Polar Biol
DOI 10.1007/s00300-010-0932-7

SHORT NOTE

Antarctic DNA barcoding; a drop in the ocean?

Rachel A. Grant · Huw J. Griffiths · Dirk Steinke · Victoria Wadley · Katrin Linse

Received: 10 May 2010 / Revised: 9 November 2010 / Accepted: 10 November 2010

Abstract Coordinated, circum-Antarctic sampling expeditions during International Polar Year 2008/09 have given access to comprehensive collections suitable for DNA barcoding. Collaborations between the Census of Antarctic Marine Life (CAML), the Marine Barcode of Life project and the Canadian Centre for DNA Barcoding have enabled the Antarctic scientific community to initiate large-scale DNA barcoding projects to record the genetic diversity of Antarctic marine fauna, coordinated by the CAML Barcoding Campaign. A total of 20,355 marine specimens from more than 2,000 morphospecies covering 18 phyla are in the processing pipeline, and to date, 11,530 sequences have been processed with the remainder due by the end of 2010. Here, we present results on the current geographic and taxonomic coverage of DNA barcode data in the Southern Ocean and identify the remaining gaps. We show how DNA barcoding in the Antarctic is answering important questions regarding marine genetic diversity and challenging current assumptions of species distribution at the poles.

Keywords Circumpolarity · Marine · Cryptic species · Biodiversity · Southern Ocean

R. A. Grant
Scott Polar Research Institute, University of Cambridge, Lensfield Road Cambridge, Cambridge CB2 1ER, UK

R. A. Grant (✉) · H. J. Griffiths · K. Linse
British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK
e-mail: rachelannegrant@gmail.com

D. Steinke
University of Guelph, Biodiversity Institute of Ontario, Canadian Centre for DNA Barcoding, Guelph, ON N1G 2W1, Canada

V. Wadley
Department of the Environment and Heritage, Australian Antarctic Division, 203 Channel Highway, Kingston, TAS 7050, Australia

Introduction

Molecular barcoding is an important tool that can be used to assess unknown biodiversity, facilitate population differentiation, speciation and phytogeographic investigations, as well as helping elucidate taxonomy and identify species. For a more detailed overview of the importance of barcoding see Grant and Linse (2009) and Bucklin et al. (2011). In 2009, Grant and Linse reported on the severe lack of Antarctic barcoding studies, particularly for marine invertebrates, which make up the majority of Antarctic fauna (Clarke and Johnston 2003). Grant and Linse (2009) found that there were less than 100 cytochrome c oxidase (COI) gene sequences for marine invertebrates available in the public domain. Additionally, these barcodes showed strong geographic and taxonomic bias as the majority of sequences came from the Weddell Sea and Antarctic Peninsula and only two phyla, crustaceans and molluscs, were covered.

The Census of Antarctic Marine Life (CAML) aims to "investigate the distribution of Antarctica's vast marine biodiversity to develop a benchmark for the benefit of humankind". CAML is committed to DNA barcoding as a means of furthering this aim. In 2008, CAML initiated the Antarctic Barcoding Campaign, hosted within the Barcode of Life Data System (BOLD; www.boldsystems.org), a specialist database designed specifically to hold DNA barcode information (Ratnasingham and Hebert 2007) and

Published online: 03 December 2010

the Scientific Committee on Antarctic Research Marine Biodiversity Information Network (SCAR-MarBIN; www.scarmarbin.be), which holds distribution records and has also been adapted to hold genetic data with deeplinking to sequences stored in GenBank.

During the second International Polar Year 2007/2008, 18 international marine Antarctic expeditions were linked with CAML, resulting in circum-Antarctic sampling from the shallow shelf to the deep-sea. These expeditions provided wide-ranging collections of both benthic and pelagic species suitable for molecular barcoding studies. Here, we report on the present status of the Antarctic Barcoding Campaign and discuss the remaining challenges.

