IRis

Northeastern University

Biology Master's Theses

Department of Biology

January 01, 2009

Evolution and function of the globin gene complexes of Antarctic Dragonfishes (Notothenioidei: Bathydraconidae)

Joyce Yuk-Ting Lau
Northeastern University

Recommended Citation

Lau, Joyce Yuk-Ting, "Evolution and function of the globin gene complexes of Antarctic Dragonfishes (Notothenioidei: Bathydraconidae)" (2009). *Biology Master's Theses.* Paper 9. http://hdl.handle.net/2047/d20000336

This work is available open access, hosted by Northeastern University.

Evolution and Function of the Globin Gene Complexes of Antarctic Dragonfishes (Notothenioidei: Bathydraconidae)

A dissertation presented

by

Joyce Yuk-Ting Lau

to
The Department of Biology

In partial fulfillment of the requirements for the degree of
Master of Science

in the field of

Biology

Northeastern University,
Boston, Massachusetts
May, 2009

Evolution and Function of the Globin Gene Complexes of Antarctic Dragonfishes (Notothenioidei: Bathydraconidae)

A dissertation presented

by

Joyce Yuk-Ting Lau

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biology
in the Graduate School of Arts and Sciences of
Northeastern University, May, 2009

Abstract:

As the Southern Ocean cooled to –1.8ºC over the past 40 million years, the teleostean clade Notothenioidei became less reliant on hemoglobin and red blood cells, a trend which culminates in the white-blooded, erythrocyte-null Antarctic icefishes. The adult α- and β-globin genes of red-blooded notothenioids are linked in 5'-to-5' orientation such that the intergenic sequences direct divergent transcription of the loci. To gain further insight into their evolution and function, we have compared the intergenic regions of the adult α/β-globin gene complexes from three species of Antarctic rockcods (Nototheniidae), a basal notothenioid group, and from eight species of Antarctic dragonfishes (Bathydraconidae), the red-blooded notothenioids most closely related to the icefishes, as well as the α- and β-globin of 13 species and 10 genera of dragonfish.

The ancestral nototheniid intergene appears to be ~3 kb in length, although one species contains a duplication of ~1 kb. The bathydraconid intergenes resolve into three distinct subclasses [long (3.8 kb), intermediate (3.0 kb), and short (1.5-2.3 kb)] that correspond to the subfamilies proposed for the taxon: Gymonodraconinae, Bathydraconinae, and Cygnodraconinae. Using luciferase reporter technology, we assessed the promoter/enhancer activities of the intergenes from the rockcod *Notothenia coriiceps* and the dragonfishes *Akarotaxis nudiceps* (Bathydraconinae) and *Gerlachea australis* (Cygnodraconinae) in the erythropoietic microenvironment of differentiated MEL cells. We found that the *N. coriiceps* intergene directs high-level transcription in both orientations, the *A. nudiceps* intergene is active only in α orientation, and the short intergene of *G. australis* supports only weak transcription.

Phylogenetic analysis indicated that dragonfish are paraphyletic. Our α, β-globin, and intergenic data agree with the subfamilies proposed by mitochondrial data but Cygnodraconinae and Gymnodraconinae are more closely related and Bathydraconinae should be split into two subclades. The icefish representative *N. ionah* was found to have diverged from the dragonfish, after the initial split of the dragonfish family such that *N. ionah* is more closely related to Cygnodraconinae and Gymnodraconinae. Studies of the β-globin gene resulted in the discovery of pseudogenes in a number of dragonfish. This raises the question as of multiplicity of β-globin in dragonfish as well as other notothenioids. Our results show that dragonfish can be resolved into four groups, with icefish more related to a subset of the dragonfish family. Functional and phylogenetic results are consistent with the hypothesis that the notothenioid globin loci are in evolutionary flux, probably due to relaxation of selection pressure for hemoglobin expression.

Acknowledgements

I would like to thank everyone in my life, who have helped shape my life academically and personally, so that I could become who I am presently. There are a few people I would especially like to thank

Dr. H. William Detrich, who has given me the opportunity to learn so much about biology, about how to write, about the Antarctic, and about myself. You have given me a chance to think and grow, and I will remember all the good advice you've given me.

Ms. Sandra K. Parker, who has provided me not only with technical assistance and advice, but who has been a wonderful friend, putting up with my "when I grow up, I want to be…" and all my strange quirkiness.

Dr. Steven Vollmer, who has taught me and helped me so much with the interesting world of phylogeny.

Dr. Erin Cram, who has given me wonderful advice and suggestions and has helped all graduate students through the weekly graduate colloquium so that we can become better and more comfortable speakers.

To all the graduate students in the Detrich lab, past and present, who some way or another, has helped me complete this thesis and all the graduate students of Northeastern Biology who have made me come to love Boston.

To my friends and family, my mom, my sister Dawn, and my grandparents who have always taught me to find what I love, to enjoy what I do, and to live life to the fullest. Finally, Chris, your love has always made everything in my life even better and you have stood by me to support my dreams and aspirations.

Thank you all for sharing your lives with mine.

Table of contents

Abstract	**3**
Acknowledgements	**5**
Table of Contents	**6**
List of Abbreviations	**7**
List of Figures	**9**
List of Tables	**10**
Chapter 1.	**11**
Abstract	**12**
Introduction	**13**
Methods	**16**
Results	**23**
Discussion	**28**
Contributions	**32**
Chapter 2.	**33**
Abstract	**34**
Introduction	**35**
Methods	**38**
Results	**43**
Discussion	**46**
Contributions	**51**
References	**52**
Figures and Tables	**57**

List of Abbreviations

Abbreviations	Proper Name
AMP	Ampicillin
APF	Antarctic polar front
BAC	Bacterial artificial chromosome
bp	Base pair(s)
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphates
DLR	Promega dual-luciferase reporter
EKLF	Erythroid *Krüppel*-like factor
FBS	Fetal bovine serum
Indel	Insertions and deletions
IPTG	Isopropyl-thio-β-D-glactoside
kb	Kilobases
LB	Lurai-Bertani medium; 1% bacto-trptone, 0.5% bacto-yeast extract, 0.5% NaCl (pH7.5)
LCR	Locus control region
MEL	Murine erythroleukemia cells
ML	Maximum likelihood
MP	Maximum parsimony
mtDNA	Mitochondrial DNA

Mya	Million years ago
NF-E2	Nuclear factor, erythroid 2
NJ	Neighbor-Joining
NNI	Nearest-neighbor interchange
TBR	Tree bisection-reconnection
PAC	P1 derived artificial chromosome
PCR	Polymerase chain reaction
PBS	Phosphate-buffered saline
RLU	Relative light units
X-gal	5-bromo-4chloro-3-indolyl-b-D-galactopyranoside

List of Figures

Fig 1. Nucleotide sequence alignments for nototheniid and bathydraconid globin intergenic regions

Fig 2. Structure of the adult globin intergenic regions from three nototheniids

Fig 3. Structure of the adult globin intergenic regions of eight bathydraconids

Fig 4. Sequence identity matrix for globin intergenic regions from 11 notothenioid fishes

Fig 5. Sequence comparison of putative nototheniid and bathydraconid enhancer motifs found in the 1-C intergenic element

Fig 6. Activities of the wild-type globin intergenic regions of *N. coriiceps*, *A. nudiceps*, and *G. australis* in a hematopoietic microenvironment

Fig 7. Representation of α-globin, β-globin and intergenic region orientation and location of primers

Fig 8. Maximum likelihood trees for α-globin, β-globin and intergenic region and corresponding bootstrap values

List of Tables

Table 1. Primer sequences used the amplify globin intergenic regions during first-round cloning

Table 2. Forward and reverse primer pairs used for first-round intergene amplification

Table 3. Primers sequences used to amplify the intergenic regions during second-round cloning

Table 4. Forward and reverse primers for amplification of second-round intergene amplification

Table 5. Sample source, location, species, family, accession numbers and primers

Chapter 1

Structural and functional variability of the globin intergenic regulatory regions of the Antarctic dragonfishes (Notothenioidei: Bathydraconidae): A locus in evolutionary flux

Abstract

As the Southern Ocean cooled to –1.8ºC over the past 40 million years, the teleostean clade Notothenioidei became less reliant on hemoglobin and red blood cells, a trend which culminates in the white-blooded, erythrocyte-null Antarctic icefishes. The adult α- and β-globin genes of red-blooded notothenioids are linked in 5'-to-5' orientation such that the intergenic sequences direct divergent transcription of the loci. To gain further insight into their evolution and function, we have compared the intergenic regions of the adult α/β-globin gene complexes from three species of Antarctic rockcods (Nototheniidae), a basal notothenioid group, and from eight species of Antarctic dragonfishes (Bathydraconidae), the red-blooded notothenioids most closely related to the icefishes. The ancestral nototheniid intergene appears to be ~3 kb in length, although one species contains a duplication of ~1 kb. The bathydraconid intergenes resolve into three distinct subclasses [long (3.8 kb), intermediate (3.0 kb), and short (1.5-2.3 kb)] that correspond to the subfamilies proposed for the taxon: Gymonodraconinae, Bathydraconinae, and Cygnodraconinae. Using luciferase reporter technology, we assessed the promoter/enhancer activities of the intergenes from the rockcod *Notothenia coriiceps* and the dragonfishes *Akarotaxis nudiceps* (Bathydraconinae) and *Gerlachea australis* (Cygnodraconinae) in the erythropoietic microenvironment of differentiated MEL cells. We found that the *N. coriiceps* intergene directs high-level transcription in both orientations, the *A. nudiceps* intergene is active only in α orientation, and the short intergene of *G. australis* supports only weak transcription. Our results are consistent with the hypothesis that the notothenioid globin loci are in evolutionary flux, probably due to relaxation of selection pressure for hemoglobin expression.

