Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Deep-Sea Research II 58 (2011) 250–263

Contents lists available at ScienceDirect

Deep-Sea Research II

journal homepage: www.elsevier.com/locate/dsr2

The actinopterygian diversity of the CEAMARC cruises: Barcoding and molecular taxonomy as a multi-level tool for new findings

A. Dettai [a,*], A.-C. Lautredou [a], C. Bonillo [a], E. Goimbault [a], F. Busson [b], R. Causse [b], A. Couloux [c], C. Cruaud [c], G. Duhamel [b], G. Denys [b], M. Hautecoeur [b], S. Iglesias [b], P. Koubbi [d], G. Lecointre [a], M. Moteki [e], P. Pruvost [b], S. Tercerie [a], C. Ozouf [a]

[a] Muséum national d'Histoire naturelle (MNHN), Département Systématique et Evolution, UMR 7138, CP 26, 43, rue Cuvier, 75231 Paris, Cedex 05, France
[b] Muséum national d'Histoire naturelle (MNHN), Département des milieux et peuplements aquatiques, UMR 7208, CP 26, 43, rue Cuvier, 75231 Paris, Cedex 05, France
[c] Genoscope, Centre National de Séquençage, 2, rue Gaston Crémieux, CP5706, 91057 Évry, Cedex, France
[d] Université Pierre et Marie Curie Paris VI, UMR 7093, Laboratoire d'océanographie de Villefranche-sur-Mer, 06230 Villefranche sur mer, France
[e] Department of Ocean Sciences, Tokyo University of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo 108-8477, Japan

ARTICLE INFO

Article history:
Received 18 May 2010
Accepted 18 May 2010
Available online 1 June 2010

Keywords:
Ichthyology
Identification
Molecular taxonomy
Antarctic zone
Barcode
Southern Ocean
Dumont d'Urville Sea

ABSTRACT

In the winter 2007–2008, the CAML-CEAMARC cruises prospected in the Eastern part of the Antarctic continental shelf (Dumont d'Urville Sea, off Terre Adélie). The Australian R/V "Aurora Australis" and the Japanese R/V "Umitaka Maru" sampled in locations and at depths previously uninvestigated in this region. In total, 538 teleost specimens collected during these cruises were sequenced for the mitochondrial cytochrome oxidase I gene (COI), with the goal of barcoding a representative sampling from the campaign. The efficiency of barcoding for identification has been questioned for some taxonomic groups, thus we compared the COI results for a few of the families and genera included here (genus *Trematomus*, Artedidraconidae, Liparidae) to results for other markers for the same specimens. To better explore intra- and interspecific variability, sequences from previous campaigns and public databases were added to the analysis for these groups. The congruence among the results for different genes (COI, cytochrome b, D-loop and the nuclear rhodopsin retrogene) and morphological identification was used to assess the efficiency of the COI dataset at recovering species delimited using other data. Where discrepancies were present among the different data sources, a morphological re-identification was performed.

The partial COI sequence yields reliable identification in most Antarctic teleost families when using their position in the clusters on a NJ tree. However, for several groups of species neither COI nor the other molecular markers investigated nor morphology recover unambiguously the currently accepted species. The taxonomy of these groups needs to be reconsidered. Identification through sequence similarity using the Barcode of Life Data System (BOLD) works for some groups, but is hampered by the incompleteness of the taxonomic coverage for antarctic teleosts. For four families (Artedidraconidae, Zoarcidae, Liparidae and Channichthyidae), several interspecific divergences were very small, and of the same magnitude as intraspecific divergences for other antarctic species. Despite these small divergences, almost all the species investigated in artedidraconids have molecular synapomorphies in the COI sequences, and a barcoding gap from the closest species. In the genus *Trematomus*, almost all species are well separated except for two pairs of closely related species that could not be distinguished by the other molecular markers either. For the typically hard to identify zoarcids and liparids, the results of barcoding are in agreement with in-depth morphological study. Once a reasonably complete reference dataset is available, barcoding will be invaluable to discriminate species from one another in these families. A careful comparison of the morphological and molecular results for our specimens allowed us to add numerous well-identified specimens (including some rare species) and sequences to BOLD. It helped to pinpoint the specimens that needed to be re-identified morphologically, and highlighted groups where barcoding is most helpful for specimen identification (*Chionodraco* species). This large-scale project underlines the need for further taxonomic work in antarctic actinopterygians.
© 2010 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +33 1 40 79 37 50; fax: +33 1 40 79 38 44.
E-mail address: adettai@mnhn.fr (A. Dettai).

0967-0645/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dsr2.2010.05.021

Author's personal copy

1. Introduction

With over 30 000 valid species, and more than 300 described each year (Froese and Pauly, 2009), actinopterygian fish remain the last great challenge in our discovery of vertebrate species diversity. They hold a crucial place in marine ecosystems and possess great economic importance. Precise and reliable identification is needed as a basis for scientific studies (Bely and Weisblat, 2006; Bortolus, 2008) as well as for fraud detection (Lockley and Bardsley, 2000; Wong and Hanner, 2008). Moreover, identification must be based on a sound knowledge of the taxonomy, as a faulty delineation of the species limits often precludes identification altogether. Our knowledge of this group comes principally from morphological studies, but in recent years, molecular taxonomy studies have begun to prove their worth, especially when combined with morphology. For instance, they have helped to detect cryptic species (Kon et al., 2007; Zemlak et al., 2009; Steinke et al., 2009) and, conversely, to relate morphologically different life stages and sexes to a single species (Johnson et al., 2009). However, identification remains primarily based on morphology. It can be limited in the case of incomplete specimens (i.e. stomach contents), for determining eggs, larvae or juveniles (Koubbi et al., 2009), or simply because of the sheer diversity of species. Some specimens can be very hard to identify even for specialists, and there is a dire lack of experts on many groups.

Molecular identification based on mitochondrial DNA has been around for several decades (see Ward et al., 2009), but has recently taken a new dimension through larger scale projects with a standardised approach and high quality control (FishTrace www.fishtrace.org, and especially the Barcode of Life (BOL) http://www.barcodinglife.org/). These rely on the sequencing of standardized gene regions (cytochrome b and rhodopsin for FishTrace, cytochrome oxidase I for BOL). Identification is then performed through a comparison to publicly accessible reference datasets, in which sequences are linked to voucher specimens. More stringent control, as well as the link with vouchers, add a reliability and an *a posteriori* controllability that is absent (Harris, 2003) from sequences deposited in other databases. The link between a sequence and its voucher specimen allows to recheck the specimen, should the systematics of a group or an identification been questioned.

The Barcode of Life project is the largest in scale. It uses a database with an associated data analysis system, the Barcode of Life Data System (BOLD, Ratnasingham and Hebert, 2007). The project has received much attention, and has been presented as a powerful tool for molecular taxonomy (Hebert et al., 2003 and others), not without generating heated debates about limits and advantages of the approach itself (see for instance DeSalle et al., 2005; Rubinoff et al., 2006; DeSalle, 2006; Buhay, 2009) and about the use of a cut-off value to differentiate inter- and intraspecific divergence levels (Meyer and Paulay, 2005; Hickerson et al., 2006). Both the evaluation of the approach and the development of ameliorations are still underway; however, it looks promising for numerous taxa. The part of the project devoted to fish diversity (www.fishbol.org) is very active (Ward et al., 2009) and the number of included species from all over the world rises steadily.

The Southern Ocean ecosystem is one of the places that will be most impacted by global warming (Clarke et al., 2005; Thatje, 2005; Aronson et al., 2009). While the changes are less noticeable in the Eastern Antarctic region, they are already visible in the Antarctic Peninsula (Steig et al., 2009; Naish et al., 2009). Monitoring of these changes requires a biodiversity baseline inventory as soon as possible, and a large amount of taxonomic work is still needed for the region, including for fish. An additional and reliable tool for identification would therefore be truly plutot que very welcome.

We explore here the efficiency of identification through barcoding with COI for Antarctic actinopterygian fish, as well as the use of this marker for preliminary studies in molecular taxonomy. For these purposes, we sequenced a large number of specimens from two of the Collaborative East Antarctic Marine Census (CEAMARC) cruises. These cruises were carried out by the Australian R/V "Aurora Autralis" and the Japanese R/V "Umitaka Maru" during the Antarctic summer 2007/2008. The project is part of the CAML initiative in the framework of the International Polar Year. Before CEAMARC, the coastal (0–200 m) fish fauna of this area had been investigated starting in the sixties (morphological studies, Hureau, 1966), and most recently for both morphological and molecular studies, by the IPEV French programme ICOTA (Ichtyologie Côtière en Terre Adélie). Only 21 teleost species had been recorded, mainly notothenioids. During CEAMARC, demersal fish were collected on board the R/V "Aurora Australis" (1172 actinopterygian specimens, 65 species) and pelagic fish and ichthyoplankton on board the R/V "Umitaka Maru" (totalling more than 350 000 actinopterygian specimens and 49 species), down to 2400 m deep. This brings the number of morphologically identified teleost species recorded in the area to at least 91, including one new and several rare species.

Whether for enriching the reference database or for biodiversity exploration, collecting cruises are highly efficient in gathering high quality material suitable for both morphological and molecular works. Yet, very often, the diversity collected is such that finding competent taxonomists for all fish groups is a very long and arduous process. Cruises in the Southern Ocean are an ideal case to explore the relevance of barcoding all specimens from a campaign because the number of actinopterygian groups (Eastman, 1993; Eastman and Clarke, 1998) is relatively low and therefore precise identification can be more easily obtained. On the CEAMARC cruises, specialists for almost all the sampled groups were involved, making reliable and fast identifications possible.

