Assessing the utility of eDNA as a tool to survey reef-fish communities in the Red Sea

Joseph D. DiBattista[1*], Darren J. Coker[2], Tane H. Sinclair-Taylor[2], Michael Stat[1], Michael L. Berumen[2], and Michael Bunce[1]

[1]Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia

[2]Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

*corresponding author: Joseph D. DiBattista, Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia

E-mail: josephdibattista@gmail.com

ORCID: 0000-0002-5696-7574

Keywords: Biodiversity, Coral reef, Environmental DNA, Metabarcoding, Next-generation sequencing

Abstract

Relatively small volumes of water may contain sufficient environmental DNA (eDNA) to detect target aquatic organisms via genetic sequencing. We therefore assessed the utility of eDNA to document the diversity of coral-reef fishes in the central Red Sea. DNA from seawater samples was extracted, amplified using fish-specific 16S mitochondrial DNA primers, and sequenced using a metabarcoding workflow. DNA sequences were assigned to taxa using available genetic repositories or custom genetic databases generated from reference fishes. Our approach revealed a diversity of conspicuous, cryptobenthic, and commercially relevant reef fish at the genus level, with select genera in the family Labridae over-represented. Our approach, however, failed to capture a significant fraction of the fish fauna known to inhabit the Red Sea, which we attribute to limited spatial sampling, amplification stochasticity, and an apparent lack of sequencing depth. Given an increase in fish species descriptions, completeness of taxonomic checklists, and improvement in species-level assignment with custom genetic databases as shown here, we suggest that the Red Sea region may be ideal for further testing of the eDNA approach.

Introduction

Coral reefs represent less than 0.02% of the surface area of our oceans and yet host more than 25% of the recognized marine species (Spalding et al. 2001). In the Red Sea, coral reefs are characterized by their high and unique biodiversity (DiBattista et al. 2016b), historical isolation (DiBattista et al. 2016a), and extreme seasonal and spatial fluctuations in abiotic and biotic factors (Raitsos et al. 2013). Yet compared to other major global reef systems (e.g., Great Barrier Reef), the associated communities of the Red Sea are considerably understudied (Berumen et al. 2013).

Community assemblages of reef fishes have historically been documented through diver-operated underwater visual census methods (UVC; Roberts et al. 1992) or baited remote underwater video surveys (BRUVs; Harvey et al. 2012). Even though these standardized survey methods are ideal for identifying large, mobile, and conspicuous species, the smaller, more cryptic species are often missed (see Ackerman and Bellwood 2000). In recent years, genomic approaches have shown considerable potential for identifying diversity in marine environments, although many of these studies are based on fixed, benthic sampling methods (e.g., Leray and Knowlton 2015). Environmental DNA (eDNA; i.e., genetic material sourced from microbes or that which has been shed from multicellular organisms), on the other hand, has proven useful in monitoring the presence/absence of invasive species and assessing biodiversity in aquatic environments (e.g., Thomsen et al. 2012).

The majority of macrobial eDNA investigations to date have been species-specific, but multi-species PCR in combination with high-throughput sequencing (metabarcoding) can extend these methods to reveal a greater diversity of species in our oceans (e.g., Miya et al. 2015). Even though the sensitivity of these high-throughput assays is likely influenced by the frequency of cell shedding, water flow (currents or tides), and rates of DNA degradation, in most cases this approach has advantages when combined with existing monitoring efforts, including operating at a reduced cost. Given that small volumes of water may contain sufficient eDNA to sequence and detect target aquatic organisms (Thomsen et al. 2012; Miya et al. 2015), our study is the first to apply this non-invasive survey method and assess its utility to document the diversity of fishes inhabiting Red Sea coral reefs.

Materials and methods

Study site

68 Sampling was conducted in the central Red Sea over a 3-d period in November 2015 near
69 Thuwal, Saudi Arabia (Fig. 1). Seven sampling sites were located on inshore, midshore, and
70 offshore reefs, with six replicate seawater samples (500 mL per sample) taken from both the
71 exposed and sheltered sides of each reef (a total of 12 replicates from each reef). Four
72 replicate samples were also taken from the King Abdullah University of Science and
73 Technology (KAUST) boat harbor, where there are relatively fewer reef-fish species.
74 Multiple precautions were taken to minimize contamination with exogenous eDNA,
75 including rinsing all water bottles and the filtering apparatus in a 10% bleach solution prior to
76 use and between each sampling. Moreover, seawater was collected only at the surface (top 20
77 cm), on the windward side of the research vessel, and within 5 m of the reef, changing latex
78 gloves in between samples. To minimize eDNA degradation, the samples were filtered
79 immediately on the research vessel using a peristaltic pump (Sentino microbiology pump,
80 Pall Life Sciences, MI) and 47 mm (0.45 µm) micronylon sterilized membranes (Pall Life
81 Sciences), which were then placed in individual plastic bags and stored on ice prior to being
82 frozen at –20 °C until eDNA extraction. All extractions were performed within 5 d of
83 collection to limit the decay of template material on the membranes.
84