1. Introduction

The Southern Ocean surrounding Antarctica began to cool ~41 to 30 million years ago (Mya) following the opening of the Drake Passage between South America and the Antarctic Peninsula (DeWitt, 1971; Eastman, 1993; Scher and Martin, 2006) and the establishment of the Antarctic Circumpolar Current. By the time the Southern Ocean attained its present frigid temperatures (–2 to +2ºC) ~14-10 Mya, most temperate, shallow-water marine teleosts had become locally extinct (Eastman, 2005), which created new ecological niches for other taxa, such as the perciform suborder Notothenioidei, that were diversifying *in situ*. Today the modern fish fauna of the Southern Ocean is taxonomically restricted, highly endemic, and singularly dominated by notothenioid species, which constitute 77% of species and 90% of biomass on the Antarctic continental shelf (Eastman, 2005).

The notothenioids have undergone rapid morphological, molecular, and ecological diversification as they adapted to a cold, highly oxygenated environment (Cheng and Detrich, 2007). Musculoskeletal diversification away from the ancestral benthic condition toward pelagic or partially pelagic zooplanktivory enabled the taxon to exploit unfilled niches in the water column (Eastman, 1993; Eastman, 2005). The acquisition of novel antifreeze proteins (Chen et al., 1997; Cheng and Chen, 1999) by notothenioids and their loss of an inducible heat shock response (Hofman et al., 2000; Buckley and Hofman, 2004) are examples of novel traits that evolved in a perennially icy environment. The universal loss of the oxygen-transport protein hemoglobin by icefishes (Notothenioidei: Channichthyidae), which is due to disruption and deletion of the globin loci (Cocca et al., 1995; Zhao et al., 1998; Near et al., 2006), and their variable loss of

myoglobin expression (Sidell and O'Brien, 2006) are perhaps the most perplexing "adaptations" of the notothenioids. Ruud (1954), who published the first systematic analysis of the blood of icefishes, concluded that the loss of hemoglobin (and erythrocytes) by icefishes could only have occurred in the extremely cold and oxygen rich waters of the Southern Ocean, an environment in which selective pressure to express an oxygen transport protein is reduced or absent. In this context, it is noteworthy that one may poison the hemoglobin of some red-blooded nototheniids with carbon monoxide without lethal effect (di Prisco, 2000).

Considering the plasticity of hemoprotein loci in icefishes, the crown group of the Notothenioidei, we chose to reassess the structure of the adult globin gene complex from a basal notothenioid family, the Nototheniidae, and to examine the structure and functionality of globin genes in the Bathydraconidae (dragonfishes), the red-blooded sister taxon to the white-blooded Channichthyidae. Relative to temperate and tropical fishes, basal notothenioid families (i.e., the Nototheniidae) have lower hematocrits and reduced red cell hemoglobin concentrations (Kunzmann et al., 1991), consistent with Ruud's hypothesis and in keeping with the results of carbon monoxide poisoning. Trends toward reduced hemoglobin multiplicity, functionality, and concentration have been reported for species in more derived notothenioid families (D'avino and di Prisco, 1987; Kunzmann et al., 1991; Bargelloni et al., 1998; di Prisco et al., 1998; di Prisco et al., 2007), which culminate in the hemoglobin-null icefishes. Here we report that the adult α and β globin genes of the Antarctic toothfish, *Dissostichus mawsoni* (Nototheniidae), are organized in 5'-to-5' linkage like those of the Antarctic yellowbelly rockcod (also known as the bullhead notothen), *Notothenia coriiceps*, and the cool-temperate New Zealand

black cod, *N. angustata*. However, the globin intergenic region of *D. mawsoni*, like that of *N. angustata*, lacks the 1.1-kb, repeat-containing intergenic duplication of *N. coriiceps*. The adult α- and β-globin genes of eight dragonfishes are similarly organized in 5'-to-5'linkage, but their intergenic regions are usually shorter than those of the nototheniids. Furthermore, the structures of the bathydraconid intergenes are concordant with the three proposed subfamily lineages of this paraphyletic taxon (Derome et al., 2002), but their suites of potential transcription-factor-binding motifs are variable. Analysis of the promoter activities of the intergenes of two species, *Gerlachia australis* (subfamily Cygnodraconinae) and *Akarotaxis nudiceps* (subfamily Bathydraconinae), reveals substantial interspecific variability of the function of these gene regulatory regions. Thus, we suggest that the globin intergenic regions of species from other notothenioid families may also show high intrafamilial variability.

2. Methods

2.1. Nototheniid genomic libraries/cloning and sequencing of adult globin gene complexes

To reevaluate the structure of the adult α/β globin gene complexes of the family Nototheniidae, we cloned them from genomic libraries of two species. The BAC library of *N. coriiceps* genomic DNA (VMRC19, available from the BACPAC Resource Center, Children's Hospital Oakland Research Institute, Oakland, CA) was prepared from high molecular weight erythrocyte DNA using the vector pCCBACE1 (Epicentre Technologies) according to standard protocols (Amemiya et al., 1996; Osoegawa et al., 1998; Danke et al., 2004; Chris T. Amemiya and H.W. Detrich, manuscript in preparation). The *D. mawsoni* genomic DNA library in Lambda FIX II, prepared from spleen, was generously provided by Dr. Christina Cheng (University of Illinois, Urbana).

The *N. coriiceps* BAC library was screened for the α/β globin gene complex by simultaneous probing of high density filters with *N. coriiceps* α- and β-globin cDNAs (GenBank acc. nos. U09186 and U09187, respectively); probes were labeled with ^{32}P by random priming. Fifty-five positive clones were screened further using genomic probes corresponding to the α-globin gene, to the β-globin gene, and to the intergenic region that links the two genes (cf. GenBank acc. no. AF049916). Four BAC clones hybridized to each of the three probes, one (BAC-B9) was chosen for *Bam*HI digestion, and the insert fragments were subcloned into pBluescript (Stratagene). DNA inserts were sequenced on both strands using automated methods by the University of Maine DNA Sequencing Facility or by MWG-The Genomic Company (Huntsville, AL).

Similarly, the *D. mawsoni* genomic library was screened for the adult globin complex using mixed α- and β-globin genomic DNA clones from *N. coriiceps*. One double-positive clone, DM7, was sequenced completely on both strands.

2.2. Collection of dragonfishes, storage of tissues, and isolation of genomic DNAs

Specimens of eight dragonfishes (Bathydraconidae) were collected by bottom trawling in the Antarctic Peninsula region, the Ross Sea, or the Weddell Sea over the course of several Antarctic expeditions. Following euthanasia of the fishes, spleens from some animals were dissected, flash frozen in liquid nitrogen, and stored at –80 ºC. Pectoral fin clips were stored in 95% ethanol at –20 ºC.

Genomic DNAs of *Akarotaxis nudiceps* (UTTC1795), *Cygnodraco mawsoni* (TJN2050), *Gerlachia australis* (UTTC1722), *Gymnodraco acuticeps* (HWD116), *Parachaenichthys charcoti* (TJN2009), *Parachaenichthys georgianus* (HWD118), *Prionodraco evansii* (TJN2300), and *Vomeridens infuscipinnis* (HWD117) were obtained from spleen samples archived at –80 ºC (HWD) or from ethanol-fixed fin clips (TJN, UTTC) using methods previously described (Near et al., 2006).

2.3. Cloning and sequencing of the intergenic regions of the bathydraconid globin loci

Because the adult α- and β-globin genes of notothenioids are tightly linked 5' to 5' (< 5 kb between start codons for species examined to date), their intergenic regions, which we define as the sequences that separate the start codons, were amplified during the first round by using the Polymerase Chain Reaction (PCR) and forward and reverse primers derived from the first exons of each gene pair. Table 1 gives the sequences of

each primer, and Table 2 indicates the primer pairs used for each species. PCR reactions (50 µl total volume) contained 200 µM dNTPs, 0.5 µM of each primer (forward from the α-exon 1 and reverse from β-exon 1), 0.5-1.0 mM $MgCl_2$, 1-2 µl of DMSO, 1× Advantage 2 PCR buffer, 0.6× Advantage 2 PCR polymerase, 0.1-0.3 µg of genomic DNA, and sterile water to volume. Touchdown PCR (Don et al., 1991; Roux, 1995a) was performed for 10 cycles using the following parameters: 1) denaturation at 95 ºC for 30 s; 2) primer annealing for 1 min, ramping temperature from 65 °C to 55 ºC in one degree increments/cycle; and 3) extension at 68 °C for 5 min. Subsequently, 20 cycles of conventional PCR were performed using the following profile: 1) denaturation at 94 °C for 30 s; 2) primer annealing at 58 °C for 1 min; and 3) extension at 68 °C for 5 min. PCR products were electrophoresed on 1% agarose gels, and the desired DNA fragments were purified by use of the Qiagen Qiaquick gel purification kit.

DNA fragments were subcloned by ligation into the PGEM-T Easy vector (Promega, Madison, WI), and the recombinant plasmids were transformed into competent NEB 5-alpha *E. coli* cells (New England BioLabs, Ipswich, MA). Transformed cells were spread onto LB-Amp plates and grown overnight in the presence of X-Gal and IPTG. White colonies were picked and grown overnight in LB (Ruther, 1980). Plasmids were purified from the liquid cultures following the Wizard Plus Miniprep (Promega, Madison, WI) protocol. Plasmid inserts were sequenced by automated methods on both strands as described under section 2.1.

2.4. Sequence analysis of globin intergenic regions

Programs of the DNASTAR Lasergene package were used to align and compare the intergenic DNA sequences and to calculate pairwise sequence identities; minor adjustments were made using BioEdit (Hall, 1999). To identify potential promoter and enhancer motifs, sequences were analyzed with Matinspector and the ElDorado Genomes database (Genomatix, Munich, Germany). The intergenic sequences were also scanned for the presence of the direct and inverted repetitive elements described for *N. coriiceps* by Lau et al. (2001).

2.5. GenBank data deposition

Using genomic DNA as template, we (Near et al., 2006) had previously amplified and sequenced the adult α-globin genes of *A. nudiceps* (DQ317934), *G. australis* (DQ317938), *P. evansii* (DQ317941), *C. mawsoni* (DQ317939), and *G. acuticeps* (DQ317940). In this study we completed the adult globin loci of these species by sequencing the linked β-globin genes and the intergenic regions. The corresponding GenBank files have been updated to incorporate the new sequence information. The complete sequences of the α/β globin loci of *P. georgianus*, *V. infuscipinnis*, and *D. mawsoni* have been deposited in GenBank under the accession numbers FJ92838, FJ792839, and FJ896404 respectively.

