Journal of Aquatic Ecology

Journal homepage: http://jae.hormozgan.ac.ir

Identification of Engraulids (*Encrasicholina punctifer* and *E. pseudoheteroloba*) using the COI gene as a DNA barcoding marker

Mahboobeh Afrand[1], Iman Sourinejad[1,4*], Seyed Abolhassan Shahzadeh Fazeli[2,3], Arash Akbarzadeh[1,4]

[1] *Department of Fisheries, Faculty of Marine Science and Technology, University of Hormozgan, Bandar Abbas, Iran.*

[2] *Molecular Bank, Iranian Biological Resource Center, ACECR, Tehran, Iran.*

[3] *Department of Molecular and Cellular Biology, Faculty of Basic Sciences and Advanced Technologies in Biology, University of Science and Culture, ACECR, Tehran, Iran.*

[4] *Department of Modern Technologies, Mangrove Forest Research Center, Bandar Abbas, Iran.*

Keywords:

DNA Barcoding
Cytochrome oxidase subunit1 gene
Encrasicholina
Persian Gulf

Abstract

The genus *Encrasicholina* includes economically and ecologically important pelagic small sized anchovies. These fish are endemic to Indo-Pacific waters and are also found in the Persian Gulf. Due to harmful human activities in different forms, the ecosystem of Persian Gulf has suffered great damages in recent decades and its biodiversity is threatened. Therefore, its sustainable development and the preservation of its biodiversity is essential. The first step to achieve this goal is to identify the accurate species biodiversity. DNA barcoding method based on mitochondrial DNA is widely applied to identify the phylogenetic relationships between the species. In this study, we examined the utility of mitochondrial cytochrome oxidase subunit I (COI) for identification of Encrasicholina genus in the Persian Gulf. Sequences sent to the "Barcod of Life Data systems" and NCBI confirmed the morphological identification results. In fact, the larvae that their morphological characteristics were insufficient to identify, were validly identified with COI gene sequences. The intra-species genetic distance based on K2P was % 0.00, while the average interspecific distance was % 18.2. Phylogenetic trees showed that the studied Anchovies were in two sister clades. The results of this study suggested that the COI gene has the ability to distinguish the fish species. The results of molecular analyses were in accordance with the results of morphology.

1. Introduction

The genus *Encrasicholina* (family: Engraulidae) includes six species of pelagic small sized (less than 10 cm standard length) anchovies fish inhabiting marine and estuarine waters in the Indo-Pacific region (Whitehead *et al*., 1988; Hata and Motomura, 2016). Three species of genus *Encrasicholina* including *E. punctifer*, *E. heteroloba* and *E. devisi* have been

* **Corresponding Author, E-mail**: sourinejad@hormozgan.ac.ir

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

reported from Persian Gulf (Owfi *et al*., 2010). According to the recent classification, *E. pseudoheteroloba* (Hardenberg 1933) is recognized as a valid species, while it was previously regarded as a junior synonym of *E. heteroloba*. *E. heteroloba* is also regarded as a senior synonym of *E. devisi* (Hata and Motomura, 2017).

As these fishes have the schooling behavior, they are easily caught by nets (Cole and McGlade, 1998). They are one of the most important commercially caught fishes in Iranian waters of Persian Gulf. The species *E. punctifer* has been reported as the most dominant species in Hormozgan waters (Van Zaling *et al*., 1993; Salarpouri *et al*., 2008). Persian Gulf is under tremendous pressure from hydrocarbon pollution, heavy constructions, offshore activities, severe drought caused by construction of dams, and changes associated with global warming. These factors are threatening the biodiversity of this unique ecosystem and would cause species extinction crisis (Khan, 2007; Sheppard *et al*., 2010). Among marine organisms, the pelagic species including Anchovies are more affected by environmental changes in oceans and seas (Garza Gil *et al*., 2011).

