**Mediterranean Marine Science**

Vol 16, No 1 (2015)

Vol 16, No 1 (2015)


# DNA barcoding of twelve shrimp species (Crustacea: Decapoda) from Turkish seas reveals cryptic diversity

*R. BILGIN, M. A. UTKAN, E. KALKAN, S. U. KARHAN, M. BEKBÖLET*

doi: 10.12681/mms.548

Research Article

*Mediterranean Marine Science*
Indexed in WoS (Web of Science, ISI Thomson) and SCOPUS
The journal is available on line at http://www.medit-mar-sc.net
DOI: http://dx.doi.org/10.12681/mms.548

# DNA barcoding of twelve shrimp species (Crustacea: Decapoda) from Turkish seas reveals cryptic diversity

**R. BILGIN[1], M.A. UTKAN[1], E. KALKAN[1], S.Ü. KARHAN[2] and M. BEKBOLET[1]**

[1] Institute of Environmental Sciences, Boğaziçi University, 34342 Bebek, Istanbul, Turkey
[2] Division of Hydrobiology, Department of Biology, Faculty of Science, Istanbul University, 34134 Vezneciler, Istanbul, Turkey

Corresponding author: rasit.bilgin@boun.edu.tr

Handling Editor: Marco Oliverio

Received: 19 July 2013; Accepted: 23 May 2014; Published on line: 30 October 2014

**Abstract**

DNA barcoding is a useful tool for the identification and potential discovery of new species. In this study, DNA barcoding was employed by sequencing the mitochondrial cytochrome oxidase subunit I gene (COI) to characterize the genetic diversity of 12 shrimp species inhabiting Turkish coastal waters and, when possible, to compare with the genetic data available from different parts of the Mediterranean and eastern Atlantic. This study also comprises the first DNA barcoding study performed in the Turkish Seas using COI. A total of 40 shrimp specimens were collected and analyzed from 9 sites. Generally, the barcoding gap criterion was successful in identifying species; hence, COI appeared to be a good marker of choice for DNA barcoding in this group. Out of the 12 species investigated, five were barcoded for the first time. For six species, two intraspecific clades were retrieved after the analyses. The results suggest the presence of cryptic diversity in a genetically understudied marine area, the Turkish coastal waters, and further investigation of these species using population genetics, taxonomic approaches and nuclear markers is likely to result in designation of new species.

**Keywords:** DNA barcoding, cytochrome oxidase subunit I, species delimitation, Decapoda, shrimp, Turkey.

## Introduction

Recording the current state of biological diversity through surveying and categorizing species is important for understanding the human impact on terrestrial and marine ecosystems, and for exploring the continuous change in biodiversity, at any geographical scale. Species identification by morphological approaches can require a high degree of specialized experience, which is difficult and impractical for those interested in surveying a broad diversity of organisms. As an alternative to morphological approaches, microgenomic systems can help to rapidly classify biological diversity through the analysis of small segments of the genome, and represent a practical approach to the analysis of biological diversity (Hebert *et al.*, 2003; Radulovici *et al.*, 2010).

Hebert *et al.* (2003) proposed the use of cytochrome c oxidase subunit I (COI) as a standard diagnostic region to help identify species, define species boundaries and aid in species delimitation, as a part of the DNA Barcoding framework. The method involves building a reference database (the Barcode of Life Database, BOLD) where data about the specimens (e.g. photographic, geographic, taxonomic including locations of the voucher specimens) are combined with molecular data (Hebert *et al*., 2013). Subsequently, sequenced DNA barcodes from unknown specimens can be compared against this reference library to identify them. The method also makes it possible to reinforce classical taxonomy, assist with the discovery of new species and characterization of the taxonomic and genetic diversity of different geographic regions, and help resolve cryptic species complexes (Hebert & Gregory, 2005; Costa *et al.*, 2007; Hajibabaei *et al.*, 2007). More than four hundred DNA barcoding papers published since 2003 (Taylor & Harris, 2012), and more than 2 million barcodes deposited to BOLD as of January 2014 testifies to the popularity of the method, though not without criticism. Among others, these criticisms include the questionable utility of the method in certain groups (e.g. amphibians, Vences *et al*., 2005), problems with the use of distance-based methods in species identification (DeSalle *et al.*, 2005; Kelly *et al*., 2007), skewed species coverage (Kvist, 2013), and the current inability to incorporate next generation sequencing methods into its work flow (Taylor & Harris, 2012).

Genetic barcoding has previously been employed as a molecular tool to reinforce classical taxonomy in the Decapoda. For instance, Pardo *et al.* (2009) determined diagnostic morphological characters of three sympatric crab species of the genus *Cancer* by comparing COI gene sequences of larvae and adults identified by morphology. Jones & Macpherson (2007) examined galatheid squat

lobsters of the genus *Munidopsis* in the Eastern Pacific using morphological characters and mitochondrial COI sequences. They described seven new *Munidopsis* species from the region that revealed high gene flow among distant conspecific populations, and they also confirmed the existence of closely-related sibling species in the genus. In other studies, DNA barcoding was used as an identification tool. To illustrate, Shih *et al.* (2009) used genetic barcodes to distinguish six species of fiddler crabs of the genus *Uca* in the Indian Ocean. In another study, Radulovici *et al.* (2009) barcoded 87 crustacean species, most of which were shrimps, and the study showed the utility of the barcoding approach in diagnosing marine crustaceans. Genetic barcoding can also be particularly useful for identifying marine invertebrate juveniles and larvae, which are difficult to distinguish morphologically. This is true for many crustacean decapods, which are clearly distinguishable as adults, but whose larval and juvenile forms are difficult to identify at species level, and DNA barcoding is a good tool for species identification of such forms (Bucklin *et al.*, 2010). These studies suggest that genetic barcoding can be an effective method for identifying and resolving relationships within this group of organisms.

Despite the species mentioned above and more, it should be noted that around 14,750 extant decapod species have been described (De Grave *et al.*, 2009), but only 17.3 % of these have COI barcodes assigned (Matzen da Silva *et al.*, 2011; http://www.boldsystems.org/index.php/TaxBrowser_TaxonPage?taxid=336). Comprising about a quarter of the described species (De Grave *et al.*, 2009), shrimps represent a significant portion of the decapod diversity. In this study, our aims were both to add new barcoding data to BOLD, and to see how our morphological identification compared against subsequent specimen identification using BOLD.

The barcoding approach was used to characterize the genetic diversity of 12 shrimp species (Decapoda: Dendrobranchiata, Stenopodidea, Caridea), which are amongst the most common and widespread shrimps inhabiting Turkish coastal waters, making it the first COI DNA barcoding study in Turkey for any group of marine invertebrates. Turkey is bordered on three sides by water, which includes the Black Sea, the Sea of Marmara, the Aegean and the Levantine Seas (Fig. 1). Although a diverse marine system, the Turkish coasts have been relatively understudied compared to the rest of the Mediterranean Sea (Patarnello *et al.*, 2007). Using 40 new sequences combined with sequence data available in BOLD and GenBank, we also made phylogeographic comparisons when possible.

