



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Biologia

Scuola di Dottorato di Ricerca in Bioscienze e Biotecnologie

Indirizzo Biologia Evoluzionistica

Ciclo XXVI

**Transcriptomics and population
differentiation in two notothenioid Antarctic
fish**

Trascrittomica e differenziamento di popolazione in due specie di pesci
nototenioidei antartici

Direttore della Scuola: Ch.mo Prof. Giuseppe Zanotti

Coordinatore d'indirizzo: Ch.mo Prof. Andrea Pilastro

Supervisore: Ch.mo Prof. Lorenzo Zane

Dottoranda: Cecilia Agostini

CONTENTS

LIST OF ABBREVIATIONS	5
LIST OF PAPERS.	7
ABSTRACT	9
GENERAL INTRODUCTION	11
AIMS OF THE THESIS.	15
STUDY SPECIES.	16
The <i>Chionodraco</i> genus	16
<i>Pleuragramma antarcticum</i>	18
PAPER EXTENDED ABSTRACTS.	21
Paper I	21
Paper II	23
Paper III	24
Paper IV	24
Paper V	25
Paper VI	27
Paper VII	30
GENERAL CONCLUSIONS.	32
PAPERS.	35
Paper I	37
Paper II	39
Paper III	41
Paper IV	43
Paper V	45
Paper VI	47
Paper VII	49
RIASSUNTO	51
REFERENCES	53
ACKNOWLEDGEMENTS	63

LIST OF ABBREVIATIONS

ABC = approximate Bayesian computation

AFGP = antifreeze glycoprotein

AP = Antarctic Peninsula

ATP = adenosine triphosphate

BLAST = basic local alignment search tool

cDNA = complementary DNA

DNA = deoxyribonucleic acid

e.g. = *exempli gratia*

EST = expressed sequence tag

HSP = heat-shock protein

HSR = heat-shock response

HW = Hardy-Weinberg

i.e. = *Id est*

mRNA = messenger RNA

mtDNA = mitochondrial DNA

MYA = million years ago

PCR = polymerase chain reaction

PhD = Philosophiae Doctor

RNA = ribonucleic acid

SL = standard length

TLP = trypsinogen-like protease

LIST OF PAPERS

This thesis is based on seven papers (three primer notes and four scientific articles), six of which have been published in scientific journals and one is under revision:

- I. Coppe A*, Agostini C*, Marino IAM, Zane L, Bargelloni L, Bortoluzzi S, Patarnello T (2013). **Genome evolution in the cold: Antarctic icefish muscle transcriptome reveals selective duplications increasing mitochondrial function.** *Genome Biology and Evolution*, 5(1): 45-60. DOI: 10.1093/gbe/evs108. *Equally contributing first authors
- II. Marino IAM, Agostini C, Papetti C, Bisol PM, Zane L, and Patarnello T. **Isolation, characterization and multiplexing of expressed sequence tag-linked microsatellite loci for the Antarctic icefish *Chionodraco hamatus*.** In: Molecular Ecology Resources Primer Development Consortium, Agostini C, Agudelo PA *et al.* (2011). Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2010-30 November 2010. *Molecular Ecology Resources*, 11(2), 418-421. DOI: 10.1111/j.1755-0998.2010.02970.x
- III. Papetti C, Marino IAM, Agostini C, Bisol PM, Patarnello T, Zane L (2011). **Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei.** *Conservation Genetics Resources*, 3(2): 259-262. DOI: 10.1007/s12686-010-9336-9
- IV. Agostini C, Marino IAM, Patarnello P, Zane L. **Isolation, characterization and multiplexing of sixteen EST-linked microsatellite loci in the Antarctic silverfish *Pleuragramma antarcticum*.** In: Molecular Ecology Resources Primer Development Consortium, Agostini C, Albaladejo RG *et al.* (2013). Permanent Genetic Resources added to Molecular Ecology Resources Database 1 April 2013-31 May 2013. *Molecular Ecology Resources* 13, 966-968. DOI: 10.1111/1755-0998.12140
- V. Agostini C, Papetti C, Patarnello T, Mark FC, Zane L, Marino IAM (2013). **Putative selected markers in the *Chionodraco* genus detected by interspecific outlier tests.** *Polar Biology*, 36(10): 1509-1518. DOI: 10.1007/s00300-013-1370-0
- VI. Marino IAM, Benazzo A, Agostini C, Mezzavilla M, Hoban SM, Patarnello T, Zane L, Bertorelle G (2013). **Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation.** *Molecular Ecology*, 22(20): 5148-5161. DOI: 10.1111/mec.12458
- VII. Agostini C, Patarnello T, Ashford J, Torres J, Zane L. **Has climate change promoted genetic fragmentation in ice-dependent *Pleuragramma antarcticum*?** *Journal of Biogeography*, under revision

For **Paper I**, I was involved in the annotation of the transcriptome and in the process of identification of duplicated genes; I performed the enrichment analysis using David, I participated in the interpretation of results and I had the main responsibility for writing the

paper. For **Paper II**, I was involved in primer isolation and I performed most of the experimental work. For **Paper III**, I performed the cross-species amplification of primers in the *Chionodraco* genus and I obtained descriptive statistics of cross-amplification. For **Papers IV, V and VII**, I had the main responsibility for experiments, analyses and for writing the paper. For **Paper VI**, I performed most of the experimental work and I was involved in part of the analyses (analysis of the genetic variability level and of the pattern of genetic differentiation using F-statistics and the program Structure).

ABSTRACT

Antarctic notothenioids radiated over past millions of years in extremely cold waters, they display a wide range of adaptations to withstand the cold and now dominate the Antarctic fish fauna. These fish may be extremely vulnerable to climate change with possible cascading effects on the entire Antarctic marine ecosystem. Therefore, crucial tasks are the concomitant study of the genomic basis of cold adaptation, the analysis of differentiation processes resulting from past and present climate change and a close survey of the current level of genetic variation and population structure. We considered four species of the Notothenioidei suborder: the three recently derived species of the *Chionodraco* genus, namely *Chionodraco hamatus*, *Chionodraco rastrispinosus* and *Chionodraco myersi*, and *Pleuragramma antarcticum*. The *Chionodraco* genus belongs to the family Channichthyidae (icefish), unique among vertebrates for the lack of hemoglobin and myoglobin expression in skeletal muscle. Oxygen delivery to tissues is ensured by a marked remodelling of the cardio-vascular system and by exceptionally high mitochondrial densities in the muscle. *P. antarcticum* (Nototheniidae) is the only notothenioid with a complete pelagic life cycle; it is dependent on sea ice and plays a key role in the trophic web of the Antarctic marine ecosystem. Analyses performed in this PhD can be grouped in two major lines of research: 1) the deepening of the knowledge on the genetic and genomic basis of icefish adaptation to the cold; 2) the analysis of patterns of intra- and inter-specific genetic differentiation with particular emphasis on how past and present environmental conditions have shaped and are influencing the fish genetic structure. With regard to the first line of research, we reconstructed and annotated the first normalized transcriptome of *C. hamatus* skeletal muscle and we exploited deep sequencing information of this energy-dependent tissue to test the hypothesis of duplication of genes involved in mitochondrial function. Using a dedicated bioinformatic pipeline we identified 124 duplicated genes specific to the icefish lineage. Significantly more duplicates were found in *C. hamatus* when transcriptome data were compared with whole genome data of model fish. Duplicated genes were significantly enriched in proteins with mitochondrial localization, involved in mitochondrial function and biogenesis. The combination of high mitochondrial densities and the maintenance of duplicated genes involved in mitochondrial function might confer a selective advantage in cold conditions and in the absence of oxygen-carrying proteins, by improving oxygen diffusion and energy supply to aerobic tissues. With regard to the second line of research, in the *Chionodraco* genus we investigated the pattern of intra- and inter-specific genetic differentiation. We found intraspecific homogeneity, but three distinct gene pools corresponding to the nominal species. We searched for putative outlier loci detecting a high level of genetic differentiation between the three species and we identified three loci, possibly influenced by natural selection, showing sequence similarity to a calmodulin transcript, an antifreeze glycoprotein/trypsinogen-like protease gene and to the mRNA of a key component of the super elongation complex. Selective pressures acting on specific loci might reflect past evolutionary processes leading to species divergence and local adaptation. The extent and timing of interspecific gene exchange was also considered to clarify the role of glacial cycles in promoting the divergence and the introgression of the *Chionodraco* species. We

found evidences of past and present introgression: several individuals in each species showed mixed ancestry; evolutionary scenarios excluding hybridization or including it only in ancient times had small or zero posterior probabilities; data supported a scenario of interspecific gene flow associated with the two most recent interglacial periods. These findings may indicate an increased opportunity for speciation in allopatric refugia during glacial periods, followed by secondary contacts and hybridization during warmer intervals. With regard to *P. antarcticum*, we investigated its population genetic structure along the Antarctic Peninsula (AP), a region highly impacted by regional warming. We found a single gene pool and absence of inter-annual variability in the south-western AP, while significant differences were detected on a geographic scale from samples collected off the tip of the AP, with a signal of increased fragmentation over time. The reduced level of gene flow along the shelf, the increase of differentiation through time, and the inability to capture *P. antarcticum* in the central-western AP for two consecutive years, all suggest that this sea ice dependent species has been affected by climate change with possible cascading effects on the Antarctic marine food web.

GENERAL INTRODUCTION

The suborder Notothenioidei

The Southern Ocean is one of the most extreme habitats in the world. Antarctic waters are isolated from other water masses by the Antarctic Convergence and constitute a unique marine environment characterized by cold water temperatures (reaching $-1.9\text{ }^{\circ}\text{C}$), ice-coverage for most part of the year, and wide seasonal fluctuations in primary production. The Southern Ocean hosts a unique cold-adapted fauna, including a group of 134 fish species of the perciform suborder Notothenioidei (Lecointre 2012). The suborder is constituted of eight families: the three most basal families (Bovichtidae, Pseudaphritidae and Eleginopidae) comprehend 13 species, which are primarily sub-Antarctic, *i.e.* live outside the Antarctic Convergence; the five most derived families (Nototheniidae, Artedidraconidae, Harpagiferidae, Bathydraconidae and Channichthyidae) constitute the so called Antarctic clade, which account for 121 species (Lecointre 2012). The most part of Antarctic notothenioids are endemic to the Southern Ocean and dominate Antarctic shelf areas in terms of species diversity (46%), abundance (92%) and biomass (91%) (Eastman 2005).

The genetic and genomic bases of cold adaptation

Antarctic notothenioids radiated over the past 22 millions of years (Near *et al.* 2012) in isolated and extremely cold waters and acquired a wide range of adaptations to withstand the cold and reduce energetic requirements. The peculiar thermal history of Antarctic notothenioids resulted in extreme stenothermality of the extant species, which are specialized to cope with a low and narrow range of temperature. In the course of evolutionary time, genes with novel functions have been acquired, while other genes have been amplified or modified to allow optimal function at low temperatures; finally, because of the extreme stability of thermal conditions, some protein-coding genes and gene regulatory mechanisms needed to cope with climate change have been lost. Among newly minted genes remarkable is the acquisition of antifreeze glycoproteins (AFGPs), which prevent ice expansion and freezing of body fluids at subzero temperatures (Raymond and DeVries 1972, 1977). The AFGP gene evolved from a trypsinogen-like protease (TLP) gene, presumably through the *de novo* amplification of a tripeptide (Thr-Ala-Ala) coding element to create the entire AFGP coding region (Chen *et al.* 1997). AFGPs are thought to have evolved once in the clade of Antarctic notothenioids allowing survival of their ancestor in a freezing polar environment and the subsequent diversification of the clade into recurrently opening niches (Near *et al.* 2012). The loss of AFGP has also been documented in some notothenioid species that secondarily occupied warmer sub- or non-Antarctic waters outside the Antarctic Convergence (Cheng and Detrich 2007).

Adaptation to the cold involved modification of pre-existing genes and proteins, such as enzymes (*e.g.* lactate dehydrogenase, Field and Somero 1998), tubulins (Detrich *et al.* 2000) and channel proteins of the endoplasmatic reticulum (Römisch *et al.* 2003); however, adaptation also entailed large genomic rearrangements and regulation of gene

expression. For instance, 189 genes were found to be overexpressed in the Antarctic notothenioid *Dissostichus mawsoni* compared to temperate water teleosts (Chen *et al.* 2008) and 48% of these genes were homologs to genes up-regulated in carp during cold acclimation (Gracey *et al.* 2004), suggesting constitutive up-regulation in Antarctic notothenioids to increase their fitness in cold conditions. Many of these overexpressed genes are involved in antioxidant functions and were found to be duplicated in Antarctic notothenioids indicating a “genomic response” to increase antioxidant defence through recurrent duplications (Chen *et al.* 2008). Notably, antioxidant response could be critical in oxygen-rich Antarctic waters, which expose marine organisms to high levels of reactive oxygen species and oxidative damage (Abele and Pantarulo 2004). On the other hand, gene expression analyses showed the loss of the inducible heat-shock response (HSR) in Antarctic notothenioids since no increased expression of any size class of heat-shock proteins (HSPs) was detected after stress exposure (Hofmann *et al.* 2000, Buckley *et al.* 2004, Buckley and Somero 2009). However, inducible HSP70 was found to be expressed under normal physiological conditions in Antarctic notothenioids, suggesting that HSR has become a constitutive process probably aimed at mitigating problems of protein folding and denaturation due to constant exposure at low temperatures (Place *et al.* 2004, Place and Hofmann 2005, Todgham *et al.* 2007).

Among gene losses remarkable is the absence of hemoglobin and functionally active erythrocytes in the family Channichthyidae (icefish) (Ruud 1954) due to a large deletion of the adult $\alpha\beta$ -globin gene cluster (Cocca *et al.* 1995, Zhao *et al.* 1998). Additionally, notothenioids do not express myoglobin in skeletal muscle and some icefish species also in cardiac myocytes (Sidell *et al.* 1997, Moylan and Sidell 2000, Grove *et al.* 2004, Borley and Sidell 2010); as a consequence, they worsen their hemoglobin-less condition since myoglobin typically assists in storage and delivery of oxygen in muscle tissue (Ordway and Garry 2004). High oxygen solubility at low temperatures has possibly relaxed selective pressure for oxygen-binding proteins; however, a large set of compensatory adaptations has evolved in icefish to ensure adequate oxygen delivery to tissues (Kock 2005a); these include a hypertrophic heart, an increased blood volume, well-perfused gills, cutaneous uptake of oxygen, reduced metabolic rate, and high mitochondrial densities in muscle tissues.

Most of the studies investigating the genetic bases of cold adaptation in Antarctic notothenioids focused on single genes and proteins, while genome-wide resources are still scarce and currently exist for a few species. Whole genome sequencing is still not available for any notothenioid, while sequenced transcriptomes have been reported for six species. Chen and colleagues (2008) used Sanger sequencing to obtain non-normalized EST (expressed sequence tag) libraries of four *D. mawsoni* tissues (brain, liver, ovary and head kidney) and they performed a comparative analysis of transcriptome profiles to identify transcriptional differences between the cold-adapted polar fish and some temperate/tropical teleosts. Shin and colleagues (2012) used 454 technology to sequence normalized cDNA libraries of *Notothenia coriiceps*, *Chaenocephalus aceratus* and *Pleuragramma antarcticum*; a qualitative comparison of the transcriptomes showed that the three Antarctic fish express more ubiquitin-conjugating proteins than temperate fish. Reference

transcriptomes were obtained also for the cryopelagic *Pagothenia borchgrevinki* (Bilyk and Cheng 2013) and for *Trematomus bernacchii*; in the latter species, the tissue specific response to thermal stress was also analysed (Huth and Place 2013). The availability of well-annotated transcriptomes and other genome-wide resources from a variety of Antarctic notothenioids represents the groundwork for the understanding of the genetic and genomic bases underling notothenioid adaptive traits and may give insights into past evolutionary processes leading to cold adaptation.

The evolutionary history and adaptive radiation of Antarctic notothenioids

The successful adaptation and current dominance of the notothenioid lineage in the waters of the Antarctic shelf is tightly linked to tectonic and paleo-climatic events of Antarctica. Approximately 40 million years ago (MYA) Antarctica was not the cold and isolated place that we know today and a cosmopolitan and temperate teleost fish fauna (including herrings, wrasses and billfishes) populated the waters of the shelf (Eastman 2005). Thermal isolation of the continent was achieved by the formation of the Antarctic Circumpolar Current after the opening of the Tasman Seaway and of the Drake Passage (Scher and Martin 2006). The establishment of cold conditions was gradual and started around 35 MYA with a period of global cooling and widespread glaciations (Zachos *et al.* 2001, DeConto and Pollard 2003). Under these new environmental conditions the Eocene temperate fish fauna was gradually extirpated and Antarctic notothenioids, thanks to freezing avoidance allowed by AFGPs, were able to invade newly empty and ice-associated niches (Matschiner *et al.* 2011, Near *et al.* 2012). This group of fish experienced a spectacular radiation to more than 100 species starting from the intense cooling event of Late Miocene, dated approximately 14 MYA, that initiated the formation of subzero polar conditions in Antarctica (Shevenell *et al.* 2004). The common ancestor of Antarctic notothenioids was benthic and the subsequent radiation of the clade probably proceeded through colonization of novel ecological niches along the benthopelagic axis. Alterations of buoyancy, which was achieved by a combination of reduced skeletal mineralization (Eastman 1997, Albertson *et al.* 2010) and the evolution of large lipid deposits (DeVries and Eastman 1978, Eastman 1993), allowed feeding in the whole water column and permitted the occupancy of pelagic, semipelagic, cryopelagic and epibenthic habitats (Eastman 2000, 2005, Eastman and Barrera-Oro 2010). Species diversification during the late Miocene (11.6-5.3 MYA) was possibly influenced by periodical fluctuations in ice activity (Tripathi *et al.* 2009), leading to recurrent extinctions of some populations of the shelf and subsequent colonization of newly exploitable niches. Moreover, repeated advances and retreats of ice sheets on the shelf probably led to geographic isolation and allopatric speciation (Near *et al.* 2012). In a similar way, alternating warming and cooling periods during the Pleistocene glacial cycles could have promoted diversification in some groups and secondary contact and hybridization in others.

For its peculiar paleo-climatic history and current species richness, the Antarctic continental shelf has been recognized as an area of rich marine biodiversity and a “species

flock generator” (Eastman and McCune 2000, Lecointre *et al.* 2013). Antarctic notothenioids meet all the criteria required to be defined a species flock: monophyly, high species diversity, high level of endemism, morphological and ecological diversity, and habitat dominance in terms of biomass (Eastman and McCune 2000). Moreover, inside the clade, at least three nested species flocks were recognized: the Trematominae, the Channichthyidae and the Artedidraconidae families (Lecointre *et al.* 2013). Mechanisms underling this fascinating bulk of biodiversity are still largely unexplored. In this respect, the investigation of past differentiation processes, using different types of molecular markers, may shed light on both neutral and selective processes leading to the divergence of gene pools and the evolution of new species.

Antarctic notothenioids and climate change

Understanding how Antarctic notothenioids responded to past climate events is important also in the light of recent global warming and may help to predict possible future scenarios. However, while past climate changes were slow enough to provide chance for biological adaptation, rates of current warming have largely exceeded those experienced in the past (Pachauri and Reisinger 2008) and may have a strong and possibly devastating impact on the Antarctic marine biodiversity. Antarctic notothenioids have evolved a wide range of adaptations to withstand the cold and thermostable Southern Ocean and cold-adapted genomes possibly entail limited adaptive capacities to environmental change (Hoffmann and Sgrò 2011). Moreover, rising temperature may affect species distribution and abundance, reduce connectivity and effective size of populations and finally decrease the level of genetic variation and thus the potential to adapt to environmental change.

Response of organisms to climate change will depend on the rate of temperature changes but also on the biological and ecological traits of species. Life cycles of Antarctic organisms are tightly linked to the high seasonal variation in productivity that characterizes polar regions. Species are forced to synchronize the peak of their energetic requirements within a narrow time window of the year. Climate change may disrupt the tight trophic interactions between phytoplankton and zooplankton and between predators and prey at different level of the food web, thus impairing the subtle and delicate link between biological and environmental cycles characterizing the Antarctic marine ecosystem (Moline *et al.* 2008). Along the Antarctic Peninsula (AP), which is one of the fastest warming areas on the planet (Turner *et al.* 2005, Zazulie *et al.* 2010), biological responses associated to climate change have already been documented; these responses are mainly consequences of alterations in sea ice dynamics, *i.e.* shifts in the extent of glaciers and duration of sea ice coverage (Clarke *et al.* 2007, Schofield *et al.* 2010).

Concluding, the concomitant study of the genomic basis of cold adaptation, the analysis of differentiation processes resulting from past and present climate change and a close survey of the current level of genetic variation and population structure of Antarctic notothenioids are crucial tasks as these fish may be extremely vulnerable to climate change and their threat and decline may have cascading effects on the entire Antarctic marine ecosystem.

AIMS OF THE THESIS

This PhD thesis is composed of two major lines of research, which represent different but complementary scientific approaches for the study of notothenioid evolutionary history, adaptation to the cold and possible consequences of recent climate change. Firstly, we focused on the genomic bases of peculiar phenotypic traits specific to the icefish lineage, which represent an excellent model for the study of cold adaptation. Then, we moved on to a more recent timescale and searched for loci putatively under selection in the three recently derived species of the *Chionodraco* genus; these loci could reflect past selective forces possibly playing a role in the process of species diversification. We also used neutral loci to evaluate how past climate cycles could have influenced the divergence and the introgression of species. Finally, we moved on to the present time and evaluate the population structure of *P. antarcticum* along the AP, an area highly impacted by regional warming (Turner *et al.* 2005, Zazulie *et al.* 2010).

Part A

The first line of research is aimed at deepening the knowledge on the genetic and genomic basis of icefish adaptation to the cold. In particular, we focused on the skeletal muscle, which is the primary site of energy production in the form of ATP, a process that is supposed to be hard in cold conditions and in the absence of oxygen-binding proteins. First, we reconstructed the normalized transcriptome of *Chionodraco hamatus* skeletal muscle. Then, we exploited deep sequencing information of this energy-dependent tissue to test the hypothesis of duplication of genes involved in mitochondrial function and biogenesis (**Paper I**).

Part B

The second line of research is focused on the pattern of genetic differentiation at the population level and between the investigated species; there is particular emphasis on how past and present environmental conditions have shaped and are influencing the fish genetic structure. As a preliminary step, we obtained several panels of microsatellite markers, which could be amplified in the studied species (**Papers II, III and IV**). Loci were isolated from both genomic DNA and from an EST library, and they were used as main tool to get the input dataset for analyses. With regard to the *Chionodraco* genus, molecular markers were used to investigate the pattern of intra- and inter-specific genetic differentiation. We searched for putative outlier loci showing a high level of genetic differentiation between the three *Chionodraco* species and thus possibly influenced by selective forces acting during their shallow evolutionary history (**Article V**). The extent and timing of interspecific gene exchange was also considered to clarify the role of glacial cycles in promoting the divergence and the introgression of species (**Article VI**). With regard to *P. antarcticum*, the main goal was the investigation of its population genetic structure along the AP shelf. In this region, rapid climate warming is threatening *P. antarcticum*, which is an ice-dependent and keystone species of the Antarctic marine ecosystem (**Article VII**).

STUDY SPECIES

The *Chionodraco* genus

The *Chionodraco* genus belongs to the family Channichthyidae (also known as icefish), which is a recently diverged group in the phylogeny of Antarctic notothenioids. Radiation of species within the icefish clade have occurred during the last six millions of years (Near *et al.* 2012) and the family is now constituted of 11 genera and 16 species, with a majority of monospecific genera (Near *et al.* 2003). One exception is the *Chionodraco* genus for which three species have been described: *C. hamatus* (**Figure 1**), *Chionodraco myersi* and *Chionodraco rastrispinosus* (Fischer and Hureau 1985, Eastman and Eakin 2000). Phylogenetic analyses using molecular data showed that the *Chionodraco* genus is a recently derived clade in the phylogenetic tree of the family (Chen *et al.* 1998, Near *et al.* 2003). The three species display similar morphology (Fischer and Hureau 1985), but have distinct and monophyletic groups of mtDNA lineages with *C. hamatus* and *C. rastrispinosus* being sister species and *C. myersi* more distantly related (Chen *et al.* 1998, Near *et al.* 2003, Patarnello *et al.* 2003). Based on mtDNA and nuclear genes estimates, the pairwise divergence of the *Chionodraco* species is similar and species differentiation started between 2 and 1.8 MYA (Near *et al.* 2012).

C. hamatus and *C. myersi* are distributed all around the continent with different depth preferences, *i.e.* 4-600 m and 200-800 m respectively (Gon and Heemstra 1990, Eastman 1993). *C. rastrispinosus*, on the contrary, is found only in the waters around the AP, the South Shetland Islands and the South Orkney Islands (Kock 1992) at depths from 0 to 1000 m (Iwami and Kock 1990). With regard to the bathymetric distribution, larvae, postlarvae and juveniles prefer subsurface waters (upper 100 m of the water column), usually not far from the coast (Loeb *et al.* 1993, Kellermann 1996), and typically feed on early life stages of krill and fish larvae (Kellermann 1989, La Mesa *et al.* 2011). When fish grow older (beyond 2-3 years of age) they become more sedentary and stay close to the sea floor most of the time sitting on their pelvic fins and waiting for preys, primarily krill and fish (Kock 1992, Hagen *et al.* 2001).

Most icefish reach sexual maturity after they become 5-8 years old (Kock and Kellermann 1991), at 4 years old in *C. rastrispinosus* (La Mesa and Ashford 2008). Spawning takes place between autumn (*C. hamatus* and *C. rastrispinosus*) and winter (*C. myersi*) (Kock 2005a) and lasts for about 2-3 months (Kock and Kellermann 1991). Fecundity of the *Chionodraco* species is low, as is typical of species from cold environments; they spawn between 1,464 and 6,389 demersal eggs, rich in yolk and of large size, with a diameter in the order of 3.8-5.0 mm (Kock 2005a). It is unknown if icefish choose particular areas for spawning or if they spawn all along the shelf in suitable grounds and depth. Nest guarding by males has been described for some icefish species (Kock 2005a, b) and it was suggested that the *Chionodraco* species might also exhibit parental care in order to assist eggs (Pshenichnov 2004). Hatching time is known for *C. rastrispinosus* and occurs in late winter/spring (Kock and Kellermann 1991). Larvae are comparatively large at hatching (13-17 mm); fish grow fast when they are immature (6-10 cm per year), but no more than 2-3 cm per year after they have attained sexual maturity (Kock 2005a, b). Icefish are

among the largest of Antarctic fish; observed maximum size is 39 cm in *C. myersi* and 49 cm in *C. hamatus* and *C. rastrospinosus* (Kock 2005a). Reported maximum ages are 10, 12 and 16 years for *C. hamatus*, *C. rastrospinosus* and *C. myersi* respectively (La Mesa and Ashford 2008).

The *Chionodraco* species, as representatives of the family Channichthyidae, display unique features that make them an excellent model to study cold adaptation. Icefish are a unique example of adult vertebrates lacking hemoglobin and functionally active erythrocytes, and, as a consequence, the oxygen-carrying capacity of their blood is only 10% that of red-blooded species (Ruud 1954, Kock 2005a). Additionally, icefish do not express myoglobin in skeletal muscle, and six members of the family have lost the ability to produce the protein in cardiac myocytes either, exacerbating their hemoglobinless condition (Sidell *et al.* 1997, Moylan and Sidell 2000, Grove *et al.* 2004, Borley and Sidell 2010). Adequate oxygen delivery to tissues, even in the absence of respiratory pigments, is ensured in icefish thanks to the evolution of peculiar phenotypic traits such as hypertrophic heart, high cardiac output, increased blood volume, enlarged vessels lumina, low blood viscosity and pressure, well-perfused gills, improved skin and fin vascularization, cutaneous uptake of oxygen, and reduced metabolic rate (Kock 2005a). One of the most remarkable icefish features is an exceptionally high mitochondrial density in the heart and skeletal muscle, which improves oxygen storage and diffusion in cells (O'Brien and Mueller 2010).



Fig. 1. Specimen of *Chionodraco hamatus*.

Pleuragramma antarcticum

P. antarcticum (**Figure 2**), also known as Antarctic silverfish, belongs to the family Nototheniidae and is a basal species in the phylogeny of Antarctic notothenioids (Near *et al.* 2012). It is the only species of the suborder in which all developmental stages, from eggs to adults, inhabit the water column (DeWitt *et al.* 1990, Vacchi *et al.* 2004). Like all the other notothenioids, *P. antarcticum* lacks a swim bladder so that the evolution of a pelagic lifestyle involved a suite of adaptations to achieve neutral buoyancy, such as lipid deposition and reduced skeletal mineralization (Eastman 1993). Pedomorphism is one of the mechanisms involved in the achievement of neutral buoyancy leading to the persistence of notochord, retention of cartilage and delayed ossification of bones (Eastman 1997, Albertson *et al.* 2010).

P. antarcticum is characterized by a circum-Antarctic distribution, including the Scotia Arc and adjacent islands, and it inhabits both open waters and areas covered by pack ice at depths from 0 to 700 m (DeWitt *et al.* 1990). Antarctic silverfish displays stratification by depth, with different stages of development found at different levels in the water column (La Mesa *et al.* 2010). Eggs are pelagic and are found in coastal waters within the platelet ice layer under the sea ice cover, while early larvae are typically distributed at depth of 0-100 m. Post-larvae and juveniles live in deeper (down to 400 m) and offshore waters, potentially contributing to species dispersal by Antarctic ocean currents. Adults are generally found deeper than 400 m and, contrary to other notothenioids, display vertical migration of as much as 350 m in the presence of seasonal light, moving to shallower waters at night and deeper during the day. Avoidance of predation by visually oriented predators, such as penguins and seals, has been suggested as the most likely explanation of this behaviour (Robison 2003).

Antarctic silverfish reaches sexual maturity at 6-7 years old corresponding to a standard length (SL) of about 13 cm (Faleeva and Gerasimchuk 1990, Ferrando *et al.* 2010). Females display high fecundity values, spawning between 4,315 and 17,774 pelagic eggs, which measure 2.2-2.5 mm when embryonated (Kock and Kellermann 1991, Vacchi *et al.* 2004). Spawning occurs during the austral winter and is confined to cold high latitude waters of the continental ice shelves (La Mesa and Eastman 2011). Eggs are incubated between 3-4 months (Kock and Kellermann 1991) and hatching typically occurs between October and December concurrent with phytoplankton and zooplankton blooms in coastal areas, which offer favourable food conditions for newly hatched larvae (La Mesa *et al.* 2010). Larvae show a SL of 0.7-1 cm at hatching (Vacchi *et al.* 2004, La Mesa and Eastman 2011) and grow about 0.15-0.24 mm per day (up to about 2 cm SL); the growth rate decreases considerably to about 0.07 mm per day in post-larvae (from 2 to about 6 cm SL, cohorts 0+ and 1+ based on Hubold and Tomo 1989) and juveniles (La Mesa and Eastman 2011). Compared to other Antarctic notothenioids, growth rate and maximum size of *P. antarcticum* are small, while its life span is long; the species can live for 12-14 years reaching a maximum total length of about 25 cm (DeWitt *et al.* 1990, La Mesa and Eastman 2011).

P. antarcticum is the dominant pelagic fish in most shelf areas of the Southern Ocean, such as the Weddell and the Ross Seas, where it accounts for over 90% of the local fish communities both in number and biomass (DeWitt 1970, Hubold and Ekau 1987). The Antarctic silverfish plays a pivotal role in the marine food web of high latitude shelf waters, being one of the major links between lower and higher trophic levels (La Mesa and Eastman 2011). Post-larvae of *P. antarcticum* typically feed on copepods, while juveniles and adults show a wider food spectrum which consists almost exclusively of zooplankton, such as copepods, euphausiids, amphipods and mysids (La Mesa and Eastman 2011). For its wide distribution and high biomass, *P. antarcticum* represents a major component of the diet of typical Antarctic marine top predators, such as whales, seals, penguins, flying birds and fishes (La Mesa *et al.* 2004, Smith *et al.* 2007).

Antarctic silverfish appears to be highly vulnerable to recent climate warming, being particularly sensitive to any climatic or oceanic change that could reduce the extent and duration of sea ice cover (Vacchi *et al.* 2012). Although the exact relationship between *P. antarcticum* and sea ice remains to be fully investigated, there is substantial evidence that indicates a strong dependence of the species on sea ice in various phases of its life history. The spawning behaviour of *P. antarcticum* has not been directly observed in the field; however, migration of thousands of individuals to coastal areas under the sea ice cover have been observed from June to October, suggesting that spawning aggregations may form under the sea ice layer in winter for reproduction (Daniels and Lipps 1982). A homing behaviour was also proposed by Koubbi and colleagues (2011), according to which adults would return to spawn at the same place where they were born, driven by geographical and/or environmental conditions experienced in their early life. Moreover, eggs with embryos and newly hatched larvae were found in the Ross Sea floating within the platelet ice layer below the solid cover of pack ice (Vacchi *et al.* 2004). Vacchi and colleagues (2012) also provided evidence that the early life cycle of *P. antarcticum* begins in association with the surface sea ice with eggs and newly hatched larvae constituting a major component of the cryopelagic community. Early life stages appear to be well adapted to the ice environment: the egg chorion protects developing embryos from freezing (Cziko *et al.* 2006); early larvae, which lack significant antifreeze capacity, display underdeveloped gill structures that reduce exposition to ice (Bottaro *et al.* 2009); Regoli and colleagues (2005) also showed a marked temporal increase of antioxidant defences in embryos of *P. antarcticum* sampled in the platelet layers, possibly representing an adaptive counteracting responses to the high oxidative conditions of platelet ice. Other considerations point toward a major role of sea ice in the early life cycle of Antarctic silverfish. The underside of sea ice and the platelet ice layer are sites of enhanced primary production and constitute the primary habitat for many invertebrate grazing species using sea ice as a major source of food and refuge from predation (Arrigo *et al.* 2010, Bluhm *et al.* 2010, Caron and Gast 2010). Granata and colleagues (2009) showed that the diet of *P. antarcticum* post-larvae is mainly composed of sea ice-associated grazers, thus indicating a direct trophic role of sea ice in the early life history of the silverfish. Moreover, the structure of irregularly disk-shaped ice platelets may also protect eggs and early larvae from predation (La Mesa *et al.* 2010). Concluding, although new data need to be acquired

by long-term monitoring surveys, available evidence points toward a tight interaction between *P. antarcticum* and the sea ice environment, serving as both a spawning area and a nursery ground.



Fig. 2. Specimen of *Pleuragramma antarcticum*.

PAPER EXTENDED ABSTRACTS

PAPER I: Genome evolution in the cold: Antarctic icefish muscle transcriptome reveals selective duplications increasing mitochondrial function

In this paper we reconstructed and annotated the first skeletal muscle transcriptome of the icefish *C. hamatus* providing a new resource for icefish genomics. The transcriptome was obtained by 454-sequencing of a normalized cDNA library and consists of 23,968 good-quality contigs, of which 48% was *de novo* annotated. A database is available at <http://compgen.bio.unipd.it/chamatusbase/>.

We exploited deep sequencing of the energy-dependent muscle tissue to test the hypothesis of duplications of genes involved in mitochondrial function and biogenesis. Energy production, in the form of ATP, is the primary function of skeletal muscle mitochondria. ATP production may be challenging in icefish because of the absence of oxygen binding proteins in skeletal muscle (Ruud 1954, Borley and Sidell 2010); moreover, icefish live in a cold environment and low temperatures typically reduce the kinetic energy of molecules and reaction rates, modify molecular interactions and diffusion rates, and affect membrane architecture.

To identify duplicated genes specific of the icefish genome we developed a bioinformatic approach that univocally assigns *C. hamatus* transcripts to orthology groups of five model teleosts with complete genome information (**Figure 3A-B**). Fish orthology groups were extracted from Ensembl Compara gene trees (**Figure 3C**) and the number of *C. hamatus* duplicates was recorded in each group.

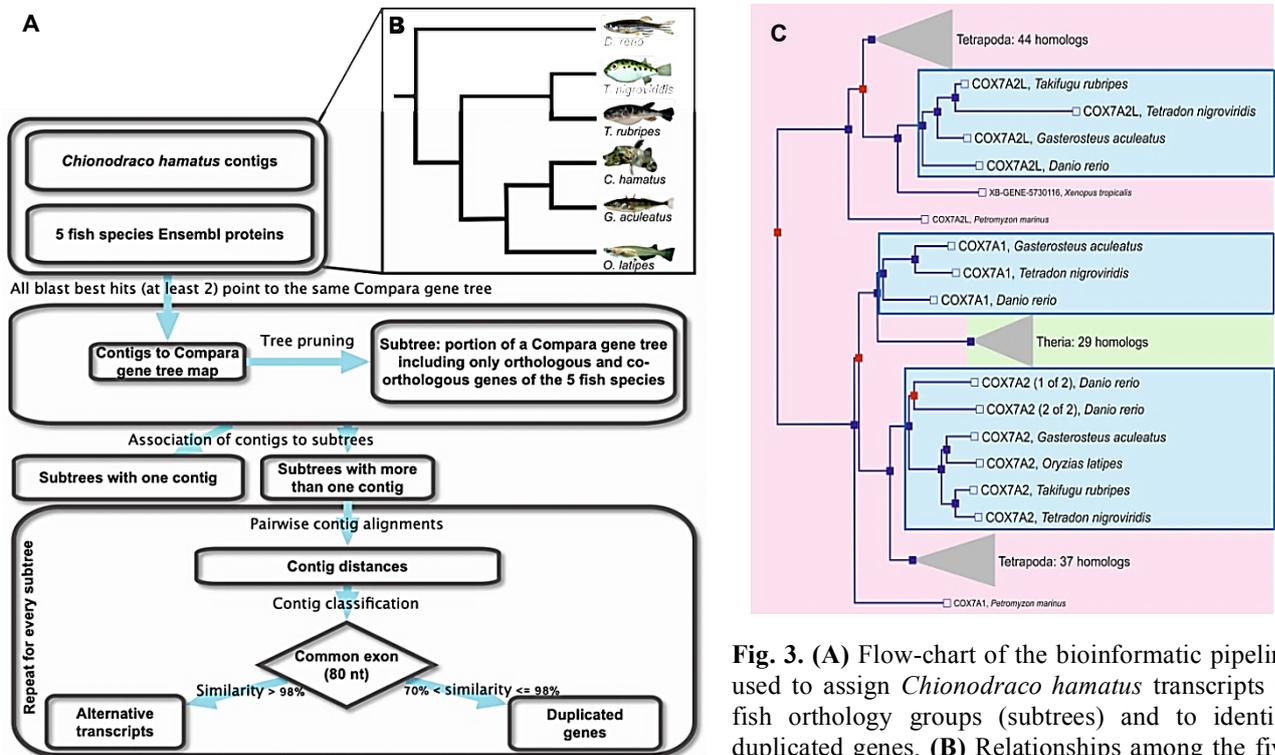


Fig. 3. (A) Flow-chart of the bioinformatic pipeline used to assign *Chionodraco hamatus* transcripts to fish orthology groups (subtrees) and to identify duplicated genes. (B) Relationships among the five model teleosts and *C. hamatus*. (C) Example of a Compara gene tree resulting in three orthology groups in the teleost lineage.

We found 124 orthology groups with *C. hamatus* lineage-specific duplications. The number of duplicates per gene ranged from 2 to 8 (2.3 on average). Significantly more duplicates were found in *C. hamatus* when transcriptome data were compared with whole genome data of four model fish species (*Gasterosteus aculeatus*, *Oryzias latipes*, *Takifugu rubripes* and *Tetraodon nigroviridis*) (**Figure 4A**). This difference is remarkable considering that it was obtained comparing whole genome data from model species (all transcripts, no matter of expression characteristics), with *C. hamatus* duplicated genes expressed in one tissue (skeletal muscle). For *Danio rerio*, which is known to have a highly duplicated genome, the statistical significance was achieved when the comparison was limited to muscle expressed genes (**Figure 4B**). It is worth noting that we are not comparing the level of retention of duplicated genes resulted from whole genome duplication events, but the amount of duplicated genes arisen thereafter and independently in each lineage.

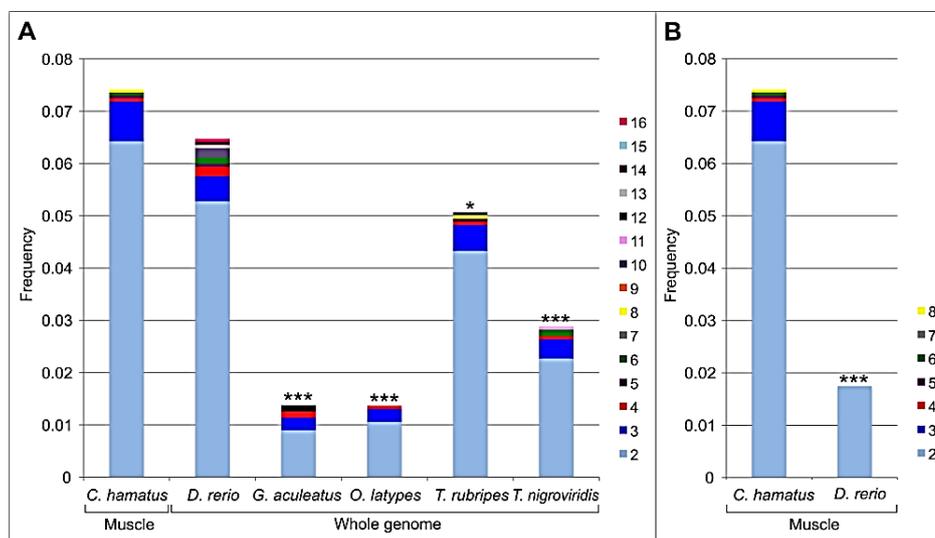


Fig. 4. Comparison of duplicated gene frequencies in different fish species. Bar plots represent, for each species, the fraction of analyzed subtrees with at least one gene represented, including two or more duplicated genes. **(A)** The proportion of *Chionodraco hamatus* duplicated genes, inferred in this study from the analysis of the muscle transcriptome, is pairwise compared with the proportion of duplicated genes of five fish species for which the whole genome is available. Test of equal proportions: ***P value < 0.001, **P value < 0.01, and *P value < 0.05, after Bonferroni correction for multiple tests. **(B)** Comparison of the proportion of duplicated genes expressed in muscle in *C. hamatus* and in *Danio rerio*.

Functional enrichment analyses showed that *C. hamatus* duplicated genes are significantly enriched in proteins involved in mitochondrial function and biogenesis. Enriched functional terms are related to key cellular processes, such as protein translation, oxidative phosphorylation, and organelle organization and biogenesis. Accordingly, at least 34 out of 124 duplicated genes encode proteins with mitochondrial localization, leading to a significant enrichment for mitochondrial proteins in duplicated genes ($p=0.0235$). Some interesting duplicated genes involved in mitochondrial function are:

- 12 structural subunits of the electron transport chain complexes;
- Cox17, Cox19 and Coa5, which are involved in the assembly of cytochrome c oxidase complex;

- Mterfd1, a regulator of mitochondrial transcription, which is required for the assembly of mitochondrial respiratory complexes;
- Oxa11, which mediates the insertion of both nuclear- and mitochondrial-encoded proteins into the mitochondrial inner membrane;
- Atp5s, a subunit of the ATP-synthase, which couples proton translocation through the F₀ subunit with ATP synthesis at F₁. Over-expression of Atp5s affects mitochondria morphology (inner membrane cristae result thicker and more packaged) determining increased coupling of respiration and ATP production;
- Romo1, which inhibits nuclear DNA synthesis and mitosis and induces mitochondrial fragmentation;
- Bnip3, a redox sensor protein activated in response to oxidative stress, which seems to be essential to regulate mitochondrial turnover in the myocardium, where high mitochondrial volume densities are required to satisfy the large ATP demand.

Gene duplications identified in this study are in line with a previously reported trend of genomic expansion accompanying the evolutionary history of Antarctic notothenioid (Chen *et al.* 2008, Detrich and Amemiya 2010). The combination of high mitochondrial densities and the maintenance of duplicated genes involved in mitochondrial biogenesis and aerobic respiration might confer a selective advantage by improving oxygen diffusion and energy supply to aerobic tissues (O'Brien and Mueller 2010).

PAPER II: Isolation, characterization and multiplexing of expressed sequence tag-linked microsatellite loci for the Antarctic icefish *Chionodraco hamatus*

The EST database obtained from *C. hamatus* skeletal muscle and analysed in **Paper I**, was screened for microsatellites sequences. Twelve loci were selected for primer design and optimization in order to set up one multiplex PCR reaction. Out of the 12 primer sets, 11 gave high quality amplification products using as template genomic DNA of 23 *C. hamatus* individuals. All loci proved to be polymorphic, with allele number ranging from 2 to 9, and observed and expected heterozygosity from 0.0435 to 0.8261, and from 0.0435 to 0.8580, respectively. No significant deviation from HW (Hardy-Weinberg) proportions was detected and no evidence for linkage disequilibrium was found.

We cross-amplified all 11 loci in 20 individuals of each of the two congeneric species, *C. myersi* and *C. rastrispinosus* (Channichthyidae) using the same amplification conditions. All loci gave high quality amplification products and were found to be polymorphic in both species.

These molecular markers were used to investigate the population genetic structure and evolutionary history of the three *Chionodraco* species in **Papers V** and **VI** and could be useful for future studies also in other Antarctic notothenioids.

PAPER III: Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei

This work reports on the isolation of novel, polymorphic molecular markers for population genetics analysis in *P. antarcticum*. Microsatellites were isolated from *P. antarcticum* genomic DNA by an enriched library using the FIASCO (Fast Isolation by AFLP of Sequences COntaining repeats) protocol (Zane *et al.* 2002). Out of the 81 sequenced clones, 22 sequences contained a microsatellite with appropriate flanking regions and sequence quality for primer design. Primer pairs were tested on genomic DNA of 40 *P. antarcticum* individuals. Among the 22 primer pairs tested, all produced amplified product, and 9 produced polymorphic and high quality profiles. The number of alleles scored for the 9 polymorphic loci ranged from 7 to 39, whereas the observed and expected heterozygosity ranged from 0.4000 to 0.9750 and from 0.3943 to 0.9782, respectively. All 9 loci resulted in linkage equilibrium, while 2 loci showed HW disequilibrium and a general excess of homozygotes. This pattern could be explained by a Wahlund effect, due to mixing of geographically or temporally differentiated genetic pools, or by the presence of null alleles.

Cross-amplification was tested in 3 species of the genus *Chionodraco* (Channichthyidae), *C. rastrispinosus*, *C. hamatus* and *C. myersi*. Of the 9 loci tested, 3 cross-amplified and were found to be polymorphic in 20 individuals of each species.

One of these microsatellites was used to investigate the population genetic structure and evolutionary history of the three *Chionodraco* species in **Papers V** and **VI**. By contrast, these loci were not used for the population genetic analysis performed in **Paper VII** because we decided to use EST-linked loci, which showed a better performance of amplification, also based on analyses of **Papers V** and **VI**.

PAPER IV: Isolation, characterization and multiplexing of sixteen EST-linked microsatellite loci in the Antarctic silverfish *Pleuragramma antarcticum*

P. antarcticum, plays a major role in the Antarctic marine ecosystem and, being sea ice dependent, is endangered by current climate change. Genetic markers, such as microsatellites, represent an important tool to assess the pattern of population differentiation among existing stocks of *P. antarcticum* and a valuable strategy to monitor the level of genetic variability within local populations.

In this note, the EST database obtained from *C. hamatus* skeletal muscle and analysed in **Paper I**, was screened for microsatellite sequences. Forty-one *C. hamatus* microsatellites were selected for primer design and optimization for cross-amplification in *P. antarcticum* in order to set up two multiplex PCR reactions. Selected loci were primarily tested on DNA samples from 1 *C. hamatus* individual and from 20 *P. antarcticum* individuals. Out of the 41 primer sets, 31 gave high quality amplification products of the expected size in *C. hamatus*, while 19 loci successfully cross-amplified in *P. antarcticum*. Moreover, the panel

of 12 polymorphic microsatellites isolated from the same EST database (**Paper II**) was tested in *P. antarcticum*. Five loci gave a successful amplification performance and were used to set up multiplex PCRs. In total, 24 out of 43 microsatellites amplified in *C. hamatus* (56%) successfully cross-amplified in *P. antarcticum*. This is an appreciable performance of cross-amplification considered the divergence time of the two species and the data obtained from other studies (Barbara *et al.* 2007, Reid *et al.* 2012). Our result also point out the importance of transcriptome sequencing and derived EST-linked molecular markers, which are localized in transcribed regions and are thus expected to be more conserved between species.

The 24 selected loci were cross-amplified together in two multiplex PCR reactions using labelled forward primers and, at this step, 8 loci were further discarded because of genotyping problems. The remaining 16 loci were polymorphic, produced high quality profiles, showing no evidence of null alleles, stuttering and large allele dropout, and were in linkage and HW equilibrium. Number of alleles per locus ranged from 2 to 16, observed and expected heterozygosity ranged from 0.1500 to 0.8500 and from 0.1449 to 0.9231, respectively.

These 16 loci were used to analyse the population genetic structure of *P. antarcticum* along the AP (**Paper VII**).

PAPER V: Putative selected markers in the *Chionodraco* genus detected by interspecific outlier tests

The identification of loci under selection (outliers) is a major challenge in evolutionary biology, being critical to comprehend evolutionary processes leading to population differentiation and speciation, and for conservation purposes, also in light of recent climate change. However, detection of selected loci can be difficult when populations are weakly differentiated. This is the case of marine fish populations, often characterized by high levels of gene flow and connectivity, and particularly of fish living in the Antarctic marine environment, characterized by a complex and strong circulating system promoting individual dispersal all around the continent.

In this study, we analysed the level of genetic differentiation (described by the F_{ST} index) between the three recently derived species of the *Chionodraco* genus (Notothenioidei, Channichthyidae), namely *C. hamatus*, *C. myersi* and *C. rastrispinosus*. The three species are endemic to the Southern Ocean and they are thought to have diverged between 2 and 1.8 MYA (Near *et al.* 2012). With the final aim of identifying outlier loci putatively under selection in the *Chionodraco* genus, neutrality tests were applied to search for loci showing a high level of genetic differentiation between the three *Chionodraco* species and thus possibly influenced by selective forces acting during their shallow evolutionary history.

Analyses were performed using a panel of 21 microsatellite loci, which were successfully cross-amplified in 108 specimens of the three species. The panel included 10 microsatellites originally isolated from genomic DNA (Type II loci) of *C. rastrispinosus* (Papetti *et al.* 2006), *C. aceratus* (Susana *et al.* 2007), and *P. antarcticum* (**Paper III**), and

11 EST-linked loci (Type I loci) isolated from a normalized cDNA library from *C. hamatus* muscle (**Paper I and II**). We used both Type I and Type II because the two classes of genetic markers could reveal different aspects of the differentiation processes of the investigated species.

Type I loci showed a better amplification performance with no detection of null alleles and no departure from HW equilibrium, which can be explained by the high conservation of primer annealing regions residing into coding sequences. On the other hand, Type II loci showed some deviations from HW equilibrium but a higher level of genetic polymorphism, as indicated by values of indices of genetic variation (number of alleles, allelic richness, observed and unbiased expected heterozygosity), which significantly exceeded those calculated with Type I markers. This result possibly reflects the presence of functional constraints and of the effects of purifying selection at Type I loci. F_{ST} values calculated at Type II loci exceeded those calculated at Type I loci in all pairwise comparisons between species only when they were standardized following Hedrick (2005) (F'_{ST}) (**Table 1**). This finding confirms the higher power of neutral genomic loci to detect differentiation and further points out a possible bias in non-standardized F_{ST} estimates.

In line with results of Patarnello and colleagues (2003), F_{ST} values indicated the presence of three distinct gene pools corresponding to the three nominal species, with *C. hamatus* and *C. rastrispinosus* being more closely related (**Table 1**).

Neutrality tests applied in interspecific comparisons identified three outlier loci: Cr38 and Ch684, in all pairwise comparisons including *C. rastrispinosus*, and Ch8501 in all pairwise comparisons including *C. myersi*. Single-locus F_{ST} and F'_{ST} values at outlier loci detected a high level of genetic differentiation, specifically in those pairwise comparisons in which they were identified as outlier by neutrality tests (**Table 2**); in these comparisons, F_{ST} and F'_{ST} estimates clearly exceeded the respective F_{ST} and F'_{ST} 95% CI between species calculated with 18 putative neutral loci (**Table 3**). Interestingly, locus Ch684 showed sequence similarity to a calmodulin transcript, Cr38 to an AFGP/TLP gene and Ch8501 to a non-annotated fish mRNA. Assuming that the higher divergence found at outlier loci can be attributed to directional selection, selective pressures could have operated in the branch leading to the divergence of *C. myersi* for Ch8501 and in the branch leading to the divergence of *C. rastrispinosus* for loci Cr38 and Ch684. These selective pressures acting on specific loci might reflect past evolutionary processes leading to species divergence and local adaptation in the *Chionodraco* genus.

Table 1. Species pairwise F_{ST} and F'_{ST} values. All pairwise comparisons were performed using four datasets: all 21 loci considered in this study, the 18 neutral loci indicated by the neutrality tests (loci Cr38, Ch684, and Ch8501 were excluded), 9 neutral Type II loci (locus Cr38 was excluded), and 9 Type I loci (loci Ch684 and Ch8501 were excluded). All reported values are statistically significant (p values < 0.0001). F_{ST} : actual estimate of population differentiation; F'_{ST} : standardized measure of population divergence. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrispinosus*.

Sample pairs	21 loci		18 neutral loci		9 neutral Type II loci		9 neutral Type I loci	
	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}
Ch-Cm	0.2100	0.4996	0.1597	0.4066	0.1863	0.6198	0.1211	0.2501
Ch-Cr	0.1672	0.3821	0.1114	0.2919	0.0914	0.3551	0.1399	0.2765
Cm-Cr	0.2004	0.4645	0.1196	0.3030	0.1028	0.3429	0.1417	0.2893

Table 2. Species pairwise F_{ST} and F'_{ST} values at outlier loci indicated by neutrality tests: Cr38, Ch684 and Ch8501. All reported values are statistically significant (p -values < 0.0001). Pairwise comparisons in which each locus was identified as an outlier are reported in bold. F_{ST} : actual estimate of population differentiation; F'_{ST} : standardized measure of population divergence. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrispinosus*.

Sample pairs	Cr38		Ch684		Ch8501	
	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}
Ch-Cm	0.3923	0.8538	0.1180	0.2730	0.8378	0.9751
Ch-Cr	0.6338	0.9192	0.6142	0.8536	0.0060	0.0065
Cm-Cr	0.5460	0.9524	0.4324	0.7675	0.7892	0.9617

Table 3. Species pairwise F_{ST} and F'_{ST} 95% confidence interval (CI) calculated at 18 putative neutral loci. F_{ST} : actual estimate of population differentiation; F'_{ST} : standardized measure of population divergence. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrispinosus*.

Sample pairs	F_{ST} 95%CI	F'_{ST} 95%CI
Ch-Cm	0.0920 to 0.2320	0.2342 to 0.5906
Ch-Cr	0.0600 to 0.1670	0.1598 to 0.4375
Cm-Cr	0.0570 to 0.1920	0.1419 to 0.4890

PAPER VI: Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation

Fish of the perciform suborder Notothenioidei now dominate high-latitude coastal shelf areas of Antarctica in terms of species abundance and biomass (Eastman and McCune 2000). Notothenioid dominance is probably related to tectonic, climatic and oceanographic events that isolated the Antarctic shelf from other oceans, and to the establishment of cold conditions. Antarctic notothenioids experienced a spectacular radiation to more than 100 species during Late Miocene and alternating warming and cooling periods of recent glacial cycles may have affected species distributions, promoted ecological divergence into

recurrently opening niches and/or possibly brought allopatric species into contact (Lautrédou *et al.* 2012, Near *et al.* 2012).

In this study, we investigated patterns of genetic differentiation and hybridization in the three species of the genus *Chionodraco*. Our main objectives were to determine the extent and timing of gene exchange between recently diverged species and thus to elucidate the role of warm interglacial and cold glacial phases in promoting the divergence and the introgression of species, respectively.

We considered six population samples (N = 108) of *C. hamatus*, *C. myersi* and *C. rastrospinosus* collected between 1988 and 2006 at four different locations (the Weddell Sea, the Ross Sea, Elephant Island and Joinville Island). Analyses were performed using 18 microsatellite markers, *i.e.* the same panel of microsatellites used in **Paper V** excluded those resulted under selection.

The dataset obtained with the 18 cross-amplifying microsatellite markers was analysed using multiple statistical approaches. In particular, (i) we analysed the nuclear genetic structure within and between the *Chionodraco* species calculating F-statistics, performing the exact test of genetic differentiation and using the software STRUCTURE, (ii) we inferred the most likely genetic composition of contemporary single individuals and identified individuals with mixed ancestry using STRUCTURE, (iii) we compared five evolutionary scenarios assuming different degrees of association between recent glacial cycles and interspecific migration rates, using Approximate Bayesian Computation (ABC).

Supporting results obtained in **Paper V**, F-statistics, the exact test of genetic differentiation and the analyses with STRUCTURE (**Figure 5**) all indicated homogeneity of population samples belonging to the same species; additionally, three genetically differentiated gene pools corresponding to the nominal species were found, also in line with previously obtained results. The study of the genetic composition of contemporary single individuals in each species indicated 29 individuals with mixed ancestry, and 2 individuals with about 50% of their genetic composition attributed to a different species (**Figure 5**). The average contribution of another species into the hybrids was estimated between 11% and 14%, and *C. rastrospinosus* was found to be the most introgressing species. Interestingly, we also observed a sample from Elephant Island morphologically classified as *C. rastrospinosus* and with nuclear composition largely composed of this species, but characterized by a *C. hamatus* mtDNA haplotype.

Analyses using ABC showed that evolutionary scenarios excluding hybridization or including it only in ancient times have small or zero posterior probabilities. By contrast, the most probable scenario indicated interspecific gene flow associated with the two most recent interglacial periods, the Eemian and the Holocene (**Figure 6**). Estimates of the number of migrants exchanged between each pair of species per generation indicated low levels of migrant exchange between *C. myersi* and *C. hamatus*. Larger values were obtained in other comparisons and especially for the pair *C. myersi* and *C. rastrospinosus*.

Our results supported the hypothesis that gene flow among the *Chionodraco* species occurred in recent and probably also in ancient times. Complete reproductive isolation

does not seem to have been reached, but also hybridization (especially during interglacial periods) does not seem to have been, as yet, extensive enough to prevent or reverse species diversification. These findings are partially compatible with the mechanism suggested to explain the recent Notothenioid diversification: an increased opportunity for speciation in allopatric refugia during glacial periods, followed by colonization of coastal habitats with empty niches due to eradication of benthic fauna, and further ecological divergence (Near *et al.* 2012). The admixture we observed in the *Chionodraco* data, indicating incomplete isolation, might be low enough to allow adaptation to ecological factors and/or might suggest hybrid genomic incompatibilities to maintain species boundaries. Exploration of adaptive differences and reinforcement mechanisms will require further study.

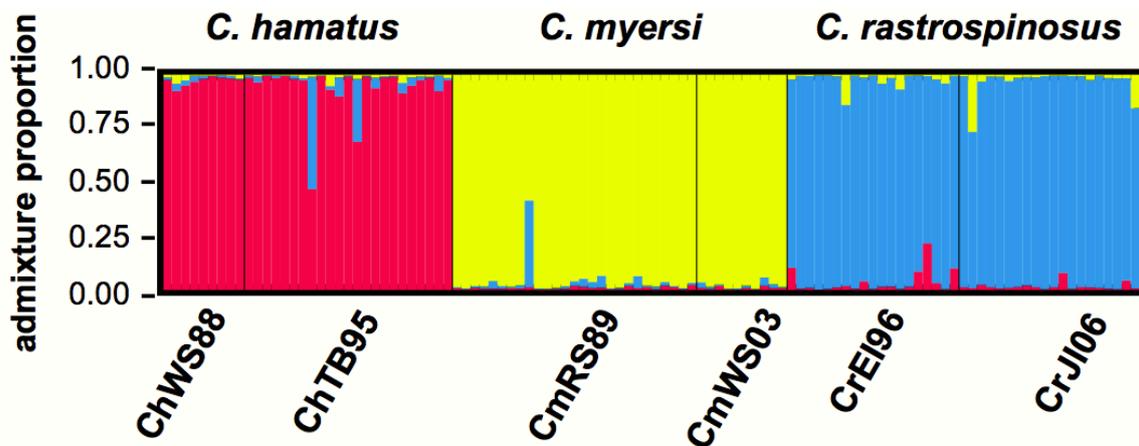


Fig. 5. Population structure estimated using the program STRUCTURE from 108 individuals from six localities using all 18 loci. Each individual is represented by a vertical line, which is partitioned into K coloured segments, the length of each colour being proportional to the estimated membership coefficient from cluster 1 (red, *Chionodraco hamatus*), cluster 2 (yellow, *Chionodraco myersi*) and cluster 3 (blue, *Chionodraco rastrospinosus*). Black lines separate different populations (labelled below the figure).

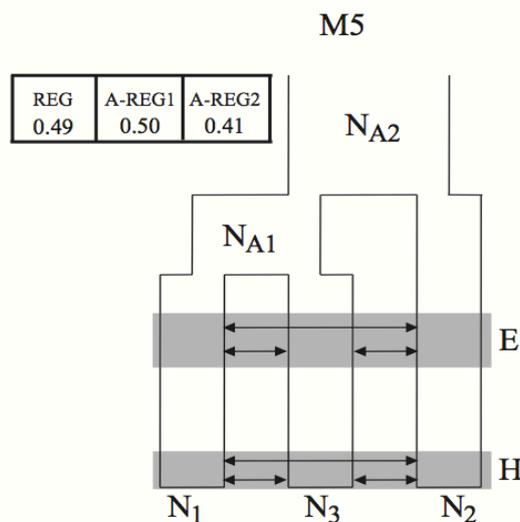


Fig. 6. Most probable evolutionary scenario resulted from the comparison of five models with the ABC approach. The small table reports the posterior probabilities computed using the regression approach (REG), the adjusted regression approach with bandwidth equal to 0.05 (A-REG1) and the adjusted regression approach with bandwidth equal to 0.3 (A-REG2).

PAPER VII: Has climate change promoted genetic fragmentation in the ice-dependent fish *Pleuragramma antarcticum*?

The study regarding *P. antarcticum* is part of a wider collaborative project between our research group and those of Prof. Joseph Torres (College of Marine Science, University of South Florida, St. Petersburg, FL, USA) and Prof. Julian Ashford (Center for Quantitative Fisheries Ecology, Old Dominion University, Norfolk, VA, USA). The final goal is to investigate the population structure of *P. antarcticum* along the AP by using two different approaches: during my PhD I analysed the pattern of genetic differentiation by using microsatellite markers (results reported in **Paper VII**), while our American collaborators focused on stable isotope and trace element chemistry from otoliths (Ferguson *et al.* 2011).

A single genetic study investigated the population structure of *P. antarcticum* in the Southern Ocean using mitochondrial DNA sequencing, but was unable to discriminate between hypotheses of panmixia, with occasional fluctuations of gene pools, and strong population structure (Zane *et al.* 2006). The aim of the present study is the investigation of the population structure of *P. antarcticum* in a restricted geographic area, the AP shelf, which is considered a true “hot-spot” of recent climate warming (Turner *et al.* 2005, Zazulie *et al.* 2010).

We analysed a total of 562 individuals from nine population samples, collected between 2001 and 2012, which are representative of four different locations along the AP shelf: Charcot Island, Marguerite Bay, Joinville Island and Larsen Bay (see **Table 4** and **Figure 7** for sample information). During both 2010 and 2011 sampling cruises, no individual was caught in the vicinity of Anvers Island, Renaud Island and in Croker Passage, despite its occurrence has been historically reported in this area (Kellermann 1996) (**Figure 7**).

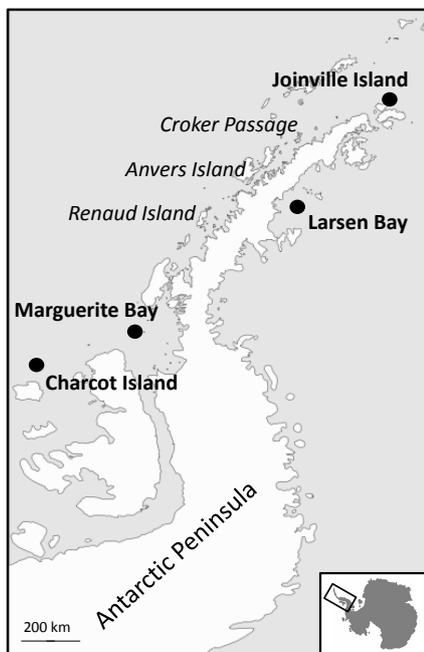


Table 4. *Pleuragramma antarcticum* population samples collected between 2001 and 2012 at four different locations off the Antarctic Peninsula. Reported are collection site, sampling year, sampling campaign, and number of individuals analysed at 16 EST-linked microsatellites (N).

Site	Year	Campaign	Coordinates	N
Charcot Island	2010	NBP 10-02	70°07'S, 76°02'W	60
Marguerite Bay	2001	SO GLOBEC-Cruise 1	67°57'S, 68°21'W	28
	2002	SO GLOBEC-Cruise 3	68°08'S, 68°01'W	49
	2010	NBP 10-02	67°49'S, 68°09'W	60
	2011	LMG Cruise 11-01, Palmer LTER	67°39'S, 70°04'W	83
Joinville Island	2007	ANTXXIII/8AWI	62°35'S, 54°45'W	34
	2010	NBP 10-02	63°30'S, 56°40'W	148
	2012	ANT-XXVIII/4 AWI	62°14'S, 55°18'W	54
Larsen Bay	2007	ANTXXIII/8AWI	65°30'S, 61°40'W	46

Fig. 7. Approximate sampling location of *Pleuragramma antarcticum* (circles). Reported in italics are also areas where the species was not detected in 2010 and 2011.

Population samples were analysed with 16 EST-linked microsatellite loci (**Paper IV**), which were mined from the EST database obtained from *C. hamatus* skeletal muscle and analysed in **Paper I**. Loci were successfully cross-amplified and genotyped in all 562 considered individuals, they were shown to be putative neutral markers and were both in linkage and in HW equilibrium in all population samples.

Indices of genetic diversity (number of alleles, allelic richness, and observed and expected heterozygosity) indicated a similar level of variation in all analysed samples, both on a temporal and geographical scale (one-way ANOVA, $p > 0.05$).