85 **DNA extraction**

86 Genomic DNA was extracted in a sterile environment at KAUST using a DNeasy blood and
87 tissue kit (Qiagen Inc., CA) following the manufacturer's protocol with a few modifications:
88 (1) 360 µL of ATL tissue lysis buffer and 40 µL proteinase K was added to 1/3 of the filter
89 membrane that had been cut up in a UV hood; (2) the filter membrane was digested for 3 h;
90 (3) the supernatant following digestion was removed and transferred to a clean 2-mL tube
91 where 400 µL of absolute ethanol and 400 µL of AL lysis buffer was added; (4) two separate
92 microcentrifuge spins with the digested solution were used to pass all of the DNA through the

93 filter column; (5) all samples were eluted in one step using 100 μL of AE buffer; and (6)
94 extraction controls, for which all steps remained the same except for the addition of the filter
95 membrane, were included for each set of 12 samples.
96
97 **Fusion-tag qPCR**
98 Quantitative PCR (qPCR) experiments were conducted in a separate, dedicated laboratory at
99 Curtin University in Australia where all benches and utensils were routinely cleaned with
100 bleach and UV sterilized. This approach allowed us to: (1) recover fish 16S mitochondrial
101 DNA barcode sequences of ~200 bp from mixed samples with the primers 16SF/D (5'
102 GACCCTATGGAGCTTTAGAC 3') and 16S2R-degenerate (5'
103 CGCTGTTATCCCTADRGTAACT 3') (modified from Deagle et al. 2007); (2) add multiple
104 forward ($N = 8$) and reverse ($N = 25$) fusion tags to the qPCR products that each contained a
105 unique 10 bp Illumina adaptor sequence; and (3) estimate final concentrations of the DNA
106 amplicons using SYBR Green as a reporter of fluorescence. All qPCRs were run in triplicate
107 and included "no template" controls to check for sample cross-contamination. qPCR
108 reactions (25 mL) consisted of the following: 10x PCR Buffer (Bioline, London, U.K.), 0.25
109 mM $MgCl_2$, 10 mM dNTPs, 0.4 mg mL^{-1} bovine serum albumin, 1.25 U Ampli-Taq Gold
110 (Applied Biosystems, CA), 20 mM primers, and 2 μL undiluted DNA extract. Cycling
111 conditions were as follows: 95 °C for 5 min, followed by 45 cycles of 95 °C for 30 s, 54 °C
112 for 30 s, and 72 °C for 45 s, with a final step of 72 °C for 10 min. Failed amplifications were
113 not repeated, resulting in fewer than 12 replicates per sampling station. DNA amplicons from
114 each sample were pooled in equimolar concentration based on qPCR endpoints and
115 quantification on a LabChip GX Touch (PerkinElmer Health Sciences, MA), size-selected
116 using a Pippen Prep (Sage Science, MA), and then purified using a QIAquick PCR
117 Purification Kit (Qiagen Inc., CA) following the manufacturer's protocol. The final library

118 was quantified on a LabChip GX Touch followed by sequencing in the TrEnD Laboratory at
119 Curtin University on an Illumina MiSeq platform using 300 cycle MiSeq V2 reagent kits and
120 nano flow cells. All unfiltered sequence data is accessible from DRYAD
121 (doi:10.5061/dryad.1pm20).
122

123 **Bioinformatic filtering**

124 Sequences were assigned to samples, quality filtered, and trimmed in Geneious Pro v 4.8.4
125 (Drummond et al. 2009) using the following workflow: (1) reads containing imperfect
126 Illlumina adaptor sequences were discarded; (2) reads containing imperfect forward or
127 reverse fish 16S primer sequences were discarded; (3) reads were de-multiplexed based on
128 the combined forward and reverse MID sequences; (4) reads with imperfect MID barcode
129 sequences were discarded; (5) reads smaller than 160 bp were discarded; (6) singletons were
130 discarded; and (7) chimeric sequences were flagged and discarded using USEARCH v
131 8.0.1623 (Edgar 2010).
132

133 **Taxonomic assignment**

134 We interrogated the NCBI BLASTn database (March 2016) on the Magnus Cray XC40
135 system located at the Pawsey Supercomputing Centre at Technology Park in Western
136 Australia, with our 16S sequences in FASTA format. BLASTn results were imported into
137 MEtaGenome ANalyzer (MEGAN) v 5.11.3 (Huson and Weber 2013), and taxonomic
138 identities were assigned to genus or species (where possible) based on the lowest common
139 ancestor (LCA) algorithm using the following settings: minimum number of reads = 2;
140 minimum bit score = 300; top percent to be considered = 10%; and minimum complexity =
141 not considered. A similarity of 95% was accepted for a genus-level match and 98% for a
142 species-level match (see also Meyer and Paulay 2005). Taxonomic assignments to species