Based on CAML's definition of Antarctic molecular barcodes (sequences from a predetermined list of suitable genes from georeferenced marine specimens), the CAML Barcoding Campaign not only holds data for the COI gene used by the Marine Barcode of Life Project (MarBOL), but also information from further nuclear and mitochondrial genes. To date, these non-COI barcodes are accessible in SCAR-MarBIN.

In collaboration with MarBOL and the Canadian Centre for DNA Barcoding (CCDB), the CAML scientific community has been submitting Antarctic specimens to the CCDB for barcoding since January 2009. The processed sequences are hosted in BOLD. The expectation is that the remaining samples will be processed by the end of 2010.

The current status of Antarctic molecular barcoding

For purposes of display, georeferenced locations of BOLD and SCAR-MarBIN barcodes in the Southern Ocean were plotted on Antarctic maps (Fig. 1). Then, the depth distribution of the Antarctic barcodes was analysed by splitting the Ocean into 1000 m depth zones (Fig. 2). For the geographic analysis, the Southern Ocean was divided into boxes of 3° latitude × 3° longitude. 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 number counts (Figs. 3, 4). The taxonomic classification of morphospecies was provided by the scientists submitting the samples, following the Register of Antarctic Marine Species (RAMS), which is reviewed by an international team of more than 70 taxonomists.

To date, a total of 20,355 marine specimens from more 2,330 morphospecies covering 18 phyla are in the pipeline, comprising 11,530 sequences which have already been processed, and the remainder which are due by the end of 2010. The actual number of species is likely to be higher, as many that were sent for barcoding were not identified to species level at that point and high numbers of cryptic species are expected. The number of species barcoded may be as high as 5,000. Comparison with other Census barcoding projects is difficult as many results have not yet been published. The Arctic Ocean Diversity barcoding project has generated 4,345 barcodes for 630 Arctic marine species, representing approximately 13% of the estimated 5,000 metazoan Arctic species covering 11 phyla (Hardy et al. 2010).

Marine specimens from more than 2000 unique locations in the Southern Ocean have been sampled for molecular barcoding (Fig. 1). The collection sites cover a wide geographic range, with intensive collections in the Scotia, eastern Weddell, Ross and Dumont D'Urville seas. However, despite the intensive CAML sampling programme, there are still gaps around the Antarctic continent, especially in the East Antarctic from 6° to 42°E, from 78° to 114°E as well as in the Bellingshausen Sea from 78° to 102°W and in the Amundsen Sea from 132° to 171°E. The large majority of barcode samples were collected at depths shallower than 1,000 m, while barcodes from deep-water Antarctic specimens are still sparse (Fig. 2).

In total, 18 of 34 marine phyla are represented by at least one Antarctic barcode sequence. The number of phyla sequenced per 3 × 3 area box is generally low to medium, ranging from one to six (non-COI data) or seven (COI/BOLD data) (Figs 3, 4a). Out of a possible 108 classes of organisms in SCAR-MarBIN's Register of Antarctic Marine Species (RAMS) (correct on 11.3.10), 38 had at least one species in the class barcoded, but never more than 16 classes per box were sequenced (Fig. 4b). The remaining 70 classes have no sequence information. Important classes without Antarctic barcode sequence are Calcarea (calcareous sponges), Stauromedusae (stalked jellyfishes), Pterobranchia, Ciliophora, Dinoflagellata, Monoplacophora and Aplacophora (both Mollusca), Priapulida, Protozoa, Tardigrada and Zooflagellata. At genus level, most boxes show a low number of genera barcoded (Fig. 4c), although the numbers reach 98 genera per box in some areas. Similar patterns are seen at the species level where few segments currently have high number of species barcoded (Fig 4d). The sites with relatively high numbers of barcoded samples occur in the Ross and Dumont D'Urville seas as well as the Antarctic Peninsula, where specific projects have generated a high number of samples. Figure 5 shows the number of species sequenced in each Phylum.