Species identification and correct classification are very crucial for the resolution of many biological issues, such as diversity discovery, bio-security assurance, species conservation and pandemics prevention (Dayrat, 2005; Frézal and Leblois 2008). Some species in the genus *Encrasicholina* are distributed in a same wide geographical range. This makes them difficult to discriminate (Whitehead *et al*., 1988). Moreover, identifying species in different stages of development is difficult due to changes in morphology and may lead to misidentification (Ko *et al*., 2013). Furthermore, conditions such as the small size, soft body and the quick spoilage after catch make it difficult to distinguish them based on morphological features.

DNA sequence analysis has been used to help species identifications. Different sequences have been used for different taxonomic groups and in different laboratories (Ward *et al*., 2005). Hebert *et al*. (2003) proposed the term "DNA barcode" for accurate and reliable identification of species across a broad range of taxa by using a standard sequence that corresponds to a single sufficient homologous gene region which can be amplified by a polymerase chain reaction (PCR) with "universal primers. A mitochondrial DNA genome known as cytochrome oxidase subunit I (CO1) has been selected as a global bioidentification system for the vast majority of animal species. DNA barcoding method is a powerful tool for identifying larval forms or even fish eggs (Ko *et al*., 2013; Baldwin *et al*., 2011). For synchronization of the data collection of samples and doing analysis with barcode data, the website Barcode of Life Data Systems (BOLD) (http://www.boldsystems.org) has been created (Ratnasingham and Hebert, 2007). BOLD is an available database that helps in management, analysis, publication and searching of DNA barcodes (Zhang and Hanner 2011).

No documented molecular study has been carried out to validate the morphological identification of the species of genus *Encrasicholina* in the Persian Gulf. This study aimed to identify the genetic characteristics of *Encrasicholina* sp. using DNA barcoding based on CO1

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

gene as well as to compare the DNA sequences of the species of *Encrasicholina* in this study with the available sequences in NCBI and BOLD.

2. Materials and methods

2-1. Sample collection and morphological identification

The samples were collected from Bandar Aftab (26° 43' 37" N, 53° 56' 47" E) and Qeshm Island (26° 53' 59" N, 56°10' 9" E) in the Persian Gulf in May and December 2015. All specimens were fixed in 96% ethanol and transferred to the laboratory. Then, they were phenotypically identified according to FAO species identification sheets (Whitehead *et al*., 1988) and other references (Randall, 1995; Bianchi, 1985; Carpenter *et al*., 1997; Hata and Motomura, 2017) . The species were identified by the presence of a short isthmus muscle not reaching anteriorly to the posterior margin of the gill membrane, an exposed urohyal, prepelvic scutes, dorsal-fin rays (unbranched), anal-fin rays (unbranched), and gill rakers on 1st gill arch (lower).

2-2. DNA extraction, amplification, and sequencing

DNA was isolated from muscle tissue of six individuals of *E. punctifer*, *E. pseudoheteroloba* and larvae using the Blood and Cultured Cell Genomic DNA Extraction Kit (cat, NO. MBK0017) according to the Iranian Biological Resources Center's instructions. Using a primer pair LCO1490 (5'GGTCAACAAATCATAAAGATATTGG3') and HCO2198 (5'TAAACTTCAGGGTGACCAAAAAATCA3') a 658 bp fragment of the COI gene was amplified (Folmer *et al*. 1994). The PCR thermal regime consisted of 37.80 µL of ultrapure water, 2 µL of 10× PCR buffer, 3 µl MgCl2 (25 mM), 0.7 µL dNTP (10 mM), 0.5 µL of each primer (10 p mol), 0. 5 µL of Taq polymerase (5 U/µl) and about 100-ng template DNA (the final volume= 50 µL).DNA samples were subjected to the following PCR conditions (Hebert *et al*. 2003): one cycle of 3 min at 94 °C; five cycles of 40 s at 94 °C, 1.5 min at 45°C and 1 min at 72°C; 35 cycles of 40 s at 94°C, 1.5 min at 51°C and 1 min at 72°C, a final cycle of 5 min at 72°C. The PCR products were electrophoresed on 1% agarose gel for band characterization. The most intense products were selected and then purified with PCR Purification Kit (Cat, NO. MBK0017) following the Iranian Biological Resources Center's instructions. The purified PCR products were sequenced.