Population pairwise F_{ST} showed no genetic differentiation between temporal samples from both Marguerite Bay and Joinville Island, indicating genetic stability on a temporal scale. On a geographic scale, a substantial genetic homogeneity was found between samples collected off Charcot Island and in Marguerite Bay, indicating a single gene pool in this area. Yet in contrast, significant differentiation was detected between Joinville Island and the south-western AP (Charcot Island and Marguerite Bay), indicating that Antarctic silverfish forms separate gene pools in the two areas, thus discounting the hypothesis of panmixia. Moreover, the population sample collected at Joinville Island in 2007 showed no differences when compared to any of the analysed samples, while samples taken in 2010 and 2012 showed significant differences with respect to both Charcot Island and Marguerite Bay, with Joinville Island 2012 displaying the highest genetic differentiation. Additionally, if we clustered individuals from Joinville Island 2010 based on a standard length of 10 cm (transition from cohort 3+ to 4+, Hubold and Tomo 1989), the “Joinville Island 2010 small cluster” showed a higher genetic differentiation than the “Joinville Island 2010 large cluster” when pairwise compared to samples from the south-western AP. Similarly, the Larsen Bay 2007 sample showed significantly higher F_{ST} values when compared to more recent samples from Joinville Island and the south-western AP (2010, 2011 and 2012) than to older samples (2001, 2002 and 2007). A hierarchical AMOVA also showed significant genetic differentiation among geographic groups. These results not only show genetic differentiation on a local geographic scale in *P. antarcticum*, despite being the only Antarctic notothenioid with a completely pelagic life cycle and thus the highest potential for dispersal, but also suggest that the level of genetic differentiation has increased on a relatively recent time scale.

Assignment tests and the migration rate analysis, using a Bayesian approach, revealed a low but significantly higher flow of migrants moving southward along the western AP, following the anti-clockwise Coastal Current, than in the opposite direction. However, gene flow seems to have been insufficient to prevent population differentiation.

Concluding, reduced level of gene flow along the shelf, the increase of population fragmentation with time, and the inability to capture *P. antarcticum* in the central region of the western AP for two consecutive years, all suggest that this sea ice dependent species could be highly vulnerable to climate change with possible cascading effects on the Antarctic marine food web.

GENERAL CONCLUSIONS

In this section I will briefly discuss the results of analyses performed during my PhD. A more detailed and comprehensive presentation and discussion of specific results is reported in the attached papers.

Antarctic notothenioids represent an exceptional “evolutionary case” for studying adaptive radiation in an extreme marine environment, adaptation to the cold, and for determining the biological effects of recent climate change. During this PhD we considered these different aspects of notothenioid evolutionary history focusing on key species of the notothenioid suborder and by using various scientific approaches, such as next generation sequencing, bioinformatic tools, and software and statistical methods for population genetics with microsatellite markers. We started from the investigation of the genomic basis of cold adaptation in the icefish, for which energy production in cold conditions is particularly challenging because of the absence of oxygen-binding proteins, and we found several duplicated genes involved in mitochondrial function and biogenesis (**Paper I**). Subsequently, we moved on to a more recent timescale and found that several microsatellite regions, in linkage with transcribed sequences, are subjected to directional selection in the three recently derived species of the *Chionodraco* genus (**Paper V**); these loci possibly reflect past selective forces playing a role in the process of species diversification. We also analysed the pattern of genetic differentiation among the three *Chionodraco* species and found evidence of interspecific gene flow associated with the two most recent interglacial periods (**Paper VI**). Finally, we moved to the present time by investigating the population structure of *P. antarcticum* along the AP, and we found some evidence that this sea ice dependent species has been affected by regional warming (**Paper VII**). The principal findings of my thesis are the follows.

In **Paper I**, we focused on the genetic and genomic adaptations to extreme cold in the icefish *C. hamatus*. The first skeletal muscle transcriptome of the species was reconstructed and characterized. Deep sequencing information, by means of a dedicated bioinformatic pipeline, revealed 124 putative duplicated genes specific to the *C. hamatus* lineage. *C. hamatus* displayed a significantly higher proportion of lineage-specific duplicates than model fish species. This difference is striking considering we compared whole-genome data from model species (all transcripts, no matter of expression characteristics), with *C. hamatus* duplicated genes expressed in muscle. Moreover, it is worth noting that we did not compare the level of retention of duplicated genes resulted from whole genome duplication events, but the amount of duplicated genes arisen thereafter and independently in each lineage. Obtained results are in line with a previously reported trend of genomic expansions accompanying the evolutionary history of Antarctic notothenioids (Chen *et al.* 2008, Detrich and Amemiya 2010). We also provided new indications about the functional role of gene duplicates. A global functional enrichment analysis of duplicated genes and the localization and biological role of single-encoded proteins supported our initial hypothesis of duplicated genes involved in mitochondrial function and biogenesis. The progressive expansion of the mitochondrial compartment, which is observed in the evolutionary history of Antarctic notothenioids (Johnston *et al.*

1998), culminates in the icefish lineage, in which energy production is expected to be exceptionally challenging because of the absence of oxygen-carrying proteins (O'Brien and Mueller 2010). Icefish presumably rely on high mitochondrial densities to achieve high concentrations of aerobic enzymes, enhanced lipid-mediated oxygen diffusion, and short oxygen route from capillaries to mitochondria, thus ensuring energy supply and oxygen diffusion to aerobic tissues (O'Brien and Mueller 2010). Therefore, the maintenance of duplicated genes in the icefish genome, may be associated, at least in part, to a strong selective pressure for increased mitochondrial density and function.

The second part of my thesis is focused on the pattern of intra- and inter-specific genetic differentiation in the *Chionodraco* genus and in the Antarctic silverfish *P. antarcticum*, with particular emphasis on how past and present environmental conditions have shaped and influenced the fish genetic structure. In **Paper V**, we investigated the level of genetic differentiation among the three *Chionodraco* species with the final aim of detecting outlier loci putatively under selection; these loci could reflect past selective forces acting during the shallow evolutionary history of the *Chionodraco* genus and possibly playing a role in the process of species diversification. The analysis of the intra- and inter-specific level of genetic differentiation, which was performed by using the F_{ST} index, indicated intraspecific homogeneity and the presence of three distinct gene pools corresponding to the nominal species. *C. hamatus* and *C. rastrospinosus* resulted to be more closely related, in line with previously results obtained with mitochondrial DNA haplotypes (Patarnello *et al.* 2003). Neutrality tests indicated that selective pressures might have played a role in the process of species divergence. Indeed, three outlier loci, displaying a higher than expected divergence, were identified in pairwise comparisons including *C. rastrospinosus*, and *C. myersi*. The three loci showed sequence similarity to known sequences: locus Ch684 to a calmodulin transcript, Cr38 to an AFGP/TLP gene, and Ch8501 to the mRNA of a key component of the super elongation complex (AF4/FMR2 family, member 4). Notably, calmodulin and other genes involved in the calcium signalling pathway are known to be related to cold tolerance in many plants and ectothermic organisms (Chinnusamy *et al.* 2007, Doherty *et al.* 2009, Teets *et al.* 2013). Yang and colleagues (2013) recently found that calmodulin is expressed at high levels in the Antarctic notothenioid *D. mawsoni* and that the introduction of the calmodulin gene of this cold-adapted fish into tobacco increases the cold tolerance of plants under low-temperature stress. Another interesting gene is the AFGP/TLP; this chimeric gene is supposed to be the ancestral intermediate between the current AFGP gene and its precursor, *i.e.* the pancreatic TLP gene (Chen *et al.* 1997, Cheng and Chen 1999). The sequence similarity of the locus Ch8501 with the member 4 of the AF4/FMR2 family was found, after the publication of **Paper V**, by BLAST search (Altschul *et al.* 1990) against the updated NCBI nucleotide database (90% sequence identity over 641 nucleotides). This gene is a key component of the super elongation complex, which is required to increase the catalytic rate of RNA polymerase II transcription by suppressing transient pausing by the polymerase at multiple sites along the DNA (Luo *et al.* 2012). Based on the results of neutrality tests and assuming that the higher divergence found at outlier loci can be attributed to directional selection, selective pressures detected in this study and acting on specific loci might reflect past evolutionary

processes in the *Chionodraco* genus. Therefore, selective forces, as those observed in **Paper I** leading to the retention of duplicates in the icefish genome and probably related to cold adaptation, may have acted on a more recent time scale (the last 2 millions of years) promoting the divergence and local adaptation in the *Chionodraco* species.

The genetic differentiation pattern found in **Paper V** was corroborated by results obtained in **Paper VI**. The exact test of differentiation, F_{ST} statistics and the model-based clustering method implemented in STRUCTURE all indicated that *C. hamatus*, *C. myersi* and *C. rastrispinosus* are genetically differentiated and internally homogeneous. Moreover, we provided the first evidence of interspecific hybridization among icefish species, with introgression occurring primarily during warmer interglacial phases. The individual assignment analysis using STRUCTURE showed that a substantial fraction of contemporary individuals in each species has mixed ancestry. Accordingly, the probabilistic comparison of multiple evolutionary scenarios using ABC indicated gene flow among the *Chionodraco* species associated with the two most recent interglacial periods, the Eemian and the present-day Holocene. By contrast, evolutionary scenarios excluding hybridization or including it only in ancient times, showed small or zero posterior probabilities. These findings are in line with the mechanism suggested by Near and colleagues (2012) to explain the recent Notothenioid diversification: an increased opportunity for speciation in allopatric refugia during glacial periods, followed by colonization of newly available shelf areas during warmer intervals, possibly favouring ecological diversification but also secondary contact and hybridization between diversified groups. Therefore, not only selective forces, as those detected in **Paper V** acting on specific loci and possibly related to local adaptation, but also recent paleoclimatic events have led to the pattern of genetic differentiation that we currently observed in the *Chionodraco* genus.

Results obtained in **Paper VII**, also suggest that environmental conditions have influenced the genetic structure, in this case at the intraspecific level, of the pelagic Antarctic notothenioid *P. antarcticum*. Here we focused on a more recent timescale (between 2001 and 2012) and on a local geographic scale, *i.e.* the AP, which is one of the most rapidly warming regions of the planet (Turner *et al.* 2005, Zazulie *et al.* 2010). Genetic differentiation analysis showed a single gene pool and an absence of inter-annual variability in the south-western AP, while significant genetic differences were detected from samples collected off the tip of the AP, with a signal of increased fragmentation over time. This result, together with the apparent disappearance of *P. antarcticum* from the central-western AP, which is the area most affected by sea ice melting, suggests that this sea ice dependent species has been affected by climate change with possible cascading effects on the Antarctic marine food web.

Concluding, results obtained in this PhD thesis, not only deepened the knowledge of the genomic changes associated to cold adaptation, but also indicate a strong relationship between past and present environmental conditions and the genetic structure of Antarctic notothenioids, thus suggesting a need for new predictions on the effects of global warming to avoid the loss of biodiversity from this fascinating past evolutionary radiation.

PAPERS

PAPER I

Genome Evolution in the Cold: Antarctic Icefish Muscle Transcriptome Reveals Selective Duplications Increasing Mitochondrial Function

Alessandro Coppe^{1,†}, Cecilia Agostini^{2,†}, Ilaria A.M. Marino², Lorenzo Zane², Luca Bargelloni¹, Stefania Bortoluzzi^{2,*}, and Tomaso Patarnello¹

¹Department of Comparative Biomedicine and Food Science, University of Padova, Agripolis, Legnaro (Padova), Italy

²Department of Biology, University of Padova, Padova, Italy

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: stefania.bortoluzzi@unipd.it.

Accepted: November 23, 2012

Abstract

Antarctic notothenioids radiated over millions of years in subzero waters, evolving peculiar features, such as antifreeze glycoproteins and absence of heat shock response. Icefish, family Channichthyidae, also lack oxygen-binding proteins and display extreme modifications, including high mitochondrial densities in aerobic tissues. A genomic expansion accompanying the evolution of these fish was reported, but paucity of genomic information limits the understanding of notothenioid cold adaptation. We reconstructed and annotated the first skeletal muscle transcriptome of the icefish *Chionodraco hamatus* providing a new resource for icefish genomics (<http://compgen.bio.unipd.it/chamatusbase/>, last accessed December 12, 2012). We exploited deep sequencing of this energy-dependent tissue to test the hypothesis of selective duplication of genes involved in mitochondrial function. We developed a bioinformatic approach to univocally assign *C. hamatus* transcripts to orthology groups extracted from phylogenetic trees of five model species. *Chionodraco hamatus* duplicates were recorded for each orthology group allowing the identification of duplicated genes specific to the icefish lineage. Significantly more duplicates were found in the icefish when transcriptome data were compared with whole-genome data of model species. Indeed, duplicated genes were significantly enriched in proteins with mitochondrial localization, involved in mitochondrial function and biogenesis. In cold conditions and without oxygen-carrying proteins, energy production is challenging. The combination of high mitochondrial densities and the maintenance of duplicated genes involved in mitochondrial biogenesis and aerobic respiration might confer a selective advantage by improving oxygen diffusion and energy supply to aerobic tissues. Our results provide new insights into the genomic basis of icefish cold adaptation.

Key words: deep sequencing, orthology/paralogy relationships, gene duplication, genome evolution, cold adaptation, Channichthyidae.

Introduction

Chionodraco hamatus (Notothenioidei, Perciformes) is an Antarctic teleost belonging to the family Channichthyidae (icefish). All members of the family, with the exception of one species, are endemic to the Southern Ocean and are some of the most stenothermal species on Earth. They evolved in the persistently cold and oxygen-rich Antarctic waters, acquiring unique adaptations at the morphological, physiological, and biochemical level. Icefish, similar to other Antarctic notothenioids, lack a swim bladder and produce antifreeze glycoproteins (AFGPs), a key innovation preventing blood and body fluids freezing at the ambient temperature of -1.86°C

(Cheng and Detrich 2007). Icefish are a unique example of adult vertebrates lacking hemoglobin and functionally active erythrocytes, and, as a consequence, the oxygen-carrying capacity of their blood is only 10% that of red-blooded species (Ruud 1954). Fifteen of the 16 members of the family completely lack the adult β -globin gene and retain only a small 3'-fragment of the α -globin gene (Near et al. 2006). As Ruud (1954) suggested, this disadaptive phenotypic trait could have evolved only in the extreme and stable environmental conditions of Antarctic waters, where the higher oxygen solubility at low temperatures may have relaxed selection pressure for oxygen-binding proteins, allowing the successful

© The Author(s) 2012. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

diversification of icefish over the past 7.8–4.8 millions of years (Near et al. 2012).

Icefish do not express myoglobin in skeletal muscle, and six members of the family have lost the ability to produce the protein in cardiac myocytes either, exacerbating their hemoglobinless condition (Borley and Sidell 2010). Adequate oxygen delivery to tissues, even in the absence of respiratory pigments, is ensured in icefish thanks to the evolution of peculiar phenotypic traits such as hypertrophic heart, high cardiac output, increased blood volume, enlarged vessels lumina, low blood viscosity and pressure, well-perfused gills, improved skin and fin vascularization, cutaneous uptake of oxygen, and reduced metabolic rate (Kock 2005). One of the most remarkable icefish features is an exceptionally high mitochondrial density in the heart and skeletal muscle, which improves oxygen storage and diffusion in cells (O'Brien and Mueller 2010). These unique features make *C. hamatus* an excellent model to study cold adaptation. Several of the phenotypic modifications required in the constantly freezing temperatures of the Southern Ocean appear to involve stable genomic modifications (e.g., deletion of the β -globin locus), which might represent a significant limitation to adaptation to warmer environments (Patarnello et al. 2011). Understanding the genomic constraints on the evolutionary potential of Antarctic species might help to predict how well they will cope with climate change. Several molecular mechanisms ranging from point mutations of preexisting genes to large genomic rearrangements might underlie such modifications. Of all possible mechanisms, gene duplication has long been recognized to play a major role in the evolution of novel functions, a role recently confirmed by eukaryotic whole-genome sequence analysis (Ohta 1989; Lynch 2007). Cheng et al. (Cheng 1996; Cheng et al. 2003) were, to our knowledge, the first to link cold adaptation in Antarctic notothenioids to gene duplication, showing that repeated duplication of the primordial AFGP gene produced the AFGP expanded family. Duplications of other genes, with a putative role in adaptation to low temperatures, were subsequently identified, that is, metallothioneins (Bargelloni et al. 1999), pepsin enzymes (Carginale et al. 2004), and hepcidins (Xu et al. 2008). More recently, Chen et al. (2008) used an array-based comparative genomic hybridization approach to compare three high Antarctic, evolutionarily derived notothenioid species to two non-Antarctic, phylogenetically basal notothenioids, showing that 101 protein-coding genes (implied in various biological functions including extracellular matrix remodeling, protein folding and degradation, response to stress conditions, defense from oxidative and apoptotic processes, Ras/MAPK and TGF- β signal transduction cascades, RNA transcription, binding and processing, cellular structural components, and innate immunity) were duplicated in the Antarctic species. The expression of some duplicated genes was found to be upregulated in the red-blooded Antarctic notothenioid *Dissostichus mawsoni* compared with temperate

water teleosts. A dosage effect was thus hypothesized as the mechanism involved in the maintenance of duplicated gene copies. A recent analysis of whole-genome size in 11 red- and white-blooded notothenioid species confirmed that the evolution of phylogenetically derived notothenioid families, such as the Channichthyidae, might have been accompanied by a genome expansion (Detrich and Amemiya 2010), without further indication, however, of which genes or gene families are involved in such expansion.

The working hypothesis of this study is that targeted gene duplications occurred at loci involved in mitochondrial function and biogenesis. Energy production, in the form of ATP, is the primary function of skeletal muscle mitochondria. ATP production may be challenging in a cold environment because low temperatures typically reduce the kinetic energy of molecules and reaction rates, modify molecular interactions and diffusion rates, and affect membrane architecture. During cold acclimation, many fish enhance oxidative capacity by a variety of physiological mechanisms (Somero 2004; O'Brien 2011). Maintaining energy production at appropriate levels for physiological activities is a constant problem for Antarctic fish, which experience a narrow temperature range reaching -1.86°C , which is lethal for most ectothermic organisms. There is evidence that temperature compensatory adaptations occur at the biochemical level because key enzymes involved in energy metabolism (cytochrome c oxidase, citrate synthase, and lactate dehydrogenase) show a (proportionally) higher enzymatic activity at a common low temperature in Antarctic notothenioids than in temperate and tropical species (Crockett and Sidell 1990; Kawall et al. 2002). This may be obtained by higher enzyme concentrations and/or by increased enzyme efficiency at low temperatures. The increased activity of lactate dehydrogenase, for example, can be explained by a compensatory high catalytic rate (Fields and Somero 1998). For other enzymes, such as cytochrome c oxidase, increased enzyme concentration compensates the higher activation energies in cold-adapted notothenioids compared with temperate and tropical fish (Mark et al. 2012). Notably, an increase in aerobic enzymes concentration is observed in many fish during cold acclimation (see e.g., Lucassen et al. 2006; Orczewska et al. 2010).

Compensation of ATP production through aerobic respiration is not the only adaptation in Antarctic fish that centers on mitochondria. As mentioned earlier, higher cellular densities of these organelles have been reported to favor oxygen diffusion, especially in the hemoglobinless icefish. Therefore, a constantly higher demand for mitochondrial biogenesis might be present in these unique animals. Stably higher enzyme concentrations and increased supply for structural components of mitochondria might be achieved through different molecular mechanisms at the genetic level, for instance, mutations affecting transcriptional regulation, mRNA, and/or protein stability. Alternatively, gene amplification at target loci might selectively increase expression levels.

We tested the hypothesis of selective duplication of genes involved in mitochondrial function and biogenesis in the Antarctic fish genome through massive parallel sequencing of the skeletal muscle transcriptome of *C. hamatus*, which belongs to the notothenioid family showing the most extreme physiological modifications in response to cold. Icefish-specific gene duplications were then identified by comparative transcript annotation against five model fish genomes, using a dedicated bioinformatic pipeline that interrogates an existing database of orthology groups, Ensembl Compara. Finally, a functional enrichment analysis was used to evaluate the biological role of proteins encoded by duplicated loci.

Materials and Methods

Transcriptome Sequencing, Assembly, and Annotation

A normalized cDNA library, constructed from muscle tissue of four *C. hamatus* individuals collected at Terranova Bay, was sequenced by 454 technology (see [Supplementary Material](#) online for details). Raw sequencing data were preprocessed using LUCY (<http://lucy.sourceforge.net/>, last accessed December 12, 2012) to remove poly-A tails escaped to restriction and to trim adaptors and sequences smaller than 60 bases or with Phred quality lower than 30. Transcriptome assembly was performed using MIRA 3 (Chevreux et al. 1999). Two runs of assembly were carried out by MIRA in “EST” and “accurate” usage mode, respectively. Settings adopted were those defined for the 454 sequencing technology. The second run was performed on previously obtained contigs plus discarded reads, which were used as input for MIRA 3. Contigs shorter than 100 nucleotides (nt) or with average quality lower than 30 were discarded.

A first de novo functional annotation of the *C. hamatus* transcriptome was obtained by similarity using Basic Local Alignment Search Tool (BLAST), Blast2GO, and custom-made scripts. BLAST (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>, last accessed December 12, 2012) was run in local mode, and assembled contigs were compared against the National Center for Biotechnology Information (NCBI) nonredundant protein database (*nr*) and to the NCBI nucleotide database (*nt*). Alignments with an *E* value $< 10^{-3}$ were considered significant. Blast2GO suite (<http://www.blast2go.org/>, last accessed December 12, 2012) was used for functional annotation of transcripts through the mapping of gene ontology (GO) terms to contigs having BLAST hits, obtained by BLASTX search against *nr*. Only ontologies retrieved from hits with an *E* value $< 10^{-6}$, an annotation cut off > 55 , and a GO weight > 5 were used for annotation.

Definition of Homology Relationships of *C. hamatus* Transcripts

A custom-made bioinformatic pipeline based on Perl and Python scripts was developed for the definition of homology

relationships of *C. hamatus* transcripts (fig. 1A). In particular, all queries to Ensembl database were carried out using the Ensembl Perl API version 62. Local BLAST queries were accomplished by using the BLAST+ program, version 2.2.22, made available by NCBI.

Similarity Mapping of Contigs to Ensembl Compara Trees

Contigs were used as query for local BLASTX search against Ensembl (version 62) proteins from five fully sequenced fish species: *Danio rerio*, zebrafish; *Gasterosteus aculeatus*, stickleback; *Takifugu rubripes*, Japanese pufferfish; *Oryzias latipes*, medaka; and *Tetraodon nigroviridis*, green spotted pufferfish (fig. 1B). BLAST hits were filtered by *E* value and subject coverage; in particular, only hits with an *E* value of 10^{-3} or less and covering at least 50% of the subject sequence were retained. Moreover, for each contig, a maximum of 20 BLAST hits were considered for further analysis.

BLASTX results were then used to associate *C. hamatus* contigs to Ensembl Compara gene trees. For each contig, and for each species, the Ensembl protein best hit was associated to the encoding transcript, to the gene, and consequently to a Compara gene tree, if any. Only nodes of the tree belonging to the five considered fish species were taken into account. If the best hit could not be associated to a Compara gene tree, the successive BLAST hit was used. This procedure was carried out in parallel for each of the five fish species. The five different contig-Compara gene tree maps were merged to obtain the mapping between contigs and Compara gene tree: A contig was considered associated to a Compara gene tree only if all BLAST best hits from different species (at least 2) pointed to the same Compara gene tree.

Identification of Transcripts Coming from Lineage-Specific Duplications

Compara trees may comprise wide groups of homologous genes, including orthologs and paralogs. To identify lineage-specific duplicates, custom subtree structures were pruned. A subtree was defined as the portion of a Compara tree including only orthologous and co-orthologous genes of the five fish model species (fig. 2).

Sets of *C. hamatus* contigs mapping to the same subtree were further analyzed to discriminate between alternative transcripts of the same gene and transcripts from duplicated genes. Sequences within each set were pairwise compared using BLAST, to identify significant alignments (*E* value $< 10^{-3}$). Pairs of contigs sharing a significant alignment of at least 80 nt (corresponding to the first quartile of exon length distribution in *G. aculeatus*, according to UCSC Genome Browser exon information data, i.e., 75% of exons are longer than 80 nt), with a similarity of at least 70% were considered for discrimination among alternative transcripts and duplicated loci. Contigs with a sequence similarity

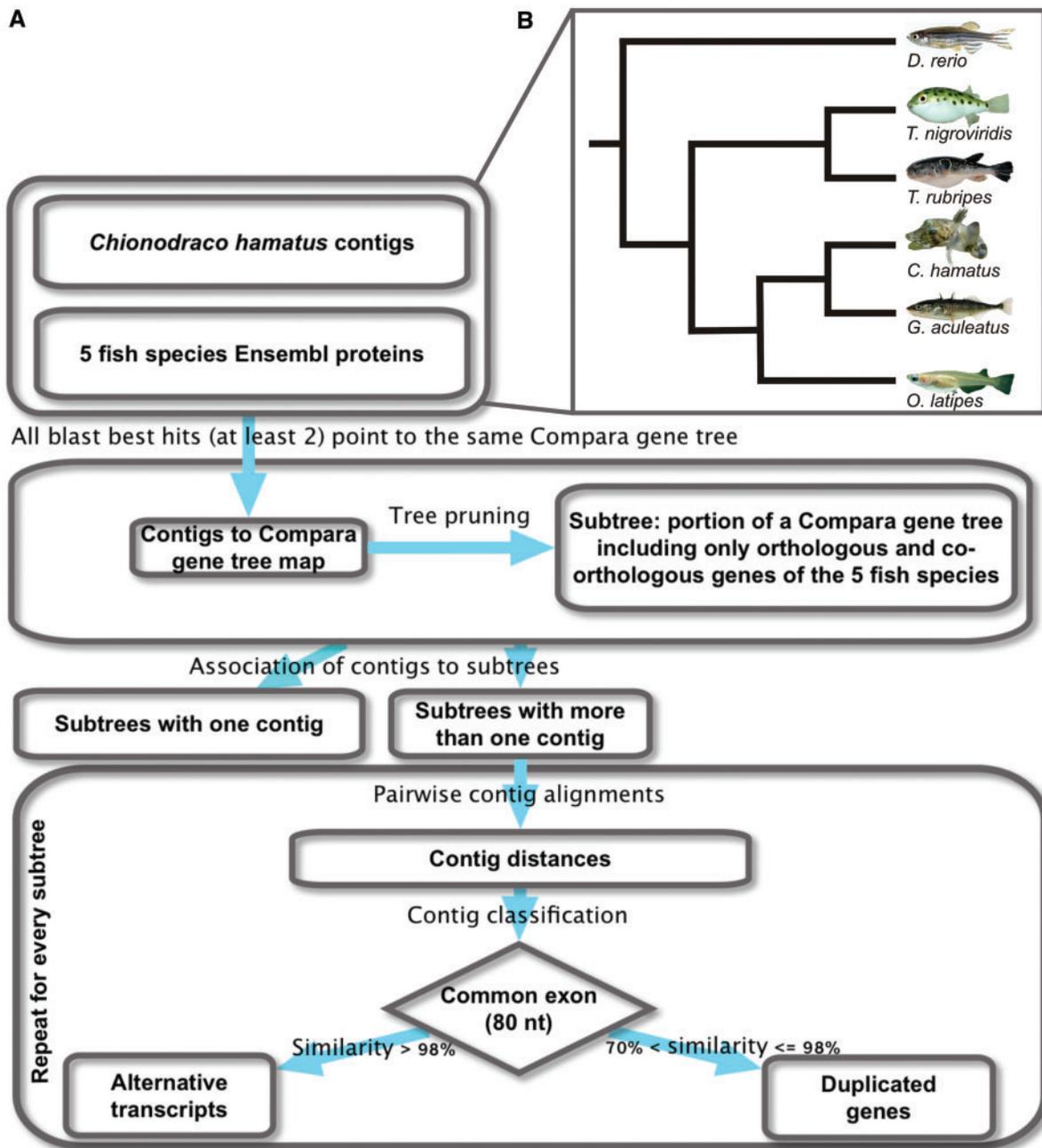


FIG. 1.—Computational pipeline work flow and reference Ensembl Compara fish species tree. (A) Flow chart showing main steps of the computational pipeline devised to assign *Chionodraco hamatus* transcripts to subtrees, that is, unique sets of orthologous or co-orthologous genes in the teleost lineage, and to identify duplicated genes. (B) Schematic phylogeny of the five model species considered as reference plus *C. hamatus* (following Ensembl Compara species tree, based on NCBI taxonomy, and Matschiner et al. (2011) for *C. hamatus* branch position).

higher than 98% were classified as alternative transcripts of the same gene, considering that less than 2% sequence divergence might be due to allele variation and/or sequencing errors within the same locus. Alignment mismatches occurring in homopolimeric stretches longer than 3 nt were not considered for the calculation of sequence similarity between

contigs, thus eliminating the confounding effect originating from the well-known error bias in 454 sequencing technology. To identify *C. hamatus* lineage-specific duplicates, all subtrees were selected in which the number of *C. hamatus* duplicates exceeded those of the closest species *G. aculeatus* (Matschiner et al. 2011).

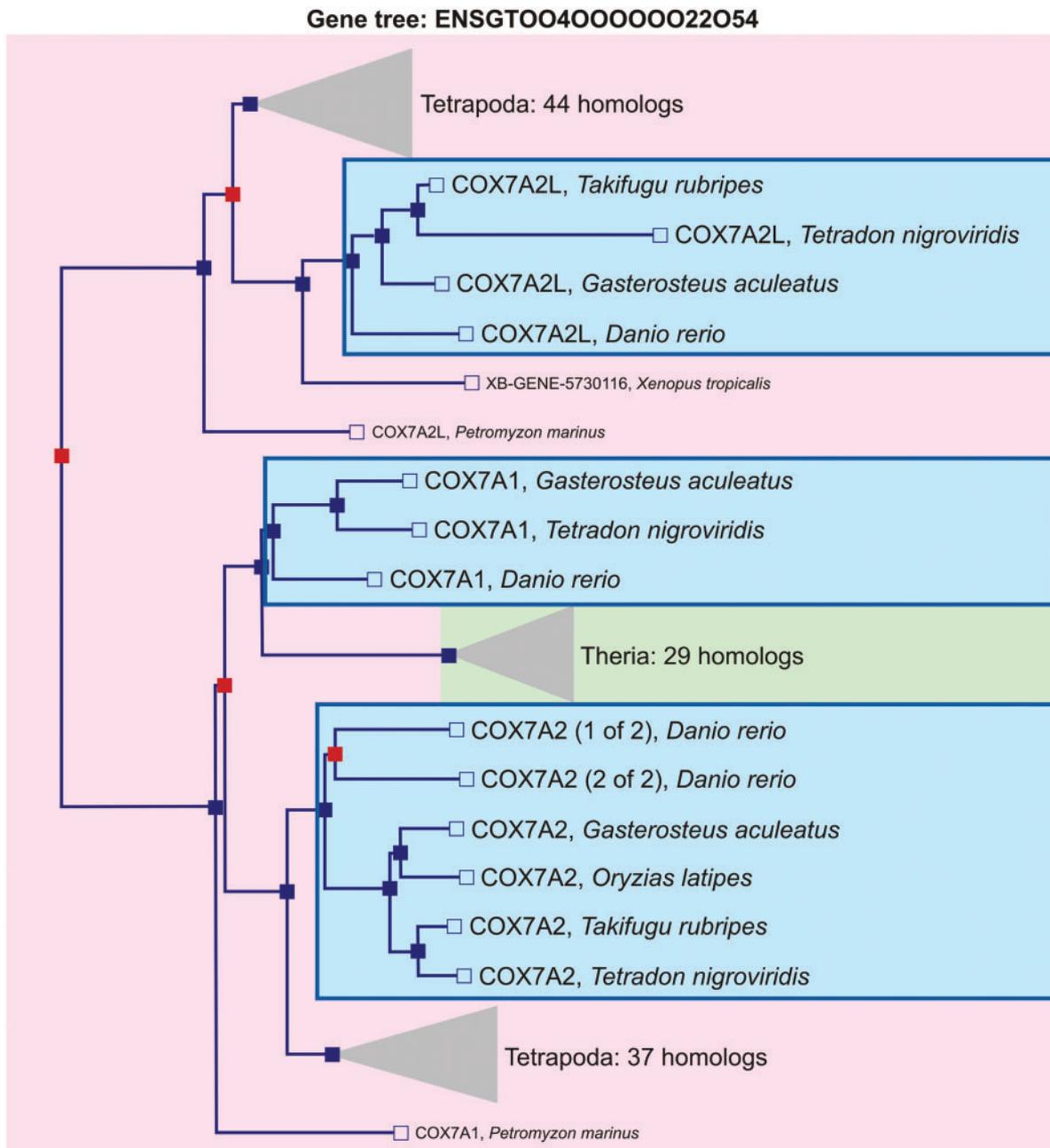


Fig. 2.—An example of Compara gene tree topology resulting in three subtree structures. Blue and red nodes correspond to speciation and duplication nodes, respectively. The subtree (encircled by a blue line) is defined as the portion of a Compara gene tree including orthologous and co-orthologous genes in the teleost lineage.

Functional Enrichment Analysis of *C. hamatus* Putatively Duplicated Genes

Functional enrichment analysis of *C. hamatus* lineage-specific duplicates was performed using DAVID v6.7 (<http://david.abcc.ncifcrf.gov>, last accessed December 12, 2012). The only fish species that is represented in the DAVID knowledge-base is zebrafish, therefore *D. rerio* Entrez gene IDs corresponding to subtrees were used as sequence identifiers to carry

out the analysis. Gene functional enrichment was determined between a “Gene List,” which included the zebrafish gene IDs corresponding to the subtrees to which mapped the putatively duplicated genes in the icefish genome, and a “Background,” which contained the zebrafish gene IDs corresponding to all the considered subtrees. *D. rerio* gene IDs corresponding to each subtree were counted only once regardless the number of *C. hamatus* putative gene copies. Thresholds for Gene

count and EASE score, which provides the significance of gene-term enrichment with a modified Fisher's exact test, were set equal to 2 and 0.05, respectively.

MitoCarta-Based Identification of Duplicated Genes Encoding Mitochondrial Proteins

A list of genes encoding proteins with mitochondrial localization was downloaded from MitoCarta (Pagliarini et al. 2008), an inventory of mouse genes encoding proteins with strong support for mitochondrial localization. As MitoCarta uses only gene symbols, Ensembl BioMart (<http://www.ensembl.org/biomart/martview>, last accessed December 12, 2012) was used to obtain Ensembl gene IDs for mouse mitochondrial proteins. A custom pipeline was developed to obtain, from Ensembl Compara, the list of fish orthologous Ensembl gene IDs for every mitochondrial mouse gene and to tag *C. hamatus* transcripts, and hence *C. hamatus* genes, as mitochondrial. The pipeline uses results from BLAST against Ensembl fish proteins to associate Ensembl gene IDs to *C. hamatus* transcripts and employs this information to tag *C. hamatus* transcripts and consequently genes as mitochondrial by using the mitochondrial fish genes obtained in the previous step. Finally, a hypergeometric distribution was used to test the enrichment for mitochondrial genes in the set of *C. hamatus* duplicated genes when compared with the set of all reconstructed *C. hamatus* genes. All computations were performed with the R statistical software.

The *C. hamatus* Skeletal Muscle Transcriptome Database

A database, freely available online at <http://compgen.bio.unipd.it/chamatusbase/> (last accessed December 12, 2012), was implemented using SQLite and Django web framework. The database is filled with different layers of information regarding the *C. hamatus* transcriptome sequences and analysis results. For each contig, different data and bioinformatic analysis results are provided. All data in the database can be searched by keywords, GO terms, or by sequence using a BLAST web interface. Moreover, a query system for massive data retrieval is available.

Results

Chionodraco hamatus Skeletal Muscle Transcriptome 1.0

A total of 382,386 raw reads were obtained, with an average sequence length of 297 nt. After removal of short reads and trimming of low-quality sequence regions, the remaining 341,171 reads (SRA052291) were assembled into contigs with two runs of assembly (supplementary table S1, Supplementary Material online). After length- and quality-based filtering, a final set of 23,968 contigs (with mean length, average quality, and read coverage, respectively, 595 nt, 48, and 13) represents the *C. hamatus* muscle

transcriptome (supplementary table S2 and fig. S1A–C, Supplementary Material online).

For 9,556 transcripts (40%), BLAST identified significant similarity with known proteins. In total, 109,307 *nr* hits were found, with an average of 11.4 hits per contig (94% of best hits E value $<10^{-6}$ and 89% E value $<10^{-9}$). In parallel, 7,155 contigs (30%) had significant hits among known nucleotide sequences (94% of best hits E value $<10^{-6}$ and 88% E value $<10^{-9}$). Comprehensively, 48% of the *C. hamatus* transcriptome (11,509 contigs) resulted to be similar to known protein or DNA sequences.

In total, 7,293 contigs were associated to one or more of 5,805 unique GO terms, for a total of 61,170 term occurrences (5.9 ± 1.8 terms per contig). Supplementary figure S1D, Supplementary Material online, reports the slim-based term classification of GO terms.

Identification of Lineage-Specific Duplicated Genes in *C. hamatus*

Using the bioinformatics pipeline depicted in figure 1A, assembled contigs were first univocally assigned to one Ensembl Compara gene tree. To this end, results of contigs BLAST against five fish species Ensembl proteins were used (table 1). Each tree corresponds to a homology group. Then, the structure of each gene tree was split to obtain one or more new entities, called subtrees, defined as a portion of a gene tree including orthologous or co-orthologous genes of five fish model species (*D. rerio*, *G. aculeatus*, *T. rubripes*, *O. latipes* and *Tet. nigroviridis*) (fig. 2). A total of 2,401 *C. hamatus* contigs were assigned to 1,712 subtrees, that is, unique sets of orthologous or co-orthologous genes in the teleost lineage. For 1,232 subtrees, a single icefish contig was found, whereas 480 subtrees were associated with two or more *C. hamatus* transcripts for a total of 1,169 transcripts.

Subsequently, a highly stringent filtering process to exclude potential alternative transcripts/alleles from the same locus was applied (fig. 1A), yielding 353 transcripts, which could be considered the product of distinct genetic loci and were associated with 127 subtrees. In 124 subtrees, the number of *C. hamatus* gene copies exceeded that observed in the closest species, *G. aculeatus*, suggesting that for each orthology group, one or more gene duplication events occurred after the separation between the stickleback and icefish lineages. The number of gene copies ranged from 2 to 8 (2.26 on average), for a total of 276 putatively duplicated genes (supplementary table S3, Supplementary Material online). For each subtree, the proportion of *C. hamatus* duplicated genes was compared with the proportion of duplicated genes in each of the five considered fish genomes. Because duplicated genes of each species were counted in each subtree, only lineage-specific duplications were considered. Using a test of equal proportion, *C. hamatus* showed a significantly higher proportion (P value with Bonferroni correction for multiple

Table 1Results of *Chionodraco hamatus* Contigs BLASTX Comparison against Ensembl Proteins of the Five Model Teleosts

Species	Contigs with Significant Hits (E value $<10^{-3}$)	Contigs with Significant Hits Covering At Least 50% of the Subject Sequence	Total Number of Protein Hits
<i>Danio rerio</i>	9,203	5,084	165,094
<i>Gasterosteus aculeatus</i>	9,411	4,776	113,220
<i>Takifugu rubripes</i>	8,958	4,485	148,743
<i>Oryzias latipes</i>	9,070	4,740	106,126
<i>Tetraodon nigroviridis</i>	8,764	4,627	98,293

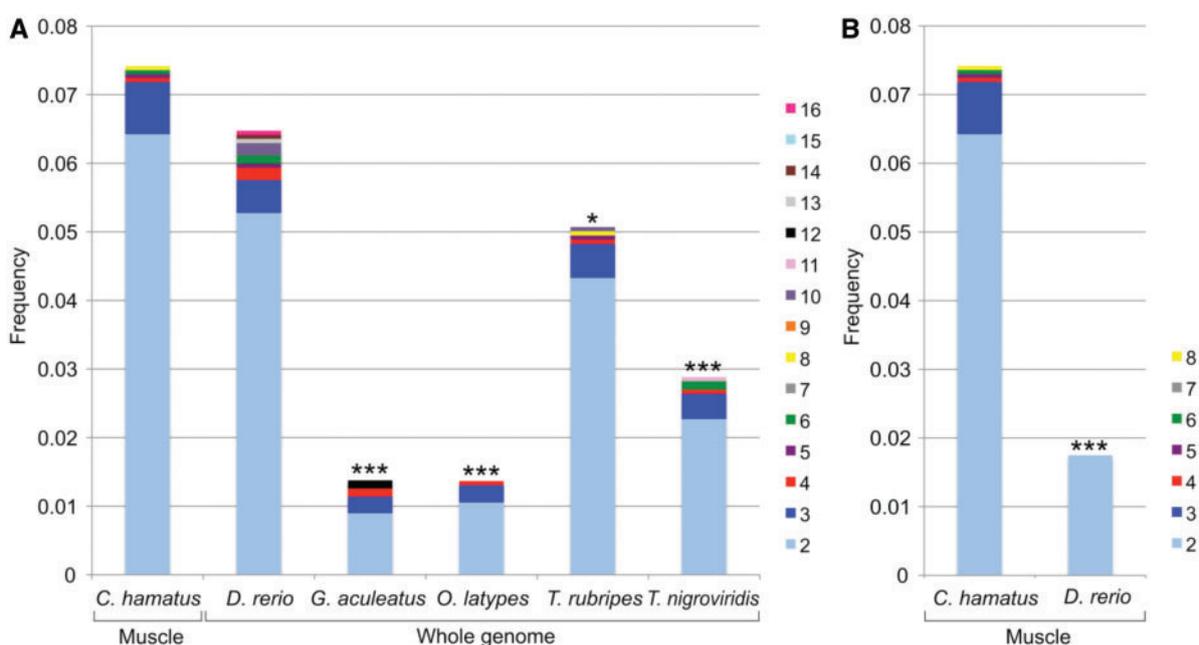


Fig. 3.—Comparison of duplicated gene frequencies in different fish species. (A) Bar plots represent, for each species, the fraction of analyzed subtrees with at least one gene represented, including two or more duplicated genes. The proportion of *Chionodraco hamatus* duplicated genes, inferred in this study from the analysis of the muscle transcriptome, is pairwise compared with the proportion of duplicated genes of the five fish species for which the whole genome is available. Test of equal proportions: *** P value <0.001 , ** P value <0.01 , and * P value <0.05 , after Bonferroni correction for multiple tests. (B) Comparison of the proportion of duplicated genes expressed in muscle in *C. hamatus* and *D. rerio*.

testing <0.05) of duplicated genes in all comparisons with the other species, except against *D. rerio* (fig. 3A). The evidence obtained for a significantly higher occurrence of gene duplications in *C. hamatus* is likely underestimated, because gene copy number was inferred based on transcriptome data from a single tissue and compared with whole-genome information for fish model species. In fact, limiting the comparison to zebrafish genes expressed in the skeletal muscle (expression data are available for *D. rerio* only and provided by ZFIN) produced a significant difference in the proportion of duplicated loci (fig. 3B).

Functional Enrichment Analysis of Duplicated Loci

Out of 124 subtrees/orthology groups that showed at least one icefish-specific gene duplication, 6 subtrees could not be

associated to any zebrafish gene ID (five of these subtrees did not include any *D. rerio* gene) and one additional subtree was not associated with any functional information in DAVID. Using zebrafish functional annotations for each subtree, a total of 11 terms resulted to be significantly over-represented (P value <0.05), which include 2 KEGG pathways, 1 SP-PIR keyword, 1 biological process (BP) GO term, 6 molecular function (MF) GO terms, and 1 cellular component (CC) GO term (table 2). Overall, most enriched terms are related to two key cellular processes, namely protein translation and oxidative phosphorylation.

A total of 1,098 mouse genes encoding proteins with strong support for mitochondrial localization were downloaded from MitoCarta. When converting gene symbols given by MitoCarta to Ensembl gene IDs, a total of 1,057

Table 2Functional Enrichment Analysis of 124 *Chionodraco hamatus* Putatively Duplicated Genes Using DAVID Web Tool

Category	Term	Gene Count	P	FE
GOTERM_MF	GO:0003735 structural constituent of ribosome	16	0.0003	2,6
GOTERM_MF	GO:0005198 structural molecule activity	17	0.0006	2,4
SP_PIR_KEYWORDS	Ribosomal protein	15	0.0008	2,6
GOTERM_BP	GO:0006412 translation	17	0.0017	2,2
KEGG_PATHWAY	dre03010 ribosome	12	0.0076	2,3
KEGG_PATHWAY	dre00190 oxidative phosphorylation	12	0.0250	2,0
GOTERM_CC	GO:0005840 ribosome	17	0.0280	1,6
GOTERM_MF	GO:0016675 oxidoreductase activity, acting on heme group of donors	4	0.0440	4,7
GOTERM_MF	GO:0015002 heme-copper terminal oxidase activity	4	0.0440	4,7
GOTERM_MF	GO:0016676 oxidoreductase activity, acting on heme group of donors, oxygen as acceptor	4	0.0440	4,7
GOTERM_MF	GO:0004129 cytochrome-c oxidase activity	4	0.0440	4,7

NOTE.—Reported are GO terms, KEGG pathways, and SP-PIR keywords significantly over-represented (P value <0.05) in the set of duplicates. For each enriched term, the number of genes involved in the term (Gene count), enrichment P value, and fold enrichment (FE) are indicated.

(96%) genes were obtained. A custom pipeline was developed to obtain, from Ensembl Compara, the list of 5,826 fish orthologs to mitochondrial mouse genes. The set of 5,826 fish mitochondrial orthologs was composed by 1,378 *D. rerio* genes, 1,115 *G. aculeatus* genes, 1,114 *T. rubripes* genes, 1,061 *O. latipes* genes, and 1,158 *Tet. nigroviridis* genes. Subsequently, a mini pipeline was developed to tag *C. hamatus* contigs as mitochondrial. Using results of BLAST against Ensembl fish proteins, *C. hamatus* contigs were associated to an Ensembl gene ID and tagged as mitochondrial if the corresponding Ensembl gene ID was present among the 5,826 fish mitochondrial genes. As a result, 34 of 124 putative duplicated genes and 274 of 1,588 nonduplicated genes were tagged with mitochondrial localization (table 3 reports the set of 34 *C. hamatus* putative duplicated genes encoding proteins with mitochondrial localization). A hypergeometric distribution was then used to test the enrichment for mitochondrial genes in the set of 124 putative duplicated *C. hamatus* genes when compared with the set of all reconstructed *C. hamatus* genes. A P value of 0.0235 indicates a significant enrichment of mitochondrial genes in the group of duplicated *C. hamatus* genes.

Discussion

This study describes the *C. hamatus* skeletal muscle transcriptome. Deep sequencing information, by means of a dedicated bioinformatic pipeline, disclosed putative duplicated genes specific to the *C. hamatus* lineage. This unique investigation licensed exploration of the genetic and genomic adaptations to extreme cold conditions. A global functional enrichment analysis of putative duplicated genes and the localization and biological role of single-encoded proteins supported our hypothesis that duplicated genes were involved in mitochondrial function and biogenesis.

Gene and genome duplications are thought to have played a major role in genome evolution and in the generation of the enormous phenotypic variability that exists among living species. In the teleost lineage, duplicated genes were shown to have significantly impacted the evolution of key traits, including immune response (Van der Aa et al. 2009), development (Zou et al. 2009), sensory systems (Rennison et al. 2012), adaptation to different biological environments (Nomiyama et al. 2008), endocrine system (Arterbery et al. 2011), and body plans (Guo et al. 2010). DNA sequences generated by duplication increase genome redundancy, a distinctive trait of eukaryotic genomes, which offers the raw genetic material for the evolution of genetic novelties. Four major outcomes of gene duplication have been proposed: 1) both duplicates may be preserved in a genome with one copy evolving new functional properties and the other copy retaining the original function (“neofunctionalization”) (Ohno 1970); 2) the two copies may be maintained for a selective advantage in increasing the synthesis of the ancestral gene product (“dosage effect”) (Ohno 1970); 3) one copy may accumulate degenerative mutations and be silenced (“nonfunctionalization”) (Ohno 1972); and 4) both copies may be partially impaired by deleterious mutations, so that the ancestral function is divided between duplicated loci (“subfunctionalization”) (Force et al. 1999).

Duplication events may involve either the whole genome (whole-genome duplication [WGD]) or single genes or genomic regions (segmental duplication). WGDs occurred in many lineages during evolution (Van de Peer et al. 2009). Two WGD events appear to have occurred early in the vertebrate evolution, whereas the third round of WGD involved the teleost lineage soon after its divergence from tetrapods (Dehal and Boore 2005; Meyer and Van de Peer 2005). A vast literature exists regarding WGD events and their evolutionary implications in vertebrates. Much less attention has been paid to recent gene duplicates and their evolutionary

Table 3Thirty-Four *Chionodraco hamatus* Putative Duplicated Genes Encoding Proteins with Mitochondrial Localization

Gene Symbol	Gene Description	Number of <i>C. hamatus</i> Duplicated Gene Copies
<i>acn9</i>	<i>ACN9</i> homolog (<i>Saccharomyces cerevisiae</i>) (Source: ZFIN; Acc: ZDB-GENE-041010-94)	2
<i>atp5s</i>	<i>ATP synthase, H⁺ transporting, mitochondrial F₀ complex, subunits</i> (Source: ZFIN; Acc: ZDB-GENE-040426-959)	2
<i>bola1</i>	<i>bolaA</i> homolog 1 (<i>Escherichia coli</i>) (Source: HGNC Symbol; Acc:24263)	2
C13H6orf57	Chromosome 6 open reading frame 57 (Source: HGNC Symbol; Acc:20957)	2
C4H1orf31	chromosome 1 open reading frame 31 (Source: HGNC Symbol; Acc:18025)	2
<i>coa5</i>	<i>Cytochrome C oxidase assembly factor 5</i> (Source: HGNC Symbol; Acc:33848)	3
<i>cox17</i>	<i>COX17 cytochrome c oxidase assembly homolog</i> (<i>S. cerevisiae</i>) (Source: ZFIN; Acc: ZDB-GENE-040912-91)	8
<i>cox4i1</i>	<i>Cytochrome c oxidase subunit IV isoform 1</i> (Source: ZFIN; Acc: ZDB-GENE-030131-5175)	2
<i>cox6a2</i>	<i>Cytochrome c oxidase subunit VIa polypeptide 2</i> (Source: HGNC Symbol; Acc:2279)	3
<i>cox7a2</i>	<i>Cytochrome c oxidase, subunit VIIa 2</i> (Source: ZFIN; Acc: ZDB-GENE-050522-153)	2
<i>etfa</i>	<i>Electron-transfer-flavoprotein, alpha polypeptide</i> (Source: ZFIN; Acc: ZDB-GENE-030131-4449)	2
<i>fdx1</i>	<i>Ferredoxin 1</i> (Source: ZFIN; Acc: ZDB-GENE-071015-2)	2
<i>glrx2</i>	<i>Glutaredoxin 2</i> (Source: ZFIN; Acc: ZDB-GENE-040718-101)	2
<i>mrpl-13</i>	<i>Mitochondrial ribosomal protein L13</i> (Source: ZFIN; Acc: ZDB-GENE-050522-167)	2
<i>mrpl-17</i>	<i>Mitochondrial ribosomal protein L17</i> (Source: ZFIN; Acc: ZDB-GENE-040912-133)	2
<i>mrpl-2</i>	<i>Mitochondrial ribosomal protein L2</i> (Source: HGNC Symbol; Acc:14056)	2
<i>mrpl-48</i>	<i>Mitochondrial ribosomal protein L48</i> (Source: ZFIN; Acc: ZDB-GENE-041008-125)	2
<i>mrpl-52</i>	<i>Mitochondrial ribosomal protein L52</i> (Source: HGNC Symbol; Acc:16655)	2
<i>mrpl-55</i>	<i>Mitochondrial ribosomal protein L55</i> (Source: HGNC Symbol; Acc:16686)	2
<i>mterfd1</i>	<i>MTERF domain containing 1</i> (Source: ZFIN; Acc: ZDB-GENE-040718-359)	2
<i>ndufa11</i>	<i>zgc:66391</i> (Source: ZFIN; Acc: ZDB-GENE-040426-1631)	2
<i>ndufa5</i>	<i>NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13 kDa</i> (Source: HGNC Symbol; Acc: 7688)	2
<i>ndufa6</i>	<i>NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6</i> (Source: ZFIN; Acc: ZDB-GENE-040426-1124)	2
<i>ndufb5</i>	<i>NADH dehydrogenase (ubiquinone) 1 beta subcomplex 5</i> (Source: ZFIN; Acc: ZDB-GENE-011010-1)	2
<i>ndufb8</i>	<i>NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8</i> (Source: ZFIN; Acc: ZDB-GENE-040426-1858)	2
<i>ndufs7</i>	<i>NADH dehydrogenase (ubiquinone) Fe-S protein 7, (NADH-coenzyme Q reductase)</i> (Source: ZFIN; Acc: ZDB-GENE-041111-261)	2
<i>oxa1l</i>	<i>Oxidase (cytochrome c) assembly 1-like</i> (Source: ZFIN; Acc: ZDB-GENE-071004-49)	2
<i>romo1</i>	<i>Reactive oxygen species modulator 1</i> (Source: HGNC Symbol; Acc:16185)	3
<i>sdhc</i>	<i>Succinate dehydrogenase complex, subunit C, integral membrane protein</i> (Source: ZFIN; Acc: ZDB-GENE-040801-26)	2
<i>timmi9</i>	<i>Translocase of inner mitochondrial membrane 9 homolog</i> (Source: ZFIN; Acc: ZDB-GENE-021206-14)	2
<i>tmem14c</i>	<i>Transmembrane protein 14C</i> (Source: ZFIN; Acc: ZDB-GENE-060825-156)	3
<i>uqcrb</i>	<i>Ubiquinol-cytochrome c reductase binding protein</i> (Source: ZFIN; Acc: ZDB-GENE-050522-542)	2
<i>uqcrc2a</i>	<i>ubiquinol-cytochrome c reductase core protein IIa</i> (Source: ZFIN; Acc: ZDB-GENE-040718-405)	2
<i>zgc:56493</i>	<i>zgc:56493</i> (Source: ZFIN; Acc: ZDB-GENE-030131-8581)	5

NOTE.—The number of *C. hamatus* putative duplicated gene copies is reported for each gene, identified by a gene symbol, and a gene description (provided by Ensembl BioMart using *Danio rerio* Ensembl gene ID of the related subtree).

consequences, which may reveal specific adaptation processes in individual species or lineages, as in *C. hamatus*. A genome-wide analysis of four mammalian genomes (human, macaque, mouse, and rat), focusing on recent gene duplicates, revealed that 10% of all lineage-specific duplicated pairs evolved under positive selection, which is a higher percentage than

a comparable set of single-copy genes (Han et al. 2009), suggesting that analysis of recent duplicates may reveal significant footprints of adaptive change. Human-specific duplicates identified in the study were involved in the development of neural and cognitive functions (Han et al. 2009).

The prediction of orthology and paralogy relationships between genes is a central issue in comparative genomics. Typically, orthologs are defined as homologous sequences, which diverge as a consequence of a speciation event, whereas paralogs are homologous sequences originating from a duplication node. Various methods have been proposed for the inference of reliable orthology/paralogy relationships at the genome-wide level. Some methods rely on pairwise sequence comparison between genomes, whereas an increasing number of web tools are based on phylogenetic approaches, following the evolutionary basis of the original orthology/paralogy definition (Altenhoff and Dessimoz 2012). In Ensembl Compara, gene phylogenies are computed using the TreeBest method, which reconciles calculated gene tree topologies with known species trees to determine whether a node corresponds to a duplication or a speciation event (Vilella et al. 2009). Ensembl Compara is a wide-ranging collection of gene trees, spanning 91% of genes across vertebrates and including five well-annotated fish species: *D. rerio*, *G. aculeatus*, *T. rubripes*, *O. latipes*, and *Tet. nigroviridis*.

The final aim of this study was to distinguish between duplicated genes originating from WGD events and *C. hamatus* lineage-specific duplications, which may offer new insights into the genetic/genomic mechanisms of cold adaptation in icefish. We used a targeted computational pipeline relying on Ensembl Compara phylogenetic resources and exploiting deep sequencing information of the *C. hamatus* muscle transcriptome. Using a quite stringent similarity-based approach (all BLAST best hits from different species pointing the same Compara gene tree), a subset of 2,401 *C. hamatus* transcripts was univocally assigned to custom subtree structures, which are portions of an Ensembl Compara gene tree including unique sets of orthologous or co-orthologous genes in the teleost lineage. By definition, duplicated genes resulted from WGD events should end up in different subtrees, with the latter including only duplicated genes arisen independently in each of the model species lineages. This is true also for *C. hamatus* mapped transcripts, because duplicated genes were identified and counted inside each subtree.

We found 124 subtrees with *C. hamatus* lineage-specific duplications, with the number of duplicates per gene ranging from 2 to 8. Moreover, *C. hamatus* displayed a significantly higher proportion of lineage-specific duplicates than the other model fish species, excepting *D. rerio*, which, however, is known to have a highly duplicated genome. This difference is striking considering that it was obtained comparing whole-genome data from model species (all transcripts, no matter of expression characteristics), with *C. hamatus* duplicated genes expressed in one tissue (skeletal muscle). Limiting the comparison to muscle expressed genes in *D. rerio*, *C. hamatus* displayed a significantly higher proportion of duplicated genes. It is worth noting that we are not comparing the level of retention of duplicated genes resulted from WGD events but the

amount of duplicated genes arisen thereafter and independently in each lineage.

We reasoned that identified duplicated genes could be simply genes free to duplicate with minimal negative outcomes. In this case, we would expect that the same genes might be duplicated in other fish. From a qualitative point of view, the overlap between the set of *C. hamatus* duplicated genes and that of zebrafish is narrow (out of 124 subtrees with *C. hamatus* lineage-specific duplications, only 6 also include zebrafish duplicates) showing that these two sets are different in composition. Moreover, in none of these six subtrees, we observed more than one zebrafish gene copy expressed in muscle. Thus, although zebrafish duplicated genes evolved as tissue-specific isoforms, *C. hamatus* duplicates might play a role in muscle.

Functional enrichment analyses, based on GO terms, pathways and MitoCarta information, showed that *C. hamatus* duplicated loci are enriched in mitochondrial function and respiratory activities. In fact, "oxidative phosphorylation" was one of the most represented KEGG pathways identified by functional enrichment analysis of duplicates. Moreover, enrichment analysis for mitochondrial proteins using MitoCarta information confirmed that at least 34 out of 124 duplicated genes encode proteins with mitochondrial localization, leading to a significant enrichment of mitochondrial function in duplicated genes. Additionally, gene-by-gene manual annotation based on relevant literature showed that many proteins encoded by duplicated loci are involved in mitochondrial biogenesis and function. Thus, these results not only support a trend for genome expansion accompanying the evolution of phylogenetically derived notothenioid families reported by Chen et al. (2008) and Detrich and Amemiya (2010) but also confirm the working hypothesis of targeted gene duplications at loci involved in mitochondrial function and biogenesis. This observation might explain, at least in part, the high mitochondrial density documented in Antarctic notothenioids. These fish possess an increased mitochondrial abundance compared with Perciformes from warmer climates (Johnston et al. 1998), with an extreme value recorded for the active and pelagic *Pleuragramma antarcticum* (Johnston et al. 1988).

There are three lines of evidence showing that a high mitochondrial density is advantageous at low temperatures: 1) it facilitates ATP production thanks to higher concentrations of enzymes involved in aerobic metabolism; 2) it enhances intracellular oxygen diffusion through the mitochondrial lipid membrane as oxygen solubility in phospholipids is approximately four times higher than in water; and 3) it reduces the mean diffusion distance for oxygen transfer from capillaries to mitochondria (O'Brien 2011). The lack of myoglobin expression in skeletal muscle is displayed by both white- and red-blooded notothenioids (Sidell et al. 1997; Moylan and Sidell 2000). All icefish completely lack hemoglobin (Ruud 1954), and six species do not express myoglobin in cardiac myocytes (Borley and Sidell 2010). Thus, oxygen supply and

ATP generation are extremely challenging in white-blooded icefish. As reviewed by O'Brien and Mueller (2010), mitochondrial density in icefish muscle reaches values that largely exceed those reported for red-blooded notothenioids, both in skeletal muscle and in the myocardium, with a particular increase in nonexpressing myoglobin icefish. Icefish mitochondria were indeed found to be structurally different from that of red-blooded notothenioids, displaying a particularly high lipid-to-protein ratio (O'Brien and Mueller 2010), suggesting a major role of lipids in facilitating oxygen storage and diffusion in cells in the absence of oxygen-carrying proteins (Sidell 1998).

Together, the progressive expansion of the mitochondrial compartment in Antarctic notothenioids, the enrichment analyses performed at the gene category level, and a close examination of recent literature regarding single protein functions indicate a role of duplicated genes in mitochondrial function and biogenesis. Among identified mitochondrial proteins, 12 are structural subunits of the electron transport chain complexes localized in the mitochondrial inner membrane. In addition, multiple gene copies (from 3 to 8) were detected for *COX17 cytochrome c oxidase assembly homolog (cox17)*, *COX19 cytochrome c oxidase assembly homolog (cox19)*, and *cytochrome c oxidase assembly factor 5 (coa5)*, which are involved in the assembly of complex IV (Horng et al. 2004; Rigby et al. 2007; Huigsloot et al. 2011). MTERF domain containing 1 (*Mterfd1*) is an important regulator of mitochondrial transcription and is required for the normal assembly of mitochondrial respiratory complexes (Park et al. 2007). The duplicated *electron-transfer-flavoprotein alpha polypeptide (efta)* gene codes for one of the two subunits of an electron-transferring flavoprotein, a mitochondrial matrix enzyme located at a key metabolic branch point, which shuttles electrons from various primary dehydrogenases to the main respiratory chain (Toogood et al. 2007). *Translocase of inner mitochondrial membrane 9 homolog (timm9)* encodes a subunit of the Tim9-Tim10 complex (or Tim9-Tim10-Tim12 complex), a mitochondrial intermembrane chaperone that mediates the translocation of nuclear-encoded hydrophobic precursor proteins into the intermembrane space and is thus essential for the import of inner membrane proteins (Schmidt et al. 2010). Moreover, *oxidase (cytochrome c) assembly 1-like (oxa1l)* codes for an integral protein of the mitochondrial inner membrane and represents the key component of the machinery that mediates insertion of both nuclear and mitochondrial translation products into the inner membrane (Hell et al. 1998; Ott and Herrmann 2010). In mammals, Oxa1l cross-links to six mitochondrial ribosomal proteins (MRPs) (*Mrpl-13*, *Mrpl-20*, *Mrpl-28*, *Mrpl-48*, *Mrpl-49* and *Mrpl-51*) (Haque et al. 2010); notably, *mrpl-13* and *mrpl-48* genes were duplicated in our analysis.

Mitochondrial ribosomes seem to have diverged substantially during eukaryotic evolution because protein composition is highly variable between different species, both in the

number and in the type of contained MRPs (Smits et al. 2007). Newly acquired MRPs, which were shown to have originated by several gene duplication events (Smits et al. 2007), may have contributed to the introduction of novel and/or more specialized functions. *Mrpl-13* and *mrpl-48*, which originated by duplication of *mrps-10* in the metazoan lineage, and four additional genes encoding MRPs were found to be duplicated in our analysis. All of them are components of the large ribosomal subunit involved in peptide bonds formation and thus more prone to cooperate with co- and/or post-translational interacting factors. As Gruschke and Ott (2010) suggested, modifications at this level may result in novel ways to organize the biogenesis of mitochondrially encoded proteins, that is, the major components of the mitochondrial respiratory complexes.

Another interesting duplication is *reactive oxygen species modulator 1 (romo1)*, encoding the first mitochondrial inner membrane protein known to be involved in the regulation of mitochondrial fission in mammalian cells (Zhao et al. 2009). Enhanced expression of *romo1* was shown to inhibit nuclear DNA synthesis and mitosis and, in particular, to induce mitochondrial fragmentation, resulting in numerous sphere-shaped organelles with no noticeable internal structural alterations (Zhao et al. 2009). The duplication of *romo1* may be associated with the need to coordinate cell cycle with mitochondrial proliferation. Increasing mitochondrial abundance without a concomitant increase in cell division would be a valuable strategy to obtain higher mitochondrial densities in cells.

An additional putative duplicated gene, possibly involved in the regulation of mitochondrial morphology, encodes the ATP synthase, H⁺ transporting, mitochondrial F₀ complex, and subunit s (*Atp5s*). *Atp5s* is an essential subunit of the ATP synthase complex that couples proton translocation through the F₀ subunit with ATP synthesis at F₁ subunit (Belogrudov and Hatefi 2002). The protein is also responsible for preventing passive proton diffusion through F₀ and for regulating proton electrochemical gradient across the inner membrane (Belogrudov 2008, 2010). *Atp5s* may enhance ATP production under proton-limited conditions and improve the efficiency of oxidative phosphorylation in animal mitochondria (Belogrudov 2009). Icefish mitochondria display an increased coupling of proton transport and ATP synthesis compared with red-blooded notothenioids, which, in turn, show a higher coupling than temperate teleosts (Mueller et al. 2011). Belogrudov (2010) showed that overexpression of *atp5s* affects mitochondria morphology in cultured animal cells, with thicker and more packaged cristae of the inner membrane, likely as a consequence of enhanced protein-protein interactions. This structural modification is supposed to increase the rate of exchange of ATP synthasome substrates across the inner membrane and/or the rate of substrate conduction along the respiratory chain. This might explain the 2-folds higher respiratory control ratio found in *atp5s*-overexpressing

cells, suggesting increased coupling between respiration and ATP production (Belogrudov 2010).

The duplication of *dynactin 6* (*dctn6*) also supports the hypothesis of increased mitochondrial biogenesis coupled with enhanced lipid synthesis in the icefish lineage. Dynactin is a major component of the activator complex that stimulates dynein-mediated vesicle transport, a highly ATP-dependent and key cellular biological process. Upregulation of this gene, which is not present in MitoCarta, is associated with the dramatic proliferation of mitochondria in skeletal muscle of mice following the genetic inactivation of the mitochondrial adenine nucleotide translocator-1, which exports mitochondrial ATP to the cytosol (Murdock et al. 1999). Mouse *Dctn6* is understood to play a role in mitochondrial lipid synthesis due to protein sequence similarity to an enzyme involved in bacterial lipid A biosynthesis (Murdock et al. 1999). The mitochondrial fatty acid synthetic pathway is present in many eukaryotic organisms and has been shown to be of particular importance for normal mitochondrial morphology and function (Hiltunen et al. 2010). Duplication of *dctn6* may be relevant for cold adaptation because it is probably a multifunctional protein involved in vesicle transport regulation, vasopressin-regulated water reabsorption, and in mitochondrial biogenesis, via lipid biosynthesis. However, it should be emphasized here that very little is known about how the biosynthesis of mitochondrial phospholipids is integrated into the process of mitochondrial biogenesis (O'Brien and Mueller 2010).

