



## Establishing a reference collection and DNA barcoding the coastal fishes of the United Arab Emirates

WILLIAM B. LUDT <sup>1,2</sup>

RIMA W. JABADO <sup>3</sup>

SHAMSA M. AL HAMELI <sup>4</sup>

LAYNE FREEMAN <sup>5</sup>

GENTA TERUYAMA <sup>5</sup>

PROSANTA CHAKRABARTY <sup>5</sup>

SHAIKHA S. AL DHAHERI <sup>4</sup>

<sup>1</sup>*Department of Ichthyology, Natural History Museum of Los Angeles County,  
900 Exposition Blvd, Los Angeles, CA 90007, USA  
corresponding author: wludt@nhm.org*

<sup>2</sup>*National Museum of Natural History, Smithsonian Institution,  
PO Box 37012, Washington, DC 20013-7012, USA*

<sup>3</sup>*Elasmo Project, PO Box 29588, Dubai, United Arab Emirates*

<sup>4</sup>*Environment Agency - Abu Dhabi, PO Box 45553, Abu Dhabi, United Arab Emirates*

<sup>5</sup>*Ichthyology Section, 119 Foster Hall, Museum of Natural Science,  
Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA*

**Citation:** Ludt, W.B., Jabado, R.W., Al Hameli, S.M., Freeman, L., Teruyama, G., Chakrabarty, P. & Al Dhaheri, S.S. (2020) Establishing a reference collection and DNA barcoding the coastal fishes of the United Arab Emirates. *Journal of the Ocean Science Foundation*, 35, 54–64.

**doi:** <https://doi.org/10.5281/zenodo.3934741>

## Abstract

The Arabian Gulf is a semi-enclosed sea largely isolated from surrounding bodies of water. Due to several physical attributes, this basin comprises a variety of habitats that are influenced by strong seasonal fluctuations in abiotic conditions, which in turn puts considerable stress on local species. The biodiversity of this body of water has been historically understudied and is under increasing threats from overexploitation, development, and climate change. Documenting the current state of this biodiversity is therefore of paramount importance. The coastal waters of the United Arab Emirates make up the majority of the southern portion of the Gulf, and the species richness of these waters has never been formally documented. Here, we present the findings of an inshore and offshore biodiversity survey that recently sampled along the entirety of the southern coast of the Gulf. We focused on the non-elasmobranch ichthyological biodiversity (i.e. bony fishes), and established a regional collection housed at the Environmental Agency-Abu Dhabi (EAD) for future researchers to use and to reference. Additionally, we mtDNA barcoded a subset of the specimens collected using the mitochondrial cytochrome oxidase subunit 1 (COI) marker. In total, 631 specimens were collected from 164 different localities representing 158 species and 60 families. We sequenced 465 of these individuals to assess the match to all sequences in the BOLD database and their species identification. The results suggest possible cryptic species or strong population structuring that warrant future taxonomic exploration. This study represents the largest bony-fish barcoding effort for the region to date and provides data that will be useful for scientists, non-governmental organizations, and policymakers moving forwards.

**Key words:** ichthyology, coral-reef fishes, Indian Ocean, conservation, Arabian Gulf, UAE, marine biology.

## Introduction

Documenting biodiversity is of primary importance on our dynamic planet, allowing us to identify shifting global patterns in species distributions and abundance. This is particularly true for understudied marine systems, as rapid development, exploitation, and habitat alterations can have unexpected consequences on marine ecosystem functions (Worm et al. 2006). The Arabian Gulf (also referred to as the Persian Gulf; henceforth the Gulf) is one of these understudied marine regions that is being heavily influenced by anthropogenic factors (Price 1993, Sale et al. 2011, Vaughan & Burt 2016, Ludt et al. 2018). The Gulf is a semi-enclosed body of water that lies on a northwest to southeast axis roughly from 24° N to 30° N, and is largely isolated and shallow (average depth is 35 m; Carpenter et al. 1997). However, deeper waters occur on the eastern margins and extend south through the Strait of Hormuz, which is the only connection to the Sea of Oman and the larger Indian Ocean. Due to the Gulf's physical location in an arid environment, evaporation exceeds both freshwater input and precipitation, resulting in an influx of surface water from the Sea of Oman, which in part drives a cyclonic circulation pattern (Reynolds 1993, Kämpf & Sadrinasab 2006, Pous et al. 2015). The combination of these physical attributes results in highly variable seasonal environmental conditions, with temperature and salinity ranging between 11.4–16.2° C and 32–41.5 psu, respectively (Coles & Fadlallah 1991, Sheppard et al. 2010, Ludt et al. 2018). These factors put considerable stress on the species inhabiting the Gulf and have in turn influenced the diversity of species in this body of water.