143 with LCA were further evaluated against expert knowledge of species distributions and the
144 most up-to-date checklist of Red Sea fishes (DiBattista et al. 2016b and references therein).
145 To be conservative, matches at the species level were not accepted if that taxon was not
146 known from the Red Sea or if 16S sequences from all species within that genus were not
147 available on NCBI or in our custom genetic database.
148 Given multiple hits to poorly archived fish 16S sequences on NCBI, an issue raised by
149 Pleijel et al. (2008), in addition to fish species not found in the Red Sea (16 of the 25 species
150 assigned), a modified BLASTn search was performed against a curated custom genetic
151 database of Red Sea fishes, within which we selected a single representative 16S sequence
152 from all reef-fish species found in the Red Sea that were available on NCBI. Where possible,
153 sequences were extracted from datasets published by trusted sources and those that had
154 vouchered their whole specimens at museums (Electronic supplementary material, ESM, File
155 S1). We additionally generated 16S sequences for 38 fish species sampled in the Red Sea to
156 add to the custom database, which provided a total of 474 of the 1071 known reef-fish
157 species from the region (DiBattista et al. 2016b).
158

159 Results and discussion

160 Our 16S eDNA metabarcoding approach at seven reef sites and a boat harbor in the central
161 Red Sea revealed between 26 to 46 genera of fish (out of 511 known genera; DiBattista et al.
162 2016b) based on 250,145 total DNA sequences. These genera represent diverse conspicuous
163 (e.g., *Chaetodon*), cryptobenthic (e.g., *Eviota*), numerically abundant (e.g., *Pseudanthias*),
164 and commercially relevant shorefish (e.g., *Cephalopholis*, *Epinephelus*, *Lethrinus*, and
165 *Lutjanus*) known to inhabit the Red Sea (DiBattista et al. 2016b). Our eDNA assignment
166 approach using the NCBI database and our custom genetic database identified 12 or 6 out of
167 the 21 most abundant conspicuous reef-fish genera identified by a recent UVC study (Roberts

168 et al. 2016), respectively, and between 17% to 42% of these genera were only detected at a
169 single site. This difference in detection capability may be due to the limited spatial scale of
170 our sampling (<1% of the eastern Red Sea coastline) or our lack of ability to resolve species
171 based on short but amplifiable 16S fragments of DNA.

172 DNA sequencing was successful for 53 of the 88 replicate seawater samples collected at
173 reefs in the central Red Sea (Fig. 1). Of these replicates, there was a modest bias towards
174 amplification success on the sheltered side versus the exposed side of the reefs (31 vs 22),
175 which may indicate increased suspension of organic material or water residency time in the
176 former; sampling effort should be focused here in future studies. However, given the lack of
177 data for the exposed side of the reefs, we combined data from both sides of the reef for all
178 downstream analyses. DNA degradation, primer biases, or a lack of template DNA may have
179 all played a role in the samples that failed to amplify. Moreover, for eDNA surveys such as
180 this, it is difficult to know a priori what the DNA template concentration will be in the water
181 column. We recovered limited amounts of template, a finding that contrasts with other studies
182 (Miya et al. 2015), but is important to consider when designing eDNA surveys in an
183 environment for the first time.

184 Our non-invasive survey method revealed a trend in sampling effort at some (but not all)
185 reefs based on the number of cumulative fish genera assigned per replicate (Fig. 2a). For
186 example, no new genera were assigned after as little as one replicate out of four total replicate
187 seawater samples taken from KAUST harbor, whereas new genera continued to be assigned
188 for two reefs that had at least eight or nine replicates included in the analysis. Indeed,
189 taxonomic assignment to genus had not yet reached saturation at four of the eight reefs based
190 on the upward trajectory of the accumulation curves. This suggests that 12 or more replicates,
191 based on the volume of seawater that we filtered in this study (500 mL), may be required. In
192 contrast, there was no clear bias in this “saturation trend” among inshore, midshore, or

193 offshore reefs. No relationship was observed between the number of 16S sequences generated
194 per reef and the number of genera assigned ($r^2 = 0.037$, $p = 0.17$**;** Fig. 2b), although when we
195 removed two outlier data points that appeared to drive the regression (>2 SD, both from Al
196 Fahal South), this relationship became significant ($r^2 = 0.45$, $p < 0.0001$), indicating a clear
197 benefit of more replicates and increased sequencing coverage per replicate, also highlighted
198 by rarefaction analysis of randomly selected samples (ESM Fig. S1).
199 When the number of unique sequences per genus was averaged across each reef we
200 identified over-represented genera (e.g., *Chlorurus*, *Halichoeres*, *Scarus*) at some, but not all
201 reefs based on assignments made using the NCBI database (Fig. 3a) and our custom genetic
202 database (Fig. 3b). This may indicate biases in the workflow that preferentially target these
203 genera or increased DNA concentrations of these genera present in seawater samples.
204 Moreover, some of the common genera in the Red Sea were notably absent from our data
205 (e.g., *Apogon, Coris, Gobiodon, Pseudocheilinus, Pygoplites*; Roberts et al. 2016). Low
206 values for the average number of unique sequences in most cases (Fig. 3a: mean = 8.08 ±
207 3.65 SE, median = 0.67; Fig. 3b: mean = 5.02 ± 2.65, median = 0.65) further suggests that
208 there may be a clear benefit of increased coverage per reef by either increasing the number of
209 replicates (>12 replicates), increasing the volume of water filtered (>500 mL), decreasing the
210 pore size of the filter membrane to capture smaller particles (<0.45 µm), or generating higher
211 numbers of metabarcoding reads. The benefit of generating higher numbers of metabarcoding
212 reads is supported by our rarefaction analysis of randomly selected samples from a single
213 reef, where there was considerable variability in the level of sequencing at which saturation
214 was reached for assignment of genera (ESM Fig. S1). Based on heatmaps (Fig. 3), there was
215 also an indication that different reefs had different compositions of fish. Although previous
216 aquatic studies have attempted to infer the abundance of taxa from the number of sequence
217 reads (amphibians: Pilliod et al. 2013; fish: Mahon et al. 2013; Lacoursière-Roussel et al.

