RESEARCH ARTICLE

Looking back on a decade of barcoding crustaceans

Michael J. Raupach[1], Adriana E. Radulovici[2]

1 *Molecular Taxonomy of Marine Organisms, German Centre of Marine Biodiversity Research (DZMB), Senckenberg am Meer, Südstrand 44, 26382 Wilhelmshaven, Germany* **2** *Biodiversity Institute of Ontario (BIO), University of Guelph, 50 Stone Road E, Guelph (ON) N1G 2W1, Ontario, Canada*

Corresponding author: *Michael J. Raupach* (mraupach@senckenberg.de)

Abstract

Species identification represents a pivotal component for large-scale biodiversity studies and conservation planning but represents a challenge for many taxa when using morphological traits only. Consequently, alternative identification methods based on molecular markers have been proposed. In this context, DNA barcoding has become a popular and accepted method for the identification of unknown animals across all life stages by comparison to a reference library. In this review we examine the progress of barcoding studies for the Crustacea using the Web of Science data base from 2003 to 2014. All references were classified in terms of taxonomy covered, subject area (identification/library, genetic variability, species descriptions, phylogenetics, methods, pseudogenes/numts), habitat, geographical area, authors, journals, citations, and the use of the Barcode of Life Data Systems (BOLD). Our analysis revealed a total number of 164 barcoding studies for crustaceans with a preference for malacostracan crustaceans, in particular Decapoda, and for building reference libraries in order to identify organisms. So far, BOLD did not establish itself as a popular informatics platform among carcinologists although it offers many advantages for standardized data storage, analyses and publication.

Introduction

The accurate diagnosis of species represents a pivotal component for many topics, including large-scale biodiversity studies and conservation planning. Traditionally, species are identified using morphological characters. This approach requires a certain level of training in observing morphology and it usually leads to a narrow specialization in identifying organisms belonging to a restricted group of taxa (e.g. a carcinologist will likely have difficulties in identifying polychaetes and the other way around). Therefore, a routine and correct morphological identification of many taxa can be challenging, time-consuming and typically requires highly trained specialists. This is especially true for larval stages, juveniles and females which are often not included in species descriptions, resulting in a quite difficult task of assigning correct species names to specimens. In many cases morphological variability and phenotypic plasticity may also complicate a correct species determination. Furthermore, we observe a decline of taxonomists that are able to identify and characterize species of many taxa (e.g. de Carvalho et al. 2007).

As consequence of the rise of molecular biology in the last decades, the application of DNA sequence data represents a promising and effective alternative approach to identify specimens throughout all life stages (Olson et al. 1991, Caterino and Tishechkin 2006, Shank et al. 2006, Bracken-Grissom et al. 2012, Torres et al. 2014 but see Page and Hughes 2011). For animals, mitochondrial DNA (mtDNA) became highly attractive for molecular species identification due to several characteristics: generally high substitution rates, lack of introns, large copy numbers in each cell, and an almost exclusive maternal and haploid inheritance with no recombination (Ballard and Whitlock 2004, Ballard and Rand 2005, Bernt et al. 2013). In this context, a fragment of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene was proposed as so-called "DNA barcode" for animal species identification more than a decade ago (Hebert et al. 2003a). The efficacy of DNA barcoding is based on a simple assumption: each species will most likely have similar DNA barcodes representing their intraspecific variability whereas the genetic variation between species exceeds the variation within species (Hebert et al. 2003a, 2003b, 2004). In contrast to DNA taxonomy which focuses on the classification of both known and undescribed species based on sequence data only (Tautz et al. 2003, Vogler and Monagham 2007), the central aim of DNA barcoding is two-fold: 1) to assign unknown specimens to already described and classified species, and 2) to enhance the discovery of new species and facilitate identification, particularly in cryptic, microscopic, and other organisms with complex or inaccessible morphology (Hebert et al. 2003a, 2003b). Based on these assumptions, the public Barcode of Life data base (BOLD; www.boldsystems.org) acts as the core data retrieval interface, allowing researchers to collect, manage, and analyze DNA barcode data (Ratnasingham and Hebert 2007). As one of various analytical tools implemented in BOLD, barcodes can be analyzed using the Barcode Index Number (BIN) system (Ratnasingham and Hebert 2013). This approach allows a comparison of specimens identified by morphological and molecular characters.

Not surprisingly, DNA barcoding has been criticized from its beginning. In various cases, DNA barcoding was considered as a useless and expensive identification method (e.g. Will et al. 2005, Cameron et al. 2006, Ebach 2011, Taylor and Harris 2012). Other studies query methodological problems of the analysis of DNA barcodes, for example the inappropriate use of neighbor-joining trees or of fixed distance thresholds (e.g. Will and Rubinoff 2004, Goldstein and DeSalle 2010, Collins and Cruickshank 2013). Finally, another major criticism of this approach was that a single molecular marker such as COI will not necessarily provide sufficient information to deliver the resolution needed to diagnose the large number of species targeted by the initiative (e.g. DeSalle et al. 2005, Prendini 2005, Will et al. 2005). In fact, various aspects can limit the use of COI and mitochondrial DNA in general for successful species delineation. Recent speciation events, heteroplasmy, incomplete lineage sorting as consequence of phylogeographic processes, or the presence of mitochondrial pseudogenes (also known as nuclear mitochondrial DNA or numts) (e.g. Funk and Omland 2003, Bucklin et al. 2011). Furthermore, low evolutionary rates for mitochondrial genes have been demonstrated for various taxa (e.g. anthozoans and some sponges) (e.g. Shearer et al. 2002, Shearer and Coffroth 2008, Sinniger et al. 2008, McFadden et al. 2011).

Nevertheless, DNA barcoding has been successfully applied in a large number of taxonomic groups belonging to both invertebrates (e.g. Carr et al. 2011, Hausmann et al. 2011, Woodcock et al. 2013, Layton et al. 2014, Raupach et al. 2014, Raupach et al. 2015) and vertebrates (e.g. Lijtmaer et al. 2011, Ivanova et al. 2012, Knebelsberger et al. 2014). Furthermore, DNA barcodes have become an integrative part of many recently published species descriptions (e.g. Riedel et al. 2013, Khalaji-Pirbalouty and Raupach 2014, Weis et al. 2014, Hansson et al. 2015).

Within the invertebrates, the Crustacea constitute a challenging taxon for DNA barcoding. With more than 67,000 described species so far (Ahyong et al. 2013), this taxon is species-rich, morphologically diverse and ecologically important. Various crustacean species are of high economic interest (e.g. lobsters, crabs, or shrimps) and represents the basis of extensive crustacean fisheries around the world. Crustaceans can be found in all aquatic environments, and some of them successfully colonized terrestrial habitats in various degrees (e.g. talitrid amphipods, terrestrial crabs, and woodlice). However, a correct identification to the species level is not straightforward for most crustacean taxa, especially for larval and immature stages. Even as adults, numerous species are difficult to identify using morphological characters and usually require the help of taxonomists to differentiate subtle degrees of morphological variability and polymorphism within and between species. This is especially true for small deep-sea crustaceans (e.g. isopods, amphipods and tanaids), and species of the meiofauna (e.g. harpacticoid copepods).

In this review we provide an update regarding the progress of DNA barcoding in crustaceans based on descriptive statistics. Major points of the review are: taxonomic coverage, subject areas, and the use of BOLD as a major platform for the standardization of barcoding studies.

Methods

This manuscript covers research articles published between 01-01-2003 and 31-12-2014 and available in the "Web of Science" (WoS) database maintained by Thomson Reuters (http://webofknowledge.com). WoS was searched on 15-01-2015 by using "barcod*" and "crusta*" as keywords in the topic of articles hosted by all databases associated with WoS. For comparison purposes, similar searches were conducted for other arthropod taxa on the same day: Insecta ("insect*"), Chelicerata ("chelicer*") and Myriapoda ("myriapod*") in combination with "barcod*". All crustacean references were individually and carefully checked for inconsistencies, in particular false positive results (e.g. articles dealing with other taxa than crustaceans) and duplications. Only publications of the type "article" were kept for further analyses. Language was not selected as filter criterion, and non-English publications with a title and abstract in English were included. Following a strict terminology for DNA barcoding (*sensu* Hebert et al. 2003a), all articles using a different molecular marker than COI-5P' were excluded. The taxonomic focus was inferred based on the same source (titles, abstracts, keywords) and each article received a label corresponding to one crustacean order with a few exceptions: Calanoida, Harpacticoida, Cyclopoida and Siphonostomatoida were combined into "Copepoda"; Kentrogonida, Scalpelliformes and Sessilia were combined into "Cirripedia"; and the taxon Ostracoda was left at the class level. Articles that covered more than one order and did not fall into the "Copepoda" or "Cirripedia" were classified as "Crustacea". We used the recent crustacean classification of Ahyong and co-authors (2011) throughout this review as a taxonomic framework. Based on our judgment derived from reading the title, abstract, keywords and, if necessary, portions of articles, we divided all references into six subject areas: 1) identification, library (DNA barcodes used for specimen identification and/or to develop reference barcode libraries), 2) genetic variability (DNA barcodes used for studies on intraspecific genetic variability such as phylogeographic studies), 3) species description (DNA barcodes used together with morphological characters as part of species descriptions), 4) phylogenetics (DNA barcodes used in phylogenetic studies), 5) methods (new lab protocols or new primers developed for barcoding crustaceans), and 6) numts (nuclear mitochondrial DNA sequences and their implications for barcoding crustaceans). In addition and where possible, each article received a label corresponding to the habitat investigated ("marine": oceans, seas, brackish waters; "freshwater": rivers, lakes, ponds, groundwater; "mixed": marine and freshwater). Moreover, geographic labels were assigned to each article based on the main regions covered (continents and oceans). In cases of more than one ocean or continent sampled within the same article, multiple labels were assigned.

