection on Amino Acid Sites in $ND6_{CR}$

Signal of positive selection was detected on the branch leading to the $ND6_{CR}$ (foreground) of Antarctic notothenioid species (fig. 4) using modified branch-site Model A test (Zhang et al. 2005). Table 1 shows the parameter estimates and results of LRT of the test. The foreground ω2 was 4.8 ($\omega > 1$), indicating positive selection operating on branch leading to the Antarctic clade (fig. 4). The LRT statistic ($2\Delta lnL$) was 5.5, with a significant *P* value (0.0185); thus, the null hypothesis of no selection can be rejected. The proportion of sites in site classes 2a and 2b was small ($p2 = 0.16$), and only seven amino acids in $ND6_{CR}$ are shown to be under selection ($\omega > 1$) (table 1).

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

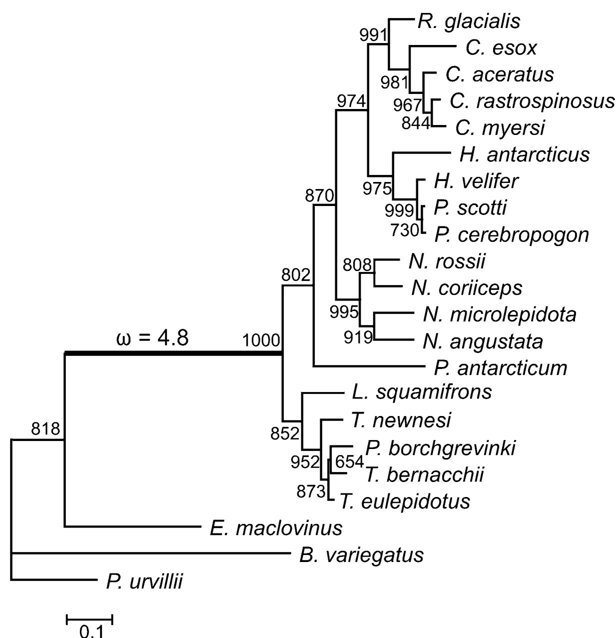

Fig. 4. ML tree of *ND6* and $ND6_{CR}$ genes from 22 notothenioid species, constructed using PHYML2.4.4 implementing the evolutionary model GTR+I+G. Node supports were evaluated with 1,000 bootstrap replicates. The tree topology was used as input tree for the Branch-site test for lineage-specific positive selection on the foreground branch (bold line) leading to the Antarctic clade, conducted in PAML 4.3. The value of foreground ω2 obtained from the test results is indicated.

Discussion

The gene order of the region spanning *ND5* through *12S rRNA* in the mt genome of *B. variegatus*, *P. urvillii*, and *E. maclovinus* representing each of the three basal, non-Antarctic notothenioid families (fig. 2; supplementary figs. S1 and S2, Supplementary Material online) that we reconstructed using sequences obtained in this study and published sequences (Papetti et al. 2007) establishes that the canonical vertebrate mt gene order is an ancestral condition in Notothenioidei. The gene order and sequences in the same region that we obtained for 19 notothenioid species representing all five Antarctic families (figs. 1 and 2; supplementary figs. S1 and S2, Supplementary Material online) establish that a rearranged mt gene order for *ND6*/$tRNA^{Glu}$ and the CR is a derived condition associated with the Antarctic notothenioid radiation. The reportedly "lost" *ND6* gene in Antarctic notothenioid fishes (Papetti et al. 2007) was not lost but had become embedded in the CR during notothenioid mt genome evolution. The $ND6_{CR}$ gene has intact gene structure (fig. 1; supplementary fig. S2, Supplementary Material online) and is transcribed (fig. 3) and expectedly would produce a functional protein. Our discovery of the functional $ND6_{CR}$ gene thus resolves the quandary of how *Antarctic* notothenioid fishes can be thriving species with demonstrated mt respiration (Weinstein and Somero 1998; Hardewig et al. 1999; Urschel and O'Brien 2009) with a "missing" *ND6*. Additionally, the Antarctic-specific mt rearrangement and translocation of *ND6* represent the newest addition to the list of dramatic genomic/molecular changes associated with Antarctic notothenioid evolution in the frigid Southern Ocean.