Since 2007, the number of Antarctic barcode sequences either processed or in the pipeline has increased significantly from 432 (Grant and Linse 2009) to 20,355. Also, the geographic coverage has increased and now covers many parts of the continental shelf around the Antarctic continent. CAML has made a significant contribution to MarBOL's worldwide target for barcoding of 50,000 marine species. At present, MarBOL has sequenced 18,270

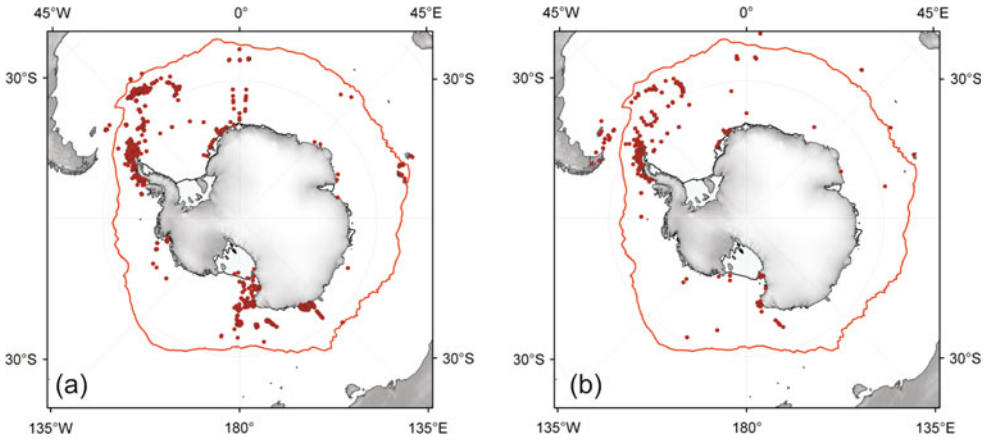

Fig. 1 Locations of barcode sequences in the Antarctic; **a** COI from *BOLD* **b** *non-BOLD* barcodes

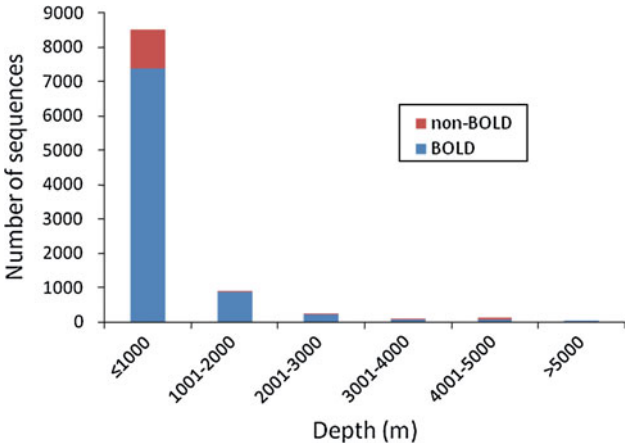

Fig. 2 Depth range of barcode specimen collection

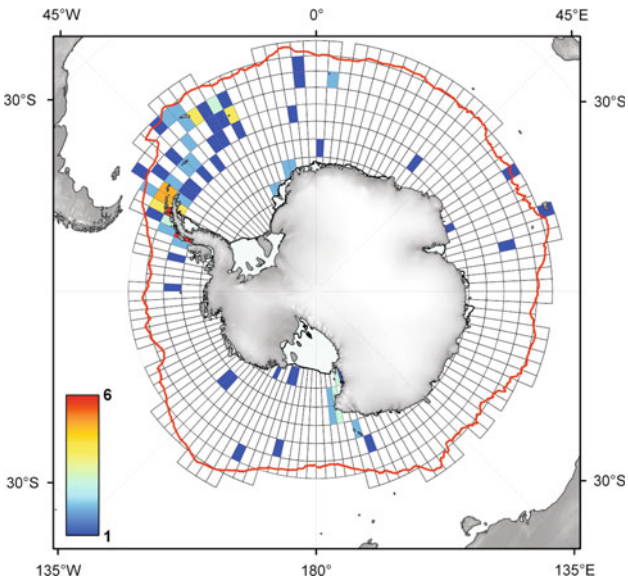

Fig. 3 Number of phyla per box with non-COI sequence

species out of 192,000 known species of marine metazoans (9%) (Bucklin et al. 2011); CAML has therefore supplied approximately 13%. The number of sequences generated and the extent of geographic coverage of the world's most remote and inhospitable continent is an outstanding achievement and a tribute to the Antarctic scientific community.