## Materials and Methods

### *Fieldwork and Species Identification*

Samples were collected from the Black Sea, the Sea of Marmara and the Mediterranean (Aegean and Levantine coasts) (Fig. 1) during marine biodiversity surveys between June 2009 and October 2010, using various methods including bottom trawls, SCUBA, and free diving. A total of 40 shrimp specimens belonging to 12 species (Table 1) were collected and analyzed from 9 locations, including three from the Black Sea (Hopa, Persembe, Kilyos), three from the Sea of Marmara (two on Burgazada and one on the Dardanelles), two from the Aegean (Ayvalik and Bodrum), and one from the Levantine coast (Kaş) (Table 1, Fig. 1). After collection, samples were preserved in 95% ethanol and are deposited at the Molecular Biology Laboratory, Institute of Environmental Sciences, Bogazici University, Turkey. The species were identified under a stereomicroscope (Olympus, SZ61) using the following literature: dendrobranchiate shrimps were identified according to Zariquiey Álvarez (1968) and Holthuis (1987), stenopodidean shrimps according to Zariquiey Álvarez (1968), and caridean shrimps according to Holthuis (1949; 1987), Nouvel & Holthuis (1957), Lewinsohn & Holthuis (1964), Zariquiey Álvarez (1968), D'Udekem d'Acoz & Wirtz (2002), and González-Ortegón & Cuesta (2006).

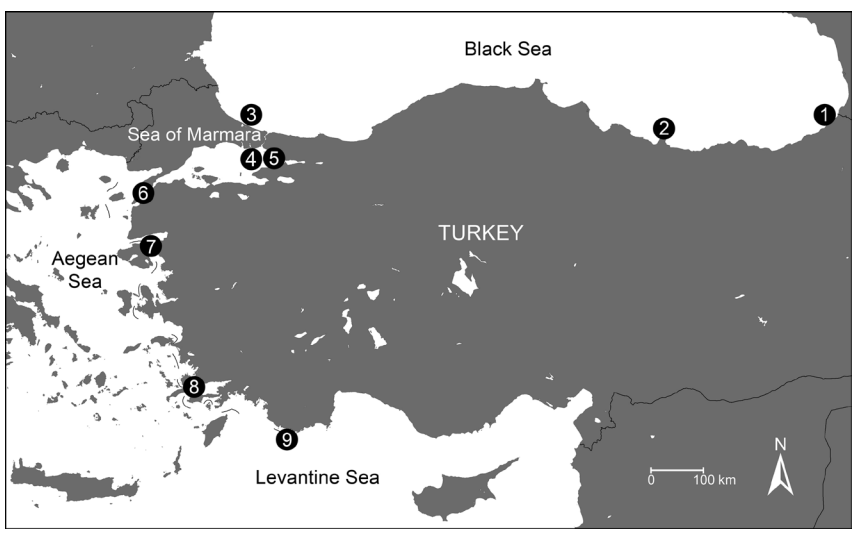


***Fig. 1:*** Map showing sampling locations. Numbers indicate the codes of the sampling sites as given in Table 1.

**Table 1.** List of sampling sites, geographic coordinates of the individuals sampled, sampling method, and sample size (N).

| Code (Map ref.) | Sampling Site | Geographic Coordinates | Sampling Method | N |
|---|---|---|---|---|
| **1** | SE Black Sea, Artvin, Hopa, Kemalpaşa fishermen's harbor | 41°28'58.56"N, 41°31'14.85"E | SCUBA diving | 1 |
| **2** | SE Black Sea, Ordu, Perşembe, Cape Yasun | 41°07'46.82"N, 37°40'31.59"E | Free diving | 1 |
| **3** | SW Black Sea, Istanbul, off Kilyos | Transect between 41°16'51.68"N, 29°01'11.24"E and 41°17'49.50"N, 28°56'59.54"E | Bottom trawl | 4 |
| **4** | NE Sea of Marmara, Prince's Islands, off Burgazada | Transect between 40°51'57.88"N, 29°4'01.03"E and 40°52'19.83"N, 29°02'58.59"E | Bottom trawl | 3 |
| **5** | NE Sea of Marmara, Prince's Islands, Burgazada coast | 40°52'39.77"N, 29°03'07.31"E | Free diving | 2 |
| **6** | SW Sea of Marmara, Dardanelles, off Yapıldak | 40°13'51.84"N, 26°32'04.78"E | SCUBA diving | 5 |
| **7** | NE Aegean Sea, Ayvalık, Cunda Island, Ortunç Cove | 36º20'01.68"N, 26º37'08.42"E | Free diving | 13 |
| **8** | SE Aegean Sea, Bodrum, Yalıçiftlik coast | 36°59'15.24"N, 27°32'55.93"E | SCUBA diving | 1 |
| **9** | Levantine coast, Antalya, Kaş, Hidayet Cove | 36°11'22.58"N, 29°36'34.24"E | SCUBA diving | 10 |

### *Laboratory Protocols*

For DNA extraction, a Roche DNA Extraction Kit (Mannheim, Germany) was used following the manufacturer's protocol with one modification. This included the first step of the protocol, where the incubation time of sample tissue with lysis buffer and proteinase K was increased to 48 hours, to enhance the lysis of the cell membranes.

PCR amplification of the COI gene was performed using three primer sets. These included the "Folmer primers" LCOI490: 5'- GGTCAACAAATCATAAAGATATTGG-3' and HCO2198: 5'- TAAACTTCAGGGTGACCAAA AAATCA-3' (Folmer *et al.*, 1994), and two sets where we used HCO2198 as the reverse primer combined with one of the two forward primers: CRUSTF1: 5'-TTTTCTACAAATCATAAAGACATTGG-3', and CRUSTF2:5'-GGTTCTTCTCCACCAACCACAARGAYATHGG-3' (Costa *et al.*, 2007). The primers set(s) that worked for amplification of different individuals in each species (given in Table 2). For the LCOI490-HCO2198 primer set, 1 µl of DNA was added to a 49 µl reaction mixture, containing 5 µl 10x high fidelity buffer, 5 µl $MgCl_2$ (25 mM), 1 µl of 10 mM dNTPs, 1.5 µl of each primer (20 µM), 34.7 µl $H_2O$ and 0.3 µl Taq DNA polymerase. Cycling parameters consisted of an initial denaturation step of 5 min at 94°C, followed by 35 cycles of 1 min at 95°C, 1 min at 45°C and 1.5 min at 72°C, with a final extension step of 7 min at 72°C. For CRUSTF1-HCO2198 and CRUSTF2-HCO2198 primer sets, the amount of DNA, $MgCl_2$ and primers were changed as follows: 1.5 µl of DNA, 4.4 µl $MgCl_2$ (25 mM), 0.5 µl of each primer (20 µM) and 36.8 µl of $H_2O$. For these primer sets, cycling parameters consisted of an initial denaturation step of 5 min at 94°C followed by 35 cycles of 1 min at 95°C, 1 min at 43°C and 1.5 min at 72°C, with a final extension step of 7 min at 72°C.