Possible Causes for Misdiagnosis of a "Missing" *ND6*

Papetti et al. (2007) inferred *ND6* gene is "missing" in Antarctic notothenioid mt genomes based on four main observations. The results in this study and other pertinent technical issues (detailed in supplementary file 1, Supplementary Material online) provide alternate explanations to their observations. 1) They found *ND6* absent at its canonical mt position, replaced by a short noncoding sequence. We confirmed this result; however, it was not due to *ND6* loss, but degeneration of the copy at the canonical *ND6* location after the *ND6*-through-CR duplication. 2) They found no *ND6* in the complete mt genome sequence of the icefish *C. rastrospinosus*. Their GenBank entry states that it is an incomplete mt genome sequence, and we found that it is missing the CR region where we discovered the embedded $ND6_{CR}$ and $tRNA^{Glu}$. 3) In mtDNA dot blot, they found no evidence of hybridization using *ND6* gene from the basal notothenioid *E. maclovinus* as a probe. We examined the primer pairs they used for amplifying mtDNA in several sections for use in the hybridization and deduced that the section spanning the rearranged CR with the embedded $ND6_{CR}$ and $tRNA^{Glu}$ was not amplified, thus, excluded from the hybridization. 4) Their RT-PCR amplifications for *ND6* transcripts produced no product. We examined the degenerate primers they used and deduced that the unsuccessful *ND6* cDNA amplification was due to the lack of primer specificity for the target sites in $ND6_{CR}$.

Mechanism of *ND6*/$tRNA^{Glu}$ and CR Rearrangement in Notothenioidei

Mt gene order rearrangements are commonly hypothesized to result from mt DNA duplication followed by random loss (Boore 2000). In this general model, a portion of

Table 1. Parameter Estimates and LRTs Statistic in the Branch of Antarctic Notothenioid $ND6_{CR}$ Genes.

Model	Parameter Estimates	*l* (lnL) Null Model	*l* (lnL) Alternative Model	2ΔlnL	Degrees of Freedom	*P* Value	Positive Selected Sites
Model A	ω0 = 0.096, p0 = 0.65 ω1 = 1.0, p1 = 0.19 ω2 = 4.8, p2 = 0.16	−5414.93	−5412.16	5.55	1	0.0185	74C, 80T, 89K, 105Q, 111P*, 124S, 126M**

*$P > 0.95$, **$P > 0.99$.