Overall, the biodiversity of the Gulf is depauperate relative to adjacent regions, with low levels of endemism and species richness (Sheppard et al. 1992, Sale et al. 2011). However, distinct habitat clusters result in a high species turnover, or  $\beta$ -diversity, within the Gulf, and comparisons with the nearby Red Sea reveal that the Gulf has a dissimilar species assemblage, making it of special conservation concern (Price 2002, Sheppard et al. 2010, Ludt et al. 2018). High primary productivity in the Gulf supports large fisheries for shrimps, lobsters, a wide variety of bony fishes, and elasmobranchs (Carpenter et al. 1997, Jabado & Spaet 2017), yet rapid development and regulatory differences between the eight countries bordering the Gulf have resulted in the overexploitation of many of these marine resources (Price 1993, Sale et al. 2011, Buchanan et al. 2019). Threats to the entire Gulf include factors such as population growth, increased oil exploration and shipping traffic, overexploited fisheries, and global climate change, making the need for categorizing the biodiversity of this region essential (Sale 2011,

Sheppard et al. 2010). This is especially true for bony fishes (infraclass Teleostei), as a recent assessment found that 8.2% of regional species are threatened: nearly double that of other regions where similar assessments have been conducted (Buchanan et al. 2019).

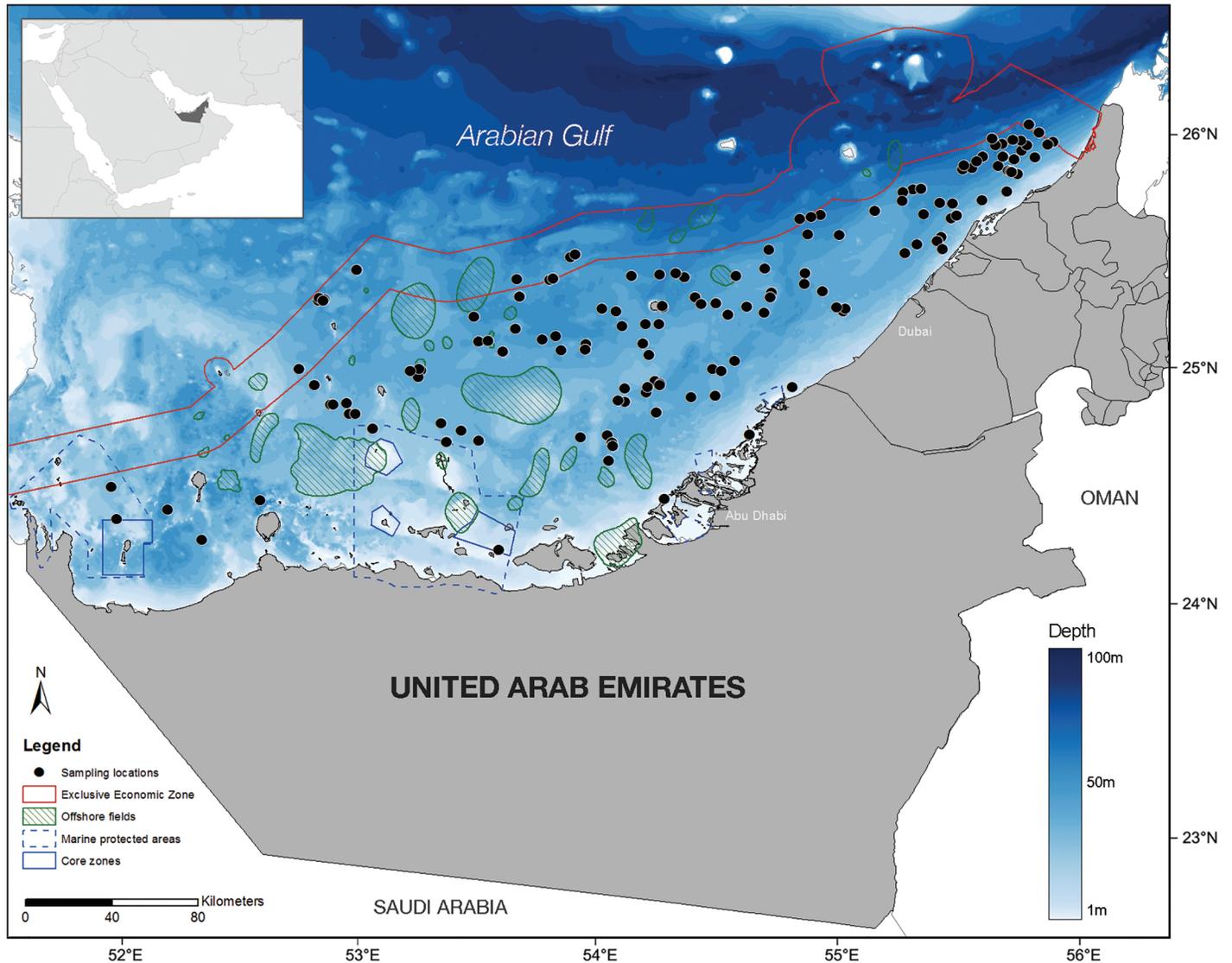
DNA barcoding, or the use of a ubiquitous gene for identifying organisms, is a commonly used approach and has a variety of applications (Hajibabaei et al. 2007). The mitochondrial cytochrome oxidase subunit I gene (COI) has become standard in many disciplines (often referred to as the “barcoding gene”), spanning the Tree of Life, and has been used to identify unknown species, or identify potential new species (Herbert et al. 2003, Herbert & Gregory 2005). For fishes, the application of barcoding has been successfully applied across the fish Tree of Life (Costa & Carvalho 2007, Ward et al. 2009) and has been useful for discovering cryptic species (e.g. Baldwin et al. 2011, Victor 2013, 2014), or identifying larvae, which can look very different from mature individuals (e.g. Pegg et al. 2006, Victor et al. 2009, Baldwin & Johnson 2014). Barcode repositories, such as the Barcode of Life Database (BOLD; [www.barcodinglife.com](http://www.barcodinglife.com)), are extremely helpful for conservation, management, identification, and other scientific purposes (Ward et al. 2009), but depend on accurate species identification. Uploading sequences from vouchered museum specimens can help ameliorate problems associated with misidentifications, as vouchers can always be reassessed by taxonomic experts (Chakrabarty et al. 2013). Furthermore, the extent that these databases can be used greatly depends on the sampling effort for a region (e.g. Weigt et al. 2012, Victor et al. 2015), and their benefits can be limited in regions of the world that have not been the focus of prior barcoding initiatives. The Gulf is one of these regions, with historic under-sampling, even in comparison to the nearby Red Sea (Vaughan & Burt 2016).