We present here the results of the barcoding of almost all specimens sampled for molecular study on the R/V "Aurora Australis", as well as some of those from the R/V "Umitaka Maru". This last cruise collected a high proportion of larvae, not all of which can be identified morphologically to the species level by the only taxonomic key available (North and Kellermann, 1990). This key describes only 58 of the 322 fish species currently known for the Southern Ocean (Koubbi et al., 2009). Therefore, misidentifications of larvae are possible and a re-examination of the specimens will need to be performed for this cruise after integrating the results of our molecular identification. While starting with molecular identification might have yielded a species list faster, it is also more destructive due to the very small size of many of the larvae and juveniles. It was therefore decided to first perform a morphological study, and to wait for a full dataset and a test of the methodology based on the sampling from the R/V "Aurora Australis" before barcoding collections from the R/V "Umitaka Maru".

2. Material and methods

2.1. Collection

Fish specimens were collected during the CEAMARC campaigns off Adélie and King George V lands (Dumont d'Urville Sea). The R/V "Aurora Australis" (AA) surveyed the benthic fauna using beam trawls on 89 stations between 139.3 and 145.53°E and between 65.44 and 67.05°S at depths ranging from 138 to 1260 m,

Author's personal copy

from December 24, 2007 to January 20, 2008. A total of 1172 fish specimens and 540 samples with vouchers were collected for barcoding. The R/V "Umitaka Maru" (UM) collected samples from 46 operations in the epi-, meso- and bathypelagic zones, between 139.5733 and 145.0082°E and 67.0572 and 61.9748°S, using IYGPT (International Young Gadoids Pelagics Trawl) and RMT (Rectangular Midwater Trawl) nets from the surface to 2000 m at 22 stations, from January 19 to February 12, 2009. A total of 219 specimens were kept for molecular analysis. A representative sampling was collected aboard both ships according to the CAML recommendations and protocols (http://www.caml.aq/barcoding/documents/Barcoding-Sampling.pdf). When possible, up to 10 specimens were kept for each species per station. For the AA, the catch was photographed fresh, and a sample (muscle or fin) was preserved for molecular study in 95% ethanol. The fish were then preserved in 10% buffered formalin, and kept as vouchers; a few smaller specimens were preserved whole in 85% ethanol. Once they arrived at the Muséum National d'Histoire Naturelle (Paris), the specimens were transferred to 85% ethanol for long-term archival in the collection. The barcode vouchers from the AA were attributed numbers MNHN2008-2592 to MNHN2008-2628 (Liparidae), MNHN2009-0020 to MNHN2009-0081 (Zoarcidae) and MNHN2009-0932 to MNHN2009-1402 (the remaining specimens). The vouchers from the UM for the specimens integrated in the present study were attributed numbers MNHN2009-902 to MNHN2009-931.

2.2. Morphological identification

Almost all the specimens from the AA were adults and could be easily sorted. Most of them were identified on board or shortly after the cruise. For Liparidae identification, the right pectoral girdle had to be removed and studied after clearing and staining (Duhamel et al. unpublished). A second round of morphological identifications was performed for specimens for which the molecular and the morphological identification were not in agreement. There were a number of juvenile stages, and even one batch of fish eggs that could not be identified morphologically after this second round. These specimens were considered as "sp." even if the molecular identification was non-ambiguous, to avoid entering into BOLD identifications that were not supported independently by morphology.

2.3. Molecular study

Muscle samples or skin samples were stored in 85% ethanol and extracted following the protocol in Winnepenninckx et al. (1993). The partial cytochrome oxidase I was amplified using primers FishF1-5′TCAACCAACCACAAAGACATTGGCAC3′, FishF2-5′TCGACTAATCATAAAGATATCGGCAC3′, FishR1-5′TAGACTTCTGGGTGGCCAAAGAATCA3′, FishR2-5′ACTTCAGGGTGACCGAAGAATCAGAA3′ (Ward et al., 2005), as well as the new TelF1-5′TCGACTAATCAYAAAGAYATYGGCAC3′ and TelR1-5′ACTTCTGGGTGNCCAAARAATCARAA3′ in a 25 µl volume 5% of DMSO, 5 µg of bovine serum albumine, 300 µM of each dNTP, 0.3 µM of Taq DNA polymerase (Qiagen), 2.5 µl of the corresponding buffer and 1.7 pM of each of the two primers. After denaturation for 2 min at 94 °C, the PCR ran for 40–50 cycles of (20 s, 94 °C; 25 s, 52 °C and 45 s, 72 °C), with a terminal elongation of 3 min at 72 °C on Biometra thermocyclers. Purification and sequencing of the PCRs were performed at the Genoscope (http://www.genoscope.cns.fr/) using the same primers. All sequences were obtained in both directions and checked manually against their chromatogram using Sequencher 4.8 (Gene Codes Corporation). They were aligned by hand using Bioedit (Hall, 1999), and were controlled for mix-ups and contaminations by pairwise sequence comparison. The sequences were deposited in the BOLD with the specimen and collection data.

To allow comparison with other barcode publications and the NJ trees provided by the BOLD, aligned sequences were analysed by the NJ distance method with the Kimura 2 parameter model (Kimura, 1980) using PAUP*4.0b10 (Swofford, 2002). At least one sequence from each of the obtained clusters was used to query the BOLD "Identify specimen" tool using the complete database of all records ("unvalidated dataset"). The intraspecific distances (mean and maximum), as well as the interspecific distance (mean and minimum, Meier et al., 2008) from the closest species cluster were calculated using MEGA (Tamura et al., 2007).

To complete the study in some of the more challenging groups investigated (genus *Trematomus*, Liparidae, Artedidraconidae), the results for the COI were compared with larger datasets including carefully chosen samples from previous campaigns (ICOTA campaigns, ICEFISH 2004, POKER 2006), as well as the available sequences from the BOLD (Duhamel and Lecointre, unpublished data; Lautredou et al., in press). These studies also include additional datasets: for the genus *Trematomus*, the rhodopsin retrogene (Lautredou et al., in press); for Artedidraconidae, D-loop, cytochrome b and the rhodopsin retrogen for almost all specimens (Lecointre et al. unpublished data), and maximum parsimony analyses (heuristic searches; TBR search, 1000 replicates), using PAUP*4.0b10 (Swofford, 2002).

3. Results and discussion

3.1. Morphological identification

Although the actinopterygian fauna of the Southern Sea is relatively well known compared to other regions (Eastman, 1993), this sector of the East Antarctic coast had not been investigated before at these depths and held promise for the discovery of new species. A new zoarcid was identified, and is currently in the process of being analysed in detail at both molecular and morphological levels. Chondrichthyans were observed in submarine videos and still images but could not be caught with the beam trawl, except for an egg containing an embryo (also currently being analysed). In terms of qualitative and quantitative species composition, the sampled fish fauna does not appear very different from other sectors of the Antarctic region according to Eastman and Eakin (2000), but there might be some genetic and morphological divergences with similar species in other sectors of the Antarctic Ocean which will deserve further comparisons.

The morphological identification of some of the species collected can be problematic. For instance, *Trematomus lepidorhinus* and *T. loennbergii* cannot be easily distinguished on the basis of morphological characters, and *T. bernacchi* and *T. pennellii* look very similar at young stages. *Chionodraco hamatus* and *C. myersi* are sometimes difficult to separate without careful examination: at first all the *Chionodraco* specimens captured were identified on board as *C. hamatus*. A new morphological identification after doubts were raised by the molecular results helped to correct the identification to *C. myersi* for a number of specimens. A few individuals from the genus *Pogonophryne* were collected. Classically used meristic characters only poorly separate the species within this genus: they can exhibit both strong growth allometry and sexual dimorphism. The study of karyotypes cannot be used as a discrimination tool either because chromosome number and formula are highly conserved within all species analysed to date (Ozouf-Costaz et al., 1991; Pisano and Ozouf-Costaz, 2002; Ozouf-Costaz, unpublished). Similarly, assigning the collected *Bathydraco*

Author's personal copy

specimens to a species on the basis of their morphology was not possible for all the individuals.

The morphological identification of both Liparidae and Zoarcidae is very difficult (Rock et al., 2008: Duhamel et al., unpublished data). Identification of the liparids required a lengthy preparation, as well as the observation of a number of complex characteristics and the comparison with numerous other specimens. The liparids from the CEAMARC and Kerguelen POKER 2006 campaigns will be fully listed and described (results for COI, morphology and life traits) elsewhere.

3.2. Molecular data and DNA quality

The COI dataset included 538 sequences, of which 500 come from the AA and 38 from the UM.

Obtaining sequences for the COI was problematic for some species, regardless of the combination of primers used. *Trematomus pennelli* and a number of bathydraconid species (see Table 1) posed recurrent problems. Multiple PCRs and sequencing attempts were necessary, and these PCRs were especially prone to contaminations. Several *Bathydraco* specimens repeatedly presented multiple peaks in the electrophoregrams, and the sequences could not be obtained at all. The problems were highly group specific, and are probably not due to low DNA quality. For several of the problematic samples, nuclear markers could be easily amplified from the same extraction. Cloning and/or new primer design for the COI will be necessary for these groups.