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

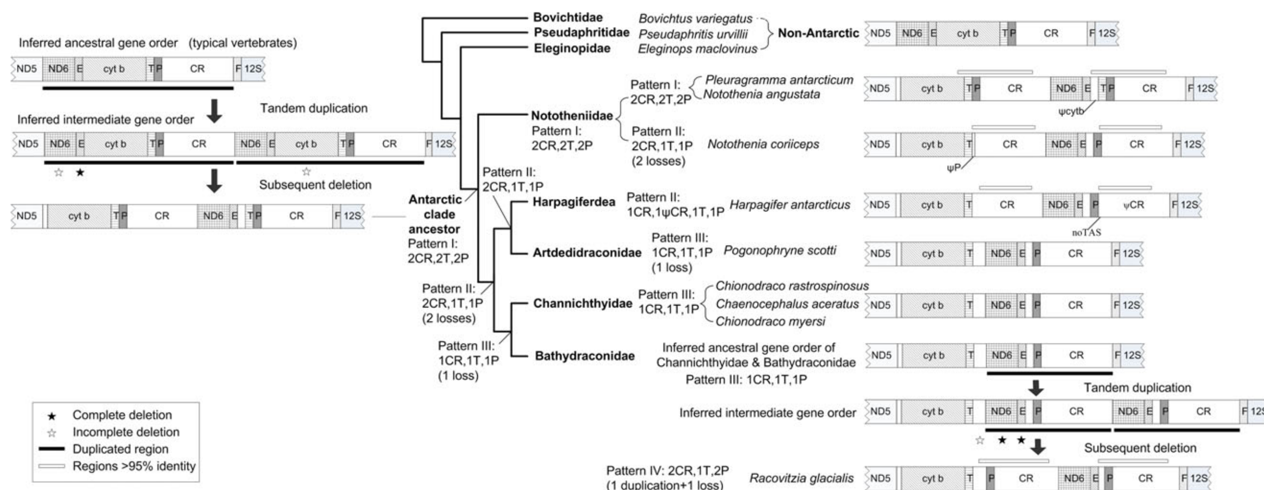

Fig. 5. Proposed mechanism of mt rearrangements in the region between *ND5* and *12S rRNA* in Antarctic notothenioids and the most parsimonious pathway leading to the four patterns of changes in this region that is consistent with notothenioid evolutionary history. A hypothesized initial duplication of the region followed by early loss/degradation of *ND6, tRNA^{Glu}*, and *Cytb* generated pattern I rearrangement in the common ancestor of the Antarctic clade, from which patterns II, III, and IV arose through successive loss/degradation of redundant CR, *tRNA^{Thr}*, and *tRNA^{Pro}* duplicates (abbreviated CR, T, P) and a secondary duplication in the bathydraconid family (see text for details). Tandem duplications are indicated by solid bars below the regions where they are inferred to occur. Regions with >95% nucleotide identity supporting the occurrence of tandem duplication are indicated with open bars. Genes/sequences that underwent complete deletion or partial deletion/degradation are indicated by solid and open stars, respectively.

the mt genome becomes duplicated; one copy of each duplicated gene subsequently loses function, degrades into a pseudogene, and/or becomes excised from the genome. Which gene copy is ultimately lost is randomly determined by the first loss-of-function mutation, and thus, a certain deletion pattern can restore the original gene order but others lead to rearrangement (Boore 2000). The extant mt gene order patterns of Antarctic notothenioids likely arose from such a process. We hypothesize that the rearranged order of *ND6*, adjacent *tRNAs*, and the CR in the Antarctic notothenioids resulted from an initial duplication of the mtDNA region encompassing *ND6_tRNA^{Glu}_Cytb_tRNA^{Thr}_tRNA^{Pro}*_CR, followed by successive degradations/deletions, ending with one copy of *ND6_tRNA^{Glu}* embedded in the CR (fig. 5).

Figure 5 depicts the most parsimonious hypothesis of the succession of evolutionary changes in the rearranged notothenioid mt region alongside a known familial phylogeny of Notothenioidei (Near et al. 2004). We posit that the ancestral Antarctic notothenioid mt genome had the typical vertebrate gene order as that of the basal non-Antarctic notothenioid species. A region-wide duplication occurred at some point in the evolutionary history of the Antarctic lineage before the divergence of the Antarctic families. The extent of the original duplication is difficult to estimate because boundaries of duplicated regions could be obscured by deletions (Mueller and Boore 2005). However, based on the extant rearrangement structures, we can reasonably infer *ND6* through CR as the minimum contiguous mt region that was duplicated, creating an intermediate of two tandem copies from which successive degradations and/or deletions of duplicated sequences ensued (fig. 5). Sequence deletion or degradation events shared by all extant Antarctic species likely occurred early, in a recent common ancestor to the Antarctic clade. In this category are the ancestral *ND6* and *tRNA^{Glu}* between *ND5* and *Cytb*, and the new *Cytb* duplicate, because they are degraded or absent in all examined representatives of the five Antarctic families (figs. 2 and 5). The ancestral canonical *ND6* was reduced to a small noncoding region of 60–236 bp in Antarctic species (figs. 2 and 5; Papetti et al. 2007). In some species, this residual sequence still bears up to 50% nucleotide identity to the corresponding segment of *ND6$_{CR}$* (results not shown), indicative of its *ND6* origin. Similarly, the new *Cytb* duplicate was reduced to a small noncoding region (*ψCytb*) in the rearranged CRs (figs. 2 and 5), with a recognizable *Cytb* origin based on sequence similarity in some species. For example, *ψCytb* is 53 bp in *P. antarcticum* and 22 bp in *N. angustata*, each sharing 100% identity with the corresponding 3′ end of the functional *Cytb* copy (fig. 5).