2-3. Data analysis

The obtained sequences from two strands were assembled and manually corrected using the SeqMan software (DNAStar). All sample sequences were identified through a BLAST search of the NCBI (National Center for Biotechnology Information) and BOLD (Barcode of Life) databases (Ratnasingham and Hebert, 2007). For comparison and validation, additional COI sequences of *Encrasicholina* species were downloaded from GenBank and BOLD and were analyzed together with the 15 sequences of *Encrasicholina* generated from this study. Sequence divergences were calculated using the Kimura two parameter (K2P) distance model (Kimura, 1980) and phylogenetic analysis of Neighbor Joining (NJ) and Maximum

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

Likelihood (ML) methods with Kimura 2-parameter evolution model and 1000x bootstrap replications were created in MEGA 6.0 software (Tamura *et al*., 2013). *Nematalosa nasus* and *Anodontostoma chacunda* (family: Clupeidae) were used as outgroups.

3. Results

According to the morphological features, anchovies sampled from Qeshm Island were identified as *E. punctifer* and *E. pseudoheteroloba*. The samples from Bandar Aftab were in larval stages, therefore we could not exactly identify them morphologically.

BOLD's search results showed probability placement percentage of 99.83% and 98.16 identification for *E. punctifer* and *E. pseudoheteroloba*, respectively (Fig. 1). Moreover, according to the blast results, the sequences of the Anchovies obtained from Bandar Aftab were identified as *E. pseudoheteroloba*. The COI species database tree was compared for similarity and relationships between *E. punctifer* (Fig. 2) and *E. pseudoheteroloba* (Fig. 3). Consensus sequences of 584 bp were used for the barcoding analysis.

Intra specific distance in both studied species was 0.00 and the intraspecies distance (shown in Table 1) ranged from 0.000 to 0.02 while interspecific variation ranged from 0.179 to 0.189. The highest intraspecies genetic distance (0.02) was detected within *E. pseudoheteroloba.*

The NJ and ML trees based on the DNA sequences obtained in the present study and five sequences derived from the database were constructed using the K2P model. Species shared similar topologies in the NJ and ML trees. The ML tree clearly distinguished two species (Fig. 4). Sequences belonging to one species can be easily identified as joint clusters with support (>50 bootstrap in ML and NJ). Outgroup used was segregated in separate clade as expected.