Targeted barcoding studies in the Gulf have been restricted to date and have primarily focused on cartilaginous (Jabado et al. 2015) or bony fishes (Asgharian et al. 2011, Rabaoui et al. 2019). While efforts should be made to expand the taxonomic scope of barcoding efforts in the region to include invertebrates, broadening the geographic range is equally important due to the high  $\beta$ -diversity, or species turnover, in the Gulf. Currently, the most inclusive barcoding studies for bony fishes in the region have been restricted to the Nayband National Park in Iran, along the central eastern part of the Gulf (Asgharian et al. 2011), and along the coast of Saudi Arabia in the western part of the Gulf (Rabaoui et al. 2019). Habitat differences between these locations and other regions mean these studies may not be representative of the entire regional bony fish assemblage, necessitating additional barcoding studies.

For aquatic organisms, one of the most physically taxing regions of the Gulf occurs along the southern extent, where extremes in temperature and salinity vary considerably throughout the year (Kämpf & Sadrinasab 2006). Understanding which species can tolerate these conditions, and how they do so, can provide essential information for understanding how species will cope with global climate change in general. The coastal waters of the United Arab Emirates (UAE), therefore, provide a rare opportunity for establishing a baseline study of the biodiversity in the southern Gulf. Trawling, which is common for shrimp fisheries throughout the region and which can highly impact the benthic fauna, is not permitted in UAE waters, making this region ideal for quantifying species richness. The Gulf has a dearth of regional natural history collections which endangers our future ability to document changes in species assemblages over time. Here, we contribute to recording the biodiversity of the Gulf by establishing a regional ichthyology collection at the Environment Agency-Abu Dhabi (EAD), and by adding the largest barcoding dataset to date for bony fishes from the Gulf region along the UAE associated with vouchered museum specimens.

## Materials and Methods

From April to June and October to December 2016, the EAD undertook comprehensive fisheries-research surveys using trawls and traps across UAE Gulf waters at depths from 1 to 46 m and within 12 nautical miles of the territorial boundary (Fig. 1). The surveys were planned to cover both spawning (September–December) and non-spawning (March–June) seasons. Inshore areas were mainly sampled using traps while offshore areas were sampled using a combination of traps (for untrawlable bottoms) and trawls. Immediately after collection, preliminary identifications were assigned to the specimens, which were then frozen onboard the research vessels for future processing. A number of samples were also collected opportunistically during fixed-stake-net surveys



**Figure 1.** A map of the study region with dots representing sampling locations throughout the Arabian Gulf; the inset provides a broader scope of where the study region is located within the Gulf.

(Al Hameli, unpublished data). Specimens were later transported to the EAD Fish Lab, thawed, identified, photographed, catalogued, and fixed in a 10% formalin solution. Prior to fixation, tissue samples were taken from the right pectoral fin and/or gills for representative species from each collecting locality. After fixation, specimens were transferred through a series of ethanol concentrations prior to long-term preservation in 75% ethanol (EtOH). All specimens are stored in a climate-controlled facility in Abu Dhabi, UAE, and all tissue samples were stored in 95% EtOH in a  $-20^{\circ}\text{C}$  freezer. The EAD catalog numbers are listed in Table S1.

A sub-sample of collected individuals, one to 5 individuals per species, was chosen for mtDNA sequencing to document the genetic diversity of fishes in the Gulf. DNA was extracted from all tissue sub-samples using a Qiagen DNeasy extraction kit following manufacturers protocols. Polymerase chain reactions (PCR) were comprised of 4mM magnesium chloride, 1X buffer solution, 120 nM of a forward and reverse primer, 0.6 mM of dNTPs, 1.25 units of Taq, and approximately 10 ng of DNA. All reactions were in 25  $\mu\text{L}$  volumes. The following BOLD primers were used for each reaction, from Ward et al. (2005): FishF1 [5'-TCAACCAACCACAAAGACATTGGCAC-3'] and FishR1 [5'-TAGACTTCTGGGTGGCCAAAGAATCA-3']. All PCR reactions were conducted with the following thermal profile: 1 minute denaturing at  $94^{\circ}\text{C}$ , followed by a 30 second annealing temperature of 48–52 $^{\circ}\text{C}$  depending on the sample, and finished with an extension of 45 seconds at  $72^{\circ}\text{C}$ , for 33 cycles following protocols outlined in Ludt et al. (2012). Samples were then purified and sequenced in both the forward and reverse direction by Macrogen USA.

