Accepted Manuscript

Title: DNA Barcoding and molecular systematics of the benthic and demersal organisms of the CEAMARC survey

Authors: Agnes Dettai, Sarah J. Adamowizc, Louise Allcock, Claudia P. Arango, David K.A. Barnes, Iain Barratt, Anne Chenuil, Arnaud Couloux, Corinne Cruaud, Bruno David, Françoise Denis, Gael Denys, Angie Díaz, Marc Eléaume, Jean-Pierre Féral, Aurélie Froger, Cyril Gallut, Rachel Grant, Huw J. Griffiths, Christoph Held, Lenaïg G. Hemery, Graham Hosie, Piotr Kuklinski, Guillaume Lecointre, Katrin Linse, Pierre Lozouet, Christopher Mah, Françoise Monniot, Mark D. Norman, Timothy O'Hara, Catherine Ozouf-Costaz, Claire Piedallu, Benjamin Pierrat, Elie Poulin, Nicolas Puillandre, Martin Riddle, Sarah Samadi, Thomas Saucède, Christoph Schubart, Peter J. Smith, Darren W. Stevens, Dirk Steinke, Jan M. Strugnell, K. Tarnowska, Victoria Wadley, Nadia Ameziane

PII: S1873-9652(11)00008-9

DOI: 10.1016/j.polar.2011.02.002

Reference: POLAR 117

To appear in: *Polar Science*

Received Date: 14 September 2010

Revised Date: 23 December 2010

Accepted Date: 10 February 2011

Please cite this article as: Dettai, A., Adamowizc, S.J., Allcock, L., Arango, C.P., Barnes, D.K., Barratt, I., Chenuil, A., Couloux, A., Cruaud, C., David, B., Denis, F., Denys, G., Díaz, A., Eléaume, M., Féral, J.-P., Froger, A., Gallut, C., Grant, R., Griffiths, H.J., Held, C., Hemery, L., Hosie, G., Kuklinski, P., Lecointre, G., Linse, K., Lozouet, P., Mah, C., Monniot, F., Norman, M.D., O'Hara, T., Ozouf-Costaz, C., Piedallu, C., Pierrat, B., Poulin, E., Puillandre, N., Riddle, M., Samadi, S., Saucède, T., Schubart, C., Smith, P.J., Stevens, D.W., Steinke, D., Strugnell, J.M., Tarnowska, K., Wadley, V., Ameziane, N. DNA Barcoding and molecular systematics of the benthic and demersal organisms of the CEAMARC survey, Polar Science (2011), doi: 10.1016/j.polar.2011.02.002

DNA Barcoding and molecular systematics of the benthic and demersal organisms of the CEAMARC survey

Agnes Dettai[a,*], Sarah J. Adamowizc[b], Louise Allcock,[c,d], Claudia P. Arango[e], David K.A. Barnes[f], Iain Barratt[d], Anne Chenuil[g], Arnaud Couloux[h], Corinne Cruaud[h], Bruno David[i], Françoise Denis[j], Gael Denys[a], Angie Díaz[k], Marc Eléaume[a], Jean-Pierre Féral[g], Aurélie Froger[a], Cyril Gallut[a], Rachel Grant[l], Huw J. Griffiths[f], Christoph Held[m], Lenaïg G. Hemery[a], Graham Hosie[n], Piotr Kuklinski[o,p], Guillaume Lecointre[a], Katrin Linse[f], Pierre Lozouet[a], Christopher Mah[q], Françoise Monniot[a], Mark D. Norman[r], Timothy O'Hara[r], Catherine Ozouf-Costaz[a], Claire Piedallu[a], Benjamin Pierrat[i], Elie Poulin[k], Nicolas Puillandre[a], Martin Riddle[n], Sarah Samadi[a], Thomas Saucède[i], Christoph Schubart[s], Peter J. Smith[t], Darren W. Stevens[t], Dirk Steinke[b], Jan M. Strugnell[u], K. Tarnowska[g], Victoria Wadley[n], Nadia Ameziane[a]

[a] Muséum national d'Histoire naturelle (MNHN), CP 26, 43, rue Cuvier, 75231 Paris Cedex 05 France.
[b] University of Guelph, Biodiversity Institute of Ontario, Canadian Centre for DNA Barcoding, Guelph, Ontario, N1G 2W1, Canada.
[c] Martin Ryan Marine Science Institute, National University of Ireland Galway, University Road, Galway, Ireland.
[d] School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK.
[e] Biodiversity Program, Queensland Museum, PO Box 3300, South Brisbane, 4101 Queensland, Australia.
[f] British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK.
[g] Université de la Méditerranée-CNRS, UMR 6540 DIMAR, COM-Station Marine d'Endoume, Marseille, France.
[h] Genoscope, Centre National de Séquençage, 2, rue Gaston Crémieux, CP5706, 91057 Évry cedex, France
[i] Université de Bourgogne-CNRS, UMR 5561 Biogéosciences, Dijon, France.
[j] Université du Maine Av. O. Messiaen 72000 Le Mans -France.
[a] Muséum national d'Histoire naturelle (MNHN), CP 26, 43, rue Cuvier, 75231 Paris Cedex 05 France.
[k] Instituto de Ecología y Biodiversidad, Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile.
[l] Scott Polar Research Institute, University of Cambridge, Lensfield Road Cambridge, CB2 1ER, UK,
[m] Alfred Wegener Institute for Polar and Marine Research, Functional Ecology, Building D, room D-2200, Am alten Hafen 26, D-27568 Bremerhaven.
[n] Australian Antarctic Division, Department of Sustainability, Environment, Water, Population and Communities, 203 Channel Highway, Kingston, Tasmania 7050, Australia.
[o] Natural History Museum, Cromwell Road, London, SW6 5BD, UK.

[p] Institute of Oceanology, Polish Academy of Sciences, ul. Powstancow Warszawy 55, Sopot 81-712, Poland.
[q] Smithsonian Institution, National Museum of Natural History, Dept Invertebrate Zool, Washington, DC 20007 USA.
[r] Museum Victoria, Melbourne, Vic 3001 Australia.
[s] Biologie 1, Universität Regensburg, 93040 Regensburg, Germany.
[t] National Institute of Water and Atmospheric Research (NIWA), Private Bag 14901, Wellington 6241, New Zealand.
[u] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

[*]Corresponding author. Muséum national d'Histoire naturelle (MNHN), Département Systématique et Evolution, UMR 7138,CP 26, 43, rue Cuvier, 75231 Paris Cedex 05 France.

E-mail address: adettai@mnhn.fr

***Keywords*:** barcode, Cytochrome Oxidase 1, CEAMARC, taxonomy, systematics

Abstract

The Dumont d'Urville Sea (East Antarctic region) has been less investigated for DNA barcoding and molecular taxonomy than other parts of the Southern Ocean, such as the Ross Sea and the Antarctic Peninsula. The Collaborative East Antarctic MARine Census (CEAMARC) took place in this area during the austral summer of 2007-2008. The Australian vessel RSV *Aurora Australis* collected very diverse samples of demersal and benthic organisms. The specimens were sorted centrally, and then distributed to taxonomic experts for molecular and morphological taxonomy and identification, especially barcoding. The COI sequences generated from CEAMARC material provide a sizeable proportion of the Census of Antarctic Marine Life barcodes although the studies are still ongoing, and represent the only source of sequences for a number of species. Barcoding appears to be a valuable method for identification within most groups, despite low divergences and haplotype sharing in a few species, and it is also useful as an preliminary taxonomic exploration method. Several new species are being described. CEAMARC samples have already provided new material for phylogeographic and phylogenetic studies in cephalopods, pycnogonids, teleost fish, crinoids and sea urchins, helping these studies to provide a better insight in the patterns of evolution in the Southern Ocean.

1. Introduction

Regardless of the type of biological study, reliable knowledge of the taxonomy and identification of the organisms under scrutiny form a fundamental basis for all other kinds of knowledge produced (Bickford et al., 2007; Bortolus, 2008). This can only be ignored at the expense of the reliability and accuracy of all subsequent data (Bely and Weissblat, 2006; Bortolus 2008). However accurate identifications can take a long time to determine depending on the taxonomic group, and most higher-level taxa (i.e. genus, family, order) still contain many un-described, or ill-defined species. There is a need to undertake a detailed re-evaluation of such taxa, including both morphological and molecular assessments, and if possible additional geographical, ecological and biological information should be gathered (Galtier et al., 2009; Padial et al., 2010).