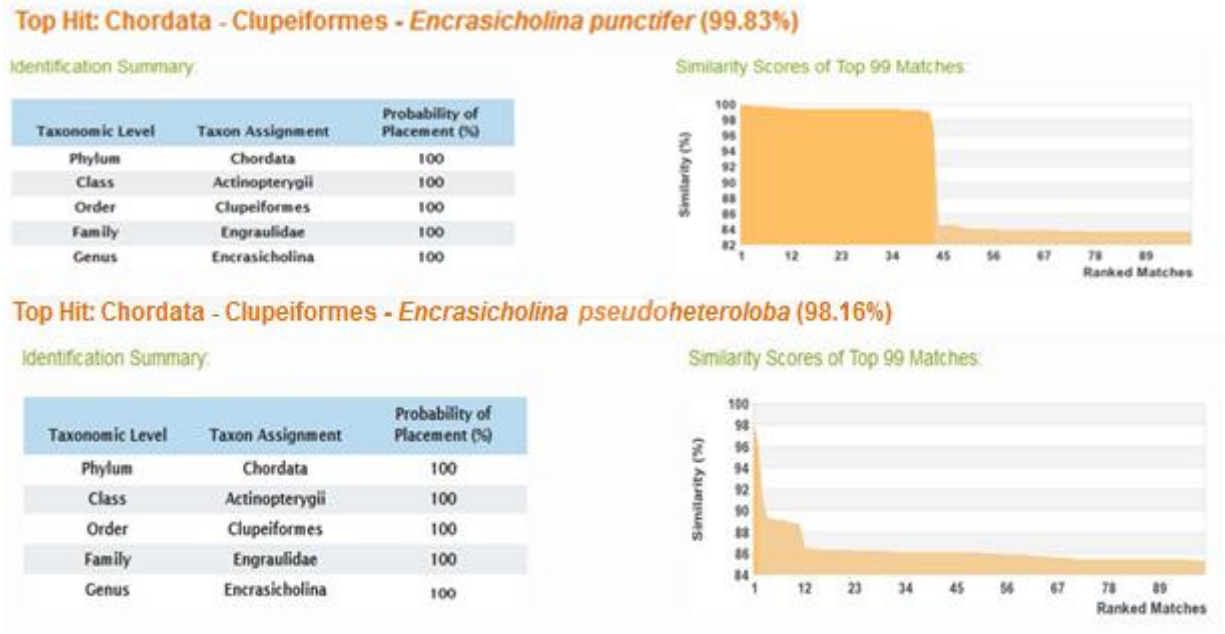

Fig. 1. BOLD's search for identification of *E. punctifer* and *E. pseudoheteroloba*