2.6. Construction of globin intergene/luciferase reporter vectors

To evaluate the activities of the α- and β-globin *cis*-regulatory elements of the intergenic regions of *G. australis* [representing the subfamily Cygnodraconinae (Derome

et al., 2002)] and *A. nudiceps* [representing the Bathydraconinae (Derome et al., 2002)], each intergene was subcloned in both orientations into the pGL3-Basic vector (Promega Dual-Luciferase Reporter; DLR). Because the original *G. australis* and *A. nudiceps* intergenic clones (pGEM-Gaus and pGEM-Anud, respectively) contained portions of the first exons of the α- and β-globin genes, we employed the nested PCR-strategy of Lau et al. (2001) in a second round of cloning to transfer only their respective intergenic sequences to the reporter plasmid. The PGL3-Basic vector was prepared to receive the intergenic sequences by sequential digestion with *Bgl*II (compatible with *Bam*HI) and *Nhe*I (compatible with *Avr*II) and dephosphorylation of the ends of the linearized vector using Antarctic phosphatase (New England Biolabs, Ipswich, MA).

The *G. australis* and *A. nudiceps* intergenes were amplified by PCR using the touchdown protocol of section 2.3 and primers (Table 3) designed with restriction sites at their 5' ends to facilitate directional insertion of the intergenes into pGL3-Basic. To obtain insertion in both orientations, two different primer pairs were used with each intergene (Table 4). The *G. australis* intergenic reporter pGausPα (i.e., with the α-promoter sequence driving luciferase reporter expression) was obtained by amplification of the template pGEM-Gaus using primers AustANheI and AustBBgl2. Following sequential digestion with *Nhe*I and *Bgl*II, the product was ligated to the complementary ends of the linearized pGL3-Basic vector. Similarly, pGausPβ (i.e., with the β-promoter sequence driving luciferase reporter expression) was produced using pGEM-Gaus as template and primers AustABgl2 and AustBNheI (AustABgl2 and AustBNheI were derived from AustANheI and AustBbgl2 by interchange of the *Nhe*I and *Bgl*II restriction sites; Table 4). The *A. nudiceps* intergenic reporters pAnudPα and pAnudPβ were

generated by the same strategy but exploited the compatible ends produced by *Nhe*I and *Avr*II and by *Bgl*II and *Bam*HI (Table 4). Recombinant plasmids were transformed into NEB 5 alpha cells for propagation. The orientation and sequence fidelity of the intergene in each of the four constructs were confirmed by automated DNA sequence analysis (section 2.1).

2.7. Transient transfection assay of promoter/reporter plasmids in murine erythroleukemia (MEL) cells

The four promoter/reporter constructs were assayed for transcription-promoting activity in the inducible hematopoietic microenvironment provided by MEL cells (generously provided by Dr. Stuart Orkin, Dana-Farber Cancer Institute, Harvard Medical School). MEL cells [2 ml at 5×10^5 /ml in DMEM + 10% FBS and penicillin (100 IU/ml)-streptomycin (100 µg/ml)] were incubated for 24 h in 6-well culture dishes (Primaria 6-well surface modified, Promega, Madison, WI) at 37 °C in a humid 5% CO_2 atmosphere. The cells were transfected with the plasmid constructs [3 µg in 100-µl sterile water, doped with 25 ng of the *Renilla* luciferase reporter plasmid pRL-SV40 (Promega], following the Turbofectin 8 (OriGene, Rockville, MD) transfection protocol (4 µl of turbofectin per µg of DNA). The cells were incubated 24 h at 37 °C (5% CO_2, humid incubator), split (1.0 ml MEL cells + 1.5 ml DMEM + 10% FBS + antibiotics), and then transferred to 6-well culture dishes. To induce hematopoietic differentiation, 1.8% DMSO was added to experimental cell cultures, which were incubated (37 °C, 5% CO_2, humid incubator) for an additional 72 h. Control cultures (no DMSO) were incubated for 48 h. Cells were harvested by centrifugation (IEC clinical centrifuge, speed 4, 4 min,

room temperature), the pellets were resuspended in 1.0 ml PBS, and cells were collected by centrifugation as before. Cell extracts (prepared by the manufacturer's protocol; Promega Dual Luciferase Reporter kit), were assayed in duplicate for both firefly and *Renilla* luciferase activities using an Optocomp I luminometer (MGM Instruments). (N.B. Control cell extracts were stored at –80 °C for 24 h prior to luciferase assay.) Three independent transfections were performed for each reporter construct. Firefly luciferase activities were normalized with respect to the *Renilla* enzyme.

Results

3.1 Reappraisal of the adult globin locus of the family Nototheniidae

We have previously reported (Lau et al., 2001) that the yellowbelly rockcod (*N. coriiceps*) and the congeneric New Zealand black cod (*N. angustata*) possess adult α- and β-globin genes that are tightly linked in 5'-to-5' orientation. The two α/β globin gene complexes are similar, with the exception of a 1.1-kb, repeat-containing segment that is present only in the Antarctic species. Furthermore, the *N. coriiceps* intergenic region functions in both orientations as a stronger promoter than the intergene of *N. angustata*, which indicates that the duplicated region contributes additional *cis*-regulatory elements that enhance transcription. As we studied the globin pseudogenes of Antarctic icefishes (Near et al., 2006) and began to characterize the functional globin intergenic regions of the dragonfishes (this report), we decided to reassess the structure of the globin gene complex of the speciose basal family Nototheniidae by recloning the *N. coriiceps* locus from a BAC library and by cloning of the locus from a second nototheniid, *D. mawsoni*.

Figure 2 shows the structure of the adult globin intergenic regions from three nototheniids: *N. coriiceps*, *N. angustata*, and *D. mawsoni*. The length (4.3 kb), organization and sequence of the *N. coriiceps* locus obtained from BAC clone B9, including its direct (NcDR1 and NcDR2) and indirect (not shown) repetitive elements, matched exactly the locus that we determined previously using a PCR-based cloning strategy and an independent *N. coriiceps* genomic DNA library in phage lambda (Lau et al., 2001). In contrast, the *D. mawsoni* locus (2.9 kb) was approximately 1.5 kb shorter and, like that of *N. angustata*, lacked the 1.1-kb, NcDR2-containing intergenic

duplication of *N. coriiceps*. (The *D. mawsoni* intergene also differed from those of the *Notothenia* spp. due to a small (~460 bp) deletion near the start of the β-globin coding sequence.) Together these results confirm our prior description of the globin intergene of *N. coriiceps* but suggest that the fundamental notothenioid α/β globin locus is similar to the shorter intergenes of *D. mawsoni* and of *N. angustata*.

3.2. Structural comparison of bathydraconid intergenic regions

Figure 1 shows the sequences of the globin intergenic regions of eight bathydraconids. The intergenic sequences of three nototheniids, *N. coriiceps*, *N. angustata*, and *D. mawsoni* are shown for comparison. The dragonfish regions varied in length from ~1.6 kb (*G. australis*) to ~3.9 kb (*G. acuticeps*), whereas the nototheniid intergenes were 3.2 kb (*N.* angustata) and 4.3 kb (*N. coriiceps*). Figure 3 indicates that the bathydraconid intergenes, which contained numerous insertions and deletions (indels) with respect to the nototheniids, resolved into three distinct subclasses [long (3.8 kb), intermediate (3.0 kb), and short (1.5-2.3 kb)] that corresponded to the subfamilies proposed for the taxon (Derome et al., 2002): Gymonodraconinae (*G. acuticeps*); Bathydraconinae (*P. evansii*, *A. nudiceps*, *V. infuscipinnis*); and Cygnodraconinae (*C. mawsoni*, *P. charcoti*, *P. georgianus, G. australis*). The correspondence of intergene subclass with species subfamily was reinforced by comparison of sequence identities (Fig. 4). Thus, the intergenes of the three species of the Bathydraconinae were 0.869-0.951 identical, whereas the range for the four species of Cygnodraconinae was 0.583-0.908. Identities measured by comparisons of intergenes across the subfamily boundaries

were substantially smaller. *G. acuticeps* was noteworthy in that it possessed an ~830-bp intergenic insertion vicinal to the α-globin gene that was not present in other species.

G. acuticeps (Gymnodraconinae) and the three species of Bathydraconinae possessed single copies of truncated versions of the *N. coriiceps* direct repeat NcDR1 (Lau et al., 2001), which includes subregions A, B, and C (Fig. 3; red, blue, and green rectangles, respectively). (Restriction site mapping and conservation of *cis*-regulatory elements demonstrated that these sequences did not correspond to NcDR2-A, -B, and -C; see Fig. 3 and Discussion.) The A elements of these four bathydraconids were ~90 bp shorter than the *N. coriiceps* 1-A, whereas their 1-B and 1-C elements were of similar in length to those of the nototheniid. By contrast, the 1-A and 1-B elements were not found in the Cygnodraconinae, which retained only part of the 1-C element (91-104 bp compared to 137 bp in NcDR1-C).

3.3. Bathydraconid intergenic enhancer motifs

Using luciferase reporter technology, Lau et al. (2001) reported that the major enhancer motif of the *N. coriiceps* globin intergene, designated NcEnh1, resided in direct repeat subregion NcDR1-C. The retention of all, or most of, the NcDR1-C element among the dragonfishes, including the Cygnodraconinae whose intergenes have undergone extensive deletion with respect to the basal nototheniids, argues strongly for the regulatory importance of NcEnh1. Figure 5 compares the sequences and putative regulatory elements of the 1-C region for three rockcods (*N. coriiceps*, *N. angustata*, and *D. mawsoni*) and for the eight dragonfishes. Like *N. angustata* and *D. mawsoni*, *G.*

acuticeps and the Bathydraconinae possessed two CCACC and two GATA motifs, elements which have been implicated in the control of globin gene transcription (Myers et al., 1986; Orkin, 1992; Pevny et al., 1995). The Cygnodraconinae differed from the other dragonfishes due to elimination of the CCACC elements. Thus, one might anticipate that the transcription-stimulating activity of the Cygnodraconinae globin intergenes would be reduced relative to those of the Bathydraconinae. The strong sequence conservation between positions 3,000 and 3,050 suggests that this region, which appears to lack conventional erythroid gene regulatory elements, might nevertheless play a structural role in forming the globin gene transcription initiation complex (see Discussion).