In order to verify the popularity and use of the BOLD workbench among crustacean barcoders, each article was searched for referencing BOLD and given a label: 'YES' or 'NO'. If a BOLD project was mentioned by code or title, subsequent steps were followed to find particular records in BOLD and import them into a dataset: 1) search by project code/title in BOLD Workbench, 2) copy all records from that project, and 3)

add records to dataset. All public records stored in BOLD and generated by crustacean barcoding studies can be retrieved by searching DS-CRST (Title: Crustacean Barcoding Studies) in BOLD or by going directly to the corresponding DOI associated with this dataset (http://dx.doi.org/10.5883/DS-CRST). By using a project code as search term, all records of that project were imported, regardless of its history (i.e., records added or removed from a project) between the publication date and January 2015). Some articles mentioned the use of BOLD without providing a project code. In such cases, we were able to find records by the process IDs mentioned in the publication or by searching BOLD based on taxa names. However, when tracking records was not a straightforward process, we excluded those studies from our BOLD-related analyses. DS-CRST in BOLD was used for standard barcoding analyses: number of species versus number of BINs, taxon ID tree and distance summary. Geographic coordinates, where available, were exported and used to create a map in QGIS (QGIS Development Team 2015).

Additional bibliographic data were compiled for all references: publication title, first authors' names, journal name, publication year, open-access feature, and the number of citations (as provided by WoS). The major results of our literature review are summarized graphically; a table containing all raw data is available as Suppl. material 1.

Results

Our search in WoS produced 243 hits associated with the terms "barcod*" and "crusta*", 1,064 references for "barcod*" and "insect*", 67 for "barcod*" and "chelicer*" and eight for "barcod*" and "myriapod*" (Fig. 1). In total, 1,382 publications were found for all Arthropoda. Our initial list of 243 crustacean references was revised and reduced to 164 publications after removing duplicates and mislabeled references. All other arthropod references were not revised in detail. The number of barcoding publications showed a fast increase from the first and singular crustacean article published in 2005 (Page et al. 2005) up to 30 publications in 2012 (Fig. 2). In 2013, a slight decrease to 29 publications was observed, followed by an increase to 31 publications in 2014. However, the frequencies of the different categories fluctuated each year (Fig. 2).

The taxonomic coverage of the 164 barcoding publications showed a strong preference for the Decapoda ($n = 60$, 36.7%), followed by the mixed taxon of "Crustacea" ($n = 28$, 17%), the Amphipoda ($n = 21$, 12.8%), Copepoda ($n = 18$, 11%), and Diplostraca ($n = 13$, 8%) (Table 1). All other crustacean taxa have been investigated by less than ten publications: Isopoda ($n = 6$, 3.7%), Anostraca and Cirripedia ($n = 5$, 3%), Stomatopoda ($n = 3$, 1.8%), and Bathynellacea ($n = 2$, 1.2%). The Euphausiacea, Ostracoda, and Tanaidacea have been analyzed only once (each with $n = 1$, 0.6%).

Our investigation also revealed that most crustacean barcoding studies focus on the identification of specimens and the expansion of reference libraries for various taxa ($n = 64$, 39.1%) (Table 2). Beside identification, DNA barcodes were frequently used in publications analyzing the genetic variability of species ($n = 44$, 26.8%) and as ad-
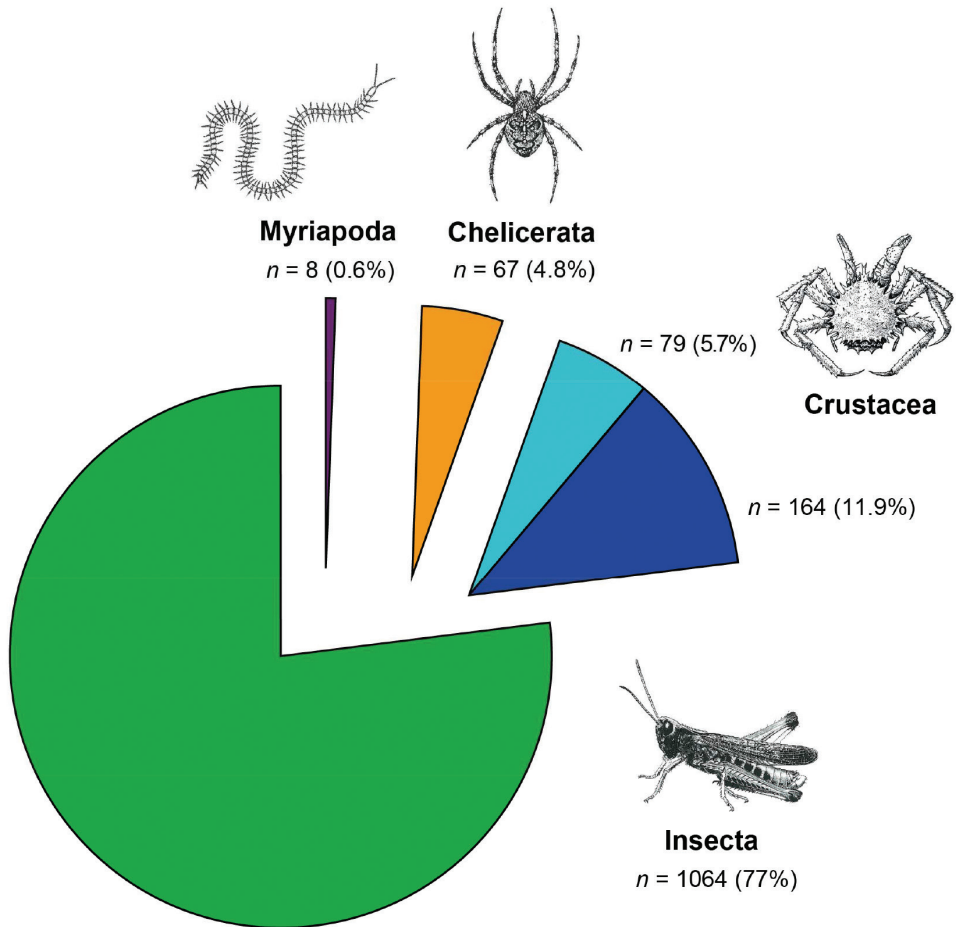
Figure 1. DNA barcoding studies of the Arthropoda. Total number and percentage values of articles published with "barcod*" and insect*" (green), "crusta*" (blue), "chelicer*" (orange), or "myriapod*" (violet) as keywords in their topic and listed in the Web of Science (period covered: 2003-2014; *n* = 1,382). For crustaceans, the total number of articles is split into: 1) the number of articles removed from our analysis (duplications and false positives) (pie sector in light blue) and 2) the core number of articles used in this review (pie sector in dark blue). Arthropod illustrations were modified from Gruner (1993) and Dathe (2003).

ditional characters in species descriptions (*n* = 32, 19.5%). Relatively small numbers of publications covered the use of DNA barcodes as part of phylogenetic reconstructions (*n* = 11, 6.7%), the publication of new protocols and methods to obtain barcode sequences (*n* = 9, 5.5%), and the study of numts (*n* = 4, 2.4%).

Approximately two thirds of the barcoding studies focused on the marine environment (*n* = 99, 60.4%) and only one third dealt with freshwater systems (*n* = 49, 29.8%) (Fig. 3). Six studies covered taxa from both marine and freshwater habitats (*n* = 6, 3.7%),
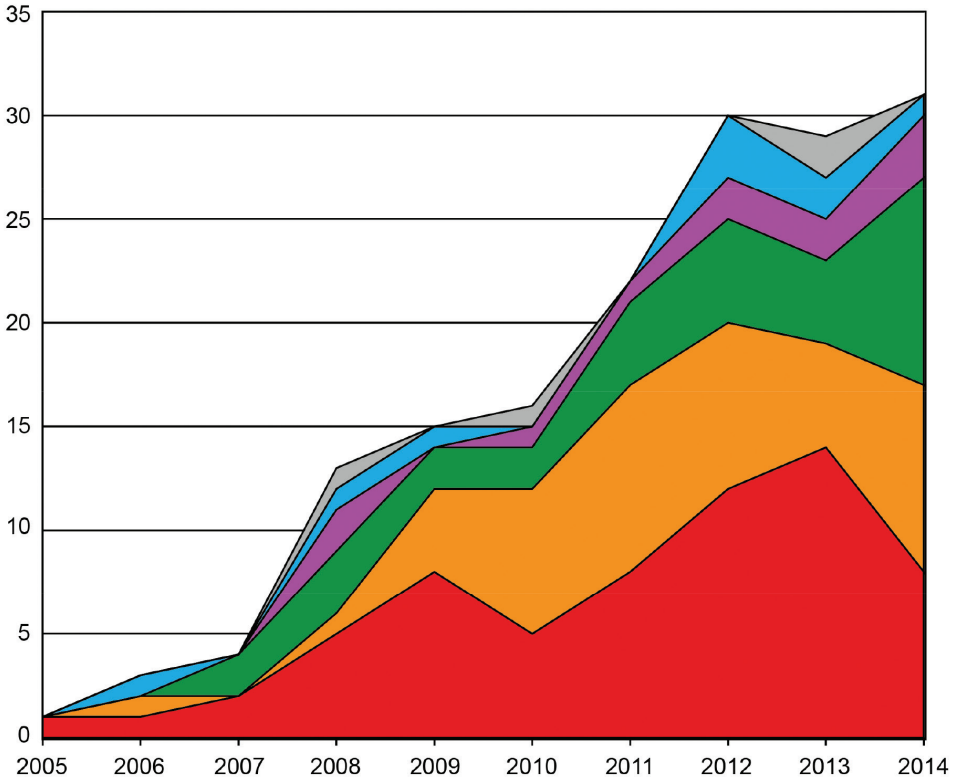
Figure 2. Subject areas of DNA barcoding studies of the Crustacea. Number of articles with "barcod*" and "crusta*" as keywords in their topic as retrieved from the Web of Science (period covered: 2003–2014; *n* = 164) and divided into six subject areas (from bottom to top): identification and barcode library (red), genetic variability (orange), species description (green), phylogenetics (violet), methods (blue), and numts (grey).

and for ten studies no classification was possible (6.1%). Interestingly, no study was found analyzing terrestrial crustaceans exclusively (e.g. woodlice) (Suppl. material 1). Our geographic investigation covered only the major divisions of land and water, namely continents and oceans. It should be noted that publications can include taxa from more than one environment or geographic region. The analyzed DNA barcoding publications covered all oceans (the Arctic, Atlantic, Indian, Pacific, and Southern Ocean), with a focus on the Pacific Ocean (*n* = 49, 25.5%), followed by the Atlantic Ocean (*n* = 28, 14.5%) (Fig. 3). In the case of continents, five were sampled: Asia (*n* = 8, 4.2%), Australia (*n* = 10, 5.2%), Europe (*n* = 17, 8.9%) as well as North and South America (*n* = 17, 8.9%; *n* = 3, 1.6%) (Fig. 3). Ten studies (5.2%) had a global geographic coverage, whereas it was impossible to place the origin of the specimens analyzed for 11 studies (5.7%), e.g. studies which used data mined from GenBank (Suppl. material 1).