218 2016), we chose to avoid this inference due to confounding factors potentially skewing the
219 proportions of reads, including fluctuating environmental conditions, variable rates of DNA
220 degradation (Dejean et al. 2011), and low template numbers (Murray et al. 2015).

221 Even with our custom genetic database, less than half of all Red Sea species were
222 represented with a 16S barcode sequence (474 of the 1071 species). Indeed, some speciose
223 genera were under-represented in our custom database, with ≤10% of the species in their
224 respective genera with a 16S barcode sequence (ESM File S1, Fig. S2). Nonetheless, almost
225 twice as many taxa were confidently assigned to species using our custom database (N = 16)
226 versus the NCBI database (N = 9), despite the same conservative criteria being used in both
227 cases (Table 1).

228 Caveats to consider when barcoding seawater samples based on 16S fish DNA include
229 misidentification in the NCBI reference database, incomplete lineage sorting or hybridization
230 between species, or (in our case) a dearth of vouchered sequences from reef fish sampled in
231 the Red Sea. Also, because 16S is a much shorter read for eDNA applications, the taxonomic
232 resolution at the species level is weak, particularly in the Red Sea. Indeed, we have
233 previously shown that intraspecific divergence of reef fish from adjacent Indian Ocean
234 populations can be comparable to interspecific comparisons (DiBattista et al. 2013) and that
235 cryptic lineages are not uncommon (DiBattista et al. 2017). These caveats are not unique to
236 our study or restricted to reef fish, but can be mitigated with increased regional barcoding
237 initiatives, and may be less of an issue for better-characterized coral reefs (e.g., Great Barrier
238 Reef). The benefits of this approach despite its equivocal outcome is that it may prove to be
239 more cost-effective than expensive field time, and also that eDNA studies are not limited by
240 poor environmental conditions (e.g., visibility, hazardous conditions) or a lack of taxonomic
241 expertise. These initial eDNA trials from the Red Sea show that the method has several

242 potential applications but that surveys will need to be carefully designed given the number of

243 potentially confounding factors.

244

245 **Acknowledgements**

246 This study was funded by KAUST award number CRG-1-2012-BER-002 and baseline

247 research funds to M.L.B., as well as a Curtin University Early Career Research Fellowship to

248 J.D.D. The authors would like to acknowledge KAUST's Coastal and Marine Resources Core

249 Lab for logistical support, as well as Matthew Power, Megan Coghlan, and Brendan

250 Chapman for DNA sequencing assistance.

251

252 ## References

253 Ackerman JL, Bellwood DR (2000) Reef fish assemblages: a re-evaluation using enclosed
254 rotenone stations. Mar Ecol Prog Ser 206:227–237
255 Berumen ML, Hoey AS, Bass WH, Bouwmeester J, Catania D, Cochran JEM, Khalil MT,
256 Miyake S, Mughal MR, Spaet JLY, Saenz-Agudelo P (2013) The status of coral reef
257 ecology research in the Red Sea. Coral Reefs 32:737–748
258 Deagle BE, Gales NJ, Evans K, Jarman SN, Robinson S, Trebilco R, Hindell MA (2007)
259 Studying seabird diet through genetic analysis of faeces: a case study on macaroni
260 penguins (*Eudyptes chrysolophus*). PLoS One 2:e831
261 Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, Taberlet P, Miaud C (2011)
262 Persistence of environmental DNA in freshwater ecosystems. PLoS One 6:e23398
263 DiBattista JD, Berumen ML, Gaither MR, Rocha LA, Eble JA, Choat JH, Craig MT,
264 Skillings DJ, Bowen BW (2013) After continents divide: comparative
265 phylogeography of reef fishes from the Red Sea and Indian Ocean. J Biogeogr
266 40:1170–1181
267 DiBattista JD, Choat JH, Gaither MR, Hobbs JP, Lozano-Cortés DF, Myers RF, Paulay G,
268 Rocha LA, Toonen RJ, Westneat M, Berumen ML (2016a) On the origin of endemic
269 species in the Red Sea. J Biogeogr 43:13–30
270 DiBattista JD, Bowen BW, Gaither MR, Hobbs J-PA, Saenz-Agudelo P, Piatek M, Rocha
271 LA, Choat JH, McIlwain J, Priest MA, Sinclair-Taylor TH, Berumen ML (2017)
272 Comparative phylogeography of reef fishes from the Gulf of Aden to the Arabian Sea
273 reveals two cryptic lineages. Coral Reefs 36:625–638
274 DiBattista JD, Roberts M, Bouwmeester J, Bowen BW, Coker DF, Lozano-Cortés DF, Choat
275 JH, Gaither MR, Hobbs JP, Khalil M, Kochzius M, Myers R, Paulay G, Robitzch V,
276 Saenz-Agudelo P, Salas E, Sinclair-Taylor TH, Toonen RJ, Westneat M, Williams S,
277 Berumen ML (2016b) A review of contemporary patterns of endemism for shallow
278 water reef fauna in the Red Sea. J Biogeogr 43:423–439
279 Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer
280 T, Wilson A (2009) Geneious v 4.8, www.geneious.com/