Identification and taxonomy based on molecular data have been around for almost as long as the multiple molecular methods that support them (Teletchea, 2009). However, larger scale projects with stringent quality control are more recent. The Barcode of Life international project (Hebert et al., 2003) and its corresponding database, the Barcode of life database BOLD (Ratnasingham and Hebert, 2007) is the largest with a focus on taxonomy. This project relies on the sequencing of standardized gene regions (e.g. cytochrome oxidase I [COI] for most metazoan groups). Molecular identification is then performed through a comparison with publicly accessible reference datasets (Ratnasingham and Hebert, 2007). Sequences in BOLD undergo more stringent control to avoid problems of unreliable data (particularly regarding the taxonomic identifications) that are encountered in other sequence repositories (Harris, 2003; Nilsson et al., 2006). In particular, individual sequences are linked to their individual voucher specimens (that is, the actual specimens from which the sequences were obtained). The storage of the voucher specimens and their link to the COI sequences allow us to perform parallel morphological and molecular studies, as well as re-examinations of the individual specimen, should its identification or the systematics of the group be questioned. The molecular reference database relies on collaboration with taxonomists to provide

reliable morphological identifications (Hajibabaei et al., 2005). In the last seven years, a number of studies have highlighted the potential usefulness of barcoding for preliminary studies in molecular taxonomy, the flagging of cryptic species, and the identification of the full range of known species in addition to molecular identification (Hebert et al., 2003; Ward et al., 2005; Bucklin et al., 2007; Ward et al., 2008, 2009; Steinke et al., 2009; Valentini et al., 2009; Hunt et al., 2010). International projects like MarBOL (http://www.marinebarcoding.org/) focus on enhancing identification for marine organisms (Bucklin et al., 2011), and are nested within the wider international Barcode of Life initiative (http://ibol.org/).

In comparison to tropical and temperate regions, biodiversity in polar regions has been poorly studied (Grant et al., in press). In addition, the West Antarctic Peninsula is, along with the Arctic, among the fastest warming places on the planet (Clarke et al., 2005; Thatje, 2005; Aronson et al., 2009). Therefore, reliable identification of specimens and flagging of taxonomic problems for more in-depth studies are more necessary than ever. Molecular barcoding can provide a valuable source of data for this purpose, and provide preliminary indication of genetic population structure within Antarctic species in and around the Southern Ocean (Ward et al., 2009; Bucklin et al., 2010). The international initiatives of the Census of Antarctic Marine Life (CAML, www.caml.aq) and International Polar Year 2007-2008 (www.ipy.org) coordinated several much-needed large-scale collaborative marine surveys in the Southern Ocean. As the Australian-French-Japanese-Belgian contribution to CAML, the Collaborative East Antarctic MARine Census (CEAMARC) took place off George V Land and Terre Adélie during the Austral summer of 2007-2008. Three vessels participated in this project with three voyages, the Australian RSV *Aurora Australis*, the French RV *Astrolabe*, and the Japanese TRV *Umitaka Maru*. While the voyages with RV *Astrolabe* and TRV *Umitaka Maru* focused on the pelagic environment (plankton, krill, fish and oceanography), the RSV *Aurora Australis* collected a wide range of benthic and demersal organisms (Beaman and O'Brien, 2009). The area was chosen because preliminary studies suggested it included a wide diversity of benthic habitats (Beaman and Harris, 2005), and because it had been subject to regular

biodiversity studies in shallower water, both ancient (see Arnaud, 1974 for an overview of the local biodiversity), and more recent, during the IPEV (Institut Paul Emile Victor) n°281 program: Ichtyologie Côtière en Terre Adélie, or ICOTA (Koubbi et al., 2001, 2009). However, the scale of the CEAMARC survey and the variety of depths investigated exceeded the geographic coverage previously investigated in the area during ICOTA. In particular, it is the first time that collections in the area exceeded depths of 200 m for molecular studies. Greater depths had already been highlighted as important but little investigated areas in previous barcoding projects surveys (Grant and Linse, 2009). The CEAMARC survey covered a marine region on the less explored eastern side of the Antarctic continent. The CEAMARC specimens therefore complement estimates from other, distant collecting sites (off Western Antarctica) of the Southern Ocean for molecular and morphological investigation of intra-specific and inter-specific diversity.

Barcoding studies, like taxonomic studies, optimally require well preserved and well documented specimens. This is not always the case in older campaigns, where the specimens have rarely been preserved in conditions suitable for molecular studies because of the wide use of formalin as a fixative (Moore, 1999), although recent studies show promising recovery of genetic material even from formalin-fixed specimens (Palero et al., 2010; Zhang 2010). In parallel, in the early period of molecular systematics, researchers generally felt confident with identifications attached to the deposited sequences, so specimens sampled for molecular studies were most often not kept for further reference. Surveys like the CEAMARC provide higher quality new samples than many existing older collections. Specimens were preserved to maximise their usability for both molecular and morphological approaches, and crucially, many aspects of the physical environment investigated during the survey are linked to the specimen data (Beaman and O'Brien, 2009). Since no single institution can provide expertise for the in-depth study of all the taxonomic groups that are collected during an extensive census of marine life like the CEAMARC, there is a chance that a part of the collected materials may remain untouched in museum collections for decades. Thus, the CEAMARC survey was organized with international collaboration in mind, and to maximise the

probability that most specimens are identified, taxonomists were contacted prior to the departure of the cruise.

The aim of this paper is to provide an overview of the present barcoding and molecular taxonomy results using the COI gene on the benthic material collected by the CEAMARC survey. It makes reference to every paper available at the end 2010 on this material; non-referenced examples are the latest developments of the studies, and unpublished to date.

2. Material and methods

The CEAMARC survey on RSV *Aurora Australis* sampled demersal and benthic organisms on the Antarctic continental shelf and slope between longitudes 138°E and 146°E. The sampling strategy was designed to cross physical (water, sediment) parameters and biological results from community to genes (Beaman and O'Brien, 2009), by keeping track of the species and specimen associations in each community as well as sampling largely some species for population genetics. Stations covered a regular grid over the study area, in order to optimize the interpolation of physical data (Figure 1). The sampling equipment included two models of beam trawl (87 collecting events using the Australian Antarctic Division beam trawl, 3.02 m wide x 1.39 m high, from between 138 to 2065 m deep and 13 collecting events using a low profile beam trawl, 4.2 m wide x 0.5 m high, from 151 to 1595 m deep), sled trawls (6 collecting events, 314 to 558 m deep), and both Smith-MacIntyre (3 collecting events; 182 to 241m deep) and Van Veen grabs (9 collecting events; 192 to 793m deep) to characterise the sediments, as well as a large number of CTDs (see Beaman and O'Brien, 2009 for a detailed description of the sampling patterns and events).

To record occurrence of elusive and fragile species and precisely categorize biological communities, videos and still pictures were collected with most trawls. The Australian Antarctic Division and the Geoscience Australia underwater video transects were studied in combination with the physical parameters measured during the survey (Post et al., 2010). The catch can therefore be

linked with the benthic assemblages characterised using the videos. Taxa for which taxonomic experts had indicated a special interest and provided a sampling protocol were well represented in number and diversity. Specimens for these taxonomic groups were isolated more efficiently, better preserved and when possible, were identified on board. Specimens of most teleost fish species and some macro-invertebrates (cephalopods, ascidians, hexacorals and sponges) were preserved on board in formalin after taking a subsample in 85% ethanol for molecular analysis. All other groups had at least some specimens preserved in 80% ethanol. Crustaceans, crinoids and a few others, were fixed in ~95% ethanol and preserved in 80% ethanol. Photographs of as many fresh specimens as possible were taken during sorting on board. Taxonomic groups were sorted to morpho-species and repackaged at the Muséum national d'Histoire naturelle in Paris (MNHN). The resulting sorted lots (containing specimens from one collection event and one morpho-species as identified by a non-specialist) were recorded in the collection databases and sent to experts in Europe, USA and Australia for confirmation and formal identification. The content of a lot changed depending on the stage of the sorting; when prepared on board a lot contained one or several specimens from the same larger-scale taxonomic group (often phylum), when sorted at the MNHN (by a competent non-specialist) a lot contained specimens from the same morpho-species. In some cases, these contained several species, as it is easy for a non-specialist to overlook the finer differences between similar looking species. Once it has been studied in depth by the specialist of the group, a lot contains one or more specimens from the same species. All the specimens used for barcode studies were isolated in separate lots, or individually labelled within multi-specimen lots.