Resulting forward and reverse sequences were aligned, edited, and exported as a consensus sequence in Geneious v6.0.5 (Biomatters). These consensus sequences were then aligned using the Geneious alignment plugin, with a 65% cost-similarity matrix. The resulting alignment was verified by eye and was also translated into amino acids to ensure that no stop codons were present in the middle of the protein coding sequence. Any questionable results were re-checked before proceeding. All subsequent analyses used the nucleotide alignment, as opposed to the amino acid translation. We constructed a maximum likelihood tree using the program RAxML v8.2.1 on the CIPRES Scientific Gateway Portal (Miller et al. 2010, Stamatakis 2014) to visualize sequence clustering (but not assess sequence matches). A GTRGAMMA substitution model was used with the rapid bootstrapping option. The resulting tree was then visualized with the program FigTree v1.4.3. All of the sequences obtained in this study are deposited in GenBank (accession numbers MT076467–MT076931).

To assess the match between our species identifications and the BOLD database identifications, we tested our sequences on the BOLD identification engine, which lists matching sequences on BOLD using a similarity (p-distance) algorithm, including all public and private sequences on the database. Separate from the similarity list and the ID tree (by K2P method), the BOLD species-level ID engine assigns the sequence to a BIN, a cluster determined by a specific clustering algorithm used by BOLD. The species ID for the BIN can then be evaluated by the species names for each record, which includes all IDs by various contributors of varying taxonomic expertise (sometimes none), whether correct or misidentified or identified to a higher taxonomic level only.

## Results

In total, 631 specimens were collected from 164 different localities spread across the coast of the UAE, representing 158 species from 60 different families (Table 1, Table S1). The most species-rich family collected was the Carangidae, for which 20 species were identified, followed by the Nemipteridae, with 9 different species collected. Roughly half of the families (29) are represented by only a single species in this collection. This is not unusual for bony fishes, as 39 families are only reported to have a single species present in the Gulf (Froese & Pauly, 2019). However, there were several families that were only represented by a single species in our dataset despite the presence of several known species in the region (e.g. Batrachoididae, Triglidae, Sillaginidae). While most species were represented by multiple individuals, for 38 species we had only a single individual. All specimens were sorted and are now stored in 369 lots in the EAD fish collection.

After subsampling the species collected in the original surveys, and the few species added opportunistically, 465 tissue samples were successfully sequenced, representing 152 species. Six species were identified but not sequenced due to logistical issues or failed DNA amplification: *Alepes djedaba*, *Herklotsichthys lossei*, *Hippocampus suezensis*, *Parapercis robinsoni*, *Parupeneus rubescens*, and *Plectorhinchus sordidus*. Sequence length ranged from 538–663 bp. Initial alignment length was 735 bp, but was trimmed to 665 bp to remove ragged edges. The resulting tree largely groups all samples within their respective families (Fig. S1).

A large majority of our species identifications agreed closely with the species identification in the BOLD database: 128 species IDs (403 sequences) matched a BIN containing the same species name, with an average sequence match of 99.86%. Of these 128, 72 species IDs (221 sequences) matched to BINs containing multiple species names including ours (a single different species name, or even a synonym, by any contributor makes a BIN non-unanimous). Only a small subset of our species identifications did not agree with any name in the BIN: 8 species (21 sequences) matched to a BIN with multiple names, none of which was the same as ours, but in all cases the match was to a potential sibling species (two of the 8 to the same genus but labeled sp.).

Only ten of our species (31 sequences) did not match within 3% of any sequence on BOLD: of those, 4 species were still matching closest to the same nominal species: *Antennatus* cf. *nummifer* (6% divergent), *Apogonichthyoides* cf. *nigripinnis* (7% divergent), *Callionymus filamentosus* (6% divergent), and *Rogadius pristiger* (by 4% divergent). These cases clearly indicated the presence of cryptic species or unresolved species complexes. The 6 remaining species were distant from any sequence on BOLD, but still the nearest BIN included IDs of congeners. These could reflect unsampled species, cryptic species, or unresolved taxonomy: *Colletteichthys occidentalis* (8%), *Callionymus persicus* (7%), *Pomadasy argenteus* (6%), *Upeneus doriae* (5%), *Pseudochromis persicus* (8%), and *Choridactylus multibarbus* (9%).

**TABLE 1.** Bony-fish families documented and collected along the coast of the United Arab Emirates for this study, with the number of species and specimens for each family.