For the smallest species, and specimens, a few individuals were directly preserved in ethanol, including larvae and a cluster of unidentified eggs. Three separate extractions of individual eggs were performed. Whilst all three yielded DNA, PCRs and sequences, only one of the sequences was of high quality, although the other two were still good enough to be used for identification purposes. This might be due to the developmental stage of the embryo, and an eventual correlation will have to be investigated systematically in a future study.

We performed a comparison of the PCR results on both muscle and skin from an *Edentoliparis terraenovae* (Liparidae) preserved whole in ethanol. While the skin gave a slightly fainter PCR than muscle, nuclear as well as mitochondrial markers could be obtained, and presented no contaminations. This provides a good alternative to muscle or fin samples in liparids, as the loose, scaleless "skin" surrounding the individuals can easily be removed and provides a sufficient amount of material even for small specimens, without damaging the specimen itself.

The specimens were fixed and kept in formalin until they arrived at MNHN approximately 4–5 months after being collected out their collection. Not surprisingly (Formalin Workshop Summary, 2006), test PCRs from extractions of formalin treated samples yielded only very faint bands from which no sequence could be obtained. But 85% ethanol without fixation is not an ideal long term conservation for whole specimens (Moore, 1999), and precludes some later uses. The fixation of the whole fish in formalin after removing a separate sample for molecular analysis currently seems the best option for an optimal study of the specimen.

3.3. Clusters and identifications

For a successful molecular identification, all the specimens from a given species must cluster together in the analysis (unique COI clusters for each species, Steinke et al., 2009). There must also be some level of difference between intraspecific and interspecific variability and/or at least a possible diagnosis through observation of a single site or of a combination of sites in the alignment of sequences (Meyer and Paulay, 2006).

The NJ analysis yielded 68 clear clusters. Most of them included only specimens with the same morphological identification, but a few included two different identifications. This can be interpreted as either a contamination problem, a morphological identification problem (identification error), or a taxonomic problem (the identification was correct in regard to the current state of knowledge; however, the "real" species limits are not in agreement with what we believe we know about it). In actinopterygians, this is not a trivial problem, and several studies have flagged a high proportion of common species with possible problems (Kon et al., 2007; Zemlak et al., 2009). To determine in each case which of these hypotheses was correct, we went back to the corresponding specimens for a new, blind, morphological identification, and re-sequenced some of the samples. The majority of the problematic sequences were in fact due to erroneous first identifications onboard ships. Most were restricted to a few clusters and the mix-ups were not random, but between species that could be misidentified for each other at a stage of their development.

A few contaminations could be detected (*T. pennellii* and some *Bathydraco* species) through comparison of sequences. They were eliminated from the dataset and re-sequenced when possible.

Some of the specimens collected on the UM shared clusters with specimens from the AA (bathydraconids, channichthyids). But in general, different species were collected on the UM and on the AA, and only 27 were shared. Branch lengths among the clusters from the UM were longer than among those from the AA, as the sampling is less dense species-wise but more diverse groups were present at the depths investigated (see Fig. 1A and B).

Intraspecific variability is generally similar across species. Two species have a divergence between some specimens that is greater than 2%: *Trematomus eulepidotus* (si218n1689 vs. the others) and *Bathylagus antarcticus* (UM7142P48 vs. si271n2105), where some specimens differed from each other by 2.5% (see Table 1). Moreover, all the fish included in this analysis were collected over a relatively restricted geographic area, and the maximal values for intraspecific divergence calculated here might not represent the real diversity of the species. The divergence from specimens identified as *B. antarcticus* in the BOLD is even higher, reaching 2.7% (specimen UM7142P48 vs. FNZB320-0).

The range of interspecific differences is much more variable depending on the groups. The use of the means for intra- and interspecific divergence comparison does not allow detection of the problematic cases, as already noticed by Meier et al. (2008), so we instead compared minima for interspecific divergences to maxima for intraspecific divergences. If the whole dataset is considered, there is a clear overlap between intra- and interspecific variabilities, as the smallest interspecific divergences are well below 1% (Artedidraconidae, Liparidae, Zoarcidae and Channichthyidae) but the largest intraspecific divergences reach above 2% (Table 1). When smaller groups are considered (family, genera), the largest intraspecific divergence is almost always smaller than the smallest interspecific divergence, except for a few species pairs (*Pogonophryne species*, *Artedidraco shackletoni, Trematomus loennbergii* and *T. lepidorhinus*) that will be discussed in the present study. However, our sampling does not contain all the species for every group and the observed gap between inter- and intraspecific divergences might only reflect an incompleteness of sampling. While there is almost no overlap, there is nonetheless no consistent barcoding gap and cut-off value for antarctic actinopterygians. Such a value would have to be established on a case by case basis.

3.4. Separation of morphologically similar species

In one case, there were two clusters for a single identification (*C. hamatus*) separated by a cluster from another species

Author's personal copy

Table 1
List of the species with repartition of the 538 specimens included in this paper (Fig. 1a, b) and inter- and intraspecific divergences. All identifications were checked morphologically, except for some of the specimens from the UM that need to be further investigated. The inter- and intraspecific variabilities were calculated using the K2p model. The average and the smallest interspecific distance from the closest cluster are indicated. Interspecific distances smaller than 2% and intraspecific distances larger than 2% are in **bold**. Overlaps between inter and intraspecific distances are underlined. For UM, as most specimens are unidentified larvae, the total number of collected specimens per species is not known. These cases are indicated by ?.

Order	Family	Genus and species	Main identifiers	nb of samples for DNA	Nb. of seq. obtained	Cruise	Intra- specific var. av. (max)	Inter-specific dist to closest av. (min)
Serraniformes								
	Notothenioidei							
	Artedidraconidae		C. Ozouf,					
		Artedidraco loennbergi	P. Pruvost,	21	20	AA	0.002 (0.007)	**0.016 (0.013)**
		Artedidraco orianae	R. Causse,	12	12	AA	0.002 (0.007)	**0.022 (0.018)**
		Artedidraco shackletoni	G. Denys	22	22	AA	0.002 (0.008)	**0.010 (0.008)**
		Artedidraco skottsbergi		11	11	AA	0.003 (0.005)	**0.016 (0.013)**
		Artedidraco sp.		5	5	AA		
		Dolloidraco longedorsalis		17	14	AA	0.001 (0.003)	**0.012 (0.010)**
		Histiodraco velifer		5	5	AA	0.003 (0.005)	**0.012 (0.010)**
		Pogonophryne scotti		9	9	AA	0.003 (0.008)	**0.008 (0.005)**
					1	UM		
		Pogonophryne cluster1		4	4	AA	0.003 (0.007)	0.007 (0.005)
		Pogonophryne cluster2		7	6	AA	0.002 (0.003)	0.007 (0.005)
	Bathydraconidae		C. Ozouf,					
		Acanthodraco dewitti	P. Pruvost,	1	0	AA	–	–
		Akarotaxis nudiceps	R. Causse,	15	2	AA	0.006 (0.006)	0.033 (0.030)
		Bathydraco antarcticus	G. Denys	2	1	AA	0.003 (0.006)	0.003 (0.006)
		Bathydraco marri		3	3	AA		
		Bathydraco sp.		3	3	AA		
		Cygnodraco mawsoni		7	7	AA	0.000 (0.001)	0.055 (0.052)
		Gerlachea australis		18	18	AA	0.003 (0.007)	0.055 (0.052)
		Gymnodraco acuticeps		2	2	AA	–	0.056 (0.050)
		Prionodraco evansii		8	8	AA	0.002 (0.007)	0.053 (0.050)
		Racovitzia glacialis		18	18	AA	0.002 (0.004)	0.053 (0.050)
		Vomeridens infuscipinnis		14	14	AA	0.000 (0.002)	0.026 (0.024)
	Channichthyidae		C. Ozouf,					
		Chaenodraco wilsoni	P. Pruvost,	4	4	AA	0.007 (0.010)	0.026 (0.022)
		Chionobathyscus dewitti	R. Causse,	3	3	AA	0.000 (0.000)	**0.018 (0.017)**
		Chionodraco hamatus	G. Denys	9	9	AA	0.004 (0.008)	**0.022 (0.017)**
		Chionodraco myersi		12	12	AA	0.003 (0.007)	**0.022 (0.017)**
		Cryodraco antarcticus		21	19	AA	0.002 (0.003)	**0.018 (0.017)**
		Dacodraco hunteri		5	5	AA	0.003 (0.005)	0.055 (0.049)
		Neopagetopsis ionah		8	7	AA	0.002 (0.003)	0.050 (0.049)
		Pagetopsis maculatus		6	5	AA	0.001 (0.002)	0.022 (0.022)
		Pagetopsis macropterus		5	5	AA	0.000 (0.000)	0.022 (0.022)
		gen. sp.		3	3	AA		
	Nototheniidae							
		Pleuragramma antarcticum		24	23	AA	0.002 (0.008)	0.107 (0.103)
		Trematomus bernacchii	C. Ozouf,	5	5	AA	0.001 (0.002)	0.050 (0.047)
		Trematomus eulepidotus	P. Pruvost,	42	42	AA	**0.007 (0.025)**	0.056 (0.045)
		Trematomus hansoni	R. Causse,	1	1	AA	–	0.058 (0.056)
		Trematomus newnesi	G. Denys	1	1	AA	–	0.089 (0.084)
		Trematomus pennellii		7	3	AA	0.000 (0.000)	0.048 (0.047)
		Trematomus scotti		23	21	AA	0.001 (0.006)	0.104 (0.101)
		Trematomus lepidorhinus/loennbergii			43	AA	0.004 (0.015)	0.056 (0.045)
		Trematomus tokarevi		4	4	AA	0.002 (0.004)	0.048 (0.047)
	Liparidae		G. Duhamel					
		Careproctus continentalis		1	1	AA	–	0.037 (0.037)
		Careproctus longipectoralis		2	2	AA	0.000 (0.000)	0.037 (0.037)
		Edentoliparis terraenovae		2	2	AA	0.005 (0.07)	0.046 (0.044)
				1	1	UM		