Specimens were, and continue to be, identified morphologically by taxonomic specialists (Allcock et al., 2010; Arango et al., 2010; Dettai et al., in press; Duhamel et al., 2010; Monniot et al., in press). The barcode data was produced using the standardized 650 bp fragment of the cytochrome c oxidase I (COI) gene as the primary marker. In collaboration with MarBOL, the Canadian Centre for DNA Barcoding (CCDB) and the Census of Antarctic Marine Life (CAML), part of the molecular work for the CEAMARC samples has been done through submission of the

samples to the CCDB since January 2009 (Allcock et al., 2010; Arango et al., 2010). Some barcode data were sequenced using local sequencing facilities, especially (but not exclusively) the Genoscope (French National Center for Sequencing). Specific preservation, amplification, sequencing and analysis protocols have been adapted for each of the groups under study (Dettai et al., in press; Chenuil et al., 2010; Monniot et al., in press), following the recommendations for the Barcode of Life Project (Hajibabaei et al., 2005; Ivanova et al., 2006; Ward et al., 2008, 2009, MarBOL online instructions), including photographs of the sequenced specimens. Previously described primers (Folmer et al., 1994; Lee et al., 2004; Ward et al., 2005, 2008, 2009; Stefaniak et al., 2009; Hoareau and Boissin, 2010), as well as new primers were used for amplification of COI (Dettai et al., in press; Diaz et al., in press; Monniot et al., in press). Additional DNA markers (Table1) are being sequenced using local resources for most taxa, in order to provide a complementary input for complex cases and integrative taxonomy (Chenuil et al., 2010; Lautredou et al., 2010; Lecointre et al., in press). For the taxonomic groups for which molecular data are available already, several return verifications were performed between the molecular and the morphological datasets to check for mistakes and investigate further unexpected results. The size of the data list of species continues to grow as sample analysis and identification continues. All sequences are being deposited in BOLD and all data including a species list are contributed to the SCAR Marine Biodiversity Information Network (SCAR-MarBIN) (De Broyer and Danis, 2009; www.scarmarbin.be).

To provide an overview of the CEAMARC barcoding results, the COI sequences of the 2678 CEAMARC specimens currently integrated in BOLD were analysed by depth, location of collection and taxonomic group. The depth distribution of the CEAMARC barcodes was analysed by splitting the Southern Ocean into 500m depth zones (Grant et al., in press). Broad phylogenetic coverage was plotted by breaking the specimens into groups based on phylum. Geo-referenced locations of CEAMARC specimens in the Southern Ocean were also plotted (Fig. 1). For the phylogeographic analysis, the Southern Ocean was divided into boxes of 0.5° latitude x 0.5° longitude (Grant et al.,

in press) while taking into account the smaller geographical scale of the sole CEAMARC survey. The number of phyla, classes, genera and species were counted and colour-coded with white representing absence, blue for low taxon number counts and red for high taxon number counts in each box (Fig. 2).

3. Results

3,630 preliminary lots containing from one to dozens of specimens from 19 phyla were prepared on board RSV *Aurora Australis*. Teleost fishes (975 lots), echinoderms (839 lots), molluscs (421 lots), crustaceans (327 lots) and pycnogonids (360 lots) dominated the sampling. The number of lots increased at each step of the sorting and identification process. For teleosts, for which experts were present on board, the increase was small, from 975 lots determined on board to 998 after the in-depth identification process. For ascidians, the number of lots increased from 89 to 271 and for pycnogonids the number of lots increased from 360 lots to 650 lots. Mistakes were made in the first groupings, for instance with holothuroids, hemichordates, nemerteans, echiuran and sipunculid worms were included in some ascidian lots. The currently tallied number of specimens is 29,576, including formalin preserved specimens. This number is expected to increase as the last remaining taxa are still undergoing in-depth study. Table 1 lists the current number of lots and specimens for each taxonomic group.

The CEAMARC voyages contributed 2,678 specimens to date to the BOLD. However, most specimens from the survey are currently being identified, sequenced and analysed, and have not been entered yet into BOLD. The number of barcodes currently obtained for each taxonomic group is given in Table 1. Barcodes are available for all the stations sampled by the RSV *Aurora Australis* (Fig 1).

Almost all currently barcoded East Antarctic samples were collected during the CEAMARC survey. To date CEAMARC is the sole source of barcodes for a considerable number of Antarctic

species in BOLD (Table 1). For instance, this is the case for 22 of the 72 teleost species, especially within Liparidae and Zoarcidae, and CEAMARC samples provide more than half of the sequences for 14 more species. In molluscs (Polyplacophora, Scaphopoda, Nudibranchia, Bivalvia, Gastropoda), searches in the BOLD database found sequences divergent by less than 2% for only 27% of CEAMARC sequence clusters. The extreme situation is Ascidiacea, where few of the more than 4000 described species have been sequenced to date. From the 31 species identified morphologically, 26 had no specimens entered in BOLD. Only two species collected during the survey already had sequences present in BOLD (*Cnemidocarpa verrucosa* (Lesson, 1830) and *Corella eumyota* Traustedt, 1882), and sequences from CEAMARC specimens diverged from the databased sequences by more than 14% (Monniot et al., in press).

New species are being described in several taxonomic groups (table 1). Most descriptions contain the reference of the barcode sequence for the type specimen in addition to the morphological description (Zoarcidae sp., *Nymphon* spps). Two of the new ascidian species have been sequenced for COI (Monniot et al., in press) and show more than 20% divergence from the other species sequenced in the same family. The specimens from the new species were collected between 400 and 820m.

Of the CEAMARC specimens already barcoded, the majority (59%) were collected between 138 and 500 m, with less than 2% being collected below 1550 m. These numbers are similar to those obtained across all samples (56% between 138 and 500 m, 2.5% below 1500 m). Specimens already integrated in BOLD came from four phyla; Mollusca (5%), Arthropoda (19%), Chordata (mainly teleost fish) (24%) and Echinodermata (53%). The difference with the proportions in the overall collection is due to the different states of progress of the various projects, since some taxa were only recently sent to the specialists of the group. Due to these group-specific differences, few (max=4) phyla are generally represented by barcode sequences per 0.5° x 0.5° area boxes (fig 2). However, there is a number of stations where the number of species for which sequences are available is also very low, while the sequencing is completed for a number of taxonomic groups

(teleost fishes, cephalopods, crinoids). The number of genera per box (max=29) and of species per box (max=36) were positively correlated, but were in poor agreement with the number of phyla per box and of class per box.

4. Discussion

Whilst amplifying and sequencing COI is generally straightforward, technical problems arose within many taxonomic groups, and can bias the representation of taxa in BOLD. This is especially true for such a large barcoding effort covering multiple phyla. Even with multiple pairs of primers, the sequences for some groups remain highly challenging to obtain. In these cases, the use of COI for routine identification by users might be compromised by the difficulty of obtaining the sequences in the first place. This may be the case for the ascidians, in which only a minority of the collected species could be amplified and sequenced easily. The problems are partially due to composition biases and poly-Ts in a number of species (Monniot et al., in press). The amplification of some mollusc groups like Bivalvia also posed considerable problems. Preliminary studies using CEAMARC specimens of *Euphausia superba* (Dana, 1852) and *E. crystallorophias* Holt and Tattersall 1906, found several distinct copies of COI in some of the individuals. Some of these copies appear to be nuclear pseudo-genes, as already described in other crustacean groups (Buhay, 2009).

4.1. Identification

Results from multiple studies including or based on CEAMARC specimens point out that for many groups, barcoding is a valuable tool for the identification of marine specimens from the Southern Ocean (Arango et al., 2010; Dettai et al., in press; Lautredou et al., 2010; Monniot et al., in press; Smith et al., 2008). A small number of juvenile specimens and eggs which could not be identified morphologically were identified by comparing COI sequences against the reference sequences held in BOLD (Dettai et al., in press), confirming that molecular identification can help

to assign these teleost specimens to the adults of their species (Webb et al., 2006; Ward et al., 2009; Valdez-Moreno et al., 2010). Eighteen larger or degraded crinoid specimens misidentified as *Solanometra antarctica* (Carpenter, 1880) could be attributed to *Florometra mawsoni* Clark, 1937 after comparison of their COI sequences to sequences from a large, well identified East Antarctic (type locality for *F. mawsoni*) based dataset. Despite the presence of pseudo-genes, CEAMARC *Euphausia superba* and *E. crystalloraphias* specimens can be distinguished using COI data, which is particularly important as these species are often damaged in the trawls. Less fragile morphological characteristics are also being investigated.

The efficiency of identification through barcoding depends on the completeness of the reference database (Ekrem et al., 2007; Puillandre et al., 2009), which is taxon dependent (Grant et al., in press). In some groups like Cephalopoda, Asteroidea, Pycnogonida or Teleostei (Dettai et al., in press), the coverage is good (see Table 1), with a majority of species collected during the CEAMARC already represented in BOLD are already included in the database. In other groups like crinoids, non-cephalopod molluscs, bryozoans or ascidians (Monniot et al., in press), few if any of the sequences obtained for CEAMARC specimens had a close match with BOLD COI sequences or even with 18S rDNA in GenBank. For these groups, the acquisition and inclusion of many reference sequences that represent the diversity of the Southern Ocean will be necessary before routine identification can be performed.