3.4. Promoter/enhancer activity of two bathydraconid globin intergenes

To assess the capacity of bathydraconid globin intergenes to drive transcription, we cloned those of two species, *A. nudiceps* (Bathydraconinae) and *G. australis* (Cygnodraconinae), into the luciferase-reporter vector pGL3-Basic for analysis in the inducible MEL cell hematopoietic microenvironment. The functional comparison of *N. coriiceps*, *G. australis*, and *A. nudiceps* intergenic regions was performed in MEL cells because comparable cell lines from the hematopoietic tissues (*e.g.*, pronephric kidney and spleen) of fishes, Antarctic or temperate, are not available. Our analysis assumes, therefore, that the promoter and enhancer elements of the fish intergenic regions are able to interact productively with murine hematopoietic transcription factors. We also assume that the physical states (*e.g.,* supercoiling) of the six plasmids are comparable and without effect on promoter activities. First, as a positive control to facilitate inter-comparability

of the current studies with our previous work (Lau et al., 2001), we made use of our extant *N. coriiceps* constructs, pNcorPα1 and pNcorPβ. Figure 6 shows that the *N. coriiceps* intergenic region in both orientations supported high-level expression of luciferase, but only following induction of the MEL-cell erythroid phenotype. The luciferase levels produced by pNcorPα1 and pNcorPβ approached those driven by the strong SV-40 enhancer and were ~100-fold greater than those obtained with the promoterless and enhancerless negative control, pGL3-Basic. The concordance of the normalized luciferase activities of these four reporter constructs with our previous results (Lau et al., 2001) was excellent, which indicated that our MEL cell assay is robust.

The results obtained for the bathydraconid intergenes were striking (Fig. 5). The *A. nudiceps* intergene in α orientation (pAnudPα) produced strong and statistically significant luciferase activity after erythroid induction ($p < 0.05$, two-tailed *t*-test), whereas the pAnudPβ gave only a background signal. *G. australis*, by contrast, showed a small, but statistically significant induction of luciferase activity ($p < 0.05$, two-tailed *t*-test) for the β orientation (pGausPβ), but little inducible activity was observed for pGausPα. Testing the induced signals by one-way ANOVA indicated that the enhancer activities of pNcorPα1, pGausPα, and pAnudPα were statistically different ($p < 0.05$), as were the enhancers in β orientation. Thus, the longer *A. nudiceps* globin intergene behaved asymmetrically as a strong unidirectional enhancer, whereas the shorter *G. australis* intergene functioned as a weak enhancer in β orientation.

Discussion

4.1. The notothenioid globin intergenic region – a reassessment

We have described previously the structural and functional properties of the globin intergenic regions of the Antarctic yellowbelly rockcod *N. coriiceps* and of the New Zealand black cod *N. angustata*, a cool-temperate congener (Lau et al., 2001). Because our prior analysis of the globin pseudogenes of the icefishes (Near et al., 2006) and our new dragonfish globin intergenic data indicated that this locus is quite malleable in notothenioid fishes, we returned to the basal family Nototheniidae to re-examine our baseline for comparison of the globin gene complex. Our results suggest strongly that the ancestral notothenioid α/β globin locus possessed an intergenic region of ~3 kb, similar to those of *D. mawsoni* and of *N. angustata*. Thus, we hypothesize that the larger, repeat-containing intergene of *N. coriiceps* represents a gain-of-function insertion of additional enhancer elements that may not characterize the Nototheniidae as a whole. In this context, sequencing of the globin gene complexes of *N. rossii*, a congener of *N. coriiceps* with a similarly active lifestyle, and of nototheniids with sluggish behavior (e.g., *Gobionotothen gibberifrons*) should resolve this issue.

4.2. The bathydraconid globin intergenes – regulatory regions in flux

The Notothenioidei suborder shows a marked tendency toward reduction of hemoglobin multiplicity, functionality, and concentration as one moves from basal to more derived families (D'avino and di Prisco, 1987; Kunzmann et al., 1991; Bargelloni et al., 1998; di Prisco et al., 1998; di Prisco et al., 2007), which is consistent with a

relaxation of selective pressure for hemoglobin function. Nevertheless, we were surprised to observe the extent to which the α/β-globin gene complexes vary in the family Bathydraconidae, the sister group to the hemoglobin-null icefishes. Although our taxonomic sampling is incomplete (eight of the 16 recognized species), our results suggest that each of the subfamilies deduced by mtDNA analysis (Derome et al., 2002) possesses an intergenic region of distinct length and sequence substructure. The intergene of the single representative of the Gymnodraconinae, *G. acuticeps*, has an insertion of ~830 bp vicinal to the α-globin gene that has not been found in other notothenioids to date but is otherwise similar to those of the Bathydraconinae. The Cygnodraconinae intergenic regions are the shortest in length (1.5-2.3 kb) and vary with respect to several indels.

G. australis stands out from the other species examined in that its globin intergene is exceptionally small, which may indicate that this 1.5-kb segment represents the minimal combination of promoter and enhancer elements necessary to drive notothenioid globin gene transcription. The maintenance of a truncated segment of the 1-C sequence element by the *G. australis* intergenic region (and partial to complete 1-C sequences in all other bathydraconids) reinforces our conclusion, deduced from analysis of the *N. coriiceps* and *N. angustata* intergenes, that it contains an important transcriptional regulatory element. However, the *G. australis* intergene supported only low levels of transcription with respect to *N. coriiceps*, even though the former contains two additional NF-E2 sites near the proximal promoter of the β-globin gene (Fig. 3). This discrepancy may be explained by the absence of the CCACC elements in the intergenes of the Cygnodraconinae (Fig. 5).

The *A. nudiceps* intergenic region, by contrast, is quite similar in structure to that of *D. mawsoni* and of *N. angustata*, the major difference being a truncation of the 1-A element (Figs. 2 and 3). Despite this overall similarity, the *A. nudiceps* intergene functions as a strong promoter/enhancer only in α orientation (Fig. 6). Our working hypothesis is that this differential activity might be due to the gain of one GATA motif immediately upstream of the α-globin coding sequence (Fig. 3) with respect to those of *N. coriiceps* and of *D. mawsoni*. Alternatively, the cause of the high-level α-transcriptional activity of the *A. nudiceps* intergene and its negligible activity in β orientation may reside in small indels in promoter-proximal sequences with respect to those of the other species.

Our luciferase reporter assays of the *N. coriiceps*, *A. nudiceps*, and *G. australis* intergenes were reproducible in multiple, independent experiments, which gives us confidence that we are not reporting artifactual results. We have validated the sequences of our expression plasmids both after initial construction and by re-sequencing after our functional assays. Thus, we face a conundrum; all dragonfishes express hemoglobins in their red blood cells, yet their α/β intergenic sequences do not appear sufficient, in and of themselves, to support its production. How may we reconcile the evolutionary conservation of most of the 1-C element by bathydraconids and nototheniids, a clear phylogenetic footprint, with the functional variability observed for the bathydraconid intergenes in our MEL cell transcription-reporter assay?

4.3. Notothenioid intergenes and the locus control region (LCR)

The minimal transcription-promoting activity of the *G. australis* globin intergene, and the directionally biased activity of the *A. nudiceps* region (strong in α orientation, nonexistent in β) demonstrate that the globin intergenic region alone cannot be solely responsible for the coordinated transcription of the globin genes in notothenioid fishes. In mammals and birds, the developmentally regulated expression of the genes of the α- and β-globin gene complexes is regulated by locus control regions (LCRs) located at a distance (tens of kb) from their respective globin genes (Dean, 2006; Liang et al., 2008; Miele and Dekker, 2008); the LCRs interact with regulatory sequences proximal to the coding sequences to confer tissue- and developmental-stage-dependent transcription. Recently, the linked α/β-globin gene pairs of two model fishes, medaka and zebrafish, have also been shown to be regulated by LCRs (Maruyama, 2004; Hsia et al., 2006; Maruyama, 2007). Therefore, we propose that the dragonfish intergenes, which contain the conserved 1-C element, must interact with an as-yet-to-be-identified notothenioid LCR to confer regulated globin gene expression *in vivo*. The work presented here sets the stage for identification and characterization of the notothenioid LCR.

Contributions

We gratefully acknowledge the logistic support provided to our Antarctic field research program, performed at Palmer Station and in the seas of the Palmer Archipelago, by the staff of the Office of Polar Programs of the National Science Foundation, by the personnel of Raytheon Polar Services Company, and by the captains and crews of the *ARSV Laurence M. Gould* and the *RVIB Nathaniel B. Palmer* (ICEFISH 2004 Cruise, www.icefish.neu.edu). We also thank Christopher D. Jones and the US Antarctic Marine Living Resources Program for providing additional support of our collection of specimens from the South Shetland Islands and the officers and crew of the *R/V Yuzhmorgeologiya* for logistic field support. Dr. Christina Cheng (University of Illinois, Urbana) generously provided the *D. mawsoni* genomic DNA library in Lambda FIX II. MEL cells were generously provided by Dr. Stuart Orkin, Dana-Farber Cancer Institute, Harvard Medical School. A number of samples were provided by Dr. Thomas J. Near, Yale University (refer to table x). In addition to the sequencing *P. charcoti* intergene, BAC library screening and sequencing of *N. coriiceps* and *D. mawsoni* intergene were completed by Ms. Sandra Parker. Ms. Katie O'Neill and Mr. Corey Allard provided general laboratory support.This work was supported by NSF grants OPP-0132032, OPP-0336932, and ANT-0635470 to H.W.D.