The vast majority of publications (*n* = 129, 78.7%) did not mention BOLD in their text (label 'NO' in Suppl. material 1). The remaining 35 publications (21.3%)

Table 1. Number of publications of the Crustacea using DNA barcodes. "Barcod*" and "crusta*" were used as keywords in the Web of Science (2003–2014). For comparison, the most recent species count per taxon is given in a separate column (based on Ahyong et al. 2011).

Taxon		Publications	(%)	Number of described species
Malacostraca	Decapoda	60	36.7	14,895
	Amphipoda	21	12.8	9,896
	Isopoda	6	3.7	10,661
	Stomatopoda	3	1.8	460
	Bathynellacea	2	1.2	241
	Euphausiacea	1	0.6	87
	Tanaidacea	1	0.6	1,069
Maxillopoda	Copepoda	18	11	15,976
	Cirripedia	5	3	1,306
Branchiopoda	Diplostraca	13	8	821
	Anostraca	5	3	313
Ostracoda		1	0.6	7,577
"Crustacea"		28	17	n. a.
Total		**164**	**100**	

Table 2. Subject area and taxonomic rank of DNA barcoding studies of the Crustacea. Number of articles were retrieved by using "barcod*" and "crusta*" as keywords in the topic of articles hosted by the Web of Science (period covered: 2003–2014).

	Identification, library	Genetic variability	Species description	Phylogenetics	Methods	numts
Decapoda	26	11	15	5	1	2
Amphipoda	4	15	1	1		
Isopoda		2	3		1	
Stomatopoda	3					
Bathynellacea		1	1			
Euphausiacea	1					
Tanaidacea	1					
Copepoda	4	5	6	3		
Cirripedia	2		2	1		
Diplostraca	2	8	3			
Anostraca	1	1	1	1	1	
Ostracoda	1					
"Crustacea"	19	1			6	2
Total	**64**	**44**	**32**	**11**	**9**	**4**

used BOLD as part of their study with project titles/codes (n = 23, 14%), or with projects created *a posteriori*, similar to the workflow of sequence publication in Gen-Bank (n = 3, 1.8%). A handful of articles used BOLD exclusively for data mining or as an identification engine for DNA sequences or mentioned BOLD as part of current
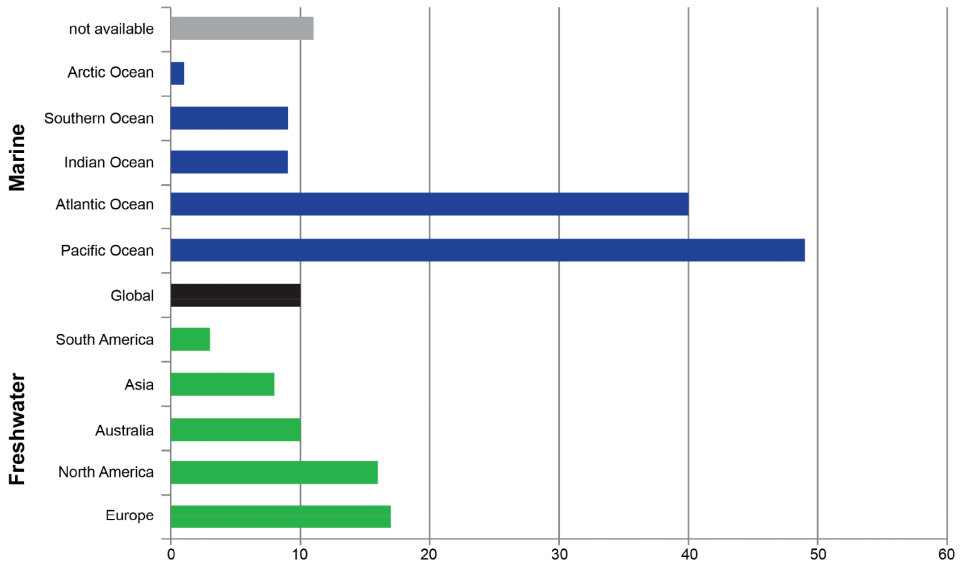
Figure 3. Geographic and habitat focus of the analyzed DNA barcoding studies of the Crustacea. Studies were listed in the Web of Science (period covered: 2003–2014, *n* = 164), with the number of publications shown on the X axis. Green bars indicate freshwater studies, dark blue bars marine studies. A black bar represents studies that were performed on a global scale. For 11 studies, no classification was possible (grey bar). Note that publications can include taxa from more than one habitat or region.

or proposed DNA barcoding workflows. A total of 6,270 records were successfully tracked and imported into DS-CRST (Fig. 4). Approximately half of the records belonged to Malacostraca (*n* = 3,208, 51.2%), followed by Branchiopoda (*n* = 1,802, 28.8%), Maxillopoda (*n* = 728, 11.6%), and Ostracoda (*n* = 532, 8.5%). In total, 5,740 records (91.5%) had species names (Linnaean names or interim names) while 530 crustaceans (8.5%) remained unidentified (March 2015). Data owners inserted 860 species names whereas BOLD assigned 1,109 BINs to the entire dataset (Fig. 5). Furthermore, 413 records (6.6%) lacked details about the country of sample collection, 845 records (13.5%) lacked GPS coordinates whereas 3,573 specimens (57%) provided no image for the voucher. Records with collection details were divided between Canada (*n* = 2,293, 36.6%) and Mexico (*n* = 1,305, 20.8%) plus another 38 countries with much fewer records (Fig. 6). In addition to 6,270 DNA barcodes, some records used supplementary genetic markers (12S, 16S, and/or 18S rDNA). A number of 1,338 records (21.3%) had no successful chromatogram ("trace") associated, one COI sequence (0.02%) had stop codons and 45 records (0.7%) had been flagged as misidentification or contamination between the publication date and March 2015. A total of 2,082 records (33.2%) were non-barcode compliant (i.e., one of the following criteria was not fulfilled: country, two trace files, a fragment length of at least 500 base pairs, and less than 1% ambiguities).
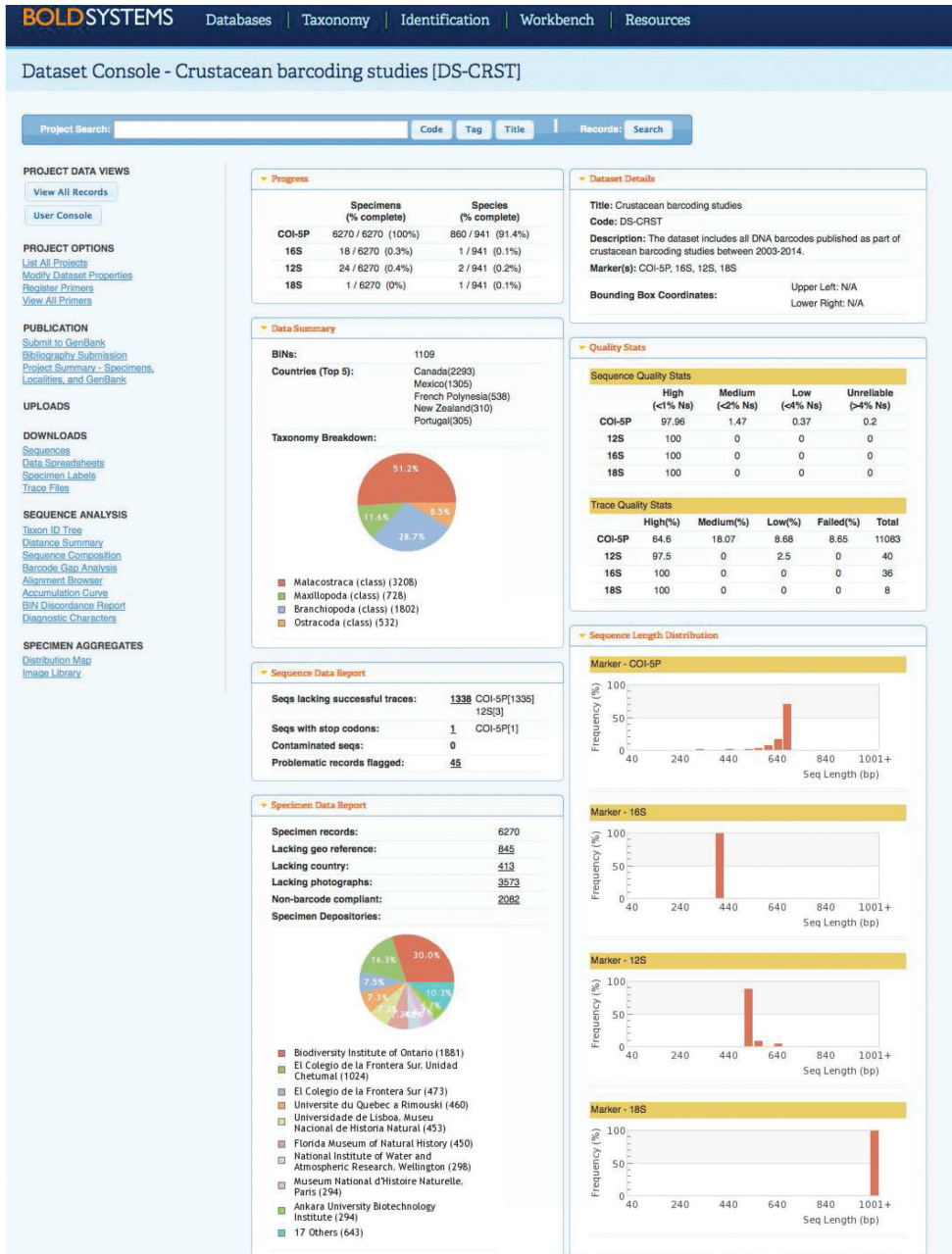
Figure 4. Project console for DS-CRST in BOLD. Various statistics for the current status of specimens are displayed: record count, species count, taxonomy breakdown, specimen depositories, country of collection, sequence count, flagged records count, trace count, image count. Note that BOLD is a dynamic environment and updates will be reflected on the project console.
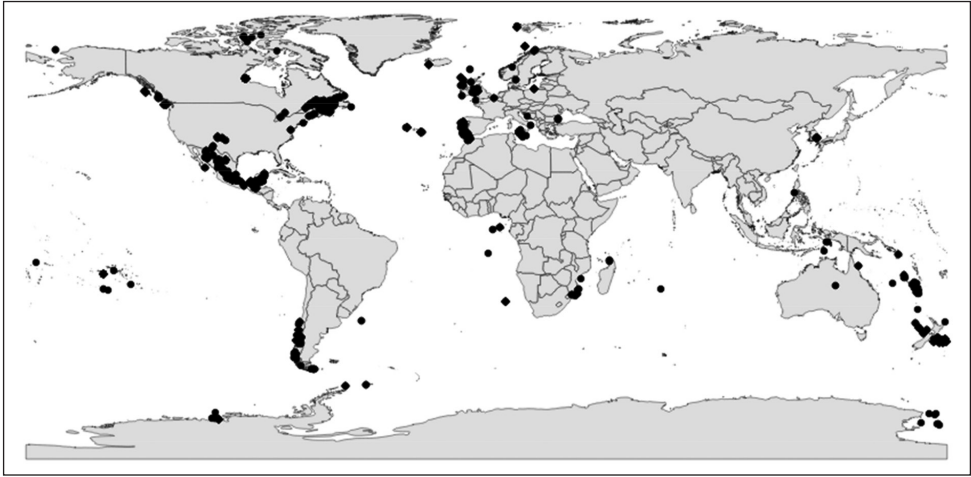
Figure 5. Example for a BIN page in BOLD. The amphipod *Rhachotropis aculeata* (Lepechin, 1780) is registered in the BIN registry as BOLD:AAB3310. Note that BOLD is a dynamic environment and updates will be reflected on the BIN page, including BIN changes.