		Paraliparis antarcticus		9	8	AA	0.001 (0.002)	**0.014 (0.013)**
		Paraliparis charcoti		3	3	AA	0.001 (0.002)	**0.014 (0.013)**
		Paraliparis leobergi		7	7	AA	0.002 (0.003)	**0.008 (0.007)**
		Paraliparis mawsoni		5	5	AA	0.001 (0.002)	0.057 (0.056)
		Paraliparis valentinae		1	1	AA	–	0.027 (0.027)
	Zoarcidae		S. Iglesias					
		Lycenchelys aratrirostris		12	12	AA	0.001 (0.002)	**0.016 (0.015)**
		Lycenchelys sp.		1	1	AA	–	**0.013 (0.012)**
		Lycenchelys tristichodon		7	7	AA	0.001 (0.002)	0.026 (0.024)
		Lycodapus pachysoma		2	2	AA	0.000 (0.000)	0.063 (0.062)
		Lycodichthys antarcticus		7	7	AA	0.004 (0.010)	**0.013 (0.012)**
		Oidiphorus mcallisteri		1	1	AA	–	0.034 (0.033)
		Ophthalmolycus amberensis		9	8	AA	0.001 (0.003)	0.031 (0.029)
		Pachycara brachycephalum		8	7	AA	0.000 (0.000)	**0.005 (0.005)**
		Pachycara sp.		1	1	AA	–	**0.005 (0.005)**
		gen. sp. Nov.		1	1	AA	–	0.026 (0.026)
Gadiformes			C. Ozouf					
	Macrouridae							
		Macrourus whitsoni		24	24	AA	0.001 (0.009)	0.214 (0.209)
		Cynomacrurus piriei		?	1	UM	–	0.214 (0.209)
	Muraenolepididae							
		Muraenolepis sp.		5	4	AA	0.000 (0.000)	0.240 (0.220)
Stephanoberyci-formes			P. Koubbi					
	Melamphaidae							
		Poromitra crassiceps		?	1		–	0.204 (0.204)
Lophiiformes								
	Oneirodidae		P. Koubbi					
		Oneirodes notius	M. Bruni	?	1	UM	–	0.237 (0.237)
Osmeriformes			P. Koubbi					
	Osmeridae							
		Bathylagus antarcticus		1	1	AA	**0.019 (0.025)**	0.056 (0.053)
				?	4	UM		
		Bathylagus tenuis ?	mol. ID: only oneNeopagetopsis ionah species identified on board	?	5		0.000 (0.000)	0.056 (0.053)
Myctophiformes								
	Myctophidae		P.A. Hulley					
		Gymnoscopelus opisthopterus		2	2	AA	0.001 (0.002)	0.028 (0.025)
				?	3	UM		
		Gymnoscopelus bolini		?	1	UM	–	0.057 (0.055)
		Gymnoscopelus braueri		?	3	UM	0.007 (0.011)	0.028 (0.025)
		Nannobrachium achirus		?	2	UM	0.004 (0.004)	0.178 (0.175)
		Electrona antarctica		2	2	AA	0.000 (0.000)	0.109 (0.109)
				?	3	UM		
		Electrona carlsbergi		?	1	UM	–	0.109 (0.109)
Aulopiformes								
	Notosudidae	**Scopelarchidae ?**	P. Koubbi				–	0.203 (0203)
		Gen. sp.		?	1	UM		
	Paralepididae		P. Koubbi					
		Notolepis coatsi		?	1	UM	–	0.223 (0.223)

Author's personal copy

Author's personal copy

(*Chaenodraco wilsoni*). Searches in the BOL D yielded two different identifications, *C. hamatus* for one cluster and *C. myersi* for the second. The intra- and interspecific divergences also show a clear barcoding gap. A re-examination of the specimens allowed their sorting into the same two groups. The sorting of these two species morphologically requires care and time, and COI sequences could be a valuable tool to hint at the need to re-identify a collection. Similarly, at first all the specimens of *Bathylagus* were identified as *B. antarcticus*. However, the sequences formed two clusters on the NJ tree, and a search with the BOLD yielded the identification *Bathylagus tenuis* for one of them. These specimens will have to be re-examined more closely.

According to the Catalog of Fishes (Eschmeyer and Fricke, 2009), 13 species are currently included in the monophyletic group formed by genera *Trematomus* and *Pagothenia*. As only 9 were present in the CEAMARC sampling (including two species represented by a single individual), specimens from previous campaigns were added. This provided a dataset comprising all but one species *(Pagothenia brachysoma*), and therefore stood a better chance of detecting potential problems due to recently diverged species. After checking a few discordant specimens that had identification or entry errors, all morphologically identified species formed distinct COI clusters with intraspecific distances below 0.5% on average and distances between clusters above 4.5% (Table 1, Fig. 2) except two pairs, *T. loennbergii–T. lepidorhinus* and *T. bernacchii–T. vicarius* (this pair will not be discussed here because *T. vicarius* is restricted to Southern Georgia, and is considered by several authors as a subspecies of *T. bernacchii*). The rhodopsin retrogene, a nuclear marker, yielded congruent clusters but with lower divergences among sequences (Lautredou et al., in press). Using COI, the divergence between specimens first identified as *T. loennbergii* and *T. lepidorhinus* appears at the same level as the intraspecific divergences in the remainder of the genus. These specimens do not form distinct clusters according to the morphologically identified species in NJ trees. However, there are a few shared molecular similarities among specimens that separate them into two groups. The same specimens (save one) were also grouped together with several molecular synapomorphies by the MP analysis of the rhodopsin retrogene (Lautredou et al., in press).

A new morphological identification did not recover these *Trematomus* clusters, and the characters usually employed for the identification of these two species were not correlated with them either. This suggests that while there might be some degree of separation involved, we cannot reach a conclusion from the currently available data. Further work with morphology, ultravariable nuclear markers (microsatellites for instance), and specimens from around the Antarctic Continent is necessary. New microsatellites have recently been defined on *Trematomus* species and could be used (Van de Putte et al., 2009), but they were not yet tested on the problematic pairs of species. Until further studies are performed, the use of "*T. loennbergii/ T. lepidorhinus* group" would best reflect the state of our knowledge. The results of Kuhn and Near (2009) are in agreement with ours, although their nuclear marker (ribosomal protein S7 intron 1) does not recover species clusters for five out of ten species. Nonetheless, COI sequences as implemented in the BOLD can be used routinely for molecular identification of the other *Trematomus* species, including the morphologically similar juveniles of *T. bernacchii* and *T. pennellii*. A few juvenile specimens that could not be identified to the species level could therefore be placed with confidence within species clusters by the molecular analysis. The sequences derived from the unidentified eggs always clustered within the sequences from *T. eulepidotus.* Individuals from this species had indeed been identified on videos from the beam trawl from the station where the eggs were collected. This is promising for future behavioural and ecological studies.

Rock et al. (2008) suggested that artedidraconids had an identification problem, and possibly a taxonomic one too, due to low COI divergence; however, they based their work on few species and specimens. Our study included 106 artedidraconids from at least 7 identified species. Species had unique molecular clusters (Steinke et al., 2009), except for several *Pogonophryne* species identified on board as *P. cf marmorata, P. cf lanceobarbata* and *P. sp.* on the basis of Eakin (1990) only. All specimens identified as *P. scotti* clustered together. The *A. loennbergi* GBGC723609 sequence retrieved from the BOLD groups with our *A. skottsbergi* sequences, not with our *A. loennbergi* sequences, and this needs to be investigated. The intraspecific distances (see Fig. 3) were less than the interspecific distances for most identified artedidraconid species, but the difference was far from the tenfold difference suggested by Ward et al. (2009) and lower also than the difference identified by Steinke et al. (2009). Almost all interspecific distances were smaller than 2%, and several were less than 0.8%. This is also observed on the trees generated for the group during identification of sequences using the BOLD. As this is within the range suggested by these authors for intraspecific divergence, we compared the COI NJ tree with the clustering of the same specimens for other markers: cytochrome b and the partial rhodopsin retrogene (Lecointre et al., unpublished data). Congruent clusters were recovered, even for the nuclear marker (except for the *Pogonophryne* group where there was a complete lack of resolution). The perfect congruence between multiple sources of independent data (mitochondrial, nuclear and morphological) suggests a lack of variability of COI linked to a very recent divergence or a slow rate of molecular evolution, rather than the absence of distinct evolutionary and genetic units. The *Pogonophryne* group (with the exception of *P. scotti*) needs to be further studied with more variable markers as well as by in-depth morphology to establish the number of distinct evolutionary units and their identification. Several *Pogonophryne* species have been placed into synonymy (Balushkin and Eakin, 1998) and new species have been described (Eakin and Eastman, 1998; Eakin et al., 2001), so great care should be taken before confirming species identification. We chose to leave these few problematic specimens as sp. for now. However, COI can be used to identify all the other species of artedidraconids present in our sampling. The position in the NJ COI tree perfectly reflects the morphological identification for seven of the species, and there are diagnostic sites or site combinations for all of these.