Chapter 2

Bathydraconid Phylogeny based on Globin Genes

Abstract

Channichthyidae (icefish), one of seven subfamilies within the notothenioid suborder, are the only vertebrates that do not produce hemoglobin. Bathydraconidae (dragonfish), the closest red-blooded relative to the icefish, act as an intermediate between the hemoglobin-less icefish and the rest of the notothenioid suborder. While the monophly of icefish is well established, morphological and genetic data indicate that dragonfish are paraphyletic. Utilizing the globin gene cluster (α-globin, intergene, and β-globin), by constructing phylogenies from 13 species and 10 genera of Bathydraconidae, this study aims to resolve the evolutionary relationship among the dragonfish.

Our phylogenetic analyses of the dragonfish generally agree with the subfamilies previously proposed by mitochondrial data, however, our study indicates that Cygnodraconinae and Gymnodraconinae are more closely related and Bathydraconinae should be split into two subclades. The icefish representative *N. ionah* was found to have diverged from the dragonfish, after the initial split of the dragonfish family such that N. ionah is more closely related to Cygnodraconinae and Gymnodraconinae. β-globin pseudogenes also exist in a number of dragonfish lineages. This raises the question about the history of the β-globin gene family within dragonfish as well as other notothenioids. We conclude that dragonfish can be resolved into four groups and that the icefish are more related to a subset of the dragonfish family.

Introduction:

The Southern Ocean surrounding Antarctica began to cool approximately 41 to 30 million years ago (Mya) following the opening of the Drake Passage between South America and the Antarctic Peninsula which established the Antarctic Circumpolar Current (DeWitt, 1971; Eastman, 1993; Scher and Martin, 2006). By the time the Southern Ocean attained its present frigid temperatures (–2 to +2ºC) approximately 14-10 Mya, most temperate, shallow-water marine teleosts, had become locally extinct (Eastman, 2005). This lead to an adaptive radiation of the Notothenioidei fish, including the families Channichthyidae (icefish) and Bathydraconidae (Antarctic dragonfish), which are now the predominant fish fauna of Antarctic coastal waters (Eastman, 2005).

The Channichthyidae (icefish), which lack hemoglobin (Ruud, 1954; Tokita et al., 2002), is one of seven subfamilies within the notothenioid suborder and is most closely related to Bathydraconidae (dragonfish). Icefish typically contain remnants of the adult α-globin gene, while the adult β-globin gene has been lost (Cocca et al., 1995; Zhao et al., 1998). Thus, hemoglobin expression by icefishes no longer occurs due to the absence of intact globin gene loci. Icefish live in cold oxygenated waters and have evolved adaptations to adjust for oxygen transport such as increased blood volume and cardiac output, enlarged hearts, scaleless bodies for cutaneous uptake of oxygen, and increased gill surface area (Eastman, 1993; Eggington, 1996; Kock, 2004; Sidell and O'Brien, 2006). It is proposed, that due to the extremely low temperatures of Antarctic waters, blood with high hematocrit (proportion of cells, including erythrocytes, in blood to overall blood volume) would be very viscous, thus posing a substantial energetic

challenge to the notothenioids (Wells et al., 1990). As a group, Notothenioidei have relatively low hematocrits with icefish being the extreme case. Bathydraconids the closest red-blooded (hemoglobin possessing) relative of icefishes, have reduced hematocrit levels relative to other species in the Nototheniidae family (Kunzmann et al., 1991). Dragonfish act as a globin expressing intermediate between icefish and the rest of the Notothenioidei sub-order making them an ideal group to study the evolution of globin gene loss in icefish.

The monophyly of the icefishes is well established based on both morphological and mitochondrial DNA (mtDNA) sequence data (Near et al., 2003), however the relationship within the dragonfish remains unresolved and paraphyletic (species do not all share the same last common ancestor) based on both molecular and morphological data (Derome et al., 2002; Eastman and Lannoo, 2003; Iwami, 2004). MtDNA sequence data indicates there are three subclades within the Bathydraconidae, the Bathydraconinae, the Gymnodraconinae, and the Cygnodraconinae, whereas α-globin data shows two major groups (Derome et al., 2002; Near et al., 2006). Morphological data from jaw musculature and brain structure supports two groups, but do not agree on the species grouping (Eastman and Lannoo, 2003; Iwami, 2004). Although studies have not agreed upon the number of subgroups needed and which species should be grouped together, it is clear that this paraphyletic group must be resolved into its monophyletic constituents (Derome et al., 2002; Eastman and Lannoo, 2003; Iwami, 2004; Near et al., 2006).

Here I aim to expand upon the globin genes of dragonfish and to help resolve the current paraphyletic status of the Bathydraconidae, and analyze its relation to the icefish using the α-globin, β-globin, and intergenic phylogenetic data. Of the 16 species in 11

genera of Bathydraconidae, I am including phylogenetic analyses for three gene regions for 13 species and 10 genera in my study to represent Bathydraconidae (refer to table 1).

Methods:

2.1. Collection of fishes and storage of tissues

The amount of biological information known about bathydraconids is limited, but we know that they do not possess swim bladders, and are generally found between 500-700 m depths (Gon and Heemstrea, 1990). Specimens of dragonfishes (Bathydraconidae: Notothenioidei) were collected by bottom trawling in the Antarctic Peninsula region, the Ross Sea, or the Weddell Sea over the course of several Antarctic expeditions (table 1). Following euthanasia of the fishes, spleens from some animals were dissected, flash frozen in liquid nitrogen, and stored at –80 ºC. Pectoral fin clips were stored in 95% ethanol at –20 ºC.

2.2. Bathydraconid genomic DNAs

Genomic DNAs of *Akarotaxis nudiceps* (UTTC1795)*, Bathydraco macrolepis* (UTTC1787)*, Bathydraco marri* (UTTC1712, YFTC 13884)*, Bathydraco scotiae* (YFTC 13892), *Bathydraco antarcticus* (YFTC 13912), Cygnodraco *mawsoni* (TJN2050)*, Gerlachia australis* (UTTC1722)*, Gymnodraco acuticeps* (HWD116)*, Parachaenichthys charcoti* (TJN2009)*, Parachaenichthys georgianus* (HWD 118)*, Pogonophyrne scotti* (YFTC2025), *Prionodraco evansii* (TJN2300), *Psilodraco breviceps* (YFTC 11037) *Racovitzia glacialis* (UTTC 1789) and *Vomeridens infuscipinnis* (HWD 117) were obtained from spleen samples archived at –80 ºC (HWD) or from ethanol-fixed fin clips (TJN, UTTC) using methods previously described (Near et al., 2006) (Table 1).

2.3. Cloning of the bathydraconid globin loci

Because the adult α-and β-globin genes of notothenioids are tightly linked 5' to 5' (<5 kb between start codons), their intergenic regions, which we define as the sequences that separate the start codons, were amplified using the Polymerase Chain Reaction (PCR) and forward and reverse primers derived from the first exons of each gene pair (figure 1). Table 1b gives the sequences of each primer, and Table 1a indicates the primer pairs used for amplifiying the intergene of each species. α- and β-globin genes were amplified in separate reactions according to primers pairs in Table 1a. PCR reactions (50 μl total volume) contained 200 μM dNTPs, 0.5 μM of each primer (forward from the α-exon 1 and reverse from β-exon 1), 0.5-1.0 mM $MgCl_2$, 1-2 μl of DMSO, 1× Advantage 2 PCR buffer, 0.6× Advantage 2 PCR polymerase, 0.1-0.3 μg of genomic DNA and sterile water to volume. Touchdown PCR (Don, 1991; Roux, 1995b) was performed for 10 cycles using the following parameters: 1) denaturation at 95 ºC for 30 s; 2) primer annealing for 1 min, ramping temperature from 65 °C to 55 ºC in one degree increments/cycle; and 3) extension at 68 °C for 5 min. Subsequently, 20 cycles of conventional PCR were performed using the following profile: 1) denaturation at 94 °C for 30 s; 2) primer annealing at 58 °C for 1 min; and 3) extension at 68 °C for 5 min. PCR products were electrophoresed on 1% agarose gels, and the desired DNA fragments were purified by use of the Qiagen Qiaquick gel purification kit.

DNA fragments were subcloned by ligation into the PGEM-T Easy vector (Promega, Madison, WI), and the recombinant plasmids were transformed into competent NEB 5-alpha *E. coli* cells (New England BioLabs, Ipswich, MA). Transformed cells were spread onto LB-Amp plates and grown overnight in the presence of X-Gal and

IPTG. White colonies were picked and grown overnight in LB (Ruther, 1980). Plasmids were purified from the liquid cultures following the Wizard Plus Miniprep (Promega, Madison, WI) protocol. Plasmid inserts were sequenced by automated methods on both strands by the University of Maine DNA Sequencing Facility or by MWG-The Genomic Company (Huntsville, AL). With the exception of *V. infusicpinnis*, all species with successfully sequenced intergenes, were found to contain globin and intergenic region overlap. Species with unknown intergenes, had α and β-globin genes amplified by internal α and β-globin primers (see table 1).

2.4. Compilation of sequence data

BioEdit (Hall, 1999) was used to compile all sequence data, then CLUSTAL (Higgins and Sharp, 1988) was used to align the α, β-globin gene, and intergenic DNA sequences, and manual adjustments were made by eye. Additional sequence data was obtained from the NCBI database (table 1) and sequence files with intergenes were deposited in GenBank (refer to Chapter 1 section 2.4). All gaps which were shared by two or more species and greater than 3 nucleotide lengths were coded as a single nucleotide change by editing sequences to show a nucleotide transition for α- and β-globin, and the intergenic region. MEGA was used to calculate GC contents and lengths of exons, introns, and intergenes are listed in figure 7.