The level of divergence between species and the overlap between intra-specific and inter-specific variation is a crucial parameter for species identification. Distance based species recognition depends on the barcode "gap" between intra- and inter-specific variation, and while the barcode "gap" has provided a valuable tool for species identification and for highlighting cryptic species (Ward et al., 2005, 2009), substantial overlap in intra- and inter-specific variation has been reported in some groups. A clear barcoding gap seems to be present in CEAMARC mysids, with intra-specific variability representing around 1/10th of inter-specific variability that has been suggested to be optimal (Hebert et al., 2004). The difference is much smaller in teleosts overall

(Dettai et al., in press). Distinct haplotypes but small inter-specific divergences have been found in several taxa, like Artedidraconidae (some species are separated by less than 1%), and *Paraliparis* in teleosts (Dettai et al., in press; Duhamel et al., 2010; Lecointre et al., in press) or *Pareledone* species in cephalopods (Allcock et al., 2010). In such cases, Ward et al. (2009) recommended checking the clusters with a more rapidly evolving marker like the mitochondrial control region. For the artedidraconids, clusters are the same using the mitochondrial cytochrome b and D-loop, and, more importantly, using a nuclear marker (Lecointre et al., in press), suggesting that the COI database is useful for the assignment of a specimen to a species in this group despite the very low divergence. Similarly, two morphological species of the echinoderm genus *Ctenocidaris* appear to have a divergence of about 1.4%. Antarctic cnidarids, like Antarctic schizasterids, may have an extremely slow rate of evolution (Lockhart, 2006; Chenuil et al., 2009). The same taxa often also show an overlap of the range of intra-specific variation with the range of inter-specific variation, and both these ranges can vary greatly, even among closely related species (Allcock et al., 2010; Dettai et al., in press; Undheim et al., 2010). In Southern Ocean cephalopods, within species with a large geographic range like *Pareledone aequipapillae* Allcock, 2005 the divergence of haplotypes is as large as the apparent divergence between some more geographically restricted species-pairs (Allcock et al., 2010). Hence intra-specific diversity of a species cannot be deduced from data available for other species and the current practice in large-scale barcode sampling to collect five specimens per species might be insufficient to assess the intra-specific diversity and the existence of a barcoding gap (Allcock et al., 2010).

4.2. Taxonomic problems and cryptic species

Identification through barcoding requires specimens from the same species to cluster together using the barcode markers. When knowledge of the species delineation is faulty (either cryptic species or single species where several species were recognized previously), the COI clusters will disagree with the identification based on the morphological identification criteria. Further studies

are necessary to identify whether the source of the problem is with the species delineation or with COI as a marker, through insufficient variability or nuclear copies (Buhay, 2009; Ward et al., 2009). Preliminary studies including CEAMARC specimens have detected COI haplotype sharing between different morphologically defined species in various taxonomic groups. Haplotype sharing can reflect either very recent divergence among species, incomplete lineage sorting, incorrect delimitation of the species boundaries, or mitochondrial introgression (Galtier et al., 2009; Ward et al., 2009). Either way, it poses problems for identification, and needs further investigations integrating multiple sources of data. For instance the teleost fish *Trematomus loennbergii* Regan, 1913 and *T. lepidorhinus* (Pappenheim, 1911) cannot be distinguished using COI nor any nuclear marker tested to date (Lautredou et al., 2010). Although both are currently considered valid species, the morphological distinctions also proved problematic, so no conclusion can be drawn at the time on the status of these species until morphology and more variable nuclear markers are investigated (Lautredou et al., 2010). Species from the echinoid genus *Notocidaris* collected during the CEAMARC and analysed with COI and the internal transcribed spacer (ITS) of the ribosomal DNA cluster revealed a striking pattern: the three species of *Notocidaris* share the same haplotypes and cannot be distinguished by molecular markers (fig. 3). Identification is based on spine morphology (David et al., 2005). Intra and inter specific variations of spine morphology do not overlap between *Notocidaris gaussensis* Mortensen, 1909 , and the *Notocidaris platyacantha* (H.L. Clark, 1925) */N. remigera* Mortensen, 1950 group. However, there is no easily visible morphological gap among morphologies, so that extreme individuals might be misidentified. So far those results do not allow the conclusion that the morpho-species are not valid, especially for the separation of *Notocidaris gaussensis* from the two other species, since such a pattern can be observed for recently diverged species and thereby reflect ancestral polymorphism or introgression. However the lack of both morphological and molecular divergences between *Notocidaris platyacantha* and *N. remigera* might indicate that they form a single group, as discussed by David et al. (2005). The three echinoids *Sterechinus diadema* (Studer, 1876), *S. agassizii* Mortensen, 1910 and *S. antarcticus* Koehler, 1901

also share COI haplotypes, despite the geographic distance between the analysed samples, and might represent a single species (Diaz et al., 2010).

Conversely, studies have suggested for quite some time the presence of cryptic species in many Antarctic groups (Allcock et al., 1997; Linse et al., 2007; Wilson et al., 2007; Hunter and Halanych 2008; Leese and Held 2008; Krabbe et al., 2010), with all the problems this can cause in further studies (Bickford et al., 2007). The CEAMARC samples of *Promachocrinus kerguelensis* Carpenter, 1879 were the first crinoids sequenced from Terre Adélie. Preliminary studies show that they display a very high intra-specific diversity (average 3%, max 7,8%), and intra-specific genetic clusters similar to those of Wilson et al. (2007) for the same species, suggesting the presence of cryptic species in this region and over the whole Southern Ocean. In Ascidiacea, two sequenced specimens of *Pyura bouvetensis* (Michaelsen, 1904) differed by more than 13%, although both were collected by the CEAMARC survey. More specimens need to be sequenced to assess intraspecific molecular variability, about which little is known in Antarctic ascidians (Monniot et al., in press). The deep divergence might indicate the presence of cryptic species in this case. In the case of the deep divergence between specimens already present in BOLD and our specimens (*Cnemidocarpa verrucosa* and *Corella eumyota*), the same might be true. Several cryptic species were identified among cephalopods by Allcock et al. (2010). In the nudibranch *Austrodoris kerguelenensis* (Bergh, 1884), preliminary results using BOLD and the CEAMARC samples display eight distinct molecular clusters for a single supposed species (specimens from East Antarctica and the Antarctic Peninsula). Within the Bryozoa, some species (e.g. within those of the *Cellarinella* genus) exhibit huge morphological plasticity on the zooid level yet not on the general appearence of the colony (Hayward, 1995). Whether this is due to cryptic speciation or morphological variation it is yet to be further investigated using barcoding results, especially based on the CEAMARC samples.

4.3. Integrating the CEAMARC sequences into a larger picture

So far 11,323 specimens have been submitted to CCBD for CAML, with the relevant data

hosted in the BOLD under the CAML campaign. The 2762 specimens contributed to date from the CEAMARC voyage represent 24% of the CAML campaign, and provide 1% of MarBOL's overall global target of 250,000 COI sequences. Grant et al. (in press) showed that for Antarctic barcoding only a few specimens have yet been collected in deep water. This is particularly true for the CEAMARC, where all specimens were collected between 200m and 2060m depths, although it was the first survey below 200m in the area. When compared with numbers of sequences for phyla, classes, genera and species from other parts of the Southern Ocean (figure 2 and Grant et al., in press), the CEAMARC area appears to be one of the richest single spot (with the Ross Sea and some parts of the Antarctic Peninsula) for its sequence yield.

COI data can be used for other purposes than identification, as has been pointed out by numerous studies (Hebert et al., 2003; Hebert et al., 2004; Bucklin et al., 2007; Smith et al., 2008; Steinke et al., 2009). Methods for molecular taxonomy studies go well beyond the simple distance methods and trees to answer the best practice of current phylogenetic and population genetics, and ideally include morphology as well as multiple mitochondrial and nuclear markers (Padial et al., 2010). However, a preliminary barcode study with multiple well-identified specimens per species appears a valuable first step combining clusters of genetic similarity supported by morphology, to investigate putative species monophyly, and highlight interesting and unexpected patterns that need to be explored further (Duhamel et al., 2010; Dettai et al., in press; Gonzaléz-Wevar et al., 2010; Gonzaléz-Wevar et al., in press). The history of a single gene is not always the history of species (Doyle, 1992; Maddison, 1997), and like any other marker the results from COI need to be compared to those from additional DNA markers, using phylogenetic inference methods (maximum parsimony, maximum likelihood or bayesian inference). In Notothenioidei (Teleostei), a Bayesian analysis of the partial COI gene of a sample including a large proportion of the CEAMARC specimens produced topologies that are more congruent with nuclear markers than the other mitochondrial markers (Near and Cheng, 2008), suggesting that the COI might be a good phylogenetic marker for this group.

The sampling of liparids collected by the CEAMARC survey was exceptionally rich both in species and in specimen number, as these are mostly small and solitary teleosts. Preliminary studies based on COI gave an unexpected pattern for the family, with complex sister-group relationships between species from both hemispheres (Dettai et al., in press; Duhamel et al., 2010). The analysis of the CEAMARC-collected schizasterid echinoderms inserted in a larger Antarctic and Sub-Antarctic sampling using multiple reconstruction methods and multiple markers (table 1) revealed the non-monophyly of the genera *Abatus*, *Amphipneustes* and *Tripylus* (Chenuil et al., 2009). This initiated a re-evaluation of the diagnostic morphological characters used for the family, a necessity also supported by some of the previous morphological work (Madon, 1998; David et al., 2005).