Figure 6. Sampling locations for crustaceans used in this review. GPS data was taken from the dataset DS-CRST in BOLD. Note that only 5,425 out of 6,270 records had GPS coordinates and are reflected here.

We found 76 different journals publishing articles dealing with DNA barcoding and crustaceans. Most studies were published in Zootaxa (n = 23, 14%), followed by the Journal of Crustacean Biology and PLOS ONE (each with n = 9, 5.5%), Molecular Ecology Resources (n = 7, 4.3%), Crustaceana and Invertebrate Systematics (each with n = 6, 3.7%). A number of 50 journals (65.8%) had only one article dealing with crustacean barcoding. Only 33 articles (20.1%) were open access as they were published in open access journals (e.g. PLOS ONE , ZooKeys) or in subscription journals where authors chose to publish their work as open-access (Suppl. material 1). The author list revealed a total number of 700 authors with 125 being first authors. The most prolific first author of crustaceans and DNA barcodes was Arthur Anker (7 articles in total, 4.3%), followed by Tomislav Karanovic (4 articles, 2.4%) and Ann Bucklin, Manuel Elías-Gutiérrez, Laetitia Plaisance, and Chien-Hui Yang, each with three first-authored papers involving DNA barcoding of crustaceans. The most cited article by far was written by Song and co-authors (2008) discussing the effects of numts for DNA barcoding (292 citations), followed by a publication of Lefébure and co-authors (2006) discussing threshold calculations for a successful species identification (185 citations), Witt and co-authors (2006) with one of the first articles on the role of DNA barcoding in highlighting the existence of cryptic species (172 citations), and Costa and co-authors in 2007 with the first comprehensive study testing the efficacy of DNA barcoding for crustacean species identification (165 citations) (Table 3). In the case of phylogenetic analyses using DNA barcode data the most cited article was published by Matzen da Silva and co-authors (2011a), focusing on the Malacostraca (21 citations). Finally, Lai and co-authors (2010) included DNA barcodes in their revision of the *Portunus pelagicus* (Linnaeus, 1758) species complex. This article was cited 23 times.

Table 3. Most cited crustacean barcoding articles per subject area. Data obtained from Web of Science based on a query with 'barcod*' and 'crusta*' as keywords in the topic of articles published between 2003 and 2014. Citations are given as the total number of citations since publication and the average number of citations per year (in brackets).

Subject area	Title	Authors	Journal	Year	Citations
Identification, library	Biological identifications through DNA barcodes: the case of the Crustacea	Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN	Canadian Journal of Fisheries and Aquatic Sciences	2007	165 (18.3)
Genetic variability	DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation	Witt JDS, Threloff DL, Hebert PDN	Molecular Ecology	2006	172 (17.2)
Species description	A revision of the *Portunus pelagicus* (Linnaeus, 1758) species complex (Crustacea: Brachyura: Portunidae), with the recognition of four species	Lai JC, Ng PKL, Davie PJF	Raffles Bulletin of Zoology	2010	23 (3.8)
Phylogenetics	Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca)	Matzen da Silva J, Creer S, dos Santos A, Costa AC, Cunha MR, Costa FO, Carvalho GR	Public Library of Science ONE	2011	21 (4.2)
Methods	Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation	Lefébure T, Douady CJ, Gouy M, Gibert J	Molecular Phylogenetics and Evolution	2006	185 (18.5)
numts	Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified	Song H, Buhay JE, Whiting MF, Crandall KA	Proceedings of the National Academy of Sciences of the USA	2008	292 (36.5)

Discussion

During the past few years, crustaceans have become a popular target for DNA barcoding among the Arthropoda, being outnumbered only by barcoding studies of the Insecta (Fig. 1). Although the observed ratio of barcoding articles of insects compared to barcoding publications for crustaceans is high (6:1), this is not surprising since insects represent the most species-rich taxon on earth (app. 1 million species described and app. 5 million species estimated) (Chapman 2009). Crustacean publications showed a continuous increase starting with the first publication in 2005. In recent years, the numbers of crustacean publications seem to have reached a plateau with approximately 30 publications per year.

Although we used a highly popular database which indexes scientific literature, we are aware that an unknown number of references are missing from our study. This is mainly caused by two reasons: 1) the term "DNA barcoding" was not used in the publication although COI sequences were applied for species identification (e.g. Shih and Cai 2007), and 2) some journals might not be indexed in WoS yet. Despite this somewhat incomplete sampling of literature, we think that our review reflects the application of DNA barcodes in carcinology in a representative way.

Taxonomic overview

A rapid investigation of the taxonomic diversity covered in the 164 barcoding publications showed the highest frequency for Malacostraca (n = 94, 57.4%), the class with the largest number of crustacean species (Ahyong et al. 2011, Appeltans et al. 2012) and the most familiar ones (e.g. lobsters, crabs, shrimps, krill, beach hoppers, woodlice). Within the Malacostraca, the ecologically and economically important Decapoda were most popular for barcoding studies (n = 60 articles, 36.7%), followed by the Amphipoda (n = 21, 12.8%), a species-rich group inhabiting most aquatic habitats and even some terrestrial habitats with high humidity (e.g. supralittoral, rainforests) (Table 1). Other malacostracan orders seem to be less popular for specific DNA barcoding despite high or moderate numbers of known species, e.g. the Isopoda (10,000 species, 6 publications, 3.7%) or Tanaidacea (1,000 species, 1 publication, 0.6%). So far, no study focused specifically on the Cumacea (1,500 species) or Mysida (app. 1,200 species). We hope that scientists working on these taxa become more aware of the benefits of DNA barcoding as part of their studies, inducing an increase in the number of publications in the near future. The Maxillopoda, the second most species-rich crustacean class representing much of the marine and freshwater zooplankton, was covered in 23 studies (14%). Copepods were most popular among the maxillopods (n = 18, 11%), as it can be expected for a species-rich group (app. 16,000 species) with ecological importance in planktonic food-webs, as opposed to Cirripedia covered by only five publications (3%). The third most popular crustacean class was the Branchiopoda (n = 18, 11%), a group of crustaceans frequently encountered in freshwater habitats. Surprisingly, the species-rich class of Ostracoda (app. 7,500 species) has been covered as an exclusive taxon in only one publication (0.6%) until now. Furthermore, 28 publications (17%) had a mixture of different taxonomic groups (i.e. multiple orders were sampled) and were labeled as "Crustacea". These were usually subject-oriented (e.g. reviews on various topics) rather than taxon-oriented publications. The remaining two classes of crustaceans, Remipedia and Cephalocarida, have not been targeted by DNA barcoding studies yet (January 2015). A search using the taxon names and "COI" in GenBank returned 24 hits for the Remipedia and 20 for the Cephalocarida. Not surprisingly, these species-poor taxa (Remipedia: 18 species, Cephalocarida: 13 species; Ahyong et al. 2011) are also less important from an economic or ecological perspective. Although we do not expect comprehensive barcoding studies for species-poor taxa in the near future, we believe they might be targeted as part of comprehensive regional studies.

Subject areas of DNA barcoding publications

In contrast to the total number of publications, which revealed a steady increase followed by a relative plateau, the trend for the six subject areas (see methods) showed large fluctuations from year to year (Fig. 2). Overall, our analyses revealed that most barcoding studies focused on species identification linked to building or expanding existing reference libraries of COI sequences ($n = 64$, 39.1%), followed by analyses of the intraspecific genetic variability ($n = 44$, 26.8%) and by species descriptions that use DNA barcodes as additional characters ($n = 32$, 19.5%) (see Table 2). Less common were studies using DNA barcodes in molecular phylogenetics ($n = 11$, 6.7%), new methods and protocols ($n = 9$, 5.5%) or the possible effects of numts for barcoding studies of crustaceans ($n = 4$, 2.4%). We provide more details for each subject area in the following paragraphs.