2.5. Phylogenetic analysis

Due to the fact that icefish are known to lack the β globin gene, it was evident that the selective pressure of the α and β globin genes would likely be different, and thus, gap

coded α and β globin genes (inclusive and exclusive of introns), and the intergenic region were analyzed separately. α-globin data included the nototheniids: *N. coriiceps, N. angustata,* icefish: *N. ionah,* Artedidraconid: *P. scotti* and the dragonfishes: *A. nudiceps, B. antarcticus, B. marri, B. scottiae, C. mawsoni, G. australis, G. acuticeps, P. charcoti, P. georgianus, P.evansii, P. breviceps,* and *V. infucipinnis*. Analysis of β-globin data omitted *N. ionah* because it was a pseudogene, and also *G. acuticeps* after the ML topology was found to be compromised due to the long branch of *G. acuticeps*. Intergenic data did not contain all available species due to difficulty in amplification of certain intergenes and therefore excluded Artedidraconid: *P. scotti,* icefish: *N. ionah,* and the dragonfishes: *B. antarcticus, B. marri, B. scottiae,* and *P. breviceps*. Phylogenies were constructed for each gene utilizing the maximum likelihood (ML), maximum parsimony (MP) and neighbor-joining (NJ) methods. In addition, α-and β-globin gene alignments, which included exon and intron data, and exon data alone were analyzed. NJ searches were run using Kimura-2 parameters. Maximum parsimony searches were conducted with random stepwise addition (n=500) and Tree bisection-reconnection (TBR) branch swapping. For the maximum likelihood searches, MODELTEST 3.7 was used to determine the optimal molecular model for the ML analysis (Posada and Crandall, 1998). Using the estimated model parameters, ML search was conducted with 10 random-addition heuristic searches, and nearest-neighbor interchange (NNI) branch swapping methods for all three analyses. Bootstrap support values were taken for all trees estimated from 500 replications. For α-and β-globins, inclusion of intron data did not change the branch relationship among taxa, but did decrease resolution (in terms of bootstrap support), most likely due to homoplasies.

Likewise, MP, NJ and ML trees produced similar topologies. Therefore, only ML phylogenies for each of the three gene regions α, β-globin, and the intergene excluding introns are presented here.

Results

3.1. Gene sequence variation

α, β-globins and intergenic sequences analyzed through MEGA, indicated low sequence variation in GC content and exon lengths (figure 7). Sequence data indicates that *B. marri*, *B. scotiae*, and *B. antarcticus* contain β-globin pseudogenes in addition to the functional β-globin genes. The β-pseudogene of the three bathydraco species contains a 10 bp deletion in exon-2. The position of these psuedogenes is uncertain in respect to the real β-globins, which are adjacent to the intergene.

Intergenic regions varied in length from 1.5 to 4.3 kb. Length variation was due to insertions and deletions (indels), which were coded to provide more data in creating phylogenetic trees. The total number of positions analyzed after gaps and missing data were eliminated was 1965.

3.3. Phylogeny of α, β-globins and the intergene

Maximum likelihood trees for α-globin have high bootstrap values supporting the branches (figure 8) of two major clades, and 2 subclades within each major clade of bathydraconidae. Within major clade 1, is subclade 1 (*G. australis*, *C. mawsoni*, *P. charcoti*, *P. georgianus*), and subclade 2 (*G. acuticeps*, *P. breviceps*). Within major clade 2, is subclade 3 (*P. evansii*, *R. glacialis*, *V. infuscipinnis*), and subclade 4 (*A. nudiceps*, *B. marri*, *B. scotiae*, *B. antarcticus*). Although intergenic data is not inclusive of all species used in α-globin analysis (due to amplification difficulties), it has the same groupings with high bootstrap values for its nodes (exception being *P. evansii* and *A. nudiceps* position). The β-globin ML analysis (figure 8), which included the functional

β-globin of *B. antarticus*, *B. marri*, and *B. scotiae*, and excluded their β-globin psuedogenes as well as *N. ionah*, topology is similar to α-globin data but with less resolution. *G. acuticeps* had a long branch which may have confounded results and therefore, the ML search was run without *G. acuticeps*. β-globin data had considerably lower resolution with lack of strong clade structures and low bootstrap values. Inclusion of introns produced considerably lower bootstrap values for α- and β-globin, but did not change the overall topology of the tree.

3.4. Comparison of α, β-globin gene, and intergenic data

The two nototheniids were included to be outgroups and the Artedidraconid *Pogonophryne scotti*, a sister taxa of the dragonfish was included as an extra control (Kawakita et al., 2003; Near et al., 2006). While α-globin displays a ML tree that concurs with *P. scotti* as a sister taxa, β-globin shows discrepancy by grouping *P. scotti* with *P. breviceps*. β-globin does not fully support the clades of α-globin or intergenic data, but does not conflict with the topologies. Bootstrap values for the ML tree of β-globin are generally lower than the α-globin and intergenic branches and lower resolution is seen in the β-globin trees. Although intergenic data is not comprehensive of all species available for α-globin and β-globin gene data, comparison of ML trees shows the intergenic tree generally support the clades proposed in the α-globin (the exception being *P. evansii* and *A. nudiceps*).

3.5. N. ionah an icefish

Immediately after the divergence of the dragonfish α-globin into two major clades, *N. Ionah* α-globin seems to have branched out from major clade 1. This indicates that perhaps the major clade 1 (currently classed as Cygnodraconinae and Gymnodraconinae) bathydraconids are more closely related to icefish than the major clade 2 bathydraconids (currently classed as Bathydraconinae). In ML analysis of β-globin inclusive of the pseudogenes of *N. Ionah* (β-noto), *B. marri*, *B. scotiae* and *B. antarcticus*, it was found the three bathydraco pseudo β-globins were more closely related to *N. ionah* than to the subfamily members of bathydraconinae including the functional β-globin genes of the three bathydracos.

Discussion

4.1. Non-coding sequences

Kawakita et al. have previously described the use of insertions and deletions (indels) for tree resolution in bumble bees (Kawakita et al., 2003). The relatively well resolved branches of each clade for the intergenic region indicate that indels coded as a single nucleotide changes may in fact be informative, since, uncoded intergene produced collapsed branches in trees. Including intron data did not change tree topology, however, bootstrap values and resolution decreased most likely due to homoplasy and lack of selective pressure on introns.

4.2. Phylogeny comparisons

Morphological and molecular studies of the dragonfish have all concluded that the current phylogeny is paraphyletic (McDonald et al., 1992; Derome et al., 2002; Eastman and Lannoo, 2003; Iwami, 2004). Morphological data have shown that in brain morphology, there are two different classes of brain morphology in modern dragonfish (Eastman and Lannoo, 2003), and in jaw musculature, *G. acuticeps* is significantly different from that of the other species within the family (Iwami, 2004). Our α-globin ML tree agrees with most of the studies (the exception being *P. evansii* disagreeing with brain morphology), with the two major clades corresponding to the brain morphology studies and *G. acuticeps* clearly in a separate subclade with *P. breviceps*.

Evidence from previous studies, our intergenic tree topology, and high bootstrap values support the ML tree derived from α-globin sequence. In addition, when *G. acuticeps* is omitted, β-globin data does not conflict with α-globin further indicating α-

globin is likely to be in congruence with the species phylogeny. We had expected our results to most strongly support the three subfamilies Cygnodraconinae, Gymnodraconinae, and Bathydraconinae proposed by mitochondria data (Derome et al., 2002). Although phylogenetic analysis of the globin gene and intergenic data does not conflict with the proposed subfamilies, our phylogenies indicate that dragonfish should be divided into 2 major clades and four subclades. Our α-globin phylogeny indicates that Cygnodraconinae and Gymnodraconinae should be more closely related (indicated by subclade one and two within the major clade 1), and further divides the Bathydraconinae into two subclades (indicated by major clade 2 divided into two subclades). In the mitochondrial study, *Bathydraco marri*, *Bathydraco macrolepis*, *R. glacialis*, and *P. evansii* were used (Derome et al., 2002). Our study was more comprehensive of this group, and included all the species in the mitochondrial study, except *B. macrolepis*, and in addition, included *A. nudiceps*, *B. scotiea*, *B. antarticus* and *V. infuscippinis*. The increased number of species for the Bathydraconinae subfamily is mostly likely the reason for our increased resolution and division of the group into two subclades. Based on our findings, dragonfish can be resolved into two monophyletic groups (group 1, Gymnodraconinae and Cygnodraconinae, and group 2, Bathydraconinae) that agree with the three proposed subfamilies of Derome et al. (2002) and the subfamily Bathydraconinae should be further divided into two sub-groups.

4.3. Low resolution in β-globin data

The β-globin tree, which has low resolution and bootstrap values, does not strongly support α-globin and intergenic trees. β-globin shows inconsistency with *P.*

scotti, a representative of the sister group Artedraconidae, grouping closely with *P. breviceps*, a dragonfish. *G. acuticeps* was found to have long branch length, and was thus omitted, resulting in a tree with higher bootstrap values (it is interesting to note, that *G. acuticeps* jaw musculature is significantly different from other dragonfish). Comparing the topology with the mtDNA and α-globin tree, the Bathydraconinae, Cygnodraconinae, and Gymnodraconinae subfamilies (Derome et al., 2002) are still retained but less distinct clades are visible and support for *V. infuscippinis*, *P. evansii*, and *A. nudiceps* branches is low. However, the β-globin tree does not conflict overall with the α-globin tree indicating that although β-globin displays low resolution, it does not disagree with the proposed dragonfish species phylogeny of the α-globin.

Analysis of pseudogenes has shown the three bathydraco β-globin pseudogenes are more closely related to *N. ionah* than to any of the other functional β-globins of its own subfamily including the functional β-globins of the same species. This is not surprising due to the fact that the obtained β-globin of the bathydraco species was a pseudogene and all the pseudogenes contained a deletion in exon 2 (although the bathydracos deletion was at a different position from *N. ionah*). This indicates that perhaps the pseudogenes originated from a common ancestor. However, further analysis of β-globin pseudogene phylogenies of dragonfish and other families within the Notothenioidae is needed.