However, the major challenge still facing Antarctic barcoding is not the number of sequences that can be obtained, but ensuring that as many geographic areas, depths as well as taxonomic levels are covered as possible. Areas such as the Antarctic Peninsula, the Dumont D'Urville, Weddell and Ross seas still host the large majority of barcode sequences, but areas outside of these are effectively barcode deserts at the present time. Despite recent Antarctic deep-water efforts (Brandt et al. 2007), the present depth distribution for barcodes reflects the overall sampling effort in the Antarctic, with dominance of sampling on the shelf and upper slope, while most of the Southern Ocean is, in fact, deep-sea (Brandt et al. 2009, Griffiths 2010). The current paucity of deep-water samples means it is likely that new levels of diversity will be encountered with increased sampling, given that bathymetry is a significant mechanism leading to separation and genetic divergence (France and Kocher 1996; Chase et al. 1998; Zardus et al. 2006). Although a large number of sequences from many higher level taxa exist, there are large taxonomic gaps, particularly in the sponges (which are difficult to barcode; Erpenbeck et al. 2006), nematodes, mammals and birds. However, now that CAML Barcoding has effectively tested the usefulness of barcoding in Antarctic marine biology, it is highly likely that efforts will continue, both in collaboration with MarBOL and through private projects, eventually providing a comprehensive genetic inventory of the Southern Ocean.

Antarctica as a cryptic species hotspot

DNA barcoding in the Antarctic is providing important insights into Antarctic marine genetic diversity. For