The PCR products were purified using a Roche PCR cleanup kit (Mannheim, Germany), and later were commercially sequenced by Macrogen Inc. (South Korea). The obtained sequences were cleaned manually with Sequencher v. 4.1 (Gene Codes Corp.) and aligned using the same program. These sequences were deposited in GenBank (Accession Nos: KJ841671 – KJ841710) and the Consortium for the Barcode of Life Database (BOLD) (Sample Ids: BUESS001 – BUESS040).

### *Analytical methods*

For the 40 new sequences, intraspecific and interspecific distances were calculated in the MEGA v. 5 (Kumar *et al.*, 2008) using Kimura 2-parameter distances. DnaSP v. 4 (Rozas *et al.*, 2003) was used to build the haplotype data files and TCS 1.21 (Clement *et al.*, 2000) was used to prepare haplotype networks. We also built neighbour-joining (NJ) trees with the sequences for each species using MEGA v. 5. The clades in these species corresponded to haplogroups in haplotype networks; therefore, we only provide the figures for the latter, as we also present the trees based on the 'identify specimen' option in BOLD for each species.

The 40 new sequences obtained were identified to species level by comparing the sequences under the 'identify specimen' option of BOLD (Ratnasingham & Hebert,

2007), and we specifically used the 'All Barcode Records on BOLD' database. This method constructs neighbour-joining trees (referred to as 'BOLD trees' from here onwards) and use Kimura 2-parameter distances to assign an unidentified specimen to the nearest species. The database was accessed between March 2011 and January 2014. BOLD trees can only be constructed using a single unknown sequence at a time. Based on our own NJ trees, for species that formed a single clade, only one individual was compared whereas for the species which formed two or more different clades, one sequence from each clade was compared (see Results & Discussion for clade definitions). Although there are over 2,000,000 sequences in BOLD available for building trees within the BOLD database, some of these sequences are not directly available for download and some locality information is not publicly available. For only one species, *Palaemon elegans*, it was possible to obtain barcode sequences from BOLD, and the locality information (GPS coordinates) was provided courtesy of Joanna Matzen, making it possible to build haplotype networks with geographic information of BOLD data for this species. For five species (*Parapenaeus longirostris*, *Palaemon adspersus*, *P. serratus*, *Eualus cranchii*, and *Crangon crangon*), although the DNA barcodes were not public, locality information was coarsely available from BOLD, generally at the level of country of origin, which was used in conjunction with the BOLD trees, making it possible to undertake some phylogeographic inferences. For the rest of the species for which neither DNA barcodes nor locality information was present in BOLD, only the 'identify specimen' option was used.

## Results and Discussion

### *The barcoding gap*

Forty new COI sequences from 12 different crustacean decapod species found in Turkish coastal waters were generated for this study. With 25 sequences obtained from BOLD/GenBank (Table 2), a total of 65 samples were analyzed. The new sequences that we obtained do not likely represent pseudogenes, as there were no double peaks in the chromatograms, no stop-codons in the amino acid translations, and no indels in the alignments. Comparing the 40 new sequences, the mean intraspecific divergence was 1.3%. The mean interspecific distance was 21.03%. A plot of the frequency of intraspecific and interspecific distances is presented in Figure 2. The minimum intraspecific distance was 0.73% for *Palaemon adspersus*, and the maximum was 2.75% for *Crangon crangon*. The minimum interspecific distance was 12.6% between *Palaemon serratus* and *P. adspersus*, and the maximum was 26.8% between *Synalpheus gambarelloides* and *P. adspersus*.

DNA barcoding is considered to be a suitable method for differentiating species in a group when interspecific variation exceeds intraspecific variation by one order of magnitude, which is called the 'barcoding gap' (Wiemers & Fiedler, 2007). The mean interspecific distance at 21% is more than an order of magnitude larger than the mean intraspecific divergence of 1.3%, demonstrating that the species investigated in this study are suitable for identification through DNA barcoding.

### *Performance of BOLD in species identification and clustering at higher levels*

The performance of BOLD in identifying the species we analyzed, and the clustering of the species with conspecifics and congenerics was variable. Our analysis with *Processa edulis*, collected from the Levantine coast, indicates that our sequence is the first DNA barcode in BOLD. The sample that is barcoded from Turkey fell within the *Processa clade* (Fig. 3A). Our sequence for *Gnathophyllum elegans* also clustered with another *G. elegans* barcode (Fig. 3B). In addition to *P. edulis* and *G. elegans*, for six other species the barcodes we generated fell within the correct clade for that species. These species were *Parapenaeus longirostris* (98.72% match with BOLD), *Palaemon adspersus* (99.67%), *P. elegans* (100%), P. serratus (89.7%), *Eualus cranchii* (85.3%), *Sicyonia carinata* (99.33%), and *Crangon crangon* (98.16%). For these eight species, the COI marker works well for DNA barcoding, with minimal gaps in the database.

For the other species, the BOLD trees did not provide the expected results. For instance, our sequence for *Processa acutirostris*, collected from the Levantine coast, represents the first DNA barcode in BOLD. However, while the highest similarity of this sequence was with *Processa modica* (83.03%), the BOLD tree did not place it within the *Processa* clade (Fig. 3E), which suggests a requirement for further investigation. Another such species is *Synalpheus gambarelloides*. The sample for this species, which was collected from the Levantine coast, fell within a complex clade that seems to be extensively paraphyletic. A general look indicates general paraphyly in the genus *Synalpheus* in the BOLD trees, and the same was true for the sequence

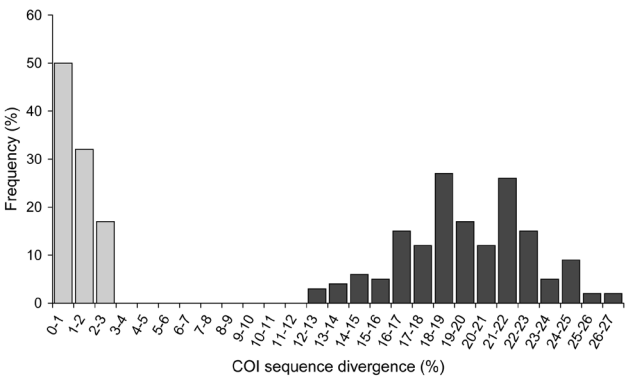


***Fig. 2:*** Frequency distribution of mean divergences for COI sequences for 40 samples, calculated by the Kimura 2-parameter model. Two taxonomic levels are represented: within species (grey bars) and between species (black bars).