It is important to note that, *B. antarcticus*, *B. marr*i, *B. scottiae* and *D. mawsoni* all contain a β-globin pseudogene in addition to a coding β-globin gene, while *N. ionah* (an icefish) only contains β-globin pseudogenes. Thus it is possible that many of the dragonfish have undiscovered second β-globin genes and the phylogeny could therefore

be a mix of β-globin genes. Considering icefish have completely lost functional β-globin, and dragonfish are its close sister group, it is possible that decreased dependence on hemoglobin as an oxygen carrier and thus low selective pressure on the globin genes, specifically β-globin, has caused an increase in corrupt β-globin. Further studies on rates of nonsynonymous vs synonymous rates of mutation may provide insight on selection pressure of β-globin. cDNA sequencing may also be useful in confirming the functionality of the β-globin sequences we've obtained.

4.4. Relationship between icefish and dragonfish

N. ionah, an icefish which has partial α-globin and two β-globin pseudogenes (β-noto and β-chan) was used to determine where icefish diverged from its sister group, the dragonfish. α-globin data confirms it is closest to the dragonfish, but is more closely related to the subfamilies Cygnodraconinae and Gymnodraconinae than the Bathydraconinae. This is not surprising, since the icefish (hemoglobinless) are thought to be closest to the dragonfish (decrease hemoglobin expression), however, it is seems that icefish have evolved from a particular group of dragonfish, after the dragonfish diverged into two groups.

β-globin of *N. ionah* could not be used for species phylogeny due to the fact that *N. ionah* possesses two pseudogenes and no functional β-globin gene. However, preliminary analysis of β-noto (the longer pseudogene) has implicated that it closer to the three bathydraco pseudogenes indicating that β-noto of icefish and the β –globin pseudogenes of dragonfish share a common ancestor. Further analysis will be required, in search of other β-globin pseudogenes both within the dragonfish family as well as the

other families within the suborder of Notothenioidae to determine the relationship β-globin pseudogenes found in the three bathydracos and *N. ionah*.

Conclusion:

Our studies show paraphyly in Bathydraconidae and support the splitting of the group into 2 major groups, to resolve the paraphyly. In addition, further resolution by dividing the current Bathydraconinae subfamily into two subfamilies was obtained by increasing the number of Bathydraconinae species analyzed. The icefish *N. ionah* α-globin data indicates that icefish are most closely related to dragonfish, but are specifically closer to the Cygnodraconinae and Gymnodraconinae subclades than to the Bathydraconinae group of dragonfish. However, more comprehensive studies of molecular and morphological data should be utilized to reclassify this group. Decreased resolution in β-globin dragonfish trees suggest that the decrease dependence on hemoglobin as an oxygen carrier may have allowed more rapid and independent divergence of the globin genes. Given the existence of β-pseudogenes, further studies will be needed for β-globin data, to determine if psuedogenes exist for the other dragonfish, as well as what drives the diversity of the β-globin genes. The possibility of more than one β-globin in each species will need to be further explored to give a better understanding of the phylogenies derived from the β-globin data.

Contributions:

We gratefully acknowledge the logistic support provided to our Antarctic field research program, performed at Palmer Station and in the seas of the Palmer Archipelago, by the staff of the Office of Polar Programs of the National Science Foundation, by the personnel of Raytheon Polar Services Company, and by the captains and crews of the *ARSV Laurence M. Gould* and the *RVIB Nathaniel B. Palmer* (ICEFISH 2004 Cruise, www.icefish.neu.edu). We also thank Christopher D. Jones and the US Antarctic Marine Living Resources Program for providing additional support of our collection of specimens from the South Shetland Islands and the officers and crew of the *R/V Yuzhmorgeologiya* for logistic field support. Dr. Christina Cheng (University of Illinois, Urbana) generously provided the *D. mawsoni* genomic DNA library in Lambda FIX II. A number of samples were provided by Dr. Thomas J. Near, Yale University (refer to table x). In addition to the sequencing *P. charcoti* intergene, BAC library screening and sequencing of *N. coriiceps* and *D. mawsoni* intergene a number of the α- and β-globin genes were completed by Ms. Sandra Parker. Mr. Corey Allard assisted in sequencing a number of the α- and β-globin genes. Dr. Steven Vollmer assisted with phylogenetic analysis. This work was supported by NSF grants OPP-0132032, OPP-0336932, and ANT-0635470 to H.W.D.

Bibliography:

Amemiya, C.T., Ota, T.Litman, G.W., 1996. Construction of P1 artificial chromosome (PAC) libraries from lower vertebrates. In: Lai, E. and Birren, B. (Lai, E. and Birren, B.s), Analysis of Nonmammalian Genomes. Academic Press, New York, pp. 223-256.

Bargelloni, L., Marcato, S.Patarnello, T., 1998. Antarctic fish hemoglobins: Evidence for adaptive evolution at subzero temperature. *PNAS* 95, 8670-8675.

Buckley, B.A.Hofman, G.E., 2004. Magnitude and duration of thermal stress determine kinetics of hsp gene regulation in the goby Gillichthys mirabilis. *Physiological and biochemical zoology* 77, 570-581.

Chen, L., Devries, A.L.Cheng, C.C., 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *PNAS* 94, 3811-3816.

Cheng, C.H.Chen, L., 1999. Evolution of an antifreeze glycoprotein. *Nature* 401, 443-444.

Cheng, C.H.Detrich, H.W., III, 2007. Molecular ecophysiology of Antarctic notothenioid fishes. *Philosophical transactions of the royal society of London. Series B, Biological sciences* 362, 2215-2232.

Cocca, E.et al., 1995. Genomic remnants of α-globin genes in the hemoglobinless antarctic icefishes. *Proc. Natl. Acad. Sci USA* 92, 1817-1821.

D'avino, R.di Prisco, G., 1987. *Antarctic fish hemoglobin: an outline of the molecular structures and oxygen binding properties-I. molecular structure* 908, 579-584.

Danke, J.et al., 2004. Genome resource for the indonesian coelacanth, *Latimeria menadoensis*. *Journal of Experimental Zoology Part A: Comparative Experimental Biology* 301A, 228-234.

Derome, N., Chen, W., Dettai, A., Bonillo, C.Lecointre, G., 2002. Phylogeny of Antarctic dragonfishes (Bathydraconidae, Notothenioidei, Teleostei) and related families based on their anatomy and two mitochondrial genes. *Molecular Phylogenetics and Evolution* 24, 139-152.

DeWitt, H.H., 1971. Coastal and deep-water benthic fishes of the Antarctic.

di Prisco, G., 2000. Life style and biochemical adaptation in antarctic fishes. *Journal of Marine Systems* 27, 253-265.

di Prisco, G., Eastman, J.T., Giordano, D., Parisi, E.Verde, C., 2007. Biogeography and adaptation of Notothenioid fish: Hemoglobin function and globin-gene evolution. *Gene*

di Prisco, G., Pisano, E.Clarke, A., 1998. Fishes of Antarctica: A biological overview. Springer, pp. 343-349.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K.Mattick, J.S., 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 19, 4008.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., Mattick, J.S., 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucl. Acids Res.* 19, 4008.

Eastman, J.T., 1993. Antarctic fish biology: evolution in a unique environment. Academic Press, San Diego.

Eastman, J.T., 2005. The nature of the diversity of Antarctic fishes. *Polar biology* 28, 93-107.

Eastman, J.T.Lannoo, M.J., 2003. Diversification of Brain and Sense Organ Morphology in Antarctic Dragonfish (Perciformes: Notothenioidei: Bathydraconidae). *Journal of Morphology* 258, 130-150.

Eggington, S., 1996. Blood rheology of Antarctic fishes: viscosity adaptations at very low temperature. *Journal of Fish Biology* 48, 513-521.

Gon, O.Heemstrea, P.C., 1990. Fishes of the Antarctic Ocean. J.L.B. Smith Institute of Ichthyology, Grahamstown.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41, 95-98.

Higgins, D.G.Sharp, P.M., 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237-244.

Hofman, G.E., Buckley, B.A., Airaksinen, S., Keen, J.E.Somero, G.N., 2000. Heat-shock protein expression is absent in the antarctic fish Trematomus bernacchii (family Nototheniidae). *Journal of Experimental Biology* 203, 2331-2339.

Iwami, T., 2004. Comparative morphology of the adductor mandibulae musculature of notothenioid fishes (Pisces, Perciformes). *Antarctic Science* 16, 17-21.

Kawakita, A., Sota, T., Ascher, J.S., Ito, M., Tanaka, H.Kato, M., 2003. Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (Bombus). *Mol. Biol. Evol.* 20, 87-92.

Kock, K.H., 2004. Antarctic icefishes (Channichthyidae): a unique family of fishes. A review. Part II. *Polar biology* 28, 897-909.

Kunzmann, A., Caruso, C.Prisco, G., 1991. Haematological studies on a high-Antarctic fish: *Bathydraco marri* Norman. *J. Exp. Mar. Biol. Ecol.* 152, 243-255.

Lau, D.T., Saeed-Kothe, A., Parker, S.K.Detrich, H.W., III, 2001. Adaptive Evolution of Gene Expression in Antarctic Fishes: Divergent Transcription of the 5'-to-5' Linked Adult α1– and β–Globin Genes of the Antarctic Teleost *Notothenia coriiceps* is Controlled by Dual Promoters and Intergenic Enhancers. *Amer. Zool.* 41, 113-132.

McDonald, M.A., Smith, M.H., Smith, M.W., Novak, J.M., Johns, P.E.Devries, A.L., 1992. Biochemical systematics of notothenioid fishes from Antarctica. *Biochemical systematics and ecology* 20, 233-241.

Myers, R.M., Tilly, K.Maniatis, T., 1986. Fine structure genetic analysis of a beta-globin promoter. *Science* 232, 613-618.

Near, T.J., Parker, S.K.Detrich, H.W., III, 2006. A Genomic Fossil Reveals Key Steps in Hemoglobin Loss by the Antarctic Icefishes. *Mol Biol Evol* 23, 2008-2016.

Near, T.J., Pesavento, J.J.Cheng, C.C., 2003. Mitochondrial DNA, morphology, and the phylogenetic relationships of Antarctic icefishes (Notothenioidei: Channichthyidae). *Molecular Phylogenetics and Evolution* 28, 87-98.