The most recent common ancestor of the Antarctic clade, thus, would have the following complement of CR and genes—two copies of CR, *tRNA^{Thr}*, and *tRNA^{Pro}* (abbreviated 2CR, 2T, and 2P) and one copy each of *ND6$_{CR}$* and *tRNA^{Glu}* in the order shown in figure 5, designated as rearrangement pattern I. The *ND6$_{CR}$* and *tRNA^{Glu}* genes are preserved in all Antarctic lineages that evolved during the subsequent adaptive radiation. For the redundant copy of CR, *tRNA^{Thr}*, and *tRNA^{Pro}*, the pattern of preservation/loss varies among species but appears consistent within each family, and the progression of loss or modification across family roughly follows notothenioid evolutionary history (figs. 2 and 5). Pattern I and three other patterns (II, III, and IV) of rearrangement are recognizable in the complete or near-complete mt region from *ND5* through *12S rRNA* we reconstructed for the nine species

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

representing all five Antarctic families (fig. 2). Each pattern potentially represents a different stage in the evolutionary progression of loss/degeneration in the Antarctic clade. *Pleuragramma antarcticum* and *N. angustata* of the basal Antarctic family Nototheniidae retain pattern I rearrangement (2CR, 2T, and 2P) of the hypothesized common Antarctic ancestor. Within Nototheniidae, further loss of one copy each of $tRNA^{Thr}$ and $tRNA^{Pro}$, represented by *N. coriiceps*, generates pattern II (2CR, 1T, 1P). The degradation of the 5′ $tRNA^{Pro}$ in *N. coriiceps* is visible as a 17-bp residue (ψP) (fig. 4) that shares 58% identity to the corresponding portion in the downstream functional $tRNA^{Pro}$. The same two losses ($tRNA^{Thr}$ and $tRNA^{Pro}$) occurred independently in the common ancestor to the other four Antarctic families—Harpagiferidae, Artedidraconidae, Channichthyidae, and Bathydraconidae. In Harpagiferidae, the single representative species *H. antarcticus* retains pattern II rearrangement with slight degeneration of the 3′ CR by the loss of the TAS blocks resulting in the set, 1CR, 1ψCR, 1T, and 1P. In Artedidraconidae, represented by *P. scotti*, drastic 5′ CR decay in the ancestral pattern II leads to the extant pattern III rearrangement (1CR, 1T, 1P). This 5′ CR loss is seen in two other artedidraconids (fig. 2), supporting its occurrence before the divergence of the family. The common ancestor to Channichthyidae and Bathydraconidae independently underwent the same 5′ CR loss and thus had pattern III rearrangement (1CR, 1T, 1P), which persists in the channichthyid lineage (figs. 2 and 5). The sister family to the channichthyids, Bathydraconidae, represented by a single species in this study, surprisingly has an increased gene/sequence set—2CR, 1T, 2P (pattern IV). We hypothesize that pattern IV resulted from a secondary duplication of the ancestral pattern III forming two tandem copies, followed by the loss of the 5′ copy of $ND6_{CR}$/ $tRNA^{Glu}$ genes (fig. 5). Evidence supporting the duplication comes from the high sequence identity (~98%) shared by the small noncoding region between $tRNA^{Thr}$ and 5′ $tRNA^{Pro}$ and that between $tRNA^{Glu}$ and 3′ $tRNA^{Pro}$, as well as between the two copies of CR (99%). Whether the duplication occurred in the most recent common ancestor of Bathydraconidae, or only in the lineage leading to *R. glacialis*, will require characterizing the CR of additional bathydraconid species.