**Table 2.** Systematic list of species investigated, number of new sequences from each species (N), number of sequences obtained from BOLD/GenBank, names of primers used, and sites at which the samples collected.

| Species | N | Bold / GenBank | Primers | Sampling Sites (Map ref.) |
|---|---|---|---|---|
| Suborder DENDROBRANCHIATA | | | | |
| Superfamily PENAEOIDEA | | | | |
| Family PENAEIDAE | | | | |
| ***Parapenaeus longirostris*** (Lucas, 1846) | 3 | 6 | CRUSTF1-LCOI490 | 4 |
| Family SICYONIIDAE | | | | |
| ***Sicyonia carinata*** (Brünnich, 1768) | 2 | – | HCO2198-LCOI490 | 9 |
| Suborder PLEOCYEMATA | | | | |
| Infraorder STENOPODIDEA | | | | |
| Family STENOPODIDAE | | | | |
| ***Stenopus spinosus*** Risso, 1827 | 3 | – | CRUSTF1-LCOI490 | 9 |
| Infraorder CARIDEA | | | | |
| Superfamily PALAEMONOIDEA | | | | |
| Family GNATHOPHYLLIDAE | | | | |
| ***Gnathophyllum elegans*** (Risso, 1816) | 1 | – | HCO2198-LCOI490 | 8 |
| Family PALAEMONIDAE | | | | |
| ***Palaemon adspersus*** Rathke, 1837 | 10 | – | HCO2198-LCOI490<br>CRUSTF1-LCOI490 | 7 |
| ***Palaemon elegans*** Rathke, 1837 | 12 | 9 | CRUSTF2-LCOI490<br>HCO2198-LCOI490 | 1, 2, 5–7, 9 |
| ***Palaemon serratus*** (Pennant, 1777) | 1 | 5 | HCO2198-LCOI490 | 7 |
| Superfamily ALPHEOIDEA | | | | |
| Family ALPHEIDAE | | | | |
| ***Synalpheus gambarelloides*** (Nardo, 1847) | 1 | – | CRUSTF1-LCOI490 | 9 |
| Family HIPPOLYTIDAE | | | | |
| ***Eualus cranchii*** (Leach, 1817) | 1 | 3 | HCO2198-LCOI490 | 9 |
| Superfamily PROCESSOIDEA | | | | |
| Family PROCESSIDAE | | | | |
| ***Processa acutirostris*** Nouvel & Holthuis, 1957 | 1 | – | HCO2198-LCOI490 | 9 |
| ***Processa edulis*** (Risso, 1816) | 1 | – | CRUSTF2-LCOI490 | 9 |
| Superfamily CRANGONOIDEA | | | | |
| Family CRANGONIDAE | | | | |
| ***Crangon crangon*** (Linnaeus, 1758) | 4 | 2 | HCO2198-LCOI490<br>CRUSTF2-LCOI490<br>CRUSTF1-LCOI490 | 3 |

we generated for our *S. gambarelloides* samples (Fig. 3C). The problem associated with *Synalpheus* seemed to extend to other species in different genera as well. *Periclimenaeus schmitti* of the family Palaemonidae being in the same clade as *Processa edulis* and two other congeneric species of the family Processidae suggests a potential mistake in the BOLD for *Periclimenaeus schmitti* (Fig. 3A). The same is true for a *Synalpheus* species (family Alpheidae), which fell into the same clade with species from the genus *Gnathophyllum* (family Gnathophyllidae) (Fig. 3B). To give another example, *Stenopus spinosus* was seen to be basal to a clade that included species from six different genera, rather than with *Stenopus zanzibaricus* or *S. hispidus*, which were the second and third closest matches to the species, respectively (Fig. 3D). The *S. zanzibaricus* samples in BOLD being from French Polynesia, and *S. hispidus* being from French Polynesia and the western coast of US, probably also contributed to this discrepancy. Paraphyletic relationships or mtDNA introgression could explain these patterns; however, taxonomic misidentifications in the BOLD database are also likely to be another potential source of these misplacements.

### *Cryptic diversity as revealed by BOLD comparisons*

In six species for which there were at least four samples, a phylogeographic pattern of two clades/haplogroups was observed. This suggests the possibility of cryptic species and/or current or historical genetic isolation of populations within species. These six species had data available from BOLD with locality information from the Mediterranean and/or Atlantic, which made it possible to make more global comparisons. These species were *Parapenaeus longirostris, Palaemon adspersus, P. elegans, P. serratus, Eualus cranchii,* and *Crangon crangon*.

In species for which geographic data were available, we were able to conduct more in-depth analyses. The analysis of Palaemon elegans sequence data from Turkey showed the presence of two groups, differentiated

by 45 bp (Fig. 4A). One group (marked in grey) comprised three haplotypes, H1–H3, and another haplotype (in white), H4. A tree constructed using H1 and H4 using the 'identify specimen' option in BOLD, as sequence data were not available from BOLD for direct comparison, also showed that the two haplotypes fell into two different clades, shown in the respective grey and white colours in Fig. 4B. The geographic distribution of the individuals belonging to these clades is shown in Fig. 4C. The individuals belonging to the 'grey clade' were found in Portugal (Azores), Poland and around most of Turkey (including the Black Sea, the Sea of Marmara, the Aegean and the

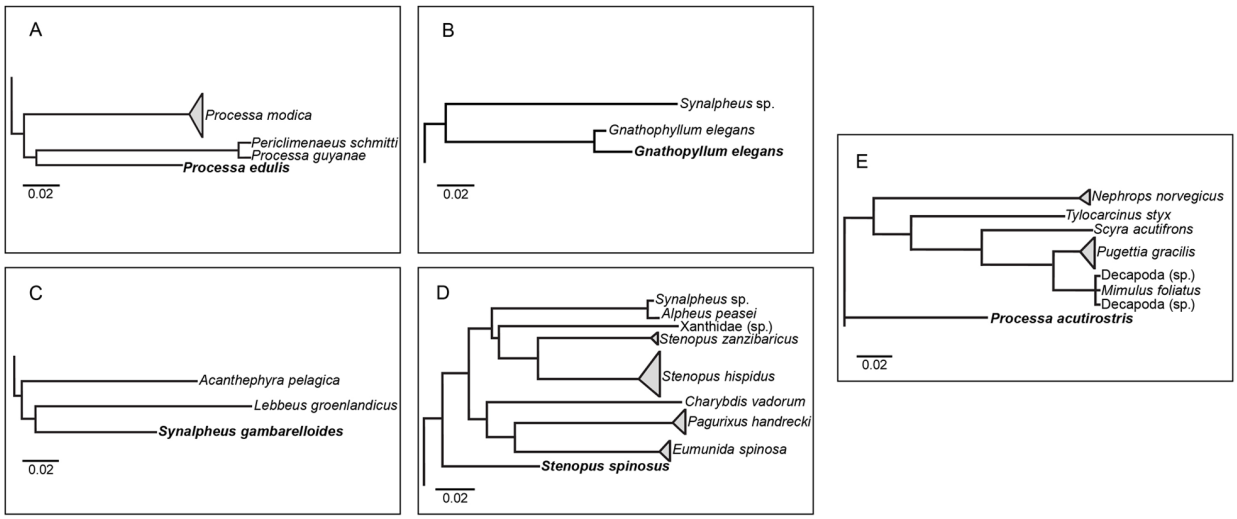


***Fig. 3:*** Sections of BOLD trees focusing on the species of interest: A) *Processa edulis*, B) *Gnathophyllum elegans*, C) *Synalpheus gambarelloides*, D) *Stenopus spinosus*, and E) *Processa acutirostris*. Our samples are indicated in bold.