Bathydraco species posed a similar problem (Rock et al., 2008). In our NJ tree, all specimens are grouped in a single cluster, with small divergences between specimens and no straightforward sub-clusters, although several distinct species had been identified morphologically. It is difficult to reach a conclusion in this case, as there are many sequences missing for this group due to technical problems with COI. More work is needed, including the sequencing of additional markers. Cytochrome b appears promising (Rock et al., 2008), but for reliable molecular taxonomy nuclear markers would also be needed.

Liparid morphological identification requires a high level of expertise and the study of osteological characters. Specimens from the same species were always grouped together, and distances between species clusters were above 2% except between *Paraliparis charcoti, P. antarcticus* and *P. leobergi.* However, there is always a clear-cut difference between intra- and interspecific divergence levels, and molecular diagnostic sites exist for all the species included in our sampling. Molecular identification using COI looks promising to help sort liparid species based on our data. Liparids are the second most species-rich group in the Southern Ocean (Eastman and Clarke, 1998), so establishing a

Author's personal copy

well-identified reference dataset is required before barcoding can be used routinely on this group, as already recognised by Rock et al. (2008).

Zoarcids raise a similar problem, and again, while COI looks promising, its results need to be compared to those of other markers and of morphology. There was a single discrepancy between molecular clusters obtained with barcoding and those resulting from morphological identification. Specimens identified as *Lycodichthys antarcticus* do not form a single cluster, as the basis is unresolved and includes the single barcoded specimen of *Lycenchelys* sp. (Fig. 1a). The specimens of *Lycenchelys* sp. are morphologically highly different from *L. antarcticus* and clearly represent a different species. The two molecular clusters obtained for *L. antarcticus* cannot at present be distinguished morphologically.

3.5. Identifications using the BOLD

The completeness of the reference database is one of the most important parameters for the obtention of reliable identification (Ekrem et al., 2007). It is not possible to obtain the correct identification if at least one sequence for the species is not present in the database, or if the available sequence is not identified to the species level. Many species collected during the CEAMARC cruises were represented by few or no specimens in BOLD. *Trematomus tokarevi, Pagetopsis maculatus, A. shackletoni* and *A. orianae,* as well as half of the liparid species were not represented by any sequences. Highly similar sequences were available for several others species, but without a species level identification: this was the case for the rest of the liparids (save *Paraliparis antarcticus)* and most of the zoarcids. For a number of other species, there were less than three sequences available in the BOLD (e.g. *Cygnodraco mawsoni, Gymnodraco acuticeps, Racovitzia glacialis, Vomeridens infuscipinnis, Artedidraco loennbergii* and *T. pennellii*). These generally posed no problem for the recovery of the identification using the search tool of the BOLD (except for the artedidraconids). The addition of the CEAMARC specimens brings all these species up to at least three specimens. Lastly, for most of the remaining species, there were no sequences available for the region sampled in our study. The majority of the specimens included at present in BOLD for the Southern Ocean come either from around the Antarctic Peninsula or the Ross Sea (FAO regions 48 and 88). The region investigated during the cruises is crucial for a complete circumpolar representation of the Antarctic Ocean and a better representation of intraspecific variability. It completes the dataset for the FAO region 58, moving several species towards the goal of at least 5 specimens per species and per FAO region.

While not all species of our sampling are represented in the database, there are closely related species for all. Moreover, the divergences are moderate within the included families, so family-level identification could be performed relatively well, contrary to many other groups (Ekrem et al., 2007).

The low level of divergence among clusters in some groups raises problems. When species have an insufficient level of interspecific divergence, an incorrect identification can look realistic even if the species is not present in the database. This is the case, for instance, for *A. shackletoni* and *A. orianae*, to which *Dolloidraco longedorsalis* is very close (see Fig. 3). However, for most of these species, the COI sequences do contain sufficient information for species identification through research of diagnostic sites in a sequence alignment or use of the clusters on an NJ tree (artedidraconids, for instance). A way to flag these groups in the database would be of interest, so a warning would appear with the results.

Attempted identifications of our only sequence of *Oneirodes notius* and the unidentified notosudid from the UM using the

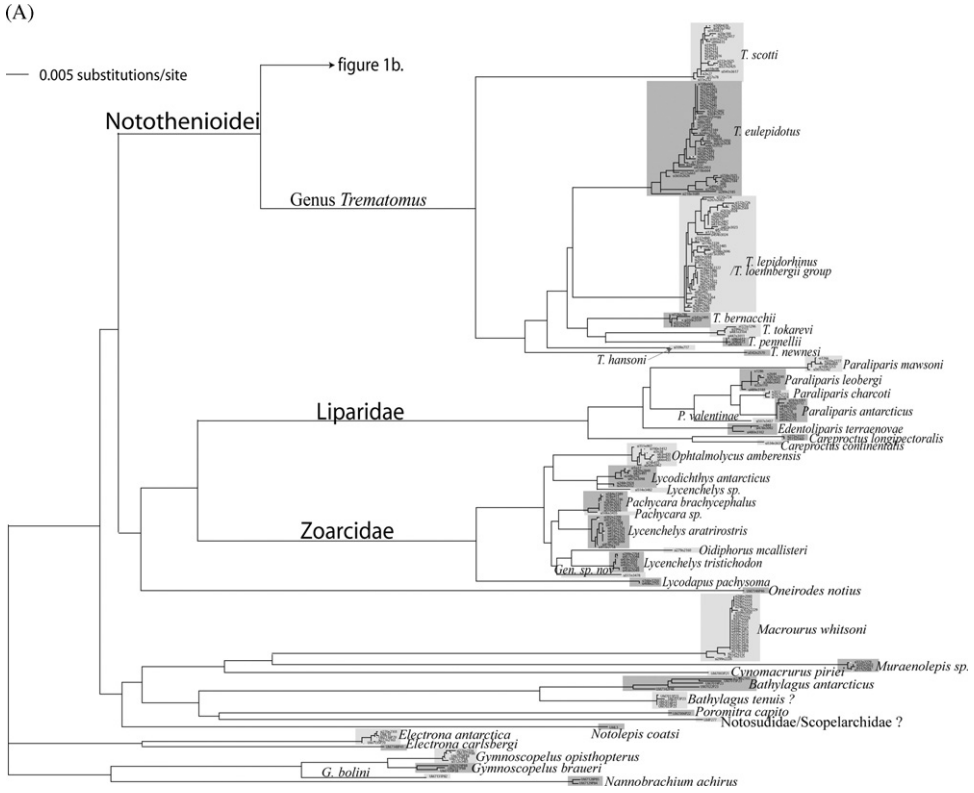

Fig. 1. NJ distance tree (K2p model) of the partial COI (652 bp, "barcode region") of 538 sequences from the CEAMARC cruises. The morphological identifications are given to the right. Arrows indicate juvenile specimens identified solely through their position in the tree.

Author's personal copy

Fig. 1. (*Continued*)

BOLD-ID tool encounter a similar problem. While for the first the best result is the single other available *O. notius* sequence of BOLD (100%), three other species are listed with a divergence of less than 1%, each with one or two sequences only. As no additional data is available on this genus, it is impossible to assess whether this is due to a low divergence for COI, a misidentification, or a faulty knowledge of the systematics of the group (i.e. several species recognised where a single evolutionary unit is present). For the unidentified notosudid specimen, the best hit was a single sequence of *Benthalbella elongata* (Scopelarchidae), with 98.54% similarity, but the next best hits are multiple sequences from other species of the same genus. It obviously needs to be re-identified and the situation clarified.

3.6. Small scale molecular taxonomy using COI

The initial lumping of barcoding and molecular taxonomy (see, for instance, Hebert et al., 2003) has lessened somewhat. Still, the sheer number of species for which data is available on COI would