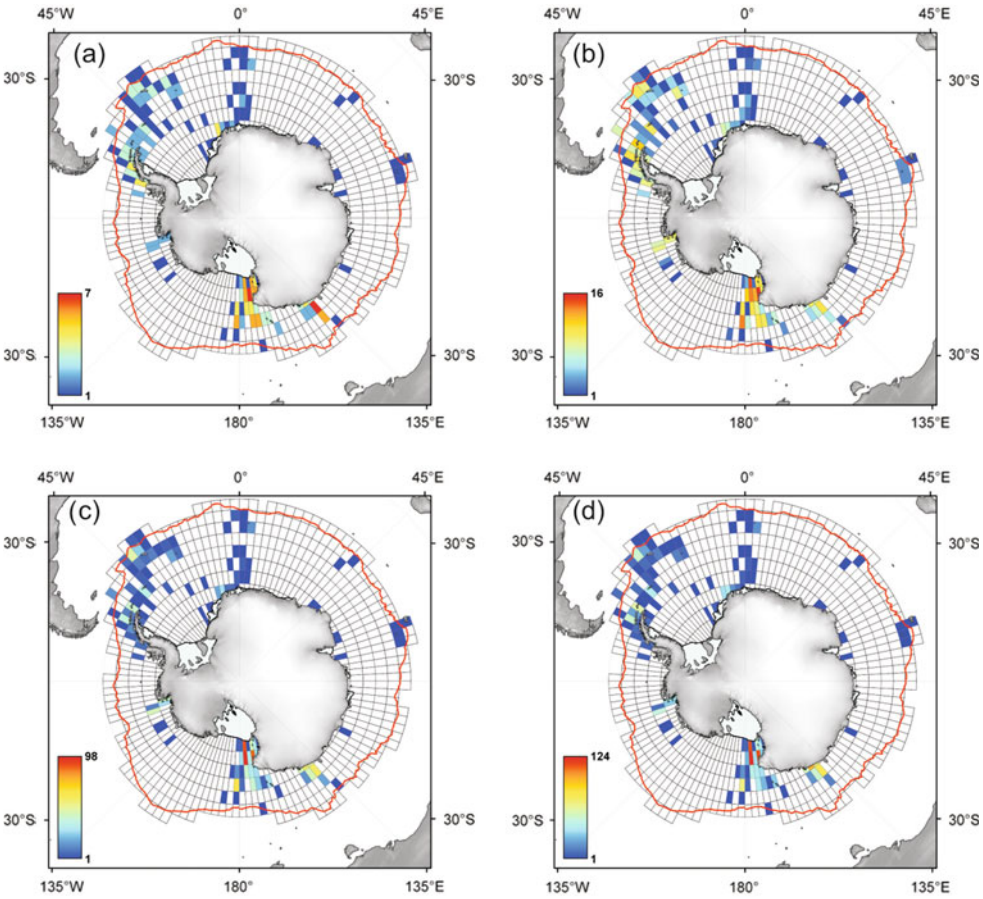

Fig. 4 Number of taxa per box with COI barcode in *BOLD*. **a** Phyla, **b** Class, **c** Genus and **d** Species

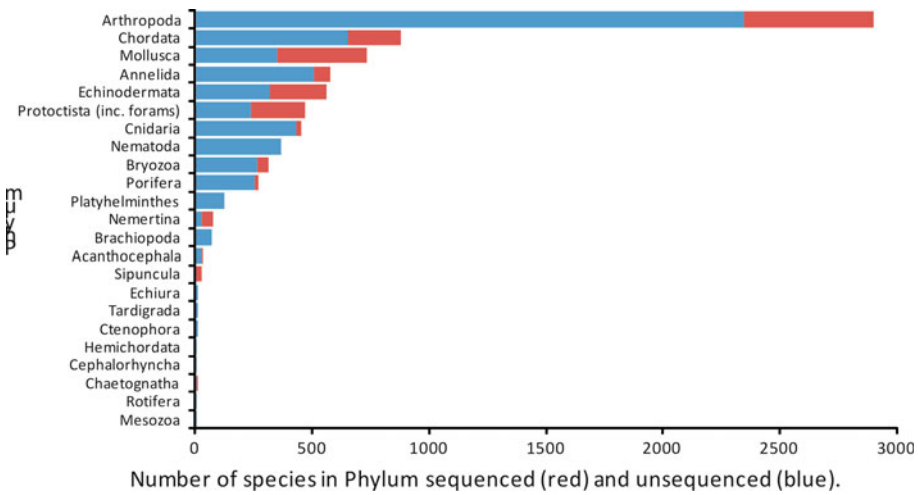

Fig. 5 Number of species sequenced (*red*) and unsequenced (*blue*) per Phylum

example, barcoding has facilitated the discovery of numerous new species as well as cryptic species complexes (species with a high level of genetic divergence but a very similar morphology), which seem to be a common feature of Antarctic marine fauna. In the past few years, there have been numerous reports of cryptic species in the Antarctic, for example in ophiuroids (Hunter and Halanych 2008), crinoids (Wilson et al. 2007) and pycnogonids (Krabbe et al. 2009). The number will surely keep on rising as more specimens are investigated by molecular methods.

Barcoding is also challenging current assumptions of species distribution at the poles. Several species that were previously thought to have circum-Antarctic distributions have been found, after genetic analysis, to be made up of

several previously unknown species (e.g. Amphipoda (Lörz et al. 2009), Ostracoda (Brandão et al. 2010)). Brandão et al. (2010) tested a model predicting the circum-Antarctic distribution of benthic species in the Southern Ocean using the ostracod genus *Macroscapha.* While previous reports named five circum-Antarctic species, their recent taxonomic revision based on a more detailed morphological species character definition recognised 20 morphospecies with narrow ranges of depth and geographic distributions. Molecular evidence from the COI gene supports both the bathymetrically and geographically narrow species classifications. On the basis of these results, Brandão et al. (2010) suggest a re-evaluation of current assumptions of circum-Antarctic benthic species' distributions in the Southern Ocean.