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

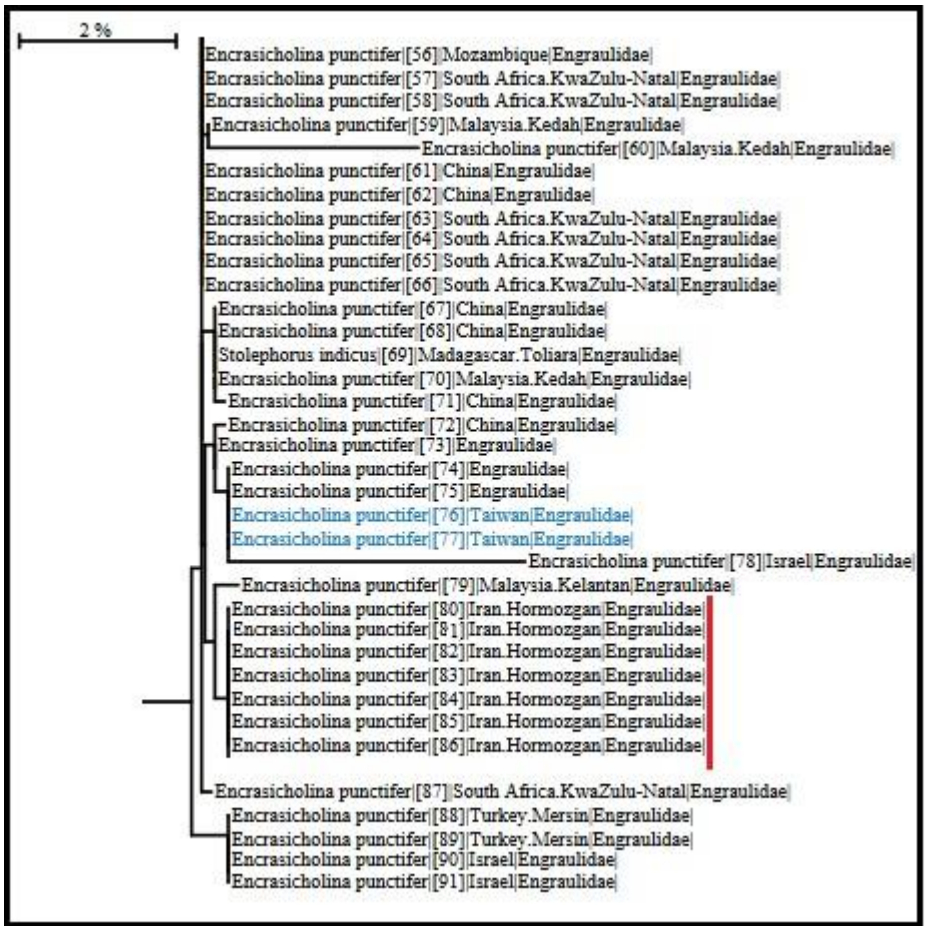

Fig. 2. COI database tree for *E. punctifer*

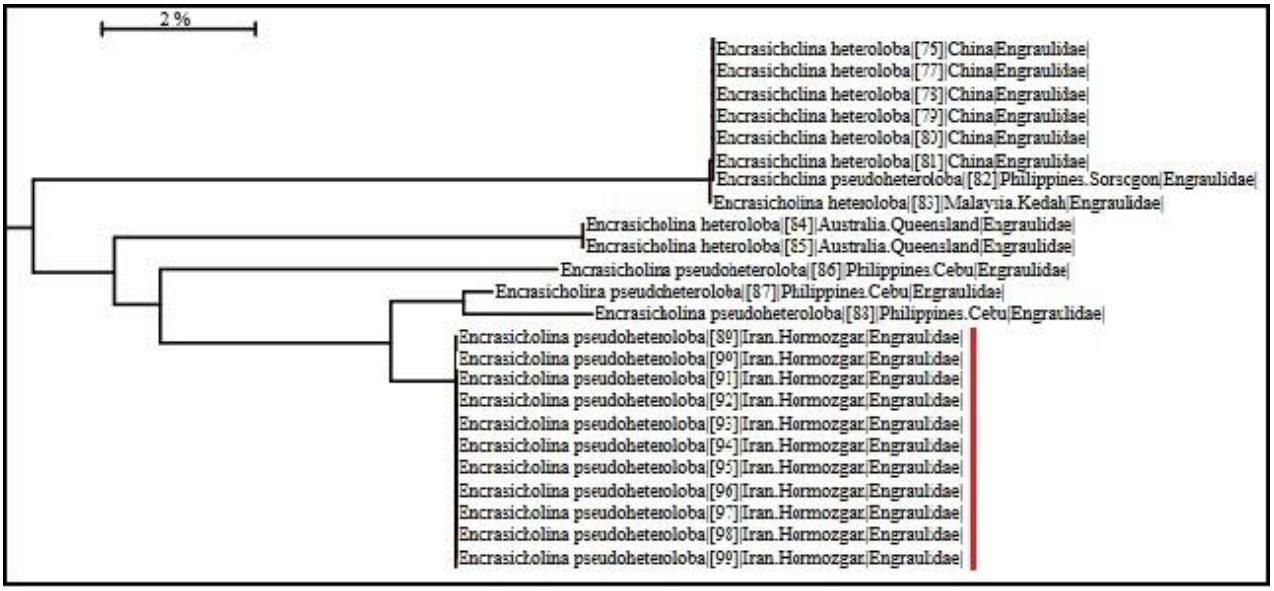

Fig. **3.** COI database tree for *E. pseudoheteroloba*

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

Table 1. Estimates of Pairwise Genetic Distances (below) and Standard error (above) of *Encrasicholina* Species under Kimura 2-Parameter Model.

species	1	2	3	4	5	6
1. *Encrasicholina punctifer*. Iran	0.000	0.002	0.002	0.002	0.021	0.022
2. *Encrasicholina punctifer*. Mozambique	0.002		0.000	0.000	0.021	0.022
3. *Encrasicholina punctifer* .South Africa	0.002	0.000		0.000	0.021	0.022
4. *Encrasicholina punctifer*. Taiwan	0.002	0.000	0.000		0.021	0.022
5. *Encrasicholina pseudoheteroloba* .Iran	0.182	0.179	0.179	0.179	0.000	0.006
6. *Encrasicholina pseudoheteroloba* . Thailand	0.189	0.186	0.186	0.186	0.020	