Oxygen solubility increases at lower temperatures, thus Antarctic fish are constantly exposed to high oxygen tensions, which enhance reactive oxygen species (ROS) production and levels of oxidative stress. Chen et al. (2008) have documented a potential evolutionary response in Antarctic notothenioids using an array-based comparative genomic hybridization technique. Nine genes involved in stress response, anti-ROS, and antiapoptotic processes were found to be duplicated in three Antarctic notothenioids when comparing the hybridization intensity of their DNA with that of two non-Antarctic and phylogenetically basal notothenioids (Chen et al. 2008). These nine genes were not detected as duplicated in our analysis, possibly because of their missed transcription in skeletal muscle, transcriptome incompleteness, the conservative approach followed by our pipeline, or because duplicated gene copies were not divergent enough to be detected by sequence analysis. However, several genes associated with anti-ROS activities or upregulated in the presence of high ROS levels were found to be duplicated. Among them, *BCL2/adenovirus E1B interacting protein 3* (*bnip3*) is activated in ventricular myocytes in response to stress when mitochondria are severely damaged, inducing mitochondrial dysfunction and subsequent cell death (Regula 2002; Gustafsson 2011). Recently, a *bnip3* apoptosis-independent function was discovered in the myocardium, where high mitochondrial volume densities are required to satisfy the large ATP demand. *Bnip3* expression

seems to be essential to regulating mitochondrial turnover by selective autophagy (Gustafsson 2011). Long-term studies in aged *bnip3*-knockout mice showed that *Bnip3* deficiency promotes the accumulation of dysfunctional mitochondria and the development of cardiac disease (Dorn 2010). Given that *bnip3* functions as a redox sensor activated in response to localized accumulation of oxidative stress (Kubli et al. 2008), clearly a constitutively increased expression of this gene in Antarctic notothenioids, via gene duplication, may improve mitochondrial turnover efficiency in extremely oxidizing conditions.

Another duplicated gene with anti-ROS function is *glutaredoxin 2* (*glrx2*), encoding a glutathione-dependent oxidoreductase activated by oxidative stress, which facilitates the maintenance of mitochondrial redox homeostasis, protects cells from oxidative damage, and inhibits mitochondrially mediated apoptosis (Enoksson et al. 2005; Lillig et al. 2005). Additionally, *zgc:56493* (for which five gene copies were detected) and *zgc:92254*, which code for a thioredoxin and a glutathione S-transferase, respectively, are multifunctional proteins displaying antioxidative and antiapoptotic properties (Arner and Holmgren 2000; Burmeister et al. 2008; Piaggi et al. 2010). *Microsomal glutathione S-transferase 3* (*mgst3*) encodes an enzyme displaying glutathione-dependent peroxidase activity toward lipid hydroperoxides (Jakobsson et al. 1997). This function may be essential to preserve icefish mitochondrial membranes which, because of their increased phospholipidic component, are particularly susceptible to lipid peroxidation (Mueller et al. 2011). *Mgst3* was also found overexpressed in carp in response to cold stress (Gracey et al. 2004).

Other noteworthy duplicated genes are: 1) *bolA homolog 1* (*Escherichia coli*) (*bola1*), which is well studied in *E. coli* and is involved in the response and adaptation to various stress conditions, among which are pH variation and oxidative stress (Adnan et al. 2011); 2) *cystatin A* (*csta*), encoding a cysteine proteinase that restrains UVB-induced apoptosis of keratinocytes and confers enhanced resistance to high salt concentrations, drought, oxidative, and cold stresses in plants (Takahashi et al. 2007; Zhang et al. 2008); 3) *nuclear protein 1* (*nupr1*), encoding a key player in cellular stress response (Goruppi and Iovanna 2010), which is upregulated in the liver of rainbow trout during stress response and recovery (Momoda et al. 2007); 4) *ring finger protein 7* (*rf7*), encoding a redox-inducible and apoptosis-protective antioxidant protein, which also decreases endogenous ROS production and oxidative DNA damage after exposure to heat shock when overexpressed (Lee et al. 2008; Sun 2008); 5) *defender against cell death 1* (*dad1*), which codes for a subunit of the oligosaccharyltransferase complex and is upregulated in zebrafish embryonic cell line by XBP-1S, a key transcriptional regulator of the unfolded protein response that allows the recovery of endoplasmic reticulum (ER) function (Hu et al. 2007); and 6) *chaperone transport regulator-like 1* (*chac1*), which encodes a

component of the mammalian unfolded protein response pathway (Mungrue et al. 2009). Duplicated genes were found for four additional proteins involved in protein folding: Prefoldin subunit 2 (Pfdn2), which is a subunit of the molecular chaperone prefoldin involved in the folding of newly synthesized proteins, mainly actins and tubulins (Pfdn2 subunit is thought to facilitate the formation of the prefoldin complex and also contain a DNA binding motif) (Martin-Benito et al. 2002; Miyazawa et al. 2011); Thioredoxin domain containing 9 (Txn2c9), which, curiously, has an action antagonistic to that of prefoldin and the combined action of the two proteins is thought to be crucial to establishing a functional cytoskeleton (Stirling et al. 2006); Ubiquitin fusion degradation 1-like (Ufd1l), which is part of a multiprotein complex involved in the recognition and export of polyubiquitin-tagged misfolded proteins from the ER to the cytoplasm, where they are degraded by the 26S proteasome (Bays and Hampton 2002); Ufd1l subunit is the one that directly interacts and stimulates the activity of ubiquitin ligase gp78, which mediates the degradation of misfolded ER proteins (Cao et al. 2007) and is also induced in yeast in response to cold (Schade et al. 2004); *ubiquitin-like 5 (ubl5)*, which is upregulated in response to mitochondrial stress, is required for mounting mitochondrial unfolded protein response, leading to the induction and subsequent import into mitochondria of mitochondrial chaperones (Haynes and Ron 2010).

Duplicated genes were found for the heat shock factor binding protein 1 (Hsbp1), a negative regulator of the heat shock response (HSR) (Satyal et al. 1998). The HSR typically consists in the upregulation of heat shock proteins (HSPs), after induction by the heat shock factors in response to elevated temperatures or other stress. Hsbp1 is a nuclear-localized protein that converts the active trimeric state of Heat shock factor 1 (Hsf1) to the inert monomeric state. Hsbp1 prevents both Hsf1 DNA-binding and transactivation activity and, consequently, blocks HSR activation (Satyal et al. 1998). In Antarctic notothenioids, both Hsp70 mRNA and Hsf1 with DNA-binding activity were found, but no increased expression of Hsp70, or of any other HSP size class, was detected after stress exposure (Buckley et al. 2004; Buckley and Somero 2009). Hsbp1 might have a role in preventing HSPs induction in Antarctic notothenioids. Interestingly, another putative duplicated gene is *BCL2-associated athanogene 2 (bag2)*, encoding a regulator of Hsp70 family chaperones through the inhibition of their chaperone activity by promoting substrate release (Takayama et al. 1999). Bag2 specifically inhibits ubiquitinylation of misfolded proteins by CHIP, an Hsp70-associated ubiquitin ligase, thus providing a regulatory mechanism for preventing uncontrolled degradation of chaperone substrates (Dai et al. 2005).

Finally, in addition to the six MRPs already discussed, 14 cytoplasmic ribosomal proteins were identified among duplicated genes, and, consistently, "structural constituent of ribosome" and "ribosome" were the most represented MF term

and KEGG pathway, respectively, found by the enrichment analysis of duplicates in DAVID. Duplication of ribosomal protein genes might reflect the need to enhance protein synthesis capacity at low temperatures. A similar response, the upregulation of ribosomal proteins, was observed in *Saccharomyces cerevisiae* after cold acclimation (Tai et al. 2007) and in the head kidney of a red-blooded notothenioid, *Dis. mawsoni*, compared with temperate and tropical teleosts (Chen et al. 2008).

Our results identified duplicated genes specific to the icefish lineage and showed that they are mainly involved in mitochondrial biogenesis and aerobic respiration. Thus, our findings confirm the importance of genomic expansions in the evolution of Antarctic notothenioids (Chen et al. 2008; Detrich and Amemiya 2010) and provide new indications about the functional role of gene duplicates. Energy production might be exceptionally challenging in extreme cold conditions and icefish, in the absence of oxygen-carrying proteins, rely on high mitochondrial densities to achieve high concentrations of aerobic enzymes, enhanced lipid-mediated oxygen diffusion, and short oxygen route from capillaries to mitochondria, thus ensuring energy supply and oxygen diffusion to aerobic tissues (O'Brien and Mueller 2010). Therefore, the preferential maintenance of duplicated genes involved in mitochondrial biogenesis and aerobic respiration, which implies a strong selective pressure for increased mitochondrial function acting for millions of years during the evolutionary history of Antarctic notothenioids, seems to underlie icefish adaptation to the cold. Gene duplications specific to the icefish lineage might have provided the raw genetic material for the natural selection of cold-adapted phenotypes able to survive at subzero temperatures. The acquisition of new functional properties or fitness benefits due to an increased synthesis of the ancestral gene product might explain the preservation of mitochondrial-related duplicates, providing an updated framework for the experimental investigation of the genomic and physiological bases of cold adaptation.

Supplementary Material

Supplementary methods, figure S1, and tables S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors are grateful to Vittorio Varotto (Department of Biology, University of Padova, Padova, Italy) for collecting *C. hamatus* samples. This work was supported by the Italian National Program for Antarctic Research (PNRA) to T.P., L.B., and L.Z. and by the University of Padova to S.B. A.C. and I.A.M.M. are recipients of post-doc grants from the University of Padova (grant numbers GRIC11Z79P and CPDR084151/08). C.A. is a PhD student in Evolutionary

Biology at the University of Padova, with a program partially supported under NSF grant 0741348.

Literature Cited

- Adnan M, Morton G, Hadi S. 2011. Analysis of *rpoS* and *bolA* gene expression under various stress-induced environments in planktonic and biofilm phase using $2\text{-}^{\text{D}}\text{DCT}$ method. *Mol Cell Biochem.* 357:275–282.
- Altenhoff AM, Dessimoz C. 2012. Inferring orthology and paralogy. *Methods Mol Biol.* 855:259–279.
- Arner ESJ, Holmgren A. 2000. Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem.* 267:6102–6109.
- Arterbery AS, et al. 2011. Evolution of ligand specificity in vertebrate corticosteroid receptors. *BMC Evol Biol.* 11:14.
- Bargelloni L, et al. 1999. Metallothioneins in Antarctic fish: evidence for independent duplication and gene conversion. *Mol Biol Evol.* 16: 885–897.
- Bays NW, Hampton RY. 2002. *Cdc48-Ufd1-Npl4*: stuck in the middle with Ub. *Curr Biol.* 12:R366–R371.
- Belogradov G. 2008. The proximal N-terminal amino acid residues are required for the coupling activity of the bovine heart mitochondrial factor B. *Arch Biochem Biophys.* 473:76–87.
- Belogradov G. 2009. Recent advances in structure-functional studies of mitochondrial factor B. *J Bioenerg Biomembr.* 41:137–143.
- Belogradov G. 2010. Coupling factor B affects the morphology of mitochondria. *J Bioenerg Biomembr.* 42:29–35.
- Belogradov G, Hatefi Y. 2002. Factor B and the mitochondrial ATP synthase complex. *J Biol Chem.* 277:6097–6103.
- Borley KA, Sidell BD. 2010. Evolution of the myoglobin gene in Antarctic icefishes (Channichthyidae). *Polar Biol.* 34:659–665.
- Buckley BA, Place SP, Hofmann GE. 2004. Regulation of heat shock genes in isolated hepatocytes from an Antarctic fish, *Trematomus bernacchii*. *J Exp Biol.* 207:3649–3656.
- Buckley BA, Somero GN. 2009. cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish *Trematomus bernacchii* to alter gene expression in response to heat stress. *Polar Biol.* 32:403–415.
- Burmeister C, et al. 2008. Oxidative stress in *Caenorhabditis elegans*: protective effects of the omega class glutathione transferase (GSTO-1). *FASEB J.* 22:343–354.
- Cao J, et al. 2007. *Ufd1* is a cofactor of *gp78* and plays a key role in cholesterol metabolism by regulating the stability of HMG-CoA reductase. *Cell Metab.* 6:115–128.
- Carginale V, Trinchella F, Capasso C, Scudiero R, Parisi E. 2004. Gene amplification and cold adaptation of pepsin in Antarctic fish: A possible strategy for food digestion at low temperature. *Gene* 336: 195–205.
- Chen Z, et al. 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc Natl Acad Sci U S A.* 105: 12944–12949.
- Cheng C-HC. 1996. Genomic basis for antifreeze glycopeptide heterogeneity and abundance in Antarctic notothenioid fishes. In: Ennion SJ, Goldspink G, editors. *Gene expression and manipulation in aquatic organisms.* Cambridge: Cambridge University Press. p. 1–20.
- 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.
- Cheng C-HC, Detrich HW III. 2007. Molecular ecophysiology of Antarctic notothenioid fishes. *Philos Trans R Soc Lond B Biol Sci.* 362: 2215–2232.
- Chevreaux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. German Conference on Bioinformatics, Hanover, Germany. p. 45–56.
- Crockett EL, Sidell BD. 1990. Some pathways of energy metabolism are cold adapted in Antarctic fishes. *Physiol Zool.* 63:472–488.
- Dai Q, et al. 2005. Regulation of the cytoplasmic quality control protein degradation pathway by BAG2. *J Biol Chem.* 280:38673–38681.
- Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3:1700–1708.
- Detrich HW III, Amemiya CT. 2010. Antarctic notothenioid fishes: genomic resources and strategies for analyzing an adaptive radiation. *Integr Comp Biol.* 50:1009–1017.
- Dorn GW 2nd. 2010. Mitochondrial pruning by Nix and BNip3: an essential function for cardiac-expressed death factors. *J Cardiovasc Transl Res.* 3:374–383.
- Enoksson M, et al. 2005. Overexpression of glutaredoxin 2 attenuates apoptosis by preventing cytochrome c release. *Biochem Biophys Res Commun.* 327:774–779.
- Fields PA, Somero GN. 1998. Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc Natl Acad Sci U S A.* 95:11476–11481.
- Force A, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Goruppi S, Iovanna JL. 2010. Stress-inducible protein p8 is involved in several physiological and pathological processes. *J Biol Chem.* 285: 1577–1581.
- Gracey AY, et al. 2004. Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proc Natl Acad Sci U S A.* 101:16970–16975.
- Gruschke S, Ott M. 2010. The polypeptide tunnel exit of the mitochondrial ribosome is tailored to meet the specific requirements of the organelle. *Bioessays* 32:1050–1057.
- Guo B, Gan X, He S. 2010. Hox genes of the Japanese eel *Anguilla japonica* and Hox cluster evolution in teleosts. *J Exp Zool B Mol Dev Evol.* 314:135–147.
- Gustafsson AB. 2011. *Bnip3* as a dual regulator of mitochondrial turnover and cell death in the myocardium. *Pediatr Cardiol.* 32: 267–274.
- Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. 2009. Adaptive evolution of young gene duplicates in mammals. *Genome Res.* 19: 859–867.
- Haque ME, Spemulli LL, Fecko CJ. 2010. Identification of protein-protein and protein-ribosome interacting regions of the C-terminal tail of human mitochondrial inner membrane protein Oxa1L. *J Biol Chem.* 285:34991–34998.
- Haynes CM, Ron D. 2010. The mitochondrial UPR—protecting organelle protein homeostasis. *J Cell Sci.* 123:3849–3855.
- Hell K, Herrmann JM, Prati E, Neupert W, Stuart RA. 1998. Oxa1p, an essential component of the N-tail protein export machinery in mitochondria. *Proc Natl Acad Sci U S A.* 95:2250–2255.
- Hiltunen JK, Chen Z, Haapalainen AM, Wierenga RK, Kastaniotis AJ. 2010. Mitochondrial fatty acid synthesis—an adopted set of enzymes making a pathway of major importance for the cellular metabolism. *Prog Lipid Res.* 49:27–45.
- Hornig YC, Cobine PA, Maxfield AB, Carr HS, Winge DR. 2004. Specific copper transfer from the Cox17 metallochaperone to both Sco1 and Cox11 in the assembly of yeast cytochrome C oxidase. *J Biol Chem.* 279:35334–35340.
- Hu MC, et al. 2007. XBP-1, a key regulator of unfolded protein response, activates transcription of IGF1 and Akt phosphorylation in zebrafish embryonic cell line. *Biochem Biophys Res Commun.* 359:778–783.
- Huigsloot M, et al. 2011. A mutation in *C2orf64* causes impaired cytochrome c oxidase assembly and mitochondrial cardiomyopathy. *Am J Hum Genet.* 88:488–493.
- Jakobsson PJ, Mancini JA, Riendeau D, Ford-Hutchinson AW. 1997. Identification and characterization of a novel microsomal enzyme with glutathione-dependent transferase and peroxidase activities. *J Biol Chem.* 272:22934–22939.

- Johnston IA, Calvo J, Guderley H, Fernandez D, Palmer L. 1998. Latitudinal variation in the abundance and oxidative capacities of muscle mitochondria in perciform fishes. *J Exp Biol.* 201:1–12.
- Johnston IA, Camm JP, White M. 1988. Specializations of swimming muscles in the pelagic Antarctic fish *Pleuragramma antarcticum*. *Mar Biol.* 100:3–12.
- Kawall H, Torres J, Sidell B, Somero G. 2002. Metabolic cold adaptation in Antarctic fishes: evidence from enzymatic activities of brain. *Mar Biol.* 140:279–286.
- Kock K-H. 2005. Antarctic icefishes (Channichthyidae): a unique family of fishes: A review, Part I. *Polar Biol.* 28:862–895.
- Kubli DA, Quinsay MN, Huang C, Lee Y, Gustafsson AB. 2008. Bnip3 functions as a mitochondrial sensor of oxidative stress during myocardial ischemia and reperfusion. *Am J Physiol Heart Circ Physiol.* 295: H2025–H2031.
- Lee SJ, et al. 2008. Regulation of heat shock-induced apoptosis by sensitive to apoptosis gene protein. *Free Radic Biol Med.* 45:167–176.
- Lillig CH, et al. 2005. Characterization of human glutaredoxin 2 as iron-sulfur protein: a possible role as redox sensor. *Proc Natl Acad Sci U S A.* 102:8168–8173.
- Lucassen M, Koschnick N, Eckerle LG, Portner HO. 2006. Mitochondrial mechanisms of cold adaptation in cod (*Gadus morhua* L.) populations from different climatic zones. *J Exp Biol.* 209:2462–2471.
- Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.
- Mark FC, et al. 2012. Mitochondrial function in Antarctic nototheniids with ND6 translocation. *PLoS One* 7:e31860.
- Martin-Benito J, et al. 2002. Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO J.* 21:6377–6386.
- Matschiner M, Hanel R, Salzburger W. 2011. On the origin and trigger of the notothenioid adaptive radiation. *PLoS One* 6:e18911.
- Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* 27:937–945.
- Miyazawa M, et al. 2011. Prefoldin subunits are protected from ubiquitin-proteasome-mediated degradation by forming complex with other constituent subunits. *J Biol Chem.* 286:19191–19203.
- Momoda TS, et al. 2007. Gene expression in the liver of rainbow trout, *Oncorhynchus mykiss*, during the stress response. *Comp Biochem Physiol Part D Genomics Proteomics.* 2:303–315.
- Moylan TJ, Sidell BD. 2000. Concentrations of myoglobin and myoglobin mRNA in heart ventricles from Antarctic fishes. *J Exp Biol.* 203: 1277–1286.
- Mueller IA, Grim JM, Beers JM, Crockett EL, O'Brien KM. 2011. Inter-relationship between mitochondrial function and susceptibility to oxidative stress in red- and white-blooded Antarctic notothenioid fishes. *J Exp Biol.* 214:3732–3741.
- Mungrue IN, Pagnon J, Kohannim O, Gargalovic PS, Lusic AJ. 2009. CHAC1/MGC4504 is a novel proapoptotic component of the unfolded protein response, downstream of the ATF4-ATF3-CHOP cascade. *J Immunol.* 182:466–476.
- Murdock DG, Boone BE, Esposito LA, Wallace DC. 1999. Up-regulation of nuclear and mitochondrial genes in the skeletal muscle of mice lacking the heart muscle isoform of the adenine nucleotide translocator. *J Biol Chem.* 274:14429–14433.
- Near TJ, Parker SK, Detrich HW III. 2006. A genomic fossil reveals key steps in hemoglobin loss by the antarctic icefishes. *Mol Biol Evol.* 23: 2008–2016.
- Near TJ, et al. 2012. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proc Natl Acad Sci U S A.* 109:3434–3439.
- Nomiyama H, et al. 2008. Extensive expansion and diversification of the chemokine gene family in zebrafish: identification of a novel chemokine subfamily CX. *BMC Genomics* 9:222.
- O'Brien KM. 2011. Mitochondrial biogenesis in cold-bodied fishes. *J Exp Biol.* 214:275–285.
- O'Brien KM, Mueller IA. 2010. The unique mitochondrial form and function of Antarctic channichthyid icefishes. *Integr Comp Biol.* 50: 993–1008.
- Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag.
- Ohno S. 1972. So much "junk" DNA in our genome. *Brookhaven Symp Biol.* 23:366–370.
- Ohta T. 1989. Role of gene duplication in evolution. *Genome* 31:304–310.
- Orczewska JJ, Hartleben G, O'Brien KM. 2010. The molecular basis of aerobic metabolic remodeling differs between oxidative muscle and liver of threespine sticklebacks in response to cold acclimation. *Am J Physiol Regul Integr Comp Physiol.* 299:R352–R364.
- Ott M, Herrmann JM. 2010. Co-translational membrane insertion of mitochondrially encoded proteins. *Biochim Biophys Acta.* 1803:767–775.
- Pagliarini DJ, et al. 2008. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134:112–123.
- Park CB, et al. 2007. MTERF3 is a negative regulator of mammalian mtDNA transcription. *Cell* 130:273–285.
- Patarnello T, Verde C, di Prisco G, Bargelloni L, Zane L. 2011. How will fish that evolved at constant sub-zero temperatures cope with global warming? Notothenioids as a case study. *Bioessays* 33:260–268.
- Piaggi S, et al. 2010. Glutathione transferase omega 1-1 (GSTO1-1) plays an anti-apoptotic role in cell resistance to cisplatin toxicity. *Carcinogenesis* 31:804–811.
- Regula KM. 2002. Inducible expression of BNIP3 provokes mitochondrial defects and hypoxia-mediated cell death of ventricular myocytes. *Circ Res.* 91:226–231.
- Rennison DJ, Owens GL, Taylor JS. 2012. Opsin gene duplication and divergence in ray-finned fish. *Mol Phylogenet Evol.* 62:986–1008.
- Rigby K, Zhang L, Cobine PA, George GN, Winge DR. 2007. Characterization of the cytochrome c oxidase assembly factor Cox19 of *Saccharomyces cerevisiae*. *J Biol Chem.* 282:10233–10242.
- Ruud JT. 1954. Vertebrates without erythrocytes and blood pigment. *Nature* 173:848–850.
- Satyal SH, Chen D, Fox SG, Kramer JM, Morimoto RI. 1998. Negative regulation of the heat shock transcriptional response by HSBP1. *Genes Dev.* 12:1962–1974.
- Schade B, Jansen G, Whiteway M, Entian KD, Thomas DY. 2004. Cold adaptation in budding yeast. *Mol Biol Cell.* 15:5492–5502.
- Schmidt O, Pfanner N, Meisinger C. 2010. Mitochondrial protein import: from proteomics to functional mechanisms. *Nat Rev Mol Cell Biol.* 11: 655–667.
- Sidell BD. 1998. Intracellular oxygen diffusion: the roles of myoglobin and lipid at cold body temperature. *J Exp Biol.* 201:1119–1128.
- Sidell BD, et al. 1997. Variable expression of myoglobin among the hemoglobinless Antarctic icefishes. *Proc Natl Acad Sci U S A.* 94:3420–3424.
- Smits P, Smeitink JA, van den Heuvel LP, Huynen MA, Ettema TJ. 2007. Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res.* 35:4686–4703.
- Somero GN. 2004. Adaptation of enzymes to temperature: searching for basic "strategies." *Comp Biochem Physiol B Biochem Mol Biol.* 139: 321–333.
- Stirling PC, et al. 2006. PhLP3 modulates CCT-mediated actin and tubulin folding via ternary complexes with substrates. *J Biol Chem.* 281: 7012–7021.
- Sun Y. 2008. RNF7 (RING finger protein-7). *Atlas Genet Cytogenet Oncol Haematol.* 12:289–291.
- Tai SL, Daran-Lapujade P, Walsh MC, Pronk JT, Daran JM. 2007. Acclimation of *Saccharomyces cerevisiae* to low temperature: a chemostat-based transcriptome analysis. *Mol Biol Cell.* 18: 5100–5112.
- Takahashi H, et al. 2007. Cystatin A suppresses ultraviolet B-induced apoptosis of keratinocytes. *J Dermatol Sci.* 46:179–187.

- Takayama S, Xie ZH, Reed JC. 1999. An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. *J Biol Chem.* 274: 781–786.
- Toogood HS, Leys D, Scrutton NS. 2007. Dynamics driving function: new insights from electron transferring flavoproteins and partner complexes. *FEBS J.* 274:5481–5504.
- Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet.* 10:725–732.
- Van der Aa LM, et al. 2009. A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. *BMC Biol.* 7:7.
- Vilella AJ, et al. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19:327–335.
- Xu Q, et al. 2008. Adaptive evolution of hepcidin genes in Antarctic notothenioid fishes. *Mol Biol Evol.* 25:1099–1112.
- Zhang XX, Liu SK, Takano T. 2008. Two cysteine proteinase inhibitors from *Arabidopsis thaliana*, AtCYSa and AtCYSb, increasing the salt, drought, oxidation and cold tolerance. *Plant Mol Biol.* 68: 131–143.
- Zhao J, et al. 2009. The novel conserved mitochondrial inner-membrane protein MTGM regulates mitochondrial morphology and cell proliferation. *J Cell Sci.* 122:2252–2262.
- Zou S, Kamei H, Modi Z, Duan C. 2009. Zebrafish IGF genes: gene duplication, conservation and divergence, and novel roles in midline and notochord development. *PLoS One* 4:e7026.

Associate editor: Laura Landweber

Supplementary Materials

1. Sampling, cDNA library construction and sequencing

Total RNA was extracted from 30 mg of muscle tissue of four *Chionodraco hamatus* individuals using RNeasy mini-column kit (QIAGEN). After checking the integrity, purity and size distribution of total RNA, RNA samples were pooled and stored in three volumes of 96% ethanol and 0.1 volume of sodium acetate to obtain 5 µg of pooled RNA in a final volume of 120 µl. Pooled RNA was sent to Evrogen (Moscow, Russia; www.evrogen.com) where double-stranded cDNA was synthesized using a SMART (Switching Mechanism At 5' end of RNA Template) approach (Zhu et al. 2001), which was modified to introduce a GsuI restriction site after the mRNA poly-A tail. After normalization, using DSN method (Zhulidov et al. 2004), poly-A was trimmed using GsuI (Fermentas) and cDNA purification was performed with Agencourt AMPure XP (BECKMAN COULTER). Approximately 15 µg of normalized cDNA were used for library construction at BMR Genomics, University of Padua, Italy (<http://www.bmr-genomics.it>). Sequencing was performed using GS FLX Titanium series reagents in half a plate of a Genome Sequencer FLX instrument.

2. *C. hamatus* skeletal muscle transcriptome 1.0

The *C. hamatus* transcriptome consisted of 23,968 good quality contigs reconstructed from a large fraction (93.3%) of the original read sequence information (341,171 reads). BLAST search against protein and nucleotide databases allowed the *de novo* annotation of 48% of the transcriptome (11,509 contigs), an appreciable result if compared with the fraction of annotated contigs obtained in other sequencing projects of fish normalized transcriptomes: 36% for *Anguilla anguilla* (Coppe et al. 2010), 40% for *Poecilia reticulata* (Fraser et al. 2011), 26% for *Xiphophorus maculatus* (Zhang et al. 2011). A slim-based classification of 5,805 unique GO terms associated to 7,293 contigs showed that

most represented BP terms (“metabolism”, “development” and “cell organization and biogenesis”) and MF terms (“catalytic activity” and “binding”) likely reflect the typical physiology of skeletal muscle, mainly involved in energy production through oxidative metabolism, and the need for an increased remodeling and turnover of structural components to compensate stress associated with extreme cold conditions; a similar response was observed in four non-normalized transcriptomes of the Antarctic notothenioid *Dissostichus mawsoni* (Chen et al. 2008).

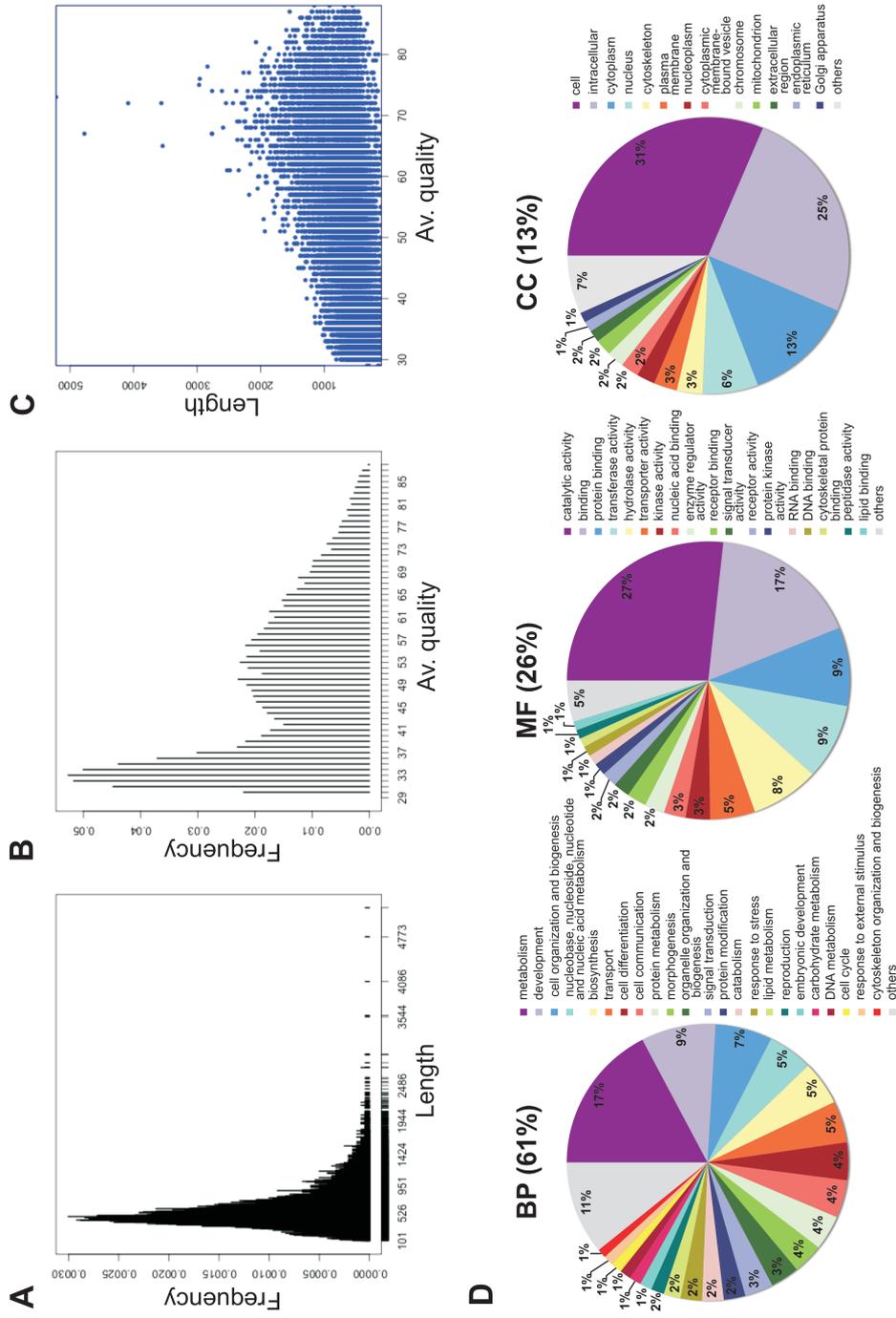


Fig. S1. Transcriptome sequences and annotation. (A) Contig length distribution; (B) contig average quality distribution; (C) relationship between contig length and average quality. (D) Outline of whole transcriptome GO functional annotation; for each ontology [biological process (BP), molecular function (MF) and cellular component (CC)], a pie chart represents frequencies of general term categories, after slim-based classification of 5,805 unique GO terms associated to 23,968 *C. hamatus* transcripts.

Supplementary tables

Table S1. *C. hamatus* muscle transcriptome reconstruction. Reported are: features of the 341,171 reads obtained by trimming 382,386 raw reads using Lucy (pre-processing data); features of the two assembly run performed with MIRA 3 assembler (assembled data). The second run assembled 28,281 contigs plus 46,569 reads discarded by the first run.

Pre-processing data		Assembled data		
			1 st run	2 nd run
Average length	319.15	Starting reads	341,171	28,281 contigs + 46,569 reads = 74,826
Median length	331	Contigs	28,281	25,398
Average quality	29.48	Assembled reads	294,602	318,440
Median quality	29.27	%	86.3	93.3

Table S2. Features of 23,968 contigs representing the *C. hamatus* muscle transcriptome. Distribution statistics for contig length, average quality and coverage (number of reads) of contig sequences.

23,968 contigs (318,440 assembled reads)	Min	1 st Q	Median	Mean	3 rd Q	Max
	Length	101	409	521	595	698
Av-quality	30	35	46	48	58	88
Coverage	2	3	5	13	11	497

Table S3. *C. hamatus* 124 putative duplicated genes. Reported are: gene symbol (provided by Ensembl BioMart using as Filter the Ensembl gene ID for *D. rerio* or that for *G. aculeatus* when *D. rerio* gene was missing in the subtree); gene description (provided by Ensembl BioMart using as Filter the Ensembl gene ID for *D. rerio* or that for *G. aculeatus* when *D. rerio* gene was missing in the subtree); Ensembl gene IDs of the five reference species included in the subtree; Ensembl gene tree ID of which the subtree is a fraction; number of *C. hamatus* putative duplicated gene copies; IDs of contigs mapping to each subtree divided by square bracket in groups of alternative transcripts of different loci.

Gene symbol	Gene description	Ensembl gene IDs of five reference species (subtree)	Ensembl gene tree ID	Number of <i>C. hamatus</i> duplicated gene copies	Contig IDs
<i>acn9</i>	<i>ACN9</i> homolog (<i>S. cerevisiae</i>) [Source:ZFIN; Acc:ZDB-GENE-041010-94]	ENSDARG00000036092 ENSGACG00000000092 ENSORLG00000010233 ENSTNIG00000018165 ENSTRUG00000005985	ENSGT003 9000001002 9	2	[chionodraco_hamatus2_c488] [chionodraco_hamatus_c8252]
<i>anapc11</i>	<i>APC11</i> anaphase promoting complex subunit 11 homolog (yeast) [Source:ZFIN; Acc:ZDB-GENE-061013-383]	ENSDARG00000070333 ENSGACG00000018183 ENSORLG00000001673 ENSTNIG00000014139 ENSTRUG00000004285	ENSGT005 5000007518 6	2	[chionodraco_hamatus2_c1104] [chionodraco_hamatus_c6662]
<i>anapc13</i>	<i>anaphase promoting complex subunit 13</i> [Source:HGNC Symbol; Acc:24540]	ENSDARG00000078435 ENSGACG00000013894 ENSORLG00000015577 ENSTRUG00000008901	ENSGT003 9000000867 3	2	[chionodraco_hamatus2_c2310] [chionodraco_hamatus_c8395]
<i>atp5s</i>	<i>ATP synthase, H⁺ transporting, mitochondrial F₀ complex, subunit s</i> [Source:ZFIN; Acc:ZDB-GENE-040426-959]	ENSDARG00000012072 ENSGACG00000011836 ENSORLG00000009104 ENSTNIG00000019337 ENSTRUG00000009762	ENSGT005 3000006368 0	2	[chionodraco_hamatus_c7200] [chionodraco_hamatus_c4284]
<i>bag2</i>	<i>BCL2-associated athanogene 2</i> [Source:ZFIN; Acc:ZDB-GENE-040426-1399]	ENSDARG00000035005 ENSGACG00000006169 ENSORLG00000010100 ENSTNIG00000016574 ENSTRUG00000015680	ENSGT003 9000001759 0	2	[chionodraco_hamatus2_rep_c3235] [chionodraco_hamatus2_c1383]
<i>bnip3</i>	<i>BCL2/adenovirus E1B interacting protein 3</i> [Source:ZFIN; Acc:ZDB-GENE-030131-8060]	ENSDARG00000045033 ENSGACG00000009527 ENSORLG00000005914 ENSTNIG00000013333 ENSTRUG00000006491	ENSGT003 9000001341 5	2	[chionodraco_hamatus2_c997] [chionodraco_hamatus_rep_c27644]
<i>bolal</i>	<i>bolA</i> homolog 1 (<i>E. coli</i>) [Source:HGNC Symbol; Acc:24263]	ENSDARG00000039915 ENSGACG00000013188 ENSORLG00000016299 ENSTNIG00000010298 ENSTRUG00000004188	ENSGT005 1000004816 5	2	[chionodraco_hamatus2_c588] [chionodraco_hamatus2_rep_c3112]
<i>bx470131.2</i>	<i>hypothetical protein LOC568697</i> [Source:RefSeq peptide; Acc:NP_001185793]	ENSDARG00000089019 ENSGACG00000017715 ENSORLG00000003159 ENSTNIG00000005911	ENSGT003 9000001513 8	2	[chionodraco_hamatus2_c106] [chionodraco_hamatus_s26060]

C10HXorf 26	<i>chromosome X open reading frame 26</i> [Source:HGNC Symbol; Acc:28790]	ENSDARG00000057299 ENSGACG00000001888 ENSORLG00000019795 ENSTNIG00000005669 ENSTRUG00000018640	ENSGT003 9000001633 3	3	[chionodraco_hamatus2_rep_c3207] [chionodraco_hamatus_c6822] [chionodraco_hamatus_rep_c27435, chionodraco_hamatus_c5563, chionodraco_hamatus_rep_c11179, chionodraco_hamatus2_c42]
C11H1orf 50	<i>chromosome 1 open reading frame 50</i> [Source:HGNC Symbol; Acc:28795]	ENSDARG00000071211 ENSGACG00000011063 ENSORLG00000012573 ENSTNIG00000007386 ENSTRUG00000001822	ENSGT003 9000000308 4	2	[chionodraco_hamatus_c843] [chionodraco_hamatus_c5716, chionodraco_hamatus_c8833]
C13H6orf 57	<i>chromosome 6 open reading frame 57</i> [Source:HGNC Symbol; Acc:20957]	ENSDARG00000039390 ENSGACG00000003861 ENSORLG00000000941 ENSTNIG00000019524 ENSTRUG00000000428	ENSGT003 9000000915 5	2	[chionodraco_hamatus_c4228] [chionodraco_hamatus2_c418]
C19H1orf 109	<i>chromosome 1 open reading frame 109</i> [Source:HGNC Symbol; Acc:26039]	ENSDARG00000012113 ENSGACG00000002689 ENSORLG00000009883 ENSTRUG000000007899	ENSGT003 9000000855 1	2	[chionodraco_hamatus_c16027] [chionodraco_hamatus_c22206]
C1H4orf4 3	<i>chromosome 4 open reading frame 43</i> [Source:HGNC Symbol; Acc:25638]	ENSDARG00000008068 ENSGACG00000017134 ENSORLG00000007775 ENSTNIG00000018663 ENSTRUG00000015324	ENSGT003 9000000417 9	2	[chionodraco_hamatus_rep_c21634] [chionodraco_hamatus2_c406, chionodraco_hamatus_c8471]
C20H15orf f29	<i>chromosome 15 open reading frame 29</i> [Source:HGNC Symbol; Acc:26199]	ENSDARG00000042522 ENSGACG00000007898 ENSORLG00000017841 ENSTNIG00000012211 ENSTRUG00000018118	ENSGT003 9000001235 1	2	[chionodraco_hamatus_c13069] [chionodraco_hamatus_c23162]
C20H6orf 162	<i>chromosome 6 open reading frame 162</i> [Source:HGNC Symbol; Acc:21401]	ENSDARG00000071040 ENSGACG00000012249 ENSORLG00000017811 ENSTNIG00000010282 ENSTRUG00000012775	ENSGT003 9000001167 4	3	[chionodraco_hamatus2_rep_c3273] [chionodraco_hamatus_c25478] [chionodraco_hamatus_c2994]
C2H1orf1 23	<i>chromosome 1 open reading frame 123</i> [Source:HGNC Symbol; Acc:26059]	ENSDARG00000063341 ENSORLG00000010467 ENSTNIG00000006952 ENSTRUG00000012792	ENSGT003 9000000152 3	2	[chionodraco_hamatus2_c403] [chionodraco_hamatus_c1202, chionodraco_hamatus_c3737]
C4H1orf3 1	<i>chromosome 1 open reading frame 31</i> [Source:HGNC Symbol; Acc:18025]	ENSDARG00000092677 ENSGACG00000019202 ENSORLG00000014590 ENSTRUG00000012436	ENSGT003 9000000409 4	2	[chionodraco_hamatus_c206, chionodraco_hamatus_c6946] [chionodraco_hamatus2_c977]
C7H17orf 61	<i>chromosome 17 open reading frame 61</i> [Source:HGNC Symbol; Acc:28618]	ENSDARG00000038901 ENSGACG00000019937 ENSORLG00000003580 ENSTNIG00000005943 ENSTRUG00000003389	ENSGT003 9000000033 4	3	[chionodraco_hamatus2_c984] [chionodraco_hamatus_rep_c22215] [chionodraco_hamatus_rep_c20780]
<i>casp3a</i>	<i>caspase 3, apoptosis- related cysteine protease a</i> [Source:ZFIN; Acc:ZDB-GENE- 011210-1]	ENSDARG00000017905 ENSGACG00000016960 ENSORLG00000008049 ENSTNIG00000007133 ENSTRUG00000015915	ENSGT005 6000007678 2	2	[chionodraco_hamatus_c23139] [chionodraco_hamatus_c13314]
<i>ccdc55</i>	<i>coiled-coil domain containing 55</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-2646]	ENSDARG00000015221 ENSGACG00000020516 ENSORLG00000003135 ENSTNIG00000016161 ENSTRUG00000000226	ENSGT006 4000009154 6	2	[chionodraco_hamatus_c961] [chionodraco_hamatus_c25503]
<i>cenpv</i>	<i>centromere protein V</i> [Source:ZFIN; Acc:ZDB-GENE- 060526-76]	ENSDARG00000092285 ENSGACG00000020259 ENSORLG00000012007 ENSTRUG00000000676	ENSGT003 9000000318 3	3	[chionodraco_hamatus2_c882] [chionodraco_hamatus_c8255] [chionodraco_hamatus2_rep_c3101, chionodraco_hamatus_c5460]

<i>chac1</i>	<i>ChaC, cation transport regulator-like 1</i> [Source:ZFIN; Acc:ZDB-GENE-030131-1957]	ENSDARG00000070426 ENSGACG00000012938 ENSORLG00000018024 ENSTNIG00000010968 ENSTRUG00000013003	ENSGT003 9000000385 5	2	[chionodraco_hamatus_c4723] [chionodraco_hamatus2_c1608]
<i>coa5</i>	<i>cytochrome C oxidase assembly factor 5</i> [Source:HGNC Symbol; Acc:33848]	ENSDARG00000094833 ENSTNIG0000001478 ENSTRUG00000014082	ENSGT003 9000000554 8	3	[chionodraco_hamatus_rep_c17901] [chionodraco_hamatus2_c1179] [chionodraco_hamatus_rep_c25493, chionodraco_hamatus_rep_c22983, chionodraco_hamatus_rep_c15830, chionodraco_hamatus2_rep_c2993]
<i>cope</i>	<i>coatomer protein complex, subunit epsilon</i> [Source:HGNC Symbol; Acc:2234]	ENSDARG00000018562 ENSGACG00000010568 ENSORLG00000017456 ENSTNIG00000003143 ENSTNIG00000015512 ENSTRUG00000011693	ENSGT003 9000000347 8	2	[chionodraco_hamatus_rep_c27202] [chionodraco_hamatus2_c1318]
<i>cox17</i>	<i>COX17 cytochrome c oxidase assembly homolog (S. cerevisiae)</i> [Source:ZFIN; Acc:ZDB-GENE-040912-91]	ENSDARG00000069920 ENSGACG00000001415 ENSORLG00000008633 ENSTNIG00000001444 ENSTRUG00000015927	ENSGT003 9000000232 9	8	[chionodraco_hamatus_rep_c26497] [chionodraco_hamatus_rep_c27756] [chionodraco_hamatus_rep_c17610] [chionodraco_hamatus2_rep_c3189] [chionodraco_hamatus_rep_c27388] [chionodraco_hamatus2_rep_c3405] [chionodraco_hamatus2_c1358] [chionodraco_hamatus2_rep_c3023]
<i>cox19</i>	<i>COX19 cytochrome c oxidase assembly homolog (S. cerevisiae)</i> [Source:HGNC Symbol; Acc:28074]	ENSDARG00000063882 ENSGACG00000005638 ENSTNIG00000007410 ENSTRUG00000017163	ENSGT003 9000001689 5	3	[chionodraco_hamatus_c3226] [chionodraco_hamatus2_rep_c3220] [chionodraco_hamatus_c9634]
<i>cox4i1</i>	<i>cytochrome c oxidase subunit IV isoform 1</i> [Source:ZFIN; Acc:ZDB-GENE-030131-5175]	ENSDARG00000032970 ENSGACG00000015963 ENSORLG00000009488 ENSTNIG00000000609 ENSTRUG00000018241	ENSGT003 9000000240 7	2	[chionodraco_hamatus2_c508] [chionodraco_hamatus_rep_c26718]
<i>cox6a2</i>	<i>cytochrome c oxidase subunit VIa polypeptide 2</i> [Source:HGNC Symbol; Acc:2279]	ENSDARG00000054588 ENSGACG00000013583 ENSORLG00000002089 ENSTNIG00000014922 ENSTRUG00000007689	ENSGT003 9000000275 6	3	[chionodraco_hamatus2_c2593] [chionodraco_hamatus_c9072, chionodraco_hamatus_c7797] [chionodraco_hamatus_c8526]
<i>cox7a2</i>	<i>cytochrome c oxidase, subunit VIIa 2</i> [Source:ZFIN; Acc:ZDB-GENE-050522-153]	ENSDARG00000053217 ENSDARG00000086341 ENSGACG00000011811 ENSORLG00000017514 ENSTNIG00000010705 ENSTRUG00000015302	ENSGT004 0000002205 4	2	[chionodraco_hamatus_rep_c20058] [chionodraco_hamatus2_c1408, chionodraco_hamatus_rep_c17229, chionodraco_hamatus2_c1716, chionodraco_hamatus_rep_c26670]
<i>cr391962.1</i>	Uncharacterized protein [Source:UniProtKB/Tr EMBL; Acc:F1R4S3]	ENSDARG00000040130 ENSGACG00000005839 ENSGACG00000012665 ENSGACG00000012682 ENSTNIG00000003673 ENSTRUG00000004908	ENSGT003 9000001435 2	6	[chionodraco_hamatus2_rep_c3480] [chionodraco_hamatus2_s3482] [chionodraco_hamatus2_rep_c3504] [chionodraco_hamatus2_rep_c3552] [chionodraco_hamatus2_rep_c3527, chionodraco_hamatus_rep_c26351] [chionodraco_hamatus_c27953]
<i>creg1</i>	<i>cellular repressor of E1A-stimulated genes 1</i> [Source:HGNC Symbol; Acc:2351]	ENSDARG00000063345 ENSGACG00000015462 ENSORLG00000001141 ENSTNIG00000006594 ENSTRUG00000003275	ENSGT003 9000000591 4	2	[chionodraco_hamatus_rep_c27970] [chionodraco_hamatus2_c1778]

<i>csta</i>	<i>cystatin A (stefin A)</i> [Source:HGNC Symbol; Acc:2481]	ENSDARG00000028164 ENSDARG00000045352 ENSDARG00000045980 ENSGACG00000003266 ENSORLG00000009242 ENSORLG00000009248 ENSORLG00000009258 ENSORLG00000009262 ENSTNIG00000012728 ENSTNIG00000012729 ENSTRUG00000000582 ENSTRUG00000017651	ENSGT003 9000001560 7	2	[chionodraco_hamatus2_rep_c3327, chionodraco_hamatus2_c263, chionodraco_hamatus2_rep_c3082] [chionodraco_hamatus_rep_c17753]
<i>dad1</i>	<i>defender against cell death 1</i> [Source:ZFIN; Acc:ZDB-GENE- 060503-233]	ENSDARG00000059791 ENSGACG00000014453 ENSORLG00000012492 ENSTNIG00000007956 ENSTRUG00000000531	ENSGT003 9000000332 4	2	[chionodraco_hamatus_c681] [chionodraco_hamatus_rep_c20380]
<i>dctm6 (1 of 2)</i>	<i>dynactin 6</i> [Source:HGNC Symbol; Acc:16964]	ENSGACG00000015891 ENSORLG00000004837 ENSTNIG00000005141 ENSTRUG00000004797	ENSGT003 9000001789 0	2	[chionodraco_hamatus_c2571] [chionodraco_hamatus2_c787]
<i>dpy30</i>	<i>dpy-30 homolog (C. elegans)</i> [Source:ZFIN; Acc:ZDB-GENE- 040718-136]	ENSDARG00000004427 ENSGACG00000019898 ENSORLG00000020166 ENSTRUG00000011982	ENSGT003 9000000880 8	2	[chionodraco_hamatus_c9482] [chionodraco_hamatus2_c733]
<i>eef1da</i>	<i>elongation factor-1, delta, a</i> [Source:ZFIN; Acc:ZDB-GENE- 040426-2740]	ENSDARG00000008109 ENSGACG00000017191 ENSORLG00000017750 ENSTNIG00000018628 ENSTRUG00000013449	ENSGT003 9000001174 7	2	[chionodraco_hamatus2_c232] [chionodraco_hamatus_c8369]
<i>esml</i>	<i>endothelial cell-specific molecule 1</i> [Source:HGNC Symbol; Acc:3466]	ENSDARG00000060189 ENSGACG00000016408 ENSORLG00000006155 ENSTNIG00000007530 ENSTRUG00000012890	ENSGT003 9000001881 0	2	[chionodraco_hamatus_c22665] [chionodraco_hamatus2_c830, chionodraco_hamatus_c4893]
<i>etfa</i>	<i>electron-transfer- flavoprotein, alpha polypeptide</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-4449]	ENSDARG000000043944 ENSGACG00000010550 ENSORLG00000007847 ENSTNIG00000008383 ENSTRUG00000008849	ENSGT003 9000001342 2	2	[chionodraco_hamatus_c8810] [chionodraco_hamatus_c149]
<i>exosc7</i>	<i>exosome component 7</i> [Source:HGNC Symbol; Acc:28112]	ENSDARG00000043336 ENSGACG00000003085 ENSORLG00000002251 ENSTNIG00000003617 ENSTRUG000000008320	ENSGT005 3000006309 3	2	[chionodraco_hamatus_c6057] [chionodraco_hamatus_c8668]
<i>fam162b</i>	<i>family with sequence similarity 162, member B</i> [Source:HGNC Symbol; Acc:21549]	ENSDARG000000063344 ENSGACG00000003283 ENSORLG00000009283 ENSTNIG00000012731 ENSTRUG00000017646	ENSGT006 4000009149 7	2	[chionodraco_hamatus2_c545, chionodraco_hamatus_c8746] [chionodraco_hamatus_c4952, chionodraco_hamatus_c24246]
<i>fdx1</i>	<i>ferredoxin 1</i> [Source:ZFIN; Acc:ZDB-GENE- 071015-2]	ENSDARG00000056410 ENSGACG00000014198 ENSORLG00000004196 ENSTNIG00000010357 ENSTRUG00000011328	ENSGT005 3000006357 7	2	[chionodraco_hamatus_c7479] [chionodraco_hamatus2_c357]
<i>fhla</i>	<i>four and a half LIM domains a</i> [Source:ZFIN; Acc:ZDB-GENE- 040206-1]	ENSDARG000000071498 ENSGACG00000017268 ENSORLG00000001563 ENSTNIG00000016850 ENSTRUG00000003991	ENSGT006 4000009129 7	2	[chionodraco_hamatus_c22365] [chionodraco_hamatus2_c127]

<i>glrx2</i>	<i>glutaredoxin 2</i> [Source:ZFIN; Acc:ZDB-GENE-040718-101]	ENSDARG00000042631 ENSGACG00000005873 ENSORLG00000012223 ENSTNIG00000012335 ENSTRUG00000003061	ENSGT003 9000000367 7	2	[chionodraco_hamatus_c6689] [chionodraco_hamatus_c8221]
<i>gnb2l1</i>	<i>guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1</i> [Source:ZFIN; Acc:ZDB-GENE-990415-89]	ENSDARG00000041619 ENSGACG00000018103 ENSORLG00000005276 ENSTNIG00000010510 ENSTRUG00000000550	ENSGT006 4000009145 9	2	[chionodraco_hamatus2_c1302] [chionodraco_hamatus_c5641]
<i>hsbp1</i>	<i>heat shock factor binding protein 1</i> [Source:ZFIN; Acc:ZDB-GENE-040426-1721]	ENSDARG00000041921 ENSDARG00000069425 ENSGACG00000013784 ENSORLG00000015789 ENSTNIG00000008805 ENSTRUG00000001395	ENSGT003 9000000629 3	2	[chionodraco_hamatus2_c369, chionodraco_hamatus_c4953] [chionodraco_hamatus_rep_c27323]
<i>imp3</i>	<i>IMP3, U3 small nucleolar ribonucleoprotein, homolog (yeast)</i> [Source:ZFIN; Acc:ZDB-GENE-040426-1062]	ENSDARG00000069287 ENSGACG00000015998 ENSORLG00000004484 ENSTRUG00000003210	ENSGT005 5000007509 0	2	[chionodraco_hamatus2_c2596] [chionodraco_hamatus_c8126]
<i>lamtor3</i>	<i>late endosomal/lysosomal adaptor, MAPK and MTOR activator 3</i> [Source:ZFIN; Acc:ZDB-GENE-050522-345]	ENSDARG00000057075 ENSORLG00000018073 ENSTNIG00000006997 ENSTRUG00000004058	ENSGT003 9000001315 9	2	[chionodraco_hamatus_rep_c9939] [chionodraco_hamatus_c8210, chionodraco_hamatus_rep_c9858, chionodraco_hamatus_rep_c9900]
<i>lin52</i>	<i>lin-52 homolog (C. elegans)</i> [Source:ZFIN; Acc:ZDB-GENE-061110-82]	ENSDARG00000070833 ENSGACG00000005856 ENSTNIG00000017427 ENSTRUG00000015313	ENSGT003 9000000840 2	3	[chionodraco_hamatus_c8545] [chionodraco_hamatus2_c180] [chionodraco_hamatus_rep_c20911]
<i>lsm7</i>	<i>LSM7 homolog, U6 small nuclear RNA associated (S. cerevisiae)</i> [Source:ZFIN; Acc:ZDB-GENE-030131-8284]	ENSDARG00000058328 ENSGACG00000010390 ENSORLG00000005164 ENSTNIG00000015525 ENSTRUG00000012826	ENSGT005 1000004787 2	4	[chionodraco_hamatus_c2586] [chionodraco_hamatus_c9465] [chionodraco_hamatus_c7965] [chionodraco_hamatus_c6321]
<i>lsm1</i>	<i>LSM domain containing 1</i> [Source:ZFIN; Acc:ZDB-GENE-030131-7762]	ENSDARG00000067677 ENSGACG00000019273 ENSORLG00000007119 ENSTNIG00000012999 ENSTRUG00000015264	ENSGT003 9000001841 8	2	[chionodraco_hamatus_c1850] [chionodraco_hamatus_c9212, chionodraco_hamatus2_c381]
<i>med9</i>	<i>mediator complex subunit 9</i> [Source:HGNC Symbol; Acc:25487]	ENSGACG00000011197 ENSORLG00000007978 ENSTNIG00000000294 ENSTRUG00000009409	ENSGT003 9000001737 9	2	[chionodraco_hamatus_c11019] [chionodraco_hamatus_c4659]
<i>mgst3</i>	<i>microsomal glutathione S-transferase 3</i> [Source:HGNC Symbol; Acc:7064]	ENSDARG00000024143 ENSGACG00000016252 ENSORLG00000010182 ENSTNIG00000007858 ENSTRUG00000000749	ENSGT003 9000000860 8	2	[chionodraco_hamatus_c5854] [chionodraco_hamatus_c2694]
<i>mrpl-13</i>	<i>mitochondrial ribosomal protein L13</i> [Source:ZFIN; Acc:ZDB-GENE-050522-167]	ENSDARG00000056903 ENSGACG00000007061 ENSORLG00000004647 ENSTNIG00000009643 ENSTRUG00000012278	ENSGT003 9000000151 5	2	[chionodraco_hamatus_c2906] [chionodraco_hamatus_c937, chionodraco_hamatus_c8628]

<i>mrpl-17</i>	<i>mitochondrial ribosomal protein L17</i> [Source:ZFIN; Acc:ZDB-GENE-040912-133]	ENSDARG00000075677 ENSGACG00000003236 ENSORLG00000013563 ENSTNIG00000008195 ENSTRUG00000001343	ENSGT003 9000001069 8	2	[chionodraco_hamatus_c8419] [chionodraco_hamatus_c119]
<i>mrpl-2</i>	<i>mitochondrial ribosomal protein L2</i> [Source:HGNC Symbol; Acc:14056]	ENSDARG00000062568 ENSGACG00000004104 ENSORLG00000012882 ENSTNIG00000013424 ENSTRUG00000000908	ENSGT003 9000000018 9	2	[chionodraco_hamatus2_c826] [chionodraco_hamatus_rep_c22384]
<i>mrpl-48</i>	<i>mitochondrial ribosomal protein L48</i> [Source:ZFIN; Acc:ZDB-GENE-041008-125]	ENSDARG00000035167 ENSGACG00000018902 ENSORLG00000008682 ENSTNIG00000007416 ENSTRUG00000005273	ENSGT003 9000001295 5	2	[chionodraco_hamatus_c839] [chionodraco_hamatus2_c2238]
<i>mrpl-52</i>	<i>mitochondrial ribosomal protein L52</i> [Source:HGNC Symbol; Acc:16655]	ENSDARG00000090568 ENSGACG00000019198 ENSORLG00000007322 ENSTNIG00000013839	ENSGT003 9000000576 3	2	[chionodraco_hamatus2_c514] [chionodraco_hamatus_c2957]
<i>mrpl-55</i>	<i>mitochondrial ribosomal protein L55</i> [Source:HGNC Symbol; Acc:16686]	ENSDARG00000043402 ENSGACG00000002136 ENSTRUG00000003082	ENSGT003 9000001030 9	2	[chionodraco_hamatus2_c2587] [chionodraco_hamatus2_c886, chionodraco_hamatus_c4550]
<i>mterfd1</i>	<i>MTERF domain containing 1</i> [Source:ZFIN; Acc:ZDB-GENE-040718-359]	ENSDARG00000003207 ENSGACG00000005407 ENSORLG00000012689 ENSTNIG00000009184 ENSTRUG000000010597	ENSGT003 9000000580 1	2	[chionodraco_hamatus_c3454] [chionodraco_hamatus_c16786]
<i>myoz1a</i>	<i>myozenin 1a</i> [Source:ZFIN; Acc:ZDB-GENE-040426-1875]	ENSDARG00000056209 ENSGACG00000010907 ENSORLG00000004086 ENSTNIG00000007456 ENSTRUG00000002720	ENSGT005 3000006318 4	2	[chionodraco_hamatus2_rep_c3055] [chionodraco_hamatus2_c452]
<i>ndufa11</i>	<i>NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 11, 14.7kDa</i> [Source:HGNC Symbol; Acc:20371]	ENSDARG00000042777 ENSGACG00000002132 ENSORLG00000002879 ENSTNIG00000007145 ENSTRUG00000016683	ENSGT003 9000001243 4	2	[chionodraco_hamatus_rep_c13626] [chionodraco_hamatus2_c2357]
<i>ndufa5</i>	<i>NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5</i> [Source:ZFIN; Acc:ZDB-GENE-050320-17]	ENSDARG00000039346 ENSGACG00000019248 ENSORLG00000013008 ENSTNIG00000001461 ENSTRUG00000010258	ENSGT003 9000000809 9	2	[chionodraco_hamatus2_rep_c3324, chionodraco_hamatus2_c785] [chionodraco_hamatus_rep_c28038]
<i>ndufa6</i>	<i>NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6</i> [Source:ZFIN; Acc:ZDB-GENE-040426-1124]	ENSDARG00000038028 ENSGACG00000007237 ENSORLG00000014731 ENSTNIG00000010838 ENSTRUG00000007261	ENSGT003 9000001889 8	2	[chionodraco_hamatus_rep_c22368] [chionodraco_hamatus_rep_c9985, chionodraco_hamatus2_rep_c3006]
<i>ndufb5</i>	<i>NADH dehydrogenase (ubiquinone) 1 beta subcomplex 5</i> [Source:ZFIN; Acc:ZDB-GENE-011010-1]	ENSDARG00000070824 ENSGACG00000009997 ENSORLG00000005708 ENSTNIG00000015544 ENSTRUG00000017599	ENSGT003 9000000998 0	2	[chionodraco_hamatus_rep_c9797] [chionodraco_hamatus2_c374]
<i>ndufb8</i>	<i>NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8</i> [Source:ZFIN; Acc:ZDB-GENE-040426-1858]	ENSDARG00000010113 ENSGACG00000002420 ENSORLG00000009284 ENSTNIG00000019397 ENSTRUG00000012214	ENSGT003 9000000062 8	2	[chionodraco_hamatus_rep_c17533, chionodraco_hamatus2_c1090, chionodraco_hamatus2_rep_c3060] [chionodraco_hamatus2_rep_c3146]

<i>ndufs7</i>	<i>NADH dehydrogenase (ubiquinone) Fe-S protein 7, (NADH-coenzyme Q reductase)</i> [Source:ZFIN; Acc:ZDB-GENE-041111-261]	ENSDARG00000074552 ENSGACG00000008341 ENSORLG00000008222 ENSTNIG00000010203 ENSTRUG00000003488	ENSGT003 9000000656 5	2	[chionodraco_hamatus_rep_c27180, chionodraco_hamatus2_c178] [chionodraco_hamatus2_rep_c3164]
<i>nol7</i>	<i>nucleolar protein 7, 27kDa</i> [Source:HGNC Symbol; Acc:21040]	ENSDARG00000075795 ENSGACG00000017989 ENSORLG00000006271 ENSTNIG00000007997 ENSTRUG00000002295	ENSGT003 9000000411 8	2	[chionodraco_hamatus_c3237] [chionodraco_hamatus_c4635]
<i>npc2</i>	<i>Niemann-Pick disease, type C2</i> [Source:ZFIN; Acc:ZDB-GENE-021206-13]	ENSDARG00000090912 ENSGACG00000007718 ENSORLG00000016360 ENSTNIG00000017398 ENSTRUG00000018269	ENSGT003 9000000622 3	2	[chionodraco_hamatus2_rep_c3313] [chionodraco_hamatus_rep_c16813, chionodraco_hamatus2_rep_c3004, chionodraco_hamatus_rep_c26481, chionodraco_hamatus2_rep_c3073, chionodraco_hamatus_c8647, chionodraco_hamatus2_rep_c2984]
<i>nupr1</i>	<i>nuclear protein 1</i> [Source:ZFIN; Acc:ZDB-GENE-030131-4653]	ENSDARG00000094557 ENSGACG00000015005 ENSORLG00000018286 ENSTRUG00000000816	ENSGT005 3000006424 2	2	[chionodraco_hamatus2_c241] [chionodraco_hamatus_c4541]
<i>ostc</i>	<i>oligosaccharyltransferase complex subunit</i> [Source:ZFIN; Acc:ZDB-GENE-040426-1079]	ENSDARG00000035838 ENSGACG00000019601 ENSORLG00000004900 ENSTNIG00000018847 ENSTRUG000000005991	ENSGT003 9000000137 6	2	[chionodraco_hamatus_rep_c11351, chionodraco_hamatus2_c762] [chionodraco_hamatus2_rep_c3191]
<i>otud1</i>	<i>OTU domain containing 1</i> [Source:HGNC Symbol; Acc:27346]	ENSGACG00000003846 ENSTNIG00000012772 ENSTRUG00000017846	ENSGT005 1000004963 5	2	[chionodraco_hamatus_c2246] [chionodraco_hamatus_c2112]
<i>oxal1</i>	<i>oxidase (cytochrome c) assembly 1-like</i> [Source:ZFIN; Acc:ZDB-GENE-071004-49]	ENSDARG00000069313 ENSGACG00000019211 ENSORLG00000007262 ENSTNIG00000013010 ENSTRUG00000015484	ENSGT005 3000006350 6	2	[chionodraco_hamatus2_c1219] [chionodraco_hamatus_c8354]
<i>pfdn2</i>	<i>prefoldin subunit 2</i> [Source:ZFIN; Acc:ZDB-GENE-060519-27]	ENSDARG00000025391 ENSGACG00000019042 ENSORLG00000007959 ENSTNIG00000003655 ENSTRUG00000004520	ENSGT003 9000000927 2	2	[chionodraco_hamatus_rep_c27535] [chionodraco_hamatus_c6908, chionodraco_hamatus2_c120]
<i>pinx1</i>	<i>pin2/trf1-interacting protein 1</i> [Source:ZFIN; Acc:ZDB-GENE-030515-2]	ENSDARG00000023532 ENSGACG00000010809 ENSORLG00000016376 ENSTNIG00000016334 ENSTRUG00000006295	ENSGT004 5000004027 9	2	[chionodraco_hamatus_c153] [chionodraco_hamatus_c4112]
<i>pold3</i>	<i>polymerase (DNA-directed), delta 3, accessory subunit</i> [Source:ZFIN; Acc:ZDB-GENE-060512-59]	ENSDARG00000076169 ENSGACG00000017641 ENSORLG00000002322 ENSTNIG00000007964 ENSTRUG00000001631	ENSGT003 9000001528 2	2	[chionodraco_hamatus_c2267] [chionodraco_hamatus_c10321]
<i>pole4</i>	<i>polymerase (DNA-directed), epsilon 4 (p12 subunit)</i> [Source:HGNC Symbol; Acc:18755]	ENSDARG00000013016 ENSGACG00000011233 ENSORLG00000017077 ENSTNIG00000015477 ENSTRUG00000000810	ENSGT005 3000006356 0	2	[chionodraco_hamatus_c7425] [chionodraco_hamatus_c8197]
<i>psenen</i>	<i>presenilin enhancer 2 homolog</i> [Source:ZFIN; Acc:ZDB-GENE-040218-1]	ENSDARG00000068698 ENSGACG00000007626 ENSORLG00000006446 ENSTNIG00000018469 ENSTRUG00000015233	ENSGT003 9000001631 9	2	[chionodraco_hamatus_c378, chionodraco_hamatus_rep_c10944] [chionodraco_hamatus_c9198]