Family	# species	# specimens	Family	# species	# specimens
Antennariidae	1	1	Mullidae	5	22
Apistidae	1	1	Nemipteridae	9	47
Apogonidae	6	24	Ostraciidae	1	2
Ariidae	1	4	Paralichthyidae	1	10
Ariommatidae	1	3	Pegasidae	1	1
Balistidae	1	5	Pinguipedidae	1	4
Batrachoididae	1	5	Platycephalidae	6	12
Belonidae	2	6	Plotosidae	1	1
Bleniidae	3	7	Pomacanthidae	1	6
Bothidae	2	6	Pomacentridae	3	16
Callionymidae	2	11	Priacanthidae	1	1
Carangidae	20	94	Psettodidae	1	6
Centriscidae	1	4	Pseudochromidae	1	5
Chaetodontidae	2	8	Rachycentridae	1	5
Clupeidae	2	3	Scombridae	2	7
Dactylopteridae	1	4	Scorpaenidae	3	8
Echeneidae	1	20	Serranidae	5	29
Ehippidae	1	4	Siganidae	2	5
Fistulariidae	1	4	Sillaginidae	1	4
Gerreidae	5	22	Soleidae	2	7
Gobiidae	2	2	Sparidae	6	16
Haemulidae	5	10	Sphyraenidae	3	11
Hemiramphidae	1	3	Synanceiidae	1	1
Labridae	3	19	Syngnathidae	1	1
Leiognathidae	6	19	Synodontidae	2	5
Lethrinidae	3	7	Terapontidae	3	10
Lutjanidae	6	29	Tetraodontidae	3	23
Microdesmidae	1	1	Triacanthidae	2	9
Monacanthidae	3	23	Trichiuridae	1	1
Monodactylidae	1	4	Triglidae	1	4

## Discussion

The Gulf presents many challenges to local marine communities. On deeper time scales, it was completely dry during the last glacial maximum (Lambeck 1996), and the communities that have formed since the basin flooded again approximately 6 kya may represent earlier, transitory assemblages that could change with age and maturity of the Gulf. Documenting these communities is therefore essential for understanding how climate change and succession dynamics vary over time in marine communities. Furthermore, as the Gulf is an area that will likely be impacted heavily due to anthropogenic influences in the coming century, recording the biodiversity of this region with high-quality museum specimens and vouchered genetic material is essential.

Here we present the largest barcoding effort to characterize the bony fish community of the Gulf to date. Previous studies to explicitly barcode bony fishes in the region reported 73 and 114 species in the eastern (Asgharian et al. 2011) and western Gulf (Rabaoui et al. 2019), respectively. Our dataset here overlaps these previous efforts by 88 species, but adds an additional 64 species and 13 families that have not been previously barcoded in the Gulf. The observed overlap in species was greater for the western Gulf (79 species in common with those in Rabaoui et al. 2019), than the eastern Gulf (40 species in common with those in Asgharian et al. 2011), with 31 wide-ranging species being found in all three regions (eastern, western, and southern Gulf). This corroborates other studies that have suggested high species turnover in the Gulf in other taxa (Price 2002) and highlights the importance of regional studies.

Single-locus mitochondrial studies are not well-suited for resolving deeper relationships among taxa, due to site saturation or other factors (Hajibabaei et al. 2006). This is apparent in this study, where certain families are not recovered as monophyletic, as is also seen in other fish-barcode datasets (Ward et al. 2005). It should be noted that we are not presenting our findings as a phylogenetic analysis. While there is clearly some phylogenetic signal in the data, we are using mitochondrial COI sequences to identify species and, to some degree, evaluating species boundaries (Ward et al. 2005). The use of COI, or other single-locus mitochondrial markers, can be useful for detecting underlying population structure and identifying potential cryptic species (Herbert et al. 2004, Zemplak et al. 2009, Baldwin et al. 2011).

One example of a potentially cryptic species in this dataset involves the whipfin mojarras, a species complex in the family Gerreidae. Gerreids in general are morphologically and phenotypically conserved, and Asgharian et al. (2011) reported two divergent clades of *Gerres filamentosus* in their barcode study, which was suggested to represent a potential cryptic species (Asgharian et al. 2010). We also recovered two distinct clades of whipfin mojarras, one not the same as those found in Asgharian et al. (2010). These multiple lineages indicate either cryptic species or unresolved population structuring in that mojarra species complex (but see Iwatsuki et al. [2015] for additional comments on this species in the Gulf).

Distinct structuring and/or unresolved species were also found within *Nemipterus peronii*, which was recovered in two separate clades, approximately 7.9% divergent from one another, both also present in BOLD (BINs AAI3070 and AAU0855). Correspondingly, in their review of the threadfin breams in that family, Hung et al. (2017) suggested the taxonomy was unresolved and additional species were likely present. Some of our sequences of other widespread species diverged significantly from conspecific lineages outside of the Gulf. Several of these corroborate findings in Rabaoui et al. (2019), such as *Triacanthus* cf. *biaculeatus*, *Lepidotrigla* cf. *bispinosa*, and *Tylosurus* sp., while other findings are novel, such as *Pegasus* cf. *volitans* (7% divergent from the nearest *Pegasus* sequence) and *Pseudorhombus* sp. (9% divergent from *P. arsius* reported from the Gulf).