Author's personal copy

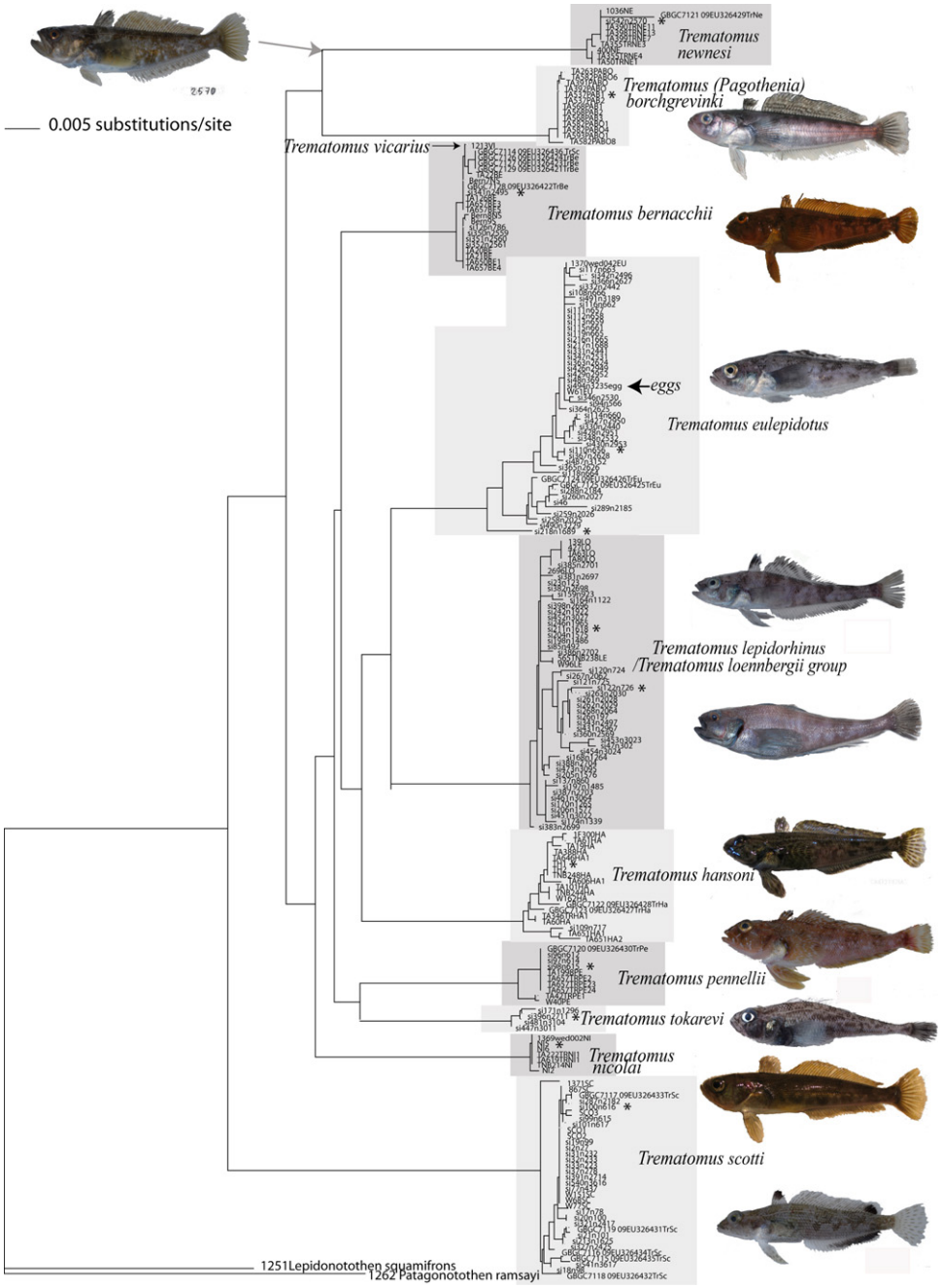

Fig. 2. NJ distance tree (K2p model) of the partial COI (571 bp, "barcode region") of 220 sequences from the CEAMARC cruises, with added specimens from ICOTA, POKER 2006, ICEFISH 2004 and from the BOL Database (from GBGC7114_09EU326436 to GBGC7129_09EU326421). The morphological identifications are given to the right. The queried specimens are indicated with *. Photo credit: S. Iglesias/CEAMARC/MNHN.

make it interesting for small scale molecular taxonomy if it proves to perform well as a phylogenetic marker. Therefore it is worth assessing the usefulness of COI for the molecular taxonomy of antarctic actinopterygians. In our opinion, the most interesting approach for small scale studies would be to use the COI for a first molecular approach of a group. Such a study should be combined with high quality identification and taxonomic knowledge, on a sampling as complete as possible. This is not different from what is best practice for any other marker, and the trees should be inferred through maximum parsimony, Bayesian inference or maximum likelihood, to avoid the numerous problems of NJ reconstructions (Leclerc et al., 1998; DeSalle et al., 2005). The areas of discrepancy between the COI trees and the morphological identifications/hypotheses can then help to pinpoint groups where a further study is needed. Sole use of a fixed depth of divergence among clusters to molecularly delineate new species would have missed most of the artedidraconid species, and some liparids, zoarcids and channichthyids in this case (false negative; Meyer and Paulay, 2005). Alternatively, if a minimum divergence level low enough to isolate these groups had been applied to the whole tree, it would have created multiple new and unwarranted groups in other clusters (false positive).

The COI findings are only the starting point for a study that should be enlarged to include independent datasets. COI, when studied alone, suffers from all the problems of a single mitochondrial marker: possible marker specific biases (Dettai and Lecointre, 2004), maternal only heredity precluding the detection of genetic exchanges, and, in some of the investigated

Author's personal copy

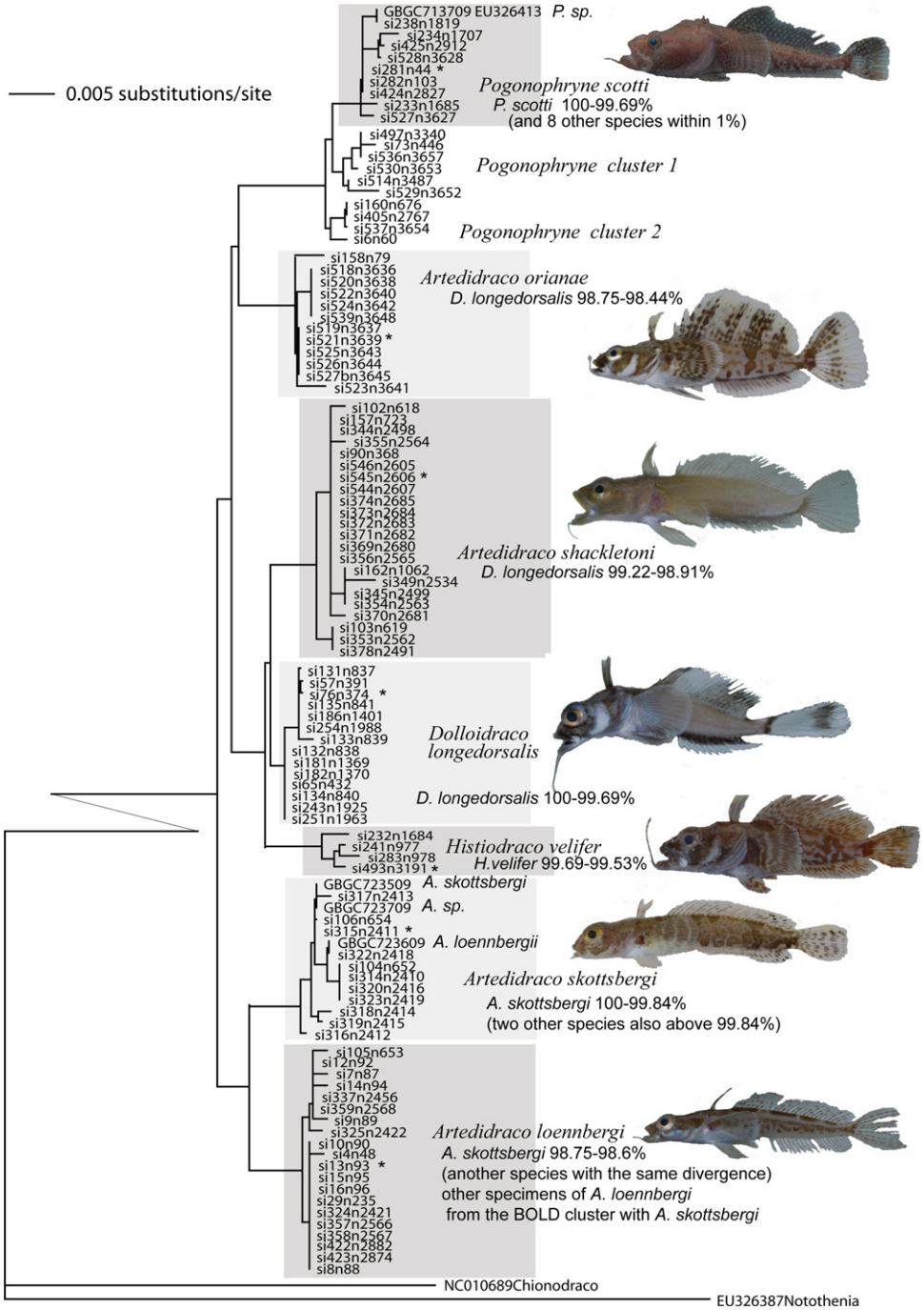

Fig. 3. NJ distance tree (K2p model) of the partial COI (652 bp, "barcode region") of 108 sequences from the CEAMARC cruises, sequences from GenBank (outgroups EU326387 *Notothenia rossii*; NC010689 *C. hamatus*) and from the BOL Database (GBGC713709_EU326413, GBGC723509_EU326315, GBGC723609_EU326314, GBGC723709_EU326313). The morphological identifications are given to the right. Below, the first results of query in BOLD with similarity range for the species in the database. The queried specimens are indicated with *. Photo credit: S. Iglesias/F. Busson/CEAMARC/MNHN.

groups, low variability or high heterogeneity in mutation rates among positions (Mueller, 2006; Rubinoff et al., 2006).

However, in all the cases investigated here, the results from COI are highly congruent with those from the morphological identification, from other mitochondrial markers (artedidraconids, liparids) or those from nuclear markers (genus *Trematomus*, notothenioids).