Hunt et al. (2010) found that the shelled pteropod *Limacina helicina,* previously thought to occur in both Arctic and Antarctic oceans actually has a 33% divergence in COI sequence between Arctic and Antarctic populations, indicating genetic divergence at the species level. Analysis of COI barcode sequences can also verify cases with genuinely wide-species distributions, e.g. circum-Antarctic or bipolar distributions. Arango et al. (2010) reported that COI sequences of the sea-spider *Nymphon australe* taken from specimens collected in the Weddell Sea, Antarctic Peninsula and East Antarctic produced single-haplotype networks. DNA barcoding is also helping to solve one of the mysteries of polar biology—are there genetically identical species at both poles? MarBOL and CAML are supporting a study to assess genetic evidence for bipolar species for which 481 barcodes are being processed from a total of 78 species, and the results are expected shortly.

The species description challenge

Given the high numbers of cryptic species complexes discovered, often related to geographic and bathymetric occurrence of specimens, it has become clear that biodiversity in the Antarctic has been severely underestimated. The challenge for Antarctic scientists is now to transfer the cryptic species information from the sequence datasets back to the taxonomic datasets and to link it to databases such as SCAR-MarBIN, and then to describe the newly discovered species. Molecular taxonomy, in combination with traditional taxonomic methods, is the way forward and offers the best chance of recording, and therefore protecting biodiversity. Studies such as Lörz et al. (2009) are excellent examples of the use of both genetic and morphological characteristics to elucidate taxonomy. Many similar studies currently in progress, across a wide range of Antarctic marine taxa, geographic areas and depths will prove useful even though the CAML Barcoding Campaign has now ended. So far, CAML Barcoding, although a significant achievement, may well be revealing just "a drop in the ocean" of total species diversity in the Southern Ocean.

The co-operation between MarBOL, CCDB and CAML has enabled the molecular analysis of thousands of specimens from the Southern Ocean. This has enabled Antarctic scientists to effectively test the paradigm of circum-Antarctic species distribution and the use of DNA barcodes for Antarctic species identification. DNA barcoding of Antarctic species will continue into 2011 and beyond via co-operation with MarBOL and other organisations, and the sequences stored both in BOLD, GenBank and with a possible link to the Southern Ocean Observing System (SOOS). In fact, the process of publishing barcode sequences from BOLD to GenBank is already underway, and eventually, all BOLD sequences will be publicly available in GenBank. CAML has gone a long way towards setting in motion a project, which may eventually provide a complete record of genetic diversity in the Antarctic.

Acknowledgments We are grateful to the team at the Canadian Centre for DNA Barcoding (supported by Genome Canada through the Ontario Genomics Institute) for their support to the CAML research community. We would also like to thank the following contributors: Louise Allcock, Nadia Ameziane, Claudia Arango, Martin Collins, Bruno Danis, Agnes Dettai, Marc Eleaume, Charlotte Havermans, Christoph Held, Joe Hoffman, Russ Hopcroft, Alexis Janosik, Tim O'Hara, Peter Smith, Stefano Schiaparelli, Ken Halanych, Guilliaume LeCointre, Florian Leese, Tim Littlewood, Sarah Mincks, Christoph Schubart, Myriam Schueller, Jan Strugnell, Odile Volontiero, Bob Ward. We also thank two anonymous referees. This paper is contribution to CAML number 29.