The CEAMARC samples, as part of a multidisciplinary, broad sampling effort, have contributed effectively to a better understanding of evolutionary patterns and processes of Antarctic and Sub-Antarctic species. In all the following examples, the addition of the CEAMARC samples were critical in increasing the scope of the study by providing several species that were not available by other sources. Moreover the location of the survey has added crucial information from a distant collecting sector to compare to the well studied Antarctic Peninsula. The collection of specimens from all around the Antarctic Ocean is a necessary step in attempting to understand species distributions by inferring gene flow between populations of supposedly circum-antarctic species, even if it cannot provide information about a possible interruption of gene exchanges at some other plausible location, like between the West and East part of the Antarctic Peninsula (Clarke et al., 2009). For instance, the pycnogonid *Nymphon australe* Hodgson, 1902 appears to be a genuinely circumpolar species, but, at the same time showing significant genetic differentiation between East and West Antarctic populations and higher or lower genetic diversity depending on the location (Mahon et al., 2008; Arango et al., 2010). Similarly, in the species of fish investigated with additional specimens and markers, specimens from very distant locations did not form geographical clades in the trees (Duhamel et al., 2010; Lautredou et al., 2010; Lecointre et al., in press; Smith et al., 2008), indicating single-species populations represented in the sampling. This is also the case

for the shallow water sterechinid *Sterechinus neumayeri* (Meissner, 1900), for which there is no phylogeographic structure between the Antarctic Peninsula and the Eastern part of the Southern Ocean (Diaz et al., in press).

Partial COI was sequenced from 350 specimens of the endemic Southern Ocean octopus genus *Pareledone* from multiple localities around the Southern Ocean (Allcock et al., 2010). A statistical parsimony haplotype network revealed no overlap of haplotypes between species, and clusters of haplotypes per species whatever the location. The CEAMARC samples of *Pareledone albimaculata* Allcock, 2005, *P. panchroma* Allcock, 2005 , *P. aequipapillae* Allcock, 2005, and *P. subtilis* Allcock, 2005 had distinct haplotypes, although these are not very different from the haplotypes of specimens from the South Shetland Islands or Ross Sea (1-5 differences). The CEAMARC haplotypes were also found at other locations (Weddell Sea, Prydz Bay, or Ross Sea) for *P. aurata* Allcock, 2005, *P. prydzensis* (Lu and Stranks, 1994) , and *P. cornuta* Allcock, 2005. Both these trends were representative of the other locations. These findings indicate restricted (more or less depending on the species) gene flow between locations (Allcock et al., 2010). *Pareledone aequipapillae* appeared to be circumpolar, with significant congruence according to a Mantel test between mean genetic distance between locations and the shortest distance between them when avoiding the warm water (Clarke et al., 2009) to the west of the Antarctic Peninsula. Adults are absent in this region (Collins et al., 2004) and *Pareledone* species do not have a planktonic dispersal phase (Barratt et al., 2008), so environments inhospitable to adults may prove a significant barrier to gene flow. Allcock et al. (2010) suggested that *P. aequipapillae* may be exhibiting an evolutionary pattern similar to that seen in ring species (e.g, Irwin et al., 2005), and that specimens of *P. aequipapillae* from the Amundsen Sea and South Shetland Islands might be reproductively isolated. These findings need to be compared to those for other groups to identify the barriers to gene flow in species with similar life and reproduction styles, and test the hypotheses (for instance Eastman and McCune, 2000; Allcock et al., 2001; Thatje, 2005; Pearse et al., 2009) of shared speciation promoters in the Southern Ocean (Diaz et al., in press; González-Wevar et al., 2010).

5. Conclusion

The Barcode of Life project directly (Hajibabaei et al., 2005) and indirectly encourages large scale molecular studies with a higher focus on quality. The structural need for a voucher specimen provides an opportunity for morphological and molecular studies using the same specimens, and for subsequent controls of the identification. Both effects were observed on the results of the CEAMARC campaign. The multiple studies performed in parallel on different organisms collected during the CEAMARC voyage allowed the detection of differences and similarities in the evolutionary patterns of very different taxonomic groups. By comparing these patterns, shared drivers of speciation in the area will be isolated (Diaz et al., in press). The first studies by non-taxonomists using data checked by barcoding techniques are now being published (Koubbi et al., 2010; Causse et al., submitted; Strugnell et al., in press), and will multiply as more results from the survey are made available, and the inclusion of the CEAMARC samples in phylogeny and phylogeography studies has also brought its first interesting results. Until recently, there were very few sequences from Antarctic taxa in the Barcode of Life Database (Grant and Linse, 2009), the number of sequences has grown considerably, with a major contribution from the CEAMARC survey.

Centralised specimen sampling and sorting followed by the work of multiple research teams with the relevant expertise, has taken advantage of the huge numbers and diversity of specimens and data collected during the survey in a relatively short amount of time. Additional sampling in the area promises new discoveries, if the number of new species collected during the CEAMARC is taken as an indication. Although deeper sampling (below 1000 m) represents only a small proportion of the sampling from the CEAMARC and the Southern Ocean overall (this study, Grant et al., in press), these first results show that rare and new species are preferentially found in the deeper sites, although there are differences between the deeper stations on the plateau and the ones

on the slope (Causse et al., submitted). The three new species of Ascidiacea all come from deeper stations (Monniot et al., in press), and depth-linked genetic structure might also be present in the area in other widely distributed taxa like pycnogonids (Arango et al., 2010).

Acknowledgements

We thank the crew and participants of the cruises involved in the capture of the samples, and especially the chief scientists of the CAML-CEAMARC cruises. The CAML-CEAMARC cruises of RSV *Aurora Australis* and TRV *Umitaka Maru* (IPY project n°53) were supported by the Australian Antarctic Division, the Japanese Science Foundation, the French polar institute IPEV (ICOTA and REVOLTA programs), the CNRS, the MNHN and the ANR (White Project ANTFLOCKs USAR n°07-BLAN-0213-01 directed by Guillaume Lecointre). This project is a contribution to the EBA-SCAR program.

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. P. Kuklinski is also supported by the Polish Ministry of Science and Higher Education (539/N-CAML/2009/0). Funding for barcoding at the CCDB was provided by Genome Canada through the Ontario Genomics Institute. DS was supported by the Alfred P. Sloan Foundation. Work on Pycnogonida is supported by the Australian Antarctic Science Grants (project AA3010), and CAML coordination grant #34 to C.P. Arango. New Zealand authors were supported by research funded by the New Zealand Government under the NZ International Polar Year-Census of Antarctic Marine Life Project. Part of the work in the MNHN was supported by the « Action Transversale du MNHN : Taxonomie moléculaire : DNA Barcode et gestion durable des collections » (MNHN-Barcode initiative; co-PI Jean-Noël Labat and Sarah Samadi); part of these data fed the MarBol project supported by the Sloan Foundation. This work was also supported by the ATM of MNHN "Biodiversité actuelle et fossile. Crises, stress, restaurations et panchronisme : le message systématique".

We thank the editors and guest editors of Polar Science, as well as two anonymous reviewers for

their constructive criticisms.

References

Allcock, A.L., Brierley, A.S., Thorpe, J.P., Rodhouse; P.G., 1997 Restricted gene flow and evolutionary divergence between geographically separated populations of the Antarctic octopus *Paraledone turqueti*. Marine Biology 129, 97–102.

Allcock, A.L., Barratt, I., Eléaume, M., Linse, K., Norman, M.D., Smith, P.J., Steinke, D., Stevens, D.W., Strugnell, J.M., 2010. Cryptic speciation and the circumpolarity debate: a case study on endemic Southern Ocean octopuses using the COI barcode of life. Deep-sea Research II. doi:10.1016/j.dsr2.2010.05.016

Allcock, A.L., Piatkowski, U., Rodhouse, P.G.K., Thorpe, J.P., 2001. A study on octopodids from the eastern Weddell Sea, Antarctica. Polar Biology 24, 832–838.

Arango, C.P., Soler Membrives, A., Miller, K.J., 2010. Genetic differentiation in the Circum-Antarctic sea spider *Nymphon australe* (Pycnogonida; Nymphonidae). Deep Sea Research II. doi:10.1016/j.dsr2.2010.05.019.

Arnaud, P.M., 1974. Contribution à la bionomie marine benthique des régions antarctiques et subantarctiques. Thétys 6(3), 465-656.

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.

Barratt, I.M., Johnson, M.P., Collins, M.A., Allcock, A.L., 2008. Female reproductive biology of two sympatric incirrate octopod species, *Adelieledone polymorpha* (Robson 1930) and *Pareledone turqueti* (Joubin 1905) (Cephalopoda: Octopodidae), from South Georgia. Polar Biology 31, 583-594.

Beaman, R.J., Harris, P.T., 2005. Bioregionalisation of the George V Shelf, East Antarctica. Continental Shelf Research, 25, 1657-1691.