Species identification and DNA barcode libraries

Species identification based on DNA barcodes relies on the existence of reference libraries which consist of COI sequences from specimens previously identified by experts based on traditional methods (i.e., morphological characters). Consequently, many barcoding studies published so far deal with the development of comprehensive barcode libraries (e.g. Dincă et al. 2010, Baird et al. 2011, Zhou et al. 2011, Raupach et al. 2014, Rougerie et al. 2014) and their use to identify unknown specimens (e.g. Holmes et al. 2009, Strutzenberger et al. 2011, Shen et al. 2013, Knebelsberger et al. 2014). Similar to this general trend, most crustacean publications reviewed here were found to fit in this category ($n = 64$, 39.1%), with a constant increase over the years (Fig. 2). In terms of crustacean diversity, most studies were performed on the Decapoda ($n = 26$, 40.6%) and the mixed group of "Crustacea" ($n = 19$, 29.9%). All the other crustacean taxa were investigated by less than five publications each (Table 2). A constantly growing library of DNA barcodes will offer numerous applications, such as seafood traceability (e.g. Haye et al. 2012, Nicolè et al. 2012, Di Pinto et al. 2013), the identification of larvae (e.g. Barber and Boyce 2006, Webb et al. 2006, Weigt et al. 2012), and tools for ecological studies in general (e.g. Valentini et al. 2009, Bowser et al. 2013, Burghart et al. 2014). Moreover, comprehensive barcode libraries will become essential for biomonitoring applications based on modern high-throughput sequencing technologies (e.g. Fonseca et al. 2010, Hajibabaei et al. 2011, Shokralla et al. 2012, Thomsen et al. 2012, Zhou et al. 2013, Leray and Knowlton 2015).

DNA barcodes and intraspecific genetic variation

The study of intraspecific genetic variation in relation to geography has become very popular in recent decades and resulted in the formation and expansion of a new research

field, namely phylogeography (Avise 2000, Hickerson et al. 2010). In the past, numerous phylogeographic studies have been published on various taxa, including crustaceans (e.g. Audzijonyte et al. 2006, Krebes et al. 2010, Campo et al. 2010, Garcia-Merchan et al. 2012, Santamaria et al. 2013). The body of sequence data generated through such phylogeographic studies was actually the background on which DNA barcoding was proposed as a method for species identification across the entire animal kingdom (Hebert et al. 2003a, 2003b). As COI sequences are used in DNA barcoding as well as in phylogeography, it is no surprise that publications with "barcod*" and "crusta*" as keywords investigate the level of genetic diversity within species as well (Fig. 2). Our review identified 44 studies for this category. Interestingly, the amphipods (n = 15, 34.1%) were more popular than decapods (n = 11, 25%) for this subject area. All other crustacean groups were present in less than ten publications per taxon (Table 2). To verify the progress in crustacean phylogeographic studies, we used phylogeograph*", "crusta*" and "cytochrome oxidase I" as keywords in WoS and retrieved 152 articles. The large discrepancy between our review and WoS is caused by the fact that the term "DNA barcode" is normally not used in phylogeographic studies as keyword. However, the variation of intraspecific genetic diversity in relation with spatial scales may have an important impact on the efficacy of DNA barcoding (Bergsten et al. 2012). Therefore we encourage researchers interested in phylogeography to address problems related to DNA barcoding as well.

New species description including DNA barcodes

Ideally, DNA barcoding and species discovery would be seen as intertwined. Whereas the main objective of DNA barcoding is to identify unknown specimens based on reference libraries, an additional outcome is reflected in the identification of unknown genetic clusters that might represent new species. As such, DNA barcodes represent powerful diagnostic supplementary characters that accelerate and revive traditional morphological taxonomy but do not replace it (DeSalle et al. 2005). It is not surprising that more and more species descriptions include barcode sequences or that entire monographs are triggered by the results of DNA barcoding (Butcher et al. 2012, Landry et al. 2013). In total, we found 32 publications incorporating DNA barcodes as part of new species descriptions of crustaceans (Table 2, Fig. 2). Again, the Decapoda were the dominant taxon (n = 15, 46.9%). Other studies focused on Copepoda (n = 6, 18.8%), Diplostraca and Isopoda (each with n = 3, 9.4%), Cirripedia (2, 6.2%), and the Amphipoda, Anostraca, and Bathynellacea (each with n = 1, 3.1%). In this context we used Thomson Reuter's Zoological Record through the Index of Organism Names (www.organismnames.com) to calculate the rate of crustacean species descriptions during the last decade. The Metrics function and the "Graphs of new taxa over time" option showed a fluctuating rate between 681 (minimum in 2014) and 1,263 (maximum in 2008) with a mean of 891 new crustacean

species being described each year, with one third representing decapods. This large discrepancy between the numbers of new species being described per year and the numbers of studies implementing DNA barcoding for species description (278:1) reflects the hesitation of taxonomists to adopt new approaches on large scale or their limited access to sequencing technologies. We hope for a change of mentality in the near future and an increased access to molecular labs as a combination of morphological and molecular data allows more detailed species descriptions as part of an aspired integrative taxonomy (e.g. Dayrat 2005, Padial et al. 2010, Schlick-Steiner et al. 2010). In addition, the new approach would also include a standardized analytical package: raw distance data (percent divergence), diagnostic characters and phylogenetic trees (Goldstein and DeSalle 2010).

DNA barcodes and phylogenetic analyses

During the last years, COI sequences combined with other mitochondrial and nuclear markers have been frequently used to reconstruct the phylogeny of various taxa of the Crustacea (e.g. Blanco-Bercial et al. 2011, Matzen da Silva et al. 2011b, Klaus et al. 2013). Similar to phylogeographic studies, the term DNA barcode is typically not used in this context. Nevertheless, we found 11 publications using the term DNA barcodes as part of molecular phylogenetic studies, with five studies analyzing relationships of the Decapoda (45.4%), three references for the Copepoda (27.3%), and one reference for the Amphipoda, Anostraca and Cirripedia (each 9.1%), respectively (Table 2). Whereas DNA barcodes may be useful to reconstruct recent radiations and/or speciation events in some cases (e.g. Schubart et al. 1998, Cristescu and Hebert 2002), the combination of mitochondrial DNA with more conserved nuclear markers (e.g. 18S or 28S rRNA genes) is essential when reconstructing higher taxa phylogenies (Schubart 2009).

Laboratory protocols and methods

Although DNA barcoding as a molecular method for species identification has been in use for more than a decade, techniques for generating, applying, and analyzing barcode data are still being improved to guarantee an efficient workflow (e.g. Lopez and Erickson 2012). We found nine studies presenting new protocols for DNA extraction or newly designed primer pairs for crustaceans. Six publications focused on various taxa of the "Crustacea" (66.7%), and one publication for each of the remaining taxa: Anostraca, Decapoda and Isopoda (each 11.1%). As DNA barcoding becomes more and more accepted in carcinology, we are convinced that the development of more specialized protocols as well as the optimization of taxa-specific primer pairs will increase in the near future (e.g. Schubart 2009), making DNA barcoding easier and more popular for carcinologists.

Nuclear copies of mitochondrial DNA: numts

The unwanted amplification of nuclear copies of mitochondrial DNA (numts) represents a problem not only for the analyses of DNA barcodes (COI sequences) but mitochondrial genes in general (Bensasson et al. 2010, Hazakani-Covo et al. 2010). Whereas numts can be useful for phylogenetic or population structure analyses in some special cases (Pons and Vogler 2005, Hazakani-Covo 2009, Soto-Calderón et al. 2014), their presence may represent a serious problem for barcode studies. Numts are known for various taxa, including mammals (e.g. Thalmann et al. 2005, Kim et al. 2006, Soto-Calderón et al. 2014), insects (e.g. Pons and Vogler 2005, Pamilo et al. 2007, Ruiz et al. 2013, Song et al. 2014), as well as crustaceans (e.g. Schneider-Broussard and Neigel 1997, Williams and Knowlton 2001, Buhay 2009, Baeza and Fuentes 2013). Until January 2015, only four studies highlighted the potential issues of numts for DNA barcoding studies of the Crustacea, with a focus on decapods (*n* = 2, 50%) and the mixed "Crustacea" (*n* = 2, 50%). Whereas most numts were found within decapods, it is actually unclear if such pseudogenes may become problematic for other crustacean taxa too. In order to minimize the risks caused by numts for DNA barcoding studies we recommend rigorous quality control of all barcode sequences. This includes a strict use of high-quality chromatograms, a translation of the barcode sequences to amino acids to detect insertions, deletions and/or in-frame stop codons, and the use of taxa-specific primers for some groups (see Song et al. 2008, Schubart 2009).