We encountered a borderline case within the artedidraconids, where the variability of COI is very low; thus, it might not be well suited for this group. However, it does contain informative sites, and is congruent with other data sources, although the inter-specific topology varies with the marker and the analysis method. *A. shackletoni* displayed a great variability in colour morphs. There is a variable density and intensity of spotting even among individuals captured in a single operation. Eastman and Eakin (1999, 2000) suggested that there could be cryptic or sibling species. We could not distinguish any distinct groups within this species with our markers, but here again their variability might have been insufficient. In contrast, *B. antarcticus* presents a large intraspecific divergence, whether among our specimens or among those present in the BOLD, compared to that of other fish species (Ward et al., 2009). However, while the morphological

Author's personal copy

identification of these specimens seemed reliable, no in-depth study has been undertaken yet. An investigation of the delineation of this species with nuclear and morphological data would be very interesting, as it might either reveal subdivisions or provide an example of deep molecular divergence within a species.

3.7. COI for phylogeny

Cytochrome oxidase I evolves rapidly, and did not perform well in studies comparing performances of mitochondrial markers for phylogenetics (Orti and Meyer, 1997; Mueller, 2006). However, comparison with other nuclear and mitochondrial markers at the intrafamilial level for antarctic teleosts (artedidraconids, genus *Trematomus*) displays a high level of congruence in the results in maximum parsimony (trees not shown, Lautredou et al., in press; Lecointre et al., unpublished data). COI is more congruent with nuclear markers (rhodopsin, and S7 as analysed by Near and Cheng, 2008) than the other mitochondrial markers used by Near and Cheng (2008). Added to the availability of a very large number of sequences with an existing link to a specimen and information about capture locality, this makes the partial COI gene sequence an interesting mitochondrial candidate for smaller scale phylogeny at least in Serraniformes. As always, the results have to be compared to those from independent markers or comparative morphology to assess reliability (Dettai and Lecointre, 2004).

3.8. Barcoding for collections

The present study underscores the need for a re-examination of the specimens in case of doubt, as a number of discrepancies between morphological and molecular identification turned out to be either identification errors or occasionally sequence contaminations (corrected through new PCR and sequencing). It is therefore critical to obtain a complete and representative collection of sequences with associated vouchers. Although there are already large and high quality collections in many institutions, most cannot be used in parallel for molecular and morphological studies. Whole specimens have generally been preserved in formalin and cannot be used for molecular studies, while for a variable proportion of the tissue collections the corresponding specimens were not kept, so new collections are needed. As pointed out by several authors (Ward et al., 2009), while sequencing COI for barcoding is not very expensive, managing the new collections and ensuring identifications is labour- and expertise-intensive, and represents the largest part of the associated cost. Molecular identification can help to obtain a highly reliable identification faster by either corroborating the identification or pinpointing the specimens that have to be rechecked without reassessing the entire collection. Moreover, the available sequences in a single repository (i.e. BOLD) "advertise" specimens of interest to taxonomists. This helps to provide better information for taxonomic revisions, but also ecological studies or species checklists, and can be a valuable investment for collection management.

4. Conclusion

COI appears to be a good choice for identification of antarctic actinopterygians, but two issues remain: practical identification problems (barcoding) and taxonomic problems.

Our results stress the necessity to check the fit between COI and other sources of data (nuclear markers, morphology) group by group, as well as the most relevant identification method before using the COI barcode as a routine identification tool. Controlling the clusters on the NJ tree can help detect problems and is a generally valuable tool, especially for closely related species with little divergence. However, full access to the sequences is also needed for a number of controls, for detection of diagnostic sites, and will become more powerful as more sequences are published and made fully available. Clustering methods can differ from phylogenetic methods under certain circumstances (Leclerc et al., 1998); analyses for a restricted sampling using other reconstruction methods would avoid the known pitfalls of NJ distance (DeSalle et al., 2005).

Identifications using BOLD, even for otherwise unproblematic groups, are highly dependent on the completeness of the database. The ongoing barcoding of the CEAMARC cruises adds a large amount of data for an under-sampled region. Even for the groups where the divergence level among species cannot be used for an identification cut-off, the position of the queried sequence in clusters or the use of diagnostic sites in COI can yield good results.

The problems detected here with some of the groups show the usefulness of additional markers, especially nuclear. Part of the data presented here is already associated with sequences from same specimens for several other markers, and other groups will be investigated in the near future (bathydraconids, zoarcids). When the morphological and the molecular identifications are not in agreement, additional markers can help conclude whether the problem is due to the limitations of COI or to an inaccurate knowledge of the taxonomy of the group. The partial rhodopsin retrogene (used by the FishTrace project), performs relatively well for this task. It is easy to amplify across a wide range of species with a single pair of primers, and we could not detect coalescence problems on the groups tested until now. Nonetheless, easy to sequence, more variable nuclear markers are badly needed.

While even restricted samplings can identify problems for some groups with the barcoding approach (for instance the bathydraconids and artedidraconids in the study from Rock et al., 2008), for other taxonomic groups, applying molecular taxonomy to a single cruise might not be sufficient to evaluate the performance of the approach. At present, sampling needs to be completed with as many species as possible; otherwise the conclusions drawn are incomplete and might even miss problems posed by closely related species pairs. This is clearly the case for the genus *Trematomus*, where only the inclusion of more species can detect potential problems missed by studies with limited sampling (Rock et al., 2008). Liparids are also an example where additional data has completely changed the biogeographic and systematic inferences that can be drawn. Studies with a limited sampling hinted at a small number of Southern Ocean clades and the monophyly of genera (Knudsen et al., 2007), whereas the inclusion of a large number of North Pacific and other specimens shows a much more complicated picture that will need further investigation (Duhamel et al., unpublished data). Distances among clusters are much less reliable to identify potentially interesting groups than morphology and anatomy, because average, minimum and maximum inter- and intraspecific distances are highly variable between related species, even when they are not overlapping (artedidraconids and trematomines, for instance). In our sampling, this approach gives rise to both false positives (Artedidraconidae) and possible false negatives (*B. antarcticus*).

Nevertheless, COI sequences and BOLD are overall valuable exploration tools for molecular taxonomy because of the unmatched species representativeness and reliability compared to other sequence databases. They will be used in the near future to identify the larvae from the R/V "Umitaka Maru". The better the specimens are identified morphologically, the more valuable the

Author's personal copy

system becomes. In such cases, the discrepancy between the molecular results and the morphological identification can help to pinpoint groups where additional taxonomic work would be warranted. This was the case in this study, highlighting the need to continue working on the taxonomy of antarctic actinopterygians, an area that superficially seemed clear-cut, but on a closer examination revealed numerous interesting grey areas.

Acknowledgments

We thank the crews and participants of the cruises involved in the capture of the samples, and especially the chief scientists of the CAML-CEAMARC cruises: G. Hosie, T. Ishimaru, M. Riddle and M. Stoddart. The CAML-CEAMARC cruises of R/V "Aurora Australis" and R/V "Umitaka maru" (IPY project no 53) were supported by the Australian Antarctic Division, the Japanese Science Foundation, the French polar institute IPEV, the CNRS, the MNHN and the ANR (White Project ANTFLOCKs USAR n07-BLAN-0213-01 directed by Guillaume Lecointre). The CAML and CNFRA allowed the first author to present part of these results at the CAML closing meeting in Genoa and at the 10th SCAR International Biology Meeting. G. Lancelot is thanked for his help for language corrections and many constructive comments. A. Froger is thanked for her help with the sequences. This project is a contribution to the EBA-SCAR programme.

This work was supported by the 'Consortium National de Recherche en Génomique', and the 'Service de Systématique Moléculaire' of the Muséum National d'Histoire Naturelle (IFR 101). It is part of the agreement number 2005/67 between the Genoscope and the Muséum national d'Histoire naturelle on the project 'Macrophylogeny of life' directed by Guillaume Lecointre.

Last, we thank the editors S. Schiaparelli and R. Hopcroft and the reviewers for their great help and comments. This is CAML contribution #40.

References

Aronson, R.B., Moody, R.M., Ivany, L.C., Blake, D.B., Werner, J.E., Glass, A., 2009. Climate change and trophic response of the Antarctic bottom fauna. PLoS One 4 (2), e4385.

Balushkin, A.V., Eakin, R.R., 1998. A new toad plunderfish *Pogonophryne fusca* sp. nova (Fam. Artedidraconidae: Notothenioidei) with notes on species composition and species groups in the genus /*Pogonophryne*/Regan. Journal of Ichthyology 38, 574–579.

Bely, A., Weisblat, D.A., 2006. Lessons from leeches: a call for DNA barcoding in the lab. Evolution and Development 8, 491–501.

Bortolus, A., 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. Ambio 37, 114–118.

Buhay, J.E., 2009. "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. Journal of Crustacean Biology 29, 96–110.

Clarke, A., Barnes, D.K.A., Hodgson, D.A., 2005. How isolated is Antarctica? Trends in Ecology and Evolution 20, 1–3.

DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation, and DNA barcoding. Philosophical Transaction of the Royal Society of London B Biological Science 360, 1905–1916.

DeSalle, R., 2006. Species discovery versus species identification in DNA barcoding efforts: Response to Rubinoff. Conservation Biology 20, 1545–1547.

Dettai, A., Lecointre, G., 2004. In search of the Notothenioid relatives. Antarctic Science 16, 71–85.