In addition to detecting cryptic species or population structuring, barcodes are also frequently used to confirm species identities. The accuracy of barcode databases is essential for this task, and misidentified sequences sharply reduce the usefulness of this technique. Indeed, the majority of the BINs to which our sequences matched were composed of nearly identical sequences identified as different species. While this occurred across many taxa we collected, it is particularly troublesome among commercially important species, as barcodes are increasingly being used to authenticate the identity of seafood being sold in restaurants or markets (reviewed in Willette et al. 2017). Of the commercially important groups in this dataset only three of 20 carangid species, two of 7 lutjanids, one of 5 serranids, and none of the three lethrinids sequenced matched BINs with a single species ID. One extreme example of this are fishes in the genus *Carangoides*, some sequences of which matched to 7 different reported species in the BOLD database. Since identifications on databases, both BOLD and GenBank, are contributed by

anyone and are not presently validated, or even revised, after submission (potentially for decades), then as sample sizes accumulate, as would occur for common and commercially important species, the misidentifications only proliferate. The only solution to this problem is sequencing vouchered museum specimens whose identity can be re-examined, and the reference collection at EAD is a step in that direction.

Ichthyology along the Arabian Peninsula has a long history (reviewed in Kuronuma & Abe [1986] and Jawad [2012]) that is scattered with the well-known names of European explorers such as Forsskål, Rüppell, Bleeker, and Günther. Over time the number of species recorded in this body of water has increased substantially, yet quantifying specifically how many species occur in the Gulf has been difficult, as most records focus on individual countries (list of country-based studies in Grandcourt [2012]). Comprehensive lists of species from the entire Gulf were first attempted by Regan (1905), with more recent Gulf-wide studies for fishes reported in Kuronuma & Abe (1986), Carpenter et al. (1997), and Eagderi et al. (2019), which list 465, 539, and 670 fish species, respectively. However, online repositories list as many as 884 fish species being present in the Gulf (e.g. Froese & Pauly 2020).

The vast majority of fish collections from the Gulf, and the Arabian Peninsula in general, are held outside of the region, and many are not associated with vouchered specimens, making it difficult to accurately assess how many species actually occur in the Gulf (Bishop 2003). Here, we establish a new regional collection containing both specimen and genetic resources in an attempt to create verifiable records for the southern Gulf. Additionally, we provide the most extensive barcoding initiative for bony fishes of the Gulf that will aid in future scientific efforts and help in accurately delimiting species. While this study also focuses on a single country, these data, in conjunction with other previous, and future, barcoding initiatives, will ultimately lead to a more accurate understanding of the overall biodiversity in the region. The Gulf is an understudied body of water with a distinct biota that needs future attention and documentation. Ultimately, we hope that the resources contained in this collection and genetic dataset will help establish a baseline from which future studies can be built.

## Acknowledgments

We would like to thank the following individuals, in no particular order, for their assistance in planning and collecting the samples used in this study: Edwin Grandcourt, Reem Al Baharna, Stanley Hartmann, Franklin Francis, Ahmed Al Shamsi, Sultan Al Ali, and Firoz Kandathil. We gratefully acknowledge the support and involvement of the New Zealand National Institute of Water and Atmospheric Research team, especially Neil Bagley, Rosemary Hurst, Keith Michael, Peter Marriott, Bilal Bjeirmi, Peter McMillan, Darren Stevens, Warrick Lyon, Darren Parsons, and Dan MacGibbon. Thanks also go to the crew of RV *Bahith I* and *Bahith II* for their assistance with the fieldwork. The authors would also like to thank Benjamin Victor and two anonymous reviewers for suggestions that improved the final manuscript.

**Author Contributions:** RWJ and SSAD conceived of the study. SSAD approved logistics associated with establishing the collection. RWJ collected the specimens. WBL and RWJ identified the specimens, preserved, and took tissue samples of the specimens. RWJ and SMAH subsampled and determined which specimens to include in the study. PC and WBL managed aspects associated with genetic lab work. WBL, LF, and GT conducted all lab work. WBL, LF, and GT processed raw sequence data. WBL analyzed the data and wrote the manuscript. All authors contributed to editing the manuscript and approved the final version.

## Archived supplementary data:

Fig. S1: Phenetic tree of sequence similarity constructed using a maximum likelihood approach for mtDNA COI sequences from the EAD survey of the bony-fish species of the Arabian Gulf.

**doi:** <https://doi.org/10.5281/zenodo.3934882>

Table S1: List of species and specimens collected and vouchered from the EAD survey of the bony-fish species of the Arabian Gulf (EAD voucher numbers; numbers in bold were specimens sequenced in this study).