<i>psmb7</i>	<i>proteasome (prosome, macropain) subunit, beta type, 7</i> [Source:ZFIN; Acc:ZDB-GENE-001208-4]	ENSDARG00000037962 ENSGACG00000018309 ENSORLG00000013170 ENSTNIG00000019654 ENSTRUG00000014190	ENSGT005 1000004653 3	2	[chionodraco_hamatus2_c96] [chionodraco_hamatus_c8165]
<i>psmd4a</i>	<i>proteasome (prosome, macropain) 26S subunit, non-ATPase, 4a</i> [Source:ZFIN; Acc:ZDB-GENE-040625-104]	ENSDARG00000036715 ENSGACG00000004332 ENSORLG00000000320 ENSTNIG00000017570 ENSTRUG00000009274	ENSGT005 3000006405 0	2	[chionodraco_hamatus_c3136] [chionodraco_hamatus_c6468]
<i>psme2</i>	<i>proteasome activator subunit 2</i> [Source:ZFIN; Acc:ZDB-GENE-991110-16]	ENSDARG00000033144 ENSGACG00000001681 ENSORLG00000013408 ENSTNIG00000008696 ENSTRUG00000017730	ENSGT005 1000004637 4	2	[chionodraco_hamatus_c8824] [chionodraco_hamatus_c2538]
<i>rbm8a</i>	<i>RNA binding motif protein 8A</i> [Source:ZFIN; Acc:ZDB-GENE-050306-51]	ENSDARG00000016516 ENSGACG00000013157 ENSORLG00000016263 ENSTNIG00000006869 ENSTRUG00000013173	ENSGT005 5000007464 1	2	[chionodraco_hamatus2_c2000] [chionodraco_hamatus_rep_c9737]
<i>rnf24</i>	<i>ring finger protein 24</i> [Source:ZFIN; Acc:ZDB-GENE-041114-40]	ENSDARG00000036965 ENSGACG00000017626 ENSORLG00000006343 ENSTNIG00000018049 ENSTRUG00000001425	ENSGT006 2000008774 5	2	[chionodraco_hamatus_c7023] [chionodraco_hamatus_c2502, chionodraco_hamatus_c6887]
<i>rnf7</i>	<i>ring finger protein 7</i> [Source:ZFIN; Acc:ZDB-GENE-050220-12]	ENSDARG00000055524 ENSGACG00000007590 ENSORLG00000011622 ENSTNIG00000005982 ENSTRUG00000006552	ENSGT003 9000001705 8	2	[chionodraco_hamatus2_c495, chionodraco_hamatus_rep_c15582, chionodraco_hamatus_rep_c26058] [chionodraco_hamatus2_rep_c3231]
<i>romo1</i>	<i>reactive oxygen species modulator 1</i> [Source:HGNC Symbol; Acc:16185]	ENSDARG00000038076 ENSGACG00000003603 ENSORLG00000020541	ENSGT003 9000000531 5	3	[chionodraco_hamatus2_rep_c3079] [chionodraco_hamatus_rep_c13875] [chionodraco_hamatus2_rep_c2985]
<i>rpl-14</i>	<i>ribosomal protein L14</i> [Source:ZFIN; Acc:ZDB-GENE-030131-2085]	ENSDARG00000036174 ENSGACG00000008265 ENSTNIG00000009252 ENSTRUG00000013316	ENSGT003 9000000788 8	2	[chionodraco_hamatus_rep_c18299] [chionodraco_hamatus2_c1587, chionodraco_hamatus_rep_c17147, chionodraco_hamatus_rep_c16334]
<i>rpl-17</i>	<i>ribosomal protein L17</i> [Source:HGNC Symbol; Acc:10307]	ENSDARG00000057556 ENSGACG00000013435 ENSORLG00000020336 ENSTNIG00000009011 ENSTRUG00000018649	ENSGT003 9000001487 3	2	[chionodraco_hamatus_c7206] [chionodraco_hamatus_c3142]
<i>rpl-22l1</i>	<i>ribosomal protein L22-like 1</i> [Source:ZFIN; Acc:ZDB-GENE-060804-3]	ENSDARG00000010244 ENSGACG00000012539 ENSORLG00000002916 ENSTNIG00000002562 ENSTRUG00000003154	ENSGT003 9000000371 9	2	[chionodraco_hamatus2_c2239] [chionodraco_hamatus_rep_c17647]
<i>rpl-23a</i>	<i>ribosomal protein L23a</i> [Source:ZFIN; Acc:ZDB-GENE-030131-7479]	ENSDARG00000006316 ENSGACG00000020625 ENSORLG00000005524 ENSTRUG00000017075	ENSGT003 9000000326 3	2	[chionodraco_hamatus2_c175] [chionodraco_hamatus2_rep_c3203]
<i>rpl-27a</i>	<i>ribosomal protein L27a</i> [Source:HGNC Symbol; Acc:10329]	ENSGACG00000006523 ENSORLG00000001144 ENSTNIG00000013740 ENSTRUG00000009256	ENSGT003 9000000553 4	2	[chionodraco_hamatus2_rep_c2980, chionodraco_hamatus_rep_c26225] [chionodraco_hamatus_rep_c27355]
<i>rpl-31</i>	<i>ribosomal protein L31</i> [Source:HGNC Symbol; Acc:10334]	ENSDARG00000053365 ENSGACG00000015169 ENSORLG00000005178 ENSTNIG00000014522 ENSTRUG00000007359	ENSGT003 9000000520 0	2	[chionodraco_hamatus_rep_c16011] [chionodraco_hamatus_rep_c12584, chionodraco_hamatus2_c1878]

<i>rpl-32</i>	<i>ribosomal protein L32</i> [Source:HGNC Symbol; Acc:10336]	ENSDARG00000054818 ENSGACG00000002699 ENSTNIG00000017891 ENSTRUG00000009522	ENSGT003 9000001472 9	2	[chionodraco_hamatus2_c1854] [chionodraco_hamatus_rep_c16067]
<i>rpl-4</i>	<i>ribosomal protein L4</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-9034]	ENSDARG00000041182 ENSGACG00000004968 ENSORLG00000006882 ENSTNIG00000008428 ENSTRUG00000014667	ENSGT003 9000001814 5	2	[chionodraco_hamatus2_c1134] [chionodraco_hamatus_c18116]
<i>rpl-6</i>	<i>ribosomal protein L6</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-8671]	ENSDARG00000058451 ENSGACG00000004055 ENSORLG00000002519 ENSORLG000000020235 ENSTNIG00000015921 ENSTRUG00000016438	ENSGT003 9000000368 2	2	[chionodraco_hamatus2_c1460] [chionodraco_hamatus_rep_c23097]
<i>rpl-711</i>	<i>ribosomal protein L7- like 1</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-970]	ENSDARG00000042864 ENSGACG00000004217 ENSORLG00000018582 ENSTNIG00000002709 ENSTNIG00000019645 ENSTRUG00000007654	ENSGT003 9000000621 3	3	[chionodraco_hamatus2_c2428] [chionodraco_hamatus2_rep_c3155] [chionodraco_hamatus_rep_c10659]
<i>rpl-9</i>	<i>ribosomal protein L9</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-8646]	ENSDARG00000037350 ENSGACG00000016153 ENSORLG00000005590 ENSTNIG00000005607 ENSTRUG00000009221	ENSGT003 9000001522 4	2	[chionodraco_hamatus_rep_c28117, chionodraco_hamatus_rep_c27206, chionodraco_hamatus2_c2004, chionodraco_hamatus_rep_c19974, chionodraco_hamatus_rep_c14854, chionodraco_hamatus_rep_c22551, chionodraco_hamatus_rep_c11717, chionodraco_hamatus_rep_c18526, chionodraco_hamatus_rep_c27617, chionodraco_hamatus_rep_c9840] [chionodraco_hamatus_rep_c22126]
<i>rpl-p1</i>	<i>ribosomal protein, large, P1</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-8663]	ENSDARG00000021864 ENSGACG00000003995 ENSORLG00000019221 ENSTNIG00000009191 ENSTRUG00000005235	ENSGT005 5000007469 8	2	[chionodraco_hamatus_c9389] [chionodraco_hamatus_c4585]
<i>rps-12</i>	<i>ribosomal protein S12</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-8951]	ENSDARG00000036875 ENSGACG00000010916 ENSORLG00000016460 ENSTNIG00000016328 ENSTRUG00000004877	ENSGT003 9000001831 8	2	[chionodraco_hamatus2_c1311, chionodraco_hamatus_rep_c25477, chionodraco_hamatus_c6715] [chionodraco_hamatus_rep_c18899]
<i>rps-28</i>	<i>ribosomal protein S28</i> [Source:ZFIN; Acc:ZDB-GENE- 030131-2022]	ENSDARG00000035860 ENSGACG00000015568 ENSORLG00000005256 ENSTNIG00000011061 ENSTRUG00000009208	ENSGT003 9000000358 0	2	[chionodraco_hamatus2_c2783] [chionodraco_hamatus_rep_c14448, chionodraco_hamatus_rep_c19035, chionodraco_hamatus_rep_c23341]
<i>sar1a</i>	<i>SAR1 gene homolog A (S. cerevisiae)</i> [Source:ZFIN; Acc:ZDB-GENE- 050417-437]	ENSDARG00000033320 ENSGACG00000008616 ENSORLG00000006779 ENSTNIG00000008635 ENSTRUG00000007269	ENSGT005 5000007469 6	2	[chionodraco_hamatus_c6318] [chionodraco_hamatus_c1762]
<i>sarnp</i>	<i>SAP domain containing ribonucleoprotein</i> [Source:HGNC Symbol; Acc:24432]	ENSDARG00000059357 ENSGACG00000011615 ENSORLG00000016477 ENSTNIG00000012432 ENSTRUG00000009497	ENSGT003 9000000294 4	2	[chionodraco_hamatus2_rep_c2995] [chionodraco_hamatus2_c1510]
<i>sdhc</i>	<i>succinate dehydrogenase complex, subunit C, integral membrane protein</i> [Source:ZFIN; Acc:ZDB-GENE- 040801-26]	ENSDARG00000038608 ENSGACG00000017685 ENSORLG00000002673 ENSTNIG00000003624 ENSTRUG00000001873	ENSGT003 9000000056 6	2	[chionodraco_hamatus_c7737] [chionodraco_hamatus_c3396, chionodraco_hamatus_c16193]

<i>sept7</i>	<i>septin 7</i> [Source:HGNC Symbol; Acc:1717]	ENSDARG00000052673 ENSGACG00000012939 ENSORLG00000016371 ENSTNIG0000000401 ENSTNIG00000011655 ENSTNIG00000011658 ENSTRUG00000008722	ENSGT006 2000008764 9	2	[chionodraco_hamatus_c25911] [chionodraco_hamatus_c15744]
<i>skal</i>	<i>spindle and kinetochore associated complex subunit 1</i> [Source:HGNC Symbol; Acc:28109]	ENSDARG00000039354 ENSGACG00000002020 ENSORLG00000018727 ENSTNIG00000004768 ENSTRUG00000017348	ENSGT003 9000001165 4	2	[chionodraco_hamatus_c23586] [chionodraco_hamatus_c6370]
<i>slc38a3</i>	<i>solute carrier family 38, member 3</i> [Source:ZFIN; Acc:ZDB-GENE-040718-395]	ENSDARG00000014587 ENSGACG00000013354 ENSORLG00000019206 ENSTNIG00000003277 ENSTRUG00000001541	ENSGT005 5000007422 2	2	[chionodraco_hamatus_c1689] [chionodraco_hamatus_c15495]
<i>snrpd2</i>	<i>small nuclear ribonucleoprotein D2 polypeptide</i> [Source:ZFIN; Acc:ZDB-GENE-040914-10]	ENSDARG00000040440 ENSGACG00000006697 ENSORLG00000013278 ENSTNIG00000011736 ENSTRUG00000000274	ENSGT003 9000001760 8	3	[chionodraco_hamatus_rep_c14049] [chionodraco_hamatus_rep_c13772] [chionodraco_hamatus2_c2510]
<i>ss18l2</i>	<i>synovial sarcoma translocation gene on chromosome 18-like 2</i> [Source:HGNC Symbol; Acc:15593]	ENSDARG00000039777 ENSGACG00000014918 ENSORLG00000014115 ENSTNIG00000010521 ENSTRUG00000015835	ENSGT005 0000004480 8	2	[chionodraco_hamatus2_c807] [chionodraco_hamatus_rep_c27636]
<i>supt4h1</i>	<i>suppressor of Ty 4 homolog 1 (S. cerevisiae)</i> [Source:HGNC Symbol; Acc:11467]	ENSDARG00000033143 ENSGACG00000020237 ENSORLG00000011097 ENSTNIG00000008553 ENSTRUG00000005839	ENSGT003 9000001855 9	2	[chionodraco_hamatus_c7008] [chionodraco_hamatus2_c1507]
<i>syf2</i>	<i>SYF2 homolog, RNA splicing factor (S. cerevisiae)</i> [Source:ZFIN; Acc:ZDB-GENE-040801-170]	ENSDARG00000004706 ENSGACG00000007305 ENSORLG00000017523 ENSTNIG00000016912 ENSTRUG00000001499	ENSGT003 9000001784 5	2	[chionodraco_hamatus2_c616] [chionodraco_hamatus_c8962]
<i>taf10</i>	<i>TAF10 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 30kDa</i> [Source:HGNC Symbol; Acc:11543]	ENSDARG00000068641 ENSGACG00000006435 ENSTNIG00000011991 ENSTRUG00000015379	ENSGT003 9000000936 8	2	[chionodraco_hamatus2_c656] [chionodraco_hamatus_c8617]
<i>taf13</i>	<i>TAF13 RNA polymerase II, TATA box binding protein (TBP)-associated factor</i> [Source:ZFIN; Acc:ZDB-GENE-030131-2873]	ENSDARG00000070834 ENSGACG00000004213 ENSORLG00000001804 ENSORLG00000002518 ENSTNIG00000002526 ENSTRUG00000004250	ENSGT003 9000001298 1	2	[chionodraco_hamatus2_c945] [chionodraco_hamatus_c8084]
<i>timt9</i>	<i>translocase of inner mitochondrial membrane 9 homolog</i> [Source:ZFIN; Acc:ZDB-GENE-021206-14]	ENSDARG00000043865 ENSGACG00000004009 ENSORLG00000010167 ENSTNIG00000002594 ENSTRUG00000005239 ENSTRUG00000009048	ENSGT004 5000004027 8	2	[chionodraco_hamatus_rep_c14546, chionodraco_hamatus2_c1315] [chionodraco_hamatus2_rep_c3192]
<i>tmem14c</i>	<i>transmembrane protein 14C</i> [Source:ZFIN; Acc:ZDB-GENE-060825-156]	ENSDARG00000075180 ENSGACG00000002339 ENSORLG00000006645 ENSTNIG00000005736 ENSTRUG00000005743	ENSGT003 9000001347 5	3	[chionodraco_hamatus2_rep_c3052] [chionodraco_hamatus_rep_c20412] [chionodraco_hamatus2_c1181]

<i>tmem176l.4</i>	<i>transmembrane protein 176l.4</i> [Source:ZFIN; Acc:ZDB-GENE-030131-3532]	ENSDARG00000074390 ENSDARG00000086230 ENSDARG00000089399 ENSDARG00000094439 ENSGACG00000013220 ENSORLG00000016141 ENSORLG00000016144 ENSTNIG00000002947	ENSGT005 1000005167 5	2	[chionodraco_hamatus2_c2286] [chionodraco_hamatus_c13859]
<i>tmem50a</i>	<i>transmembrane protein 50A</i> [Source:ZFIN; Acc:ZDB-GENE-040426-2937]	ENSDARG00000015757 ENSGACG00000007351 ENSORLG00000017537 ENSTNIG00000016914 ENSTRUG00000001698	ENSGT003 9000000743 4	2	[chionodraco_hamatus2_rep_c3078] [chionodraco_hamatus_rep_c27398, chionodraco_hamatus2_c603]
<i>trappc2</i>	trafficking protein particle complex 2 [Source:ZFIN; Acc:ZDB-GENE-060929-1266]	ENSDARG00000068831 ENSORLG00000017809 ENSTRUG00000008730	ENSGT005 1000004716 8	2	[chionodraco_hamatus_rep_c25939] [chionodraco_hamatus_c1122, chionodraco_hamatus_rep_c25003]
<i>txndc9</i>	<i>thioredoxin domain containing 9</i> [Source:ZFIN; Acc:ZDB-GENE-030131-8569]	ENSDARG00000069853 ENSGACG00000002912 ENSORLG00000012228 ENSTNIG00000013871 ENSTRUG00000014285	ENSGT003 9000001564 5	2	[chionodraco_hamatus2_c1930] [chionodraco_hamatus2_c25, chionodraco_hamatus_c9589, chionodraco_hamatus2_rep_c3162]
<i>ubl5</i>	<i>ubiquitin-like 5</i> [Source:HGNC Symbol; Acc:13736]	ENSGACG00000016584 ENSORLG00000009337 ENSTNIG00000008267 ENSTRUG00000013643	ENSGT003 9000000194 5	3	[chionodraco_hamatus_c6894] [chionodraco_hamatus_c8538] [chionodraco_hamatus2_c1927]
<i>ufd1l</i>	<i>ubiquitin fusion degradation 1-like</i> [Source:ZFIN; Acc:ZDB-GENE-040718-150]	ENSDARG00000010051 ENSGACG00000003940 ENSORLG00000002611 ENSTNIG00000015923 ENSTRUG00000016457	ENSGT003 9000000240 8	2	[chionodraco_hamatus2_c1960] [chionodraco_hamatus_c9625]
<i>uqcrb</i>	<i>ubiquinol-cytochrome c reductase binding protein</i> [Source:ZFIN; Acc:ZDB-GENE-050522-542]	ENSDARG00000011146 ENSGACG00000005419 ENSORLG00000012697 ENSTNIG00000009185 ENSTRUG00000010634	ENSGT003 9000001291 6	2	[chionodraco_hamatus_rep_c13725, chionodraco_hamatus_rep_c9822] [chionodraco_hamatus2_c811]
<i>uqcrc2a</i>	<i>ubiquinol-cytochrome c reductase core protein Ila</i> [Source:ZFIN; Acc:ZDB-GENE-040718-405]	ENSDARG00000014794 ENSGACG00000011541 ENSORLG00000008672 ENSTNIG00000017979 ENSTRUG00000016072	ENSGT005 5000007466 6	2	[chionodraco_hamatus2_c1588] [chionodraco_hamatus_c812]
<i>utp3</i>	<i>UTP3, small subunit (SSU) processome component, homolog (S. cerevisiae)</i> [Source:HGNC Symbol; Acc:24477]	ENSDARG00000056720 ENSGACG00000007417 ENSORLG00000004019 ENSTNIG00000014783 ENSTRUG00000014591	ENSGT005 0000004494 7	2	[chionodraco_hamatus_c7207, chionodraco_hamatus2_c2836] [chionodraco_hamatus_c717]
<i>zgc:56493</i>	<i>zgc:56493</i> [Source:ZFIN; Acc:ZDB-GENE-030131-8581]	ENSDARG00000031435 ENSDARG00000044125 ENSGACG00000019142 ENSORLG00000007575 ENSTNIG00000013854 ENSTRUG00000004064	ENSGT005 3000006300 8	5	[chionodraco_hamatus_rep_c16146] [chionodraco_hamatus2_rep_c3120] [chionodraco_hamatus_rep_c17579, chionodraco_hamatus2_c780] [chionodraco_hamatus_rep_c28167] [chionodraco_hamatus_c3776, chionodraco_hamatus2_rep_c3056, chionodraco_hamatus2_rep_c3001]
<i>zgc:92254</i>	<i>zgc:92254</i> [Source:ZFIN; Acc:ZDB-GENE-040718-365]	ENSDARG00000022183 ENSDARG00000033285 ENSGACG00000009287 ENSORLG00000006192 ENSORLG00000006201 ENSTNIG00000013320 ENSTRUG00000009691	ENSGT003 9000000547 9	2	[chionodraco_hamatus_c2381] [chionodraco_hamatus_c4569]

<i>znrd1</i>	<i>zinc ribbon domain containing 1</i> [Source:ZFIN; Acc:ZDB-GENE-050320-131]	ENSDARG00000036070 ENSGACG00000001469 ENSORLG00000010654 ENSTNIG00000005004 ENSTNIG00000005917 ENSTRUG00000005223	ENSGT003 9000000812 6	2	[chionodraco_hamatus_c7272] [chionodraco_hamatus_c6870]
--------------	---	--	-----------------------------	---	--

References:

- Chen Z, et al. 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc Natl Acad Sci U S A*. 105: 12944-12949. doi: 10.1073/pnas.0802432105.
- Coppe A, et al. 2010. Sequencing, de novo annotation and analysis of the first *Anguilla anguilla* transcriptome: EelBase opens new perspectives for the study of the critically endangered European eel. *BMC Genomics*. 11: 635. doi: 10.1186/1471-2164-11-635.
- Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA. 2011. Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. *BMC Genomics*. 12: 202. doi: 10.1186/1471-2164-12-202.
- Zhang Z, et al. 2011. Transcriptome analysis of female and male *Xiphophorus maculatus* Jp 163 A. *PLoS ONE*. 6: e18379. doi: 10.1371/journal.pone.0018379.
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques*. 30: 892-897.
- Zhulidov PA, et al. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res*. 32. doi: 10.1093/nar/gnh031.

PAPER II

1 **Isolation, characterization and multiplexing of expressed sequence tag-linked**
2 **microsatellite loci for the Antarctic icefish *Chionodraco hamatus***

3
4 Ilaria AM Marino¹, Cecilia Agostini¹, Chiara Papetti¹, Paolo Maria Bisol¹, Lorenzo Zane^{1*}, and
5 Tomaso Patarnello²

6
7 **1** Department of Biology-University of Padova, via U. Bassi 58/b, I-35121 Padova, Italy

8 **2** Department of Public Health, Comparative Pathology and Veterinary Hygiene-University of
9 Padova, Agripolis, via Romea 16, I-35020 Legnaro (PD), Italy

10

11

12 **Keywords:** *Chionodraco hamatus*, Antarctic icefish, Notothenioidei, Southern Ocean, EST-linked
13 microsatellite

14

15 **Running title:** Antarctic icefish EST-linked microsatellite loci

16

17

18

19 *** CORRESPONDENCE TO:**

20 Lorenzo Zane

21 Department of Biology, University of Padova

22 Via G. Colombo 3 I-35121, Padova, Italy

23 Tel. +39 049 8276220, Fax: +39 049 8276209

24 e.mail: lorenz@bio.unipd.it

25 **Abstract**

26 An Antarctic icefish (*Chionodraco hamatus*) expressed sequence tag database was screened for
27 microsatellites sequences. Primers were designed to amplify 12 loci, of which 11 produced high-
28 quality amplification products and were successfully multiplexed in a single PCR. Allele number
29 ranged from two to 9, whereas observed and expected heterozygosity from 0.0435 to 0.8261, and
30 from 0.0435 to 0.8580, respectively. No evidence for Hardy-Weinberg disequilibrium was found.
31 All loci were successfully genotyped in two related species of the same genus (*C. myersi* and *C.*
32 *rastrospinosus*). These molecular markers will be useful to investigate genetic structure of this, and
33 potentially other, icefish species.

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50 The Antarctic icefish *Chionodraco hamatus* (Notothenioidea, Channichthyidae) is, like other
51 icefish, a benthonic species characterized by peculiar adaptations to the Southern Ocean
52 environment, such as the presence of antifreeze glycoproteins that prevent the freezing of body
53 fluids (DeVries 1988) and the complete lack of hemoglobin (Ruud 1954), a unique condition among
54 vertebrates. This species has a circumpolar distribution, with the exception of the Antarctic
55 Peninsula, and occurs at depths between 4 and 600 m (Gon & Heemstra 1990, Eastman 1993).

56 In this note, we describe the isolation and characterization of 11 polymorphic EST-linked
57 microsatellites that will be useful tools for studying the population structure of *C. hamatus*. In
58 particular, these markers will be helpful to further investigate the weak genetic structure, with
59 significant differences between samples in space and time that has been found within the species
60 distribution range using the mitochondrial control region (Patarnello *et al.* 2003).

61 A total of 27800 contigs have been obtained by high-throughput sequencing of a normalized cDNA
62 library from muscle of *C. hamatus* (manuscript in preparation). These sequences were searched for
63 microsatellite repeats using the software MIRA 3 (Chevreux *et al.* 2004) with the following criteria:
64 nine repeats for di-, seven for tri-, and five for tetra-, penta-, and exanucleotide motifs. A total of
65 434 microsatellites were found, including 200 dinucleotide, 175 trinucleotide, 48 tetranucleotide, 9
66 pentanucleotide and 2 exanucleotide repeats.

67 Twelve loci were selected for primer design and optimization. Primers pairs for amplification of
68 microsatellite loci were designed with the software FASTPCR (Kalendar *et al.* 2009) to amplify
69 three different size products (small size: 85-130 bp, medium size: 130-180 bp and large size: 180-
70 280 bp). The major parameters for primer design were set as follows: primer length from 21 to 22
71 nucleotides, no secondary structures, optimum annealing temperature at 53-68 °C, and GC contents
72 from 40 to 70 %. Primer pairs were tested for compatibility for future multiplexing. The selected
73 loci were tested on DNA samples from 23 *C. hamatus* individuals collected on 1996 in Terranova
74 Bay (11th Italian Expedition, Ross Sea, National Antarctic Research Program, PNRA) and preserved

75 at -80 °C. Genomic DNA was extracted from minute sections of tissue (10-100 mg) using a salting-
76 out protocol (Patwary *et al.* 1994). Polymerase chain reaction (PCR) products were obtained
77 separately for each of the twelve loci, with the following conditions: 1X *Taq* buffer (Promega, 50
78 mM KCl, 10 mM Tris-HCl pH = 9 at 25 °C, 0.1 % TritonX-100), 1 mM MgCl₂, 150 nM of each
79 primer, 70 μM dNTPs, 0.8 U of *Taq*, 100 ng of genomic DNA and water up to 20 μL. Thermal
80 profile, on a GeneAmp PCR System 2700 Thermocycler (Applied Biosystems), was: (i) pre-
81 denaturation for 2 min at 94 °C; (ii) 30 cycles: denaturation for 30 s at 94 °C, annealing for 40 s at
82 the primer-specific annealing temperature (Table 1), extension for 50 s at 72 °C; (iii) additional
83 extension for 5 min at 72 °C. Amplified fragments were visualized in 1.8 % agarose gels.

84 Out of the 12 primer sets, 11 gave high quality amplification products and only one showed an
85 incorrect size in comparison with the expected size on the basis of the EST sequence. These 11
86 primer pairs, including 5 di-, 3 tri-, 2 tetra- and one pentanucleotide repeats, were labeled with
87 fluorescent dyes for detection with ABI PRISM automatic sequencer (Table 1) and amplified
88 together in a single multiplex reaction. PCRs were carried out in 10 μL total volume containing:
89 1X QIAGEN Multiple PCR Master mix (QIAGEN, HotStarTaq DNA Polymerase, Multiplex PCR
90 Buffer, dNTP Mix), 0.2 μM primer mix and 100 ng of DNA template. PCR conditions were as
91 follows: initial activation step for 15 min at 95 °C, 30 cycles of 30 s at 94 °C, 90 s at 57 °C and 1
92 min at 72 °C, and final elongation for 30 min at 60 °C. All individuals tested (23 out of 23) showed
93 an amplification product at all 11 loci. Alleles were designated according to a LIZ 500 (50-500 bp)
94 marker (genotyping facility at www.bmr-genomics.com). Microsatellite analysis was carried out
95 using PEAK SCANNER ver. 1.0 (Applied Biosystems). Binning was automated with the software
96 FLEXIBIN ver. 2 (Amos *et al.* 2007).

97 Indices of variability, total number of alleles per locus, observed (H_O) and expected (H_E)
98 heterozygosity were calculated using GENETIX ver. 4.05.2 (Belkir *et al.* 2001). We used GENEPOP
99 online version (Raymond & Rousset 1995, Rousset 2008) to assess genetic variation and test

100 deviations from Hardy-Weinberg proportions (Fisher's exact test) and genotypic associations for all
101 possible pairs of loci (Fisher's exact test) within the sample (nominal significant threshold $\alpha =$
102 0.01).

103 All loci proved to be polymorphic (Table 1): the number of alleles per locus ranged from 2 to 9 and
104 the H_O and H_E ranged from 0.0435 to 0.8261 (mean value = 0.4308), and from 0.0435 to 0.8580
105 (mean value = 0.4441), respectively. No significant deviation from Hardy-Weinberg proportions
106 was detected; all loci resulted in linkage equilibrium at a 1 % nominal significance level.

107 We cross-amplified all 11 polymorphic loci in 2 closely related species of the genus *Chionodraco*,
108 *C. myersi* and *C. rastrospinosus* (Channichthyidae) collected on 1990 in Terranova Bay (5th Italian
109 Expedition, Ross Sea, PNRA) and on 2007 in Joinville Island (ANT-XXIII/8 cruise, Alfred
110 Wegener Institute for Polar and Marine Research, AWI), respectively. DNA extraction and
111 amplification conditions were the same as described above. All 11 primer sets (100 %) gave high
112 quality amplification products, cross-amplified and were found to be polymorphic in all (20 out of
113 20) individuals of each species tested.

114 These molecular markers will be useful to investigate icefish genetic structure and exposition to
115 global environment change, possibly providing insights on its effective population size and
116 demographic history.

117

118 **Acknowledgements**

119 This work was supported by the Italian Antarctic Research Program (PNRA). We thank prof. S.
120 Greco for samples collection during the 5th Italian Expedition, PNRA, V. Varotto for samples
121 collection during the 11th Italian Expedition, PNRA and E. Bortolotto for samples collection during
122 the ANT-XXIII/8 cruise, AWI.

123

124 **References**

- 125 Amos W, Hoffman JI, Frodsham A, Zhang L, Best S & Hill AVS (2007) Automated binning of
126 microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, **7**, 10-14.
- 127 Belkir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (2001) *GENETIX, Logiciel Sous Windows*
128 *Pour la Genetique Des Populations*. Laboratoire Genome et Populations, Montpellier, France.
- 129 Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T & Suhai S (2004) Using
130 the miraEST assembler for reliable and automated mRNA transcript assembly and SNP
131 detection in sequenced ESTs. *Genome Research*, **14**, 1147-1159.
- 132 DeVries AL (1988) The role of antifreeze glycopeptides and peptides in the freezing avoidance of
133 Antarctic fishes. *Comparative Biochemistry and Physiology - Part B: Biochemistry &*
134 *Molecular Biology*, **90**, 611-621.
- 135 Eastman JT (1993) Antarctic fish biology. Evolution in a unique environment. Academic press, San
136 Diego, California, 322 pp.
- 137 Gon O & Heemstra PC (1990) Fishes of the Southern Ocean. J.L.B. Smith Institute of Ichthyology,
138 Grahamstown, South Africa.
- 139 Kalendar R, Lee D & Schulman AH (2009) FastPCR software for PCR primer and probe design
140 and repeat search. *Genes, Genomes and Genomics*, **3**, 1-14.
- 141 Patarnello T, Marcato S, Zane L, Varotto V & Bargelloni L (2003) Phylogeography of the
142 *Chionodraco* genus (Perciformes, Channichthyidae) in the Southern Ocean. *Molecular*
143 *Phylogenetics and Evolution*, **28**, 420-429.
- 144 Patwary MU, Kenchington EL, Bird CJ & Zouros E (1994) The use of random amplified
145 polymorphic DNA markers in genetic studies of the sea-scallop *Placopecten magellanicus*
146 (Gmelin 1791). *Journal of Shellfish Research*, **13**, 547-553.
- 147 Raymond M & Rousset F (1995) Genepop version 1.2: population genetics software for exact tests
148 and ecumenism. *Journal of Heredity*, **86**, 248-249.

- 149 Rousset F (2008) Genepop '007: a complete reimplementation of the Genepop software for
150 Windows and Linux. *Molecular Biology and Evolution*, **4**, 406-425.
- 151 Ruud JT (1954) Vertebrate without erythrocytes and blood pigment. *Nature*, **173**, 848-850.

Table 1. Amplification conditions and descriptive statistics for 11 novel *Chionodraco hamatus* EST-linked microsatellite loci. The table reports: locus names, primer pairs sequences, fluorescent dye of forward primer, repeat motifs, size range of amplified fragments in bp (s.r.), annealing temperature for single locus amplification (see note), the (N/N_a) number of individuals assayed/number of alleles detected, observed and expected heterozygosity (H_o and H_E), deviation from Hardy-Weinberg proportions (P_(HW)) and the genbank accession number of each locus (GB acc. no.).

Locus	Primers (5'-3')	Dye	Repeat motifs	s.r.	T _a (°C)	N/N _a	H _o	H _E	P _(HW)	GB acc. no.
Ch126	F: CGGTTTTTATGCATGTTGCCA	NED	(AAT) ₇	276-283	56	23/2	0.0435	0.0435	1.0000	HQ395761
	R: ACTGCTCATTACACTGGTTC									
Ch623	F: GCTGTTGATTCCTCGTGAGG	PET	(AAC) ₇	206-225	62	23/6	0.6522	0.7237	0.1102	HQ395762
	R: AAAAGTGGTCTCCGCTGCAGT									
Ch684	F: CCAACCTGAGGGCCCAACAAC	6-FAM	(AGAA) ₆	129-146	62	23/3	0.3913	0.3816	1.0000	HQ395763
	R: GGTTCATGGGCTGGCCAATC									
Ch1968	F: AGCTCAAGGTGTTCCGAGACG	PET	(ACA) ₇	97-106	60	23/4	0.3478	0.3140	1.0000	HQ395764
	R: TAGCCAGCAGCGCTAATCCTG									
Ch2309	F: CAGCTCAATTAAACGCTTTGCA	6-FAM	(CA) ₁₀	207-227	58	23/9	0.8261	0.8580	0.8887	HQ395765
	R: CGTCTCAAATGCTGCACAACC									
Ch2788	F: TGGTTTTCGATGAAGAATGCTC	VIC	(CT) ₉	264-283	56	23/5	0.5652	0.5671	0.2456	HQ395766
	R: TGATGATAATTGGTCTGGTCG									
Ch3603	F: TGAGACAAGTCAAATCCAAACC	VIC	(TTTAA) ₅	94-105	56	23/3	0.1304	0.2019	0.2144	HQ395767
	R: GGCATAAAGCTATTTGAGGCT									
Ch3866	F: AGCGTTACACACTCCCATCCGT	NED	(CA) ₉	105-112	62	23/4	0.5217	0.4338	1.0000	HQ395768
	R: CTGCACCACTTACCAGGGACG									
Ch5817	F: TTTAAAGCTGGGAAACACAGG	NED	(GT) ₉	165-192	56	23/6	0.5652	0.5778	0.8154	HQ395769
	R: GAAACTGTGACAAACACAGGCT									
Ch8461	F: ACAGAGGGAGTAAGACGCGTG	6-FAM	(CA) ₉	91-106	62	23/6	0.6522	0.7401	0.5667	HQ395770
	R: GGAAGGCTCTGTAGCTGCTGA									
Ch8501	F: ACCAATTGTTCAAAGGGACAC	PET	(CCAT) ₅	156-165	56	23/2	0.0435	0.0435	1.0000	HQ395771
	R: TTTGATATGGAGGCGTGTCT									

¹Reported annealing temperature refers to single locus PCR. Multiplex PCR was performed with annealing at 57 °C.

158 **Table 2. Cross-amplification of the *Chionodraco hamatus* EST-linked microsatellite loci in 2**
 159 **closely related species of the genus *Chionodraco*.** The table reports: locus names, species names,
 160 size range of amplified fragments in bp (s.r.), the (N/N_a) number of individuals assayed/number of
 161 alleles detected and the deviation from Hardy-Weinberg proportions (P_(HW)) for each of the cross-
 162 amplified loci.

Locus	Species					
	<i>Chionodraco myersi</i> ¹			<i>Chionodraco rastrispinosus</i> ²		
	s.r.	N/N _a	P _(HW)	s.r.	N/N _a	P _(HW)
Ch126	266-302	20/4	1.0000	276-292	20/2	1.0000
Ch623	209-218	20/4	0.7654	206-223	20/5	0.0081
Ch684	129-146	20/5	0.5467	125-138	20/3	1.0000
Ch1968	100-104	20/2	1.0000	100-106	20/3	1.0000
Ch2309	207-221	20/7	0.9290	207-219	20/4	1.0000
Ch2788	259-271	20/4	0.4582	260-277	20/6	0.1624
Ch3603	94-105	20/3	0.8182	89-115	20/4	0.4698
Ch3866	104-112	20/5	0.4718	101-112	20/4	0.1156
Ch5817	165-174	20/4	0.0015	151-180	20/5	0.7628
Ch8461	89-113	20/8	0.5565	89-104	20/7	0.3300
Ch8501	156-161	20/2	1.0000	156-169	20/4	1.0000

163 ¹ 5th Italian Expedition, Ross Sea, Terra Nova Bay 1989/1990, PNRA

164 ² ANT-XXIII/8 cruise, Joinville Island 2006/2007, AWI

PAPER III

Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei

Chiara Papetti · Ilaria Anna Maria Marino ·
Cecilia Agostini · Paolo Maria Bisol ·
Tomaso Patarnello · Lorenzo Zane

Received: 5 October 2010 / Accepted: 8 October 2010 / Published online: 29 October 2010
© Springer Science+Business Media B.V. 2010

Abstract We characterized nine polymorphic microsatellites by an enriched library from the Antarctic silverfish *Pleuragramma antarcticum*, a key species in the high Antarctic zone of the Southern Ocean. The number of alleles scored ranged from 7 to 39, whereas the observed and expected heterozygosity ranged from 0.4000 to 0.9750 and from 0.3943 to 0.9782, respectively. Cross-amplification was tested in the 3 species of the genus *Chionodraco* (Channichthyidae). These new microsatellites could potentially be useful in further investigations on *P. antarcticum* for which many questions on population structure and exposition to global environment change are still open to debate.

Keywords *Pleuragramma antarcticum* · Antarctic silverfish · Notothenioidei · Southern Ocean · Microsatellite · FIASCO protocol

The Antarctic silverfish *Pleuragramma antarcticum* Boulenger, 1902 is a key species in the high Antarctic zone of the Southern Ocean. This species has a circumpolar distribution and occurs at depths between 0 and 1,000 m (Eastman 1993). Among notothenioids, *P. antarcticum* is the only species in which all life stages occur throughout

the water column (Regoli et al. 2005; Vacchi et al. 2004). Moreover, *P. antarcticum* represents the most abundant fish in Antarctic waters and a major contribution to the diet of most Antarctic vertebrates such as whales, seals, penguins, flying birds and benthic fish (La Mesa et al. 2004). Unfortunately, this species is experiencing a dramatic decline and has disappeared from a significant part of its range probably as a result of an high sensitivity to the global climate change, especially on the western side of the Antarctic Peninsula, with its warmer water and reduction in sea-ice coverage (Massom and Stammerjohn 2010 and references therein). This work reports on the isolation of new, polymorphic, molecular markers for population genetics analysis of *P. antarcticum*. In particular, these markers will be useful to detect possible loss of genetic variability for this species and further investigate the weak genetic structure, with significant differences between samples both in space and time, that have been found within *P. antarcticum*'s distribution range using the mitochondrial control region (Zane et al. 2006).

Microsatellites were isolated by an enriched library using the FIASCO (Fast Isolation by AFLP of Sequences COntaining repeats) protocol (Zane et al. 2002) as also described in Susana et al. (2007). Genomic DNA was extracted from 10–100 mg of a muscle tissue sample collected in Terranova Bay in 1997 (Ross Sea, 13th Italian expedition, PNRA) and 90% ethanol preserved.

Among the 81 sequenced clones (sequencing facility at <http://www.bmr-genomics.com>), 22 sequences contained a microsatellite with appropriate flanking regions and sequence quality for primer design. Primer pairs for amplification of microsatellite loci were designed with the software OLIGOEXPLORER (<http://www.genelink.com>). Primer pairs were used to amplify genomic DNA from 40 *P. antarcticum* individuals all collected at Terranova Bay

C. Papetti · I. A. M. Marino · C. Agostini ·
P. M. Bisol · L. Zane (✉)
Department of Biology, University of Padova, via U. Bassi 58/B,
35121 Padova, Italy
e-mail: lorenz@bio.unipd.it

T. Patarnello
Department of Public Health, Comparative Pathology
and Veterinary Hygiene, University of Padova,
Agripolis, via Romea 16, 35020 Legnaro (Pd), Italy

Table 1 Amplification conditions and descriptive statistics for 9 novel microsatellite loci of *Pleurogramma antarcticum*

Locus	S. R.	Primers (5'-3')	Repeat motifs	Initial T_a (°C)	Core T_a (°C)	N/N_a	H_O	H_E	PHWE	GB acc. no.
26PI	144–172	F: FAM-CGGATTAACGCCACAGCAAA R: TCGTTTGGATTACTGGTCTGG	(AC) ₅ GTATAG(AC) ₁₃	57	51	40/11	0.5250	0.6582	0.0213	HQ393884
41PI	254–344	F: FAM-AGAATCACGGACCAATAGCC R: TGTGTAGTGGAAATGAAGGG	(CA) ₁₅ TATG(CA) ₃	57	51	40/32	0.9500	0.9620	0.1196	HQ393886
42PI	81–108	F: FAM-TTCGGATA TTCAGGTACAGCC R: GGTCACACTTCTACGGTATGT	(GT) ₆ GCTT(GT) ₃	57	51	40/9	0.7500	0.7025	0.3336	HQ393885
78PI	263–400	F: TAMRA-GAGACTGCGTTGGTTAGACT R: CCTTCCATTGCTTCCGAGT	(CA) ₃₀	56	52	40/38	0.6750	0.9782	0.0004	HQ393887
101PI	166–262	F: TAMRA-AAAGCCAGAGGACAGCAGGAGGA R: CCGTCGTGGAAAGGTGTTGGGA	(GT) ₄ TAT(GT) ₁₀	60	55	40/20	0.6000	0.9237	0.0001	HQ393888
132PI	168–271	F: HEX-ATATTCAGATGGCTGCTTATG R: CGTGTACAGAACAGATTACATATAA	(GA) ₃₇	58	52	40/32	0.9250	0.9703	0.2513	HQ393889
206PI	97–205	F: HEX-TGTTCCAAATCCATGAGTCCAAGC R: TACCAAAACAAAACACTACCCAAACCCA	(GT) ₁₅ G(GT) ₅	60	55	40/38	0.9250	0.9759	0.0620	HQ393890
211PI	213–331	F: HEX-TCATAGTCAGTCAGACATCAGTAT R: CGTCTCACCTCTAATCTACC	(GA) ₄₄	58	52	40/39	0.9750	0.9778	0.4743	HQ393891
221PI	172–194	F: TAMRA-AGAGGTAGGACAAAAGGACAGAT R: GAAAAGGGAAAGCATGATGATGTTGG	(AC) ₅ AT(AC) ₃ AG(AC) ₄	60	55	40/7	0.4000	0.3943	0.6512	HQ393892

S. R. size range of amplified fragments in bp, Initial T_a (°C) annealing temperature of the first cycle, Core T_a (°C) annealing temperature of following thirty cycles, N/N_a , number of individuals assayed/number of alleles detected, H_O and H_E observed and expected heterozygosities; PHWE the probability of Hardy–Weinberg equilibrium; GB acc. no. the GenBank accession number

Table 2 Cross-amplification of the *Pleuragramma antarcticum* microsatellite loci in three additional notothenioid species

Species name	Collection cruise and year	Locus 41PL			Locus 101PL			Locus 221PL		
		S.R.	N_A	PHWE	S.R.	N_A	PHWE	S.R.	N_A	PHWE
<i>Chionodraco hamatus</i>	11th Italian expedition 1995–1996 ^a	289–371	22	0.0070	181–337	23	0.5986	161–166	3	0.0200
<i>Chionodraco myersi</i>	5th Italian expedition 1989–1990 ^b	305–388	24	0.0013	181–312	23	0.0259	161–180	4	0.1734
<i>Chionodraco rastrorpinosus</i>	ANT-XIV 2 Cruise 1996 ^c	318–380	20	0.0002	188–343	27	0.0134	161–177	7	0.7494

^a 11th Italian expedition, Ross Sea, Terranova Bay 1995–1996, PNRA

^b 5th Italian expedition, Ross Sea, Terranova Bay 1989–1990, PNRA

^c ANTARKTIS expedition, PFS “Polarstern” ANT-XIV/2 Cruise 1996/1997 at the Antarctic Peninsula

S. R. size range of amplified fragments in bp, N_A number of alleles detected, PHWE the probability of Hardy–Weinberg equilibrium

in 1997 (Ross Sea, 13th Italian expedition, PNRA). PCRs were carried out in 20 μ L total volume containing: Taq buffer 1X (Promega, 50 mM KCl, 10 mM Tris–HCl pH 9 at 25°C, 0.1% TritonX-100), 1 mM MgCl₂, 150 nM of each primer, 70 μ M dNTPs, 0.04 U/ μ L of Taq polymerase (Promega) and 50 ng of genomic DNA. Loci were amplified with a touchdown PCR profile (for differences between starting and final annealing temperatures, see Table 1): (1) pre-denaturation: 94°C 2 min; (2) 8, 10 or 12 touchdown cycles: denaturation 94°C 30 s, annealing 60–56°C 30 s decreased of 0.5°C each cycle, extension 72°C 30 s; (3) 30 cycles: denaturation 95°C 30 s, annealing 55–51°C 30 s, extension 72°C 30 s; (4) additional extension for 5 min at 72°C.

Forward primers were labeled with different fluorescent dyes and a fraction of the PCR product was loaded on an ABI PRISM 3100 or 3700 automated sequencer (Rox400 as size standard, genotyping facility at <http://www.bmr-genomics.com>) and allele sizes were assigned using GENOTYPER 3.7 (Applied Biosystems). Binning was automated with the software FLEXIBIN ver. 2 (Amos et al. 2007) and all input files for further analysis were produced with CREATE (Coombs et al. 2008).

Among the 22 primer pairs tested, all produced amplified product, nine produced polymorphic microsatellite markers (Table 1), and the remaining were monomorphic or produced bad quality profiles.

The number of alleles scored for the nine polymorphic loci ranged from 7 to 39 and the observed heterozygosity ranged from 0.4000 to 0.9750. Hardy–Weinberg equilibrium (Fisher’s exact test) and genotypic disequilibrium for pairs of loci (Fisher’s exact test) within the sample were tested with the software GENEPOP, online version (Raymond and Rousset 1995) (nominal significant threshold $\alpha = 0.01$). No significant deviation from Hardy–Weinberg equilibrium was detected at seven out of nine loci. A general excess of homozygotes was observed for loci that were found to be in disagreement with the Hardy–Weinberg model. Among the general reasons, this pattern could

be explained by Wahlund effect due to mixing of geographically or temporally differentiated genetic pools or null alleles. The presence of null alleles is a common phenomenon in microsatellites (Dakin and Avise 2004) and their frequency was shown to increase by the transfer of microsatellite loci to new species (Li et al. 2003). Cross species amplification results (see below and Table 2) were not consistent with the presence of null alleles for locus 101PL. All loci resulted in linkage equilibrium at a 1% nominal significance level.

Cross-amplification of primers was evaluated in 3 species of the genus *Chionodraco* (Channichthyidae). DNA was extracted as in Patwary et al. (1994) from 95% ethanol stored muscle tissue of *Chionodraco rastrorpinosus*, *C. hamatus*, and *C. myersi* collected during three different Antarctic cruises and campaigns (Table 2). Of the nine loci tested, three cross-amplified and were found to be polymorphic in 20 individuals of each species (Table 2). The loci represent key markers for further investigations on *P. antarcticum* for which many questions on population structure and exposition to global environment change are still open to debate.

Acknowledgements This work was supported by the Italian Antarctic Research Programme (PNRA) and by National Science Foundation (award number 0741348). We thank Prof. S. Greco for samples collection during the 5th Italian expedition, PNRA, V. Varotto for samples collection during the 11th and 13th Italian expeditions, PNRA and Dr. K.H. Kock and AWI (Alfred-Wegener-Institute für Polar und Meeresforschung) who collected the 1996 sample during the ANT-XIV/2 Polarstern Cruise.

References

- Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Mol Ecol Notes* 7(1):10–14
- Coombs JA, Letcher BH, Nislow KH (2008) Create: a software to create input files from diploid genotypic data for 52 genetic software programs. *Mol Ecol Resour* 8(3):578–580
- Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* 93:504–509

- Eastman JT (1993) Antarctic fish biology. Evolution in a unique environment. Academic Press, San Diego, California, 322 pp
- La Mesa M, Eastman JT, Vacchi M (2004) The role of notothenioid fish in the food web of the Ross Sea shelf waters: a review. *Polar Biol* 27(6):321–338
- Li G, Hubert S, Bucklin K, Ribes V, Hedgecock D (2003) Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. *Mol Ecol Notes* 3:228–232
- Massom RA, Stammerjohn SE (2010) Antarctic sea ice change and variability—physical and ecological implications. *Polar Sci* 4(2):149–186
- Patwary MU, Kenchington EL, Bird CJ, Zouros E (1994) The use of random amplified polymorphic DNA markers in genetic studies of the Sea-Scallop *Placopecten magellanicus* (Gmelin, 1791). *J Shellfish Res* 13(2):547–553
- Raymond M, Rousset F (1995) Genepop (version 1.2): population genetics software for exact tests and ecumenism. *J Hered* 86:248–249
- Regoli F, Nigro M, Benedetti M, Fattorini D, Gorbi S (2005) Antioxidant efficiency in early life stages of the Antarctic silverfish, *Pleuragramma antarcticum*: responsiveness to prooxidant conditions of platelet ice and chemical exposure. *Aquat Toxicol* 75(1):43–52
- Susana E, Papetti C, Barbisan F, Bortolotto E, Buccoli S, Patarnello T, Zane L (2007) Isolation and characterization of eight microsatellite loci in the icefish *Chaenocephalus aceratus* (Perciformes, Notothenioidae, Channichthyidae). *Mol Ecol Notes* 7(5):791–793
- Vacchi M, La Mesa M, Dalu M, MacDonald J (2004) Early life stages in the life cycle of Antarctic silverfish, *Pleuragramma antarcticum* in Terra Nova Bay, Ross Sea. *Antarct Sci* 16(3):299–305
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1–16
- Zane L, Marcato S, Bargelloni L, Bortolotto E (2006) Demographic history and population structure of the Antarctic silverfish *Pleuragramma antarcticum*. *Mol Ecol* 15(14):4499–4511

PAPER IV

1 **Isolation, characterization and multiplexing of sixteen EST-linked microsatellite loci in**
2 **the Antarctic silverfish *Pleuragramma antarcticum***

3 Cecilia Agostini^{1*}, Ilaria A.M. Marino¹, Tomaso Patarnello², Lorenzo Zane¹

4 ¹Department of Biology, University of Padova, Via G. Colombo 3, 35121, Padova, Italy

5 ²Department of Comparative Biomedicine and Food Science, University of Padova,
6 Agripolis, Viale dell'Università 16, 35020, Legnaro (Padova), Italy

7
8 **Keywords:** *Pleuragramma antarcticum*, Antarctic silverfish, Notothenioidei, cross-species
9 amplification, EST-linked microsatellites, multiplex PCR

10 **Running title:** EST-linked microsatellites for *P. antarcticum*

11

12 *** Correspondence to:**

13 Cecilia Agostini

14 Department of Biology, University of Padova, Via G. Colombo 3, 35121, Padova, Italy

15 Tel. +39 049 8276220, Fax: +39 049 8276209

16 Email: ceciliagostini@gmail.com

17

18 **Abstract**

19 Antarctic silverfish, *Pleuragramma antarcticum*, plays a major role in the Antarctic marine
20 ecosystem. Being sea-ice dependent, *P. antarcticum* is endangered by climate change and an
21 appropriate panel of molecular markers is needed to monitor its population structure and
22 genetic variability. A *Chionodraco hamatus* EST database was screened for microsatellite
23 sequences. Primers were designed for 41 loci and 16 polymorphic microsatellites were
24 successfully multiplexed and genotyped in *P. antarcticum* and were in linkage and Hardy-
25 Weinberg equilibrium. Number of alleles per locus ranged from 2 to 16, observed and
26 expected heterozygosity ranged from 0.1500 to 0.8500 and from 0.1449 to 0.9231,
27 respectively.

28

29 *Pleuragramma antarcticum* (Notothenioidei, Nototheniidae), commonly known as Antarctic
30 silverfish, belongs to the perciform suborder Notothenioidei, a group of fish that evolved a
31 suite of specific adaptations to cope with the extreme cold conditions of the Southern Ocean
32 (Cheng & Detrich 2007). *P. antarcticum* is widely distributed in the shelf waters around the
33 Antarctic continent, including the Scotia Arc and adjacent islands, and inhabits both open
34 waters and areas of pack ice at depths from 0 to 900 m (Eastman 1993). Among Antarctic
35 notothenioids, it is the only species in which all developmental stages, from egg to adult, live
36 throughout the water column (DeWitt *et al.* 1990; Vacchi *et al.* 2004; La Mesa & Eastman
37 2012). For its specialization for the pelagic realm and for its overwhelming dominance in the
38 fish fauna, *P. antarcticum* has become a key species in the food web of the Southern Ocean,
39 where it serves as major component of the diet of many Antarctic top predator, such as seals,
40 whales, penguins, marine birds and other fishes (La Mesa & Eastman 2012). The ongoing
41 climate change has raised growing concern about the condition of Antarctic silverfish since
42 both adults and early life stages appear to be strongly sea-ice dependent (Vacchi *et al.* 2004;
43 La Mesa *et al.* 2010; La Mesa & Eastman 2012). Declines in local populations of Antarctic
44 silverfish were reported and might have a cascading effect from lower to higher trophic levels
45 of the Antarctic marine ecosystem (Emslie & Patterson 2007; Moline *et al.* 2008; Quetin &
46 Ross 2009). In this context, genetic markers, such as microsatellites, represent an important
47 tool to assess the pattern of population differentiation among existing stocks of *P. antarcticum*
48 and a valuable strategy to monitor the level of genetic variability within local populations.

49 Here we describe the isolation and characterization of 16 microsatellite loci from high-
50 throughput sequencing data of the species *Chionodraco hamatus* (Notothenioidei,
51 Channichthyidae) and their successful cross-amplification and genotyping in *P. antarcticum*.

52 A total of 23,968 contigs, obtained by 454-sequencing of a normalized cDNA library
53 of *C. hamatus* skeletal muscle (Coppe *et al.* 2013), were searched for SSRs (Short Tandem
54 Repeats) using the software MISA (Thiel *et al.* 2003). The following criteria were used: seven
55 repeats for di-, six for tri-, and five for tetra-, penta-, and hexanucleotide motifs, with a
56 maximal number of bases interrupting 2 SSRs in a compound microsatellite equal to 100. A
57 total of 854 microsatellites were found, including 476 dinucleotide, 324 trinucleotide, 44
58 tetranucleotide, 8 pentanucleotide and 2 hexanucleotide repeats.

59 Forty-one *C. hamatus* microsatellites were selected for primer design and optimization
60 for cross-amplification in *P. antarcticum* in order to set up two multiplex PCR reactions.
61 Starting loci were selected in excess to take into account possible low rate of cross-
62 amplification. Primer pairs were designed using FASTPCR (Kalendar *et al.* 2009) to amplify
63 three different size products (small size: 85-130 bp, medium size: 130-180 bp and large size:
64 180-280 bp). The following parameters were used: primer length from 21 to 22 nucleotides,
65 no secondary structures, optimum annealing temperature at 53-68°C, and GC contents from
66 40 to 70%. Primer pairs were tested for compatibility for future multiplexing. Selected loci
67 were tested on DNA samples from 1 *C. hamatus* individual collected in 1996 at Terranova
68 Bay (11th Italian Expedition, Ross Sea, National Antarctic Research Program, PNRA) and
69 from 20 *P. antarcticum* individuals collected in 2010 at Charcot Island (NBP 10-02 cruise
70 conducted during late March-April 2010 on board the R/V Nathaniel B. Palmer). Samples
71 were preserved in RNA later at -80 °C. Genomic DNA was extracted from 10-100 mg of
72 muscle tissue using a standard salting out protocol (Patwary *et al.* 1994). DNA solutions were
73 stored at -20°C before PCR amplification. Initially, PCR reactions were performed separately
74 for each of the thirty-five loci with the following conditions: 1X Taq buffer (Promega, 50 mM
75 KCl, 10 mM Tris-HCl pH 9 at 25°C, 0.1% TritonX-100), 1 mM MgCl₂, 150 nM of each

76 primer, 70 μ M dNTPs, 0.04 U/ μ L of Taq polymerase (Promega), 100 ng of genomic DNA
77 and water up to 20 μ L. Thermal profile, on a GeneAmp PCR System 2700 Thermocycler
78 (Applied Biosystems), was: (i) pre-denaturation for 2 min at 94°C; (ii) 30 cycles: denaturation
79 for 30 s at 94°C, annealing for 40 s at the primer-specific annealing temperature (Table 1),
80 extension for 50 s at 72°C; (iii) additional extension for 5 min at 72°C. Amplified fragments
81 were visualized in 1.8% agarose gels.

82 Out of the 41 primer sets, 31 gave high quality amplification products of the expected
83 size in *C. hamatus*, while 19 loci successfully cross-amplified in *P. antarcticum*. Moreover, a
84 panel of 12 polymorphic microsatellites isolated from the same *C. hamatus* cDNA library and
85 optimized for amplification in *C. hamatus* (Molecular Ecology Resources Primer
86 Development Consortium *et al.* 2011) were tested in *P. antarcticum*. Five loci gave a
87 successful amplification performance (Ch126, Ch623, Ch2788, Ch3866, Ch5817) and were
88 used to set up multiplex PCRs. In total, 24 out of 43 microsatellites amplified in *C. hamatus*
89 (56%) successfully cross-amplified in *P. antarcticum*. The performance of cross-amplification
90 obtained in this study is appreciable since the two species belong to different families of the
91 Antarctic notothenioid clade, with *P. antarcticum* (Nototheniidae) basal in the phylogeny and
92 *C. hamatus* (Channichthyidae) recently derived (Near *et al.* 2012). The divergence time
93 between the two species was estimated to be approximately 20 million years ago (Near *et al.*
94 2012), while their genetic divergence based on 16S mitochondrial rRNA genes, downloaded
95 from GenBank, is 9.11% over 1713 aligned nucleotides (16S rRNA for *C. hamatus*:
96 gb|AY249473.1; 16S rRNA for *P. antarcticum*: NC_015652.1). Barbara *et al.* (2007),
97 reviewing results of several primer notes reporting cross-species transfer of nuclear
98 microsatellites in a large number of taxa, found that, on average, 33% of microsatellites are
99 successfully amplified between families within order, of which 28% are polymorphic. Reid *et*

100 *al.* (2012) tested the transferability of 22 microsatellite markers among 11 species of sparids
101 and found a mean amplification success for cross-species amplification of 48% with a
102 probability of the microsatellite to be polymorphic of 32%. The authors also found a negative
103 correlation between the transferability of microsatellite loci and the genetic distance
104 calculated at the 16S rRNA gene. For sequence divergences comparable to that found
105 between *P. antarcticum* and *C. hamatus* (between 9 and 9.5%) 48% of tested microsatellites
106 were successfully cross-amplified and 22% were also polymorphic. This comparison confirms
107 the good transferability of *C. hamatus* microsatellites in *P. antarcticum*. Indeed, EST-linked
108 microsatellites, which are localized in transcribed regions, are expected to be more conserved
109 between species and, consequently, to show higher rate of cross-amplification than loci
110 isolated randomly from genomic DNA. In this sense, transcriptome sequencing data, offers a
111 great opportunity to develop molecular markers which can be used in several species, an
112 approach that allow to save both money and time.

113 Forward primers for the 24 microsatellites were labeled with fluorescent dyes for
114 detection with an ABI PRISM automatic sequencer. Loci were amplified together in two
115 multiplex reactions (Table 1). PCRs were carried out in 10 μ L total volume containing: 1X
116 QIAGEN Multiple PCR Master mix (QIAGEN, HotStarTaq DNAPolymerase, Multiplex PCR
117 Buffer, dNTP Mix), 0.2 μ M primer mix and 100 ng of DNA template. PCR conditions were
118 as follows: initial activation step for 15 min at 95°C, 30 cycles of 30 s at 94°C, 90 s at 57°C
119 and 1 min at 72°C, and final elongation for 30 min at 60°C. Alleles were designated according
120 to a LIZ 500 size standard (genotyping facility at <http://www.bmr-genomics.com>). Allele
121 sizes were assigned using PEAK SCANNER version 1.0 (Applied Biosystems). Binning was
122 automated with the software FLEXIBIN version 2 (Amos *et al.* 2007).

123 Eight loci were further discarded: loci Ch10607, Ch15722 and Ch7790 because of
124 difficulties with multiplexing and loci Ch2788, Ch2880, Ch3866, Ch4416 and Ch5817
125 because of genotyping problems. The remaining 16 loci were polymorphic and produced high
126 quality profiles. No evidence of null alleles, stuttering, and large allele dropout was found
127 using MICRO-CHECKER version 2.2.3 (Van Oosterhout *et al.* 2004) and genetic data of all
128 pairs of individuals resulted consistent with an unrelated relationship (1,000 randomizations,
129 5% significant level) using ML-RELATE (Kalinowski *et al.* 2006). Indices of genetic
130 variability, i.e. number of alleles per locus and observed and expected heterozygosity, were
131 calculated using GENETIX version 4.05.2 (Belkhir *et al.* 1996-2004), while GENEPOP
132 online version 4.0.10 (Raymond & Rousset 1995; Rousset 2008) was used to test for Hardy-
133 Weinberg equilibrium and for genotypic linkage equilibrium between pairs of loci within the
134 sample. The number of alleles per locus ranged from 2 to 16 and the observed and expected
135 heterozygosity ranged from 0.1500 to 0.8500 and from 0.1449 to 0.9231, respectively. No
136 significant deviation from Hardy-Weinberg equilibrium was detected and all loci were in
137 linkage equilibrium at a 1% nominal significance level.

138 BLASTN search (BLAST program, version 2.2.26+, made available by the National
139 Center for Biotechnology Information) of *C. hamatus* contigs corresponding to the 16 EST-
140 linked loci against the 9 polymorphic microsatellites, which were already isolated from
141 genomic DNA of the Antarctic silverfish (Papetti *et al.* 2011), showed no significant match.
142 Therefore, the 16 EST-linked loci described in this study, contribute to the development of a
143 useful panel of molecular markers for this species. Microsatellites provide a valuable tool to
144 investigate the genetic structure of a species and to evaluate the level of connectivity among
145 local populations; this task is particularly important for *P. antarcticum*, which is now

146 seriously endangered by current climate change and sea-ice melting (La Mesa & Eastman
147 2012).

148

149 **References**

150

- 151 Amos W, Hoffman JI, Frodsham A, *et al.* (2007) Automated binning of microsatellite alleles:
152 problems and solutions. *Molecular Ecology Notes* **7**, 10-14.
- 153 Barbara T, Palma-Silva C, Paggi GM, *et al.* (2007) Cross-species transfer of nuclear
154 microsatellite markers: potential and limitations. *Molecular Ecology* **16**, 3759-3767.
- 155 Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous
156 Windows TM pour la génétique des populations. *Laboratoire Génome, Populations,*
157 *Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).*
- 158 Cheng C-HC, Detrich HW, III (2007) Molecular ecophysiology of Antarctic notothenioid fishes.
159 *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*
160 **362**, 2215-2232.
- 161 Coppe A, Agostini C, Marino IAM, *et al.* (2013) Genome Evolution in the Cold: Antarctic Icefish
162 Muscle Transcriptome Reveals Selective Duplications Increasing Mitochondrial
163 Function. *Genome biology and evolution* **5**, 45-60.
- 164 DeWitt HH, Heemstra PC, Gon O (1990) Nototheniidae. In: *Fishes of the Southern Ocean* (eds.
165 Gon O, Heemstra PC), pp. 279–331. JLB Smith Institute of Ichthyology, Grahamstown.
- 166 Eastman JT (1993) *Antarctic fish biology: evolution in a unique environment* Academic Press.
- 167 Emslie SD, Patterson WP (2007) Abrupt recent shift in delta 13C and delta 15N values in Adelie
168 penguin eggshell in Antarctica. *Proceedings of the National Academy of Sciences of the*
169 *United States of America* **104**, 11666-11669.
- 170 Kalendar R, Lee D, Schulman AH (2009) FastPCR software for PCR primer and probe design and
171 repeat search. *Genes, Genomes and Genomics* **3**, 1-14.
- 172 Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum
173 likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* **6**, 576-
174 579.
- 175 La Mesa M, Catalano B, Russo A, *et al.* (2010) Influence of environmental conditions on spatial
176 distribution and abundance of early life stages of Antarctic silverfish, *Pleuragramma*
177 *antarcticum* (Nototheniidae), in the Ross Sea. *Antarctic Science* **22**, 243-254.
- 178 La Mesa M, Eastman JT (2012) Antarctic silverfish: life strategies of a key species in the high-
179 Antarctic ecosystem. *Fish and Fisheries* **13**, 241–266.
- 180 Molecular Ecology Resources Primer Development Consortium, Agostini C, Agudelo PA, *et al.*
181 (2011) Permanent Genetic Resources added to Molecular Ecology Resources Database
182 1 October 2010–30 November 2010. *Molecular Ecology Resources* **11**, 418-421.
- 183 Moline MA, Karnovsky NJ, Brown Z, *et al.* (2008) High latitude changes in ice dynamics and
184 their impact on polar marine ecosystems. *Annals of the New York Academy of Sciences*
185 **1134**, 267-319.
- 186 Near TJ, Dornburg A, Kuhn KL, *et al.* (2012) Ancient climate change, antifreeze, and the
187 evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of*
188 *Sciences of the United States of America* **109**, 3434-3439.
- 189 Papetti C, Marino IAM, Agostini C, *et al.* (2011) Characterization of novel microsatellite
190 markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species
191 amplification in other Notothenioidae. *Conservation Genetics Resources* **3**, 259-262.
- 192 Patwary MU, Kenchington EL, Bird CJ, Zouros E (1994) The use of random amplified
193 polymorphic DNA markers in genetic-studies of the sea-scallop *Placopecten*
194 *magellanicus* (Gmelin, 1791). *Journal of Shellfish Research* **13**, 547-553.
- 195 Quetin LB, Ross RM (2009) Life Under Antarctic Pack Ice: A Krill Perspective. In: *Smithsonian at*
196 *the Poles: Contributions to International Polar Year Science* (eds. Krupnik I, Lang MA,
197 Miller SE), pp. 285-298, Smithsonian Institute, Washington DC.