The mt genes *ND6*, $tRNA^{Glu}$, and $tRNA^{Pro}$ which have become translocated in Antarctic notothenioid mt rearrangements (figs. 2 and 5) are near the origin of heavy-strand replication (O_H, adjacent to CSB1 in the CR) in canonical mtDNA (Walberg and Clayton 1981). Vertebrate mt rearrangements commonly occur in regions near replication origins particularly O_H of the heavy strand. The cause is attributed to the greater probability of strand slippage, asynchronous initiation and termination, or imprecise termination during the replication of the circular mtDNA near the sites of replication origin that would result in duplication and rearrangement (Macey, Larson, Ananjeva, and Papenfuss 1997; Boore 1999, 2000). Known incidences of mt rearrangements near replication origins that shuffled the positions of *ND6*, $tRNA^{Glu}$, $tRNA^{Pro}$, and CR are found in diverse vertebrate taxa, including the lantern fish *Myctophum affine* (Miya et al. 2001), conger eel *C. myriaster* (Inoue et al. 2001), the salamanders *Aneides flavipunctatus* and *Stereochilus marginatus* (Mueller and Boore 2005), and different avian lineages (Mindell et al. 1998). The Antarctic notothenioids thus add to this list of examples.

Functionality of *tRNA* and CR Duplicates in Antarctic Notothenioids

The $tRNA^{Thr}$ and $tRNA^{Pro}$ duplicates in *P. antarcticum* and *N. angustata* and $tRNA^{Pro}$ duplicates in *R. glacialis* (fig. 2) all have viable secondary structures and same anticodon sequence and very high sequence identity (at least 97%) between the two copies. We thus infer that both copies are functional. All CR sequences in the rearranged mt genome of a species, either occurring singly or in duplicates, which we designated as complete CR, contain a full set of the conserved mt regulatory sequence modules, TAS, cTAS, and ETAS (fig. 1; supplementary fig. S1, Supplementary Material online). The TAS TACAT can base pair with its cTAS ATGTA, leading to formation of stable hairpin loops and presumably functions as a sequence-specific signal for the termination of D-loop strand synthesis (Doda et al. 1981). The ETAS functions in termination of mt DNA replication (Guo et al. 2003). An additional set of CSBs participate in the formation of a proper RNA primer for replication (Fernández-Silva et al. 2003). CSB-F like, CSB-E, CSB-D, CSB1, and CSB2 were identified in the conserved domain of all complete notothenioid CR sequences in this study, whereas CSB3 occurs only in the basal non-Antarctic *E. maclovinus* (supplementary fig. S1, Supplementary Material online). The missing CSB3 in Antarctic notothenioids may be inconsequential because the functional importance of CSB3 is dubious (Gemmell et al. 1996; Rotskaya et al. 2009), and it is also missing in *B. variegatus* and *P. urvillii* representing the other two basal non-Antarctic families (supplementary fig. S1, Supplementary Material online). The functionally important CSB1, which is always located near O_H and thought to play a role in the switch from RNA to DNA synthesis that commences at O_H (Walberg and Clayton 1981; Brown et al. 1986), however, is conserved in the complete CR of all species (supplementary fig. S1, Supplementary Material online). Thus, the single or duplicate complete CR present in all rearrangement patterns (fig. 5) have a full complement of essential regulatory elements and are assumed functional.

Positive Selection on $ND6_{CR}$

Signals of positive selection on the branch leading to $ND6_{CR}$ proteins of the species of the Antarctic radiation, and on several residues within $ND6_{CR}$ (fig. 4 and table 1), while not overwhelming, is statistically significant, suggesting diversifying adaptive change in the protein. The functional significance of the putative adaptive change is unknown. Because ND6 plays a crucial role in mt Complex I assembly (Bai and Attardi 1998), improving $ND6_{CR}$ conformational flexibility in the subzero Antarctic marine temperatures conducive to intersubunit interactions may be