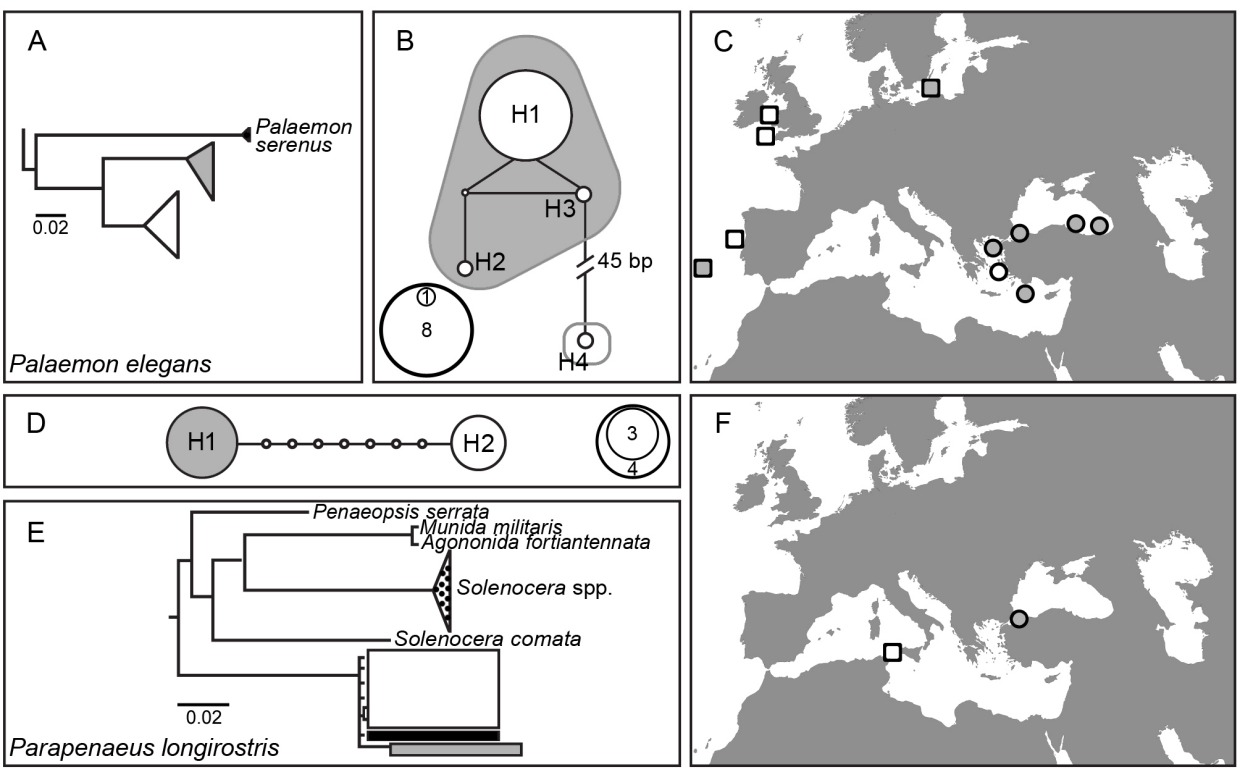


***Fig. 4:*** For *Palaemon elegans*: A) Section of BOLD tree focusing on *Palaemon elegans*, showing two clades in grey and white. B) Haplotype network – grey and white shades correspond to the clades in Fig. 4A. C) Clade distribution map – grey and white circles correspond to the distribution of the grey and white clades, respectively, in the corresponding BOLD tree and the haplotype network. The squares designate data retrieved from BOLD. - For *Parapenaeus longirostris*: D) Haplotype network - grey and white shades correspond to the clades in Fig. 4E. E) Section of BOLD tree focusing on *Parapenaeus longirostris*, showing the two clades in grey and white. F) Clade distribution map – grey and white coloured circles correspond to the distribution of the grey and white clades in the corresponding BOLD tree and the haplotype network. The square designates data retrieved from BOLD.

Levantine coasts), whereas the 'white clade' was found in the United Kingdom, Portugal (the Atlantic coast) and Turkey (the Aegean coast). Reuschel *et al.* (2010) investigated the degree of genetic variation within the widely distributed *P. elegans*, using COI sequences obtained from the Atlantic, the Baltic Sea, the Mediterranean, the Black Sea and the Caspian Sea. Three main haplogroups were identified in their study: one from the Atlantic (Type I) and two from the Mediterranean (Types II and III). The authors interpreted these differences as suggesting potential cryptic speciation. The COI sequences in this study could not be compared with the sequences in that study, as the sequences were not made available on GenBank. Our COI sequences also formed two groups, with one group consisting of samples from all around Turkey, and the second group consisting of one individual (# 38) from the Aegean. Comparing the first group with BOLD, the barcodes from Turkey clustered closely with barcodes in BOLD from Poland. On the other hand, individual #38 fell within the opposite branch (with barcodes from United Kingdom and Portugal). Hence, our comparative results support the idea of two cryptic species in *P. elegans*, as suggested by Reuschel *et al*. (2010).

Parapenaeus longirostris was the only species in which sequence data were publicly available from BOLD. These three identical sequences also included locality information; they were from Italy (Fig. 4F). The network constructed for this species showed two haplotypes, H1 and H2, from Italy and Turkey, respectively, differentiated by eight bp (Fig. 4D). The tree constructed using H2 and 'identify specimen' in BOLD (Fig. 4E) showed that the individuals with H1 (marked at the tips of the tree in grey), another individual without locality information and sequence data (marked in black) and the H2 haplotype generated in this study (marked in white) formed a single clade. The Turkish sample with the H2 haplotype had a longer branch leading to it than the other samples, reflecting the eight bp differentiation; however, the pattern here suggests that this differentiation could be due to isolation-by-distance.