- 198 Raymond M, Rousset F (1995) GENEPOP (Version-1.2) - population-genetics software for exact
199 tests and ecumenicism. *Journal of Heredity* **86**, 248-249.
- 200 Reid K, Hoareau TB, Bloomer P (2012) High-throughput microsatellite marker development in
201 two sparid species and verification of their transferability in the family Sparidae.
202 *Molecular Ecology Resources* **12**, 740-752.
- 203 Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for
204 Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- 205 Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the
206 development and characterization of gene-derived SSR-markers in barley (*Hordeum*
207 *vulgare* L.). *Theoretical and Applied Genetics* **106**, 411-422.
- 208 Vacchi M, La Mesa M, Dalu M, Macdonald J (2004) Early life stages in the life cycle of Antarctic
209 silverfish, *Pleuragramma antarcticum* in Terra Nova Bay, Ross Sea. *Antarctic Science*
210 **16**, 299-305.
- 211 Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for
212 identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology*
213 *Notes* **4**, 535-538.

214

215

216 **Acknowledgements**

217 This work was supported by the Italian Antarctic Research Program (PNRA).

218 We thank V. Varotto for providing *C. hamatus* sample during the 11th Italian

219 Expedition, Ross Sea, PNRA and E. Bortolotto and G. Santovito for providing *P.*

220 *antarcticum* samples during the NBP 10-02 oceanographic campaign, conducted during

221 late March-April 2010 on board the R/V Nathaniel B. Palmer, and coordinated by Prof.

222 Joseph J Torres (College of Marine Science, University of South Florida). This work

223 was supported by the National Program for Antarctic Research (PNRA) to TP and LZ.

224 IAMM is recipient of post-doc grant from the University of Padova (grant number

225 CPDR084151/08). CA is a PhD student in Evolutionary Biology at the University of

226 Padova, with a program partially supported under NSF grant 0741348.

227

228

229

230 **Data Accessibility:**

231 DNA sequences: Genbank accessions KC855717, KC855718, KC855719,
232 KC855720, KC855721, HQ395761, KC855722, KC855723, KC855724, KC855725,
233 KC855726, KC855727, KC855728, KC855729, KC855730, HQ395762.

234

235 **Table 1. Amplification conditions and descriptive statistics for 16 novel**
236 ***Chionodraco hamatus* EST-linked microsatellite loci cross-amplified in**
237 ***Pleuragramma antarcticum*.** The table reports: locus names, repeat motifs (in *C.*
238 *hamatus*), primer pair sequences, fluorescent dye of forward primer, size range of
239 fragments in base pairs cross-amplified in *P. antarcticum* (s.r.), annealing temperature
240 for single locus amplification (^a see note), set of loci in which each locus was grouped
241 for multiplexing, number of individuals assayed/number of alleles detected (N/N_a),
242 observed and expected heterozygosity (H_O/H_E), Hardy-Weinberg Equilibrium
243 probabilities (pHWE), and GenBank accession numbers of each locus (GB acc.no.).

244

Locus	Repeat motif	Primer sequences (5'-3')	Dye	s.r.	T _a (°C)	Loci set for multiplexing	N/Na	H ₀ /H _E	pHWE	GB acc.no.
Ch10105	(CTG) ₁₂	F: TGCCTTGGTTAGGATTAACCGT R: AGAAGTGCTCCATCAAGTCCA	6-FAM	103-111	57	1	20/4	0.3500/0.3500	0.2596	KC855717
Ch10441	(CA) ₁₀	F: GTCCTTACCTGGCAGTGACA R: GCACATCAGTCCAAAATCGGTG	PET	118-129	60	2	20/5	0.7500/0.6231	0.4694	KC855718
Ch10857	(CA) ₉	F: GCTTAAATCACCAATGTGCCA R: TGGACATGAATGTACCAAAAACG	6-FAM	175-188	57	1	20/7	0.7500/0.7218	0.6202	KC855719
Ch11230	(GA) ₇	F: ATCACACAGCTGACTGGGGCT R: TGAGACTCGTCCGGAAATGGA	NED	96-117	60	2	20/6	0.7000/0.7090	0.7440	KC855720
Ch11483	(TAA) ₇	F: ATCAAACTGAACAGCCTGTGT R: TGGAGAACAGGAATGGACAGA	NED	131-150	57	2	20/7	0.5000/0.5679	0.6364	KC855721
Ch126^b	(AAT) ₇	F: CGGTTTTTATGCATGTTGCCA R: ACTGCTCAITTCACACTGGTTC	NED	273-276	56	1	20/2	0.4000/0.4923	0.6389	HQ395761
Ch13222	(TTG) ₆	F: GTCCCATGGTGA CTGATAGGT R: ACGAGTCAATGTAACCCGGAAG	PET	86-92	57	1	20/3	0.1500/0.1449	1.0000	KC855722
Ch17977	(GAT) ₈	F: TTGACTGACAGCTTGGGTGCA R: GAATGAGCCATGTTGAGCCC	6-FAM	190-205	57	2	20/5	0.5000/0.5603	0.5922	KC855723
Ch18085	(TGA) ₆	F: TTTAGGGGTGGCACATTTGGAC R: ACAAACCTTCTCCTTGTCTTCT	VIC	101-125	57	1	20/8	0.8000/0.8128	0.8016	KC855724
Ch19846	(GT) ₁₄	F: CGACCTTTGAGAAAAGGGGCA R: CGACGTGTATCACAAAGGGTCA	6-FAM	122-135	60	2	20/7	0.7000/0.8128	0.1271	KC855725
Ch24332	(GCA) ₈	F: ATCGGTTTTCGTCCTCCACAC R: GGCTGTTTCTGGATCAGCGGT	6-FAM	79-88	60	2	20/3	0.5000/0.6090	0.0264	KC855726
Ch25478	(AC) ₈	F: ACAGTGTGTCTTGGAGAGGT R: GGAGAGAAGTACATGAGGAGGA	PET	162-168	57	2	20/3	0.4500/0.5808	0.0892	KC855727
Ch2931	(TCT)8t(TTC) ₆	F: GGGCTTTCAGGAGCATCGGGA R: ACTTGAACCTGACGTGGCAAC	VIC	89-116	60	2	20/8	0.7000/0.7795	0.3519	KC855728
Ch4796	(GAT) ₈	F: ACAA TGGTTGGTGAGAGCTCC R: TGAGTTAAGCAGAACAAAGTGC	VIC	163-193	57	2	20/11	0.8500/0.8462	0.8281	KC855729
Ch520	(TC) ₇	F: GGAACAACCTTGAGCCAAAGACA R: CCATAAAAATGCATCATTCGGT	6-FAM	263-268	57	1	20/4	0.3500/0.4333	0.5821	KC855730
Ch623^b	(AAC) ₇	F: GCTGTTTGAATCCCTCGTGAGG R: AAAAGTGGTCTCCCGTGCAGT	PET	216-265	62	1	20/16	0.8500/0.9231	0.0707	HQ395762

245 ^a Reported annealing temperature refers to single locus PCR. Multiplex PCR was performed with annealing at 57°C

246 ^b This locus was isolated in Molecular Ecology Resources Primer Development Consortium *et al.* (2011)

PAPER V

Putative selected markers in the *Chionodraco* genus detected by interspecific outlier tests

Cecilia Agostini · Chiara Papetti · Tomaso Patarnello · Felix C. Mark · Lorenzo Zane · Ilaria A. M. Marino

Received: 14 March 2013 / Revised: 20 June 2013 / Accepted: 27 June 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract The identification of loci under selection (outliers) is a major challenge in evolutionary biology, being critical to comprehend evolutionary processes leading to population differentiation and speciation, and for conservation purposes, also in light of recent climate change. However, detection of selected loci can be difficult when populations are weakly differentiated. This is the case of marine fish populations, often characterized by high levels of gene flow and connectivity, and particularly of fish living in the Antarctic marine environment, characterized by a complex and strong circulating system promoting individual dispersal all around the continent. With the final aim of identifying outlier loci putatively under selection in the *Chionodraco* genus, we used 21 microsatellites, including both genomic (Type II) and EST-linked loci (Type I), to investigate the genetic differentiation among the three recently derived *Chionodraco* species that are endemic to the freezing Antarctic waters. Neutrality tests were applied in interspecific comparisons in order to identify candidate loci showing high levels of genetic differentiation, which might reveal imprints of past selection.

Three outlier loci were identified, detecting a higher differentiation between species than did neutral loci. Outliers showed sequence similarity to a calmodulin gene, to an antifreeze glycoprotein/trypsinogen-like protease gene and to nonannotated fish mRNAs. Selective pressures acting on outlier loci identified in this study might reflect past evolutionary processes, which led to species divergence and local adaptation in the *Chionodraco* genus. Used loci will provide a valuable tool for future population genetic studies in Antarctic notothenioids.

Keywords Genome-wide selection scan · Genetic differentiation · Local adaptation · EST-linked microsatellites · Standardized F_{ST} · Antarctic icefish

Introduction

The idea that marine species are genetically homogeneous throughout their range of distribution is traditionally assumed. Large population sizes and wide spatial distributions, associated with an extensive potential for dispersal, in the absence of evident barriers to gene flow, are supposed to limit genetic structuring among local populations (Ward et al. 1994). Typically, genetic studies of natural populations employ neutral molecular markers, like SNPs and microsatellites, which permit to elucidate various aspects of species biology. These loci also allow the estimation of demographic parameters, such as effective population size and migration rate, the inference of which would be biased by the effects of natural selection acting in a locus-specific manner (Avice 1994). The growing interest in understanding the genetic bases of ecologically important traits and in studying local adaptation has, however, gradually shifted the attention to genetic markers

Electronic supplementary material The online version of this article (doi:10.1007/s00300-013-1370-0) contains supplementary material, which is available to authorized users.

C. Agostini · C. Papetti · L. Zane (✉) · I. A. M. Marino
Department of Biology, University of Padova, Padua, Italy
e-mail: lorenzo.zane@unipd.it

T. Patarnello
Department of Comparative Biomedicine and Food Science,
University of Padova, Agripolis, Legnaro, Padua, Italy

F. C. Mark
Integrative Ecophysiology, Alfred Wegener Institute for Polar
and Marine Research, Bremerhaven, Germany

influenced by natural selection, as revealed by the development of several methods aimed at detecting putative selected loci from genome-wide scans (Bowcock et al. 1991; Beaumont and Nichols 1996; Vitalis et al. 2001; Schlötterer 2002; Porter 2003; Beaumont and Balding 2004; Foll and Gaggiotti 2008).

In fish, genome scan studies accomplished so far, in which many individuals were screened using randomly selected molecular markers, reported very few instances of outlier detection; these cases include several members of the Salmonidae family (e.g., Freamo et al. 2011; Seeb et al. 2011), which are well known to be highly structured, some coral reef fishes (e.g., Fauvelot et al. 2007), and a number of other economically important species like Atlantic cod (e.g., Nielsen et al. 2009) and Atlantic herring (e.g. André et al. 2010). Difficulties in finding selected loci in marine fishes may be referred to the limited power of available programs to identify outlier loci when population differentiation is weak. Most of the existing statistical methods for detecting recent signatures of selection from allele frequency data are based on the original idea developed by Lewontin and Krakauer (1973), according to which loci showing a significantly higher or lower level of genetic differentiation than expected under an appropriate neutral population-genetics model can be considered candidates for being influenced by directional (adaptive) or balancing selection, respectively. The level of genetic differentiation is typically described by the F_{ST} index, which summarizes differences in allele frequency distribution between populations. The main difficulty of these methods is to obtain by simulations the expected F_{ST} distribution under neutrality, which should depict the expected variance of F_{ST} values across loci with different levels of genetic variability. From a statistical standpoint, outlier detection is particularly difficult when population structuring is weak and then neutral mean F_{ST} values are low (Beaumont and Balding 2004; Foll and Gaggiotti 2008).

The identification of outlier loci in marine fish is particularly challenging when environmental settings favor gene flow across populations with a consequent homogenizing effect. This is the case of Antarctic waters, where the complex ocean circulating system, which promotes dispersal and connectivity all around Antarctica, contributes to limit genetic differentiation among populations (Matschiner et al. 2009; Papetti et al. 2009, 2012 Damerou et al. 2012).

In this study, we analyzed the genetic differentiation between the three recently derived species of the *Chionodraco* genus (Notothenioidei, Channichthyidae), namely *Chionodraco hamatus*, *Chionodraco myersi*, and *Chionodraco rastrispinosus*, using a panel of 21 microsatellite loci. The three species, which are endemic to the Southern Ocean and whose taxonomical status is based on few

morphological differences and molecular data (Fischer and Hureau 1985; Patarnello et al. 2003), are thought to have diverged between 2 and 1.8 millions of years ago (additional result not reported in Near et al. 2012). Furthermore, we performed an F_{ST} -based survey to search for loci showing a high level of genetic differentiation between the three *Chionodraco* species and thus possibly influenced by selective forces acting during their shallow evolutionary history. To allow future investigations at the population level, we used a panel of loci, including both microsatellites originally isolated from genomic DNA and EST-linked loci, that cross-amplified in all the three species. Cross-amplified molecular markers, such as microsatellites, have proven to be valuable tools for species discrimination and individual assignment to nominal species also in the notothenioid family Trematominae (Van de Putte et al. 2009).

Materials and methods

Sample collection and DNA extraction

Population samples of *C. hamatus*, *C. myersi*, and *C. rastrispinosus* were collected between 1988 and 2007 at four different locations: the Weddell Sea, the Ross Sea (Terra-nova Bay), and the Antarctic Peninsula (Elephant Island and Joinville Island) (Table 1). A small piece of muscle tissue was collected from each specimen and preserved at $-80\text{ }^{\circ}\text{C}$ or in ethanol 90 % until molecular analysis.

Total genomic DNA for each individual was extracted from 10 to 100 mg of muscle tissue following a standard salting out protocol (Patwary et al. 1994). DNA solutions were stored at $-20\text{ }^{\circ}\text{C}$ before PCR amplification.

Genetic and statistical analysis: DNA amplification and genotyping

Twenty-one microsatellite loci were amplified in 108 specimens: 10 loci originally isolated from *C. rastrispinosus*, *Chaenocephalus aceratus*, and *Pleuragramma antarcticum* genomic DNA (Papetti et al. 2006, 2011; Susana et al. 2007), and 11 EST-linked loci isolated from about 24,000 contigs obtained by a high-throughput sequencing of a normalized cDNA library from *C. hamatus* muscle (Molecular Ecology Resources Primer Development Consortium et al. 2011; Coppe et al. 2013). According to O'Brien et al. (1993), microsatellite loci isolated from genomic DNA, which are anonymous genomic sequences, will be called Type II loci, while EST-linked loci, which are found inside or flanking coding gene sequence, will be defined as Type I loci. For the sake of simplicity, final conditions for all the loci used in this study are reported in Table S1 (supplementary materials).

Table 1 Summary of samples analyzed in this study

Species	Population sample	Collection cruise	Sample acronym	Sample size
<i>C. hamatus</i>	Weddell Sea	ANT-VII/4 ^a	ChWS88	9
	Ross Sea (Terranova Bay)	11th Italian expedition PNRA ^b	ChRS95	23
<i>C. myersi</i>	Ross Sea (Terranova Bay)	5th Italian expedition PNRA ^c	CmRS89	27
	Weddell Sea	ANT-XXI/2 ^d	CmWS03	10
<i>C. rastrorpinosus</i>	Elephant Island	ANT-XIV/2 ^e	CrEI96	19
	Joinville Island	ANT-XXIII/8 ^f	CrJI06	20

Reported are as follows: species name, site of collection of each population sample, collection cruise when sampling was performed, sample acronym used in this paper, and sample size

^a ANTARKTIS expedition, RV “Polarstern” ANT-VII/4 (EPOS leg3), Weddell Sea, 1988/1989, AWI

^b 11th Italian expedition, Ross Sea, Terranova Bay, 1995/1996, PNRA

^c 5th Italian expedition, Ross Sea, Terranova Bay, 1989/1990, PNRA

^d ANTARKTIS expedition, RV “Polarstern” ANT-XXI/2 (Bendex), Weddell Sea, 2003/2004, AWI

^e ANTARKTIS expedition, RV “Polarstern” ANT-XIV/2, Antarctic Peninsula, 1996/1997, AWI

^f ANTARKTIS expedition, RV “Polarstern” ANT-XXIII/8, Joinville Island, 2006/2007, AWI

Fragment analysis was performed on an ABI 3730xl automated sequencer, and microsatellite analysis was carried out using PEAK SCANNER version 1.0 (Applied Biosystems). In order to minimize the negative consequences of a poor allele calling, binning was automated with the software FLEXIBIN version 2 (Amos et al. 2007) and the final scoring was then manually checked to ensure the accuracy of the process. MICRO-CHECKER version 2.2.3 (Van Oosterhout et al. 2004) was used to test for null alleles, stuttering and large allele dropout presence, and results for loci showing null alleles were corrected using FREENA (Chapuis and Estoup 2007).

Linkage disequilibrium, Hardy–Weinberg equilibrium (HWE), and genetic diversity

Descriptive analyses, such as number of alleles (N_A), observed heterozygosity (H_{Obs}), and unbiased expected heterozygosity (H_{Exp}), were computed for each locus using GENETIX version 4.05.2 (Belkhir et al. 1996–2004), while allelic richness (A_R), the number of alleles independent of the sample size, was calculated with FSTAT version 2.9.3 (Goudet 2001). The software GENEPOP version online 4.0.10 (Raymond and Rousset 1995; Rousset 2008) was used to test for HWE and for genotypic linkage equilibrium between pairs of loci in each species. Significance of all tests was estimated by the Markov Chain method (demonstration number = 10,000; number of batches = 500; number of iterations per batch = 10,000). When needed, significance threshold ($\alpha = 0.05$) was adjusted using a standard Bonferroni correction for multiple tests (Rice 1989).

Genetic differentiation

Genetic differentiation between species and population samples was quantified by computing pairwise estimators of F_{ST} with FSTAT version 2.9.3 according to Weir and Cockerham (1984). The 95 % CI were estimated by 15,000 bootstrap replicates over loci, and p values were assessed by 1,000 permutation tests in GENETIX version 4.05.2 (Belkhir et al. 1996–2004). Statistical significance level was adjusted, when needed, against type I errors using a standard Bonferroni correction (Rice 1989). The statistic was performed for the whole set of 21 loci, for the same set after the exclusion of outlier loci found by neutrality tests (see below), and separately for Type I and Type II loci, excluding outliers, in order to notice potential differences in the level of detected genetic differentiation. Type I loci, situated in transcribed regions of the genome, are expected to show low levels of variability and to be highly conserved for the presence of functional constraints and of the effects of purifying selection. In contrast, Type II loci are expected to be highly polymorphic because microsatellites are more frequent in noncoding regions of the genome and, as a consequence, they have a higher probability to be neutral markers. Standardized measures of pairwise F_{ST} (F'_{ST}) (independent of the heterozygosity level) were calculated to facilitate the comparison of results obtained by the two sets of loci, which are expected to differ in heterozygosity values and hence in maximum levels of detectable genetic divergence (Hedrick 2005). The calculation was performed by dividing the original F_{ST} estimate by its maximum value, which was obtained using the recoded data file with the program RECODEDATA version 0.1 (Meirmans 2006).

Neutrality tests

To search for the signature of selection, we applied two different neutrality tests relying on the expectation that spatially varying divergent selection should increase the observed level of genetic differentiation between populations/species. Nevertheless, since the two tests are based on different assumptions and parameters, the concordant detection of outlier loci with more than one statistical approach will reduce the number of false positives (Vasemagi et al. 2005). Both tests were applied in all pairwise comparisons between the three species, while they were not used in intraspecific comparisons because of the limited sample size of the individual population samples.

The first method, implemented in the software LOSITAN (Antao et al. 2008), uses coalescent simulations based on the observed data and an island model of migration to generate the expected distribution of F_{ST} versus H_{Exp} with neutral markers. This distribution is then used to identify outlier loci detecting significantly high or low F_{ST} values compared to neutral expectations. These outlier loci are candidates for being subject to natural selection (Antao et al. 2008). In our analysis, 1,000,000 simulations were run assuming a stepwise mutation model (SMM). To better estimate the status of each locus, all pairwise comparisons between species were performed considering firstly Type II loci and then the same set of loci (excluding those eventually resulting under selection) plus one Type I locus at a time. This approach was adopted because the two sets of loci showed different ranges of F_{ST} and H_{Exp} values and because the program failed to simulate the F_{ST} versus H_{Exp} expected distribution when analyzing the two sets together. Moreover, given the higher probability of EST-linked loci to be under selection, considering these loci one by one would not affect the simulation process of the F_{ST} versus H_{Exp} expected distribution in a neutral scenario. In contrast, since microsatellites are more frequent in noncoding regions of the genome, we expected loci isolated from genomic DNA to be neutral markers and, as a consequence, not to affect the simulation process.

The second approach for detecting selection, implemented in the software package DETSEL (Vitalis 2003), simulates a coalescent process of divergence of two populations from a common ancestral population. The program identifies putative outlier loci relying on the population-specific parameters of population divergence, F_1 and F_2 , which are simple functions of the parameters of interest of the model. The expected distributions of F_1 and F_2 were generated maintaining the same number of allelic states as in the observed data and using different combinations of the following nuisance parameters: mutation rate under infinite allele model (IAM) 0.005, 0.001, and 0.0001; ancestral population size 500, 1,000, and 10,000;

population size before the split 50 and 500; time since an assumed bottleneck event 50, 100, and 200 generations; and time since the population split 50 and 100 generations. For each pairwise comparison, 100,000–500,000 coalescent simulations were performed. All loci lying outside the 99 % probability region, simulated in a neutral scenario, were considered candidates for being subject to natural selection. The same approach used in LOSITAN was chosen to improve the coalescent simulation process (all Type II, excluding those eventually resulting under selection, plus one Type I locus at a time).

Results

Microsatellite genetic variability

All 21 microsatellite loci were successfully amplified in the three *Chionodraco* species. One-way analysis of variance (ANOVA) showed no significant differences in H_{Obs} , H_{Exp} , and A_R across sampling locations in each of the three species (p value >0.05). All loci proved to be polymorphic. For the 10 Type II loci, N_A per locus in the three species ranged from 2 (Ca48 in *C. myersi*) to 32 (Cr171 in *C. rastrispinosus*), while N_A for the 11 Type I loci ranged from 2 (Ch126 in *C. hamatus*, Ch1968 in *C. myersi*, and Ch8501 in *C. hamatus* and *C. myersi*) to 11 (Ch8461 in *C. myersi*) (Tables S2 and S3 supplementary materials). Mean N_A for the two sets of loci in the three species (17.40 combining Type II loci and 7.36 considering Type I loci) was significantly different (p value <0.05 using one-way ANOVA) indicating a lower level of allelic variability at Type I loci. A significant decrease in variability was detected at Type I loci also when A_R , the number of alleles independent of the sample size, was considered (Table 2).

At Type II loci, H_{Obs} in the three species ranged from 0.1892 (Ca21 in *C. myersi*) to 0.9444 (Cr127 in *C. myersi*), while H_{Exp} ranged from 0.2118 (Cr38 in *C. rastrispinosus*) to 0.9710 (Cr171 in *C. rastrispinosus*) (Table S2 supplementary materials). A significant excess of homozygotes

Table 2 Mean values and standard deviations for number of alleles (N_A), allelic richness (A_R), observed (H_{Obs}) and unbiased expected heterozygosity (H_{Exp}) at 10 Type II loci and at 11 Type I loci

	10 Type II loci		11 Type I loci		p value (ANOVA)
	Mean	SD	Mean	SD	
N_A	17.40	11.19	7.36	3.04	0.0100
A_R	12.02	7.33	5.55	2.10	0.0110
H_{Obs}	0.5589	0.0200	0.4397	0.0500	<0.0001
H_{Exp}	0.6842	0.0300	0.4653	0.0500	<0.0001

p values resulted from one-way ANOVA are reported

was found at several loci: Cr15 in *C. hamatus* and *C. rastrispinosus*, Cr171 in *C. rastrispinosus*, Cr236 in *C. hamatus*, Ca21 in *C. hamatus* and *C. myersi*, and Ca86 in *C. myersi*. At Type I loci, H_{Obs} in the three species ranged from 0.0313 (Ch126 and Ch8501 in *C. hamatus*) to 0.8649 (Ch8461 in *C. myersi*), while H_{Exp} ranged from 0.0313 (Ch126 and Ch8501 in *C. hamatus*) to 0.8482 (Ch2309 in *C. hamatus*) (Table S3 supplementary materials). A significant excess of homozygotes was found at locus Ch5817 in *C. myersi*. A significant difference was detected, using one-way ANOVA, when comparing mean H_{Obs} and H_{Exp} for the two sets of loci in the three species (p value <0.05) (Table 2). This result confirmed a differential level of genetic variability at Type I loci, which showed significantly lower values of N_A , A_R , H_{Obs} , and H_{Exp} compared to Type II loci, suggesting the presence of functional constraints and of the effects of purifying selection.

Hardy–Weinberg equilibrium probabilities were calculated for each locus and for each species because of the genetic homogeneity observed in population samples within species. A significant departure from HWE was found at 7 out of 30 tests at Type II loci and at 1 out of 33 tests when considering Type I loci. In all the tests showing a departure from HWE, a significant deficit of heterozygotes was found and MICRO-CHECKER version 2.2.3 indicated the presence of null alleles. Subsequent analyses were performed both with the original and the corrected dataset suggested by MICRO-CHECKER. Furthermore, the excluding null alleles (ENA) correction method implemented in FREENA was carried out to correct for the positive bias induced by the presence of null alleles on F_{ST} estimation. Since both the corrections performed yielded comparable results, the original dataset was maintained for successive analyses. Overall, tests for linkage disequilibrium among all loci showed no significant departures from expected values (p value >0.01).

Genetic differentiation

Pairwise estimators of F_{ST} were calculated both between populations of each species and between species, which are characterized by a limited morphological and genetic differentiation (Fischer and Hureau 1985; Patarnello et al. 2003). The analysis was performed for the complete set of 21 loci, for the 18 neutral loci detected by neutrality tests (see below) and separately for Type I and Type II loci excluding the outlier loci Cr38, Ch684, and Ch8501. For all these sets of loci, pairwise F_{ST} between population samples within species were close to zero and not statistically significant (data not shown). For the limited sample size of individual population samples, and for the genetic homogeneity found in intraspecific comparisons, samples of the same species were grouped. Conversely, pairwise F_{ST}

between species were highly significant (p value <0.0001) whatever set of loci was considered (Table 3). When considering all 21 loci, a lower level of differentiation between *C. hamatus* and *C. rastrispinosus* was found; this pattern persisted considering the standardized measure of pairwise F_{ST} (F'_{ST}) and after the removal of the three outlier loci suggested by neutrality tests. In all pairwise comparisons between species, F_{ST} values decreased when outlier loci were excluded from the analysis. Single-locus F_{ST} and F'_{ST} values at outlier loci (Cr38, Ch684, Ch8501) detected a high level of genetic differentiation, specifically in those pairwise comparisons in which they were identified as outlier by neutrality tests (Table 4); in these comparisons, F_{ST} and F'_{ST} estimates clearly exceeded the respective F_{ST} and F'_{ST} 95 % CI between species calculated with 18 putative neutral loci (Table 5).

Neutrality tests

Two different neutrality tests, based on different assumptions and parameters, were applied to detect outlier loci. Simulation results from both tests identified as outlier loci Cr38 and Ch684, in all pairwise comparisons including *C. rastrispinosus*, and locus Ch8501 in all pairwise comparisons including *C. myersi* (p value <0.01). These three loci fell outside the upper 99 % CI of the F_{ST} versus H_{Exp} expected distribution obtained with LOSITAN and lay outside the 99 % probability region of population-specific parameters, F_1 and F_2 , simulated by DETSEL; thus, they are potential candidates for being subject to directional selection (two exemplifying results are shown in Figs. 1, 2).

Discussion

The identification of outlier loci in intraspecific comparisons is very challenging in marine fish because of the low levels of population genetic structuring. This task can be even more difficult when oceanographic settings favor individual dispersal and gene flow among populations; this is the case of the Antarctic marine environment, characterized by a complex system of currents. For these reasons and with the aim of detecting outlier loci putatively under selection in the *Chionodraco* species, we moved from the population level to the species level and investigated the pattern of genetic differentiation among the three *Chionodraco* species by using 21 microsatellite loci. Specifically, we used both EST-linked loci (Type I) and loci randomly isolated from genomic DNA (Type II) because the two classes of genetic markers could reveal different aspects of differentiation processes acting during the evolutionary history of the investigated species.

Table 3 Species pairwise F_{ST} and F'_{ST} values

Sample pairs	21 Loci		18 Neutral loci		9 Neutral Type II loci		9 Neutral Type I loci	
	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}
Ch–Cm	0.2100	0.4996	0.1597	0.4066	0.1863	0.6198	0.1211	0.2501
Ch–Cr	0.1672	0.3821	0.1114	0.2919	0.0914	0.3551	0.1399	0.2765
Cm–Cr	0.2004	0.4645	0.1196	0.3030	0.1028	0.3429	0.1417	0.2893

All pairwise comparisons were performed using four datasets: all 21 loci considered in this study, the 18 neutral loci indicated by the neutrality tests (loci Cr38, Ch684, and Ch8501 were excluded), 9 neutral Type II loci (locus Cr38 was excluded), and 9 Type I loci (loci Ch684 and Ch8501 were excluded). All reported values are statistically significant (p values <0.0001). F_{ST} : actual estimate of population differentiation; F'_{ST} : standardized measure of population divergence. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrorpinosus*

Table 4 Species pairwise F_{ST} and F'_{ST} values at outlier loci indicated by neutrality tests: Cr38, Ch684, and Ch8501

Sample pairs	Cr38		Ch684		Ch8501	
	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}
Ch–Cm	0.3923	0.8538	0.1180	0.2730	0.8378	0.9751
Ch–Cr	0.6338	0.9192	0.6142	0.8536	0.0060	0.0065
Cm–Cr	0.5460	0.9524	0.4324	0.7675	0.7892	0.9617

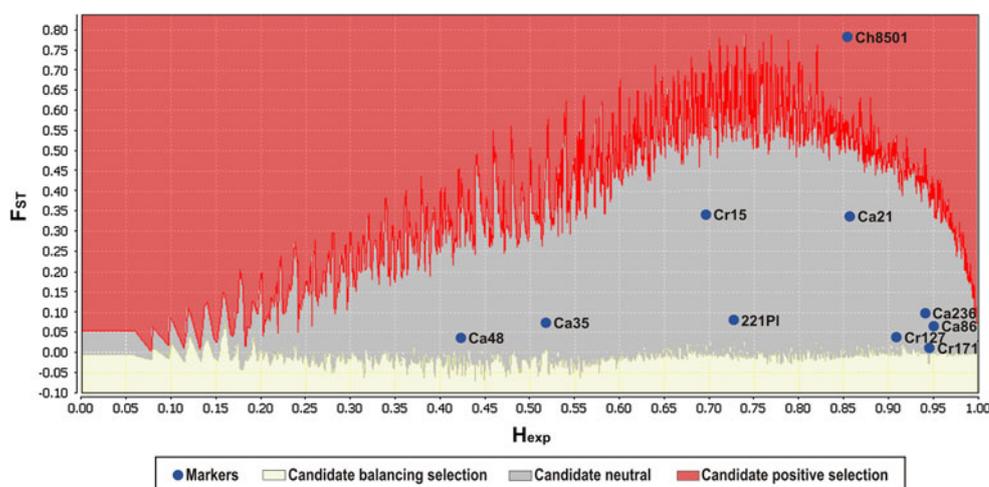
All reported values are statistically significant (p values <0.0001). Pairwise comparisons in which each locus was identified as an outlier are reported in bold. F_{ST} : actual estimate of population differentiation; F'_{ST} : standardized measure of population divergence. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrorpinosus*

Table 5 Species pairwise F_{ST} and F'_{ST} 95 % confidence interval (CI) calculated at 18 putative neutral loci

Sample pairs	F_{ST} 95 % CI	F'_{ST} 95 % CI
Ch–Cm	0.0920–0.2320	0.2342–0.5906
Ch–Cr	0.0600–0.1670	0.1598–0.4375
Cm–Cr	0.0570–0.1920	0.1419–0.4890

F_{ST} : actual estimate of population differentiation; F'_{ST} : standardized measure of population divergence. Ch: *Chionodraco hamatus*, Cm: *Chionodraco myersi*, Cr: *Chionodraco rastrorpinosus*

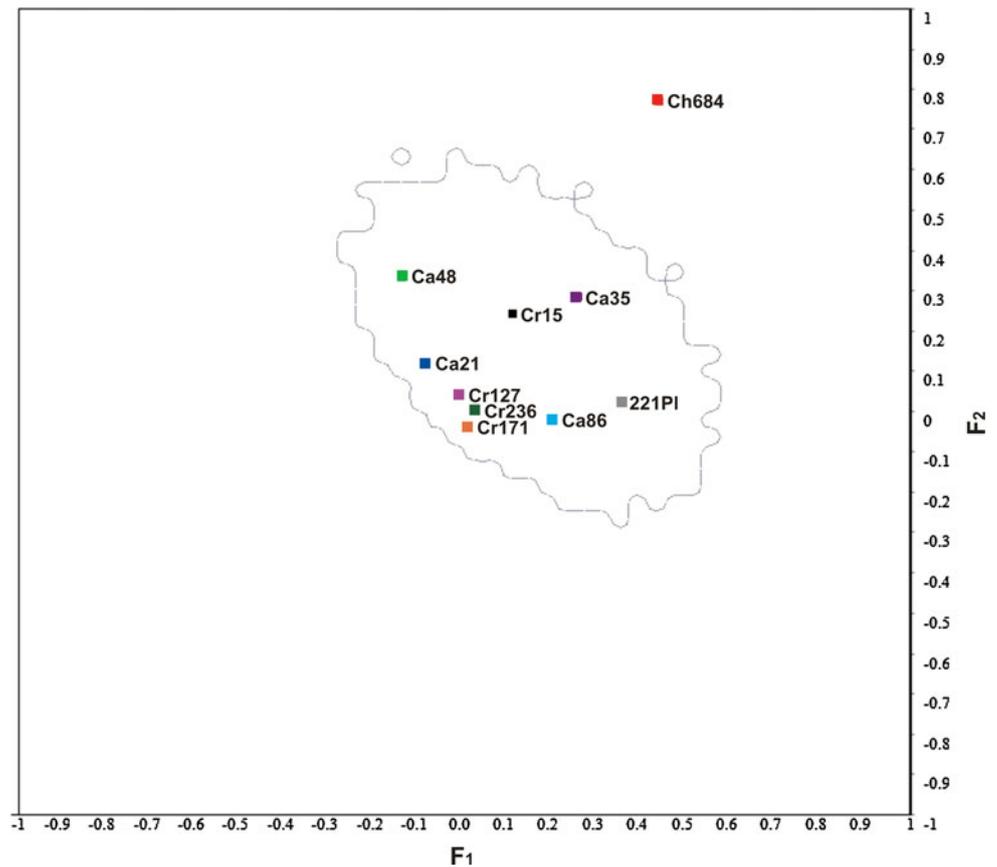
Fig. 1 Result from LOSITAN 1,000,000 simulations under the SMM for the comparison between *C. rastrorpinosus* and *C. myersi* using all Type II loci, except the outlier Cr38, plus Ch8501. Shown in gray (center) is the 99 % CI of the expected F_{ST} versus H_{Exp} (expected heterozygosity) distribution with neutral markers. Loci outside this interval are potential candidates for being subject to directional selection (red, top) or balancing selection (yellow, bottom). (Color figure online)



Type I loci showed a better amplification performance with no detection of null alleles and no departure from HWE (except 1 out of the 33 tests performed), which can be explained by the high conservation of primer annealing regions residing into coding sequences. On the other hand, Type II loci showed some deviations from HWE but a higher level of genetic polymorphism, as indicated by N_A , A_R , H_{Obs} , and H_{Exp} values, which significantly exceeded those calculated with Type I loci. A high level of variability can increase the power of molecular markers to find genetic differences, although this might be obscured by recurrent mutations (O'Reilly et al. 2004). F_{ST} measures are also known to be limited by the average homozygosity within populations (Hedrick 2005). In this study, F_{ST} values calculated at Type II loci exceeded those calculated at Type I loci in all pairwise comparisons between species only when they were standardized following Hedrick (2005). This finding confirms the higher power of neutral genomic loci to detect differentiation and further points out a possible bias in nonstandardized F_{ST} estimates.

Whatever combination of loci was considered, F_{ST} estimates indicated the presence of three distinct gene pools corresponding to the three nominal species, with *C. hamatus* and *C. rastrorpinosus* being more closely related. This result

Fig. 2 Result from DETSEL 500,000 coalescent simulations under the IAM for the comparison between *C. hamatus* and *C. rastrispinosus* using all Type II loci, except the outlier Cr38, plus Ch684. The circumscribed area represents the 99 % probability region of the expected distribution of population-specific parameters of population divergence, F_1 and F_2 , in a neutral scenario. Loci outside this area are potential candidates for being subject to directional selection



is in line with a previous mitochondrial DNA analysis showing that, despite a small genetic differentiation, all the haplotypes of *C. hamatus* and *C. rastrispinosus* clustered together in two sisters monophyletic groups, while *C. myersi* was more distantly related (Patarnello et al. 2003).

Results obtained in this study by neutrality tests suggest that, during this period of time, selective pressures might have played a role in the divergence process of the three species. Indeed, three outlier loci, displaying a higher than expected divergence, were identified: Cr38 and Ch684, in all pairwise comparisons including *C. rastrispinosus*, and Ch8501 in all pairwise comparisons including *C. myersi*. With regard to Ch684 microsatellite, BLAST search (Altschul et al. 1990) showed that it resides at the 3' region of a gene coding for calmodulin, a highly conserved calcium-binding protein that transduces calcium signals and mediates many crucial processes in eukaryotic cells, such as inflammation, metabolism, apoptosis, muscle contraction, intracellular movement, short-term and long-term memory, nerve growth, and the immune response. BLAST similarity search of Ch8501 sequence showed that this locus is instead associated with a *C. hamatus* transcript showing high similarity to other fish mRNA sequences with no functional annotation provided. FULL-LENGTH Web tool (Lara et al. 2007) indicated that the

transcript from which Ch8501 microsatellite was identified has a high probability (92 %) to be a gene coding sequence. Cr38 microsatellite was detected as an outlier besides being a Type II marker. Type II loci were randomly isolated from genomic DNA and we cannot exclude a priori the possibility to draw by chance a microsatellite located in coding regions. In addition, several studies verified the functional relevance of a considerable number of microsatellites residing outside coding gene sequences and thus under the possible influence of natural selection (see Chistiakov et al. 2006 for review). Intriguingly, BLAST search of Cr38 sequence against the NCBI nucleotide database showed a significant alignment, with 96 % sequence identity over 92 nucleotides, to a gene encoding an AFGP/TLP in the Antarctic notothenioid *Dissostichus mawsoni* (Nicodemus-Johnson et al. 2011). Notably, genes coding for notothenioid AFGPs, which are essential molecules for freezing avoidance at subzero temperatures, evolved from a pancreatic TLP, presumably through an ancestral intermediate, that is, a chimeric AFGP/TLP gene (Chen et al. 1997; Cheng and Chen 1999).

Based on the results of neutrality tests between species, and assuming that the higher divergence of outlier loci can be attributed to directional selection, selective pressures could have operated in the branch leading to the divergence

of *C. myersi* for Ch8501 and in the branch leading to the divergence of *C. rastrispinosus* for loci Cr38 and Ch684. Recently, Near et al. (2012) showed how the diversification of Antarctic notothenioids in the Southern Ocean, which was considered a direct consequence of AFGP evolution, was related both to freezing avoidance and to the colonization and adaptation to new ecological niches created by glacial and ice sheet activity during the Late Miocene (11.6–5.3 millions of years ago), thus at least 10 million years after the origin of AFGPs. A time-calibrated Bayesian phylogeny of 83 notothenioid species, performed by the same authors, revealed pulses of lineage diversification occurring in the most derived clades of Antarctic notothenioids, such as the genus *Trematomus* and the family Channichthyidae. Therefore, selective pressures acting on specific loci, as those detected by this study, might reflect past evolutionary processes leading to local adaptation and differentiation of the three *Chionodraco* species. As mentioned before, the divergence of the *Chionodraco* genus was estimated between 2 and 1.8 millions of years ago (additional result not reported in Near et al. 2012), with the separation of *C. myersi* occurring before the divergence of *C. hamatus* and *C. rastrispinosus* (Patarnello et al. 2003). The area of distribution of the three species may give some insights into the ecological cues leading to their diversification. *C. hamatus* and *C. myersi* both display a circum-Antarctic distribution, while *C. rastrispinosus* is found in waters surrounding the Antarctic Peninsula, the South Shetland Islands, and the South Orkney Islands (Kock 1992). Patarnello et al. (2003) hypothesize that the establishment of the Scotia-Weddell confluence, where out flowing Weddell Sea waters converge with the eastward flowing waters of the Scotia Sea, may represent the vicariant event leading to interrupted of gene flow between *C. hamatus* and *C. rastrispinosus*. However, our data suggest that selective pressures might have operated in the past promoting species diversification and possibly leading to local adaptation of the *Chionodraco* species.

Acknowledgments We are grateful to Vittorio Varotto, Dr. Erica Bortolotto, the Italian National Program for Antarctic Research (PNRA), and the Alfred Wegener Institute (AWI) for collecting samples during the Antarctic expeditions. This work has been funded by the Italian National Program for Antarctic Research (PNRA). CA is a PhD student in Evolutionary Biology at the University of Padova, with a program partially funded under National Science Foundation (NSF) Grant 0741348. CP has been funded by a un Senior Research Grant of University of Padova (GRIC110B82). IAMM has been funded by a University of Padova post doc Grant (CPDR084151/08).

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2
- Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Mol Ecol Notes* 7(1):10–14. doi:10.1111/j.1471-8286.2006.01560.x
- André C, Larsson LC, Laikre L, Bekkevoeld D, Brigham J, Carvalho GR, Dahlgren TG, Hutchinson WF, Mariani S, Mudde K, Ruzzante DE, Ryman N (2010) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity* (Edinb) 106(2):270–280. doi:10.1038/hdy.2010.71
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9:323. doi:10.1186/1471-2105-9-323
- Avise JC (1994) Molecular markers, natural history and evolution. Chapman & Hall, New York
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* 13(4):969–980. doi:10.1111/j.1365-294X.2004.02125.x
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond B Biol Sci* 263(1377):1619–1626
- Belkhir K, Borsari P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France)
- Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavallisforza LL (1991) Drift, admixture, and selection in human-evolution—a study with DNA polymorphisms. *Proc Natl Acad Sci USA* 88(3):839–843. doi:10.1073/pnas.88.3.839
- Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24(3):621–631. doi:10.1093/molbev/msl191
- Chen LB, 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(8):3811–3816
- Cheng CHC, Chen LB (1999) Evolution of an antifreeze glycoprotein. *Nature* 401(6752):443–444. doi:10.1038/46721
- Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* 255(1–4):1–29. doi:10.1016/j.aquaculture.2005.11.031
- Coppe A, Agostini C, Marino IAM, Zane L, Bargelloni L, Bortoluzzi S, Patarnello T (2013) Genome evolution in the cold: antarctic Icefish muscle transcriptome reveals selective duplications increasing mitochondrial function. *Genome Biol Evol*. doi:10.1093/gbe/evs108
- Damerou M, Matschiner M, Salzburger W, Hanel R (2012) Comparative population genetics of seven notothenioid fish species reveals high levels of gene flow along ocean currents in the southern Scotia Arc, Antarctica. *Polar Biol* 35(7):1073–1086. doi:10.1007/s00300-012-1155-x
- Fauvelot C, Lemaire C, Planes S, Bonhomme F (2007) Inferring gene flow in coral reef fishes from different molecular markers: which loci to trust? *Heredity* (Edinb) 99(3):331–339. doi:10.1038/sj.hdy.6801005
- Fischer W, Hureau JC (1985) FAO species identification sheets for fishery purposes. In: Fischer W, Hureau JC (eds) Southern ocean: fishing areas 48, 58, and 88, CCAMLR Convention area. Food and Agriculture Organization of the United Nations, Rome, pp 233–470
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180(2):977–993. doi:10.1534/genetics.108.092221

- Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Mol Ecol Resour* 11(Suppl 1):254–267. doi:10.1111/j.1755-0998.2010.02952.x
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995)
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59(8):1633–1638
- Kock KH (1992) Antarctic fish and fisheries. Studies in polar research. Cambridge University Press, Cambridge
- Lara AJ, Pérez-Trabado G, Villalobos DP, Díaz-Moreno S, Cantón FR, Claros MG (2007) A web tool to discover full-length sequences—Full-Lengther. In: Corchado E, Corchado JM, Abraham A (eds) Innovations in hybrid intelligent systems. vol advances in soft computing. Springer, Berlin, pp 361–368. doi:10.1007/978-3-540-74972-1_47
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74(1):175–195
- Matschiner M, Hanel R, Salzburger W (2009) Gene flow by larval dispersal in the Antarctic notothenioid fish *Gobionotothen gibberifrons*. *Mol Ecol* 18(12):2574–2587. doi:10.1111/j.1365-294X.2009.04220.x
- Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60(11):2399–2402. doi:10.1111/j.0014-3820.2006.tb01874.x
- MolecularEcologyResourcesPrimerDevelopmentConsortium, Agostini C, Agudelo PA, BÂ K, Barber PA, Bisol PM, Brouat C, Burgess TI, Calves I, Carrillo Avila M, Chow S, Cordes L, Da Silva D, Dalecky A, De Meester L, Doadrio I, Dobigny G, Duplantier JM, Evison SEF, Ford R, Fresneau D, Galetti PM, Gauthier P, Geldof S, Granjon L, GuÉRin F, Hardy GE, Hernandez Escobar C, Hima K, Hu J, Huang L, Humeau L, Jansen B, Jaquemet S, Jiang Z-Q, Jung S-J, Kim B-S, Kim C-H, Kim J-O, Lai C-H, Laroche J, Lavergne E, Lawton-Rauh A, Le Corre M, Leach MM, Lee J, Leo AE, Lichtenzveig J, Lin LIN, Linde CC, Liu S-F, Marino IAM, McKeown NJ, Nohara K, Oh M-J, Okamoto H, Oliver R, Olivera Angel M, Ornelas-GarcÍA CP, Orsini L, Ostos Alfonso H, Othman AS, Papetti C, Patarnello T, Pedraza-Lara C, Piller KR, Poteaux C, Requier JB, Roziana MK, Sembá Y, Sembene M, Shah RM, Shahril AR, Shao A, Shaw PW, Song L, Souza Ferreira R, Su Y-Q, Suzuki N, Tatar C, Taylor KM, Taylor PWJ, Thiam M, Valbuena R, Wang HE, Yang B-G, Yuan Q, Zajonz U, Zane L, Zhu L, Zhuang Z-M, Zulaiha AR (2011) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2010–30 November 2010. *Mol Ecol Resour* 11(2):418–421. doi:10.1111/j.1755-0998.2010.02970.x
- Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, Zane L, Fernandez DA, Jones CD (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proc Natl Acad Sci USA* 109(9):3434–3439. doi:10.1073/pnas.1115169109
- Nicodemus-Johnson J, Silic S, Ghigliotti L, Pisano E, Cheng CHC (2011) Assembly of the antifreeze glycoprotein/trypsinogen-like protease genomic locus in the Antarctic toothfish *Dissostichus mawsoni* (Norman). *Genomics* 98(3):194–201. doi:10.1016/j.ygeno.2011.06.002
- Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger GL, Ogdén R, Carvalho GR (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evol Biol* 9:276. doi:10.1186/1471-2148-9-276
- O'Brien SJ, Womack JE, Lyons LA, Moore KJ, Jenkins NA, Copeland NG (1993) Anchored reference loci for comparative genome mapping in mammals. *Nat Genet* 3(2):103–112. doi:10.1038/ng0293-103
- O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Mol Ecol* 13(7):1799–1814. doi:10.1111/j.1365-294X.2004.02214.x
- Papetti C, Zane L, Patarnello T (2006) Isolation and characterization of microsatellite loci in the icefish *Chionodraço rastrispinosus* (Perciformes, Notothenioidea, Channichthyidae). *Mol Ecol Notes* 6(1):207–209. doi:10.1111/j.1471-8286.2005.01194.x
- Papetti C, Susana E, Patarnello T, Zane L (2009) Spatial and temporal boundaries to gene flow between *Chaenocephalus aceratus* populations at South Orkney and South Shetlands. *Mar Ecol Prog Ser* 376:269–281. doi:10.3354/meps07831
- Papetti C, Marino IAM, Agostini C, Bisol PM, Patarnello T, Zane L (2011) Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei. *Conserv Genet Resour* 3(2):259–262. doi:10.1007/s12686-010-9336-9
- Papetti C, Pujolar JM, Mezzavilla M, La Mesa M, Rock J, Zane L, Patarnello T (2012) Population genetic structure and gene flow patterns between populations of the Antarctic icefish *Chionodraço rastrispinosus*. *J Biogeogr* 39(7):1361–1372. doi:10.1111/j.1365-2699.2011.02682.x
- Patarnello T, Marcato S, Zane L, Varotto V, Bargelloni L (2003) Phylogeography of the *Chionodraço* genus (Perciformes, Channichthyidae) in the Southern Ocean. *Mol Phylogenet Evol* 28(3):420–429. doi:10.1016/s1055-7903(03)00124-6
- Patwary MU, Kenchington EL, Bird CJ, Zouros E (1994) The use of random amplified polymorphic DNA markers in genetic-studies of the sea-scallop *Placopecten magellanicus* (Gmelin, 1791). *J Shellfish Res* 13(2):547–553
- Porter AH (2003) A test for deviation from island-model population structure. *Mol Ecol* 12(4):903–915. doi:10.1046/j.1365-294X.2003.01783.x
- Raymond M, Rousset F (1995) GENEPOP (Version-1.2)—population-genetics software for exact tests and ecumenicism. *J Hered* 86(3):248–249
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43(1):223–225
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 8(1):103–106. doi:10.1111/j.1471-8286.2007.01931.x
- Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* 160(2):753–763
- Seeb LW, Templin WD, Sato S, Abe S, Warheit K, Park JY, Seeb JE (2011) Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Mol Ecol Resour* 11(Suppl 1):195–217. doi:10.1111/j.1755-0998.2010.02966.x
- Susana E, Papetti C, Barbisan F, Bortolotto E, Buccoli S, Patarnello T, Zane L (2007) Isolation and characterization of eight microsatellite loci in the icefish *Chaenocephalus aceratus* (Perciformes, Notothenioidei, Channichthyidae). *Mol Ecol Notes* 7(5):791–793. doi:10.1111/j.1471-8286.2007.01703.x
- Van de Putte AP, Van Houdt KJK, Maes GE, Janko K, Koubbi P, Rock J, Volckaert FAM (2009) Species identification in the trematode family using nuclear genetic markers. *Polar Biol* 32(12):1731–1741. doi:10.1007/s00300-009-0672-8
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting

- genotyping errors in microsatellite data. *Mol Ecol Notes* 4(3):535–538. doi:[10.1111/j.1471-8286.2004.00684.x](https://doi.org/10.1111/j.1471-8286.2004.00684.x)
- Vasemagi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol Biol Evol* 22(4):1067–1076. doi:[10.1093/molbev/msi093](https://doi.org/10.1093/molbev/msi093)
- Vitalis R (2003) DetSel 1.0: a computer program to detect markers responding to selection. *J Hered* 94(5):429–431. doi:[10.1093/jhered/esg083](https://doi.org/10.1093/jhered/esg083)
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* 158(4):1811–1823
- Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *J Fish Biol* 44(2):213–232. doi:[10.1111/j.1095-8649.1994.tb01200.x](https://doi.org/10.1111/j.1095-8649.1994.tb01200.x)
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370

Supplementary materials

Table S1. Reaction mix and thermal profile optimized for the amplification of the 21 microsatellites in the three *Chionodraco* species. Loci Cr15, Cr38, Cr127, Cr171 and Cr236 were obtained from Papetti et al. (2006); loci Ca21, Ca35, Ca48, Ca86 were obtained from Susana et al. (2007); locus 221Pl was obtained from Papetti et al. (2011); loci Ch126, Ch623, Ch684, Ch1968, Ch2309, Ch2788, Ch3603, Ch3866, Ch5817, Ch8461 and Ch8501 were obtained from MolecularEcologyResourcesPrimerDevelopmentConsortium et al. (2011). Reported are: locus name; sequences of forward (F) and reverse (R) primer; reaction mix; annealing temperature at TD (touchdown) cycles, when TD was used; number of TD cycles, annealing temperature at fixed temperature cycles; number of cycles at fixed temperature.

Locus	Primer sequences (5'-3')	Reaction mix	TD cycles	Number of TD cycles	Fixed temperature cycles	Number of cycles at fixed temperature
Cr15 ^a	F: TAACATGCTCCCGTCTCC R: ATATCATGTGGGCAGAATGGT	A	-	-	59°C	35
Cr38 ^b	F: ACGCCATGCTAATCAGAATC R: GAGTCCCCACACATGACTGT	A	65°C	12	55°C	30
Cr127 ^b	F: CGTATAGGGCCGTACCTCA R: GCTCCATCATAGATCCAGTCA	A	65°C	12	55°C	30
Cr171 ^a	F: CTCGTGGTGGATGATATAAAA R: AAGTCATATGGTAGCACTTTAGC	A	-	-	57°C	35
Cr236 ^b	F: CTGCAGCACAAATATCTGG R: GTACACGTACGGGATTTGAA	A	65°C	12	55°C	30
Ca21 ^b	F: GGACGAAGAAAAGCAAGAAGAACGA R: CTATGTGGGC(AGC)TGTTTGTGATTAT	A	60°C	12	52°C	30
Ca35 ^b	F: CCCAACTCCTCTCAATGTCTG R: TCGGCTGTGTAACCTCCTCCA	A	60°C	12	52°C	30
Ca48 ^b	F: CATTCTCCCTGAGTTTACCAC R: TCTTCCTTTTACCTCTAC(AG)G	A	58°C	16	50°C	30
Ca86 ^b	F: AGTTATTATGGTCGTAATCGCT R: GCTGCTGGT(CT)ACAGGAGTGT	A	58°C	16	50°C	30
221Pl ^c	F: AGAGGTAGGACAAAAAGGACAGAT R: GAAAAGGGAAAAGCATGATGAGTGG	A	60°C	10	55°C	30
Ch126 ^d	F: CGGTTTTTATGCATGTTGCCA R: ACTGCTCATTACACTGGTTC	B	-	-	57°C	30
Ch623 ^d	F: GCTGTTTGATTCCCTCGTGAGG R: AAAAGTGGTCTCCGCTGCAGT	B	-	-	57°C	30
Ch684 ^d	F: CCAACTGAGGGCCCAACAAC R: GGTCTATGGGCTGGCCAATC	B	-	-	57°C	30
Ch1968 ^d	F: AGCTCAAGGTGTTTCGACAGACG R: TAGCCAGCAGCGCTAATCCTG	B	-	-	57°C	30
Ch2309 ^d	F: CAGCTCAATTAAACGCTTTGCA R: CGTCTCAAATGCTGTCACAACC	B	-	-	57°C	30
Ch2788 ^d	F: TGGTTTCGATGAAGAATGCTC R: TGATGATATTGGTCGTGGTCC	B	-	-	57°C	30
Ch3603 ^d	F: TGAGACAAGTCAAATTCCAACC R: GGCATAAAGCTATTTGAGGCT	B	-	-	57°C	30
Ch3866 ^d	F: AGCGTTACACACTCCCATCCGT R: CTGCACCACTCTACCAGGGACG	B	-	-	57°C	30
Ch5817 ^d	F: TTTAAAGCTGGGAAAACACAGG R: GAAACTGTGACAAAACACAGGCT	B	-	-	57°C	30
Ch8461 ^d	F: ACAGAGGGAGTAAGACGCGTG R: GGAAGGCTCTGTAGCTGCTGA	B	-	-	57°C	30
Ch8501 ^d	F: ACCAATTGTTCAAAGGGACAC R: TTGATATGGAGGCGTGTCT	B	-	-	57°C	30

^aA standard thermal profile was used according to the following scheme: (i) predenaturation: 94°C 3 min; (ii)

35 cycles: denaturation 94°C 30 s, annealing 57/59°C 40 s, extension 72°C 50 s; and (iii) final extension for 10 min at 72°C.

^bA touchdown thermal profile was used according to the following scheme: (i) predenaturation: 94°C 3 min; (ii) 12-16 TD cycles: denaturation 94°C 40 min, annealing 65-58°C 30 s decreased of 0.5°C each cycle, extension 72°C 1 min; (iii) 30 cycles: denaturation 94°C 40 s, annealing 55-50°C 30 s, extension 72°C 1 min; and (iv) final extension for 10 min at 72°C.

^cA touchdown thermal profile was used according to the following scheme: (i) predenaturation: 94°C 2 min; (ii) 10 TD cycles: denaturation 94°C 30 s, annealing 60°C 30 s decreased of 0.5°C each cycle, extension 72°C 30 s; (iii) 30 cycles: denaturation 94°C 30 s, annealing 55°C 30 s, extension 72°C 30 s; and (iv) final extension for 10 min at 72°C.

^dA standard multiplex thermal profile for microsatellite loci was used according to the following scheme: (i) initial activation step: 95°C 15 min; (ii) 30 cycles: denaturation 94°C 30 s, annealing 57°C 90 s, extension 72°C 60 s; and (iii) final extension for 30 min at 60°C.

Reaction mix A. PCR amplifications were performed, separately for each locus, in 20 µL total volume containing: Taq buffer 1X (Promega, 50 mM KCl, 10 mM Tris-HCl pH = 9 at 25°C, 0.1% TritonX-100), MgCl₂ 1 mM, 150 nM of each primer, 70 µM dNTPs, 0.04 U/µL of Taq polymerase (Promega) and 50 ng of genomic DNA (except for locus Ca35 for which 100 ng of genomic DNA were used).

Reaction mix B. Multiplex PCR amplifications were performed in 10 µL total volume containing: 1X QIAGEN Multiplex PCR Master mix (QIAGEN, HotStarTaq DNA Polymerase, Multiplex PCR Buffer, dNTP Mix), 0.2 µM primer mix and 100 ng of template DNA.

Table S2. Allelic variability at 10 Type II loci in the three species belonging to the *Chionodraco* genus. The table reports: locus names, number of alleles (N_A), observed heterozygosity (H_{Obs}) / unbiased expected heterozygosity (H_{Exp}) at each locus and species, and probability of Hardy-Weinberg equilibrium (pHWE) at each locus in the three species (population p-values were combined using Fisher's method) following the standard Bonferroni correction (significance threshold $\alpha = 0.05$, comparison number $k = 30$, corrected threshold = 0.0017, significant p-values are reported in **bold**).
Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrispinosus*.

Locus	N_A				H_{Obs}/H_{Exp}			pHWE		
	Ch	Cm	Cr	Tot	Ch	Cm	Cr	Ch	Cm	Cr
Cr15	8	3	6	11	0.3462/ 0.7813	0.2162/ 0.2447	0.3846/ 0.6707	<0.0001	0.5019	0.0007
Cr38	5	7	4	9	0.4063/ 0.4241	0.5676/ 0.6475	0.2308/ 0.2118	0.4391	0.0357	1.0000
Cr127	24	20	23	28	0.8750/ 0.9539	0.9444/ 0.9272	0.8158/ 0.9386	0.0410	0.8838	0.0056
Cr171	25	24	32	41	0.8065/ 0.9577	0.8919/ 0.9563	0.5385/ 0.9710	0.0095	0.3776	<0.0001
Cr236	15	18	17	22	0.6250/ 0.8963	0.7838/ 0.8949	0.7895/ 0.9189	0.0004	0.2135	0.0606
Ca21	12	6	10	17	0.3125/ 0.6419	0.1892/ 0.5902	0.4359/ 0.5245	<0.0001	<0.0001	0.2897
Ca35	6	5	5	7	0.6250/ 0.6260	0.3514/ 0.3458	0.6842/ 0.6042	0.6561	1.0000	0.6096
Ca48	6	2	3	6	0.5938/ 0.6141	0.6216/ 0.4535	0.2821/ 0.3593	0.3679	0.0284	0.1961
Ca86	11	17	20	23	0.5161/ 0.7240	0.3889/ 0.9104	0.8947/ 0.9326	0.0123	<0.0001	0.7468
221PI	3	7	8	10	0.2500/ 0.4787	0.6563/ 0.5913	0.7436/ 0.7349	0.0069	0.2845	0.2822

Table S3. Allelic variability at 11 Type I loci in the three species belonging to the genus *Chionodraco*. The table reports: locus names, number of alleles (N_A), observed heterozygosity (H_{Obs}) / unbiased expected heterozygosity (H_{Exp}) at each locus and species, and probability of Hardy-Weinberg equilibrium (pHWE) at each locus in the three species (population p-values were combined using Fisher's method) following the standard Bonferroni correction (significance threshold $\alpha = 0.05$, comparison number $k = 33$, corrected threshold = 0.0015, significant p-values are reported in **bold**). Symbol “-“ is reported when, in a population sample, only two alleles were detected but one was represented by only one copy. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrispinosus*.

Locus	N_A				H_{Obs}/H_{Exp}			pHWE		
	Ch	Cm	Cr	Tot	Ch	Cm	Cr	Ch	Cm	Cr
Ch126	2	6	3	6	0.0313/ 0.0313	0.4167/ 0.3607	0.0513/ 0.0509	-	1.0000	1.0000
Ch623	6	4	5	7	0.6250/ 0.6959	0.5946/ 0.5879	0.4872/ 0.5711	0.3530	0.7058	0.2619
Ch684	4	5	3	6	0.4375/ 0.4117	0.7027/ 0.7142	0.1795/ 0.1675	0.5162	0.3525	1.0000
Ch1968	4	2	3	4	0.3438/ 0.3070	0.1351/ 0.1277	0.3590/ 0.4172	1.0000	1.0000	0.5780
Ch2309	9	7	5	9	0.8125/ 0.8482	0.6757/ 0.7357	0.3077/ 0.3213	0.6218	0.7666	0.6395
Ch2788	5	4	8	9	0.5625/ 0.5838	0.5294/ 0.6212	0.6667/ 0.7143	0.4907	0.3678	0.4610
Ch3603	3	3	5	5	0.1250/ 0.1751	0.3784/ 0.4091	0.3333/ 0.5132	0.2264	0.8449	0.0036
Ch3866	4	5	5	6	0.5938/ 0.4980	0.4595/ 0.4883	0.5385/ 0.5594	0.5009	0.5156	0.0217
Ch5817	8	5	6	13	0.5938/ 0.5828	0.3529/ 0.6137	0.4615/ 0.5002	0.4713	<0.0001	0.8466
Ch8461	7	11	8	12	0.7188/ 0.7679	0.8649/ 0.8349	0.7949/ 0.7536	0.4690	0.7749	0.8294
Ch8501	2	2	4	4	0.0313/ 0.0313	0.2162/ 0.2369	0.1242/ 0.1282	-	0.5028	1.0000

References

- Molecular Ecology Resources Primer Development Consortium, Agostini C, Agudelo PA, BÂ K, Barber PA, Bisol PM, Brouat C, Burgess TI, Calves I, Carrillo Avila M, Chow S, Cordes L, Da Silva D, Dalecky A, De Meester L, Doadrio I, Dobigny G, Duplantier JM, Evison SEF, Ford R, Fresneau D, Galetti PM, Gauthier P, Geldof S, Granjon L, GuÉRin F, Hardy GE, Hernandez Escobar C, Hima K, Hu J, Huang L, Humeau L, Jansen B, Jaquemet S, Jiang Z-Q, Jung S-J, Kim B-S, Kim C-H, Kim J-O, Lai C-H, Laroche J, Lavergne E, Lawton-Rauh A, Le Corre M, Leach MM, Lee J, Leo AE, Lichtenzveig J, Lin LIN, Linde CC, Liu S-F, Marino IAM, McKeown NJ, Nohara K, Oh M-J, Okamoto H, Oliver R, Olivera Angel M, Ornelas-GarcÍA CP, Orsini L, Ostos Alfonso H, Othman AS, Papetti C, Patarnello T, Pedraza-Lara C, Piller KR, Poteaux C, Requier JB, Roziana MK, Semba Y, Sembene M, Shah RM, Shahril AR, Shao A, Shaw PW, Song L, Souza Ferreira R, Su Y-Q, Suzuki N, Tatarc C, Taylor KM, Taylor PWJ, Thiam M, Valbuena R, Wang HE, Yang B-G, Yuan Q, Zajonz U, Zane L, Zhu L, Zhuang Z-M, Zulaiha AR (2011) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2010–30 November 2010. *Mol Ecol Resour* 11 (2):418-421. doi:10.1111/j.1755-0998.2010.02970.x
- Papetti C, Marino IAM, Agostini C, Bisol PM, Patarnello T, Zane L (2011) Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei. *Conserv Genet Resour* 3 (2):259-262. doi:10.1007/s12686-010-9336-9
- Papetti C, Zane L, Patarnello T (2006) Isolation and characterization of microsatellite loci in the icefish *Chionodraco rastrospinosus* (Perciformes, Notothenioidea, Channichthyidae). *Mol Ecol Notes* 6 (1):207-209. doi:10.1111/j.1471-8286.2005.01194.x
- Susana E, Papetti C, Barbisan F, Bortolotto E, Buccoli S, Patarnello T, Zane L (2007) Isolation and characterization of eight microsatellite loci in the icefish *Chaenocephalus aceratus* (Perciformes, Notothenioidei, Channichthyidae). *Mol Ecol Notes* 7 (5):791-793. doi:10.1111/j.1471-8286.2007.01703.x

PAPER VI

Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation

I. A. M. MARINO,^{*1} A. BENAZZO,^{†1} C. AGOSTINI,^{*} M. MEZZAVILLA,[‡] S. M. HOBAN,[†] T. PATARNELLO,[§] L. ZANE^{*} and G. BERTORELLE[†]

^{*}Department of Biology, University of Padova, via U. Bassi 58/b, 35121 Padova, Italy, [†]Department of Life Sciences and Biotechnology, University of Ferrara, via L. Borsari 46, 44100 Ferrara, Italy, [‡]Institute for Maternal and Child Health - IRCCS Burlo Garofolo and University of Trieste, Via dell'Istria, 65/1, 34137 Trieste, Italy, [§]Department of Comparative Biomedicine and Food Science, University of Padova, viale dell'Università 16, 35020 Agripolis, Legnaro, Padova, Italy

Abstract

Determining the timing, extent and underlying causes of interspecific gene exchange during or following speciation is central to understanding species' evolution. Antarctic notothenioid fish, thanks to the acquisition of antifreeze glycoproteins during Oligocene transition to polar conditions, experienced a spectacular radiation to >100 species during Late Miocene cooling events. The impact of recent glacial cycles on this group is poorly known, but alternating warming and cooling periods may have affected species' distributions, promoted ecological divergence into recurrently opening niches and/or possibly brought allopatric species into contact. Using microsatellite markers and statistical methods including Approximate Bayesian Computation, we investigated genetic differentiation, hybridization and the possible influence of the last glaciation/deglaciation events in three icefish species of the genus *Chionodraco*. Our results provide strong evidence of contemporary and past introgression by showing that: (i) a substantial fraction of contemporary individuals in each species has mixed ancestry, (ii) evolutionary scenarios excluding hybridization or including it only in ancient times have small or zero posterior probabilities, (iii) the data support a scenario of interspecific gene flow associated with the two most recent interglacial periods. Glacial cycles might therefore have had a profound impact on the genetic composition of Antarctic fauna, as newly available shelf areas during the warmer intervals might have favoured secondary contacts and hybridization between diversified groups. If our findings are confirmed in other notothenioids, they offer new perspectives for understanding evolutionary dynamics of Antarctic fish and suggest a need for new predictions on the effects of global warming in this group.