Eakin, R.R., 1990. Artedidraconidae. In: Gon, O., Heemstra, P.C. (Eds.), Fishes of the Southern Ocean. J.L.B. Smith Institute of Ichthyology, Grahamstown, South Africa, pp. 332–356.

Eakin, R.R., Eastman, J.T., 1998. New species of *Pogonophryne* (Pisces, Artedidraconidae) from the Ross Sea, Antarctica. Copeia 1998, 1005–1009.

Eakin, R.R., Eastman, J.T., Jones, C.D., 2001. Mental barbel variation in *Pogonophryne scotti* Regan (Pisces: Perciformes: Artedidraconidae). Antarctic Science 13, 363–370.

Eastman, J.T., 1993. Antarctic Fish Biology: Evolution in a Unique Environment. Academic Press, San Diego 322pp.

Eastman, J.T., Clarke, A., 1998. Radiations of Antarctic and non-Antarctic fish. In: Di Prisco, G., Pisano, E., Clarke, A. (Eds.), Fishes of Antarctica: A Biological Overview. Springer Verlag, Milan, pp. 3–26.

Eastman, J.T., Eakin, R.R., 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, J.T., Eakin, R.R., 1999. Fishes of the genus *Artedidraco* (Pisces, Artedidraconidae) from the Ross Sea, Antarctica, with the description of a new species and a colour morph. Antarctic Science 11, 13–22.

Ekrem, T., Willassen, E., Stur, E., 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Molecular Phylogenetic and Evolution 43, 530–542.

Eschmeyer, W.N., Fricke, R. (Eds.), 2009. Catalog of Fishes electronic version (9 September 2009). 〈http://research.calacademy.org/ichthyology/catalog/fishcatmain.asp〉.

Formalin Workshop Summary, 2006. Path to effective recovering of DNA from formalin fixed biological samples in natural history collections. In: Tang, E.P.Y. (Ed.), The National Academies Press, 〈http://www.nap.edu/openbook.php?record_id=11712〉.

Froese, R., Pauly, D. (Eds.), 2009. FishBase. World Wide Web electronic publication. 〈www.fishbase.org〉, version (10/2009).

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Research Symposium Series 41, 95–98.

Harris, D.J., 2003. Can you bank on GenBank? Trends in Ecology and Evolution 18, 317–319.

Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Philosophical Transactions of the Royal Society of London B Biological Science 270 (suppl), S96–S99.

Hickerson, M.J., Meyer, C.P., Moritz, C., 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Systematic Biology 55, 729–739.

Hureau, J.-C., 1966. Biologie comparée de quelques Poissons antarctiques (Nototheniidae). Thèse, Muséum national d'Histoire naturelle, 1391p.

Johnson, G.D., Paxton, J.R., Sutton, T.T., Satoh, T.P., Sado, T., Nishida, M., Miya, M., 2009. Deep-sea mystery solved: astonishing larval transformations and extreme sexual dimorphism unite three fish families. Biology Letters 5, 235–239.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16, 111–120.

Knudsen, S.W., Møller, P.R., Gravlund, P., 2007. Phylogeny of the snailfishes (Teleostei: Liparidae) based on molecular and morphological data. Molecular Phylogenetics and Evolution 44, 649–666.

Kon, T., Yoshino, T., Mukai, T., Nishida, M., 2007. DNA sequences identify numerous cryptic species of the vertebrate: a lesson from the gobioid fish *Schindleria*. Molecular Phylogenetics and Evolution 44, 53–62.

Koubbi, P., Duhamel, G., Heck, J.-H., Beans, C., Loots, C., Pruvost, P., Tavernier, E., Vacchi, M., Vallet, C., 2009. Ichthyoplankton in the neritic and coastal zone of Antarctica and Subantarctic islands: a review. Journal of Marine Systems 78, 547–556.

Kuhn, K.L., Near, T.J., 2009. Phylogeny of Trematomus (Notothenioidei: Nototheniidae) inferred from mitochondrial and nuclear gene sequences. Antarctic Science 21, 565–570.

Lautredou, A.-C., Bonillo, C., Denys, G., Cruaud, C., Ozouf-Costaz, C., Lecointre, G., Dettai, A., in press. Molecular taxonomy and identification within the Antarctic genus Trematomus (Notothenioidei, Teleostei): How valuable is barcoding with COI? Polar Science.

Leclerc, M.C., Barriel, V., Lecointre, G., De Reviers, B., 1998. Low divergence in rDNA ITS sequences among five species of *Fucus* (Phaeophyta) suggests a very recent radiation. Journal of Molecular Evolution 46, 115–120.

Lockley, A., Bardsley, R., 2000. DNA-based methods for food authentication. Trends in Food Science Technology 11, 67–77.

Meier, R., Zhang, G., Ali, F., 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Systematic Biology 57, 809–813.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biology 3 (12), e422.

Moore, S., 1999. Fluid preservation. In: Carter, D., Walker, A.K. (Eds.), Care and Conservation of Natural History Collections. Butterworth Heinemann, Oxford, pp. 92–132.

Mueller, R.L., 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in bayesian phylogenetic analysis. Systematic Biology 55, 289–300.

Naish, T., Powell, R., Levy, R., Wilson, G., Scherer, R., Talarico, F., Krissek, L., Niessen, F., Pompilio, M., Wilson, T., Carter, L., DeConto, R., Huybers, P., McKay, R., Pollard, D., Ross, J., Winter, D., Barrett, P., Browne, G., Cody, R., Cowan, E., Crampton, J., Dunbar, G., Dunbar, N., Florindo, F., Gebhardt, C., Graham, I., Hannah, M., Hansaraj, D., Harwood, D., Helling, D., Henrys, S., Hinnov, L., Kuhn, G., Kyle, P., Läufer, A., Maffioli, P., Magens, D., Mandernack, K., McIntosh, W., Millan, C., Morin, R., Ohneiser, C., Paulsen, T., Persico, D., Raine, I., Reed, J., Riesselman, C., Sagnotti, L., Schmitt, D., Sjunneskog, C., Strong, P., Taviani, M., Vogel, S., Wilch, T., Williams, T., 2009. Obliquity-paced Pliocene West Antarctic ice sheet oscillations. Nature 458, 322–328.

Near, T.J., Cheng, C.H., 2008. Phylogenetics of notothenioid fishes (Teleostei: Acanthomorpha): inferences from mitochondrial and nuclear gene sequences. Molecular Phylogenetics and Evolution 47, 832–840.

Author's personal copy

North, A.W., Kellermann, A., 1990. Identification key and catalogue of larval antarctic fish. In: Kellermann, A. (Ed.), Ber. Polarforsch. 67. and BIOMASS Scientific Series 10.

Orti, G., Meyer, A., 1997. The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Systematic Biology 46, 75–100.

Ozouf-Costaz, C., Hureau, J.C., Beaunier, M., 1991. Chromosome studies on fish of the suborder Notothenioidei collected in the Weddell Sea during EPOS 3 cruise. Cybium 15 (4), 275–290.

Pisano, E., Ozouf-Costaz, C., 2002. Cytogenetics and evolution in extreme environment: the case of Antarctic fishes. In: Val, A., Kapoor, G. (Eds.), Fish Adaptation. Oxford IBH Publishing, New Delhi, pp. 311–338.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system. Molecular Ecology Notes 7, 355–364.

Rock, J., Costa, F.O., Walker, D.I., North, A.W., Hutchinson, W.F., Carvalho, G.R., 2008. DNA barcodes of fish of the Scotia Sea, Antarctica indicate priority groups for taxonomic and systematics focus. Antarctic Science 20, 253–262.

Rubinoff, D., Cameron, S., Will, K., 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. Journal of Heredity 97, 581–594.

Steig, E.J., Schneider, D.P., Rutherford, S.D., Mann, M.E., Comiso, J.C., Shindell, D.T., 2009. Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. Nature 457, 459–462.

Steinke, D., Zemlak, T.S., Boutillier, J.A., Hebert, P.D.N., 2009. DNA barcoding of Pacific Canada's fishes. Marine Biology 156, 2641–2647.

Swofford, D., 2002. PAUP. Phylogenetic Analysis Using Parsimony (and other Methods), Version 4.0b10. Sinauer Associates, Sunderland, MA.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24, 1596–1599.

Thatje, S., 2005. The future fate of the Antarctic marine biota? Trends in Ecology and Evolution 20, 418–419.

Van de Putte, A.P., Van Houdt, J.K.J., Maes, G.E., Janko, K., Koubbi, P., Rock, J., Volckaert, F.A.M., 2009. Species identification in the trematomid family using nuclear genetic markers. Polar Biology 32, 1731–1741.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D., 2005. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London B Biological Science 360, 1847–1857.

Ward, R.D., Hanner, R., Hebert, P.D., 2009. The campaign to DNA barcode all fishes. FISH-BOL. Journal of Fisheries Biology 74, 329–356.

Winnepenninckx, B., Backeljau, T., De Wachter, R., 1993. Extraction of high molecular weight DNA from molluscs. Trends in Genetics 9 (1993), 407.

Wong, H.K., Hanner, R.H., 2008. DNA barcoding detects market substitution in North American seafood. Food Research International 41, 828–837.

Zemlak, T.S., Ward, R.D., Connell, A.D., Holmes, B.H., Hebert, P.D., 2009. DNA barcoding reveals overlooked marine fishes. Molecular Ecology Resources 9 (Suppl. 1), 237–242.