Beaman, R.J., O'Brien, P., 2009. Collaborative East Antarctic Marine Census (CEAMARC): Post-Survey Report, RSV *Aurora Australis* Voyage 3, December 2007 – January 2008. Geoscience Australia Record 2008/05, 61pp.

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

Bickford, D., Lohman, D. J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation, TRENDS in Ecology and Evolution 22(3), 148-155.

Bortolus, A., 2008. Error cascades in the biological Sciences: the unwanted consequences of using bad taxonomy in ecology. Ambio 37 (2), 114-118.

Bucklin, A., Wiebe, P.H., Smolenack, S.B., Copley, N.J., Beaudet, J.G., Bonner, K.G., Farber Lorda, J., Pierson, J.J., 2007. DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). Journal of Plankton Research 29, 483–493.

Bucklin, A., Hopcroft, R.R., Kosobokova, K.S., Nigro, L.M., Ortman, B.D., Jennings, R.M., Sweetman, C.J., 2010. DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. Deep-Sea Research II 57 , 40–48.

Bucklin, A., Steinke, D., Blanco-Bercial, L., 2011. DNA Barcoding of Marine Metazoa. Annual Review of Marine Science 3: 18.11-18.38.

Buhay, J.E., 2009. "COI-like" Sequences are Becoming Problematic in Molecular Systematic and DNA Barcoding Studies. Journal of Crustacean Biology 29(1), 96-110.

Causse, R., Ozouf-Costaz, C., Koubbi, P., Lamy, D., Eléaume, M., Dettai, A., Duhamel, G., Busson, F., Pruvost, P., Post, A., Beaman, R. J., Riddle, M.J. Demersal ichtyofauna from the Dumont d'Urville Sea (East Antarctica) during the CEAMARC surveys in 2007-2008. Submitted to the

CEAMARC special issue of Polar Science.
Chenuil, A., Egea, E., Rocher, C., Féral J.-P., 2009. Comparing substitution rates in spatangoid sea urchins with putatively different effective sizes, and other echinoderm datasets. Proceedings of the 12th International Echinoderm conference. Balkema. P159-161.
Chenuil, A., Hoareau, T.B., Egea, E., Penant, G., Rocher, C., Aurelle, D., Mokhtar-Jamai, K., Bishop, J.D.D., Boissin, E., Diaz, A., Krakau, M., Luttikhuizen, P.C., Patti, F.P., Blavet, N., Mousset, S., 2010. An efficient method to find potentially universal population genetic markers, applied to metazoans. BMC evolutionary biology 10, 276.
Clarke, A., Barnes, D.K.A., Hodgson, D.A., 2005. How isolated is Antarctica? Trends in Ecology and Evolution 20, 1–3.
Clarke, A., Griffiths, H.J., Barnes, D.K.A., Meredith, M.P., Grant, S.M., 2009. Spatial variation in seabed temperatures in the Southern Ocean: Implications for benthic ecology and biogeography. Journal of Geophysical Research 114, G03000.
Collins, M.A., Allcock, A.L., Belchier, M., 2004. Cephalopods of the South Georgia slope. Journal of the Marine Biological Association of the United Kingdom 84, 415-419.
David, B., Choné, T., Mooi, R., De Ridder, C., 2005. Antarctic echinoidea. Synopses of the Antarctic benthos, vol.10. Koeltz Scientific Books, Königstein. 274 pp.
De Broyer, C., Danis, B. (Eds.), 2009. SCAR-MarBIN: The Antarctic Marine Biodiversity Information Network World Wide Web electronic publication. Available online at. http://www.scarmarbin.be/.
Dettai, A., Lautredou, A.-C., Bonillo, C. , Goimbault, E., Busson, F., Causse, R.,Couloux, A., Cruaud, C., Duhamel, G., Denys, G., Hautecoeur, M., Iglesias, S., Koubbi, P., Lecointre, G., Moteki, M., Pruvost, P., Tercerie, S., Ozouf, C., in press. The actinopterygian diversity of the CEAMARC cruises: Barcoding and molecular taxonomy as a multi level tool for new findings. Deep Sea Research II. doi:10.1016/j.dsr2.2010.05.021
Diaz, A., Féral, J.-P., David B., Saucède T., Poulin E., in press. Evolutionary pathways among shallow and deep sea echinoids of the genus *Sterechinus* in the Southern Ocean. Deep Sea Research II. doi:10.1016/j.dsr2.2010.10.012
Doyle, J. J., 1992. Gene trees and species trees: Molecular systematics as one-character taxomony. Systematic Botany 17, 144-163.
Duhamel, G., Hautecoeur, M., Dettai, A., Causse, R., Pruvost, P., Busson, F., Couloux, A., P. Koubbi, Williams, R, Ozouf-Costaz, C., Nowara, G., 2010. Liparids from the Eastern sector of Southern Ocean and first information from molecular studies. Cybium. 34, 319-343.
Eastman, J.T., McCune, A.R., 2000. Fishes on the Antarctic continental shelf: evolution of a marine species flock? Journal of Fish Biology, 57(supplA), 84-102.
Ekrem, T., Willassen, E., Stur, E., 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Molecular Phylogenetics and Evolution 43, 530–542.
Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology. 3: 294–299
Galtier, N., Nabholtz, B., Glémin, S., Hurst, G.D.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology 18, 4541–4550.
González Wevar C.A., Nakano T., Cañete J.I., Poulin E., 2010. Molecular phylogeny and historical biogeography of *Nacella* (Patellogastropoda: Nacellidae) in the Southern Ocean. *Molecular Phylogenetics and Evolution* 56: 115-124.
González-Wevar C.A., B. David, Poulin E., in press. Phylogeography and demographic inference in *Nacella* (Patinigera) *concinna* (Strebel, 1908) in the western Antarctic Peninsula. Deep Sea Research II. Doi:10.1016/j.dsr2.2010.05.026.
Grant, R.A., Linse, K., 2009. Barcoding Antarctic Biodiversity; current status and the CAML initiative, a case study of marine invertebrates. Polar Biology 32(11), 1629-1637.
Grant, R.A., Griffiths, H., Steinke, D., Wadley, V., Linse, K., in press. Antarctic DNA barcoding; a drop in the ocean? Polar Biology. DOI 10.1007/s00300-010-0932-7

Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Mackie, P.M., Hebert, P.D., 2005. Critical factors for assembling a high volume of DNA barcodes. Philos Trans R Soc Lond B Biol Sci. 360(1462), 1959-1967.

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

Hayward, P.J., 1995. The Antarctic cheilostomatous Bryozoa. Oxford University Press.

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.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of Birds through DNA Barcodes. PLoS Biology 2, e312.

Hoareau, T.B., Boissin, E., 2010. Design of phylum-specific hybrid primers for DNA barcoding: addressing the need for efficient COI amplification in the Echinodermata. Molecular Ecology ressources 10(6), 960-967.

Hunt, B., Strugnell, J., Bednarsek, N., Linse, K.R., Nelson, J., Pakhomov, E., Seibel, B., Steinke, D., Würzberg, L., 2010. Poles Apart: The ''Bipolar'' Pteropod Species *Limacina helicina* Is Genetically Distinct Between the Arctic and Antarctic Oceans. PloS ONE, 5(3), e9835.

Hunter, R.L., Halanych, K.M., 2008. Evaluating connectivity in the brooding brittle star *Astrotoma agassizii* across the Drake Passage in the Southern Ocean. Journal of Heredity 99, 137-148.

Irwin, D.E., Bensch, S., Irwin, J.H., Price, T.D., 2005. Speciation by distance in a ring species. Science 307, 414-416.

Ivanova, N.V., Dewaard, J.R., Hebert, P.D.N., 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes M6, 998-1002.

Koubbi P., Ozouf-Costaz C., Hureau J.-C., Lecointre G., Pisano E, Vacchi M., Razouls S. White M, Le Guilloux E., 2001. Le sous-ordre des Notothenioidei (Ostéichthyens, Perciformes), modèle unique de radiation adaptative en milieu isolé. /Rapport d'activité 2000/, Campagne d'été 2000-2001. Institut Français pour la Recherche et la Technologie Polaires, p. 64-70, 6 fig.

Koubbi, P., Ozouf-Costaz, C., Hosie, G., 2009. L'ichtyofaune de la mer de Dumont D'Urville: apports du Census of Antarctic Marine Life au programme d'Ichtyologie Côtière en Terre Adélie (ICOTA).IPEV, Rapport d'activité 2008. Campagnes d'été 2008-2009 : 44-45.

Koubbi, P., Ozouf-Costaz, C., Moteki, M., Goarant, A., Hulley, P.A., Dettai, A., Duhamel, G., Pruvost, P., Causse, R., Tavernier, E., Post, A., Beaman, R.J., Rintoul, S., Hirawake, T., Hirano, D., Ishimaru, T., Riddle, M., Hosie, G., 2010. Estimating the biodiversity of the East Antarctic shelf and oceanic zone for ecoregionalisation: example of the ichthyofauna of the CEAMARC (Collaborative East Antarctic Marine Census) CAML surveys. Polar Science 4(2), 115-133.