Keywords: Approximate Bayesian Computation, *Chionodraco*, hybridization, interglacials, microsatellites, structure

Received 27 July 2012; revision received 17 July 2013; accepted 18 July 2013

Introduction

Interspecific gene exchange can influence a species' evolutionary history in several ways: delaying initial

species' divergence, favouring homogenization during secondary contacts of divergent gene pools and thus allowing 'reverse speciation' (Taylor *et al.* 2006), providing reinforcement opportunities (Servedio & Noor 2003), or creating new adaptive variation (Abbott *et al.* 2013). Estimating the timing and degree of interspecific hybridization is therefore of general interest to both molecular ecologists and evolutionary biologists as this

Correspondence: Lorenzo Zane, Fax: +39 049 8276209; E-mail: lorenzo.zane@unipd.it

¹These authors contributed equally to this work.

can help determine the factors that regulate the initiation and cessation of gene flow. Until recently, statistical methods assumed constant migration rates across time (Hey & Nielsen 2004), but new developments allow investigating more complex situations, such as more than two species and gene flow only during specific time intervals (Beaumont 2010). In this study, we investigate the occurrence of secondary contacts and possible hybridization associated with climatic cycles, specifically gene exchange during the Pleistocene glacial cycles, which have affected the distribution of many species and ecosystems.

The Southern Ocean is one of Earth's most extreme habitats. It hosts a unique cold-adapted fauna, including a group of 134 fish species in the perciform suborder Notothenioidei (Lecointre 2012). Most notothenioids are endemic to Antarctic waters and dominate high-latitude coastal shelf areas in terms of species abundance and biomass (Eastman & McCune 2000). Notothenioid dominance is probably related to tectonic, climatic and oceanographic events that isolated the Antarctic shelf from other oceans, and to the establishment of cold conditions. The fossil record shows that a cosmopolitan and temperate teleost fish fauna (including herrings, wrasses and billfishes) existed in this area approximately 40 million years ago (MYA; Eastman 2005; Near *et al.* 2012). This fauna is thought to have been extirpated by the transition to polar conditions in the Southern Ocean, starting approximately 33.5 MYA during the Oligocene. This change probably provided ecological opportunities for Antarctic notothenioids (Eastman 2000), which were able to cope with subzero temperatures due to the presence of antifreeze glycoproteins (AFGP; Near *et al.* 2004; Matschiner *et al.* 2011).

The subsequent radiation of notothenioid families probably proceeded through colonization of novel ecological niches along the benthic-pelagic axis. Starting from a benthic ancestor, species diversification was achieved partly by buoyancy adaptations, which allowed feeding in the whole water column. Buoyancy changes occurred via a combination of reduced skeletal mineralization and lipid deposition, leading to occupancy of pelagic, semipelagic, cryopelagic and epibenthic habitats (Eastman 2000, 2005; Eastman & Barrera-Oro 2010). In summary, the diversification of notothenioids in the late Miocene (11.6–5.3 MYA) was probably influenced to some degree by decreased temperature and increased glacial activity and subsequent ecological opportunities (Lautrédou *et al.* 2012; Near *et al.* 2012).

At a more recent timescale, the patterns of species diversification within families and subfamilies are poorly understood. It is plausible that the periodic expansion of ice sheets, producing isolation in some shelf areas and large fluctuations in habitat availability,

favoured extinctions of some groups and allopatric speciation in others (Janko *et al.* 2007; Convey *et al.* 2009). Ice retreats during melting phases, on the other hand, might have promoted ecological divergence in new exploitable niches (Near *et al.* 2012), but also hybridization due to secondary contacts between formerly isolated gene pools. Increased levels of gene flow in interglacial times could underlie weak intraspecific differentiation currently observed in many notothenioids (Matschiner *et al.* 2009; Damerou *et al.* 2012; Papetti *et al.* 2012). Depending on the degree of reproductive isolation previously achieved, interspecific gene exchange may have been nonexistent, limited or extensive, with different impacts on the evolutionary trajectory of a species.

In this study, we investigated patterns of genetic differentiation and hybridization, and the possible influence of the last glaciations, in the recently diverged Antarctic notothenioids of the genus *Chionodraco*. This genus belongs to the Channichthyidae family, which is unique among vertebrates for the loss of erythrocytes and haemoglobin (Ruud 1954; Kock 2005). The Channichthyidae family has 11 genera and 16 species, with most genera being monospecific. One exception is the genus *Chionodraco* for which three species have been described: *Chionodraco hamatus*, *Chionodraco myersi* and *Chionodraco rastrispinosus* (Fischer & Hureau 1985; Eastman & Eakin 2000). The three species have distinct and monophyletic groups of mtDNA lineages (Chen *et al.* 1998; Near *et al.* 2003; Patarnello *et al.* 2003), but show similar morphology (Fischer & Hureau 1985). *C. rastrispinosus* is found only in the waters around the Antarctic Peninsula, the South Shetland Islands and the South Orkney Islands (Kock 1992) at depths from 0 to 1000 m (Iwami & Kock 1990). *C. hamatus* and *C. myersi*, on the contrary, have a wider distribution and have been reported around the entire continent with different depth preferences (4–600 m and 200–800 m, respectively; Gon & Heemstra 1990; Eastman 1993). The pairwise divergence between *Chionodraco* species is similar and, based on mtDNA and nuclear genes estimates, started around 2 MYA (Near *et al.* 2012).

Using a panel of 18 cross-amplifying microsatellite DNA markers and multiple statistical approaches, here we (i) analyse the nuclear genetic structure within and between *Chionodraco* species, (ii) infer the most likely genetic composition of contemporary single individuals and identify individuals with mixed ancestry, (iii) compare evolutionary scenarios that assume different degrees of association between recent glacial cycles and interspecific migration rates, using Approximate Bayesian Computation (ABC). Our main objectives are to determine the extent and timing of gene exchange between recently diverged species and thus to elucidate

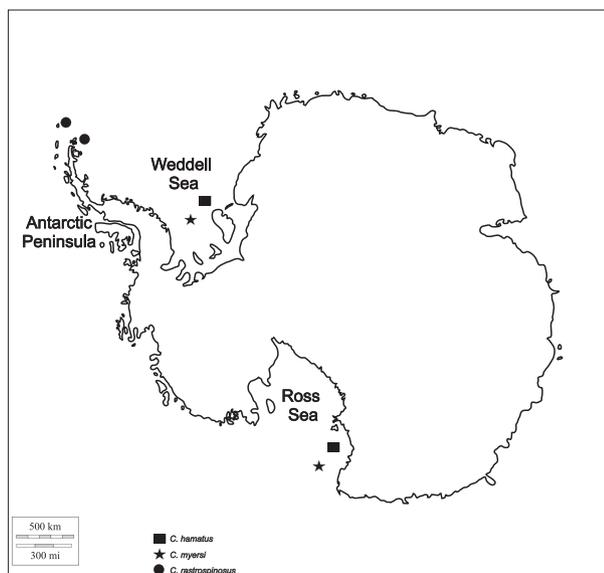


Fig. 1 Sampling locations of *Chionodraco hamatus*, *C. myersi* and *C. rastrospinosus*. Geometric symbols were used for designating the sampling site of each population of the three species.

the role of warm interglacial and cold glacial phases in promoting the divergence and the introgression of species, respectively. Our findings suggest that hybridization occurred during interglacials but did not eliminate species boundaries in the past. However, given current incomplete species isolation, ongoing and rapid global climate warming may promote continued mixing under the current increase in temperatures in the Antarctic sea, with possible evolutionary consequences such as homogenization.

Materials and methods

Sample collection and DNA extraction

Population samples ($N = 108$) of *C. hamatus*, *C. myersi* and *C. rastrospinosus* were collected between 1988 and 2006 at four different locations (Fig. 1, Table 1): the

Weddell Sea, the Ross Sea (Terranova Bay) and the Antarctic Peninsula (Elephant Island and Joinville Island). Based on morphological characters, specimens were assigned to one of the three species following Fischer & Hureau (1985). A small piece of muscle tissue was collected and preserved at $-80\text{ }^{\circ}\text{C}$ or in ethanol 90% for molecular analysis. Total genomic DNA was extracted following a standard salting-out protocol.

DNA amplification and genotyping

Eighteen microsatellite loci were amplified from 108 specimens. Nine loci were isolated from genomic DNA of *C. rastrospinosus*, *Chaenocephalus aceratus* and *Pleura-gramma antarcticum* using a Fast Isolation by AFLP of Sequences Containing Repeats protocol (FIASCO, Zane *et al.* 2002) as described by Papetti *et al.* (2006), Susana *et al.* (2007) and Papetti *et al.* (2011) and were cross-amplified in *C. hamatus*, *C. myersi* and *C. rastrospinosus*. Additionally, nine loci were isolated from about 24 000 contigs obtained by high-throughput sequencing of a normalized cDNA library from *C. hamatus* muscle and amplified as reported in Molecular Ecology Resources Primer Development Consortium *et al.* (2011). Table S1 (Supporting Information) reports details on the microsatellite loci. Fragment analysis was performed at BMR Genomics (www.bmr-genomics.it), and sizing was carried out using PEAK SCANNER, version 1.0 (Applied Biosystems). Binning was automated with the software FLEXIBIN, version 2 (Amos *et al.* 2007), and final scoring was checked to ensure the accuracy of the process. MICRO-CHECKER, version 2.2.3 (Van Oosterhout *et al.* 2004), was used to test for null alleles, stuttering and large allele drop-out presence. The possible presence of outlier markers showing extremely low or high divergence among species was tested using LOSITAN (Antao *et al.* 2008).

Basic diversity and differentiation statistics

The number of alleles (N_A), the observed heterozygosity (H_O) and the unbiased expected heterozygosity (H_E)

Table 1 Summary of samples analysed in this study. Reported are species name, area of collection, cruise when sampling was performed, sample acronym used in this article and sample size

Species	Population sample	Collection cruise	Sample acronym	Sample size
<i>Chionodraco hamatus</i>	Weddell Sea	ANT-VII/4, AWI*	ChWS88	9
	Ross Sea (Terranova Bay)	11th Italian Expedition, PNRA*	ChTB95	23
<i>Chionodraco myersi</i>	Ross Sea (Terranova Bay)	5th Italian Expedition, PNRA	CmRS89	27
	Weddell Sea	ANT-XXI/2, AWI	CmWS03	10
<i>Chionodraco rastrospinosus</i>	Elephant Island	ANT-XIV/2, AWI	CrEI96	19
	Joinville Island	ANT-XXIII/8, AWI	CrJI06	20

*AWI, Alfred Wegener Institute for Polar and Marine Research; PNRA, Italian National Program for Antarctic Research. For more details on cruises, see Agostini *et al.* 2013.

were estimated using GENETIX, version 4.05.2 (Belkiri *et al.* 2005). In addition, to correct for small sample size, allelic richness (A_R) was calculated with FSTAT, version 2.9.3 (Goudet 2001). The software GENEPOP, version online 4.0.10 (Rousset 2008), was used to test for Hardy–Weinberg equilibrium (HWE) and genotypic linkage equilibrium between pairs of loci in each population.

Differentiation within and between species was analysed with an exact test of differentiation using GENEPOP and computing F_{ST} statistics (estimated by f following Weir & Cockerham 1984) using FSTAT. Pairwise F_{ST} were determined with GENETIX, using 10 000 permutations for all comparisons.

Population structure and admixture analysis

The program STRUCTURE 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003; Hubisz *et al.* 2009) was used to estimate the most likely number of genetic clusters ignoring a priori information about species or population assignment of individuals, and to estimate individual admixture proportions. The number of inferred clusters was determined by estimating the probability of the data assuming one to six groups. For each value of K (1–6), five replicates were obtained using 10^6 steps and a burn-in period of 10^5 steps. An admixture ancestry model (i.e. allowing the genetic composition of individuals to be a mixture from different populations) was assumed, with uncorrelated allele frequencies ($\lambda = 1$). The ΔK statistic of Evanno *et al.* (2005) was used to infer the most likely number of genetic clusters in the data set.

Each individual was classified as ‘pure’ or admixed when its largest admixture proportion (q) was higher or lower than an arbitrary but commonly used threshold of 0.95 (Roberts *et al.* 2009). Ancestral polymorphisms or allelic size convergence may increase the rate of false positives in tests designed to identify hybrid individuals, and this threshold may not always be appropriate. The ability of clustering programs to identify hybrids correctly can be assessed by using simulated data sets, based on observed data (Vaha & Primmer 2006; Hoban *et al.* 2012). We performed such an assessment by simulating ‘pure’ individuals using HYBRIDLAB (Nielsen *et al.* 2006). Genotypes of ‘pure’ individuals were generated by randomly drawing alleles from the observed allele frequency distribution for each of the three species. This allowed to simulate a scenario of complete isolation between species and to estimate the expected resultant q values. We created samples of 300 individuals per species, consisting of simulated ‘pure’ *C. hamatus*, *C. myersi* and *C. rastrispinosus* genotypes and analysed them using STRUCTURE with the same settings used for the real data set. The percentage of simulated ‘pure’ individuals with $q < 0.95$ was used as an estimate of the proba-

bility to incorrectly identify a putatively admixed individual by chance given the allelic frequencies observed in the three species. The inclusion of potentially hybrid individuals in this analysis is expected to make this assessment conservative.

Approximate Bayesian Computation

We used the Approximate Bayesian Computation approach (ABC; Beaumont *et al.* 2002), as implemented in the package ABC toolbox (Wegmann *et al.* 2010) and integrated with custom-written scripts, to estimate the posterior probabilities of alternative evolutionary scenarios for the genus *Chionodraco* and to reconstruct the posterior distributions of relevant demographic parameters. ABC is becoming a standard simulation method to analyse complex models when likelihood functions cannot be theoretically specified and, with relatively little additional efforts, it allows to perform power analyses (see reviews of Beaumont 2010; Bertorelle *et al.* 2010 and Csilléry *et al.* 2010, for theoretical and practical aspects). Importantly, the simulation aspect of ABC allows incorporation of particular mutation models, so that possible bias in genetic diversity estimates in the real data, such as introduced by recurrent origins of microsatellite alleles, is also present in the simulated data sets that are compared to real data.

Five models were compared under the ABC approach (Fig. 2). All models assume the same pattern of almost simultaneous species divergence, but exhibit different migration histories. The first split, leading to *C. myersi*, was fixed to 2.0 MYA, whereas the divergence between *C. rastrispinosus* and *C. hamatus* was fixed to 1.8 MYA. These values are the best estimates obtained using a combined mtDNA and nuclear data set of more than 6000 bp (Near *et al.* 2012). The species divergence times were fixed and not estimated in our main analysis in order to reduce the number of parameters and because microsatellites are not highly informative for phylogenetic reconstructions. However, the effect of fixing the species divergence times on the model comparison was tested both considering the support intervals of these times as estimated by Near *et al.* (2012), and also by assuming fixed times up to 30% larger or smaller than the best estimates (more details in Table S2 (Supporting Information), and in Results).

Models M1 and M2 assume isolation and constant migration between species, respectively. Models M3 and M4 assume that migration between species was possible only in the two most recent interglacial periods. M3 allows for migration between 130 000 and 110 000 years ago, a time interval corresponding approximately to the Eemian interglacial (Ivy-Ochs *et al.* 2008), and M4 allows for migration between 10 000 years ago and the present,

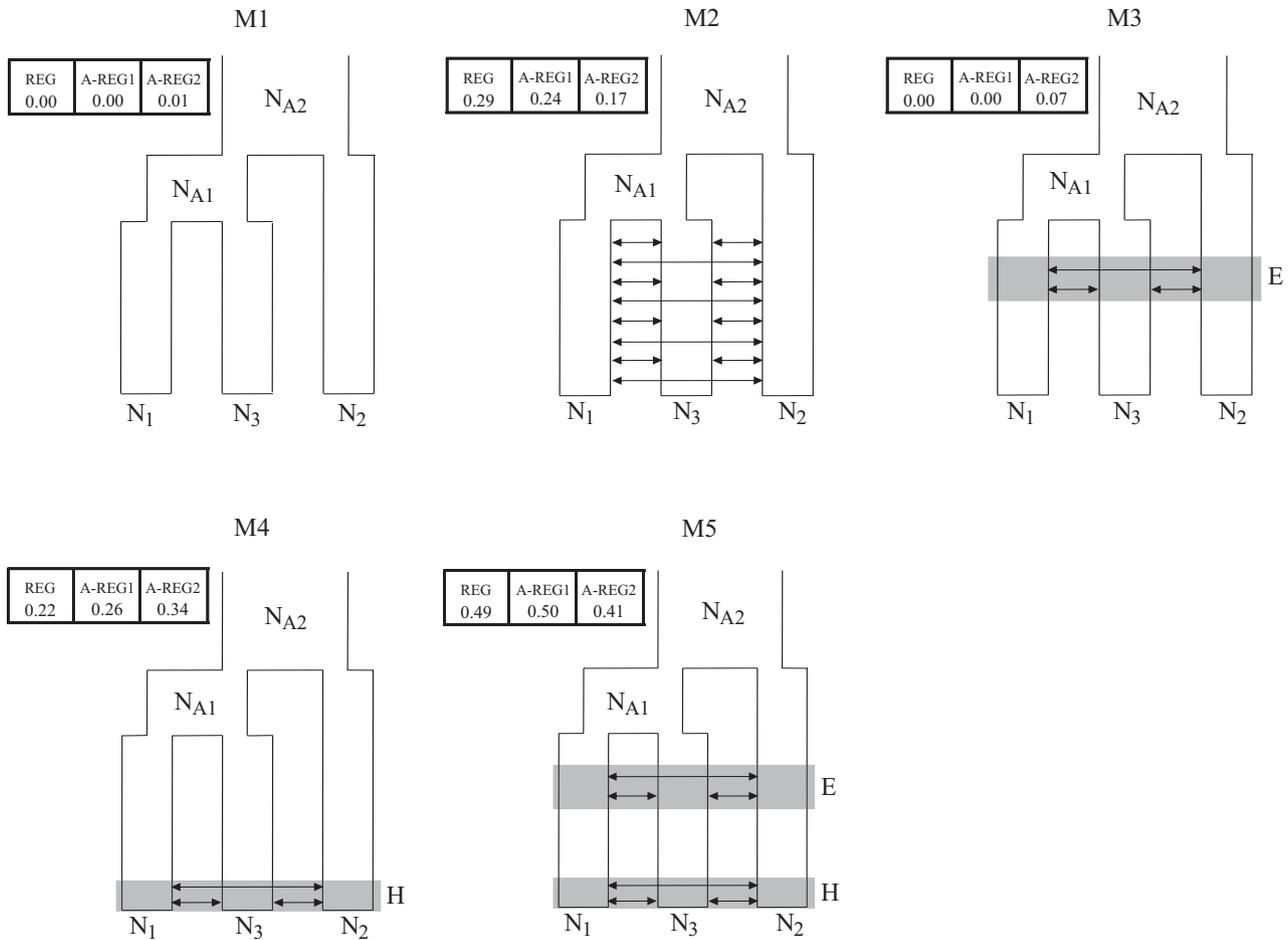


Fig. 2 The five models tested with the ABC approach. For each model, a small table reports the posterior probabilities computed using the regression approach (REG), the adjusted regression approach with bandwidth equal to 0.05 (A-REG1) and the adjusted regression approach with bandwidth equal to 0.3 (A-REG2).

corresponding approximately to the Holocene interglacial (Ivy-Ochs *et al.* 2008). M5 is the combination of M3 and M4. An additional model where migration was allowed at any time excluding the Eemian and the Holocene interglacial was used as control in the model comparison procedure. Figure 2 shows the demographic parameters of these models, that is, population sizes and migration rates. Modern species and ancestral groups had uniform prior distributions of N_e , ranging from 250 to 20 000, whereas migration rates between species had loguniform prior distribution ranging from 0 to 1% per generation (see also Table S3, Supporting Information). Population sizes, when extracted from their prior distributions in each simulation, were assumed to be constant in time. This model of demographic stability is an important assumption, so we investigated the validity of such a model for each species; demographic stability was supported or cannot be excluded by a Bayes factor analysis performed using the software BEAST (Drummond *et al.* 2012) on microsatellite data and a

mtDNA data set (Patarnello *et al.* 2003), respectively. Details of this analysis are given as (Table S4, Supporting Information).

The ABC analysis was based on a subset of eight microsatellite markers: Cr15, Cr127, Cr236, Ca21, Ca35, Ca48, Ca86 and 221PI (Table S1, Supporting Information). EST-linked markers, possibly affected by natural selection at linked genes, and Cr171, which featured an unusual allelic range (possibly affected by a specific mutation process) were discarded to avoid biases in the model choice and parameters estimation analyses. A generalized stepwise mutation model (GSM, Estoup *et al.* 2002) was used in the simulations. The average mutation rate across loci had a normal prior distribution with mean equal to 0.0005 and variance equal to 0.00013. At each locus, the mutation rate had a gamma distribution with shape parameter equal to 2 (Excoffier *et al.* 2005; Neuenschwander *et al.* 2008; Guillemaud *et al.* 2010; see also Table S3, Supporting Information). A generation time of 6 years (Kock & Kellermann 1991)

was used to scale times as needed by the coalescent model.

Summary statistics are used in ABC to compare real and simulated data. We defined a set of 29 summary statistics: the mean and standard deviation across loci of the expected heterozygosity and the allelic range within each species, the pairwise F_{ST} values between species and a set of statistics related to allele-sharing among different species and their variation across loci. Table S5 (Supporting Information) reports details on the summary statistics and their values in the real data.

One million data sets were generated for each model, and the posterior probabilities of each model were estimated using the regression approach (Beaumont 2008), testing different numbers of retained simulations (Table S6, Supporting Information). Following the methods introduced by Fagundes *et al.* (2007), we also estimated the probability that each model is the correct model given the joint observed posterior probabilities of all five models. This latter method depends on the bandwidth used to estimate the density of the combined model probabilities. Veeramah *et al.* (2012) suggest to test bandwidths between 0.05 and 0.3, but we cautiously extended this range to 0.5 as the statistical properties of this approach have not been analysed in detail yet. The posterior distributions of the parameters in the selected model were estimated using the 2000 simulations closest to the observed data and performing a locally weighted linear regression between retained parameters and summary statistics (Beaumont *et al.* 2002).

The adequacy of the selected model to reproduce simulated data sets similar to the real data set is a crucial aspect of the ABC approach and was analysed in four different ways (see Supporting Information for details): (i) computing a principal component analysis to graphically show the position of the real data set when compared to the 10 000 simulations closest to it under each demographic model; (ii) generating by simulation, with parameter values extracted from their posterior distribution under the best model, the posterior distributions of 15 summary statistics not used in the model choice analysis (use of a distinct set of summary statistics is recommended to avoid overfitting, Cornuet *et al.* 2010) and comparing them with the real values; (iii) computing a combined Bayesian P -value (Gelman *et al.* 2004; Voight *et al.* 2005) from the posterior distributions of the summary statistics obtained in point 2; (iv) performing principal component analysis to graphically show the position of the real data set when compared to the data sets generated under the best model as described in point 2.

Finally, we also evaluated the robustness of model selection by estimating the rate of true and false positives for each model. The same ABC procedure applied

to the real data set was applied to pseudo-observed data sets (PODS). For each of the five demographic models (M1–M5), 10 000 PODS were generated by simulation using the same prior distributions of the parameters defined earlier. Each PODS was then analysed as a real data set, estimating the posterior probability of each model using the regression approach. For each model, the rate of true positives was estimated as the proportion of PODS simulated under that model that correctly favours it and the rate of false positives as the fraction of the PODS simulated under a different model (any of the other four) that favours it. Two different decision probability thresholds, 0.7 and 0.5, were tested. A decision probability threshold of 0.5 means, for example, that a model is favoured only if its probability is equal or larger than 0.5. We also computed the rate of true and false positives using a relative threshold, that is, when the selection of the model is based on the probabilities rank.

Results

Outlier detection did not identify putatively selected loci (see Fig. S1–S3, Supporting Information). Only one marker (locus Ch5817) deviated from the simulated 99% confidence interval in one of the three species' comparisons. Given the large number of comparisons (54: 18 loci for three species), this is probably due to chance and basic analyses were therefore performed on the complete data set of 18 microsatellites.

Genetic variation within and between populations and species

All microsatellites were polymorphic, with a minimum of four alleles per locus (Ch1968) and a maximum of 41 (Cr171). Observed and expected heterozygosity ranged from 0.169 (locus Ch126) to 0.853 (locus Cr127) and from 0.152 (locus Ch126) to 0.955 (locus Cr171), respectively. A general excess of homozygotes was found at all loci (Table 2). At the population level, observed and expected heterozygosity ranged from 0.514 in ChTB95 to 0.536 in CrEI96 and from 0.563 in CmWS03 to 0.621 in ChTB95, whereas at species level, ranged from 0.520 in *C. hamatus* to 0.532 in *C. rastrispinosus* and from 0.594 in *C. myersi* to 0.620 in *C. hamatus*, respectively (Table 3). The average allelic richness was similar in all populations, ranging from 4.71 in CmWS03 to 5.20 in CrEI96. Similar allelic richness values were observed in the three species, ranging from 7.58 in *C. myersi* to 8.50 in *C. rastrispinosus* (Table 3).

HWE was tested separately for each of the 18 loci and for each of the six populations. Twenty-one of 108 tests showed a significant departure from HWE, with a

Table 2 *Chionodraco* genus microsatellites. For each locus, we report the overall number of alleles (N_A), the average observed (H_O) and unbiased expected heterozygosity (H_E)

Locus	N_A	H_O	H_E
Cr15	11	0.326	0.560
Cr127	28	0.853	0.929
Cr171	41	0.751	0.955
Cr236	22	0.766	0.912
Ca21	17	0.324	0.589
Ca35	7	0.577	0.541
Ca48	6	0.509	0.460
Ca86	23	0.603	0.838
221P1	10	0.563	0.605
Ch126	6	0.169	0.152
Ch623	7	0.552	0.605
Ch1968	4	0.273	0.282
Ch2309	9	0.604	0.640
Ch2788	9	0.605	0.648
Ch3603	5	0.257	0.336
Ch3866	6	0.523	0.517
Ch5817	13	0.471	0.577
Ch8461	12	0.809	0.786

Table 3 Summary of genetic variability estimates across sampling locations and species. Reported are the average allelic richness among samples (A_R), average observed (H_O) and average unbiased expected heterozygosity (H_E)

Sample	A_R	H_O	H_E
ChWS88	5.021*	0.532	0.613
ChTB95	5.133	0.514	0.621
<i>C. hamatus</i>	8.254 [†]	0.520	0.620
CmRS89	4.919	0.521	0.606
CmWS03	4.713	0.533	0.563
<i>C. myersi</i>	7.578	0.525	0.594
CrEI96	5.197	0.536	0.620
CrJI06	5.184	0.528	0.609
<i>C. rastrispinosus</i>	8.498	0.532	0.614

*For population samples analysis, allelic richness was calculated based on the smallest sample size of eight diploid individuals.

[†]For species analysis, allelic richness was calculated based on the smallest sample size of 26 diploid individuals.

nominal significance threshold of 0.05 (Table S7, Supporting Information). These departures, however, were not clustered by locus or sample. No significant linkage disequilibrium was observed.

Neither the exact test of differentiation nor F_{ST} statistics (corrected for the possible influence of null alleles) found significant differentiation between populations belonging to the same species. All samples from the

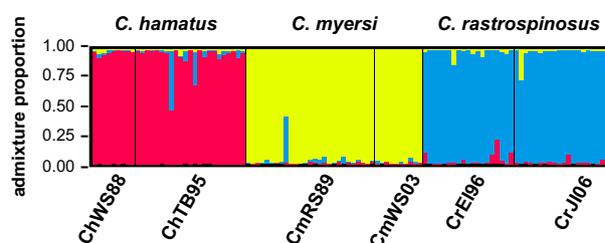


Fig. 3 Population structure estimated in the program STRUCTURE from 108 individuals from six localities using all 18 loci. Each individual is represented by a vertical line, which is partitioned into K coloured segments, the length of each colour being proportional to the estimated membership coefficient from cluster 1 (red, *Chionodraco hamatus*), cluster 2 (yellow, *C. myersi*) and cluster 3 (blue, *C. rastrispinosus*). Black lines separate different populations (labelled below the figure).

same species were thus pooled in subsequent analyses. Genetic differentiation was observed between species, with a highly significant P -value ($P < 0.001$) and a moderate overall F_{ST} value of 0.13 (95% CI = 0.077–0.186, $P < 0.001$). Pairwise F_{ST} values were highly significant, with a slightly larger point estimate (but overlapping confidence intervals) in the comparison *C. hamatus* and *C. myersi* (0.16; CI = 0.093–0.232, $P < 0.001$) compared with the values between *C. hamatus* and *C. rastrispinosus* (0.11; CI = 0.060–0.167, $P < 0.001$) and *C. myersi* and *C. rastrispinosus* (0.12 CI = 0.056–0.193, $P < 0.001$).

Number of genetic clusters and individual admixture

STRUCTURE showed that the most probable number of groups was 3, corresponding to the three species. The likelihood of K increased substantially from 1 to 3, then levelled off for higher K s, and ΔK showed a peak at $K = 3$ (Fig. S4, Supporting Information).

The largest fraction of the genetic background was correctly assigned to the corresponding morphological species. However, 29 individuals of 108 had the highest value of the admixture coefficient (q) below the threshold of 0.95, (Fig. 3, Table 4), and two individuals (one *C. hamatus* and one *C. myersi*) had about 50% of their genetic composition attributed to a different species. Using this threshold, approximately one-third of the *C. hamatus* and *C. rastrispinosus* individuals were classified as admixed to some degree, while one of seven was admixed for *C. myersi*. The average q values estimated in putative hybrid individuals are similar with 11–14% introgression. *C. rastrispinosus* appears to be the more introgressing species.

The admixture proportions in individuals simulated with HYBRIDLAB from the allelic distributions observed in each species indicated that the probability to obtain q values < 0.95 was 0.027. This result suggests

Table 4 Admixture analysis. Lines correspond to single individuals classified as ‘hybrids’, with their estimated admixture coefficient (q). Averages of q values are also reported

Samples	q value		
	Ch	Cm	Cr
ChWSA09	0.925	0.041	0.034
ChWSA10	0.949	0.027	0.024
ChTB163	0.469	0.010	0.521
ChTB165	0.932	0.051	0.016
ChTB167	0.898	0.013	0.089
ChTB171	0.687	0.020	0.293
ChTB173	0.938	0.013	0.049
ChTB176	0.912	0.038	0.050
ChTB177	0.950	0.013	0.037
ChTB193	0.925	0.005	0.071
<i>C. hamatus</i> average	0.859	0.023	0.118
CmRS18	0.013	0.585	0.402
CmRS27	0.015	0.947	0.038
CmRS33	0.012	0.933	0.055
CmRS40	0.008	0.936	0.056
CmWS770	0.019	0.940	0.041
<i>C. myersi</i> average	0.013	0.868	0.118
CrEI03	0.103	0.021	0.876
CrEI10	0.019	0.138	0.843
CrEI12	0.042	0.014	0.944
CrEI14	0.016	0.043	0.942
CrEI16	0.006	0.068	0.926
CrEI18	0.087	0.004	0.909
CrEI19	0.225	0.007	0.768
CrEI20	0.032	0.022	0.946
CrEI22	0.104	0.006	0.890
CrJI02	0.008	0.263	0.729
CrJI03	0.025	0.030	0.945
CrJI12	0.080	0.004	0.915
CrJI19	0.046	0.016	0.938
CrJI20	0.007	0.155	0.838
<i>C. rastrispinosus</i> average	0.057	0.057	0.886

that, on the average, fewer than three individuals in our data set would have been considered as hybrids ($q < 0.95$) incorrectly, if no gene exchange was occurring. Therefore, most of the 29 individuals with $q < 0.95$ in the real data set are likely to be of mixed ancestry.

ABC modelling

The model probabilities computed using the regression method clearly rejected the scenarios without migration (M1) or with migration only in the Eemian interglacial (M3; Fig. 2). The other three models had similar support, but the highest probability was observed for the more complex model where migration was allowed to occur only during both the Eemian and the Holocene interglacial periods (M5). This result was not affected

by the number of retained simulations in the ABC procedure (see Table S6, Supporting Information), by considering the errors in the species divergence times, or by fixing these times to values 20% higher or lower than the best estimates (Table S2, Supporting Information). Only a 30% increase (but not decrease) of the species' divergence times modifies the rank of the models, with M5 becoming the second after M2. M5 was selected as the best model also when pairwise compared with each of the other models (Table S8, Supporting Information). The probability of each model to be the correct one, given the joint observed posterior probabilities of all five models, also supported model M5. M5 had always the highest probability regardless of the bandwidth used to fit the set of model posterior probabilities (see Fig. S5, Supporting Information). Considering two reasonable extremes for the bandwidth (Veeramah *et al.* 2012), models from 1 to 5 had the following set of probabilities: 0, 0.24, 0, 0.26, 0.50 (bandwidth 0.05) and 0.01, 0.17, 0.07, 0.34, 0.41 (bandwidth 0.3). The probability of the control model (constant migration except during the interglacial frames) was always smaller than 0.005 (results not shown).

The adequacy of M5 to generate data sets similar to the real data was confirmed graphically by comparing the posterior distributions of 15 summary statistics (Fig. S6, Supporting Information) to these same summary statistics in the observed data, and by the PCA analysis based on these distributions (Fig. S7, Supporting Information) and on the 10 000 simulations closest to the real data under each model (Fig. S8, Supporting Information). The combined Bayesian P -value used to check the fit between simulations generated under M5, and the real data were not significant ($P = 0.098$).

The analysis of true and false positives (Table 5) in selecting one of the five demographic scenarios, based on PODS generated with known histories, suggested that the summary statistics and the ABC approach that we used have a reasonable power to identify the correct model. True positives were always frequent when the decision threshold was not too high, and false positives were always rare or very rare. Looking in more detail at the results most relevant for the inference in the *Chionodraco* genus, we note that in 40 000 PODS simulated assuming true the models M1, M2, M3 or M4 (10 000 PODS for each model), the model M5 was wrongly chosen as the most likely only in 8% of the cases.

For each parameter in model M5, estimates and credible intervals from the posterior distributions are reported in Table S9 (Supporting Information). Credible intervals were large, and posterior distributions (reported in Fig. S9, Supporting Information) were for some parameters uninformative. We therefore decided to

Table 5 The fraction of true and false positives computed simulating 10,000 PODS (pseudo-observed data sets) under each model and applying to each PODS the same ABC analysis applied to the real data. This analysis is based on the regression approach with 20 000 retained simulations. Each PODS is considered as favouring a specific model if its probability is higher compared with the other probabilities of the other models (relative threshold, or RT), or higher than the indicated threshold

Model	Threshold	True positives	False positives
M1	RT	0.91	0.08
	0.5	0.84	0.07
	0.7	0.55	0.03
M2	RT	0.77	0.04
	0.5	0.73	0.03
	0.7	0.63	0.01
M3	RT	0.72	0.11
	0.5	0.59	0.06
	0.7	0.25	0.01
M4	RT	0.54	0.08
	0.5	0.46	0.06
	0.7	0.11	0.01
M5	RT	0.47	0.08
	0.5	0.38	0.05
	0.7	0.07	0.00

combine the six parameters with the highest correlation with the summary statistics to obtain estimates about the recent migratory pattern. Modern population sizes (N_1 , N_2 and N_3) were combined with Holocene migration rates (m_{H12} , m_{H13} and m_{H23}) in three joint parameters, $m_{Hij}(N_i + N_j)$, corresponding to the number of migrants exchanged between each pair of species per generation (see Table S9, Supporting Information). Examination of such combined parameters is common in ABC, as disentangling N_e and m can be challenging. *C. myersi* and *C. hamatus* showed lower levels of migrants' exchange. Considering the median value of the posterior distributions, the estimated number of migrants per generation was around two individuals (or four chromosomes) per generation. Larger values were obtained in other comparisons and especially for the pair *C. myersi* and *C. rastrispinosus*, where the number of exchanged individuals per generation was around 11. These absolute values should be considered with caution, but the relative pattern of migration exchanges between different pairs of species was confirmed by other indices of the posterior distributions (Table S9, Supporting Information). We note that these estimates assume symmetric gene flow between pairs of species. Considering the results provided by STRUCTURE, where the fraction of hybrids was lower in *C. myersi* but the average contribution of *C. rastrispinosus* into both *C. myersi* and *C. hamatus* was higher than

vice versa, these estimates might be best considered as averages between asymmetric introgression levels.

Discussion

This study provides the first evidence of hybridization among icefish species. Interspecific hybridization had been previously suggested to explain the existence of two highly divergent β -globin pseudogenes in the icefish *Neopagetopsis ionah* (Near *et al.* 2006), but it has not yet been detected in population genetic or phylogenetic studies. Importantly, our genetic data from nuclear microsatellite markers, analysed using individual assignment and intensive simulations of alternative demographic scenarios, support the hypothesis that gene flow among the *Chionodraco* species occurred in recent and probably also in ancient times. Complete reproductive isolation does not seem to have been reached, but also hybridization (especially during interglacial periods) does not seem to have been, as yet, extensive enough to prevent or reverse species diversification. These findings are partially compatible with the mechanism suggested to explain the recent Notothenioid diversification: an increased opportunity for speciation in allopatric refugia during glacial periods, followed by colonization of coastal habitats with empty niches due to eradication of benthic fauna, and further ecological divergence (Near *et al.* 2012). The admixture we observe in the *Chionodraco* data, indicating incomplete isolation, might be low enough to allow adaptation to ecological factors and/or might suggest hybrid genomic incompatibilities to maintain species boundaries. Exploration of adaptive differences and reinforcement mechanisms will require further study.

Genetic structure and temporal pulses of introgression

Our results indicate that *C. hamatus*, *C. myersi* and *C. rastrispinosus* are genetically differentiated and internally homogeneous, with high gene flow among populations of the same species (see also Papetti *et al.* 2012 for *C. rastrispinosus*). However, species' genetic identity is clearly influenced by significant levels of genetic exchange among species. The model-based clustering method implemented in STRUCTURE showed that (i) three major genetically distinct groups are present in our total sample, corresponding to the three morphologically distinct species, and (ii) a large fraction of individuals in each species (14–36%) has mixed genetic composition, with an average contribution of another species into the hybrids estimated between 11% and 14%. Though the threshold for identifying of hybrid individuals is somewhat arbitrary, we showed using simulated genotypes that our estimate is robust. Nota-

bly, estimates seem to also suggest that introgression occurs more frequently from the species with the most restricted distribution into the two other species than vice versa, and that the two latter species, which are widely distributed around the Antarctic continent, show the lowest levels of hybridization.

The existence of contemporary introgression is shown in two putatively F1 hybrids, which had about 50% of their genetic composition attributed to the *C. rastrispinosus* species despite having been classified morphologically as *C. hamatus* and *C. myersi*, respectively. Interestingly, these were collected from the Ross Sea where *C. rastrispinosus* has never been reported. We also observed a sample from the Elephant Island morphologically classified as *C. rastrispinosus*, and with nuclear composition largely composed of this species, but characterized by a *C. hamatus* mtDNA haplotype. *C. hamatus* has never been reported in the Elephant Island area where this sample was collected. The fact that this sample showed a 100% identity with a frequent mtDNA haplotype of *C. hamatus* (AY221999; Patarnello *et al.* 2003) is consistent with the unidirectional hybridization hypothesis of Wirtz (1999), who argued that unidirectional mitochondrial introgression should occur most frequently in cases where a low-density population of one species interacts with a more common species and introgression should proceed from the former into the latter. Regarding our species, hybrid reproduction could arise by the matings of rare *C. hamatus* females with males of *C. rastrispinosus*, common in the Elephant Island area, or may have occurred in the more distant past.

The ABC analysis allowed additional inferences beyond the study of contemporary genetic parameters, and facilitated quantitative estimates about past rates and timing of migration. In our approach, similar to another ABC analysis recently published (Duvaux *et al.* 2011), we modelled pulses of migration at different time points and were able to compare probabilistically the ability of several models to generate simulated data similar to the real data observed in the *Chionodraco* species. Relatively few microsatellite markers were sufficient for the ABC approach to exclude some scenarios with large confidence and discriminate among others on the basis of their relative probabilities. We assumed fixed species' divergence times (parameters for which we had relatively robust estimates), absence of demographic expansions or bottleneck (as suggested by an independent skyline plot analysis based on both microsatellites and mtDNA), and avoided a hyperparameterization of the model by assuming symmetric gene flow among species. The power analysis we performed supports the conclusion that the data were informative in this respect. We also showed that assuming up to 20–30% earlier or later species divergence times, or con-

sidering in the simulations that divergence times can vary using an error distribution function, did not change the main conclusions. Our tests of goodness of fit also lend confidence to our conclusions.

Recent interspecific gene flow (as suggested by the STRUCTURE analysis) is strongly supported by model comparison implemented under the ABC analysis. In fact, all the scenarios excluding recent hybridization have posterior probabilities equal to 0 (or very close to it). This analysis also excludes, though with lower statistical support, the model with gene flow restricted to the last interglacial period (<10 000 years ago), and also the model with constant, uninterrupted gene flow along the evolutionary history of these species after their divergence. Highest support is given to gene flow at two time pulses during the last two major interglacial times. Note that we fixed the hypothetical migration pulses at specific time intervals, that is, the migration frames were not free parameters in the model. However, considering also the results provided by the control model with shifted migration frames, we conclude that interspecific gene flow was likely neither zero nor constant, nor limited to recent times, but was likely high in two specific time intervals (when genetic contacts among these species were probably favoured). It is important to notice that this result does not exclude hybridization in earlier glacial cycles; such a possibility might be examined in future studies with large-scale genetic data sets.

Palaeoclimatic and evolutionary explanations

Notothernioids survived the cooling climate of the Pliocene, which has been recognized as an extrinsic factor promoting differentiation (Ritchie *et al.* 1997), and the freezing and thawing cycles of the Pleistocene, which largely affected Antarctic marine ecosystems. For benthic species, the advance of glaciers on the shelf periodically eradicated local habitats (Janko *et al.* 2007), leaving behind fragmented refuge areas around sub-Antarctic islands or in high-productivity polynyas (Thatje *et al.* 2008). Re-colonization of newly available shelf areas during the warmer intervals, and new niche opportunities, may have been accompanied by ecological diversification (Near *et al.* 2004, 2012), but also favoured secondary contacts and hybridization between differentiated groups not reproductively and/or ecologically isolated. This hypothetical scenario suggested us to compare a demographic model without interspecific gene exchange to a model with constant migration, and also to three additional models with migration pulses during the earlier, later and both interglacial expansion phases. Our results support the model (M5), which assumes association between palaeoclimatic and migration cycles.

Interestingly, STRUCTURE results support a possibly somewhat asymmetric pattern of hybridization in recent times: each species shows a similar fraction of genetic introgression, but *C. myersi* and *C. hamatus* receive this external gene pool mainly from *C. rastrispinosus* and have very limited genetic exchange between each other, whereas *C. rastrispinosus* is introgressed with similar proportions from *C. myersi* and *C. hamatus*. Depth preferences (somewhat different but largely overlapping among the three species; Gon & Heemstra 1990; Iwami & Kock 1990; Eastman 1993) or ecological barriers between *C. myersi* and *C. hamatus*, do not seem able to completely explain this pattern. However, the degree of gene exchange may be related to the distribution ranges. *C. myersi* and *C. hamatus* occur around the entire Antarctic continent, whereas *C. rastrispinosus* is found only near the Antarctic Peninsula, the South Shetland Islands and the South Orkney Islands (Kock 1992). Therefore, the opportunity to hybridize appears higher between *C. myersi* and *C. hamatus* than between these two and *C. rastrispinosus*. Assuming that distributions were similar in the past, and if the three species had similar levels of postmating compatibility, it is possible that *C. myersi* and *C. hamatus* developed a higher level of premating isolation due to a higher chance of reinforcement (Noor 1999) to operate. Further, under this scenario, secondary contact between these two species and *C. rastrispinosus* during interglacial times would have been more limited and reinforcement less efficient. As a consequence, present-day interspecific pairs involving *C. rastrispinosus* would be more frequent. This is a hypothesis for future investigations with larger samples and additional genetic markers, but is consistent with our data.

In our collections, hybridization is relatively common (one of four individuals classified as hybrids at the 95% threshold and out of twelve at a 90% threshold), suggesting that intrinsic hybrid incompatibilities, if any, are not severe. Divergent ecological selection (Seehausen 2006) has been suggested to have played a role in Notothenioid radiation, with ecological opportunity (Schluter 2000) promoting diversification (Eastman 2005). The role of ecological selection is supported by parallel dietary niche evolution in multiple families (Rutschmann *et al.* 2011) and by increasing disparity of buoyancy through time, indicating repeated colonization of benthic, epibenthic, semipelagic and pelagic habitats among closely related lineages (Near *et al.* 2012). Considering this, our observed lack of intrinsic isolation suggests that the final stage of species diversification (*sensu* Streelman & Danley 2003) has not yet been reached. Therefore, altered environmental and ecological conditions promoting further gene exchange could reduce species biodiversity in this group.

In conclusion, this study reports the first evidence of interspecific hybridization among icefish species and suggests that introgression primarily occurred during warmer glacial phases. The homogenizing effect of interglacials may be one of the factors explaining the large fraction of monotypic genera observed in Channichthyids. Future research steps to help understand the role of hybridization and speciation in these fish might include a careful re-examination of the congruence between phylogenies inferred with different nuclear and mitochondrial markers, studies of reproductive isolation between taxa, assessment of geographic patterns and possible asymmetry of hybridization and its correlation with ecological factors and temporal divergence (Nosil 2012). Genomic data will probably assist in understanding the extent of introgression among different species and, especially if polymorphism associated to relevant traits will be found, in elucidating the potential of introgression in providing adaptive variation. Finally, predictive simulations of hybridization may help assess the likelihood of 'reverse speciation' under changing environmental conditions, as further warming and hybridization may limit the differentiation and the development of premating barriers, possibly resulting in homogenization and the loss of biodiversity from a spectacular past evolutionary radiation.

Acknowledgements

This work was supported by the Italian National Program for Antarctic Research (PNRA). We are grateful to Vittorio Varotto, Erica Bortolotto, Felix C. Mark (Integrative Ecophysiology, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany) and AWI for collecting samples during Antarctic expeditions. IAMM has been funded by a University of Padova post doc Grant (CPDR084151/08). CA was partially supported by the NSF Grant 0741348. SMH was supported by the University of Ferrara and the EU project 'ConGRESS'.

References

- Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.
- Agostini C, Papetti C, Patarnello T, Mark FC, Zane L, Marino IAM (2013) Putative selected markers in the *Chionodraco* genus detected by interspecific outlier tests. *Polar Biology*, doi:10.1007/s00300-013-1370-0.
- Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, **7**, 10–14.
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.
- Beaumont M (2008) Joint determination of topology, divergence time and immigration in population trees. In: *Simula-*

- tions, Genetics, and Human Prehistory (eds Matsumura S, Forster P, Renfrew C), pp. 135–154. McDonald Institute for Archaeological Research, Cambridge, UK.
- Beaumont M (2010) Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution and Systematics*, **41**, 379–406.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Belkir K, Borsa P, Goudet J, Bonhomme F (2005) GENETIX v. 4.05, logiciel sous Windows pour la génétique des populations. Laboratoire Génome et Populations. CNRS UPR 9060, Université Montpellier II.
- Bortorelle G, Benazzo A, Mona S (2010) ABC as a flexible frame work to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.
- Chen WJ, Bonillo C, Lecomte G (1998) Phylogeny of the Channichthyidae (Notothenioidei, Teleostei) based on two mitochondrial genes. In: *Fishes of Antarctica: A Biological Overview* (eds di Prisco G, Pisano E, Clarke A), pp. 287–298. Springer, Berlin.
- Convey P, Stevens MI, Hodgson DA *et al.* (2009) Exploring biological constraints on the glacial history of Antarctica. *Quaternary Science Reviews*, **28**, 3035–3048.
- Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, **25**, 410–418.
- Damerou M, Matschiner M, Salzburger W, Hanel R (2012) Comparative population genetics of seven notothenioid fish species reveals high levels of gene flow along ocean currents in the southern Scotia Arc, Antarctica. *Polar Biology*, **35**, 1073–1086.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.
- Duvaux L, Belkir K, Boulesteix M, Boursot P (2011) Isolation and gene flow: inferring the speciation history of European house mice. *Molecular Ecology*, **20**, 5248–5264.
- Eastman JT (1993) *Antarctic Fish Biology. Evolution in a Unique Environment*. Academic Press, San Diego.
- Eastman JT (2000) Antarctic notothenioid fishes as subjects for research in evolutionary biology. *Antarctic Science*, **12**, 276–287.
- Eastman JT (2005) The nature of the diversity of Antarctic fishes. *Polar Biology*, **28**, 93–107.
- Eastman JT, Barrera-Oro E (2010) Buoyancy studies of three morphs of the Antarctic fish *Trematomus newnesi* (Nototheniidae) from the South Shetland Islands. *Polar Biology*, **33**, 823–831.
- Eastman JT, Eakin RR (2000) An updated species list for Notothenioid fish (Perciformes; Notothenioidei), with comments on Antarctic species. *Archive of Fishery and Marine Research*, **48**, 11–20.
- Eastman JT, McCune AR (2000) Fishes on the Antarctic continental shelf: evolution of a marine species flock? *Journal of Fish Biology*, **57** (Suppl. A), 84–102.
- Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Ecology*, **14**, 2611–2620.
- Excoffier L, Estoup A, Cornuet JM (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, **169**, 1727–1738.
- Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17614–17619.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- Fischer W, Hureau JC (1985) FAO species identification sheets for fisheries purposes. In: *Southern Ocean (Fishing Areas 48, 58 and 88) (CCAMLR Convention Area)*, Vol. 2 (eds Fischer W, Hureau JC), pp. 233–470. Food and Agriculture Organization of the United Nations, Rome.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian Data Analysis*. Chapman and Hall/CRC Press, Boca Raton, Florida.
- Gon O, Heemstra PC (1990) *Fishes of the Southern Ocean*. JLB Smith Institute of Ichthyology, Grahamstown, South Africa.
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at: www2.unil.ch/popgen/softwares/fstat.html
- Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A (2010) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, **104**, 88–99.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.
- Hoban SM, McCleary TS, Schlarbaum SE, Anagnostakis SL, Romero-Severson J (2012) Human-impacted landscapes facilitate hybridization between a native and an introduced tree. *Evolutionary Applications*, **5**, 720–731.
- Hubisz M, Falush D, Stephens M, Pritchard J (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.
- Ivy-Ochs S, Kerschner H, Reuther A *et al.* (2008) Chronology of the last glacial cycle in the European Alps. *Journal of Quaternary Science*, **23**, 559–573.
- Iwami T, Kock KH (1990) Channichthyidae. In: *Fishes of the Southern Ocean* (eds Gon O, Heemstra PC), pp. 381–399. JLB Smith Institute of Ichthyology, Grahamstown, South Africa.
- Janko K, Lecomte G, DeVries A (2007) Did glacial advances during the Pleistocene influence differently the demographic histories of benthic and pelagic Antarctic shelf fishes? - Inferences from intraspecific mitochondrial and nuclear DNA sequence diversity. *BMC Evolutionary Biology*, **7**, 220.
- Kock KH (1992) *Antarctic Fish and Fisheries*. Cambridge University Press, Cambridge.
- Kock KH (2005) Antarctic icefishes (Channichthyidae): a unique family of fishes. A review, part I. *Polar Biology*, **28**, 862–895.
- Kock KH, Kellermann AK (1991) Reproduction in Antarctic notothenioid fish. *Antarctic Science*, **3**, 125–150.
- Lautrédu AC, Hinsinger DD, Gallut C *et al.* (2012) Phylogenetic footprints of an Antarctic radiation: the Trematominae (Notothenioidei, Teleostei). *Molecular Phylogenetics and Evolution*, **65**, 87–101.

- Lecointre G (2012) Phylogeny and systematics of Antarctic teleosts: methodological and evolutionary issues. In: *Adaptation and Evolution in Marine Environments – The Impacts of Global Change on Biodiversity*, Vol. 1 (eds di Prisco G, Verde C), pp. 97–117, Series “From Pole to Pole” Springer, Berlin.
- Matschiner M, Hanel R, Salzburger W (2009) Gene flow by larval dispersal in the Antarctic notothenioid fish *Gobionotothen gibberifrons*. *Molecular Ecology*, **18**, 2574–2587.
- Matschiner M, Hanel R, Salzburger W (2011) On the origin and trigger of the notothenioid adaptive radiation. *PLoS ONE*, **6**, e18911.
- Molecular Ecology Resources Primer Development Consortium, Agostini C, Agudelo PA *et al.* (2011) Permanent genetic resources added to molecular ecology resources database 1 October 2010 - 30 November 2010. *Molecular Ecology Resources*, **11**, 418–421.
- Near TJ, Pesavento JJ, Cheng CHC (2003) Mitochondrial DNA, morphology, and the phylogenetic relationships of Antarctic icefishes (Notothenioidei: Channichthyidae). *Molecular Phylogenetics and Evolution*, **28**, 87–98.
- Near TJ, Pesavento JJ, Cheng CHC (2004) Phylogenetic investigations of Antarctic notothenioid fishes (Perciformes: Notothenioidei) using complete gene sequences of the mitochondrial encoded 16S rRNA. *Molecular Phylogenetics and Evolution*, **32**, 881–891.
- Near TJ, Parker SK, Detrich HW III (2006) A genomic fossil reveals key steps in hemoglobin loss by the Antarctic icefishes. *Molecular Biology and Evolution*, **23**, 2008–2016.
- Near TJ, Dornburg A, Kuhn KL *et al.* (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3434–3439.
- Neuenschwander S, Lurgiader CR, Ray N, Currat M, Vonlanthen P, Excoffier L (2008) Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular Ecology*, **17**, 757–772.
- Nielsen EE, Bach L, Kotlicki P (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, **6**, 971–973.
- Noor MAF (1999) Reinforcement and other consequences of sympatry. *Heredity*, **83**, 503–508.
- Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford.
- Papetti C, Zane L, Patarnello T (2006) Isolation and characterization of microsatellite loci in the icefish *Chionodraco rastrospinosus* (Perciformes, Notothenioidei, Channichthyidae). *Molecular Ecology Notes*, **6**, 207–209.
- Papetti C, Marino IAM, Agostini C, Bisol PM, Patarnello T, Zane L (2011) Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei. *Conservation Genetics Resources*, **3**, 259–262.
- Papetti C, Pujolar JM, Mezzavilla M *et al.* (2012) Population genetic structure and gene flow patterns between populations of the Antarctic icefish *C. rastrospinosus*. *Journal of Biogeography*, **39**, 1361–1372.
- Patarnello T, Marcato S, Zane L, Varotto V, Bargelloni L (2003) Phylogeography of the *Chionodraco* genus (Perciformes, Channichthyidae) in the Southern Ocean. *Molecular Phylogenetics and Evolution*, **28**, 420–429.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Ritchie PA, Lavoue S, Lecointre G (1997) Molecular phylogenetics and the evolution of Antarctic notothenioid fishes. *Comparative Biochemistry and Physiology Part A: Physiology*, **118**, 1009–1025.
- Roberts DG, Gray CA, West RJ, Ayre DJ (2009) Evolutionary impacts of hybridization and interspecific gene flow on an obligately estuarine fish. *Journal of Evolutionary Biology*, **22**, 27–35.
- Rousset F (2008) GENEPOP’007: a complete reimplemention of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Rutschmann S, Matschiner M, Damerou M *et al.* (2011) Parallel ecological diversification in Antarctic notothenioid fishes as evidence for adaptive radiation. *Molecular Ecology*, **20**, 4707–4721.
- Ruud JT (1954) Vertebrate without erythrocytes and blood pigment. *Nature*, **173**, 848–850.
- Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, New York.
- Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 1987–1998.
- Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution and Systematics*, **34**, 339–364.
- Streelman JT, Danley PD (2003) The stages of vertebrate evolutionary radiation. *Trends in Ecology and Evolution*, **18**, 126–131.
- Susana E, Papetti C, Barbisan F *et al.* (2007) Isolation and characterization of eight microsatellite loci in the icefish *Chionocephalus aceratus* (Perciformes, Notothenioidei, Channichthyidae). *Molecular Ecology Notes*, **7**, 791–793.
- Taylor EB, Boughman JW, Groenenboom M, Sniatynski M, Schluter D, Gow JL (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, **15**, 343–355.
- Thatje S, Hillenbrand CD, Mackensen A, Larter R (2008) Life hung a thread: endurance of Antarctic fauna in glacial periods. *Ecology*, **89**, 682–692.
- Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.
- Veeramah KR, Wegmann D, Woerner A *et al.* (2012) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution*, **29**, 617–630.
- Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18508–18513.

- Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABC toolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, **58**, 1–12.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**, 1–16.

Conceived and designed the experiments: C.A., I.A.M.M. and L.Z. Performed the experiments: C.A. and I.A.M.M. Analysed the data: C.A., A.B., G.B., I.A.M.M., M.M. and L.Z. Contributed reagents/materials/analysis tools: G.B., T.P., L.Z. Wrote the paper: G.B., S.M.H., I.A.M.M. and L.Z. Critically discussed the manuscript: A.B., C.A. and S.M.H., T.P.

Data accessibility

Microsatellites genotypes data are available in the DRYAD repository with doi:10.5061/dryad.n8n6k.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Microsatellite loci optimized in the three *Chionodraco* species.

Table S2 Posterior probability for M2, M4 and M5 (the models with the highest posterior probabilities) computed varying the divergence times among species.

Table S3 Prior distributions of the model parameters in the ABC analysis.

Table S4 Bayes Factor analysis to compare long-term demographic stability with exponential growth/decline.

Table S5 Summary statistics used in ABC analysis and their value in the real data.

Table S6 Posterior probabilities of the models (regression approach) under different numbers of retained simulations.

Table S7 Hardy-Weinberg Equilibrium probabilities calculated for each of the 18 loci and for each of the six population samples.

Table S8 Posterior probability of M5 in pairwise comparisons.

Table S9 Estimated parameters in the M5 model.

Fig. S1 Analysis for identification of loci under selection in the comparison between *Chionodraco hamatus* and *C. myersi* using LOSITAN.

Fig. S2 Analysis for identification of loci under selection in the comparison between *C. hamatus* and *C. rastrispinosus* using LOSITAN.

Fig. S3 Analysis for identification of loci under selection in the comparison between *C. myersi* and *C. rastrispinosus* using LOSITAN.

Fig. S4 Estimation of the most likely number of genetic groups following Evanno *et al.* (2005).

Fig. S5 ABC analysis: posterior probabilities that each model is correct given the joint observed posterior probabilities of all five models, as a function of the bandwidth in the density estimation (adjusted regression approach).

Fig. S6 ABC analysis: posterior distributions of summary statistics obtained in 10 000 simulations under M5. Summary statistics in this analysis differ from those used for model selection.

Fig. S7 ABC analysis: Principal Component Analysis based on the 15 summary statistics in Figure S6 obtained in 10 000 simulations under M5.

Fig. S8 ABC analysis: Principal Component Analysis of the 10 000 simulations closest to the observed summary statistics under each demographic model.

Fig. S9 ABC analysis: posterior distributions of the demographic parameters under M5.

Table S1. Microsatellite loci optimized in the three *Chionodraco* species

Locus	Primer sequences (5'-3')	Touchdown (TD) cycles	Number of TD cycles	Fixed T _a	Number of cycles at fixed T _a	Accession no.
Cr15^a	F: TAACATGCTCCCCTCTCC	-	-	59 °C	35	DQ141191
	R: ATATCATGTGGGCAGAATGGT					
Cr127^b	F: CGTATAGGGCCGTACCTCA	65 °C	12	55 °C	30	DQ141192
	R: GCTCCATCATAGATCCAGTCA					
Cr171^a	F: CTCGTGGTGGATGATATAAAA	-	-	57 °C	35	DQ141187
	R: AAGTCATATGGTAGCACTTTTAGC					
Cr236^b	F: CTCGAGCACAACATATCTGG	65 °C	12	55 °C	30	DQ141185
	R: GTACACGTACGGGATTGAA					
Ca21^b	F: GGACGAAGAAAGCAAGAAGAACGA	60 °C	12	52 °C	30	EF103946
	R: CTATGTGGGC(AGC)TGTTTGTGATTAT					
Ca35^b	F: CCCAAACTCCTCTCAATGTCTG	60 °C	12	52 °C	30	EF103949
	R: TCGGCTGTGTAACCTCCTCCA					
Ca48^b	F: CATTCTCCCTGAGTTTACCAC	58 °C	16	50 °C	30	EF103951
	R: TCTTCTCTTTACCTCTAC(AG)G					
Ca86^b	F: AGTTATTATGGTCGTAATCGCT	58 °C	16	50 °C	30	EF103952
	R: GCTGCTGGT(CT)ACAGGAGTGT					
221Pl^c	F: AGAGGTAGGACAAAAAGGACAGAT	60 °C	10	55 °C	30	HQ393892
	R: GAAAAGGGAAAGCATGATGAGTGG					
Ch126^d	F: CGGTTTTTATGCATGTTGCCA	-	-	57 °C	30	HQ395761
	R: ACTGCTCATTACACTGGTTC					
Ch623^d	F: GCTGTTTGATTCCCTCGTGAGG	-	-	57 °C	30	HQ395762
	R: AAAAGTGGTCCTCCGCTGCAGT					
Ch1968^d	F: AGCTCAAGGTGTTTCGACAGACG	-	-	57 °C	30	HQ395764
	R: TAGCCAGCAGCGCTAATCCTG					
Ch2309^d	F: CAGCTCAATTAACGCTTTGCA	-	-	57 °C	30	HQ395765
	R: CGTCTCAAATGCTGTCACAACC					
Ch2788^d	F: TGGTTTCGATGAAGAATGCTC	-	-	57 °C	30	HQ395766
	R: TGATGATATTGGTCGTGGTCCG					
Ch3603^d	F: TGAGACAAGTCAAATCCAACC	-	-	57 °C	30	HQ395767
	R: GGCATAAAGCTATTTGAGGCT					
Ch3866^d	F: AGCGTTACACACTCCATCCGT	-	-	57 °C	30	HQ395768
	R: CTGCACCACTCTACCAGGGACG					
Ch5817^d	F: TTTAAAGCTGGGAAACACAGG	-	-	57 °C	30	HQ395769
	R: GAAACTGTGACAAACACAGGCT					
Ch8461^d	F: ACAGAGGGAGTAAGACGCGTG	-	-	57 °C	30	HQ395770
	R: GGAAGGCTCTGTAGCTGCTGA					

^aA standard thermal profile was used according to the following scheme: (i) predenaturation: 94°C 3 min; (ii) 35 cycles: denaturation 94°C 30 s, annealing 57/59°C 40 s, extension 72°C 50 s; and (iii) final extension for 10 min at 72°C.

^bA touchdown (TD) thermal profile was used according to the following scheme: (i) predenaturation: 94 °C 3 min; (ii) 12-16 TD cycles: denaturation 94 °C 40 min, annealing 65-58 °C 30 s decreased of 0.5 °C each cycle, extension 72 °C 1 min; (iii) 30 cycles: denaturation 94 °C 40 s, annealing 55-50 °C 30 s, extension 72 °C 1 min; and (iv) final extension for 10 min at 72 °C.

^cA TD thermal profile was used according to the following scheme: (i) predenaturation: 94 °C 2 min; (ii) 10 TD cycles: denaturation 94 °C 30 s, annealing 60 °C 30 s decreased of 0.5 °C each cycle, extension 72 °C 30 s; (iii) 30 cycles: denaturation 94 °C 30 s, annealing 55 °C 30 s, extension 72 °C 30 s; and (iv) final extension for 10 min at 72 °C.

^dA standard multiplex thermal profile for microsatellite loci was used according to the following scheme: (i) initial activation step: 95 °C 15 min; (ii) 30 cycles: denaturation 94 °C 30 s, annealing 57 °C 90 s, extension 72 °C 60 s; and (iii) final extension for 30 min at 60 °C.

Table S2. Posterior probability for M2, M4 and M5 (the models with the highest posterior probabilities) computed varying the divergence times among species

The first line refers to the most supported divergence times based on mtDNA and nuclear data (Near *et al.* 2012), i.e. 2.0 MYA for the divergence of *C. myersi* and 1.8 MYA for the subsequent divergence between *C. rastrispinosus* and *C. hamatus*. The second line correspond to the analysis where we extracted the species divergence times in each simulation from normal distributions with means and standard deviations as estimated by Near *et al.* (2012). (Mean = 2.0 MY, SD = 800 KY for the divergence of *C. myersi*; Mean = 1.8 MYA, SD = 800 KY, for the subsequent divergence between *C. rastrispinosus* and *C. hamatus*). In the other estimates we assumed a fixed but more recent (-30%) or more ancient (10 to 30%) species divergence time compared to the best estimates.