281 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST.
282 Bioinformatics 26:2460–2461
283 Harvey ES, Newman SJ, McLean DL, Cappo M, Meeuwig JJ, Skepper CL (2012)
284 Comparison of the relative efficiencies of stereo-BRUVs and traps for sampling
285 tropical continental shelf demersal fishes. Fish Res 125:108–120
286 Huson DH, Weber N (2012) Microbial community analysis using MEGAN. Methods
287 Enzymol 531:465–485
288 Lacoursière-Roussel A, Côté G, Leclerc V, Bernatchez L (2016) Quantifying relative fish
289 abundance with eDNA: a promising tool for fisheries management. J Appl Ecol
290 53:1148–1157
291 Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples
292 reveal patterns of marine benthic diversity. Proc Natl Acad Sci U S A 112:2076–2081
293 Mahon AR, Jerde CL, Galaska M, Bergner JL, Chadderton WL, Lodge DM, Hunter ME,
294 Nico LG (2013) Validation of eDNA surveillance sensitivity for detection of Asian
295 carps in controlled and field experiments. PLoS One 8:e58316
296 Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling.
297 PLoS Biology 3:e422
298 Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S,
299 Yamanaka H, Araki H, Kondoh M (2015) MiFish, a set of universal PCR primers for
300 metabarcoding environmental DNA from fishes: detection of more than 230
301 subtropical marine species. R Soc Open Sci 2:150088
302 Murray DC, Coghlan ML, Bunce M (2015) From benchtop to desktop: important
303 considerations when designing amplicon sequencing workflows. PLoS One
304 10:e0124671
305 Pilliod DS, Goldberg CS, Arkle RS, Waits LP (2013) Estimating occupancy and abundance
306 of stream amphibians using environmental DNA from filtered water samples. Can J
307 Fish Aquat Sci 70:1123–1130
308 Pleijel F, Jondelius U, Norlinder E, Nygren A, Oxelman B, Schander C, Sundberg P,
309 Thollesson M (2008) Phylogenies without roots? A plea for the use of vouchers in
310 molecular phylogenetic studies. Mol Phylogenet Evol 48:369–371
311 Raitsos DE, Pradhan Y, Brewin RJ, Stenchikov G, Hoteit I (2013) Remote sensing the
312 phytoplankton seasonal succession of the Red Sea. PLoS One 8:e64909
313 Roberts CM, Shepherd ARD, Ormond RF (1992) Large-scale variation in assemblage
314 structure of Red Sea butterflyfishes and angelfishes. J Biogeogr 1992:239–250
315 Roberts MB, Jones GP, McCormick MI, Munday PL, Neale S, Thorrold S, Robitzch VS,
316 Berumen ML (2016) Homogeneity of coral reef communities across 8 degrees of
317 latitude in the Saudi Arabian Red Sea. Mar Pollut Bull 105:558–565
318 Spalding MD, Ravilious C, Green EP (2001) World atlas of coral reefs. University of
319 California Press, Berkeley
320 Thomsen PF, Kielgest J, Iversen LL, Møller PR, Rasmussen M, Willerslev E (2012)
321 Detection of a diverse marine fish fauna using environmental DNA from seawater
322 samples. PLoS One 7:e41732
323