Krabbe, K., Leese, F., Mayer, C., Tollrian, R., Held, C., 2010. Cryptic mitochondrial lineages in the widespread pycnogonid *Colossendeis megalonyx* Hoek, 1881 from Antarctic and Subantarctic waters. Polar Biology 33(3), 281-292.

Lautredou, A.-C, Bonillo, C., Denys, G., Cruaud, C., Lecointre, G., Dettai, A., 2010. Molecular taxonomy and identification of genus *Trematomus* (Notothenioidei, Teleostei): How good is barcoding with COI ? Polar Science 4(2), 333-352.

Lecointre, G., Gallut, C., Bonillo, C., Couloux, A. , Ozouf-Costaz, C., Dettai, A., in press. The Antarctic Fish Genus *Artedidraco* is paraphyletic (Teleostei, Notothenioidei, Artedidraconidae). Polar Biology.

Lee, Y., Song, M., Lee, S., Leon, R., Godoy, S., Cañete, I., 2004. Molecular phylogeny and divergence time of the Antarctic sea urchin (*Sterechinus neumayeri*) in relation to the South American sea urchins. Antarctic Science 1, 29 - 36.

Leese, F., Held C., 2008. Identification and characterization of microsatellites from the Antarctic isopod *Ceratoserolis trilobitoides* - nuclear evidence for cryptic species. Conservation Genetics 9, 1369–1372.

Linse, K., Cope, T., Lorz, A.N., Sands, C., 2007. Is the Scotia Sea a centre of Antarctic marine diversification? Some evidence of cryptic speciation in the circum-Antarctic bivalve *Lissarca notorcadensis* (Arcoidea: Philobryidae). Polar Biology 30, 1059–1068.

Lockhart, S.J., 2006. Molecular evolution, phylogenetics and parasitism and in antarctic cidaroid echinoids. PhD thesis. University of California (Santa Cruz).
Maddison, M. P., 1997. Gene trees in species trees. Systematic Biology, 46, 523-536.
Madon, C., 1998. Disparité morphologique et architecturale des Schizasteridae incubants (Echinoidea, Spatangoida) des régions australes = Morphological and architectural disparity of brooding Schizasterids (Echinoidea, Spatangoida) from south polar regions. PhD, Université de Dijon.
Mahon, A.R., Arango, C.P., Halanych, K.M., 2008. Genetic diversity of *Nymphon* (Arthropoda:Pycnogonida: Nymphonidae) along the Antarctic Peninsula with a focus on *Nymphon australe* Hodgson 1902. Marine Biology 155, 315-323.
Monniot, F., Dettai, A., Eleaume, M., Cruaud, C., Ameziane, N., in press. Antarctic Ascidians (Tunicata) of the French-Australian survey CEAMARC in Terre Adélie. Zootaxa.
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.
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(2), 832-40.
Nilsson, R.H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.-H., Koljalg, U., 2006. Taxonomic Reliability of DNA Sequences in Public Sequence Databases: A Fungal Perspective. PloS ONE 1, e59.
Padial, J.M., Miralles, A., De la Riva, I., Vences, M., 2010. The integrative future of taxonomy. Frontiers in Zoology. 25(7), 16.
Palero, F., Hall, S., Clark, P. F., Johnston, D., Mackenzie-Dodds, J., Thatje, S., 2010. DNA extraction from formalin-fixed tissue: new light from the deep sea. Scientia Marina 74(3), 465-470.
Pearse, J.S., Mooi, R., Lockhart, S.J., Brandt, A., 2009. Brooding and Species Diversity in the Southern Ocean: Selection for Brooders Or Speciation within Brooding Clades? In: A. Krupnik, I., Lang, M.A., Miller, S.E., eds. Smithsonian at the Poles: Contributions to International Polar Year Science, 181-196.
Post, A.L., Beaman, R.J., O'Brien, P.E., Eléaume, M., Riddle, M.J., 2010. Community structure and benthic habitats across the George V Shelf, East Antarctica: trends through space and time. Deep Sea Res. II, doi: 10.1016/j.dsr2.2010.05.020.
Puillandre, N., Strong, E.E., Bouchet, P., Boisselier, M.-C., Couloux, A., Samadi, S., 2009. Identifying gastropod spawn from DNA barcodes: possible but not yet practicable. Molecular ecology Ressources 9(5), 1311-1321.
Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system. Mol. Ecol. Notes 7, 355-364.
Smith, P.J, Steinke, D., Mcveagh, S.M., Stewart, A.L., Struthers, C.D., Roberts, C.D., 2008. Molecular analysis of Southern Ocean skates (*Bathyraja*) reveals a new species of Antarctic skate. Journal of Fish Biology 73, 1170-1182.
Stefaniak, L.M., Lambert, G., Gittenberger, A., Zhang, H., Lin, S., Whitlatch, R.B., 2009. Genetic conspecificity of the worldwide populations of *Didemnum vexillum* Kott, 2002. Aquat. Invasions 4 (1), 29-44.
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.
Strugnell, J.M., Cherel, I., Cooke, I.R., Gleadall, I.G., Hochberg, F.G., Ibáñez, C.M., Jorgensen, E., Laptikhovsky, V.V., Linse, K., Norman, M., Vecchione, M., Voight, J.R., Allcock, A.L., in press.The Southern Ocean: Source and sink? Deep Sea Research II. doi:10.1016/j.dsr2.2010.05.015
Teletchea, F., 2009. Molecular identification methods of fish species: reassessment and possible applications. Reviews in Fish Biology and Fisheries 19(3), 265-293
Thatje, S., 2005. The future fate of the Antarctic marine biota? Trends in Ecology and Evolution 20,

418–419.
Undheim, E.A.B., Norman, J.A., Thoen, H.H., Fry, B.G., 2010. Genetic identification of Southern Ocean octopod samples using mtCOI. Comptes Rendus Biologies 333, 395–404.
Valdez-Moreno, M., Vasquez-Yeomans, L., Elias-Gutierrez, M., Ivanova, N.V., Hebert, P.D.N., 2010. Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. Marine and Freshwater Research 61(6), 665-671.
Valentini, A., Pompanon, F., Taberlet, P., 2009. DNA barcoding for ecologists. Trends in Ecology and Evolution 24(2),110-117.
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., Bronwyn, H.H., O'Hara, T.D., 2008. DNA barcoding discriminates echinoderm species. Molecular Ecology Resources 8, 1202-1211.
Ward, R.D., Hanner, R., Hebert, P.D., 2009. The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74, 329-356.
Webb, K.E., Barnes, D.K.A., Clark, M.S., Bowden, D.A., 2006. DNA barcoding: a molecular tool to identify Antarctic marine larvae. Deep Sea Research II 53, 1053-1060.
Wilson, N.G., Belcher, R.L., Lockhart, S.J., Halanych, K.M., 2007. Multiple lineages and an absence of panmixia in the 'circumpolar' crinoid *Promachocrinus kerguelensis* in the Atlantic sector of Antarctica. Marine Biology 152(4), 895-904
Zhang, J., 2010. Exploiting formalin-preserved fish specimens for resources of DNA barcoding. Molecular Ecology Resources 10, 935-941.

Tables and Figures

Table 1

Number of specimens per taxon and barcode sequences already generated (including sequences not yet integrated into BOLD) for the CEAMARC survey as of september 2010. Classification follows BOLD. Several projects are still ongoing. Numbers not yet available are indicated with ?. *Asteroidea specimens were identified from photographs. > indicates values that are not fixed yet and are expected to grow in the future.

Figure 1. Location of the CEAMARC samples entered in BOLD. The pattern is the same as the CEAMARC sampling pattern.

Figure 2. Geographic location and taxonomic coverage of already barcoded samples: A Phylum; B Class; C Genus; D Species. Red shows a high number of taxonomic groups per area and blue denotes a low number, white shows absence (in this case unsampled areas).

Figure 3. Haplotype network obtained from a portion of the COI gene for the five species of Cnidaridae collected during the CEAMARC campaign. Both COI and the nuclear internal transcribed spacer (ITS) give the same pattern.