Divergence Times	Posterior Probability		
	M2	M4	M5
Best estimates	0.29	0.22	0.49
Normal distribution	0.36	0.18	0.46
-30%	0.17	0.37	0.46
+10%	0.33	0.19	0.48
+20%	0.40	0.17	0.43
+30%	0.48	0.15	0.37

Table S3. Prior distributions of the model parameters in the ABC analysis

N : effective sizes (in chromosomes) of the three species (1 = *Chionodraco hamatus*; 2 = *C. myersi*; 3 = *C. rastrospinosus*) and the ancestral groups (A1 and A2, see Figure 2 in the main text); m : migration rates between pairs of species, with E and H referring to Eemian and Holocene migration rates, respectively; μ : mean mutation rate; P : mean geometric parameter of the mutation model; IR: irrelevant. Migration rates in model M2 have the same prior distributions as the migration rates specified in the table for the Eemian and the Holocene interglacial periods. Once the average mutation rate μ across loci was extracted from its normal prior distribution, single locus mutation rates were drawn from a Gamma distribution with mean μ and shape parameter equal to 2. The number of repeats by which mutants differ from ancestral alleles were generated following a Generalised Stepwise Mutation model, with uniform distribution of the mean geometric parameter P , and beta distributed single locus P values (as in Neuenschwander *et al.* 2008).

Parameter	Distribution	Mean	Mode	Quantiles				
				Min.	5%	50%	95%	Max.
N_1	uniform	19750	IR	500	2475	20250	38025	40000
N_2	uniform	19750	IR	500	2475	20250	38025	40000
N_3	uniform	19750	IR	500	2475	20250	38025	40000
N_{A1}	uniform	19750	IR	500	2475	20250	38025	40000
N_{A2}	uniform	19750	IR	500	2475	20250	38025	40000
μ	normal	$5.0 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-7}$	$3.0 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$	$7.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
P	uniform	0.40	IR	0	0.04	0.40	0.76	0.80
m_{E12}	loguniform	$7.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$	$2.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-5}$	$5.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
m_{E13}	loguniform	$7.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$	$2.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-5}$	$5.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
m_{E23}	loguniform	$7.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$	$2.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-5}$	$5.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
m_{H12}	loguniform	$7.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$	$2.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-5}$	$5.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
m_{H13}	loguniform	$7.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$	$2.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-5}$	$5.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
m_{H23}	loguniform	$7.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$	$2.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-5}$	$5.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$

Table S4. Bayes Factor analysis to compare long-term demographic stability with exponential growth/decline

We analysed the microsatellite and mitochondrial DNA variation pattern within species with the software BEAST v1.7.4 (Drummond *et al.* 2012) using both the constant population and the exponential growth/decline models. The Bayes Factor was used to find evidence for one model or the other. The log₁₀ Bayes Factor of the exponential growth/decline model over the constant population model was computed with the software Tracer 1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>) using the method suggested by Newton and Raftery (1994). According to the guidelines suggested by Kass and Raftery (1995), in the table below a Log₁₀BF > 2 should be considered as very strong evidence against the “exponential growth/decline” model. The 8 microsatellite loci selected for the Approximate Bayesian Computation were included in this analysis and a total of 249 base pairs of the Mitochondrial DNA D-loop typed in 188 individuals from the three species (Patarnello *et al.* 2003) were also analysed under the same framework. We used the Generalized Mutational Model (Fu & Chakraborty 1998) to model the evolution of microsatellites estimating the individual locus mutation rates directly from the data. For mtDNA analysis we used the HKY mutational model, as indicated by jmodeltest2 (Darriba *et al.* 2012), setting a uniform prior distribution to the mutation rate from 10⁻⁸ to 10⁻⁷ mutation per years (Ren *et al.* 2013). Beast analysis were run for 500 and 100 million steps for microsatellites and mtDNA respectively, discarding the first 10% of each run as burnin. Convergence of the runs was evaluated with Tracer and all parameters had an effective sample size of at least 100. To take into account the effect of the prior choice on results, we replicate all the analysis using a uniform prior distribution (the same used in the ABC analysis, U: 500 - 40,000) or the Jeffreys's one-on-x prior for the population size parameter.

Marker	Prior distribution	Log ₁₀ Bayes Factor		
		<i>C. hamatus</i>	<i>C. rastrorpinosus</i>	<i>C. myersi</i>
Microsatellite	Uniform	17.52	20.69	16.92
Microsatellite	Jeffreys (1/x)	12.21	38.59	28.53
mtDNA(d-loop)	Uniform	0.63	0.70	0.79
mtDNA(d-loop)	Jeffreys (1/x)	0.52	0.43	0.86

Table S5. Summary statistics used in ABC analysis and their value in the real data

Summary statistics were computed in the observed and simulated datasets using Arlequin 3.5.1.2 (Excoffier & Lischer 2010). The numbers in the third column refer to the species as in Table S3.

Heterozygosity	Average	1	0.71
		2	0.61
		3	0.71
	SD	1	0.15
		2	0.26
3		0.21	
Allelic range	Average	1	13.31
		2	12.20
		3	16.20
	SD	1	9.66
		2	10.10
3		10.60	
F_{st}	Average	1-2	0.21
		1-3	0.10
		2-3	0.12
Shared alleles	Total	1+2+3	52
		1+2	5
		1+3	9
		2+3	13
	SD	1+2+3	5.37
		1+2	0.91
		1+3	0.61
Private alleles	Total	1	19
		2	8
		3	18
	SD	1	1.99
		2	0.92
		3	1.48

Table S6. Posterior probabilities of the models (regression approach) under different numbers of retained simulations

The posterior probability of each model was initially computed with the regression approach (Beaumont 2008), where probabilities are estimated from a weighted multinomial logistic regression with the summary statistics in the retained simulations as the predictors and the model indicator as the dependent variable. The effect of considering different numbers of retained simulations is reported.

Threshold	Models				
	M1	M2	M3	M4	M5
20000	4.26×10^{-12}	0.286	1.63×10^{-4}	0.217	0.495
50000	6.67×10^{-10}	0.294	1.67×10^{-4}	0.257	0.448
75000	1.92×10^{-9}	0.307	1.92×10^{-4}	0.257	0.434

Table S7. Hardy-Weinberg Equilibrium probabilities calculated for each of the 18 loci and for each of the six population samples

Significant departures from HWE are in bold. *Locus Ch126 was monomorphic in this population sample, thus HWE cannot be calculated.

Locus	ChMW88	ChBT95	CmMR89	CmMW03	CrEI96	CrJI06
Cr15	0.142	0.000	0.249	1.000	0.010	0.012
Cr127	0.010	0.568	0.970	0.299	0.025	0.026
Cr171	0.154	0.051	0.239	1.000	0.000	0.000
Cr236	0.049	0.000	0.178	0.210	0.172	0.059
Ca21	0.069	0.000	0.000	0.003	0.196	0.274
Ca35	0.937	0.231	1.000	1.000	0.874	0.690
Ca48	1.000	0.625	0.210	0.173	0.371	0.279
Ca86	0.119	0.068	0.000	0.000	0.837	0.668
221PI	0.059	0.039	0.743	0.013	0.392	0.019
Ch126	- *	1.000	1.000	1.000	1.000	1.000
Ch623	0.803	0.111	0.792	0.733	1.000	0.009
Ch1968	1.000	1.000	1.000	1.000	0.235	1.000
Ch2309	1.000	0.887	0.723	0.557	0.135	1.000
Ch2788	0.131	0.245	0.265	0.598	0.866	0.159
Ch3603	1.000	0.215	1.000	1.000	0.001	0.472
Ch3866	0.882	1.000	0.614	0.480	0.114	0.114
Ch5817	0.493	0.817	0.000	0.006	0.620	0.766
Ch8461	0.880	0.565	0.804	0.847	0.117	0.325

Table S8. Posterior probability of M5 in pairwise comparisons

Logistic regression was used to compute the posterior probability of M5 when compared separately with each of other four models.

	Models			
	M1	M2	M3	M4
M5	0.99	0.62	0.99	0.62

Table S9. Estimated parameters in the M5 model.

Three point estimates and 95% credible intervals are given. See Table S3 for the names of the parameters. R^2 is the proportion of the parameters variance explained by the summary statistics. The posterior distributions of the parameters the model with highest posterior probability (M5) were estimated using the 2,000 simulations closest to the observed data, and performing a locally weighted linear regression between retained parameters and summary statistics (Beaumont *et al.* 2002). A logtan transformation was applied to the parameters before the regression step (Hamilton *et al.* 2005) to avoid posterior distributions outside the range of the corresponding priors. The posterior distributions for the joint parameter $m_{Hij}(N_i+N_j)$, corresponding to the effective numbers of chromosomes exchanged between populations i and j during the Holocene interglacial, are also given. The R function available from popABC website (<http://code.google.com/p/popabc/>) running under R 2.10 (R Development Core Team, 2009) was used for model comparison and parameters estimation.

	Mean	Median	Mode	95%HPD Low	95%HPD High	R^2
N_1	19534	18719	14275	4770	36776	0.62
N_2	13216	11222	6722	1464	30653	0.63
N_3	22167	22087	21402	7717	39486	0.63
N_{A1}	19563	19386	500	500	32901	$2.90 \cdot 10^{-4*}$
N_{A2}	17547	16030	500	500	30171	$1.36 \cdot 10^{-3*}$
μ	$3.82 \cdot 10^{-4}$	$3.73 \cdot 10^{-4}$	$3.44 \cdot 10^{-4}$	$1.66 \cdot 10^{-4}$	$6.19 \cdot 10^{-4}$	0.26
P	0.36	0.34	0.11	0.00	0.75	$3.29 \cdot 10^{-5*}$
m_{E12}	$1.07 \cdot 10^{-3}$	$5.51 \cdot 10^{-5}$	$1.00 \cdot 10^{-8}$	$1.00 \cdot 10^{-8}$	$6.58 \cdot 10^{-3}$	0.02*
m_{E13}	$3.48 \cdot 10^{-4}$	$4.23 \cdot 10^{-6}$	$1.00 \cdot 10^{-8}$	$1.00 \cdot 10^{-8}$	$3.12 \cdot 10^{-3}$	0.02*
m_{E23}	$8.09 \cdot 10^{-4}$	$4.28 \cdot 10^{-5}$	$1.00 \cdot 10^{-8}$	$1.00 \cdot 10^{-8}$	$5.38 \cdot 10^{-3}$	0.02*
m_{H12}	$2.02 \cdot 10^{-4}$	$7.29 \cdot 10^{-5}$	$1.00 \cdot 10^{-8}$	$1.00 \cdot 10^{-8}$	$6.62 \cdot 10^{-4}$	0.73
m_{H13}	$7.97 \cdot 10^{-4}$	$3.82 \cdot 10^{-4}$	$1.00 \cdot 10^{-8}$	$1.00 \cdot 10^{-8}$	$3.00 \cdot 10^{-3}$	0.73
m_{H23}	$1.30 \cdot 10^{-3}$	$7.04 \cdot 10^{-4}$	$3.55 \cdot 10^{-4}$	$1.00 \cdot 10^{-8}$	$4.74 \cdot 10^{-3}$	0.73
$m_{H12}(N_1+N_3)$	7.70	3.90	$1.00 \cdot 10^{-5}$	$1.00 \cdot 10^{-5}$	27.80	0.74
$m_{H13}(N_2+N_3)$	25.60	13.80	$1.00 \cdot 10^{-5}$	$1.00 \cdot 10^{-5}$	85.10	0.74
$m_{H23}(N_1+N_2)$	35.60	22.00	14.40	$1.00 \cdot 10^{-5}$	111.60	0.74

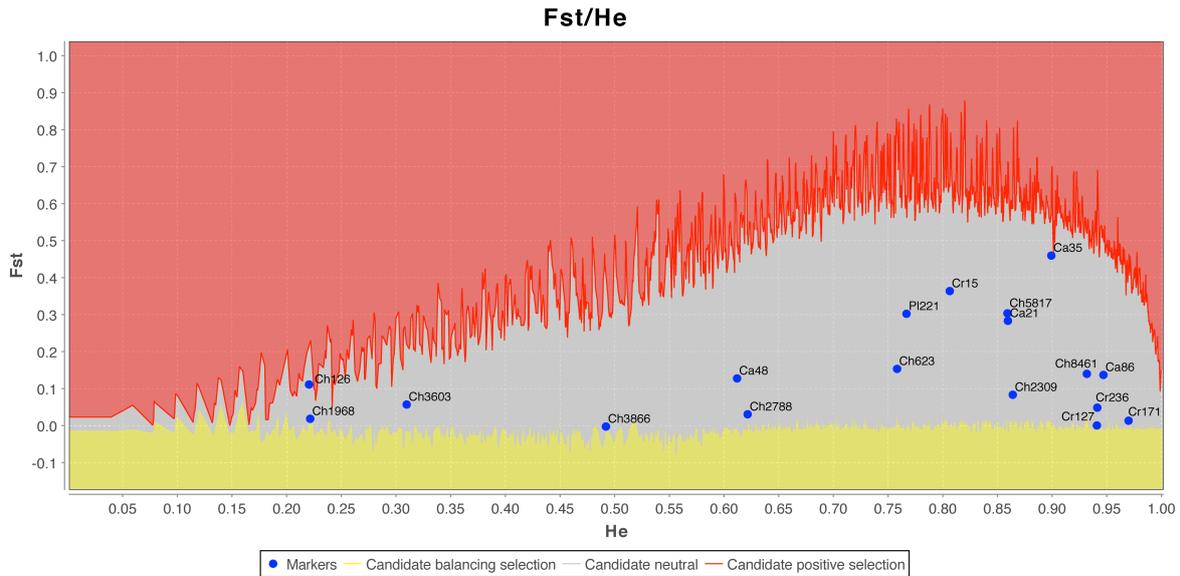


Figure S1. Analysis for identification of loci under selection in the comparison between *Chionodraco hamatus* and *C. myersi*. Results from LOSITAN 1,000,000 simulations under the SMM (Stepwise Mutation Model) using all 18 loci. Shown in grey is the 99% confidence interval of the expected F_{st} vs. H_e (expected heterozygosity) distribution with neutral markers. Loci outside this interval are potential candidates for being subject to directional selection (red) or balancing selection (yellow).

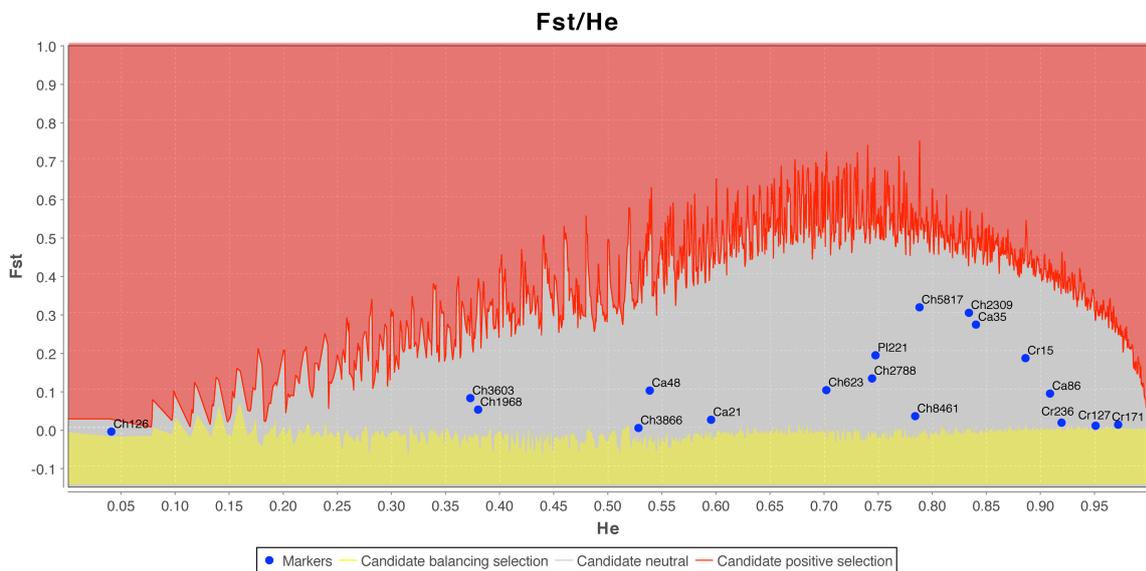


Figure S2. Analysis for identification of loci under selection in the comparison between *Chionodraco hamatus* and *C. rastrispinosus*. Results from LOSITAN 1,000,000 simulations under the SMM (Stepwise Mutation Model) using all 18 loci. Shown in grey is the 99% confidence interval of the expected F_{st} vs. H_e (expected heterozygosity) distribution with neutral markers. Loci outside this interval are potential candidates for being subject to directional selection (red) or balancing selection (yellow).

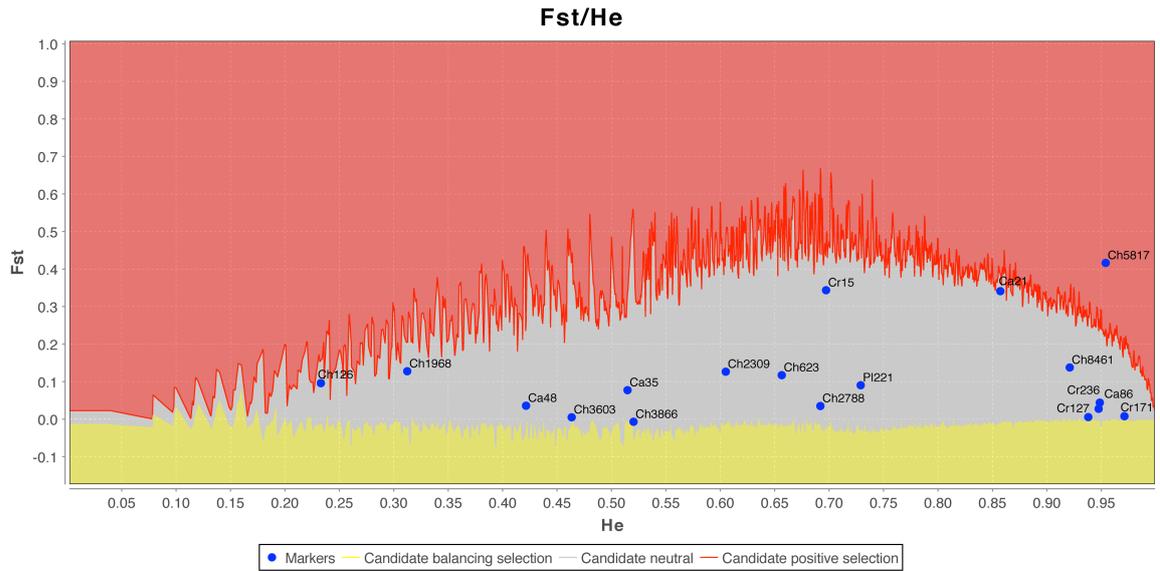


Figure S3. Analysis for identification of loci under selection in the comparison between *Chionodraco myersi* and *C. rastrospinosus*. Results from LOSITAN 1,000,000 simulations under the SMM (Stepwise Mutation Model) using all 18 loci. Shown in grey is the 99% confidence interval of the expected F_{st} vs. H_e (expected heterozygosity) distribution with neutral markers. Loci outside this interval are potential candidates for being subject to directional selection (red) or balancing selection (yellow).

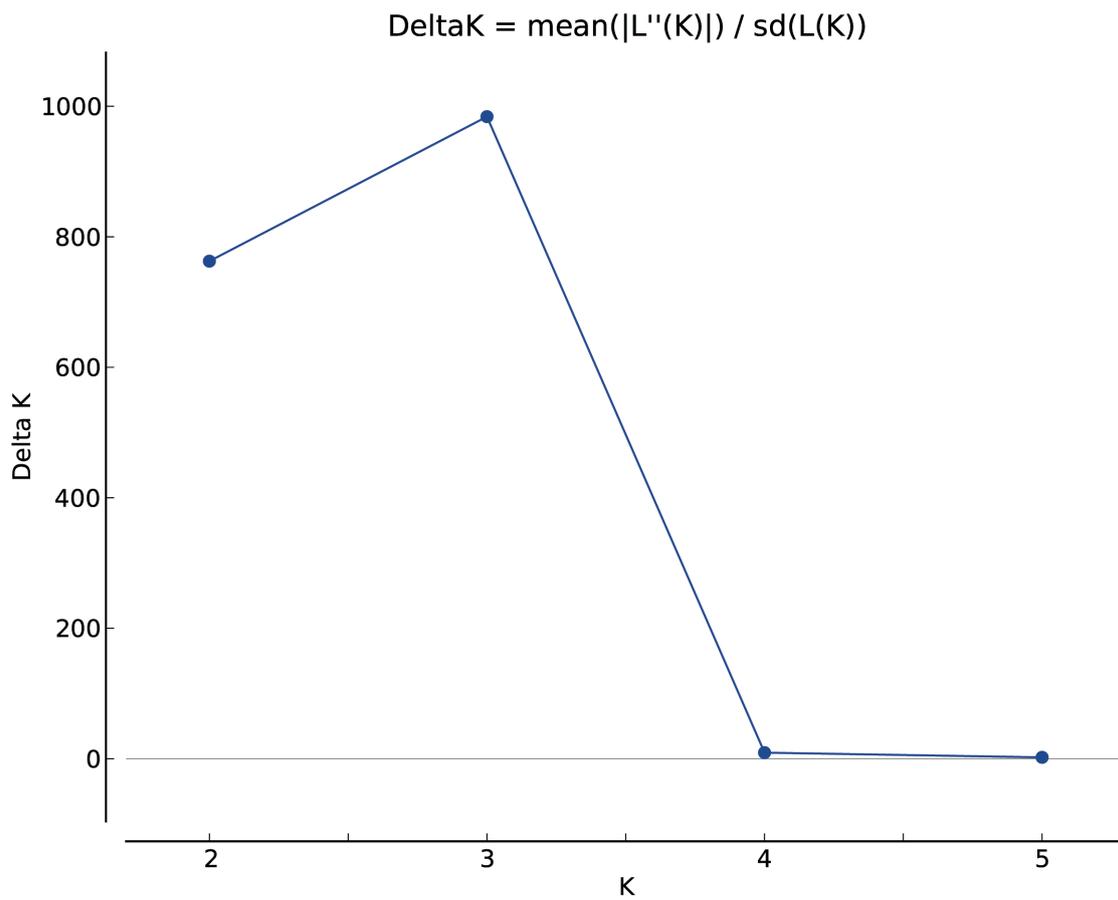


Figure S4. Estimation of the most likely number of genetic groups. Values of ΔK as rate of change of the likelihood function (estimated as in Evanno *et al.* 2005) plotted against K for the *Chionodraco* population samples.

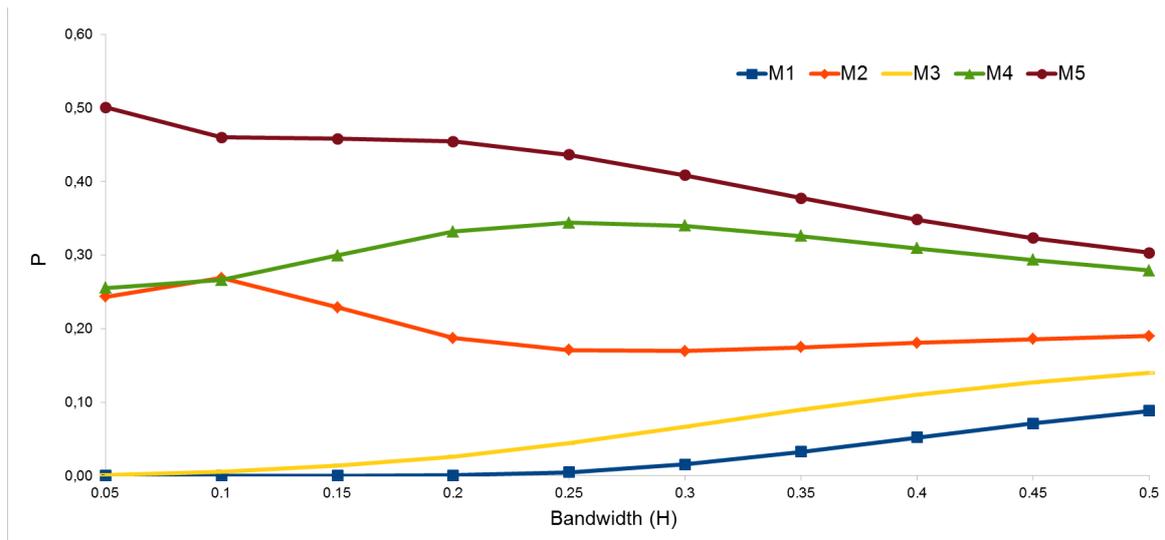


Figure S5. Posterior probabilities that each model is correct given the joint observed posterior probabilities of all five models, as a function of the bandwidth in the density estimation (adjusted regression approach)

We also computed the probability that each model is the correct model given the joint observed posterior probabilities of all five models. This method is based on 10,000 simulations of each model, with parameter values extracted from their prior distributions, and a multivariate product Gaussian kernel density estimation of the combined model probabilities (Veeramah *et al.* 2012). The density estimation depends on a bandwidth parameter, with too small values producing under-smoothing of the density and too large values producing over-smoothing. The values tested by Veeramah *et al.* (2012) in a similar ABC study were between 0.05 and 0.3, and we cautiously extended the higher value to 0.5.

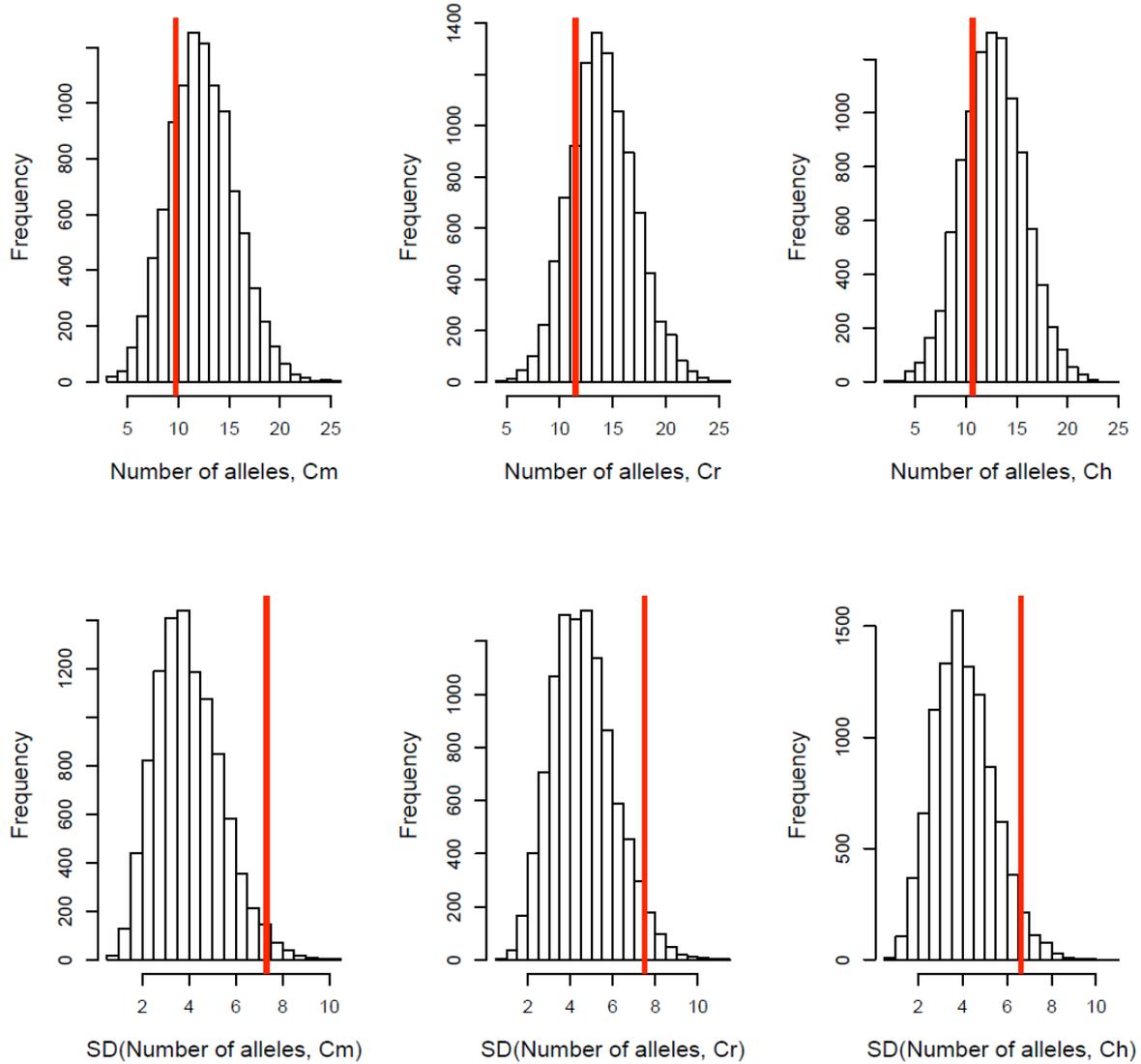


Figure S6. Posterior distributions of summary statistics. Summary statistics in this analysis differ from those used to select the most supported model. Each distribution corresponds to the posterior distribution obtained simulating M5 10,000 times with parameters extracted from the posterior distribution of the parameters. Red lines correspond to the observed (real) values. Cm, Cr, and Ch correspond to the three species. SD = Standard deviation. GW = Garza-Williamson'M index; Jost'D is an index of population differentiation (Jost 2008).

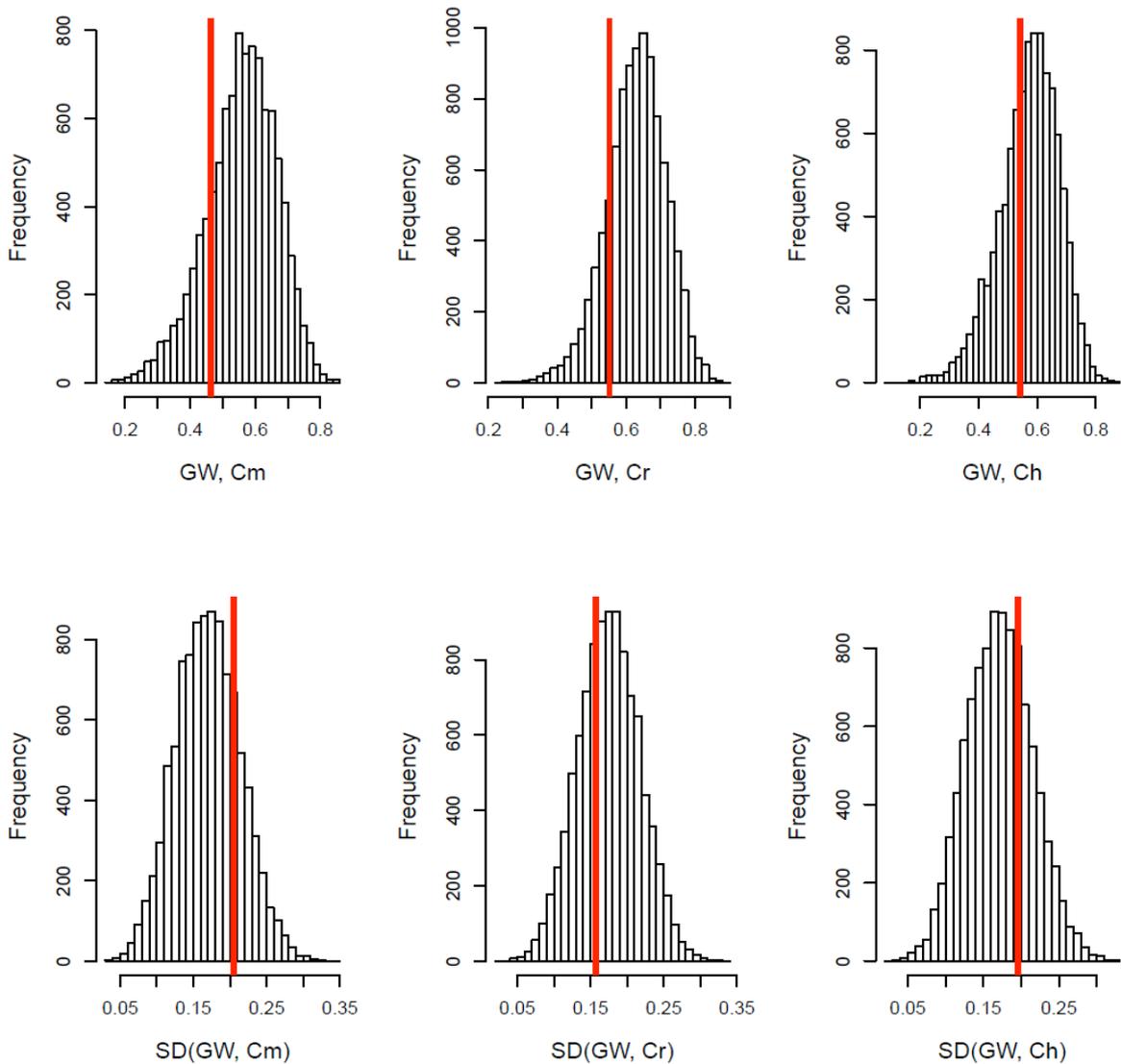


Figure S6 (continued). Posterior distributions of summary statistics. Summary statistics in this analysis differ from those used to select the most supported model. Each distribution corresponds to the posterior distribution obtained simulating M5 with parameters extracted from the posterior distribution of the parameters. Red lines correspond to the observed (real) values. Cm, Cr, and Ch correspond to the three species. SD = Standard deviation. GW = Garza-Williamson's M index; Jost'D is an index of population differentiation (Jost 2008).

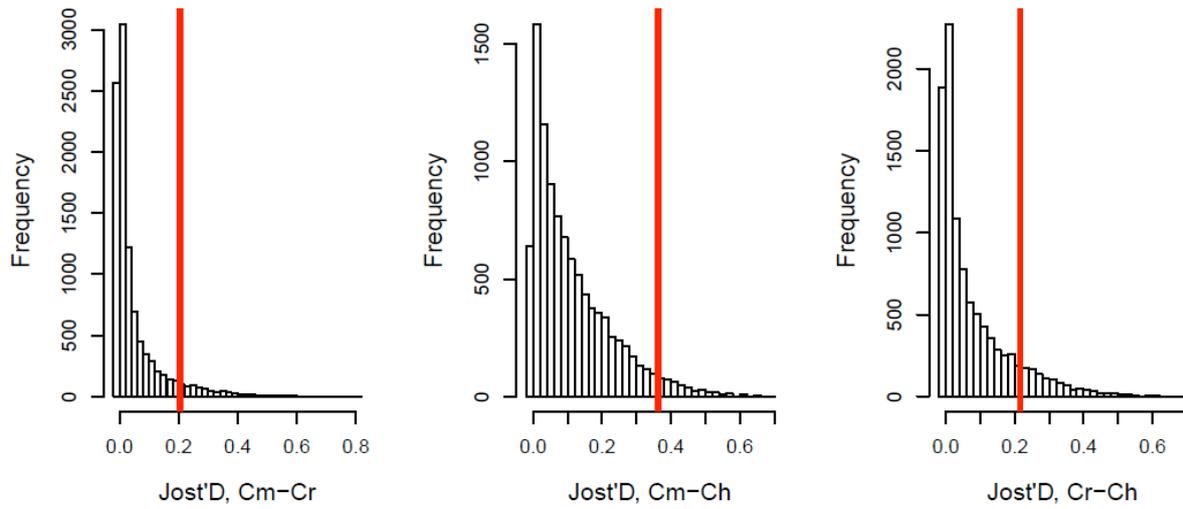


Figure S6 (continued). Posterior distributions of summary statistics. Summary statistics in this analysis differ from those used to select the most supported model. Each distribution corresponds to the posterior distribution obtained simulating M5 with parameters extracted from the posterior distribution of the parameters. Red lines correspond to the observed (real) values. Cm, Cr, and Ch correspond to the three species. SD = Standard deviation. GW = Garza-Williamson's M index; Jost'D is an index of population differentiation (Jost 2008).

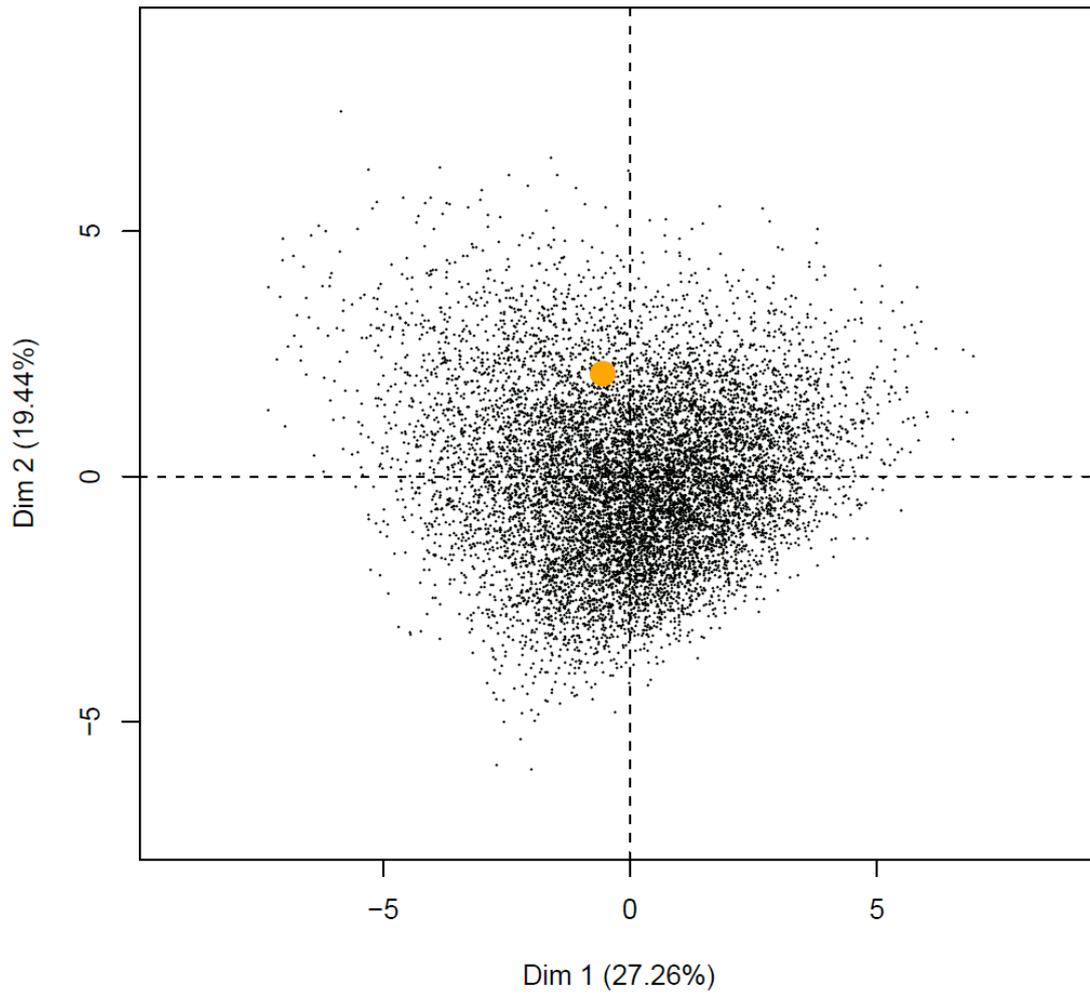


Figure S7. Principal Component Analysis based on the 15 summary statistics in Figure S6 obtained in 10,000 simulations under M5. Model parameters are sampled from their posterior distributions (Figure S6). The observed (real) data set is shown in orange.

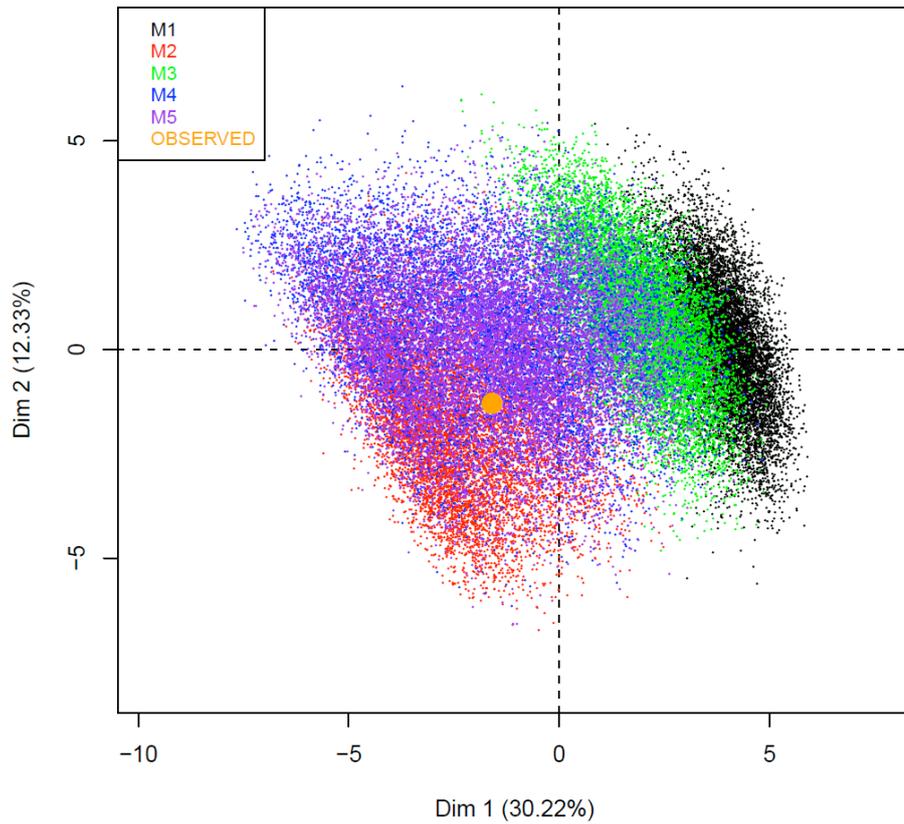


Figure S8. Principal Component Analysis of the 10,000 simulations closest to the observed summary statistics under each demographic model. Euclidean distances were computed between the summary statistics observed in the real sample and those obtained in the simulations generated under each model. The observed (real) data set is shown in orange.

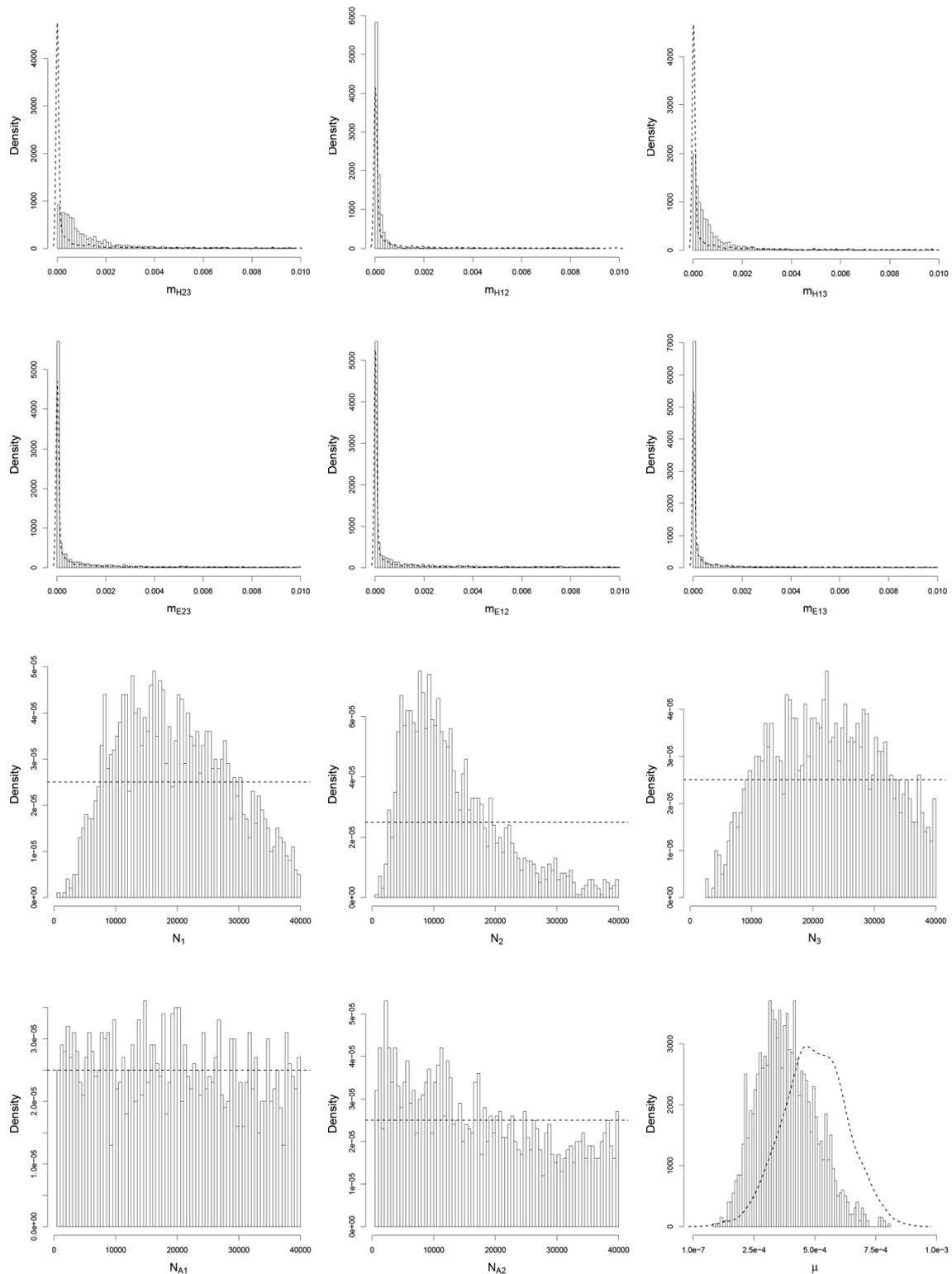


Figure S9. Posterior distributions of the demographic parameters under M5. Posterior distributions were estimated by performing a locally weighted regression adjustment on 2,000 retained simulations. Prior distributions are represented with dashed lines.

References

- Beaumont M (2008). Joint determination of topology, divergence time and immigration in population trees. In: Matsumura S, Forster P, Renfrew C, (Eds), *Simulations, genetics, and human prehistory*. Cambridge: Cambridge (UK): McDonald Institute for Archaeological Research. pp. 135-154.
- Beaumont MA, Zhang W, Balding DJ (2002). Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025-2035.
- Darriba D, Taboada GL, Doallo R, Posada D (2012). JModelTest2: more models, newheuristics and parallel computing. *Nature Methods*, **9**, 772.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *molecular Biology and Evolution*, **29**, 1969-1973.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Ecology*, **14**, 2611-2620.
- Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.
- Fu Y, Chakraborty R (1998). Simultaneous estimation of all the parameters of a stepwise mutation mode. *Genetics*, **150**, 487-497.
- Hamilton G, Stoneking M, Excoffier L (2005). Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilineal populations. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 7476-7480.
- Jost L (2008). G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015-4026.
- Kass RE, Raftery AE (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773-795.
- Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, Zane L, Fernandez DA, Jones CD (2012). Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3434-3439.
- Neuenschwander S, Largiadèr CR, Ray N, Currat M, Vonlanthen P, Excoffier L (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular Ecology*, **17**, 757-772.
- Newton MA, Raftery AE (1994). Approximate bayesian inference by weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Ser. B*, **56**, 3-48.
- Patarnello T, Marcato S, Zane L, Varotto V, Bargelloni L (2003) Phylogeography of the *Chionodraco* genus (Perciformes, Channichthyidae) in the Southern Ocean).

Molecular Phylogenetics and Evolution, **28**, 420-429.

R Development Core Team (2009). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Ren G, Liu Q, Gao T, Yanagimoto T (2013). Population demography and genetic structure of the fat greenling (*Hexagrammos otakii*) inferred from mtDNA control region sequence analyses. *Biochemical Systematics and Ecology*, **47**, 156-163.

Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, Destro-Bisol G, Soodyall H, Louie L, Hammer MF (2012). An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution*, **29**, 617-630.

PAPER VII

Has climate change promoted genetic fragmentation in the ice-dependent fish *Pleuragramma antarcticum*?

Cecilia Agostini¹, Tomaso Patarnello², Julian Ashford³, Joseph Torres⁴, Lorenzo Zane¹

¹Department of Biology, University of Padova, Via G. Colombo 3, 35121, Padova, Italy

²Department of Comparative Biomedicine and Food Science, University of Padova, Agripolis, Viale dell'Università 16, 35020, Legnaro (Padova), Italy

³Center for Quantitative Fisheries Ecology, Old Dominion University, Norfolk, VA, USA

⁴College of Marine Science, University of South Florida, St. Petersburg, FL, USA

ABSTRACT

Aim

Pleuragramma antarcticum is the only Antarctic notothenioid characterized by a complete pelagic life cycle and plays a major trophic role in the coastal Antarctic marine ecosystem. A previous genetic study investigated the population structure of this species, but used mitochondrial DNA sequencing and was unable to discriminate between hypotheses of panmixia, with occasional fluctuations of gene pools, and population structure. The aim of the present study is the investigation of the population structure of *P. antarcticum* along the Antarctic Peninsula (AP) shelf, a region of Antarctica highly impacted by regional warming.

Location

Four geographic areas along the AP shelf: Charcot Island, Marguerite Bay, Joinville Island and Larsen Bay.

Methods

A total of 562 individuals from nine population samples, representative of four geographic areas, were genotyped at 16 EST-linked microsatellites. Indices of genetic variability and Hardy-Weinberg probabilities were calculated in each population. Population structure was studied by using non-hierarchical and hierarchical F-statistics. A Bayesian method was used to estimate the level of recent gene flow between geographic localities.

Results

We found a single gene pool and an absence of inter-annual variability in the southwestern AP, while significant genetic differences were detected on a small geographic scale from samples collected off the tip of the AP, with a signal of increased fragmentation over time. Assignment tests revealed a stronger flow of migrants moving southward along the western AP, following the anti-clockwise Coastal Current, than in the opposite direction.

Main conclusions

Reduced level of gene flow along the shelf, the increase of population fragmentation with time, and the inability to capture *P. antarcticum* in the central region of the western AP for two consecutive years, all suggest that this sea-ice dependent species could be highly vulnerable to climate change with possible cascading effects on the Antarctic marine food web.

INTRODUCTION

Population connectivity versus isolation in the Southern Ocean

Both life history and large-scale ocean circulation can influence the population genetic structure of marine fish (Hauser & Carvalho, 2008) forming complex interactions that shape gene flow. In the Southern Ocean, most Antarctic notothenioids are characterized by a sedentary habit, which can limit the dispersal ability of the adult stage. Yet, many species show long larval and early juvenile pelagic stages (Eastman, 1993) that may promote connectivity between populations via two strong circumpolar currents: the clockwise flowing Antarctic Circumpolar Current (ACC) (Orsi *et al.*, 1995), which transports more water than any other ocean current (Klinck & Nowlin, 2001), and the anticlockwise Coastal Current (CC). Early studies indicated that young krill (*Euphausia superba*) entrained in fronts are advected along the ACC from the western Antarctic Peninsula (AP) to South Georgia (e.g. Hofmann *et al.*, 1998; Fach & Klinck, 2006) and simulations of the large-scale circulation predicted potential transport pathways connecting disparate areas of the Southern Ocean (e.g. Thorpe *et al.*, 2007).

Empirical studies using otolith chemistry (e.g. Ashford *et al.*, 2006, 2008, 2010b) have also found evidence of population structuring and connectivity in notothenioids related to the large-scale circulation of the Southern Ocean. For instance, Antarctic currents and fronts were shown to influence population structure of the Patagonian toothfish (*Dissostichus eleginoides*), in which adult neutral buoyancy is expected to facilitate

large-scale dispersal at low energetic cost (Eastman, 1993). In this species, otolith chemistry revealed significant differences between samples from South America and Antarctica, consistent with the hydrogeographic barrier formed by the Polar Front of the ACC. A cluster analysis also suggested the presence of three discrete groups in the Antarctic sector, forming a complex structure, which resulted from advective transport along the ACC (Ashford *et al.*, 2008).

Analyses of population structure using genetic approaches, which can directly address gene flow, have thus far shown considerable variability between Antarctic species. Matschiner *et al.* (2009) pointed out that connectivity generated by oceanographic currents may promote gene flow around Antarctica, limiting genetic differentiation. Their study, using mitochondrial DNA and microsatellite data, indicated genetic homogeneity in samples of *Gobionotothen gibberifrons* collected in the Scotia Sea and around the northwestern AP. An eastward unidirectional gene flow between sampling sites was inferred, consistent with the west-to-east direction of the ACC (Matschiner *et al.*, 2009). In addition, the authors suggested that the lack of population genetic structure could be a general pattern in Antarctic fish due to the combined effect of the ACC strength and long larval stages. Similarly, low or absent population differentiation and high levels of gene flow were found in several notothenioid species inhabiting the southern Scotia Arc and characterized by distinct life history traits and larval durations (Damerou *et al.*, 2012; Papetti *et al.*, 2012). However, in the same geographic area, genetic structuring was

found among population samples of the icefish *Chaenocephalus aceratus*; in this species, microsatellite markers showed significant differentiation both at the temporal and geographic level and, while the prevailing direction of gene flow followed the ACC, a signal of counter-current migration was also detected (Papetti *et al.*, 2009, 2012). Evidence of population structuring was also reported in other Antarctic notothenioids, characterized by both demersal and pelagic life stages (Patarnello *et al.*, 2003, 2011 for review).

A better understanding of the role of the large-scale circulation in shaping the genetic structure of Antarctic notothenioid populations is critically needed, especially in the light of recent climate change. Over the last 50 years, surface temperature data have revealed significant warming trends in west Antarctica and, above all, across the AP (Convey *et al.*, 2009), leading to reduced sea-ice extent, shifts in hydrography, and alterations of community structure and species distribution (Moline *et al.*, 2008). In addition, global warming may potentially shorten larval-stage duration and thus, dispersal ability (O'Connor *et al.*, 2007). Human-driven reductions in connectivity may lead to fragmented and isolated populations, susceptible to the erosion of genetic variability and, potentially, to local extinction.

Antarctic silverfish

This study focuses on *Pleuragramma antarcticum* (Notothenioidae, Nototheniidae), also known as Antarctic silverfish, which is characterized by a circum-Antarctic distribution. It is a

keystone species in the trophic web of the Antarctic marine ecosystem, feeding mainly on copepods and different life stages of Antarctic krill, and is a major component of the diet of almost all top predators (La Mesa *et al.*, 2004; Smith *et al.*, 2007b). *P. antarcticum* is the only Antarctic notothenioid characterized by a complete pelagic life cycle and it is found in most shelf areas of the Southern Ocean, where it is the dominant fish in the water column (La Mesa & Eastman, 2011). The Antarctic silverfish is found at depths from 0 to 700 m (DeWitt *et al.*, 1990) and displays stratification by depth (La Mesa *et al.*, 2010 for review): eggs are pelagic and are found in coastal waters within the platelet ice layer under the sea-ice cover; early larvae are typically distributed at depth of 0-100 m, while post-larvae and juveniles live in deeper (down to 400 m) and offshore waters, potentially contributing to species dispersal by transport in the CC; larger fish are generally found deeper than 400 m and display vertical migration in the presence of seasonal light. Both adults and early-life stages are dependent on sea-ice (Vacchi *et al.*, 2004), making the species potentially extremely vulnerable to recent climate warming (La Mesa & Eastman, 2011; Vacchi *et al.*, 2012).

Despite its importance to the Antarctic marine ecosystem, the population structure of *P. antarcticum*, has not yet been fully clarified. Partial sequencing of the D-loop mitochondrial region in samples collected at four locations in the Southern Ocean (Weddell Sea, Ross Sea, Elephant Island, and King George Island), showed a situation “close to panmixia” with small although significant differences detected both on a temporal and geographic scale

(Zane *et al.*, 2006). Genetic differentiation was detected between samples taken in the Weddell Sea two years apart and one of these samples was also different from a population sample collected in the Ross Sea. However, genetic homogeneity was found in most pairwise comparisons, also between samples from opposite sides of the continent. Two hypotheses were proposed to explain the pattern: first, *P. antarcticum* might have a strong population structure, which was not detected because of the small sample size and limited power of the molecular marker used; second, a general context of panmixia might exist with sporadic and weak differences between populations as a consequence of temporal and local fluctuations of gene pools (Zane *et al.*, 2006).

Here, to elucidate the population structure of Antarctic silverfish, we increased the power to detect differences by using a large panel of 16 microsatellite loci, and

we focused on a more restricted geographic area, the AP, which is considered a true “hot-spot” of recent climate warming (Zazulie *et al.*, 2010). Because both adults and early-life stages are dependent on sea-ice, the monitoring of the genetic variation level and population differentiation pattern of Antarctic silverfish is crucial to evaluate the possible impact of climate change on its population genetic structure, and the consequent implications for the Antarctic marine ecosystem.

MATERIALS AND METHODS

Sample collection and DNA extraction

A total of 562 individuals from nine population samples, representative of four geographic areas, have been considered in this study (**Fig. 1, Table 1**). Samples from Marguerite Bay 2001 and 2002 were collected as described in Donnelly & Torres (2008). In 2010 sampling was

Table 1. *Pleuragramma antarcticum* population samples collected between 2001 and 2012 at four different locations off the Antarctic Peninsula. Reported are collection site, sampling year, sampling campaign, and number of individuals analyzed at 16 EST-linked microsatellites (N).

Site	Year	Campaign	Coordinates	N
Charcot Island	2010	NBP 10-02 ^a	70°07'S, 76°02'W	60
Marguerite Bay	2001	SO GLOBEC-Cruise 1 ^b	67°57'S, 68°21'W	28
	2002	SO GLOBEC-Cruise 3 ^c	68°08'S, 68°01'W	49
	2010	NBP 10-02 ^a	67°49'S, 68°09'W	60
	2011	LMG Cruise 11-01, Palmer LTER ^d	67°39'S, 70°04'W	83
Joinville Island	2007	ANTXXIII/8AWI ^e	62°35'S, 54°45'W	34
	2010	NBP 10-02 ^a	63°30'S, 56°40'W	148
	2012	ANT-XXVIII/4 AWI ^e	62°14'S, 55°18'W	54
Larsen Bay	2007	ANTXXIII/8AWI ^e	65°30'S, 61°40'W	46

^a cruise conducted during late March-April 2010 on board the R/V Nathaniel B. Palmer

^b cruise conducted during 2001 austral fall aboard the Antarctic Research Support Vessel (A.R.S.V.) *Laurence M. Gould*, as part of the Southern Ocean Global Ocean Ecosystems Dynamics (SO GLOBEC) program

^c cruise conducted during 2002 austral fall aboard the A.R.S.V. *Laurence M. Gould*, as part of the Southern Ocean Global Ocean Ecosystems Dynamics (SO GLOBEC) program

^d cruise conducted during January 2011 aboard the A.R.S.V. *Laurence M. Gould*, as part of the Palmer Long-Term Ecological Research (LTER) program

^e AWI: Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany, aboard the German research vessel Polarstern

performed as described by Ashford *et al.* (2010a) and fishing activity was conducted at six sites along the western AP shelf: off Charcot Island, in Marguerite Bay, south of Joinville Island, south of Renaud Island, over the shelf west and south of Anvers Island, and in Croker Passage; in the three latter sites the species was not detected (**Fig. 1**). In 2011 sampling was conducted based on the Palmer LTER regional grid (900 km x 200 km west of the AP) from Anvers Island to Charcot Island (Waters & Smith, 1992), with successful collection at Marguerite Bay (**Table 1**). One additional sample from Larsen Bay (2007) and two samples from Joinville Island (2007 and 2012), which were caught during AWI (Alfred Wegener Institute) Antarctic expeditions, were included in the study. Standard length (SL) was recorded for all specimens and a small piece of muscle tissue was collected and preserved in RNA later (absolute ethanol for 2001 and 2002 samples) at -80 °C until the molecular analysis. Genomic DNA was extracted from 10-100 mg of muscle tissue following a standard salting out protocol (Patwary *et al.*, 1994). DNA solutions were stored at -20 °C before PCR amplification.

Microsatellite analysis: DNA amplification and genotyping

Population samples were analyzed with 16 EST-linked microsatellite loci: 2 from Molecular Ecology Resources Primer Development Consortium *et al.* (2011), and 14 from Molecular Ecology Resources Primer Development Consortium *et al.* (2013). All loci were mined from a database of about 24,000 contigs obtained by high-throughput sequencing of a normalized cDNA library from *Chionodraco hamatus* skeletal muscle

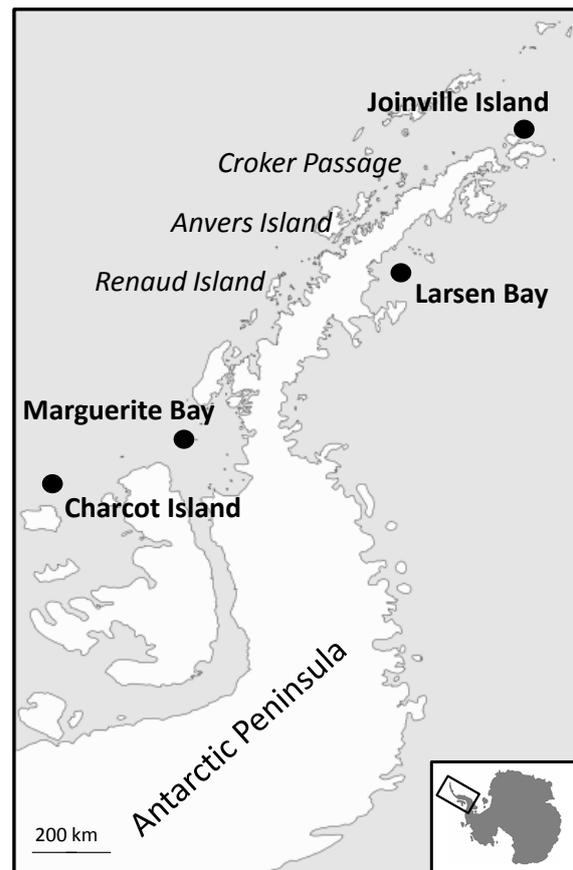


Figure 1. Approximate sampling locations of *Pleuragramma antarcticum* along the Antarctic Peninsula shelf (circles). Reported in italics are also areas where the species was not detected in 2010 and 2011.

(Coppe *et al.*, 2013). Sixteen loci were multiplexed in two PCR reactions, specifically optimized for cross-amplification in *Pleuragramma antarcticum*. Multiplexed PCRs were performed in 10 µL total volume containing: 1X QIAGEN Multiplex PCR Master mix (QIAGEN, HotStarTaq DNA Polymerase, Multiplex PCR Buffer, dNTP Mix), 0.2 µM primer mix and 100 ng of template DNA. PCR conditions were as follows: (i) initial activation step: 95 °C 15 min; (ii) 30 cycles: denaturation 94 °C 30 s, annealing 57 °C 90 s, extension 72 °C 60 s; and (iii) final extension: 60 °C 30 min.

Fragment analysis was performed on an ABI 3730xl automated sequencer and microsatellite analysis was carried out using PEAK SCANNER ver. 1.0 (Applied Biosystems). In order to minimize the negative consequences of a poor allele calling, binning was automated with FLEXIBIN ver. 2 (Amos *et al.*, 2007) and the final scoring was then manually checked to ensure the accuracy of the process.

Microsatellite analysis: data analysis

Genetic variability within each sample was assessed by computing observed heterozygosity (H_O) and unbiased expected heterozygosity (H_E) with GENETIX ver. 4.05.2 (Belkhir *et al.*, 1996-2004), while number of alleles (N_A) and allelic richness (A_R) were calculated using FSTAT ver. 2.9.3 (Goudet, 2001). Deviations from Hardy-Weinberg equilibrium (HWE) and genotypic linkage equilibrium between pairs of loci in each population were tested with GENEPOP ver. online 4.0.10 (Raymond & Rousset, 1995; Rousset, 2008). Significance of all tests was estimated by the Markov Chain method (dememorization number = 10,000; number of batches = 500; number of iterations per batch = 10,000).

The software LOSITAN (Antao *et al.*, 2008), which implements an F_{ST} outlier detection approach (Beaumont & Nichols, 1996), was used to identify loci potentially under selection. The program uses coalescent simulations based on observed data and an island model of migration to generate the expected distribution of F_{ST} vs. H_E with neutral markers. This distribution is then used to detect outlier loci, which show significantly higher or lower F_{ST} values compared with neutral expectations and are potentially subject to

natural selection (Antao *et al.*, 2008). In our analysis, 1,000,000 simulations were run assuming a Stepwise Mutation Model (SMM).

Population structure was studied by using non-hierarchical and hierarchical F-statistics (Weir & Cockerham, 1984) calculated with ARLEQUIN ver. 3.5.1.3 (Excoffier & Lischer, 2010) and significance was assessed by 10,000 permutation tests. In all calculations performed, significance threshold ($\alpha = 0.05$) was adjusted, when needed, using a standard Bonferroni correction for multiple tests (Rice, 1989).

To estimate the level of recent gene flow between geographic localities, we used the Bayesian method of Rannala & Mountain (1997), implemented in GENECLASS2 ver. 2.0.h (Piry *et al.*, 2004), to detect individuals that were immigrants, or had recent immigrant ancestry, based on their multilocus genotypes. The test has power to detect immigrant ancestors up to two generations in the past even when the overall differentiation of allele frequencies among populations is low. Because of genetic homogeneity, samples from Charcot Island and Marguerite Bay were treated as a single population sample. To avoid bias introduced by temporal variability we considered only samples collected in 2010. For each individual, the relative probability that it was a recent immigrant or a resident in the population where it was sampled was calculated. Assignment probabilities were computed by simulating 10,000 individuals using the Markov Chain Monte Carlo (MCMC) resampling algorithm developed by Paetkau *et al.* (2004) and individuals with probabilities above 0.8 were considered as correctly assigned (Schunter *et al.*, 2011). A test of equal proportions, as performed

in R (Ihaka & Gentleman, 1996), was used to examine whether a significantly higher proportion of migrants occurred at Charcot Island-Marguerite Bay than at Joinville Island.

In addition to migrant identification with GENECLASS, recent migration rates between the 2010 Charcot Island-Marguerite Bay and the Joinville Island sample were estimated by using the multilocus non-equilibrium Bayesian method implemented in BAYESASS (Wilson & Rannala, 2003). Four independent runs, initialized with different seeds, were performed, with each run made up of 50,000,000 MCMC iterations and 5,000,000 burn-in iterations.

Following authors' recommendations, the posterior mean parameter estimates of the independent runs were checked for concordance and identical results were obtained.

RESULTS

Microsatellite analysis: genetic variability and genetic population structure

Analyzed samples are reported in **Table 1**.

Both in 2010 and in 2011 sampling cruises, no silverfish were found in the central western AP at the sampling areas off Renaud and Anvers Island, and in Croker Passage (**Fig. 1**).