In *Palaemon adspersus*, the haplotype network indicated the presence of two groups in Turkey, each composed of two haplotypes (Fig. 5A). There was a network of two groups (grey and white, Fig. 5A), separated from each other by seven bp. This species, as mentioned above, is a new entry to BOLD and hence no sequence data were publicly available to build a network with sequences that were not from Turkey. When a representative sequence from each group was used under the 'identify specimen' option of BOLD, each sample fell on a different branch, sister to another unidentified bar-

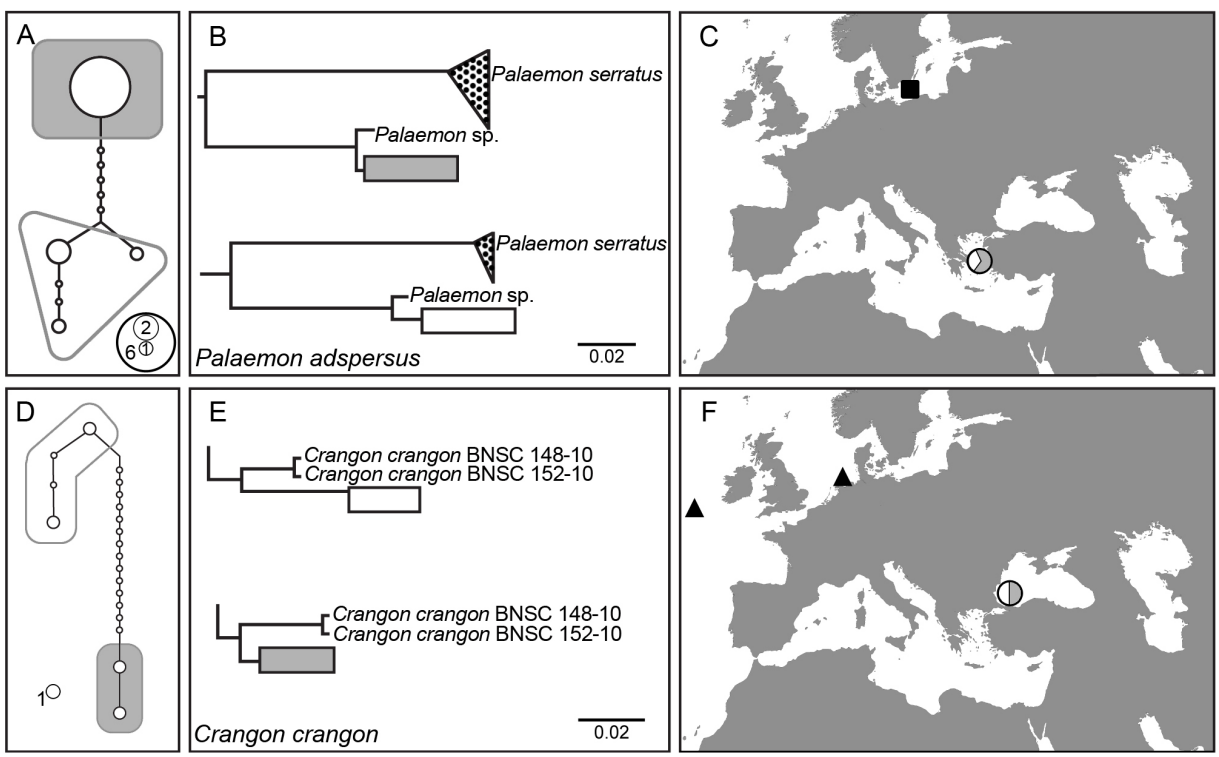


***Fig. 5:*** For *Palaemon adspersus*: A) Haplotype network - grey and white shades correspond to the clades in Fig. 5B. B) Section of BOLD tree focusing on *Palaemon adserpsus*, showing the two intraspecific clades in grey and white. C) Clade distribution map – grey and white circles correspond to the distribution of the grey and white clades, respectively in the corresponding BOLD tree and the haplotype network. The black square designates data retrieved from BOLD.- For *Crangon crangon*: D) Haplotype network - grey and white shades correspond to the clades in Fig. 5E. E) Section of BOLD trees focusing on *Crangon crangon*, showing the two intraspecific clades in grey and white. F) Clade distribution map – grey and white circles correspond to the distribution of the grey and white clades, respectively, with the pie-chart showing the relative distribution of each clade in Turkey. The triangles indicate the distribution of the other clade [composed of individuals coded *Crangon crangon* BNSC148-10 (the BOLD process-ids) and *Crangon crangon* BNSC152-10 in Fig. 5E.]

code from the genus *Palaemon* (*Palaemon* spp., Fig 5B). A comparison with BOLD suggests that a specifically unidentified barcode from BOLD, *Palaemon* sp. (Process-id: FCDPBAS02D) could actually be *P. adspersus*. The representative *P. adspersus* sample that was barcoded from Turkey fell within the *Palaemon* clade, and formed a clade together with this *Palaemon* sp. from Poland, BOLD.

In *Crangon crangon*, the pattern was similar to that in *Palaemon adspersus*. There were no public sequences available in BOLD, hence the network included only haplotypes from Turkey. These two groups (grey and white) were differentiated from each other by 15 bp (Fig. 5D). All of the Turkish individuals were collected from the pre-Bosphoric region of the Black Sea, close to the northern end of the Bosphorus Strait. When a representative haplotype from each group was used with the 'identify specimen' option of the BOLD tree, both of these haplotypes formed a sister clade to two samples from BOLD (Fig. 5E), barcoded from the North Atlantic and off the coast of Germany, respectively (Fig. 5A). A previous study using COI by Luttikhuizen *et al*. (2008) showed a total of four clades in the Atlantic, the Mediterranean, the Adriatic Sea and the Black Sea. The sequences in this study might correspond to the Mediterranean or the Black Sea group; however, we were not able to make a direct comparison with these sequences, as it was not possible to align them with our barcodes. In their study, Luttikhuizen *et al*. (2008) mentioned some double peaks they encountered in their sequences, although they argue against the possibility of nuclear mitochondrial insertions (NUMTs). Our own sequences aligned with those in BOLD and also with the unpublished *C. crangon* COI sequences of C. Schubart (pers. comm.); hence, the mismatch between these data and that of Luttikhuizen *et al*. (2008) suggest a need for further investigation. For four species we were able to prepare either a network or a BOLD tree. For *Eualus cranchii* (referred to as *Thoralus cranchii* in BOLD, synonym of *E. cranchii*) and *Palaemon serratus*, one sample was analyzed for each; hence, a network could not be constructed. In these two species, the BOLD trees showed the presence of two clades. In *E. cranchii* the specimen that we barcoded from the Mediterranean coast of Turkey (in grey) fell onto a clade that was basal to another clade (Fig. 6A) that included *E. cranchii* barcodes in BOLD (from United Kingdom, in white) and *Acanthephyra pelagica* (family Acanthephyridae) (Fig. 6B). Similarly, the sample of *P. serratus* from Turkey (Fig. 6C, in grey) formed a sister lineage to *P. serratus* samples from the United Kingdom and Portugal (Fig. 6D, in white). These results suggest a potential history of isolation between the populations of these species in the Atlantic and eastern Mediterranean, and the Aegean Sea.