References

Arango CP, Soler Membrives A, Miller KJ (2010) Genetic differentiation in the Circum-Antarctic sea spider *Nymphon australe* (Pycnogonida; Nymphonidae). Deep-Sea Res Pt II doi: 10.1016/j.dsr2.2010.05.019 (in press)

Brandão SN, Sauer J, Schön I (2010) Circumantarctic distribution in Southern Oceanbenthos? A genetic test using the genus *Macroscapha* (Crustacea, Ostracoda) as a model. Mol Phylogenet Evol 55:1055–1069

Brandt A, Brix S, Brökeland W, Cedhagen T, Choudhury M, Cornelius N, Danis B, De Mesel I, Diaz RJ, Gillan DC, Hilbig B, Howe J, Janussen D, Kaiser S, Linse K, Malyutina M, Brandão SN, Pawlowski J, Raupach M, Gooday AJ (2007) The Southern Ocean deep sea: first insights into biodiversity and biogeography. Nature 447:307–311

Brandt A, Linse K, Schueller M (2009) Bathymetric distribution patterns of Southern Ocean macrofaunal taxa: Bivalvia, Gastropoda, Isopoda and Polychaeta. Deep-Sea Res pt I 56:2013–2025

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA bar coding of marine metazoa. Ann Rev Mar Sci 3:18.11–18.38

Chase MC, Etter RJ, Rex MA, Quattro JM (1998) Bathymetric patterns of genetic variation in a deep-sea bivalve. Mar Biol 131:301–308

Clarke A, Johnston NM (2003) Antarctic marine benthic diversity. Oceanogr Mar Biol Ann Rev 41:47–114

Erpenbeck D, Hooper JNA, Wörheide G (2006) CO1 phylogenies in diploblasts and the 'Barcoding of Life'—are we sequencing a suboptimal partition? Mol Ecol Notes 6:550–553

France SC, Kocher TD (1996) Geographic and bathymetric patterns of mitochondrial 16S rRNA sequence divergence among deep-sea amphipods, *Eurythenes gryllus*. Mar Biol 126:633–643

Grant RA, Linse K (2009) Barcoding antarctic biodiversity; current status and the CAML initiative, a case study of marine invertebrates. Polar Biol 32:1629–1637

Griffiths HJ (2010) Antarctic Marine Biodiversity—What do we know about the distribution of life in the Southern Ocean? PLoS ONE 5(8):e11683

Hardy S, Carr C, Hardman M, Steinke D, Corstorphine E, Mah C (2010) Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools marine biodiversity. Springer Berlin, Heidelberg, pp 1–16

Hunt B, Strugnell J, Bednarsek N, Linse K, Nelson JR, 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:e9835

Hunter RL, Halanych KM (2008) Evaluating connectivity in the brooding brittle star *Astrotoma agassizii* across the Drake Passage in the Southern Ocean. J Hered 99:137–148

Krabbe K, Leese F, Mayer C, Tollrian R, Held C (2009) Cryptic mitochondrial lineages in the widespread pycnogonid *Colossendeis megalonyx* Hoek, 1881 from Antarctic and Subantarctic waters. Polar Biol 33:281–292

Lörz AN, Maas EW, Linse K, Coleman CO (2009) Do circum-Antarctic species exist in peracarid Amphipoda? A case study in the genus *Epimeria* Costa, 1851 (Crustacea, Peracarida, Epimeriidae). In: Bruce N (ed) Advances in the taxonomy and biogeography of Crustacea in the Southern Hemisphere, vol 18. ZooKeys, pp 91–128

Ratnasingham S, Hebert P (2007) BOLD: the barcode of life data system. Mol Ecol Notes 7:355–364

Wilson NG, Hunter RL, Lockhart SJ, Halanych KM (2007) Multiple lineages and absence of panmixia in the 'circumpolar' crinoid *Promachocrinus kerguelensis* from the Atlantic sector of Antarctica. Mar Biol 152:895–904

Zardus JD, Etter RJ, Rex ChaseMC, MA BoyleEE (2006) Bathymetric and geographic population structure in the pan-Atlantic deep-sea bivalve *Deminucula atacellana* (Schenck, 1939). Mol Ecol 15:639–651