Table1

Sheet1

	Taxon	Lots	Specimens, including non barcodable	Nb of identified species (nb of species with sequences)	Nb of non-identified species (nb of sequence clusters)	Number of new species being described	Number of obtained COI barcodes	Clusters new to BOLD	Additional marker	Status
Projects in progress										
Chordata										
	Teleost fishes	998	1172	72(72)		1	550	22	rhodopsin, Cytb, D-loop	almost completed
	Ascidiacea	271	490	31(26)		3	37	26	18S rDNA	In progress
Echinodermata										
	Asteroidea	731	821	>23 gen & 16 sp*	13 (40)	?	122	13	18S rDNA	In progress
	Crinoidea	2087	2087	7(7)		0	776	6	16S, 28S, 18S rDNA	In progress
	Echinoidea	222	470	20(15)		0	89	?	ITS, 16S, 28S rDNA	almost completed
	Ophiuroidea	130	2996			?				Starting
	Holothuroidea	544	2065			?				Starting
Annelida		112	>1500		13 families(0)	?		?	16S & 12S rDNA	Starting
Arthropoda										
	Pycnogonida	650	>1700	>30(25)	8(7)	2	~236	5-10	16S rDNA	In progress
	Malacostraca	402	2603	?	?	>	400	?		In progress
Mollusca	Non-Cephalopoda		1509		(61)	?	237	45		In progress
	Cephalopoda	251	263	8(8)		0	36	0		Over
Bryozoa		134	>71	21		4	84	?		In progress

Not started yet	**Lots**	**Specimens**
Platyhelminthes	5	6
Priapula	5	25
Sipuncula	6	9
Brachiopoda	63	394
Hemichordata	4	4

	Lots	**Specimens**
Cnidaria	343	682
Ctenophora	2	3
Porifera	273	297
Phaeophyta	5	5
Rhodophyta	13	13

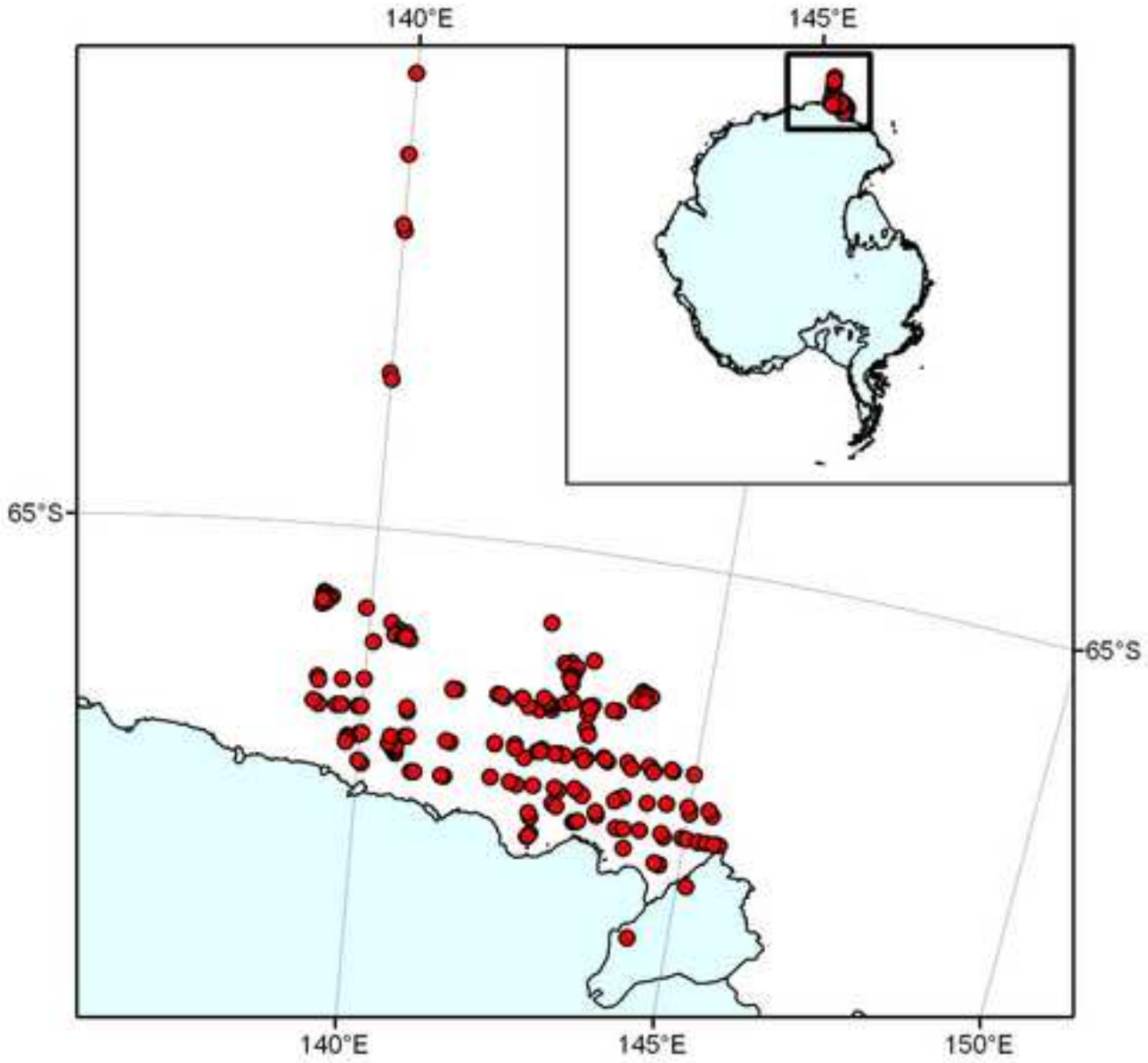

Figure2

Click here to download high resolution image

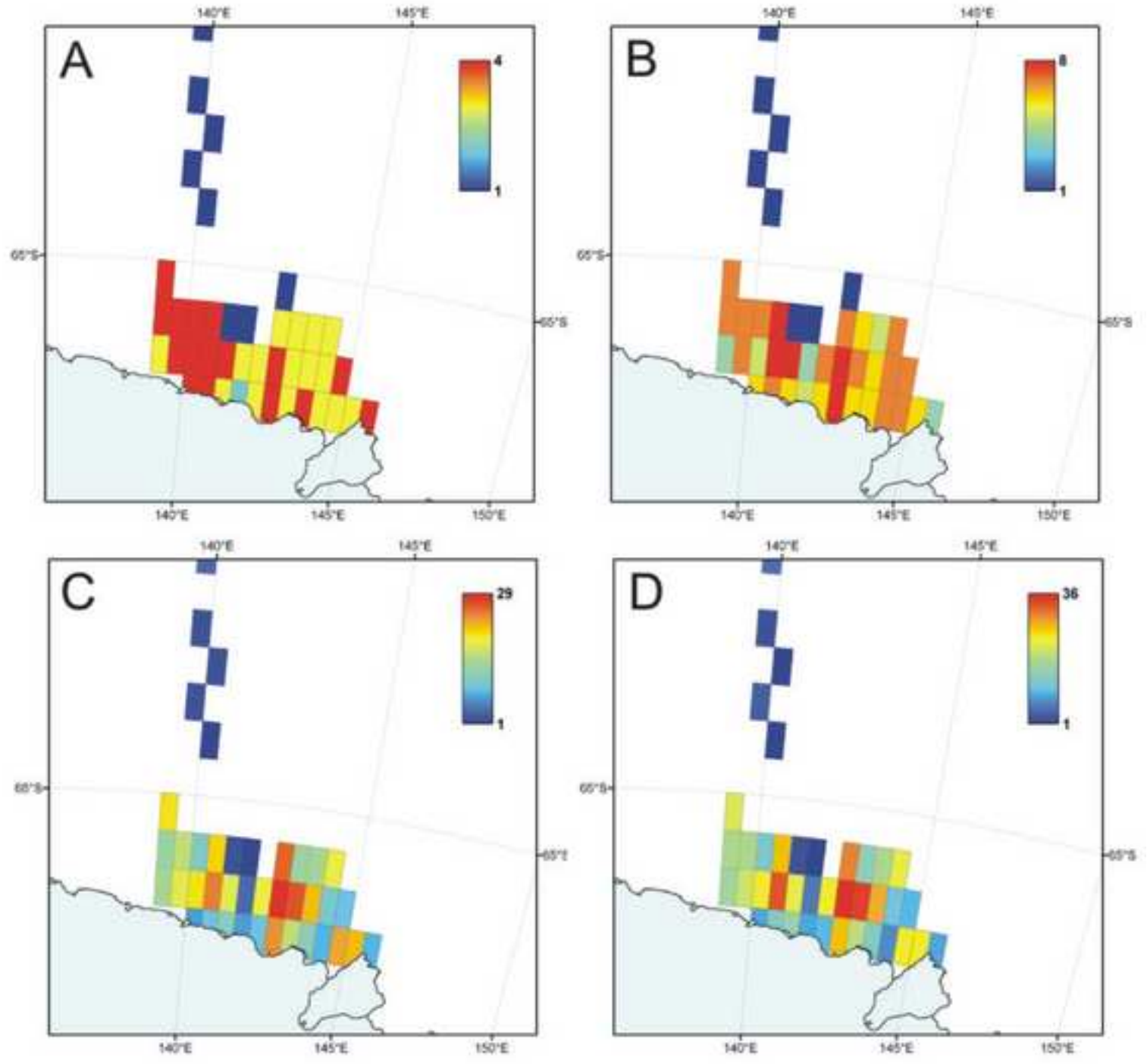

Figure3
Click here to download high resolution image