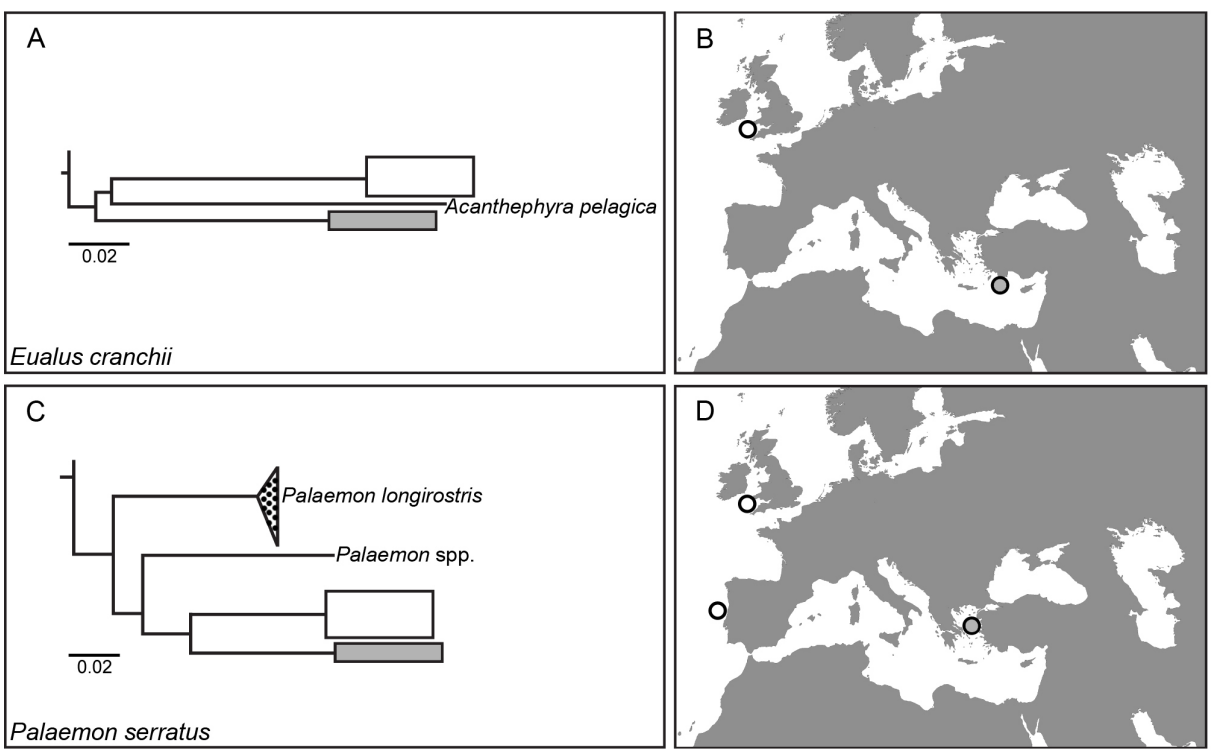


***Fig. 6:*** For *Eualus cranchii*: A) Section of BOLD tree focusing on *Eualus cranchii*, showing the two intraspecific clades in grey and white. B) Clade distribution map – grey and white circles correspond to the distribution of the grey and white clades, respectively in the corresponding BOLD tree. The data for the white clade were retrieved from BOLD. – For *Palaemon serratus* C) Section of BOLD tree focusing on *Palaemon serratus*, showing the two intraspecific clades in grey and white. D) Clade distribution map – grey and white circles correspond to the distribution of the grey and white clades, respectively in the corresponding BOLD tree. The data for the white clade were retrieved from BOLD.

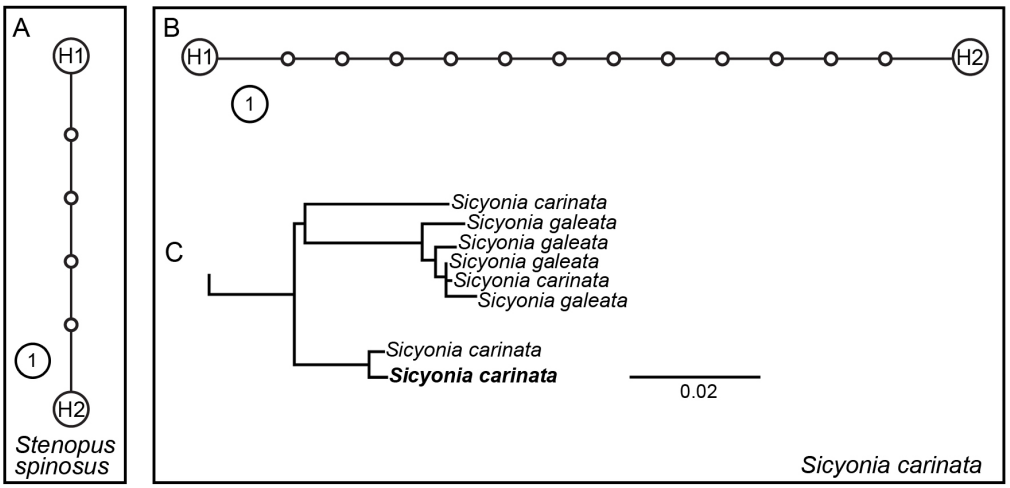


***Fig. 7:*** The haplotype network for A) *Stenopus spinosus* and B) *Sicyonia carinata* and C) the section of BOLD tree focusing on *Sicyonia carinata*. Empty circles represent hypothetical haplotypes. Our sample is indicated in bold in the BOLD tree.

In two species, *Stenopus spinosus* and *Sicyonia carinata*, as two individuals were barcoded for each, and these individuals had different sequences, it was possible to calculate haplotype networks. All of these four samples were collected from the Levantine coast. In both species two haplotypes were found, which were differentiated from each other by five and 13 bp in *Stenopus spinosus* (Fig. 7A) and *Sicyonia carinata* (Fig. 7B), respectively. Using BOLD, our *Sicyonia carinata* clustered with the sequence available in the database, showing congruence and correct identification (Fig. 7C). No subspecies are currently recognized for these two species, and hence the differentiation observed in this study could indicate cryptic speciation thus suggesting that an evaluation of the subspecific/species status may be informative. However, it should be noted that these two species had two individuals sequenced, each, and that the observed differences between the haplotypes could be indicative of a single population with high genetic diversity.

## Conclusion

The barcoding gap worked well for the 12 crustacean decapod species examined in this study, and expands the current decapod reference set through the generation of novel sequences. For five species, the unique COI barcodes generated represent the first submissions to the BOLD database for these species. From a phylogeographic perspective, for six out of the 12 species investigated, barcoding has distinguished the presence of two clades. The six species are good candidates for discovery of cryptic species or subspecies. Previous research has shown the presence of more than one clade in Turkish coastal waters in crustacean decapods (Luttikhuizen *et al.*, 2008; Reuschel *et al.*, 2010), bivalves (Kalkan & Bilgin, under review) and fish (Magoulas *et al.*, 2006), indicating the presence of a potential marine genetic break in this region. More extensive phylogeographic sampling and analyses both in Turkish coastal waters, and when possible in the Mediterranean and the Atlantic, especially using nuclear markers will be useful for taxonomic clarification in these taxa. In this study, DNA barcoding has been an effective tool for identifying known species and new clades to be examined further, which is a first step towards one of the main objectives of DNA barcoding: facilitating the discovery of new species (Hajibabaei *et al.*, 2007).