Crustacean DNA barcoding and BOLD

In March 2015, the Public Data Portal of BOLD was hosting more than 80,000 DNA barcodes representing about 5,700 crustacean species (plus a large amount of unidentified specimens) and 10,000 BINs. Only 8% (6,270 records; 860 species names) were directly associated with crustacean barcoding studies (35 publications, Suppl. material 1) as the respective authors used BOLD for their research. The remaining crustacean barcodes were associated with private projects and with published sequences mined from GenBank. By retrieving COI data from GenBank that were generated as part of non-barcoding studies but fulfill the 'barcode' requirements, BOLD is assembling all information pertaining to reference libraries in a single database, thus reducing the risk of duplication in barcoding the same taxa multiple times. Despite a decade of work in the field of DNA barcoding, only app. 7,000 crustacean species have been barcoded to date (public and private data, available from the Taxonomy Browser in BOLD). However, existing biodiversity catalogues specify a number of more than 67,000 crustacean species described worldwide (Ahyong et al. 2011) and app. 150,000 undescribed species (Chapman 2009), although recent inventories give estimate numbers as high as 200,000–360,000 species in the marine environment alone (Appeltans et al. 2013). In times of limited taxonomic expertise as well as resources and rampant accumulation of barcode data, the option of using a DNA-based registry (such as the BIN system) for crustacean diversity has clear advantages. A

fast and accurate clustering of COI sequences into groups corresponding to presumptive species (BINs) would assist in screening large amounts of data and highlighting those cases that need detailed investigation (e.g. taxonomic synonymy, cryptic diversity, specimen misidentification). For instance, 10,000 BINs are available for crustaceans in BOLD, and a rapid initial investigation would require morphological identification of roughly 10,000 specimens as opposed to 80,000 screened through DNA sequencing. Besides identifying cohesive genetic clusters, the BIN system provides a persistent catalogue of biodiversity as each BIN has a unique alphanumeric identifier. In addition, each BIN has an individual webpage in BOLD which displays all the available information: BIN member count, nearest neighbour, genetic distance summary, haplotype network, images, sampling map, specimen depositories, collectors, identifiers, data status (public or private), data owners, annotations inserted by the barcoding community and publications using a specific BIN (Fig. 5). Multiple options to download specimen and/or sequence data are also given.

A growing database such as BOLD, which follows specific high standards for data quality, will certainly be useful for large-scale analyses in crustacean phylogeography, biogeography and biodiversity assessment and will offer support for technological advances such as high-throughput sequencing.

Conclusions

Our review shows that DNA barcoding has gained popularity in carcinology and that the most popular group targeted for various related topics are the malacostracan crustaceans, in particular decapods. As the main goal of DNA barcoding is to assign unknown specimens to known species, most crustacean barcoding studies were found to build or use existing reference libraries for identification purposes and this trend will surely continue and probably increase in the future. The generation of comprehensive barcode libraries will represent a challenging but also an important task, especially for some species-rich habitats (e.g. the deep sea or coral reefs), where our general knowledge about crustacean diversity, in particular species numbers, is still poor. A second objective of DNA barcoding is to accelerate species discovery, particularly in cryptic, microscopic and other organisms with complex or inaccessible morphology. We believe that more progress will be made in this direction as well.

Crustacean taxonomy seems to be slowly incorporating DNA barcoding in the field as the top journal in this field is a taxonomic journal and the most prolific first authors have a taxonomic background. However, a larger acceptance and application is highly desirable, and therefore we encourage a stronger cooperation between "classical" taxonomists and the DNA barcoding community. Moreover, the term "DNA barcode" should only be used for COI-5P` sequences (Hebert et al. 2003a). In this context we also recommend the use of BOLD for data storage, analysis and publication. By following such standards in data generation and analysis, large comparisons across taxonomic groups would be easily drawn for better predictions of biodiversity, in particular molecular, patterns and species diversity in general.

Acknowledgements

References

Ahyong ST, Lowry JK, Alonso M, Bamber RN, Boxshall GA, Castro P, Gerken S, Karaman GS, Goy JW, Jones DS, Meland K, Rogers CR, Svavarsson J (2011) Subphylum Crustacea Brünnich, 1772. In: Zhang Z-Q (Ed.) Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. Zootaxa 3148: 165–191.

Appeltans W, Ahyong ST, Anderson G, Angel MV, Artois T, Bailly N, Bamber R, Barber A, Bartsch I, Berta A, Blazewicz-Paszkowycz M, Bock P, Boxshall G, Boyko CB, Brandão SN, Bray RA, Bruce NL, Cairns SD, Chan T-Y, Cheng L, Collins AG, Cribb T, Curini-Galletti M, Dahdouh-Guebas F, Davie PJF, Dawson MN, De Clerck O, Decock W, De Grave S, de Voogd NJ, Domning DP, Emig CC, Erseus C, Eschmeyer WN, Fauchald K, Fautin DG, Feist SW, Fransen CHJM, Furuya H, Garcia-Alvarez O, Gerken S, Gibson D, Gittenberger A, Gofas S, Gómez-Daglio L, Gordon DP, Guiry MD, Hernandez F, Hoeksema BW, Hopcroft RR, Jaume D, Kirk P, Koedam N, Koenemann S, Kolb JB, Kristensen RM, Kroh A, Lambert G, Lazarus DB, Lemaitre R, Longshaw M, Lowry JK, Macpherson E, Madin LP, Mah C, Mapstone G, McLaughlin PA, Mees J, Meland K, Messing CG, Mills CE, Molodtsova TN, Mooi R, Neuhaus B, Ng PKL, Nielsen C, Norenburg JL, Opresko DM, Osawa M, Paulay G, Perrin W, Pilger JF, Poore GCB, Pugh P, Read GB, Reimer JD, Rius M, Rocha RM, Saiz-Salinas JI, Scarabino V, Schierwater B, Schmidt-Rhaesa A, Schnabel KE, Schotte M, Schuchert P, Schwabe E, Segers H, Self-Sullivan C, Shenkar N, Siegel V, Sterrer W, Stöhr S, Swalla BJ, Tasker ML, Thuessen EV, Timm T, Todaro MA, Turon X, Tyler S, Uetz P, van der Land J, Vanhoorne B, van Ofwegen LP, van Soest RWM, Vanaverbeke J, Walker-Smith G, Walter TC, Warren A, Williams GC, Wilson SP, Costello MJ (2012) The magnitude of global marine species diversity. Current Biology 22: 2189–2202. doi: 10.1016/j.cub.2012.09.036

Audzijonyte A, Väinölä R (2006) Phylogeographic analyses of a circumarctic coastal and boreal lacustrine mysid crustacean, and evidence for fast postglacial mtDNA rates. Molecular Ecology 15: 3287–3301. doi: 10.1111/j.1365-294X.2006.02998.x

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, 464 pp.

Baeza JA, Fuentes MS (2013) Exploring phylogenetic informativeness and nuclear copies of mitochondrial DNA (numts) in three commonly used mitochondrial genes: mitochondrial phylogeny of peppermint, cleaner, and semi-terrestrial shrimps (Caridea: *Lysmata*, *Exhippolysmata*, and *Merguia*). Zoological Journal of the Linnean Society 168: 699–722. doi: 10.1111/zoj.12044

Baird DJ, Pascoe TJ, Zhou X, Hajibabaei M (2011) Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: formalin preservation vs. specimen age. Journal of the North America Benthological Society 30: 125–130. doi: 10.1899/10-013.1

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Molecular Ecology 13: 729–744. doi: 10.1046/j.1365-294X.2003.02063.x

Ballard JWO, Rand RM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. Annual Review in Ecology, Evolution and Systematics 36: 621–642. doi: 10.1146/annurev.ecolsys.36.091704.175513

Barber PH, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proceedings of the Royal Society of London Series B: Biological Sciences 273: 2053–2061. doi: 10.1098/rspb.2006.3540

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology and Evolution 16: 314–321. doi: 10.1016/S0169-5347(01)02151-6

Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monagham MT, Balke M, Hendrich L, Geijer J, Herrmann J, Foster GN, Ribera I, Nilsson AN, Barraclough TG, Vogler AP (2012) The effect of geographic scale of sampling on DNA barcoding. Systematic Biology 61: 851–869. doi: 10.1093/sysbio/sys037

Bernt M, Braband A, Schierwater B, Stadler PF (2013) Genetic aspects of mitochondrial genome evolution. Molecular Phylogenetics and Evolution 69: 328–338. doi: 10.1016/j.ympev.2012.10.020

Blanco-Bercial L, Bradford-Grieve J, Bucklin A (2012) Molecular phylogeny of the Calanoida (Crustacea: Copepoda). Molecular Phylogenetics and Evolution 59: 103–113. doi: 10.1016/j.ympev.2011.01.008

Bowser AK, Diamond AW, Addison JA (2013) From puffins to plankton: a DNA-based analysis of a seabird food chain in the Northern Gulf of Maine. Public Library of Science ONE 8: e83152. doi: 10.1371/journal.pone.0083152

Bracken-Grissom HD, Felder DL, Vollmer NL, Martin JW, Crandall KA (2012) Phylogenetics links monster larvae to deep-sea shrimp. Ecology and Evolution 2: 2367–2373. doi: 10.1002/ece3.347

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine Metazoa. Annual Review of Marine Science 3: 471–508. doi: 10.1146/annurev-marine-120308-080950

Buhay JE (2009) "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. Journal of Crustacean Biology 29: 96–110. doi: 10.1651/08-3020.1

Burghart SE, Van Woudenberg L, Daniels CA, Meyers SD, Peebles EB, Breibart M (2014) Disparity between planktonic fish egg and larval communities as indicated by DNA barcoding. Marine Ecology Progress Series 503: 195–204. doi: 10.3354/meps10752

Butcher BA, Smith MA, Sharkey MJ, Quicke DLJ (2012) A turbo-taxonomic study of Thai *Aleiodes* (*Aleiodes*) and *Aleiodes* (*Arcaleiodes*) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. Zootaxa 3457: 1–232.

Cameron S, Rubinoff D, Will K (2006) Who will actually use DNA barcoding and what will it cost? Systematic Biology 55: 844–847. doi: 10.1080/10635150600960079

Campo D, Morales J, Garcia L, Fernandez-Rueda P, Garcia-Gonzalez C, Garcia-Vasquez E (2010) Phylogeography of the European stalked barnacles (*Pollicipes pollicipes*): identification of glacial refugia. Marine Biology 157: 147–156. doi: 10.1007/s00227-009-1305-z

Carr CM, Hardy SM, Brown TM, Macdonald TA, Hebert PDN (2011) A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. Public Library of Science ONE 6: e22232. doi: 10.1371/journal.pone.0022232

Caterino MS, Tishechkin AK (2006) DNA identification and morphological description of the first confirmed larvae of Hetaeriinae (Coleoptera: Histeridae). Systematic Entomology 31: 405–418. doi: 10.1111/j.1365-3113.2006.00338.x

Chapman AD (2009) Numbers of Living Species in Australia and the World. Australian Biological Resources Study, Canberra, 84 pp.