**doi:** <https://doi.org/10.5281/zenodo.3934766>

## References

- Asgharian, H., Elahi, E., Kalirad, A. & Hosseinzadeh, H. (2010) Sequence data on four genes suggest nominal *Gerres filamentosus* specimens from Nayband National Park in the Persian Gulf represent two distinct species. *Iranian Journal of Animal Biosystematics*, 6 (2), 1–11.
- Asgharian, H., Sahafi, H.H., Ardalan, A.A., Shekarriz, S. & Elahi E. (2011) Cytochrome c oxidase subunit 1 barcode data of fish of the Nayband National Park in the Persian Gulf and analysis using metadata flag several cryptic species. *Molecular Ecology Resources*, 11 (3), 461–472.
- Baldwin, C.C., Castillo, C.I., Weigt, L.A. & Victor, B.C. (2011) Seven new species within western Atlantic *Starksia atlantica*, *S. lepicoelia*, and *S. sluiteri* (Teleostei, Labrisomidae), with comments on congruence of DNA barcodes and species. *ZooKeys*, 79, 21–72.
- Baldwin, C.C. & Johnson, G.D. (2014) Connectivity across the Caribbean Sea, DNA barcoding and morphology unite an enigmatic fish larva from the Florida Straits with a new species of sea bass from deep reefs off Curacao. *PLOS ONE*, 9 (5), e97661.
- Bishop, J.M. (2003) History and current checklist of Kuwait's ichthyofauna. *Journal of Arid Environments*, 54 (1), 237–256.
- Buchanan, J.R., Ralph, G.M., Krupp, F. et al. (2019) Regional extinction risks for marine bony fishes occurring in the Persian/Arabian Gulf. *Biological Conservation*, 230, 10–19.
- Carpenter, K.E., Krupp, F., Jones, D.A. & Zajonz, U. (1997) The living marine resources of Kuwait, Eastern Saudi Arabia, Bahrain, Qatar, and the United Arab Emirates, FAO species identification field guide for fishery purposes, FAO, 239 pp.
- Chakrabarty, P., Warren, M., Page, L. & Baldwin, C.C. (2013) GenSeq, An updated nomenclature and ranking for genetic sequences from type and non-type sources. *ZooKeys*, 346, 29–41.
- Coles, S.L. & Fadlallah, Y.H. (1991) Reef coral survival and mortality at low temperatures in the Arabian Gulf, new species-specific lower temperature limits. *Coral Reefs*, 9 (4), 231–237.
- Costa, F.O. & Carvalho, G.R. (2007) The Barcode of Life Initiative, synopsis and prospective societal impacts of DNA barcoding of fish. *Genomics, Society and Policy*, 3 (2), 29–40.
- Eagderi, S., Fricke, R., Esmaili, H.R. & Jalili, P. (2019) Annotated checklist of the fishes of the Persian Gulf, Diversity and conservation status. *Iranian Journal of Ichthyology*, 6, 1–171.
- Froese, R. & Pauly, D. (2020) *FishBase*. Available at [www.fishbase.org](http://www.fishbase.org) (last accessed 10 June 2020),
- Grandcourt, E. (2012) Reef fish and fisheries in the Gulf. In: Riegl, B.M. & Purkis, S.J. (Eds.) *Coral Reefs of the Gulf Adaptation to Climatic Extremes*. Springer, Dordrecht, Germany, pp. 127–161.
- Hajibabaei, M., Singer, G.A. & Hickey, D.A. (2006) Benchmarking DNA barcodes, an assessment using available primate sequences. *Genome*, 49 (7), 851–854.
- 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 (4), 167–172.
- Hebert, P.D., 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 (1512), 313–321.
- Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004) Ten species in one, DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*. *Proceedings of the National Academy of Sciences*, 101 (41), 14812–14817.
- Hebert, P.D., & Gregory, T.R. (2005) The promise of DNA barcoding for taxonomy. *Systematic Biology*, 54,(5), 852–859.
- Hung, K.W., Russell, B.C. & Chen, W.J. (2017) Molecular systematics of threadfin breams and relatives (Teleostei, Nemipteridae). *Zoologica Scripta*, 46 (5), 536–551.
- Iwatsuki, Y., Bogorodsky, S., Tanaka, F., Mal, A.O. & Ali, A.H. (2015) Range extension of *Gerres infasciatus* (Perciformes, Gerreidae) from the Red Sea and the Arabian Gulf, with distributional implications for the *G. filamentosus* complex. *Cybium*, 39 (2), 155–160.