Sixteen microsatellites were successfully amplified and genotyped in all 562 individuals, with no missing genotypes. This high performance of amplification indicates a good transfer ability of microsatellite loci isolated from *Chionodraco hamatus* to *Pleuragramma antarcticum* despite the relatively large phylogenetic distance between the two species (approximately 20 millions of years, Near *et al.*, 2012).

All population samples showed relatively high levels of genetic variation at 16 microsatellite loci (**Table 2; Appendix S1 in Supporting Information**) and no significant difference in N_A , A_R , H_O , and H_E was found across samples using one-way ANOVA (p-value > 0.05). No significant genotypic linkage disequilibrium was observed between any pair of loci (data not shown) and all loci were in HWE in all population samples after Bonferroni correction for multiple

Table 2. Genetic variability of *Pleuragramma antarcticum* population samples collected along the Antarctic Peninsula shelf at 16 microsatellite loci. Reported are number of alleles (N_A), allelic richness (A_R) based on a minimum sample size of 28 diploid individuals, observed heterozygosity (H_O), and unbiased expected heterozygosity (H_E). Standard deviation in parentheses. Samples: Charcot Island 2010 (CI10), Marguerite Bay 2001 (MB01), Marguerite Bay 2002 (MB02), Marguerite Bay 2010 (MB10), Marguerite Bay 2011 (MB11), Joinville Island 2007 (JI07), Joinville Island 2010 (JI10), Joinville Island 2012 (JI12), and Larsen Bay 2007 (LB07).

Population sample	$N_A \pm SD$	$A_R \pm SD$	$H_O \pm SD$	$H_E \pm SD$
CI10	8.31 (4.85)	7.01 (4.08)	0.5948 (0.2071)	0.6311 (0.2151)
MB01	7.50 (4.55)	7.50 (4.55)	0.6183 (0.2684)	0.6375 (0.2433)
MB02	8.13 (4.15)	7.01 (3.77)	0.6327 (0.2410)	0.6310 (0.2338)
MB10	8.69 (4.96)	7.33 (3.92)	0.6646 (0.2096)	0.6588 (0.2079)
MB11	9.13 (5.24)	7.07 (3.91)	0.6333 (0.2218)	0.6314 (0.2293)
JI07	7.56 (4.59)	7.20 (4.23)	0.6526 (0.2261)	0.6462 (0.2170)
JI10	10.88 (5.58)	7.39 (3.57)	0.6503 (0.1927)	0.6653 (0.1944)
JI12	7.94 (3.92)	6.84 (3.49)	0.6123 (0.2208)	0.6380 (0.2123)
LB07	7.69 (4.19)	6.95 (3.72)	0.6196 (0.2360)	0.6335 (0.2247)

tests (**Appendix S1 in Supporting Information**). No outlier loci were identified using LOSITAN, that is, none of the 16 microsatellites showed significantly higher or lower F_{ST} values compared with neutral expectations.

Population pairwise F_{ST} (**Table 3**) showed no genetic differentiation between temporal samples from both Marguerite Bay and Joinville Island, suggesting genetic stability on a temporal scale in the two locations. On a geographic scale, a substantial genetic homogeneity was found between samples collected off Charcot Island and in Marguerite Bay, indicating a single gene pool in this area. A more complex pattern was found at Joinville Island with respect to samples from the southwestern AP: Joinville Island 2007 showed no significant difference when compared to any other population

2001 and 2010 was lost after Bonferroni correction), with Joinville Island 2012 showing a higher differentiation than the 2010 sample in all pairwise comparisons. Finally, Larsen Bay 2007 was significantly different from Marguerite Bay 2011, and positive but not significant F_{ST} values after Bonferroni correction were detected against Charcot Island 2010, Marguerite Bay 2010, and Joinville Island 2012 samples (statistical significance was lost). A hierarchical AMOVA showed significant genetic differentiation among geographical groups ($F_{ST} = 0.0038$ p-value < 0.0001 , $F_{SC} = 0.0002$ p-value = 0.4233, $F_{CT} = 0.0035$ p-value = 0.0098).

Fish lengths' heterogeneity was found in the Joinville Island 2010 sample, contrary to both Charcot Island and Marguerite Bay 2010 samples for which a single length mode was found. For this reason, we tested

Table 3. Genetic differentiation among population samples based on microsatellite data. Reported are pairwise F_{ST} estimates (above diagonal) and corresponding p-values (below diagonal). Samples: Charcot Island 2010 (CI10), Marguerite Bay 2001 (MB01), Marguerite Bay 2002 (MB02), Marguerite Bay 2010 (MB10), Marguerite Bay 2011 (MB11), Joinville Island 2007 (JI07), Joinville Island 2010 (JI10), Joinville Island 2012 (JI12), and Larsen Bay 2007 (LB07).

	CI10	MB01	MB02	MB10	MB11	JI07	JI10	JI12	LB07
CI10	-	-0.0011	-0.0008	0.0007	-0.0001	-0.0009	0.0065	0.0102	0.0061
MB01	0.7027	-	-0.0015	-0.0026	-0.0015	0.0017	0.0057	0.0086	0.0063
MB02	0.6216	0.7207	-	0.0005	0.0009	0.0007	0.0037	0.0093	0.0035
MB10	0.3333	0.8018	0.4054	-	0.0004	-0.0028	0.0017	0.0062	0.0029
MB11	0.5496	0.6757	0.2342	0.2793	-	-0.0004	0.0041	0.0060	0.0089
JI07	0.7838	0.2973	0.3874	0.8559	0.5315	-	-0.0015	0.0008	0.0001
JI10	<0.0001**	0.0180*	<0.0001**	0.0541	<0.0001**	0.8108	-	0.0020	0.0020
JI12	<0.0001**	0.0090*	<0.0001**	0.0270*	<0.0001**	0.3874	0.0991	-	0.0057
LB07	0.0180*	0.0991	0.0811	0.0451*	<0.0001**	0.4955	0.0991	0.0090*	-

*P-value < 0.05 ; **p-value < 0.0014 (significance threshold after standard Bonferroni correction for multiple tests with 36 comparisons)

sample, while samples collected at Joinville Island in 2010 and 2012 were both significantly different from Charcot Island and from Marguerite Bay samples collected in 2002 and 2011 (statistical significance of F_{ST} against Marguerite Bay

for genetic differentiation between clusters of individuals from Joinville Island 2010 with different length mode. Two slightly but significantly differentiated gene pools were detected ($F_{ST} = 0.0032$, p-value = 0.0227) when individuals were clustered

based on a SL of 10 cm, which corresponds to the transition from cohort 3+ to cohort 4+ (Hubold & Tomo, 1989). The “Joinville Island 2010 small cluster” (SL<10 cm) showed a higher genetic differentiation than the “Joinville Island 2010 large cluster” (SL>10 cm) when pairwise compared to other population samples (**Table 4**). No difference in the level of genetic variation was detected between the two clusters (one-way ANOVA, p-value > 0.05) and both were in HWE (p-value > 0.05).

For the GENECLASS assignment tests, 58 individuals out of 268 fish sampled in 2010 were assigned to one of the two differentiated samples (p-value > 0.8, see methods). Of the 33 individuals collected at Charcot Island-Marguerite Bay, which were allocated with high confidence, 20 (61%) were genetically assigned to Joinville Island, indicating recent immigration from Joinville Island to Charcot Island-Marguerite Bay, while 13 (39%) were assigned to Charcot Island-Marguerite Bay, indicating no recent immigration in their ancestry. Of the 25 individuals collected at Joinville Island, which were allocated with high confidence, 6 (24%) were assigned to Charcot Island-Marguerite Bay, while 19 (76%) were assigned to Joinville Island (**Table 5**). The test of equal proportions

Table 5. Results of the GENECLASS assignment test performed on population samples collected in 2010. The table reports the number of individuals genetically assigned to Charcot Island-Marguerite Bay (CI-MB, Charcot Island 2010 and Marguerite Bay 2010 pooled samples) and Joinville Island (JI, Joinville Island 2010 sample) for each collection site.

Collected at	Assigned to			Total
	CI-MB	JI		
CI-MB	13	20		33
JI	6	19		25

Table 4. Genetic differentiation between the two clusters of individuals (“JI10 small” and “JI10 large”) within the Joinville Island 2010 (JI) population sample based on a standard length (SL) of 10 cm, and other population samples. Reported are pairwise F_{ST} estimates and corresponding p-values (in parenthesis). Samples: Charcot Island 2010 (CI10), Marguerite Bay 2001 (MB01), Marguerite Bay 2002 (MB02), Marguerite Bay 2010 (MB10), Marguerite Bay 2011 (MB11), Joinville Island 2007 (JI07), Joinville Island 2010 (JI10), Joinville Island 2012 (JI12), and Larsen Bay 2007 (LB07).

	JI10 small (SL < 10 cm)	JI10 large (SL ≥ 10 cm)
CI10	0.0083 (<0.0001**)	0.0067 (<0.0001**)
MB01	0.0082 (<0.0001**)	0.0054 (0.0270*)
MB02	0.0057 (<0.0001**)	0.0032 (0.0721)
MB10	0.0038 (0.0451*)	0.0011 (0.2883)
MB11	0.0060 (<0.0001**)	0.0045 (<0.0001**)
JI07	0.0022 (0.1441)	-0.0036 (0.9820)
JI12	0.0034 (0.0270*)	0.0026 (0.0991)
LB07	0.0050 (0.0180*)	0.0001 (0.4505)

*P-value < 0.05; **p-value < 0.0031 (significance threshold after standard Bonferroni correction for multiple tests with 16 comparisons)

found a significantly higher proportion of migrants at Charcot Island-Marguerite Bay than at Joinville Island (p-value = 0.0060) consistent with a prevailing gene flow along the anticlockwise CC. Accordingly, migration rates inferred by BAYESASS showed a higher flux of migrants from Joinville Island to Charcot Island-Marguerite Bay (0.3289 ± 0.0031) than in the opposite direction (0.0028 ± 0.0028). Notably, however, none of the individuals identified in either area as having recent immigration in their ancestry was a first generation migrant.

DISCUSSION

Antarctic silverfish show genetic structuring

Testing between hypotheses of panmixia and genetic structuring in *Pleuragramma antarcticum*, along the AP shelf using a panel of 16 EST-linked polymorphic microsatellites, we found no genetic differentiation between population samples from Charcot Island taken in 2010 and samples from Marguerite Bay collected in 2001, 2002, 2010, and 2011. This indicated a single gene pool in the southwestern AP and an absence of inter-annual variability. No genetic difference was found also between temporal samples from Joinville Island taken in 2007, 2010, and 2012. Yet in contrast, on a geographic scale significant differentiation was detected between Joinville Island and the southwestern AP (Charcot Island and Marguerite Bay), indicating that Antarctic silverfish forms separate gene pools in the two areas, thus discounting the hypothesis of panmixia.

Moreover, the population sample collected at Joinville Island in 2007 showed no differences when compared to any of the analyzed samples, while samples taken in 2010 and 2012 showed significant differences with respect to both Charcot Island and Marguerite Bay, with Joinville Island 2012 displaying the highest genetic differentiation. Additionally, if we clustered individuals from Joinville Island 2010 based on a standard length of 10 cm (transition from cohort 3+ to 4+), “Joinville Island 2010 small cluster” showed a higher genetic differentiation than “Joinville Island 2010 large cluster” when pairwise compared to samples from the southwestern AP. These data imply an increasing genetic differentiation with time

at Joinville Island, possibly arising after 2007. Similarly, the sample from Larsen Bay 2007 showed, when pairwise compared to Marguerite Bay 2001, 2002 and Joinville Island 2007, a mean F_{ST} of 0.0033 and no significant values after Bonferroni correction. Yet in contrast, a significant mean F_{ST} of 0.0051 was found when Larsen Bay was compared to Charcot Island 2010, Marguerite Bay 2010 and 2011, and to Joinville Island 2010 and 2012.

These findings suggest that some differentiation processes have been ongoing between the southwestern and the north AP since 2007, potentially adding information and clarifying the population genetic structure of Antarctic silverfish found by Zane *et al.* (2006). These authors used partial sequencing of the D-loop mitochondrial region to analyze several population samples from different locations around Antarctica and, possibly due to the low power of the molecular marker used, found no sign of between-groups variation when pooling samples from the same geographic area. In contrast, by using a large number of microsatellites and focusing on a specific geographic area we found significant genetic differentiation among geographical groups. Additionally, when population samples were pairwise compared, Zane *et al.* (2006) found weak and sporadic differences between samples which did not reveal an evident pattern of genetic structuring but rather seemed to indicate random fluctuations of gene pools over time or, alternatively, the inability of the marker to detect existing differences. Instead, our results not only show genetic differentiation on a local geographic scale, but also suggest that the level of genetic

differentiation has increased on a relatively recent time scale.

Migration patterns discount transport by the ACC

Our results indicate the existence of local differentiation in *P. antarcticum*, despite being the only Antarctic notothenioid with a completely pelagic life cycle and thus the highest potential for dispersal. Recent migration patterns along the AP shelf, as established by the assignment test and the migration rate analysis performed in this study, showed a significantly higher flow of migrants moving from Joinville Island towards Charcot Island and Marguerite Bay, than in the opposite direction. Therefore, migration of *P. antarcticum* appears to be mainly southward along the western AP shelf, inconsistent with transport by ACC flow northward along the shelf break.

This pattern is quite different from that found by Matschiner *et al.* (2009) analyzing populations of *Gobionotothen gibberifrons* sampled in the Scotia Sea and around the tip of the AP. The authors found genetic homogeneity and unidirectional west-to-east gene flow, consistent with a circulation mainly mediated by the ACC that prevented genetic differentiation. Because of its large-scale circumpolar flow, they also predicted that the lack of genetic structuring could be a general pattern in Antarctic fish. However, this is not borne out by our results. The paradox may be explained by the fact that the analyzed area of this study is different from that in the study by Matschiner *et al.* (2009). Evidence from drifters and models indicate that, in contrast to the Scotia Sea, coastal

waters from the tip of the Peninsula are transported southward along the western AP by the CC. Conceivably, the ACC or outer shelf processes might transport fish northwards along the shelf break past Anvers Island, but there are no obvious transport pathways from there to the tip of the Peninsula (Thompson *et al.*, 2009; Piñones *et al.*, 2011).

However, recent levels of southward gene flow seem to have been insufficient to prevent population differentiation. Lagrangian particle tracking simulations also showed that the circulation of the western AP shelf is influenced by several clockwise gyres, of which the main are in Crystal Sound, in the Laubeuf Fjord, and off Alexander Island (Piñones *et al.*, 2011). These gyres, which were shown to favor local retention of pelagic and planktonic species (Piñones *et al.*, 2011), might hinder the transport of larvae and juveniles of Antarctic silverfish. Changes in population structure, and relationships with transport and retention may account for reductions in the level of gene flow between the northern and the southwestern AP, in accordance with genetic differences found between population samples in this study. Notably, during sample collection in both 2010 and 2011, no Antarctic silverfish was found in the central western AP, despite its occurrence has been historically reported in this area (Kellermann, 1996). This finding, in line with the reported reduction of *P. antarcticum* larval catches central western AP (Ducklow *et al.*, 2007; Moline *et al.*, 2008) and its recent disappearance from the diet of Adelie penguins at Palmer station would support the hypothesis of recent discontinuity between populations

of the northern and the southern regions of the western AP.

Gene flow and climate change

The western AP is one of the fastest warming areas on the planet. Mid-winter temperatures have increased by 5-6 °C over the last 50 years (more than five times the global average) and seasonal sea-ice cover is declining dramatically (Smith *et al.*, 2007a). Although the exact role of sea-ice in the life cycle of Antarctic silverfish remains to be fully elucidated, observations of *P. antarcticum* associated with the underside of sea-ice and platelet layer in various phases of its life cycle, indicate a strong dependence of the species on this peculiar microhabitat (Vacchi *et al.*, 2012). Sea-ice cover is probably used as spawning ground and nursery area, offering large food availability and protecting early life stages from predation (Vacchi *et al.*, 2012).

The apparent disappearance of Antarctic silverfish in the central western AP and the increased genetic differentiation with time detected between the north and the southwestern AP, as found in this study, indicate that *P. antarcticum* may be highly vulnerable to recent global warming. Indices of genetic diversity, that showed a similar level of variation in all analyzed samples, and the reduced migration implied by the lack of first generation migrants, suggest that the observed pattern of genetic differentiation is recent. In a context of increased population fragmentation genetic drift drives alleles to fixation or extinction within a relatively short amount of time, possibly leading to erosion of genetic diversity. Low genetic variability means a lower potential to

adapt and survive to climate change and other possible environmental disturbances in the future. Our results suggest that processes promoting fragmentation and differentiation may have entered a newly dynamic phase. In consequence, this study should represent only the starting point for other in-depth analyses to survey and monitor the level of genetic variation and the pattern of genetic differentiation of Antarctic silverfish around the continental margins of Antarctica.

ACKNOWLEDGMENTS

We thank E. Bortolotto and G. Santovito for providing *P. antarcticum* samples during the NBP 10-02 oceanographic campaign, performed on board the R/V Nathaniel B. Palmer and coordinated by Prof. Joseph Torres (College of Marine Science, University of South Florida). This work was supported by the National Program for Antarctic Research (PNRA) to TP, and LZ. CA is a PhD student in Evolutionary Biology at the University of Padova, with a program partially supported under NSF grant 0741348.

REFERENCES

- Amos, W., Hoffman, J.I., Frodsham, A., Zhang, L., Best, S. & Hill, A.V.S. (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, **7**, 10-14.
- Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A. & Luikart, G. (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.
- Ashford, J., Ferguson, J., Piñones, A., Torres, J. & Fraser, W. (2010a)

- Preliminary report: Connectivity and population structure in *Pleuragramma antarcticum* along the West Antarctic Peninsula. *CCAMLR Scientific Papers WG-FSA-10/16*.
- Ashford, J., La Mesa, M., Fach, B.A., Jones, C. & Everson, I. (2010b) Testing early life connectivity using otolith chemistry and particle-tracking simulations. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 1303-1315.
- Ashford, J.R., Arkhipkin, A.I. & Jones, C.M. (2006) Can the chemistry of otolith nuclei determine population structure of Patagonian toothfish *Dissostichus eleginoides*? *Journal of Fish Biology*, **69**, 708-721.
- Ashford, J.R., Jones, C.M., Hofmann, E.E., Everson, I., Moreno, C.A., Duhamel, G. & Williams, R. (2008) Otolith chemistry indicates population structuring by the Antarctic Circumpolar Current. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 135-146.
- Beaumont, M.A. & Nichols, R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619-1626.
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).
- Convey, P., Bindschadler, R., di Prisco, G., Fahrback, E., Gutt, J., Hodgson, D.A., Mayewski, P.A., Summerhayes, C.P., Turner, J. & Consortium, A. (2009) Antarctic climate change and the environment. *Antarctic Science*, **21**, 541-563.
- Coppe, A., Agostini, C., Marino, I.A.M., Zane, L., Bargelloni, L., Bortoluzzi, S. & Patarnello, T. (2013) Genome Evolution in the Cold: Antarctic Icefish Muscle Transcriptome Reveals Selective Duplications Increasing Mitochondrial Function. *Genome Biology and Evolution*, **5**, 45-60.
- Damerau, M., Matschiner, M., Salzburger, W. & Hanel, R. (2012) Comparative population genetics of seven notothenioid fish species reveals high levels of gene flow along ocean currents in the southern Scotia Arc, Antarctica. *Polar Biology*, **35**, 1073-1086.
- DeWitt, H.H., Heemstra, P.C. & Gon, O. (1990) Nototheniidae. *Fishes of the Southern Ocean* (ed. by O. Gon and P.C. Heemstra), JLB Smith Institute of Ichthyology, Grahamstown.
- Donnelly, J. & Torres, J.J. (2008) Pelagic fishes in the Marguerite Bay region of the West Antarctic Peninsula continental shelf. *Deep-Sea Research Part II*, **55**, 523-539.
- Ducklow, H.W., Baker, K., Martinson, D.G., Quetin, L.B., Ross, R.M., Smith, R.C., Stammerjohn, S.E., Vernet, M. & Fraser, W. (2007). Marine pelagic ecosystems: the West Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**, 67-94.
- Eastman, J.T. (1993) *Antarctic fish biology: evolution in a unique environment*. Academic Press.
- Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.

- Molecular Ecology Resources*, **10**, 564-567.
- Fach, B.A. & Klinck, J.M. (2006) Transport of Antarctic krill (*Euphausia superba*) across the Scotia Sea. Part I: Circulation and particle tracking simulations. *Deep-Sea Research Part I-Oceanographic Research Papers*, **53**, 987-1010.
- Goudet, J. (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995).
- Hauser, L. & Carvalho, G.R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333-362.
- Hofmann, E.E., Klinck, J.M., Locarnini, R.A., Fach, B. & Murphy, E. (1998) Krill transport in the Scotia Sea and environs. *Antarctic Science*, **10**, 406-415.
- Hubold, G. & Tomo, A.P. (1989) Age and growth of Antarctic silverfish *Pleuragramma antarcticum* Boulenger, 1902, from the southern Weddell Sea and Antarctic Peninsula. *Polar Biology*, **9**, 205-212.
- Ihaka, R. & Gentleman, R. (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, **5**, 299-314.
- Kellermann, A.K. (1996) Midwater Fish Ecology. *Foundations for Ecological Research West of the Antarctic Peninsula* (ed. by E.E. Hofmann, R.M. Ross and L.B. Quetin), Antarctic Research Series, vol. 70, pp. 231-256, AGU, Washington, D. C., doi:10.1029/AR070.
- Klinck, J.M. & Nowlin, W.D., Jr. (2001) Antarctic Circumpolar Current. *Encyclopedia of ocean sciences* (ed. by J. Steele, S. Thorpe and K. Turekian). Academic Press, New York.
- La Mesa, M., Catalano, B., Russo, A., Greco, S., Vacchi, M. & Azzali, M. (2010) Influence of environmental conditions on spatial distribution and abundance of early life stages of Antarctic silverfish, *Pleuragramma antarcticum* (Nototheniidae), in the Ross Sea. *Antarctic Science*, **22**, 243-254.
- La Mesa, M. & Eastman, J.T. (2011) Antarctic silverfish: life strategies of a key species in the high-Antarctic ecosystem. *Fish and Fisheries*, **13**, 241-266.
- La Mesa, M., Eastman, J.T. & Vacchi, M. (2004) The role of notothenioid fish in the food web of the Ross Sea shelf waters: a review. *Polar Biology*, **27**, 321-338.
- Matschiner, M., Hanel, R. & Salzburger, W. (2009) Gene flow by larval dispersal in the Antarctic notothenioid fish *Gobionotothen gibberifrons*. *Molecular Ecology*, **18**, 2574-2587.
- Molecular Ecology Resources Primer Development Consortium, Agostini, C., Agudelo, P.A., BÂ, K., *et al.* (2011) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2010-30 November 2010. *Molecular Ecology Resources*, **11**, 418-421.
- Molecular Ecology Resources Primer Development Consortium, Agostini, C., Albaladejo, R.G., Aparicio, A., *et al.* (2013) Permanent Genetic Resources added to Molecular Ecology Resources Database April 2013-31 May 2013. *Molecular Ecology Resources*, **13**, 966-968.

- Moline, M.A., Karnovsky, N.J., Brown, Z., Divoky, G.J., Frazer, T.K., Jacoby, C.A., Torres, J.J. & Fraser, W.R. (2008) High latitude changes in ice dynamics and their impact on polar marine ecosystems. *Annals of the New York Academy of Sciences*, **1134**, 267-319.
- Near, T.J., Dornburg, A., Kuhn, K.L., Eastman, J.T., Pennington, J.N., Patarnello, T., Zane, L., Fernandez, D.A. & Jones, C.D. (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences*, **109**, 3434-3439.
- O'Connor, M.I., Bruno, J.F., Gaines, S.D., Halpern, B.S., Lester, S.E., Kinlan, B.P. & Weiss, J.M. (2007) Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *Proceedings of the National Academy of Sciences*, **104**, 1266-1271.
- Orsi, A.H., Whitworth, T. & Nowlin, W.D. (1995) On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Research Part I-Oceanographic Research Papers*, **42**, 641-673.
- Paetkau, D., Slade, R., Burden, M. & Estoup, A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55-65.
- Papetti, C., Pujolar, J.M., Mezzavilla, M., La Mesa, M., Rock, J., Zane, L. & Patarnello, T. (2012) Population genetic structure and gene flow patterns between populations of the Antarctic icefish *Chionodraco rastrispinosus*. *Journal of Biogeography*, **39**, 1361-1372.
- Papetti, C., Susana, E., Patarnello, T. & Zane, L. (2009) Spatial and temporal boundaries to gene flow between *Chaenocephalus aceratus* populations at South Orkney and South Shetlands. *Marine Ecology Progress Series*, **376**, 269-281.
- Patarnello, T., Marcato, S., Zane, L., Varotto, V. & Bargelloni, L. (2003) Phylogeography of the *Chionodraco* genus (Perciformes, Channichthyidae) in the Southern Ocean. *Molecular Phylogenetics and Evolution*, **28**, 420-429.
- Patarnello, T., Verde, C., di Prisco, G., Bargelloni, L. & Zane, L. (2011) How will fish that evolved at constant sub-zero temperatures cope with global warming? Notothenioids as a case study. *Bioessays*, **33**, 260-268.
- Patwary, M.U., Kenchington, E.L., Bird, C.J. & Zouros, E. (1994) The use of random amplified polymorphic DNA markers in genetic-studies of the sea-scallop *Placopecten magellanicus* (Gmelin, 1791). *Journal of Shellfish Research*, **13**, 547-553.
- Piñones, A., Hofmann, E.E., Dinniman, M.S. & Klinck, J.M. (2011) Lagrangian simulation of transport pathways and residence times along the western Antarctic Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, **58**, 1524-1539.
- Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L. & Estoup, A. (2004) GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity*, **95**, 536-539.
- Rannala, B. & Mountain, J.L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences*, **94**, 9197-9201.

- Raymond, M. & Rousset, F. (1995) GENEPOP (Version-1.2) - population-genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.
- Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.
- Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.
- Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintore, J., Vidal-Vijande, E., Pascual, A., Guidetti, P. & Pascual, M. (2011) Matching genetics with oceanography: directional gene flow in a Mediterranean fish species. *Molecular Ecology*, **20**, 5167-5181.
- Smith, J.A., Bentley, M.J., Hodgson, D.A. & Cook, A.J. (2007a) George VI Ice Shelf: past history, present behaviour and potential mechanisms for future collapse. *Antarctic Science*, **19**, 131-142.
- Smith, W.O., Ainley, D.G. & Cattaneo-Vietti, R. (2007b) Trophic interactions within the Ross Sea continental shelf ecosystem. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**, 95-111.
- Thompson, A.F., Heywood, K.J., Thorpe, S.E., Renner, A.H.H. & Trasviña, A. (2009) Surface Circulation at the Tip of the Antarctic Peninsula from Drifters. *Journal of Physical Oceanography*, **39**, 3-26.
- Thorpe, S.E., Murphy, E.J. & Watkins, J.L. (2007) Circumpolar connections between Antarctic krill (*Euphausia superba* Dana) populations: Investigating the roles of ocean and sea ice transport. *Deep-Sea Research Part I-Oceanographic Research Papers*, **54**, 792-810.
- Vacchi, M., DeVries, A.L., Evans, C.W., Bottaro, M., Ghigliotti, L., Cutroneo, L. & Pisano, E. (2012) A nursery area for the Antarctic silverfish *Pleuragramma antarcticum* at Terra Nova Bay (Ross Sea): first estimate of distribution and abundance of eggs and larvae under the seasonal sea-ice. *Polar Biology*, **35**, 1573-1585.
- Vacchi, M., La Mesa, M., Dalu, M. & Macdonald, J. (2004) Early life stages in the life cycle of Antarctic silverfish, *Pleuragramma antarcticum* in Terra Nova Bay, Ross Sea. *Antarctic Science*, **16**, 299-305.
- Waters, K.J. & Smith, R.C. (1992) Palmer LTER: A sampling grid for the Palmer LTER program. *Antarctic Journal of the United States*, **27**, 236-239.
- Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.
- Wilson, G.A. & Rannala, B. (2003) Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics*, **163**, 1177-1191.
- Zane, L., Marcato, S., Bargelloni, L., Bortolotto, E., Papetti, C., Simonato, M., Varotto, V. & Patarnello, T. (2006) Demographic history and population structure of the Antarctic silverfish *Pleuragramma antarcticum*. *Molecular Ecology*, **15**, 4499-4511.
- Zazulie, N., Rusticucci, M. & Solomon, S. (2010) Changes in Climate at High Southern Latitudes: A Unique Daily Record at Orcadas Spanning 1903-2008. *Journal of Climate*, **23**, 189-196.

SUPPORTING INFORMATION

Appendix S1. Number of alleles (N_A), allelic richness (A_R) based on a minimum sample size of 28 diploid individuals, observed and unbiased expected heterozygosity (H_O/H_E), and Hardy-Weinberg equilibrium probability (pHWE) for each locus across population samples. Samples: Charcot Island 2010 (CI10), Marguerite Bay 2001 (MB01), Marguerite Bay 2002 (MB02), Marguerite Bay 2010 (MB10), Marguerite Bay 2011 (MB11), Joinville Island 2007 (JI07), Joinville Island 2010 (JI10), Joinville Island 2012 (JI12), and Larsen Bay 2007 (LB07).

Locus	CI10			MB01			MB02			MB10			MB11		
	N_A	A_R	H_O/H_E pHWE												
Ch10105	6	4.74	0.3000/ 0.3111	3	3.00	0.0714/ 0.0708	5	3.96	0.2449/ 0.2411	4	3.64	0.2167/ 0.2157	5	4.20	0.2530/ 0.2332
Ch10441	9	7.08	0.6500/ 0.6555	10	10.00	0.8571/ 0.7721	9	7.85	0.6122/ 0.6840	11	9.28	0.7333/ 0.7143	12	8.04	0.6867/ 0.6873
Ch10857	9	7.92	0.6833/ 0.6920	9	9.00	0.5357/ 0.6727	7	6.77	0.5714/ 0.6737	11	8.36	0.7500/ 0.7374	10	7.83	0.7349/ 0.7187
Ch11230	7	5.78	0.7667/ 0.7248	8	8.00	0.7143/ 0.7883	7	6.11	0.7347/ 0.7410	6	5.44	0.8000/ 0.7402	7	5.81	0.7349/ 0.7376
Ch11483	10	7.88	0.4667/ 0.5641	9	9.00	0.5714/ 0.6500	9	6.91	0.4490/ 0.4673	8	6.82	0.6500/ 0.5772	8	5.31	0.5542/ 0.5645
Ch126	3	2.47	0.4500/ 0.5097	3	3.00	0.4643/ 0.5188	4	3.56	0.5918/ 0.5590	4	2.93	0.4333/ 0.5195	3	2.98	0.5783/ 0.5509
Ch13222	3	2.71	0.1667/ 0.1563	2	2.00	0.1786/ 0.1656	4	2.71	0.0612/ 0.0606	4	2.93	0.2500/ 0.2244	3	2.30	0.0964/ 0.1038
Ch17977	8	6.61	0.5833/ 0.6342	4	4.00	0.5714/ 0.5662	7	5.71	0.6531/ 0.6270	9	6.78	0.6167/ 0.6613	10	6.37	0.5783/ 0.5804
Ch18085	10	8.39	0.8333/ 0.8368	9	9.00	0.8571/ 0.8383	9	8.31	0.8571/ 0.8367	10	8.63	0.8333/ 0.8370	11	8.50	0.8675/ 0.8236
Ch19846	8	7.30	0.7000/ 0.7840	8	8.00	0.8571/ 0.7981	8	7.39	0.8571/ 0.8075	8	7.31	0.8667/ 0.8045	8	6.66	0.7108/ 0.7797
Ch24332	6	5.04	0.6000/ 0.6272	5	5.00	0.5714/ 0.6058	5	4.39	0.7347/ 0.6240	6	5.62	0.6000/ 0.6528	7	5.89	0.6867/ 0.6310
Ch25478	4	3.46	0.5667/ 0.5606	4	4.00	0.4643/ 0.5656	6	4.71	0.6122/ 0.6324	6	5.38	0.6833/ 0.6448	8	5.76	0.6386/ 0.6303
Ch2931	10	8.46	0.7333/ 0.8339	10	10.00	0.8571/ 0.8351	11	9.09	0.8367/ 0.8239	10	9.21	0.8333/ 0.8167	10	8.12	0.7349/ 0.8115
Ch4796	13	11.26	0.7500/ 0.8780	12	12.00	0.9643/ 0.8987	14	12.43	0.9388/ 0.9007	13	11.45	0.8167/ 0.8919	14	12.75	0.9277/ 0.9081
Ch520	4	3.68	0.3333/ 0.3828	4	4.00	0.3929/ 0.5182	5	4.49	0.4694/ 0.4869	5	4.68	0.6167/ 0.5619	5	4.27	0.4458/ 0.4021
Ch623	23	19.36	0.9333/ 0.9461	20	20.00	0.9643/ 0.9351	20	17.76	0.8980/ 0.9308	24	18.89	0.9333/ 0.9410	25	18.39	0.9036/ 0.9401

*p-value < 0.05; significance threshold after standard Bonferroni correction for multiple tests with 144 comparisons is 0.000347. — *Table continues on the next page*

Table S1 (continued).

Locus	JI07			JI10			JI12			LB07		
	N_A	A_R	H_O/H_E pHWE									
Ch10105	4	3.94	0.2353/ 0.2195	6	4.46	0.3176/ 0.3713	5	4.60	0.2037/ 0.2553	4	3.54	0.1739/ 0.1839
Ch10441	9	8.41	0.7353/ 0.6712	17	8.53	0.6689/ 0.6976	7	6.18	0.6667/ 0.6632	11	9.13	0.6957/ 0.7117
Ch10857	7	6.62	0.7647/ 0.7463	12	8.32	0.6959/ 0.7405	11	8.56	0.7407/ 0.7150	9	7.77	0.5435/ 0.6333
Ch11230	7	6.79	0.8824/ 0.7669	9	6.23	0.7432/ 0.7429	7	6.03	0.6852/ 0.7096	6	5.89	0.6957/ 0.7496
Ch11483	5	4.79	0.5882/ 0.4864	12	6.48	0.5541/ 0.5846	8	6.02	0.5370/ 0.5762	7	6.05	0.6087/ 0.5602
Ch126	5	4.62	0.5588/ 0.5531	6	3.78	0.6284/ 0.5478	5	3.93	0.5000/ 0.5459	3	2.98	0.6522/ 0.4928
Ch13222	3	2.82	0.2059/ 0.2366	4	2.75	0.1622/ 0.1746	3	2.51	0.1481/ 0.1397	3	2.60	0.0435/ 0.1245
Ch17977	7	6.94	0.6765/ 0.6962	9	6.47	0.6014/ 0.6320	8	6.95	0.6296/ 0.6404	6	5.44	0.5435/ 0.5781
Ch18085	10	9.47	0.7647/ 0.8301	12	8.83	0.8446/ 0.8505	8	7.66	0.8519/ 0.8217	8	7.83	0.8261/ 0.8251
Ch19846	7	6.79	0.7647/ 0.7831	10	7.67	0.8041/ 0.7946	8	7.64	0.7778/ 0.7996	9	8.23	0.8261/ 0.7874
Ch24332	4	4.00	0.6176/ 0.6449	8	6.13	0.6216/ 0.6543	6	4.96	0.5741/ 0.5974	6	5.45	0.5652/ 0.6527
Ch25478	7	6.29	0.4706/ 0.6155	9	5.94	0.6689/ 0.6320	5	4.04	0.5556/ 0.5947	5	4.58	0.6739/ 0.5984
Ch2931	10	9.63	0.7059/ 0.8507	10	8.21	0.7838/ 0.8141	8	7.31	0.6667/ 0.8238	11	10.04	0.7826/ 0.8354
Ch4796	11	10.79	1.0000/ 0.8942	14	11.28	0.8243/ 0.8884	12	10.70	0.9815/ 0.8804	10	9.16	0.9130/ 0.8645
Ch520	3	3.00	0.5000/ 0.4113	8	5.04	0.5743/ 0.5823	6	4.81	0.4444/ 0.5152	5	4.59	0.5000/ 0.6022
Ch623	22	20.28	0.9706/ 0.9333	28	18.16	0.9122/ 0.9380	20	17.49	0.8333/ 0.9301	20	17.97	0.8696/ 0.9360

*p-value < 0.05; significance threshold after standard Bonferroni correction for multiple tests with 144 comparisons is 0.000347.

RIASSUNTO

I nototenioidi antartici (subordine Notothenioidei, ordine Perciformes) si sono evoluti per milioni di anni nelle acque gelide che caratterizzano l'Oceano Meridionale; essi presentano una vasta gamma di adattamenti per resistere al freddo e ora dominano la fauna ittica antartica sia per numero di specie che per biomassa. Per la loro estrema stenotermia, questi pesci potrebbero essere fortemente vulnerabili ai cambiamenti climatici con possibili effetti a cascata sull'intero ecosistema marino antartico. Pertanto, è di fondamentale importanza investigare le basi genetiche e genomiche dell'adattamento al freddo, analizzare i processi di differenziamento derivanti dai cambiamenti climatici del passato e attuali e, allo stesso tempo, indagare il livello di variabilità e di differenziamento genetico presente a livello di specie e di popolazione.

In questo dottorato sono state prese in considerazione quattro specie di nototenioidi antartici: le tre specie di derivazione recente appartenenti al genere *Chionodraco*, *Chionodraco hamatus*, *Chionodraco rastrispinosus* e *Chionodraco myersi*, e *Pleuragramma antarcticum*. Il genere *Chionodraco* appartiene alla famiglia Channichthyidae (icefish), unica tra i vertebrati per l'assenza di emoglobina e l'incapacità di esprimere mioglobina nel muscolo scheletrico. Un adeguato rifornimento di ossigeno ai tessuti è permesso da un marcato rimodellamento del sistema cardio-vascolare e da una densità mitocondriale eccezionalmente elevata a livello muscolare. *P. antarcticum* (Nototheniidae) è l'unico nototeniideo caratterizzato da un ciclo vitale completamente pelagico, è dipendente dal ghiaccio marino in vari stadi del ciclo vitale e svolge un ruolo chiave nella catena trofica dell'ecosistema marino antartico.

Le analisi svolte in questo dottorato possono essere raggruppate in due principali linee di ricerca: 1) l'approfondimento della conoscenza sulle basi genetiche e genomiche dell'adattamento al freddo degli icefish; 2) l'analisi del pattern di differenziamento genetico presente a livello intra- e inter-specifico, ponendo particolare enfasi su come le condizioni ambientali del passato e del presente abbiano plasmato e stiano influenzando la struttura genetica delle specie.

Per quanto riguarda la prima linea di ricerca, è stato ricostruito e annotato il primo trascrittoma normalizzato del muscolo scheletrico di *C. hamatus* e l'informazione di sequenza così ottenuta è stata utilizzata per verificare l'ipotesi di duplicazione di geni coinvolti nella funzione mitocondriale. Utilizzando una pipeline bioinformatica sviluppata *ad hoc*, sono stati identificati 124 geni duplicati specifici del lineage di *C. hamatus*. La proporzione di duplicazioni lineage-specifiche identificate in *C. hamatus* è stata confrontata con quella presente in cinque specie modello di pesci teleostei. È interessante notare come *C. hamatus* abbia mostrato una proporzione significativamente maggiore di geni duplicati andando a confrontare dati ottenuti dal solo trascrittoma del muscolo con i dati genomici delle specie modello. Un'analisi di arricchimento funzionale ha mostrato come l'insieme dei geni duplicati in *C. hamatus* fosse significativamente arricchito in proteine con localizzazione mitocondriale, coinvolte nella funzione e nella biogenesi mitocondriale. La presenza di elevate densità mitocondriali e il mantenimento a livello genomico di geni duplicati con funzione mitocondriale potrebbero conferire un vantaggio

selettivo agli icefish in un ambiente freddo e in assenza di proteine di trasporto per l'ossigeno, migliorando la diffusione dell'ossigeno e la produzione di energia nei tessuti aerobici.

Per quanto riguarda la seconda linea di ricerca, è stato studiato, in primo luogo, il pattern di differenziamento genetico presente a livello intra- e inter-specifico nel genere *Chionodraco*. È stata rilevata omogeneità genetica a livello di popolazione, ma la presenza di tre *pool* genici distinti corrispondenti alle tre specie. Successivamente, sono stati ricercati putativi loci *outlier*, ovvero in grado di rilevare un elevato livello di differenziamento genetico tra specie. Sono stati identificati tre loci, probabilmente soggetti a selezione naturale, che hanno mostrato similarità di sequenza per la calmodulina, per una proteasi *antifreeze glycoprotein/trypsinogen-like* e per un componente fondamentale del *super elongation complex*. Pressioni selettive, agenti su loci specifici, potrebbero riflettere processi evolutivi del passato che hanno portato alla divergenza tra specie e all'adattamento locale.

È stata inoltre investigata la presenza e l'entità del flusso genico, passato e presente, tra le tre specie del genere *Chionodraco*, anche per chiarire il ruolo dei cicli glaciali nel processo di divergenza e di introgressione tra specie. Sono state rilevate molteplici evidenze di introgressione passata e recente: primo, diversi individui in ciascuna specie hanno mostrato una composizione genetica mista; secondo, scenari evolutivi escludenti l'ibridazione o ammettendola solo in tempi antichi hanno mostrato una probabilità a posteriori piccola o nulla; terzo, i dati sono risultati supportati da uno scenario evolutivo comprendente flusso genico interspecifico e associato ai due periodi interglaciali più recenti (Eemiano e Olocene). Questi risultati potrebbero indicare una maggiore opportunità di speciazione allopatrica in rifugi durante i periodi glaciali, seguita da contatti secondari e ibridazione durante gli intervalli più caldi.

Infine, nella specie *P. antarcticum*, è stata studiata la struttura genetica di popolazione a livello della Penisola Antartica, una regione molto influenzata dal surriscaldamento climatico. Lungo la costa sud-occidentale della penisola è stato rilevato un *pool* genico unico e assenza di variabilità su scala temporale. Differenze significative sono state evidenziate, invece, su scala geografica tra campioni raccolti nella regione sud-occidentale e quelli ottenuti dalla punta settentrionale della penisola, con un segnale di incremento del differenziamento nel tempo. Molteplici evidenze, quali il ridotto livello di flusso genico lungo la piattaforma continentale della penisola, l'aumento del differenziamento su scala temporale e l'incapacità di catturare *P. antarcticum* lungo la costa centro-occidentale per due anni consecutivi, suggeriscono che questa specie, così dipendente dal ghiaccio marino, sia stata colpita dai cambiamenti climatici con possibili effetti a cascata sull'intera catena alimentare dell'ecosistema marino antartico.

REFERENCES

- Abele D, Pantarulo S (2004) Formation of reactive species and induction of antioxidant defence systems in polar and temperate marine invertebrates and fish. *Comparative Biochemistry and Physiology. Part A, Molecular and Integrative Physiology*, **138**: 405-415.
- Agostini C, Papetti C, Patarnello T, Mark FC, Zane L, Marino IAM (2013) Putative selected markers in the *Chionodraco* genus detected by interspecific outlier tests. *Polar Biology*, **36**: 1509-1518.
- Albertson RC, Yan YL, Titus TA, Pisano E, Vacchi M, Yelick PC, Detrich HW III, Postlethwait JH (2010) Molecular pedomorphism underlies craniofacial skeletal evolution in Antarctic notothenioid fishes. *BMC Evolutionary Biology*, **10**: 4.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403-410.
- Arrigo KR, Mock T, Lizotte MP (2010) Primary producers and sea ice. In: *Sea ice (2nd Edition)*, (eds Thomas DN, Dieckmann GS), pp 283-325. Wiley-Blackwell, Oxford.
- Barbara T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*, **16**: 3759-3767.
- Bilyk KT, Cheng CH (2013) Model of gene expression in extreme cold-reference transcriptome for the high-Antarctic cryopelagic notothenioid fish *Pagothenia borchgrevinki*. *BMC Genomics*, **14**: 634.
- Bluhm BA, Gradinger RR, Schnack-Schiel SB (2010) Sea ice meio and macrofauna. In: *Sea ice (2nd Edition)*, (eds Thomas DN, Dieckmann GS), pp 357-393. Wiley-Blackwell, Oxford.
- Borley KA, Sidell BD (2010) Evolution of the myoglobin gene in Antarctic icefishes (Channichthyidae). *Polar Biology*, **34**: 659-665.
- Bottaro M, Oliveri D, Ghigliotti L, Pisano E, Ferrando S, Vacchi M (2009) Born among the ice: first morphological observations on two developmental stages of the Antarctic silverfish *Pleuragramma antarcticum*, a key species of the Southern Ocean. *Reviews in Fish Biology and Fisheries*, **19**: 249-259.
- Buckley BA, Place SP, Hofmann GE (2004) Regulation of heat shock genes in isolated hepatocytes from an Antarctic fish, *Trematomus bernacchii*. *Journal of Fish Biology*, **207**: 3649-3656.
- Buckley BA, Somero GN (2009) cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish *Trematomus bernacchii* to alter gene expression in response to heat stress. *Polar Biology*, **32**: 403-415.
- Caron DA, Gast RJ (2010) Heterotrophic protists associated with sea ice. In: *Sea ice (2nd Edition)*, (eds Thomas DN, Dieckmann GS), pp 327-356. Wiley-Blackwell, Oxford.

- Chen L, DeVries AL, Cheng CH (1997) Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America*, **94**: 3811-3816.
- Chen WJ, Bonillo C, Lecointre G (1998) Phylogeny of the Channichthyidae (Notothenioidei, Teleostei) based on two mitochondrial genes. In: *Fishes of Antarctica: A Biological Overview*, (eds di Prisco G, Pisano E, Clarke A), pp 287-298. Springer, Berlin.
- Chen Z, Cheng CH, Zhang J, Cao L, Chen L, Zhou L, Jin Y, Ye H, Deng C, Dai Z, Xu Q, Hu P, Sun S, Shen Y, Chen L (2008) Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America*, **105**: 12944-12949.
- Cheng CH, Chen LB (1999) Evolution of an antifreeze glycoprotein. *Nature*, **401**: 443-444.
- Cheng CH, Detrich HW III (2007) Molecular ecophysiology of Antarctic notothenioid fishes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **362**: 2215-2232.
- Chinnusamy V, Zhu J, Zhu JK (2007) Cold stress regulation of gene expression in plants. *Trends in Plant Science*, **12**: 444-451.
- Clarke A, Murphy EJ, Meredith MP, King JC, Peck LS, Barnes DK, Smith RC (2007) Climate change and the marine ecosystem of the western Antarctic Peninsula. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **362**: 149-66.
- Cocca E, Ratnayake-Lecamwasam M, Parker SK, Camardella L, Ciaramella M, di Prisco G, Detrich HW III (1995) Genomic remnants of alpha-globin genes in the hemoglobinless Antarctic icefishes. *Proceedings of the National Academy of Sciences of the United States of America*, **92**: 1817-1821.
- Coppe A, Agostini C, Marino IAM, Zane L, Bargelloni L, Bortoluzzi S, Patarnello T (2013) Genome evolution in the cold: Antarctic icefish muscle transcriptome reveals selective duplications increasing mitochondrial function. *Genome Biology and Evolution*, **5**: 45-60.
- Cziko PA, Evans CW, Cheng C-HC, DeVries AL (2006) Freezing resistance of antifreeze-deficient larval Antarctic fish. *The Journal of Experimental Biology*, **209**: 407-420.
- Daniels RA, Lipps JH (1982) Distribution and ecology of fishes of the Antarctic Peninsula. *Journal of Biogeography*, **9**: 1-9.
- DeConto RM, Pollard D (2003) Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO₂. *Nature*, **421**: 245-249.
- Detrich HW III, Amemiya CT (2010) Antarctic notothenioid fishes: genomic resources and strategies for analyzing an adaptive radiation. *Integrative and Comparative Biology*, **50**: 1009-1017.
- Detrich HW III, Parker SK, Williams RC Jr, Nogales E, Downing KH (2000) Cold

adaptation of microtubule assembly and dynamics. Structural interpretation of primary sequence changes present in the alpha- and beta-tubulins of Antarctic fishes. *The Journal of Biological Chemistry*, **275**: 37038-37047.

DeVries AL, Eastman JT (1978) Lipid sacs as a buoyancy adaptation in an Antarctic fish. *Nature*, **271**: 352-353.

DeWitt HH (1970) The character of the midwater fish fauna of the Ross Sea, Antarctica. In: *Antarctic Ecology*, Vol 1 (eds Holdgate MW), pp 305-314. Academic Press, London.

DeWitt HH, Heemstra PC, Gon O (1990) Nototheniidae. In: *Fishes of the Southern Ocean*, (eds Gon O and Heemstra PC), pp 279-331. JLB Smith Institute of Ichthyology, Grahamstown, South Africa.

Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF (2009) Roles for *Arabidopsis* CAMTA Transcription Factors in Cold-Regulated Gene Expression and Freezing Tolerance. *The Plant Cell*, **21**: 972-984.

Eastman JT (1993) *Antarctic Fish Biology: Evolution in a Unique Environment*. Academic, San Diego.

Eastman JT (1997) Phyletic divergence and specialization for pelagic life in the Antarctic notothenioid fish *Pleuragramma antarcticum*. *Comparative Biochemistry and Physiology*, **118A**: 1095-1101.

Eastman JT (2000) Antarctic notothenioid fishes as subjects for research in evolutionary biology. *Antarctic Science*, **12**: 276-287.

Eastman JT (2005) The nature of the diversity of Antarctic fishes. *Polar Biology*, **28**: 93-107.

Eastman JT, Barrera-Oro E (2010) Buoyancy studies of three morphs of the Antarctic fish *Trematomus newnesi* (Nototheniidae) from the South Shetland Islands. *Polar Biology*, **33**: 823-831.

Eastman JT, Eakin RR (2000) An updated species list for Notothenioid fish (Perciformes; Notothenioidei), with comments on Antarctic species. *Archive of Fishery and Marine Research*, **48**: 11-20.

Eastman JT, McCune AR (2000) Fishes on the Antarctic continental shelf: evolution of a marine species flock? *Journal of Fish Biology*, **57**(Suppl. A): 84-102.

Faleeva TI, Gerasimchuk VV (1990) Features of reproduction in the Antarctic sidestripe, *Pleuragramma antarcticum* (Nototheniidae). *Journal of Ichthyology*, **30**: 67-79.

Ferguson J, Ashford J, Piñones A, Torres J, Fraser W, Jones C, Pinkerton M (2011) Connectivity and population structure in *Pleuragramma antarcticum*. WG-FSA-11/19, CCAMLR.

Ferrando S, Hanchet S, Angiolillo M, Gambardella C, Pisano E, Tagliaferro G, Vacchi M (2010) Insights into the life cycle of the Antarctic Silverfish: Reproduction features of the Ross Sea population. *International Polar Year Oslo Science Conference*, 8-12 June 2010,

abstracts, p 1.

Fields PA, Somero GN (1998) Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **95**: 11476-11481.

Fischer W, Hureau JC (1985) FAO species identification sheets for fisheries purposes. In: *Southern Ocean (Fishing Areas 48, 58 and 88) (CCAMLR Convention Area)*, Vol 2 (eds Fischer W, Hureau JC), pp 233-470. Food and Agriculture Organization of the United Nations, Rome.

Gon O, Heemstra PC (1990) *Fishes of the Southern Ocean*. JLB Smith Institute of Ichthyology, Grahamstown, South Africa.

Gracey AY, Fraser EJ, Li W, Fang Y, Taylor RR, Rogers J, Brass A, Cossins AR (2004) Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 16970-16975.

Granata A, Zagami G, Vacchi M, Guglielmo L (2009) Summer and spring trophic niche of larval and juvenile *Pleuragramma antarcticum* in the western Ross Sea, Antarctica. *Polar Biology*, **32**: 369-382.

Grove TJ, Hendrickson JW, Sidell BD (2004) Two species of Antarctic icefishes (genus *Champscephalus*) share a common genetic lesion leading to the loss of myoglobin expression. *Polar Biology*, **27**: 579-585.

Hagen W, Kattner G, Terbrüggen A, Van Vleet ES (2001) Lipid metabolism of the Antarctic krill *Euphausia superba* and its ecological implications. *Marine Biology*, **139**: 95-104.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**: 1633-1638.

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**: 479-485.

Hofmann GE, Buckley BA, Airaksinen S, Keen JE, Somero GN (2000) Heat-shock protein expression is absent in the Antarctic fish *Trematomus bernacchii* (Family Nototheniidae). *The Journal of Experimental Biology*, **203**: 2331-2339.

Hubold G, Ekau W (1987) Midwater fish fauna of the Weddell Sea, Antarctica. In: *Proceedings of the Fifth Congress of the European Ichthyological Society, Stockholm, 1985* (eds Kullander SO, Fernholm B), pp 391-396. Swedish Museum of Natural History, Stockholm.

Hubold G, Tomo AP (1989) Age and growth of Antarctic silverfish *Pleuragramma antarcticum* Boulenger, 1902, from the southern Weddell Sea and Antarctic Peninsula. *Polar Biology*, **9**: 205-212.

Huth TJ, Place SP (2013) De novo assembly and characterization of tissue specific

- transcriptomes in the emerald notothen, *Trematomus bernacchii*. *BMC Genomics*, **14**: 805.
- Iwami T, Kock KH (1990) Channichthyidae. In: *Fishes of the Southern Ocean* (eds Gon O, Heemstra PC), pp 381-399. JLB Smith Institute of Ichthyology, Grahamstown, South Africa.
- Johnston IA, Calvo J, Guderley H, Fernandez D, Palmer L (1998) Latitudinal variation in the abundance and oxidative capacities of muscle mitochondria in perciform fishes. *The Journal of Experimental Biology*, **201**: 1-12.
- Kellermann A (1989) Food and feeding ecology of early stage *Chionodraco rastrispinosus* De Witt and Hureau 1979 off the Antarctic Peninsula. *Pesquisa Antartica Brasileira*, **1**: 25-30.
- Kellermann A (1996) Midwater fish ecology. In: *Foundation for Antarctic Research west of the Antarctic Peninsula. Antarctic Research Series*, Vol 70 (eds Ross RM, Hofmann E, Quetin LB), pp 231-256. American Geophysical Union, Washington DC.
- Kock KH (1992) *Antarctic Fish and Fisheries*. Cambridge University Press, Cambridge.
- Kock KH (2005a) Antarctic icefishes (Channichthyidae): a unique family of fishes: A review, Part I. *Polar Biology*, **28**: 862-895.
- Kock KH (2005b) Antarctic icefishes (Channichthyidae): a unique family of fishes: A review, Part II. *Polar Biology*, **28**: 897-909.
- Kock KH, Kellermann A (1991) Reproduction in Antarctic notothenioid fish - a review. *Antarctic Science*, **3**: 125-150.
- Koubbi P, O'Brien C, Loots C, Giraldo C, Smith M, Tavernier E, Vacchi M, Vallet C, Chevallier J, Moteki M (2011) Spatial distribution and inter-annual variations in the size frequency distribution and abundances of *Pleuragramma antarcticum* larvae in the Dumont d'Urville Sea from 2004 to 2010. *Polar Science*, **5**: 225-238.
- La Mesa M, Ashford J (2008) Age and growth of ocellated icefish, *Chionodraco rastrispinosus* DeWitt and Hureau, 1979, from the South Shetland Islands. *Polar Biology*, **31**: 1333-1342.
- La Mesa M, Catalano B, Greco S (2011) Larval feeding of *Chionodraco hamatus* (Pisces, Channichthyidae) in the Ross Sea and its relation to environmental conditions. *Polar Biology*, **34**: 127-137.
- La Mesa M, Catalano B, Russo A, Greco S, Vacchi M, Azzali M (2010) Influence of environmental conditions on spatial distribution and abundance of early life stages of Antarctic silverfish, *Pleuragramma antarcticum* (Nototheniidae), in the Ross Sea. *Antarctic Science*, **22**: 243-254.
- La Mesa M, Eastman JT (2011) Antarctic silverfish: life strategies of a key species in the high-Antarctic ecosystem. *Fish and Fisheries*, **13**: 241-266.
- Lautrédou AC, Hingsinger DD, Gallut C, Cheng CH, Berkani M, Ozouf-Costaz C, Cruaud C, Lecointre G, Dettai A (2012) Phylogenetic footprints of an Antarctic radiation: the

Trematominae (Notothenioidei, Teleostei). *Molecular Phylogenetics and Evolution*, **65**: 87-101.

Lecointre G (2012) Phylogeny and systematics of Antarctic teleosts: methodological and evolutionary issues. In: *Adaptation and Evolution in Marine Environments - The Impacts of Global Change on Biodiversity*, Vol 1 (eds di Prisco G, Verde C), pp 97-117. Series "From Pole to Pole" Springer, Berlin.

Loeb VJ, Kellermann AK, Koubbi P, North AW, White MG (1993) Antarctic larval fish assemblages: a review. *Bulletin of Marine Science*, **53**: 416-449.

Luo Z, Lin C, Shilatifard A (2012) The super elongation complex (SEC) family in transcriptional control. *Nature Reviews. Molecular Cell Biology*, **13**: 543-547.

Marino IAM, Benazzo A, Agostini C, Mezzavilla M, Hoban SM, Patarnello T, Zane L, Bertorelle G (2013). Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation. *Molecular Ecology*, **22**: 5148-5161.

Matschiner M, Hanel R, Salzburger W (2011) On the origin and trigger of the notothenioid adaptive radiation. *PLoS ONE*, **6**: e18911.

Molecular Ecology Resources Primer Development Consortium, Agostini C, Agudelo PA, BÂ K, *et al.* (2011) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2010-30 November 2010. *Molecular Ecology Resources*, **11**: 418-421.

Molecular Ecology Resources Primer Development Consortium, Agostini C, Albaladejo RG, Aparicio A, *et al.* (2013) Permanent Genetic Resources added to Molecular Ecology Resources Database April 2013-31 May 2013. *Molecular Ecology Resources*, **13**: 966-968.

Moline MA, Karnovsky NJ, Brown Z, Divoky GJ, Frazer TK, Jacoby CA, Torres JJ, Fraser WR (2008) High latitude changes in ice dynamics and their impact on polar marine ecosystems. *Annals of the New York Academy of Sciences*, **1134**: 267-319.

Moylan TJ, Sidell BD (2000) Concentrations of myoglobin and myoglobin mRNA in heart ventricles from Antarctic fishes. *The Journal of Experimental Biology*, **203**: 1277-1286.

Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, Zane L, Fernández DA, Jones CD (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **109**: 3434-3439.

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

O'Brien KM, Mueller IA (2010) The unique mitochondrial form and function of Antarctic channichthyid icefishes. *Integrative and Comparative Biology*, **50**: 993-1008.

Ordway GA, Garry DJ (2004) Myoglobin: An essential hemoprotein in striated muscle. *The Journal of Experimental Biology*, **207**: 3441-3446.

- Pachauri RK, Reisinger A (2008) *Climate change 2007. Synthesis report. Contribution of Working Groups I, II and III to the fourth assessment report*. IPCC, Geneva, Switzerland.
- Papetti C, Marino IAM, Agostini C, Bisol PM, Patarnello T, Zane L (2011) Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei. *Conservation Genetics Resources*, **3**: 259-262.
- Papetti C, Zane L, Patarnello T (2006) Isolation and characterization of microsatellite loci in the icefish *Chionodraco rastrospinosus* (Perciformes, Notothenioidea, Channichthyidae). *Molecular Ecology Notes*, **6**: 207-209.
- Patarnello T, Marcato S, Zane L, Varotto V, Bargelloni L (2003) Phylogeography of the *Chionodraco* genus (Perciformes, Channichthyidae) in the Southern Ocean. *Molecular Phylogenetics and Evolution*, **28**: 420-429.
- Place S, Hofmann G (2005) Constitutive expression of a stress-inducible heat shock protein gene, hsp70, in phylogenetically distant Antarctic fish. *Polar Biology*, **28**: 261-267.
- Place SP, Zippay ML, Hofmann GE (2004) Constitutive roles for inducible genes: evidence for the alteration in expression of the inducible hsp70 gene in Antarctic notothenioid fishes. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, **287**: 429-436.
- Pshenichnov LV (2004) Some peculiarities of *Chionobathyscus dewitti* biology in the Ross Sea. WG-FSA-04/90, 6 pp, CCAMLR, Hobart, Australia.
- Raymond JA, DeVries AL (1972) Freezing behavior of fish blood glycoproteins with antifreeze properties. *Cryobiology*, **9**: 541-547.
- Raymond JA, DeVries AL (1977) Adsorption-inhibition as a mechanism of freezing resistance in polar fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **74**: 2589-2593.
- Regoli F, Nigro M, Benedetti M, Fattorini D, Gorbi F (2005) Antioxidant efficiency in early life stages of the Antarctic silverfish, *Pleuragramma antarcticum*: responsiveness to pro-oxidant conditions of platelet ice and chemical exposure. *Aquatic Toxicology*, **75**: 43-52.
- Reid K, Hoareau TB, Bloomer P (2012) High-throughput microsatellite marker development in two sparid species and verification of their transferability in the family Sparidae. *Molecular Ecology Resources*, **12**: 740-752.
- Robison BH (2003) What drives the diel vertical migrations of Antarctic midwater fish? *Journal of the Marine Biological Association of the United Kingdom*, **83**: 639-642.
- Römisch K, Collie N, Soto N, Logue J, Lindsay M, Scheper W, Cheng CH (2003) Protein translocation across the endoplasmic reticulum membrane in cold-adapted organisms. *Journal of Cell Science*, **116**: 2875-2883.
- Ruud JT (1954) Vertebrates without erythrocytes and blood pigment. *Nature*, **173**, 848-850.

- Scher HD, Martin EE (2006) Timing and climatic consequences of the opening of Drake Passage. *Science*, **312**: 428-430.
- Schofield O, Ducklow HW, Martinson DG, Meredith MP, Moline MA, Fraser WR (2010) How do polar marine ecosystems respond to rapid climate change? *Science*, **328**: 1520-1523.
- Shevenell AE, Kennett JP, Lea DW (2004) Middle Miocene Southern Ocean cooling and Antarctic cryosphere expansion. *Science*, **305**: 1766-1770.
- Shin SC, Kim SJ, Lee JK, Ahn do H, Kim MG, Lee H, Lee J, Kim BK, Park H (2012) Transcriptomics and comparative analysis of three Antarctic notothenioid fishes. *PLoS One*, **7**: e43762.
- Sidell BD, Vayda ME, Small DJ, Moylan TJ, Londraville RL, Yuan ML, Rodnick KJ, Eppley ZA, Costello L (1997) Variable expression of myoglobin among the hemoglobinless Antarctic icefishes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**: 3420-3424.
- Smith WO, Ainley DG, Cattaneo-Vietti R (2007) Trophic interactions within the Ross Sea continental shelf ecosystem. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **362**: 95-111.
- Susana E, Papetti C, Barbisan F, Bortolotto E, Buccoli S, Patarnello T, Zane L (2007) Isolation and characterization of eight microsatellite loci in the icefish *Chaenocephalus aceratus* (Perciformes, Notothenioidei, Channichthyidae). *Molecular Ecology Notes*, **7**: 791-793.
- Teets NM, Yi SX, Lee RE Jr, Denlinger DL (2013) Calcium signaling mediates cold sensing in insect tissues. *Proceedings of the National Academy of Sciences of the United States of America*, **110**: 9154-9159.
- Todgham AE, Hoaglund EA, Hofmann GE (2007) Is cold the new hot? Elevated ubiquitin-conjugated protein levels in tissues of Antarctic fish as evidence for cold-denaturation of proteins in vivo. *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology*, **177**: 857-866.
- Tripathi AK, Roberts CD, Eagle RA (2009) Coupling of CO₂ and ice sheet stability over major climate transitions of the last 20 million years. *Science*, **326**: 1394-1397.
- Turner J, Colwell SR, Marshall GJ, Lachlan-Cope TA, Carleton AM, Jones PD, Lagun V, Reid PA, Iagovkina S (2005) Antarctic climate change during the last 50 years. *International Journal of Climatology*, **25**: 279-294.
- Vacchi M, DeVries AL, Evans CW, Bottaro M, Ghigliotti L, Cutroneo L, Pisano E (2012) A nursery area for the Antarctic silverfish *Pleuragramma antarcticum* at Terra Nova Bay (Ross Sea): first estimate of distribution and abundance of eggs and larvae under the seasonal sea-ice. *Polar Biology*, **35**: 1573-1585.
- Vacchi M, La Mesa M, Dalu M, Macdonald J (2004) Early life stages in the life cycle of Antarctic silverfish, *Pleuragramma antarcticum* in Terra Nova Bay, Ross Sea. *Antarctic*

Science, **16**: 299-305.

Yang N, Peng C, Cheng D, Huang Q, Xu G, Gao F, Chen L (2013) The over-expression of calmodulin from Antarctic notothenioid fish increases cold tolerance in tobacco. *Gene*, **521**: 32-37.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**: 686-693.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**: 1-16.

Zane L, Marcato S, Bargelloni L, Bortolotto E, Papetti C, Simonato M, Varotto V, Patarnello T (2006) Demographic history and population structure of the Antarctic silverfish *Pleuragramma antarcticum*. *Molecular Ecology*, **15**: 4499-4511.

Zhao Y, Ratnayake-Lecamwasam M, Parker SK, Cocca E, Camardella L, di Prisco G, Detrich HW III (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. *The Journal of Biological Chemistry*, **273**: 14745-14752.

Zazulie N, Rusticucci M, Solomon S (2010) Changes in Climate at High Southern Latitudes: A Unique Daily Record at Orcadas Spanning 1903-2008. *Journal of Climate*, **23**: 189-196.

ACKNOWLEDGEMENTS

I'd like to thank all the people who helped me during my PhD.

First of all, many thanks to my supervisor Lorenzo Zane for his supervision, help and support.

Thanks to Tomaso Patarnello for his help, support and for the American experience.

I thank all the people in the laboratory, in particular Elisa Boscari, Ilaria Marino, Chiara Papetti, and Marti Pujolar.

Thanks to Alessandro Coppe for its constant help and support.

Thanks to Joseph Torres and Julian Ashford for giving me the opportunity to collaborate.

Thanks to Stefania Bortoluzzi for her “bioinformatic help”.

Thanks to all other lab members, colleagues and lab-neighbours: Leonardo Congiu, Andrea Pilastro, Alessandro Grapputo, Marta Paterno, Andrea Di Nisio, Alessandro De Vigili, Clelia Gasparini, Silvia Cattelan, Federica Barbisan, Michele Vidotto, Claudia Saccoman, Andrea Bisognin, Enrico Gaffo.

And of course, many many thanks to my mother Floriana, my father Paolo, my brother Tairè, and my dear friends Dedè and Anna Lisa, for supporting me in everything.