324

325

326

327 ## Figure legends

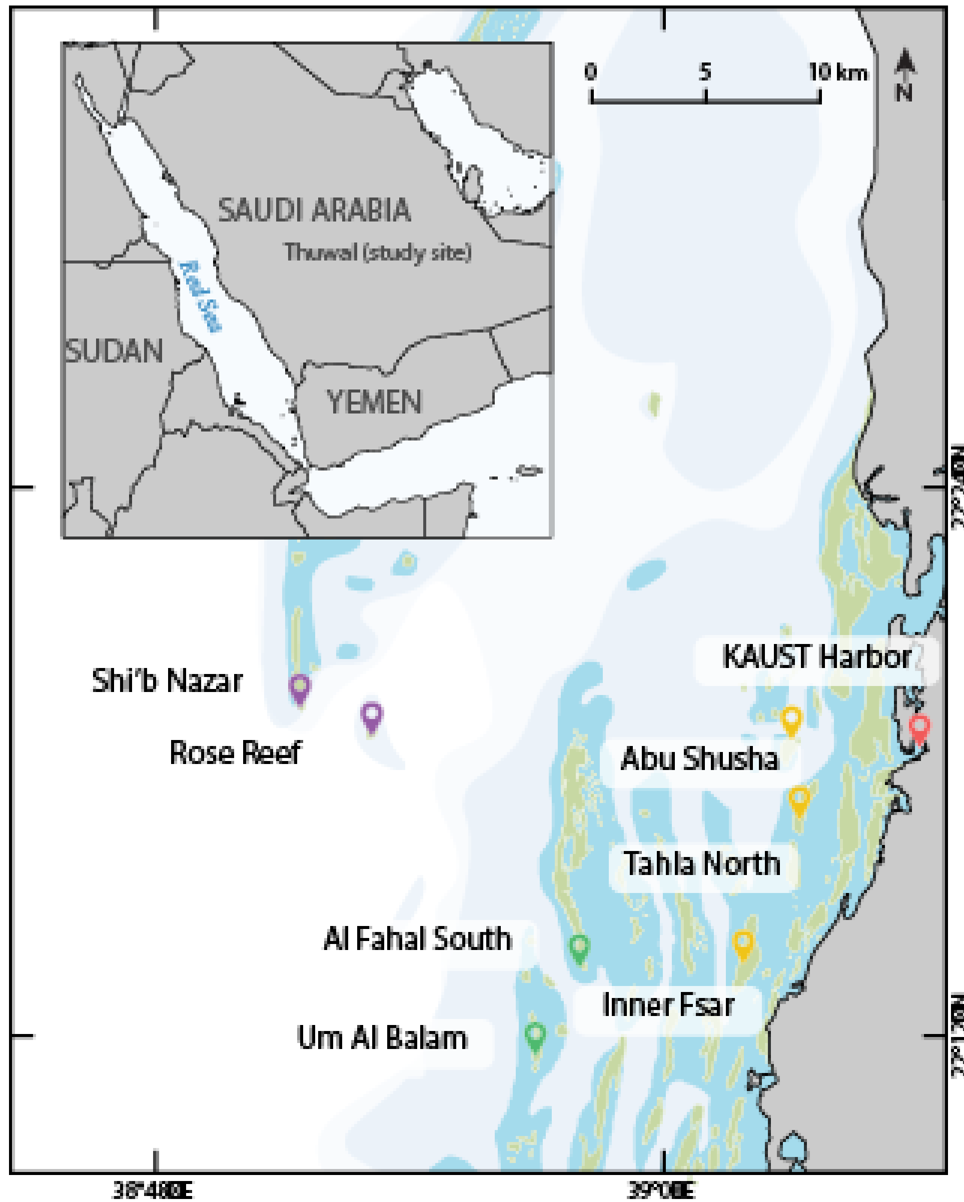
328 **Fig. 1** Seawater sampling sites at inshore (*yellow*), midshore (*green*), and offshore (*purple*)
329 coral reefs, and a boat harbor (*red*) along the central Red Sea coastline near Thuwal, Saudi
330 Arabia

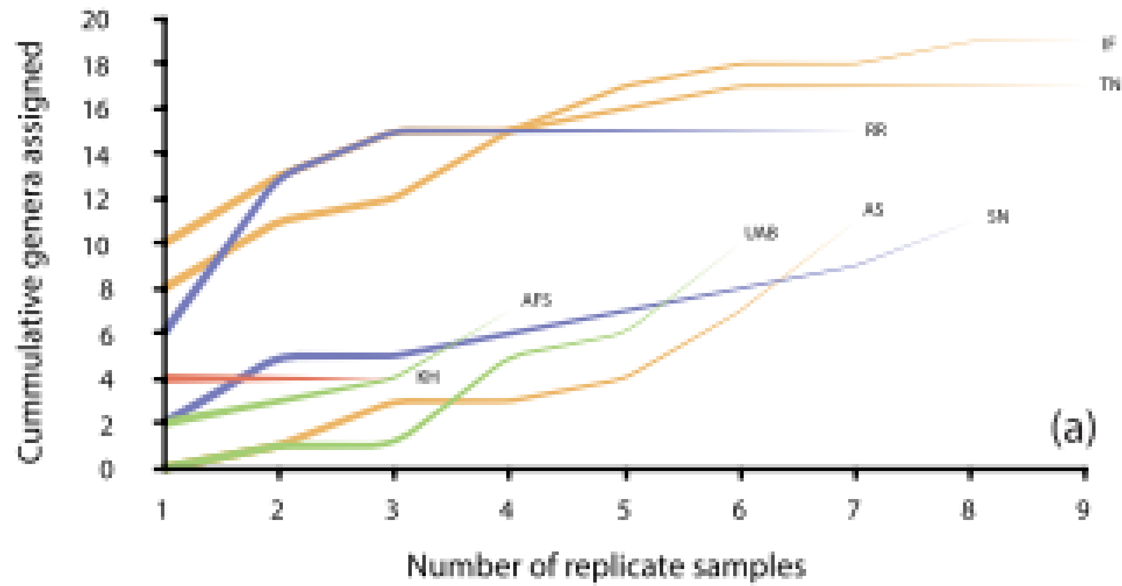
331 **Fig. 2** Number of cumulative (**a**) or absolute (**b**) genera assigned per replicate as a function of
332 amplifiable 16S mitochondrial DNA sequences detected at seven inshore (*yellow*), midshore
333 (*green*), and offshore (*purple*) coral reefs, and a boat harbor (*red*) along the central Red Sea
334 coastline near Thuwal, Saudi Arabia. SN = Shi'b Nazar; RR = Rose Reef; UAB = Um Al
335 Balam; AFS = Al Fahal South; IF = Inner Fsar; TN = Tahla North; AS = Abu Shusha; KH =
336 KAUST Harbor