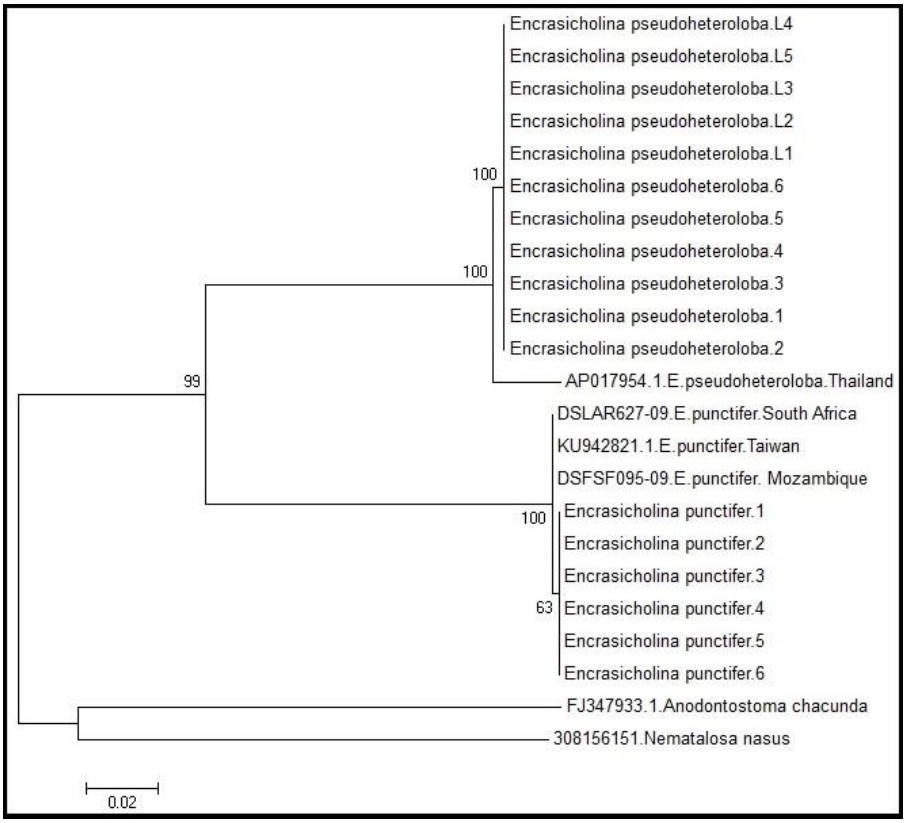

Fig. 4. Phylogenetic consensus tree of Encrasicholina species constructed based on COI gene ML method. *Anodontostoma chacunda* and *Nematalosa nasus* used as outgroup

4. Discussion

This study was the first to determine the efficacy of DNA barcoding in identification of genus *Encrasicholina* in the Persian Gulf. The mitochondrial cytochrome oxidase I (COI) region of 17 samples was successfully amplified using PCR but one sample failed. No stop codons, deletions or insertions were observed in any of the sequences, suggesting that all of the sequences constituted functional mitochondrial COI gene (Zhang, 2011; Zhu *et al*., 2013).

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

The results of molecular identifications verified the morphological according to the sequences available in gene banks (NCBI and BOLD). Although morphological characteristics of fish larvae captured from Bandar Aftab were insufficient for accurate identification, but these fish larvae were validly identified using barcoding technique. DNA barcoding is believed to be one of the best methods to confirm larval fish species identification (Ko *et al*., 2013).

The similarities of the sequences of the two identified species were compared to the sequences available in database and a phylogram was created. The phylogram showed that the sequences of different geographically regions for the same species invariably get clustered in the same clade (visible in both species). Therefore, CO1 gene could be used as a global DNA marker in fish barcoding. Present results correspond with Ward *et al* (2008) findings.

In the present study, there were no variations among the sequence within *E. punctifer* and *E. pseudoheteroloba* in Persian Gulf (Table 1). In addition, the results of genetic variation showed no significant differentiations between sequences obtained from *E. punctifer* and *E. pseudoheteroloba* in Persian Gulf and the sequences from the same species in other geographic areas (0.000-0.002). Similarly, low or zero conspecific variations were observed in other barcoding studies on several teleost fish (Lakra *et al*., 2009; Zemlak *et al*., 2009; Ward *et al*., 2005).