## Acknowledgements

We would like to thank Elizabeth Hemond, Baki Yokeş and Volkan Demir for their help with the fieldwork, and Elizabeth Hemond, Baki Yokeş and two anonymous reviewers for providing comments on an earlier version of the manuscript. We also would like to thank Joana Matzen for providing the locality information for the *Palaemon elegans* sequences from BOLD. This study was supported by a grant (No: 09S101) from the Research Fund of Boğaziçi University in Istanbul to RB.

## References

Bucklin, A., Steinke, D., Blanco-Bercial, L., 2010. DNA barcoding of marine metazoa. *Annual Review of Marine Science*, 3, 471-508.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657-1659.

Costa, F.O., Boutillier, J., Ratnasingham, S., Dooh, R.T., Hajibabaei, M., *et al.*, 2007. Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, 64, 272-295.

De Grave, S., Pentcheff, N.D., Ahyong, S., Chan, T., Crall, K.A., *et al.*, 2009. A classification of living and fossil genera of decapod crustaceans. *Raffles Bulletin of Zoology*, 21, 1-109.

DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: Taxonomy, species delimitation, and DNA barcoding. *Philosophical Transactions of the Royal Society of London B*, 360, 1905-1916.

D'Udekem d'Acoz, C., Wirtz, P., 2002. Observations on some interesting coastal Crustacea Decapoda from the Azores, with a key to the genus *Eualus* Thallwitz, 1892 in the Northeastern Atlantic and the Mediterranean. *Arquipélago. Ciências biológicas e marinhas/Life and marine sciences*, 19A, 67-84.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, 294-299.

González-Ortegón, E., Cuesta, J.A., 2006. An illustrated key to species of *Palaemon* and *Palaemonetes* (Crustacea: Decapoda: Caridea) from European waters, including the alien species *Palaemon macrodactylus*. *Journal of the Marine Biological Association of the United Kingdom,* 86, 93-102.

Hajibabaei, M., Singer, G.A., Hebert, P.D., Hickey, D.A., 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics,* 23, 167-172.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 270, 313-321.

Hebert, P.D.N., Gregory, R.T., 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology*, 54, 852-859.

Hebert, P.D.N., deWaard, J.R., Zakharov, E.V., Prosser, S.W.J., Sones, J.E., *et al*., 2013. A DNA 'barcode blitz': rapid digitization and sequencing of a natural history collection. *PLoS ONE*, 8(7), e68535.

Holthuis, L.B., 1949. The caridean Crustacea of the Canary Islands. *Zoologische Mededelingen,* 30, 227-255.

Holthuis, L.B., 1987. Vrais Crabes. In: *Fiches FAO d'identification des espèces pour les besoins de la pêche Méditerranée et Mer Noire Zone de pêche 37 Végetaux et invertébrés* (eds. Fisher W, Schneider M, Bauchot ML). FAO, Rome, 760 pp.

Jones, W.J., Macpherson, E., 2007. Molecular phylogeny of the East Pacific squat lobsters of the genus *Munidopsis* (Decapoda: Galatheidae) with the descriptions of seven new species. *Journal of Crustacean Biology,* 27, 477-501.

Kelly, R. P., Sarkar, I. N., Eernisse, D. J., DeSalle, R., 2007. DNA barcoding using chitons (genus *Mopalia*). *Molecular Ecology Notes*, 7, 177-183.

Kumar, S., Nei, M., Dudley, J., Tamura, K., 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics,* 9, 299-306.

Kvist, S., 2013. Barcoding in the dark?: a critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Molecular Phylogenetics and Evolution*, 69, 39-45.

Lewinsohn, C., Holthuis, L.B., 1964. New records of decapod Crustacea from the Mediterranean coast of Israel and the eastern Mediterranean. *Zoologische Mededelingen,* 40, 45-63.

Luttikhuizen, P.C., Campos, J., Bleijswijk, J., Peijnenburg, K.T.C.A., van der Veer, H.W., 2008. Phylogeography of the common shrimp, *Crangon crangon* (L.) across its distribution range. *Molecular Phylogenetics and Evolution,* 46, 1015-1030.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S., Patarnello, T., 2006. Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular Phylogenetics and Evolution,* 39, 734-746.

Matzen da Silva, J., Creer, S., Dos Santos, A., Costa, A.C., Cunha, M.R., *et al*., 2011. Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS One,* 6, e19449.

Nouvel, H., Holthuis, L.B., 1957. Les Processidae (Crustacea: Decapoda: Natantia) des eaux européennes. *Zoologische Verhandelingen,* 32, 1-53.

Pardo, L.M., Ampuero, D., Véliz, D., 2009. Using morphological and molecular tools to identify megalopae larvae collected in the field: the case of sympatric *Cancer* crabs. *Journal of the Marine Biological Association of the United Kingdom,* 89, 481-490.

Patarnello, T., Volckaert, F.A.M.J., Castilho, R., 2007. Pillars of Hercules: is the Atlantic- Mediterranean transition a phylogeographical break? *Molecular Ecology,* 16, 4426-4444.

Radulovici, A.E., Sainte-Marie, B., Dufresne, F., 2009. DNA barcoding of marine crustaceans from the Estuary and Gulf of St Lawrence: a regional-scale approach. *Molecular Ecology Resources,* 9, 181-187.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, 7, 355-364.

Reuschel, S., Cuesta, J.A., Schubart, C.D., 2010. Marine biogeographic boundaries and human introduction along the European coast revealed by phylogeography of the prawn *Palaemon elegans*. *Molecular Phylogenetics and Evolution*, 55, 765-775.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19, 2496-2497.

Shih, H.T., Kamrani, E., Davie, P.J.F., Liu, M.Y., 2009. Genetic evidence for the recognition of two fiddler crabs, *Uca iranica* and *U. albimana* (Crustacea: Brachyura: Ocypodidae), from the northwestern Indian Ocean, with notes on the *U. lactea* species-complex. *Hydrobiologia*, 635, 373-382.

Taylor, H., Harris, W., 2012. An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, 3, 377-388.

Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, 4, 16.

Vences, M., Thomas, M., Bonett, R.M., Vieites, D.R. 2005. Deciphering amphibian diversity through DNA barcoding: Chances and challenges. *Philosophical Transactions of the Royal Society of London B*, 360, 1859-1868.

Zariquiey Álvarez, R., 1968. Crustáceos Decápodos Ibéricos. *Investigación Pesquera,* 32, 1-510.