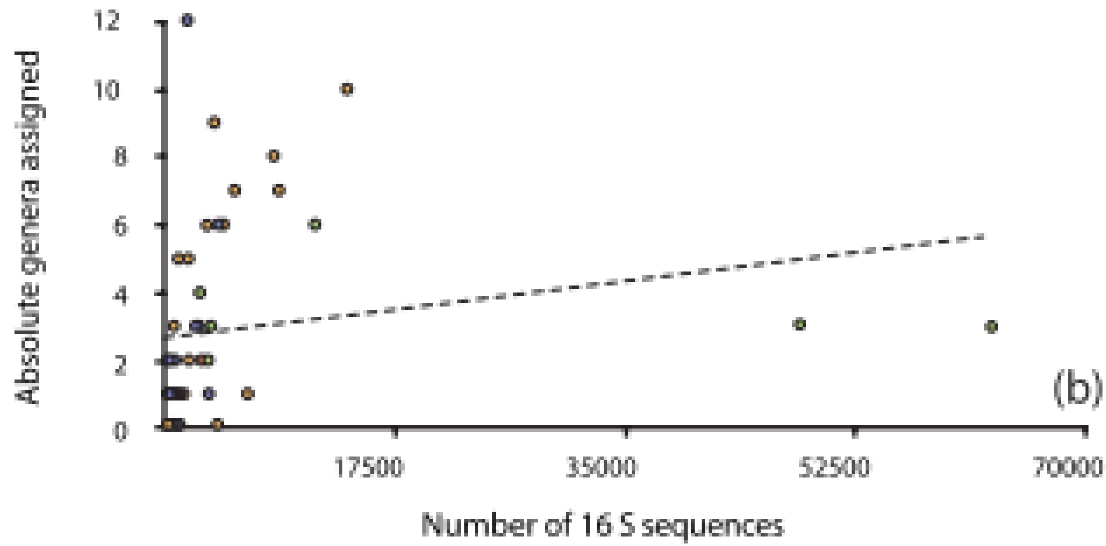
337 **Fig. 3** Heatmap showing the mean number of unique 16S mitochondrial DNA sequences for
338 each genus (**a**, **b**) detected at seven inshore (*yellow*), midshore (*green*), and offshore (*purple*)
339 coral reefs, and a boat harbor (*red*) along the central Red Sea coastline near Thuwal, Saudi
340 Arabia. Analyses were independently conducted using the National Center for Biotechnology
341 Information database (**a**) or our custom genetic database (**b**). *Numbers in parentheses* are the
342 number of seawater replicates that amplified per sample site. See **Fig. 1** for summary of
343 sample sites; abbreviations as in **Fig. 2**