Collins RA, Cruickshank RH (2013) The seven deadly sins of DNA barcoding. Molecular Ecology Resources 13: 969–975. doi: 10.1111/1755-0998.12046

Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN (2007) Biological identifications through DNA barcodes: the case of the Crustacea. Canadian Journal of Fisheries and Aquatic Sciences 64: 272–295.

Cristescu MEA, Hebert PDN (2002) Phylogeny and adaptive radiation of the Onychopoda (Crustacea, Cladocera): evidence from multiple gene sequences. Journal of Evolutionary Biology 15: 838–849. doi: 10.1046/j.1420-9101.2002.00466.x

Dathe HH (2003) Lehrbuch der Speziellen Zoologie. Begründet von A. Kaestner. Band I. Wirbellose Tiere. 5. Teil: Insecta. Spektrum Akademischer Verlag, Heidelberg, 962 pp. [In German]

Dayrat B (2005) Towards integrative taxonomy. Biological Journal of the Linnean Society 85: 407–415. doi: 10.1111/j.1095-8312.2005.00503.x

de Carvalho MR, Bockmann FA, Amorim DS, Brandão CRF, de Vivo M, de Figueiredo JL, Britski HA, de Pinna MCC, Menezes NA, Marques FPL, Papavero N, Cancello EM, Crisci JV, McEachran JD, Schelly RC, Lundberg JG, Gill AC, Britz R, Wheeler QD, Stiassny MLJ, Parenti LR, Page LM, Wheeler WC, Faivovich J, Vari RP, Grande L, Humphries CJ, DeSalle R, Ebach MC, Nelson GJ (2007) Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. Evolutionary Biology 34: 140–143. doi: 10.1007/s11692-007-9011-6

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 360: 1905–1916. doi: 10.1098/rstb.2005.1722

Dincă V, Zakhharov EV, Hebert PDN, Vila R (2010) Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. Proceedings of the Royal Society of London Series B: Biological Sciences 278: 347–355. doi: 10.1098/rspb.2010.1089

Di Pinto A, Di Pinto P, Terio V, Bozzo G, Bonerba E, Ceci E, Tantillo G (2013) DNA barcoding for detecting market substitution in salted cod fillets and battered cod chunks. Food Chemistry 141: 1757–1762. doi: 10.1016/j.foodchem.2013.05.093

Ebach MC (2011) Taxonomy and the DNA barcoding enterprise. Zootaxa 2742: 67–68.

Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lambshead PJD, Thomas WK, Creer S (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nature Communications 1: 98. doi: 10.1038/ncomms1095

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Reviews in Ecology and Evolutionary Systematics 34: 397–423. doi: 10.1146/annurev.ecolsys.34.011802.132421

Garcia-Merchan VH, Robainas-Barcia A, Abello P, Macpherson E, Palero F, Garcia-Rodriguez M, de Sola LG, Pascual M (2012) Phylogeographic patterns of decapod crustaceans at the Atlantic-Mediterranean transition. Molecular Phylogenetics and Evolution 62: 664–672. doi: 10.1016/j.ympev.2011.11.009

Goldstein PZ, DeSalle R (2010) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. Bioessays 33: 135–147. doi: 10.1002/bies.201000036

Gruner H-E (1993) Lehrbuch der Speziellen Zoologie. Begründet von A. Kaestner. Band I. Wirbellose Tiere. 4. Teil: Arthropoda (ohne Insecta), Gustav Fischer Verlag, Jena, 1279 pp. [In German]

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. Public Library of Science ONE 6: e17497. doi: 10.1371/journal.pone.0017497

Hansson C, Smith MA, Janzen DH, Hallwachs W (2015) Integrative taxonomy of New World *Euplectrus* Westwood (Hymenoptera, Eulophidae), with focus on 55 new species from Area de Conservación Guanacaste, northwestern Costa Rica. ZooKeys 485: 1–236. doi: 10.3897/zookeys.485.9124

Hausmann A, Haszprunar G, Hebert PDN (2011) DNA barcoding the Geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions. Public Library of Science ONE 6: e17134. doi: 10.1371/journal.pone.0017134

Haye PA, Segovia NI, Vera R, Gallardo MA, Gallardo-Escárate C (2012) Authentication of commercialized crab-meat in Chile using DNA barcoding. Food Control 25: 239–244. doi: 10.1016/j.foodcont.2011.10.034

Hazakani-Covo E (2009) Mitochondrial insertions into primate nuclear genomes suggest the use of *numts* as a tool for phylogeny. Molecular Biology and Evolution 26: 2175–2179. doi: 10.1093/molbev/msp131

Hazakani-Covo E, Zelle RM, Martin W (2010) Molecular poltergeists: mitochondrial DNA copies (*numts*) in sequenced nuclear genomes. Public Library of Science Genetics 6: e1000834. doi: 10.1371/journal.pgen.1000834

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B: Biological Sciences 270: 313–321. doi: 10.1098/rspb.2002.2218

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London Series B: Biological Sciences 270 (Supplement): S96–S99. doi: 10.1098/rsbl.2003.0025

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Pro-

ceedings of the National Academy of Science of the United States of America 101: 14812–14817. doi: 10.1073/pnas.0406166101

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution 54: 291–301. doi: 10.1016/j.ympev.2009.09.016

Holmes BH, Steinke D, Ward RD (2009) Identification of shark and fin rays using DNA barcoding. Fisheries Research 95: 280–288. doi: 10.1016/j.fishres.2008.09.036

Ivanova NV, Clare EL, Borisenko A (2012) DNA barcoding in mammals. Methods in Molecular Biology 858: 153–182. doi: 10.1007/978-1-61779-591-6_8

Khalaji-Pirbalouty V, Raupach MJ (2014) A new species of *Cymodoce* Leach, 1814 (Crustacea: Isopoda: Sphaeromatidae) based on morphological and molecular data, with a key to the Northern Indian Ocean species. Zootaxa 3826: 230–254. doi: 10.11646/zootaxa.3826.1.7

Kim J-H, Antunes A, Suo J-L, Menninger J, Nash WG, O`Brien SJ, Johnson WE (2006) Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the *Panthera* genus species. Gene 366: 292–302. doi: 10.1016/j.gene.2005.08.023

Klaus S, Selvandran S, Goh JW, Wowor D, Brandis D, Koller P, Schubart CD, Streit B, Meier R, Ng PKL, Yeo DCJ (2013) Out of Borneo: neogene diversification of Sundaic freshwater crabs (Crustacea: Brachyura: Gecarcinucidae: *Parathelphusa*). Journal of Biogeography 40: 63–74. doi: 10.1111/j.1365-2699.2012.02771

Knebelsberger T, Landi M, Neumann H, Kloppmann M, Sell A, Campbell P, Laakmann S, Raupach MJ, Carvalho G, Costa FO (2014) A reliable DNA barcode reference library for the identification of the European shelf fish fauna. Molecular Ecology Resources 14: 1060–1071. doi: 10.1111/1755-0998.12238

Krebes L, Blank M, Jürss K, Zettler ML, Bastrop R (2010) Glacial-driven vicariance in the amphipod *Gammarus duebeni*. Molecular Phylogenetics and Evolution 54: 372–385. doi: 10.1016/j.ympev.2009.07.034

Lai JCY, Ng PKL, Davie PJF (2010) A revision of the *Portunus pelagicus* (Linnaeus, 1758) species complex (Crustacea: Brachyura: Portunidae), with the recognition of four species. The Raffles Bulletin of Zoology 58: 199–237.

Landry JF, Nazari V, Dewaard JR, Mutanen M, Lopez-Vaamonde C, Huemer P, Hebert PDN (2013) Shared but overlooked: 30 species of Holarctic Microlepidoptera revealed by DNA barcodes and morphology. Zootaxa 3749: 1–93. doi: 10.11646/zootaxa.3749.1.1

Layton KKS, Martel AL, Hebert PDN (2014) Patterns of DNA barcode variation in Canadian marine molluscs. Public Library of Science ONE 9: e95003. doi: 10.1371/journal.pone.0095003

Lefébure T, Douady CJ, Gouy M, Gibert J (2006) Relationships between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. Molecular Phylogenetics and Evolution 40: 435–447. doi: 10.1016/j.ympev.2006.03.014

Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proceedings of the National Academy of Science of the United States of America 112: 2076–2081. doi: 10.1073/pnas.1424997112

Lijtmaer DA, Kerr KCR, Barreira AS, Hebert PDN, Tubaro PL (2011) DNA barcode libraries provide insight into continental patterns of avian diversification. Public Library of Science ONE 6: e20744. doi: 10.1371/journal.pone.0020744

Lopez I, Erickson DL (2012) DNA Barcodes. Methods and Protocols. Humana Press, New York, 470 pp.

Matzen da Silva JS, Creer S, dos Santos A, Costa AC, Cunha MR, Costa FO, Carvalho GR (2011a) Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). Public Library of Science ONE 6: e19449. doi: 10.1371/journal.pone.0019449

Matzen da Silva JS, dos Santos A, Cunha MR, Costa FO, Creer S, Carvalho GR (2011b) Multigene molecular systematics confirm species status of morphologically convergent hermit crabs. Public Library of Science ONE 6: e28233. doi: 10.1371/journal.pone.0028233

McFadden CS, Benayahu Y, Pante E, Thoma JN, Nevarez A, France SC (2011) Limitations of mitochondrial gene barcoding in Octocorallia. Molecular Ecology Resources 11: 19–31. doi: 10.1111/j.1755-0998.2010.02875.x

Nicolè S, Negrisola E, Eccher G, Mantovani R, Patarnello T, Erickson DL, Kress WJ, Barcaccia G (2012) DNA barcoding as a reliable method for the authentication of commercial seafood products. Food Technology and Biotechnology 50: 387–398.

Olson RR, Runstadler JA, Kocher TD (1991) Whose larvae? Nature 351: 357–358. doi: 10.1038/351357b0

Padial JM, Miralles A, de la Riva I, Vences M (2010) The integrative future of taxonomy. Frontiers in Zoology 7: 16. doi: 10.1186/1742-9994-7-16

Page TJ, Choy SC, Hughes JM (2005) The taxonomic feedback loop: symbiosis of morphology and molecules. Biology Letters 1: 139–142. doi: 10.1098/rsbl.2005.0298

Page TJ, Hughes JM (2011) Neither molecular nor morphological data have all answers; with an example from *Macrobrachium* (Decapoda: Palaemonidae) from Australia. Zootaxa 2874: 65–68.