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

a possibility. However, ND6 likely has other functional roles because several mutations in human ND6 cause optic neuropathy (Chinnery et al. 2001), and one that is associated with hypoxia-sensitive phenotype in human glioma cells (DeHaan et al. 2004). Additionally, Complex I, along with Complex III, are believed to be the major source of mt/cellular reactive oxygen species (ROS) because large changes in the potential energy of electrons (relative to reduction of O_2) occur at these two sites, which can result in premature electron leakage and free oxygen radical formation (Turrens 2003; Balaban et al. 2005). Also, the rate of ROS production increases with cellular O_2 tension; in human, hyperoxia was shown to increase H_2O_2 production by lung mitochondria (Turrens et al. 1982). Thus, high *in vivo* O_2 tension in Antarctic notothenioids resulting from cold, oxygen-rich ambient water would likely lead to greater ROS production and oxidative stress (Abele and Puntarulo 2004). The amino acid changes in the $ND6_{CR}$ protein may have a role in modulating Complex I redox potential and ROS production. All these are hypotheses subject to experimental testing. The detection of positive selection on $ND6_{CR}$, however, brings to light the need for including mt-encoded proteins in studies of evolutionary adaptation to environments of changing oxygenation.

Conclusion

In this study, we discovered the reportedly missing mt *ND6* and *tRNA*^{Glu} genes in Antarctic notothenioids (Papetti et al. 2007) was not a result of gene loss but their translocation through mt DNA rearrangement to previously uncharacterized portions of the CR. This drastic mt rearrangement is a derived synapomorphy of the species of the Antarctic radiation only, whereas basal non-Antarctic notothenioids have the canonical vertebrate mt gene order. Thus, we have identified an additional major molecular change associated with notothenioid evolution within the frigid polar environment besides antifreeze protein gain and hemoprotein loss. It is unclear what drove the initial mt rearrangement (the hypothesized tandem duplication spanning *ND6* through CR) in the ancestor of the Antarctic clade. The conventional thinking regarding structural alteration of mt genome is that it may result from selection for, or at least absence of selection against rearrangement, at the cell or the organism level. Unlike deletions, duplications of portions of the mt genome generally have no pathological consequence (in human) (Tang et al. 2000), suggesting that there may not be strong organismal-level selection against the initial region-wide duplication in the mt genome of the Antarctic notothenioid ancestor. Thus, the initial duplication in the Antarctic ancestor could have occurred simply by chance due to the affected region being near the O_H and prone to replication overrun. However, the fixation of the duplications and rearrangements in the population would require that they be selectively advantageous. The rearranged mitochondria must outcompete the canonical mitochondria at the intracellular level for the new haplotype to become fixed in the population. The process expectedly has to be driven by positive selection, otherwise the redundant gene copies would eventually be eliminated. The single most distinctive difference between the isolated Antarctic marine environment and non-Antarctic environments is its chronically cold and oxygen-rich condition. Thus, it is reasonable to suggest that the effect of rising marine oxygenation on mt respiration as Antarctic sea-level glaciation commenced might be involved in the retention of the initial region-wide duplication in the mt genome of the Antarctic notothenioid ancestor. Two copies of CRs with two O_H and duplicate *tRNAs* might increase transcriptional and translational efficiency in the production of mitochondria-encoded proteins of the respiratory chain to handle increased cellular oxygen tension. However, the duplicate CRs and *tRNAs* have not been maintained throughout Antarctic notothenioid evolution, but were reduced back to single copies in many species. The invariant feature among all the genic and noncoding sequence modifications in the rearranged region of all the Antarctic notothenioids we examined is the preservation of the $ND6_{CR}$ gene and *tRNA*^{Glu}. More importantly, $ND6_{CR}$ protein sequences have diverged substantially from canonical ND6 sequences of the basal non-Antarctic notothenioids inhabiting temperate waters, and signal of positive selection is detected on the $ND6_{CR}$ lineage and several residues in the protein, suggesting diversifying adaptive change in operation. What functional significance or improvement the putative adaptive amino acid changes confer would require further study, and a few testable hypotheses were suggested above. The identification of positive selection on $ND6_{CR}$ underscores the importance of including the mt genome in studies of evolutionary adaptation to changing environments. For Antarctic notothenioids, it would be interesting to examine whether the other mt proteins (mt and nuclear encoded) are under selection, to determine if there is a co-ordinated enhancement of mt protein functioning in the extreme cold and oxygen-rich polar marine environment.