- Jabado, R.W., Al Ghais, S.M., Hamza, W., Shivji, M.S. & Henderson, A.C. (2015) Shark diversity in the Arabian/Persian Gulf higher than previously thought, insights based on species composition of shark landings in the United Arab Emirates. *Marine Biodiversity*, 45 (4), 719–731.
- Jabado, R.W. & Spaet, J.L. (2017) Elasmobranch fisheries in the Arabian Seas Region, Characteristics, trade and management. *Fish and Fisheries*, 18 (6), 1096–1118.
- Jawad, L.A. (2012) History of the Study of the Fish Fauna of Iraq. *Water Research and Management*, 2 (3), 11–20.
- Kämpf, J. & Sadrinasab, M. (2006) The circulation of the Persian Gulf, a numerical study. *Ocean Science*, 2 (1), 27–41.
- Kuronuma, K. & Abe, Y. (1986) *Fishes of the Arabian Gulf*. Kuwait Institute for Scientific Research, Kuwait City, Kuwait, 356 pp.
- Lambeck, K. (1996) Shoreline reconstructions for the Persian Gulf since the last glacial maximum. *Earth and Planetary Science Letters*, 142, 43–57.
- Ludt, W.B., Bernal, M.A., Bowen, B.W., & Rocha, L.A. (2012) Living in the past, phylogeography and population histories of Indo-Pacific wrasses (genus *Halichoeres*) in shallow lagoons versus outer reef slopes. *PLOS ONE*, 7(6), e38042.
- Ludt, W.B., Morgan, L., Bishop, J. & Chakrabarty, P. (2018) A quantitative and statistical biological comparison of three semi-enclosed seas, the Red Sea, the Persian (Arabian) Gulf, and the Gulf of California. *Marine Biodiversity*, 48 (4), 2119–2124.
- Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010, 1–8.
- Pegg, G.G., Sinclair, B., Briskey, L. & Aspden, W.J. (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef- Australia. *Scientia Marina*, 70 (S2), 7–12.
- Pous, S., Lazure, P. & Carton, X. (2015) A model of the general circulation in the Persian Gulf and in the Strait of Hormuz, Intraseasonal to interannual variability. *Continental Shelf Research*, 94, 55–70.
- Price, A.R.G. (1993) The Gulf, Human impacts and management initiatives. *Marine Pollution Bulletin*, 27, 17–27.
- Price, A.R.G. (2002) Simultaneous ‘hotspots’ and ‘coldspots’ of marine biodiversity and implications for global conservation. *Marine Ecology Progress Series*, 241, 23–27.
- Rabaoui, L., Yacoubi, L., Sanna, D., Casu, M., Scarpa, F., Lin, Y.J., Shen, K.N., Clardy, T.R., Arculeo, M. & Qurban, M.A. (2019) DNA barcoding of marine fishes from Saudi Arabian waters of the Gulf. *Journal of Fish Biology*, 95 (5), 1286–1297.
- Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes*, 7 (3), 355–364.
- Ratnasingham, S. & Hebert, P.D.N. (2013) A DNA-Based Registry for All Animal Species, The Barcode Index Number (BIN) System. *PLOS ONE*, 8 (8), e66213.
- Regan, C.T. (1905) On fishes from the Persian Gulf, the Sea of Oman, and Karachi, collected by Mr. FW Townsend. *Journal of the Bombay Society of Natural History* 16, 318–333.
- Reynolds, R.M. (1993) Physical oceanography of the Gulf, Strait of Hormuz, and the Gulf of Oman- Results from the Mt Mitchell expedition. *Marine Pollution Bulletin*, 27, 35–59.
- Sale P.F., Feary, D.A., Burt, J.A. et al. (2011) The growing need for sustainable ecological management of marine communities of the Persian Gulf. *Ambio*, 40 (1), 4–17.
- Sheppard, C.R.C., Price, A. & Roberts, C. (1992) *Marine ecology of the Arabian region, patterns and processes in extreme tropical environments*. Elsevier Science, London, UK, 359 pp.
- Sheppard, C., Al-Husaini, M., Al-Jamali, F. et al. (2010) The Gulf, a young sea in decline. *Marine Pollution Bulletin*, 60 (1), 13–38.
- Stamatakis A. (2014) RAxML version 8, a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30 (9), 1312–1313.
- Vaughan, G.O. & Burt, J.A. (2016) The changing dynamics of coral reef science in Arabia. *Marine Pollution Bulletin*, 105 (2), 441–458.

- Victor, B.C., Hanner, R., Shivji, M., Hyde, J. & Caldow, C. (2009) Identification of the larval and juvenile stages of the Cubera Snapper, *Lutjanus cyanopterus*, using DNA barcoding. *Zootaxa*, 2215 (2), 24.
- Victor, B.C., Valdez-Moreno, M. & Vásquez-Yeomans, L. (2015) Status of DNA barcoding coverage for the tropical western Atlantic shorefishes and reef fishes. *DNA Barcodes*, 3 (1), 85–93.
- Victor, B.C. (2013) The Caribbean Roughhead Triplefin (*Enneanectes boehlkei*): DNA barcoding reveals a complex of four West Indian sympatric cryptic species (Teleostei: Blennioidei: Tripterygiidae). *Journal of the Ocean Science Foundation*, 7, 44–73. <https://doi.org/10.5281/zenodo.1041966>
- Victor, B.C. (2014) Three new endemic cryptic species revealed by DNA barcoding of the gobies of the Cayman Islands (Teleostei: Gobiidae). *Journal of the Ocean Science Foundation*, 12, 25–60. <https://doi.org/10.5281/zenodo.1049119>
- Ward, R.D., Hanner, R. & Hebert, P.D.N. (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, 74, 329–356.
- Ward, R.D., Zemplak, T.S., Innes, B.H., Last, P.R. & Hebert, P.D. (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B, Biological Sciences*, 360, 1847–1857.
- Weigt, L.A., Baldwin, C.C., Driskell, A., Smith, D.G., Ormos, A. & Reyier, E.A. (2012) Using DNA Barcoding to Assess Caribbean Reef Fish Biodiversity: Expanding Taxonomic and Geographic Coverage. *PLOS ONE*, <https://doi.org/10.1371/journal.pone.0041059>
- Willette, D.A., Simmonds, S.E., Cheng, S.H., Esteves, S., Kane, T.L., Nuetzel, H., Pilaud, N., Rachmawati, R. & Barber, P.H. (2017) Using DNA barcoding to track seafood mislabeling in Los Angeles restaurants. *Conservation Biology*, 31 (5), 1076–1085.
- Worm, B., Barbier, E.B., Beaumont, N. et al. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314 (5800), 787–790.
- Zemplak, 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, 237–242.