Pamilo P, Viljakainen L, Viljakainen A (2007) Exceptionally high density of NUMTs in the honeybee genome. Molecular Biology and Evolution 24: 1340–1346. doi: 10.1093/molbev/msm055

Pons J, Vogler AP (2005) Complex pattern of coalescence and fast evolution of a mitochondrial rRNA pseudoegene in a recent radition of tiger beetles. Molecular Biology and Evolution 22: 991–1000. doi: 10.1093/molbev/msi085

Prendini L (2005) Comment on 'Identifying spiders through DNA barcoding'. Canadian Journal of Zoology 83: 498–504. doi: 10.1139/Z05

QGIS Development Team (2015) QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355–364. doi: 10.1111/j.1471-8286.2007.01678.x

Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. Public Library of Science ONE 8: e66213. doi: 10.1371/journal.pone.0066213

Raupach MJ, Hendrich L, Küchler SM, Deister F, Morinière J, Gossner MM (2014) Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. Public Library of Science ONE 9: e106940. doi: 10.1371/journal.pone.0106940

Raupach MJ, Barco A, Steinke D, Beermann J, Laakmann S, Mohrbeck I, Neumann H, Kihara TC, Pointner K, Radulovici A, Segelken-Voigt A, Wesse C, Knebelsberger T (2015) The application of DNA barcodes for the identification of marine crustaceans from the North Sea and adjacent regions. Public Library of Science ONE 10: e0139421. doi: 10.1371/journal.pone.0139421

Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M (2013) One hundred and one new species of *Trigonopterus* weevils from New Guinea. ZooKeys 280: 1–150. doi: 10.3897/zookeys.280.3906

Rougerie R, Kitching IJ, Haxaire J, Miller SE, Hausmann A, Hebert PDN (2014) Australian Sphingidae – DNA barcodes challenge current species boundaries and distributions. Public Library of Science ONE 9: e101108. doi: 10.1371/journal.pone.0101108

Ruiz C, May-Itzá WdJ, Quezada-Euán JJG, de la Rúa C (2013) Presence of nuclear copies of mitochondrial origin (NUMTs) in two related species of stingless bee genus *Melipona* (Hymenoptera: Meliponini). Journal of Zoological Systematic and Evolutionary Research 51: 107–113. doi: 10.1111/jzs.12011

Santamaria CA, Mateos M, Taiti S, DeWitt TJ, Hurtado LA (2013) A complex evolutionary history in a remote archipelago: phylogeography and morphometrics of the Hawaiian endemic *Ligia* isopods. Public Library of Science ONE 8: e85199. doi: 10.1371/journal.pone.0085199

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. Annual Review of Entomology 55: 421–438. doi: 10.1146/annurev-ento-112408-085432

Schneider-Broussard R, Neigel JE (1997) A large-subunit mitochondrial ribosomal DNA sequence translocated to the nuclear genome of two stone crabs (*Menippe*). Molecular Biology and Evolution 14: 156–165.

Schubart CD, Diesel R, Hedges SB (1998) Rapid evolution to terrestrial life in Jamaican crabs. Nature 393: 363–365. doi: 10.1038/30724

Schubart CD (2009) Mitochondrial DNA and decapod phylogenies: the importance of pseudogenes and primer optimization. In: Martin JW, Crandall KA, Felder DS. Decapod Crustacean Phylogenetics. Crustacean Issues 18: 47–66.

Shank TM, Lutz RA, Vrijenhoek RC (2006) Molecular systematics of shrimp (Decapoda: Bresiliidae) from deep-sea hydrothermal vents, I: enigmatic "small orange" shrimp from the Mid-Atlantic Ridge are juvenile *Rimicaris exoculata*. Molecular Marine Biology and Biotechnology 7: 88–96.

Shearer TL, van Oppen MJH, Romano SL, Wörheide G (2002) Slow mitochondrial DNA sequence evolution in Anthozoa (Cnidaria). Molecular Ecology 11: 2475–2487. doi: 10.1046/j.1365-294X.2002.01652.x

Shearer TL, Coffroth MA (2008) Barcoding corals: limited by interspecific divergence, not intraspecific variation. Molecular Ecology Resources 8: 247–255. doi: 10.1111/j.1471-8286.2007.01996.x

Shen Y-Y, Che X, Murphy RW (2013) Assessing DNA barcoding as a tool for species identification and data quality control. Public Library of Science ONE 8: e57125. doi: 10.1371/journal.pone.0057125

Shih H-T, Cai Y (2007) Two new species of the land-locked freshwater shrimps genus, *Neocaridina* Kubo, 1938 (Decapoda: Caridea: Atyidae), from Taiwan, with notes on speciation on the island. Zoological Studies 46: 680–694.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. Molecular Ecology 21: 1794–1805. doi: 10.1111/j.1365-294X.2012.05538.x

Sinniger F, Reimer JD, Pawlowski J (2008) Potential of DNA sequences to identify zoanthids (Cnidaria: Zoantharia). Zoologica Scripta 25: 1253–1260 doi: 10.2108/zsj.25.1253

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the National Academy of Sciences of the United States of America 105: 13486–13491. doi: 10.1073/pnas.0803076105

Song H, Moulton MJ, Whiting MF (2014) Rampant nuclear insertion of mtDNA across diverse lineages within Orthoptera (Insecta). Public Library of Science ONE 9: e110508. doi: 10.1371/journal.pone.0110508

Soto-Calderón ID, Clark NJ, Wildschutte JVH, DiMattio K, Jensen-Seaman MI, Anthony NM (2014) Identification of species-specific nuclear insertions of mitochondrial DNA (numts) in gorillas and their potential use as population genetic markers. Molecular Phylogenetics and Evolution 81: 61–70. doi: 10.1016/j.ympev.2014.08.018

Strutzenberger P, Brehm G, Fiedler K (2011) DNA barcoding-based species delimitation increases species count of *Eois* (Geometridae) moths in a well-studied tropical mountain forest up to 50%. Insect Science 18: 349–362. doi: 10.1111/j.1744-7917.2010.01366.x

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends in Ecology and Evolution 18: 70–74. doi: 10.1016/S0169-5347(02)00041-1

Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. Molecular Ecology Resources 12: 377–388. doi: 10.1111/j.1755-0998.2012.03119.x

Thalmann O, Serre D, Hofreiter M, Lukas D, Erikson J, Vigilant L (2005) Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. Molecular Ecology 14: 179–188. doi: 10.1111/j.1365-294X.2004.02382.x

Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Orlando L, Willerslev E (2012) Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology 21: 2565–2573. doi: 10.1111/j.1365-294X.2011.05418.x

Torres AP, Palero F, dos Santos A, Abelló P, Blanco E, Boné A, Guerao G (2014) Larval stages of the deep-sea lobster *Polycheles typhlops* (Decapoda, Polychelidae) identified by DNA analysis: morphology, systematic, distribution and ecology. Helgoland Marine Research 68: 379–397. doi: 10.1007/s10152-014-0397-0

Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. Trends in Ecology and Evolution 24: 110–117. doi: 10.1016/j.tree.2008.09.011

Vogler AP, Monagham MT (2007) Recent advances in DNA taxonomy. Journal of Systematic and Evolutionary Research 45: 1–10. doi: 10.1111/j.1439-0469.2006.00384.x

Webb KE, Barnes DKA, Clark MS, Bowden DA (2006) DNA barcoding: a molecular tool to identify Antarctic marine larvae. Deep-Sea Research II: Topical Studies in Oceanography 53: 1053–1060. doi: 10.1016/j.dsr2.2006.02.013

Weigt LA, Baldwin CC, Driskell A, Smith DG, Ormos A, Reyier EA (2012) Using DNA barcoding to assess Caribbean reef fish biodiversity: expanding taxonomic and geographic coverage. Public Library of Science ONE 7: e41059. doi: 10.1371/journal.pone.0041059

Weis A, Meyer R, Dietz L, Dömel JS, Leese F, Melzer RR (2014) *Pallenopsis patagonica* (Hoek, 1881) – a species complex revealed by morphology and DNA barcoding, with description of a new species of *Pallenopsis* Wilson, 1881. Zoological Journal of the Linnean Society 170: 110–131. doi: 10.1111/zoj.12097

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20: 47–55. doi: 10.1111/j.1096-0031.2003.00008.x

Will KP, Mishler PD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology 54: 844–851. doi: 10.1080/10635150500354878

Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. Molecular Biology and Evolution 18: 1484–1493. doi: 10.1093/oxfordjournals.molbev.a003934

Witt JDS, Threloff DL, Hebert PDN (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Molecular Ecology 15: 3073–3082. doi: 10.1111/j.1365-294X.2006.02999.x

Woodcock TS, Boyle EE, Roughley RE, Kevan PG, Labbee RN, Smith ABT, Goulet H, Steinke D, Adamowicz SJ (2013) The diversity and biogeography of the Coleoptera of Churchill: insights from DNA barcoding. BMC Ecology 13: 40. doi: 10.1186/1472-6785-13-40

Zhou X, Robinson JL, Geraci CJ, Parker CR, Flint Jr OS, Etnier DA, Ruiter D, DeWalt RE, Jacobus LM, Hebert PDN (2011) Accelerated construction of a regional DNA-barcode reference library: caddisflies (Trichoptera) in the Great Smoky Mountains National Park. Journal of the North American Benthological Society 30: 131–162. doi: 10.1899/10-010.1

Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. GigaScience 2: 4. doi: 10.1186/2047-217X-2-4

Supplementary material 1

A decade of DNA barcoding of crustaceans: input file
Authors: Michael J. Raupach, Adriana Radulovici
Data type: data table
Explanation note: Raw data related to 164 publications on crustacean barcoding as retrieved from Web of Science: bibliography, citations, habitat type, geographical area, BOLD use.