According to the results of phylogeny tree either using NJ or ML, in *E. pseudoheteroloba* clade, there were two sub-clades with the bootstrap value of 100% (Persian Gulf samples clad and clad sample of Thailand). Moreover, in *E. punctifer* clade, there were two sub-clades supported by the bootstrap value of 100% (Persian Gulf samples clad and Taiwan, South Africa and Mozambique samples clad). Although DNA barcoding analysis seeks only to delineate species borders, there is also some phytogeography information in CO1 sequences (Ward *et al*., 2005, Bucklin *et al*., 2007). Unfortunately, *E. devisi* species that is very similar to *E. pseudoheteroloba* (Whitehead *et al*., 1988; Hata and Motomura, 2017) was not found in sampling and its CO1 gene sequence was not available in NCBI. Lavoue *et al* (2017) stated that these two species are synonymous and *E. heteroloba* was misidentified as *E. devisi*.

4. Conclusion

We confirmed that the COI gene is suitable for DNA barcoding in the genus *Encrasicholina* from the Persian Gulf. The identified sequence would be of great value for future studies on phylogeny, management and conservation of resources in this valuable species.

Acknowledgements

This research was supported through funding of the Iranian Biological Resource Center. We wish to thank Dr. Ali Salarpouri from the Persian Gulf and Oman Sea Ecological Research Institute for his kind help in preparing samples.

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

References

Baldwin, C.C., Brito, B.J., Smith, D.G., Weigt, L.A., Escobar-Briones, E. 2011. Identification of early life-history stages of Caribbean Apogon (Perciformes: Apogonidae) through DNA Barcoding. Zootaxa. 3133: 1-36.

Bianchi, G. 1985. Field guide to the commercial marine and brackish water species of Pakistan. 234p.

Bucklin, A., Wiebe, P.H., Smolenack, S.B., Copley, N.J., Beaudet, J.G., Bonner, K.G., Pierson, J.J. 2007. DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). Journal of Plankton Research. 29(6): 483-493.

Carpenter, K.E. 1997. Living marine resources of Kuwait, Eastern Saudi Arabia, Bahrain, Qatar, and the United Arab Emirates. Food and Agriculture Org.

Cole, J., McGlade, J. 1998. Clupeoid population variability, the environment and satellite imagery in coastal upwelling systems. Reviews in fish biology and fisheries. 8(4): 445-471.

Dayrat, B. 2005. Towards integrative taxonomy. Biological journal of the Linnean society. 85(3): 407-415.

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(5): 294-299.

Frézal, L., Leblois, R. 2008. Four years of DNA barcoding: current advances and prospects. Infection, Genetics and Evolution. 8(5): 727-736.

Garza-Gil, M.D., Prada-Blanco, A., Vázquez-Rodríguez, M.X., Varela-Lafuente, M. 2011. Global warming and its economic effects on the anchovy fishery and tourism sector in North-Western Spain. In Tech Open Access Publisher.

Hata, H., Motomura, H. 2017. Validity of Encrasicholina pseudoheteroloba (Hardenberg 1933) and redescription of Encrasicholina heteroloba (Rüppell 1837), a senior synonym of Encrasicholina devisi (Whitley 1940) (Clupeiformes: Engraulidae). Ichthyological Research. 64: 18-28.

Hata, H., Motomura, H. 2016. Two new species of the genus Encrasicholina (Clupeiformes: Engraulidae): E. intermedia from the western Indian Ocean and E. gloria from the Persian Gulf, Red Sea and Mediterranean. Raffles Bulletin of Zoology. 64: 79-88.

Hebert, P.D., Cywinska, A., Ball, S.L. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B: Biological Sciences. 270(1512): 313-321.

Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., Zhang, J. 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLoS one, 3(6): e2490.

Khan, N.Y. 2007. Multiple stressors and ecosystem-based management in the Gulf. Aquatic Ecosystem Health & Management. 10(3): 259-267.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 16(2): 111-120.