Supplementary Material

Supplementary tables S1 and S2, figures S1 and S2 and file 1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Arthur DeVries for his help in the collecting the Ross Sea notothenioid species. This work was supported by National Science Foundation (OPP 0636696 to C.-H.C.C.). The sequences reported in this study have been deposited at National Center for Biotechnology Information under accession numbers GU214209-GU214230.

References

Abele D, Puntarulo S. 2004. Formation of reactive species and induction of antioxidant defence systems in polar and temperate marine invertebrates and fish. *Comp Biochem Physiol A Mol Integr Physiol*. 138:405–415.

Amer SAM, Kumazawa Y. 2007. The mitochondrial genome of the lizard *Calotes versicolor* and a novel gene inversion

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

in South Asian draconine agamids. *Mol Biol Evol*. 24: 1330–1339.

Bai Y, Attardi G. 1998. The mtDNA-encoded ND6 subunit of mitochondrial NADH dehydrogenase is essential for the assembly of the membrane arm and the respiratory function of the enzyme. *EMBO J*. 17:4848–4858.

Balaban RS, Nemoto S, Finkel T. 2005. Mitochondria, oxidants, and aging. *Cell* 120:483–495.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res*. 27:1767–1780.

Boore JL, 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. pp 133–148. in D Shankoff, and JH Nadeau, editors. Comparative genomics: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families. Dordrecht (The Netherlands): Kluwer Academic Publishers.

Boore JL, Daehler LL, Brown WM. 1999. Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* (Amphioxus). *Mol Biol Evol*. 16:410–418.

Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisa E. 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J Mol Evol*. 192: 503–511.

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.

Chen Z, Cheng C-HC, Zhang J, et al. (15 co-authors). 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc Natl Acad Sci USA*. 103:10491–10496.

Cheng C-HC, Chen L, Near TJ, Jin Y. 2003. Functional antifreeze glycoprotein genes in temperate-water New Zealand nototheniid fish infer an Antarctic evolutionary origin. *Mol Biol Evol*. 20:1897–1908.

Chinnery PF, Brown DT, Andrews RM, Singh-Kler R, Riordan-Eva P, Lindley J, Applegarth DA, Turnbull DM, Howell N. 2001. The mitochondrial ND6 gene is a hot spot for mutations that cause Leber's hereditary optic neuropathy. *Brain* 124:209–218.

Cocca E, Ratnayake-Lecamwasam M, Parker SK, Camardella L, Ciaramella M, di Prisco G, Detrich IHW. 1995. Genomic remnants of alpha-globin genes in the hemoglobinless antarctic icefishes. *Proc Natl Acad Sci USA*. 92:1817–1821.

DeHaan C, Habibi-Nazhad B, Yan E, Salloum N, Parliament M, Allalunis-Turner J. 2004. Mutation in mitochondrial complex I ND6 subunit is associated with defective response to hypoxia in human glioma cells. *BMC Mol Cancer*. 3:19.

DeVries AL. 1971. Glycoproteins as biological antifreeze agents in Antarctic fishes. *Science* 172:1152–1155.

Doda JN, Wright CT, Clayton DA. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci USA*. 78:6116–6120.

Eastman JT. 2005. The nature of the diversity of Antarctic fishes. *Polar Biol*. 28:94–107.

Fernández-Silva P, Enriquez JA, Montoya J. 2003. Replication and transcription of mammalian mitochondrial DNA. *Exp Physiol*. 88:41–56.

Gemmell NJ, Western PS, Watson JM, Graves JA. 1996. Evolution of the mammalian mitochondrial control region–comparisons of control region sequences between monotreme and therian mammals. *Mol Biol Evol*. 13:798–808.

Guo X, Liu S, Liu Y. 2003. Comparative analysis of the mitochondrial DNA control region in cyprinids with different ploidy level. *Aquaculture* 224:25–38.

Hardewig I, Peck LS, Pörtner HO. 1999. Thermal sensitivity of mitochondrial function in the Antarctic notothenioid, *Lepidonotothen nudifrons*. *J Comp Physiol B*. 169:597–604.