Orkin, S.H., 1992. GATA-binding transcription factors in hematopoietic cells. *Blood* 80, 575-581.

Osoegawa, K.et al., 1998. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52,

Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H.Costantini, F., 1995. Development of hematopoietic cells lacking transcription factor GATA-1. *Development* 121, 163-172.

Posada, D.Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

Roux, K.H., 1995a. Optimization and troubleshooting in PCR. *Genome Research* 4, 185-194.

Roux, K.H., 1995b. Optimization and troubleshooting in PCR. *Genome Res.* 4, 185-194.

Ruther, U., 1980. Construction and properties of a new cloning vehicle, allowing direct screening for recombinant plasmids. *Molecular and General Genetics MGG* 178, 475-477.

Ruud, J.T., 1954. Vertebrates without erythrocytes and blood pigment. *Nature* 173, 848-850.

Scher, H.D.Martin, E.E., 2006. Timing and Climatic Consequences of the Opening of Drake Passage. *Science* 312, 428-430.

Sidell, B.D.O'Brien, K.M., 2006. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefish. *Journal of Experimental Biology* 209, 1791-1802.

Tokita, M., Ishii, S.Iwami, T., 2002. Phylogenetic analysis of Antarctic notothenioid fishes based on two-dimensional gel electrophoresis. *Polar Biology* 25, 163-168.

Wells, R.M.G., Macdonald, J.A.di Prisco, G., 1990. Thin-blooded antarctic fishes: a rheological comparison of the hemoglobin-free icefishes, *Chionodraco kathleenae* and *Cygnodraco antarcticus*, with a red-blooded nototheniid, *Pagothenia bernacchii*. *Journal of Fish Biology* 17, 517-527.

Zhao, Y.et al., 1998. The Major Adult alpha -Globin Gene of Antarctic Teleosts and Its Remnants in the Hemoglobinless Icefishes. Calibration of the Mutational Clock for Nuclear Genes. *J Biol. Chem* 273, 14745-14752.

Table 1

Primer sequences used the amplify globin intergenic regions during first-round cloning

Primer	Sequence
A1ASL	GATGCACTCACCTGCTCAGAGCATC
AG1	CACAGAGCCTTGACTGCTGCCTTGT
AG2	GACTTGCYGATCTTGCTCCACAGAGC
B1AS2L	CTGCTGAGAGCCTTGGGTCCGATGT
BG1	ATGTGAGCTCGGTTTTGCTCCACT
BRALL	GCACAAGCTGCTGGTTTAGTG
CMARF	GTTTCCTTCACACTCAGAAG
OBR4	TTAGTGGTACTGCTTTCCCAGG

Table 2

Forward and reverse primer pairs used for first-round intergene amplification

Species	Forward primer	Reverse primer
A. nudiceps	AG1	BG1
C. mawsoni	A1ASL	B1AS2L
G. australis	A1ASL	BRALL
G. acuticeps	A1ASL	B1AS2L
P. charcoti	A1ASL	B1AS2L
P. georgianus	A1ASL	B1AS2L
V. infuscipinnis	AG1	BG1

Table 3

Primers sequences used to amplify the intergenic regions during second-round cloning

Primer	Sequence
AustAblg2	CGCAGATCTCTTGCTTGCTTATTTCTTCCT
AustBbgl2	GCGAGATCTGGTGGCTGATGTTTAATGGAT
AustAnheI	CGCGCTAGCCTTGCTTGCTTATTTCTTCCT
AustBnheI	GTAGCTAGCGGTGGCTGATGTTTAATGGAT
NudAavrII	CGGCCTAGGCTTGCTTGCTTATTTCTTCCT
NudBavrII	GTGCCTAGGGATGGCTGATGTTTAATGGGT
NudAbamHI	CACGCGGATCCCTTGCTTGCTTATTTCTTCCT
NudBbamHI	GTCTCGGATCCGATGGCTGATGTTTAATGGGT

Table 4

Forward and reverse primers for amplification of second-round intergene amplification

Species and direction	Forward primer	Reverse primer	Digest enzymes
A. nudiceps α	NudAbamHI	NudBavrII	*Bam*HI, *Avr*II
A. nudiceps β	NudAavrII	NudBbamHI	*Avr*II, *Bam*HI
G. australis α	AustAnheI	AustBbgl2	*Nhe*I, *Bgl*II
G. australis β	AustAblg2	AustBnheI	*Bgl*II, *Nhe*I

Table 5

1A Sample source, locations, species, family, accession numbers and primers used for PCR products. PCR products and primer pairs in bold are newly sequenced, while non-bolded PCR products were completed in previous studies. [1] Denotes Bathydraconidae species in subfamily Gymnodraconinae. [2] Denotes Bathydraconidae species in subfamily Cygnodraconinae. [3] Denotes Bathydraconidae species in the subfamily Bathydraconinae.

1B name of primer pairs and corresponding primers and sequences used for amplification of α-globin, β-globin and intergenic regions.

Fig. 1. Nucleotide sequence alignments for nototheniid and bathydraconid globin intergenic regions. Potential TATA, CCAAT, EKLF, GATA, CCACC, C-Myb, and NF-E2 motifs are indicated. Abbreviations: Anud, *A. nudiceps*; Cmaw, *C. mawsoni*; Dmaw, *D. mawsoni*; Gacu, *G. acuticeps*; Gaus, *G. australis*; Nang, *N. angustata*; Ncor, *N. coriiceps*; Pcha, *P. charcoti*; Peva, *P. evansii*; Pgeo, *P. georgianus*; Vinf, *V. infuscipinnis*.

Fig. 2. Structure of the adult globin intergenic regions from three nototheniids. The direct repeats of the *N. coriiceps* intergene (NcDR1 and NcDR2) and their subregions (A, B, and C) are shown; *N. angustata*, and *D. mawsoni* lack the 1.1-kb duplication that contains NcDR2. Mapped onto the globin intergenic regions are potential *cis*-acting regulatory elements: T, TATA box; C, CCAAT box; E, EKLF motif; G, GATA motif; K, CCACC (reverse compliment of GGTGG) elements recognized by CAC-binding proteins and Sp1; M, C-Myb motif; N, NF-E2 motif. These intergenes support divergent transcription of the α- and β-globin genes (left and right respectively.).

Fig. 3. Structure of the adult globin intergenic regions of eight bathydraconids. The dragonfish intergenes fall into three size classes that correspond to three proposed subfamilies: Gymonodraconinae (*G. acuticeps*); Bathydraconinae (*P. evansii*, *A. nudiceps*, *V. infuscipinnis*); and Cygnodraconinae (*C. mawsoni*, *P. charcoti*, *P. georgianus, G. australis*). Potential *cis*-acting regulatory regions are indicated (for abbreviations, see legend to Fig. 1.) The intergenic regions of *N. coriiceps* and of *N. angustata* are shown for comparison.

Fig. 4. Sequence identity matrix for globin intergenic regions from 11 notothenioid fishes Pairwise sequence identities for three nototheniids and eight bathydraconids (subclades Gymnodraconinae, Bathydraconinae, and Cygnodraconinae) were scored using CLUSTALW (section 2.4). Bold font shows sequence identities within each taxon. Abbreviations: Anud, *A. nudiceps*; Cmaw, *C. mawsoni*; Dmaw, *D. mawsoni*; Gacu, *G. acuticeps*; Gaus, *G. australis*; Nang, *N. angustata*; Ncor, *N. coriiceps*; Pcha, *P. charcoti*; Peva, *P. evansii*; Pgeo, *P. georgianus*; Vinf, *V. infuscipinnis*.

Fig. 5. Sequence comparison of putative nototheniid and bathydraconid enhancer motifs found in the 1-C intergenic element. Yellow highlights CCACC motifs (reverse compliment of GGTGG), and green highlights GATA elements. Grey shading shows intergenic sequence variation, including deletions (dashes). Species names are abbreviated as in Fig. 3.

Fig. 6. Activities of the wild-type globin intergenic regions of *N. coriiceps*, *A. nudiceps*, and *G. australis* in a hematopoietic microenvironment. The strengths of the promoters of the intergenic regions were assayed by transfection of MEL cells with the firefly luciferase reporter constructs (see section 2.7). MEL cells were induced to differentiate by treatment with DMSO. To control for variable transduction efficiencies, background-corrected firefly luciferase activities were normalized with respect to the activity of *Renilla* luciferase expressed by the reporter plasmid pRL-SV40. Promoter constructs: Basic, the promoterless/enhancerless vector PGL3-Basic; SV40, the strong SV-40 viral promoter/enhancer in pGL3-Basic; Ncorα, pNcorPα1 with the *N. coriiceps* α1-globin promoter sequences vicinal to the *luc*+ reporter gene; Ncorβ, pNcorPβ with the *N. coriiceps* β-globin promoter sequences adjacent to the *luc*+ reporter gene; Gausα, Gausβ, Anudα, and Anusβ represent reporter plasmids pGausPα, pGausPβ, pAnudPα, and pAnudPβ, respectively. Bars represent the mean ± SEM (n = 3). Asterisks: the activities of the reporter constructs in differentiated MEL cells were significantly greater than in undifferentiated cells ($p < 0.05$, two tailed *t*-test). Squares: the activities of Ncorα, Gausα, and Anudα were significantly different in differentiated MEL cells ($p < 0.05$, one-way ANOVA). Triangles: the activities of Ncorβ, Gausβ, and Anudβ were significantly different in differentiated MEL cells ($p < 0.05$, one-way ANOVA).

Fig.7. Representation of α-globin, β-globin and intergenic region orientation and location of primers. Table corresponds to exons (in black block) numbered 1-3 5’ to 5’ from the intergene outwards and indicates the lengths for the exons of α-globin, β-globin and the intergenic. GC content was calculated by use of MEGA.

Fig.8. Maximum likelihood trees for α-globin, β-globin and intergenic region and corresponding bootstrap values. An α-globin ML tree was created with a JC+I+G model (ln L=1174.0106), an intergenic ML tree was created with a TVM+G model (ln L=12517.5732), and a β-globin ML tree was created with a HKY+I model (ln L=1219.592).