Ko, H.L., Wang, Y.T., Chiu, T.S., Lee, M.A., Leu, M.Y., Chang, K.Z., Shao, K.T. 2013. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS One. 8(1): e53451.

Lakra, W.S., Verma, M.S., Goswami, M., Lal, K.K., Mohindra, V., Punia, P., Hebert, P. 2011. DNA barcoding Indian marine fishes. Molecular Ecology Resources. 11(1): 60-71.

Lavoué, S., Bertrand, J.A., Wang, H.Y., Chen, W.J., Ho, H.C., Motomura, H., Hata, H., Sado, T., Miya, M. 2017. Molecular systematics of the anchovy genus Encrasicholina in the Northwest Pacific. PloS one. 12(7): p.e0181329.

Nelson, J.S. 2006. Fishes of the World. 4rd edition. John Wiley and Sons. 624 p.

Owfi, F. 2010. Systematic review and biogeography of fishes Persian Gulf and Oman Sea, Fatemi, M., Coad, B. Tehran: Islamic Azad University - Science and Research Branch.

Randall, J.E. 1995. Coastal fishes of Oman. University of Hawaii Press.

Ratnasingham, S., Hebert, P.D. 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Molecular Ecology Notes. 7(3): 355-364.

Salarpour, A., Darvishi, M., Behzadi, S., Seraji, F. 2008. Reproduction and feeding of buccaneer anchovy (Encrasicholina punctifer) from coastal waters of Qeshm Island, the Persian Gulf. Iranian Scientific Fisheries Journal. 17(3): 45 -54.

Sheppard, C., Al-Husiani, M., Al-Jamali, F., Al-Yamani, F., Baldwin, R., Bishop, J., Jones, D.A. 2010. The Gulf: a young sea in decline. Marine Pollution Bulletin. 60(1): 13-38.

Steinke, D., Zemlak, T.S., Hebert, P.D. 2009. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. PLoS one. 4(7): e6300.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 30(12): 2725-2729.

[DOR: 20.1001.1.23222751.1396.7.3.4.3]

Van Zailinge, N.P., Owfi, F., Ghasemi, S., Khorshidian, K., Niamaimandi, N. 1993. Resources of small pelagics in Iranian waters, a review. FAO/ UNDP Fisheries Development Project Iran/83/013: 370.

Ward, R.D., Costa, F.O., Holmes, B.H., Steinke, D. 2008. DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but Zeus faber and Lepidopus caudatus each probably constitute two species. Aquatic Biology. 3(1): 71-78.

Ward, R.D., Woodwark, M., Skibinski, D.O.F. 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of Fish Biology. 44(2): 213-232.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D. 2005. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 360(1462): 1847-1857.

Whitehead, P.J.P., Nelson, G.J., Wongratana, T. 1988. FAO species catalogue. Vol 7. Clupeoid fishes of the world (Engraulidae). An annotated and ilustrated catalogue of the herrings, sardines, pilchards, sprats, anchovies and wolfherrings. Part 2-Engraulidae. FAO Fish Synop.

Zemlak, T.S., Ward, R.D., Connell, A.D., Holmes, B.H., Hebert, P.D. 2009. DNA barcoding reveals overlooked marine fishes. Molecular Ecology Resources. 9(s1): 237-242.

Zhang, J. 2011. Species identification of marine fishes in China with DNA barcoding. Evidence-Based Complementary and Alternative Medicine. 2011.

Zhang, J.B., Hanner, R. 2011. DNA barcoding is a useful tool for the identification of marine fishes from Japan. Biochemical Systematics and Ecology. 39(1): 31-42.

Zhu, S.R., Fu, J.J., Wang, Q., Li, J.L. 2013. Identification of Channa species using the partial cytochrome c oxidase subunit I (COI) gene as a DNA barcoding marker. Biochemical Systematics and Ecology. 51: 117-122.

[DOR: 20.1001.1.23222751.1396.7.3.4.3]