Inoue JG, Miya M, Tsukamoto K, Nishida M. 2001. Complete mitochondrial DNA sequence of *Conger myriaster* (Teleostei: anguilliformes): novel gene order for vertebrate mitochondrial genomes and the phylogenetic implications for anguilliform families. *J Mol Evol*. 52:311–320.

Janke A, Feldmaier-Fuchs G, Thomas WK, Von-Haeseler A, Paabo S. 1994. The marsupial mitochondrial genome and the evolution of placental mammals. *Genetics* 137:243–256.

Kurabayashi A, Sumida M, Yonekawa H, Glaw F, Vences M, Hasegawa M. 2008. Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. *Mol Biol Evol*. 25:874–891.

Lee WJ, Kocher TD. 1995. Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics* 139:873–887.

Macey JR, Larson A, Ananjeva NB, Fang Z, Papenfuss TJ. 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol Biol Evol*. 14:91–104.

Macey JR, Larson A, Ananjeva NB, Papenfuss TJ. 1997. Replication slippage may cause parallel evolution in the secondary structures of mitochondrial transfer RNAs. *Mol Biol Evol*. 14:30–39.

Mindell DP, Sorenson MD, Dimcheff DE. 1998. Multiple independent origins of mitochondrial gene order in birds. *Proc Natl Acad Sci USA*. 95:10693–10697.

Miya M, Kawaguchi A, Nishida M. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Mol Biol Evol*. 18: 1993–2009.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol Biol Evol*. 26:121–138.

Mueller RL, Boore JL. 2005. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Mol Biol Evol*. 22:2104–2112.

Near TJ, Pesavento JJ, Cheng C-HC. 2004. Phylogenetic investigations of Antarctic notothenioid fishes (Perciformes: notothenioidei) using complete gene sequences of the mitochondrial encoded 16S rRNA. *Mol Phylogenet Evol*. 32:881–891.

Papetti C, Lio P, Ruber L, Patterson J, Zardoya R. 2007. Antarctic fish mitochondrial genomes lack ND6 gene. *J Mol Evol*. 65:519–528.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst Biol*. 53:793–808.

Rotskaya UN, Rogozin IB, Vasyunina EA, Kolosova NG, Malyarchuk BA, Nevinsky GA, Sinitsyna OI. 2009. Analysis of mitochondrial DNA somatic mutations in OXYS and Wistar strain rats. *Biochemistry (Mosc)*. 74:430–437.

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.

Tang Y, Schon EA, Wilichowski E, Vazquez-Memije ME, Davidson E, King MP. 2000. Rearrangements of human mitochondrial DNA (mtDNA): new insights into the regulation of mtDNA copy number and gene expression. *Mol Biol Cell*. 11: 1471–1485.

Turrens JF. 2003. Mitochondrial formation of reactive oxygen species. *J Physiol*. 552:335–344.

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024

Turrens JF, Freeman BA, Crapo JD. 1982. Hyperoxia increases H_2O_2 release by lung mitochondria and microsomes. *Arch Biochem Biophys*. 217:411–421.

Urschel M, O'Brien KM. 2009. Mitochondrial function in Antarctic notothenioid fishes that differ in the expression of oxygen-binding proteins. *Polar Biol*. 32:1323–1330.

Walberg MW, Clayton DA. 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Res*. 9:5411–5421.

Weinstein RB, Somero GN. 1998. Effects of temperature on mitochondrial function in the Antarctic fish *Trematomus bernacchii*. *J Comp Physiol B*. 168:190–196.

Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution. pp. 173–216. in DR Wolstenholme, and KW Jeon, editors. Mitochondrial genomes. San Diego (CA): Academic Press.

Yamanoue Y, Miya M, Matsuura K, Yagishita N, Mabuchi K, Sakai H, Katoh M, Nishida M. 2007. Phylogenetic position of tetraodontiform fishes within the higher teleosts: Bayesian inferences based on 44 whole mitochondrial genome sequences. *Mol Phylogenet Evol*. 45:89–101.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.

Downloaded from https://academic.oup.com/mbe/article/27/6/1391/1115839 by guest on 18 April 2024