344
345
346
347
348
349
350
351

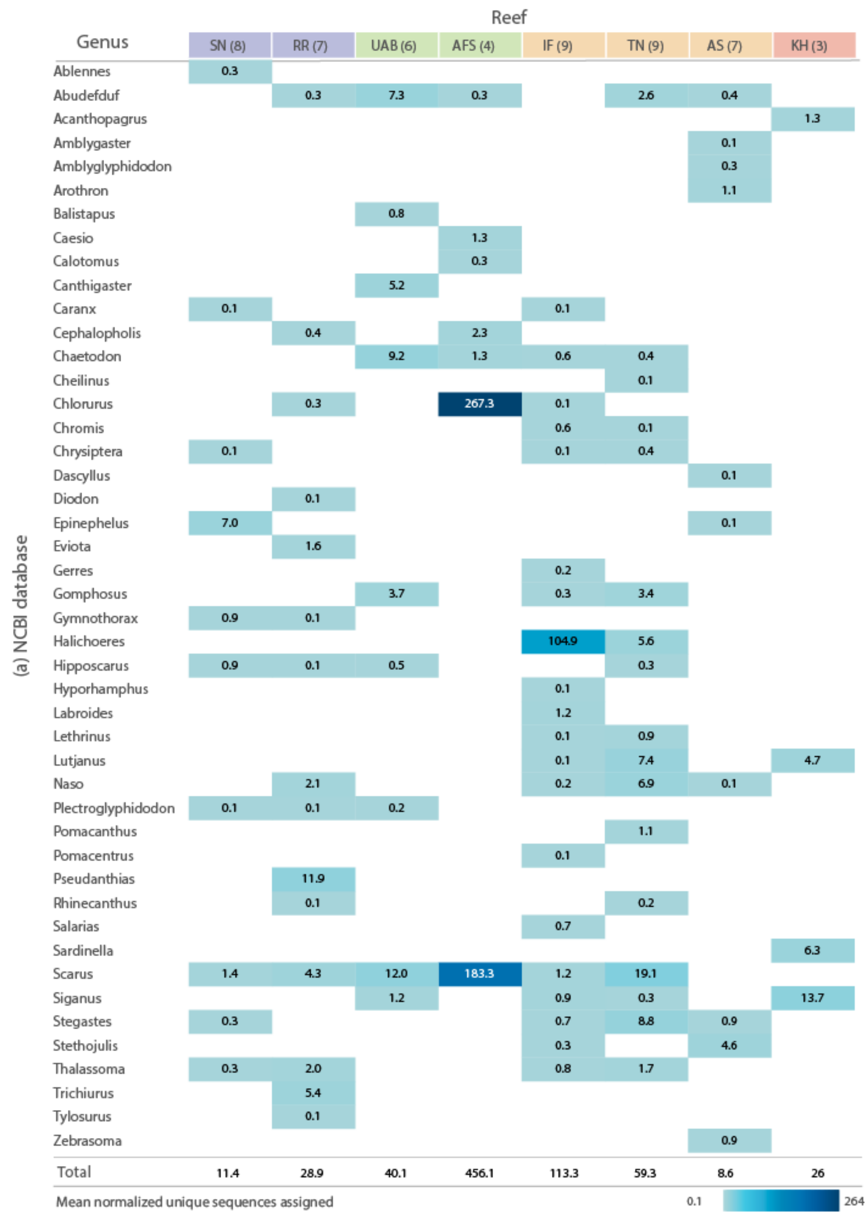
352



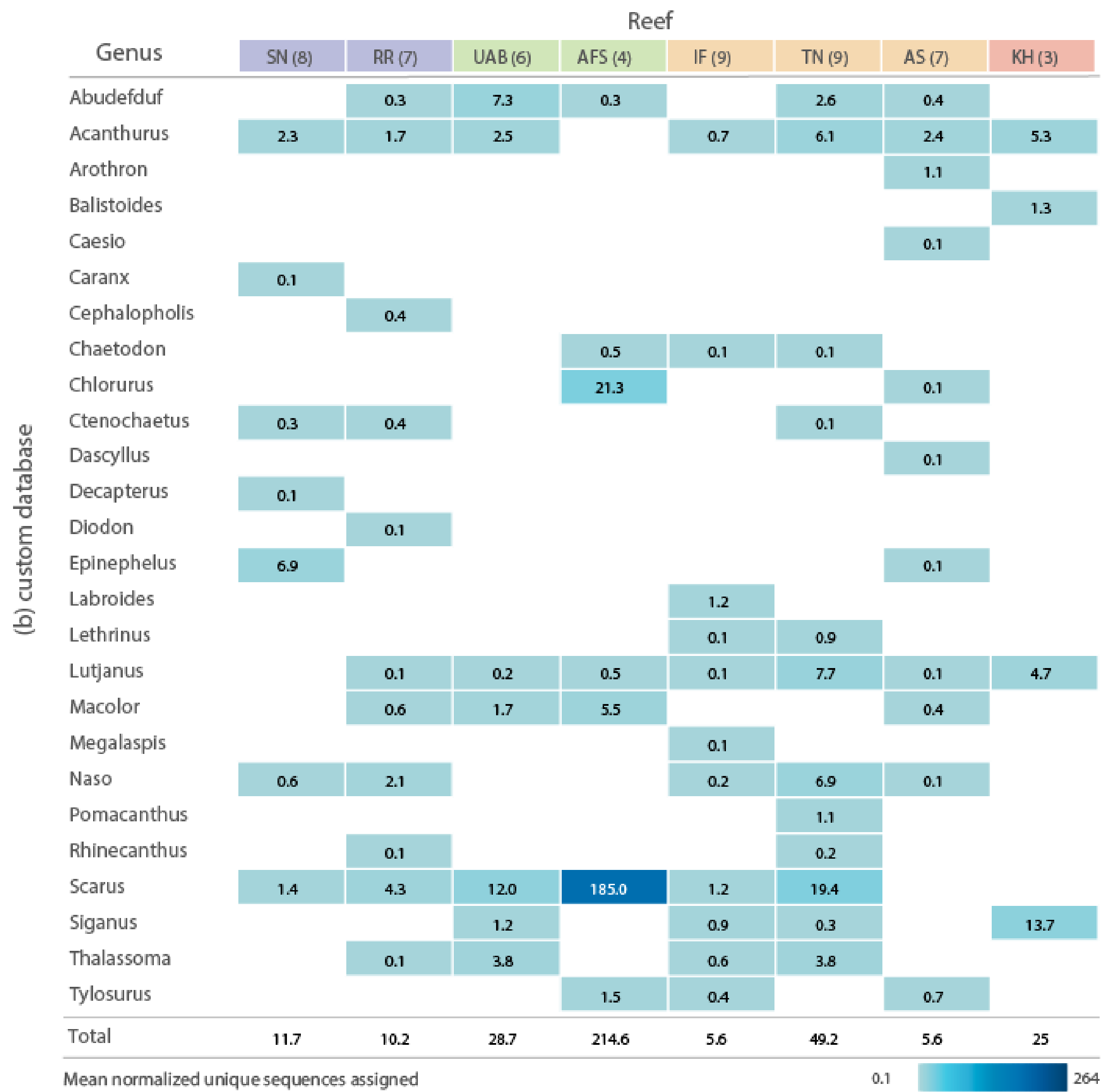




353



354



(b) custom database

Genus	SN (8)	RR (7)	UAB (6)	AFS (4)	IF (9)	TN (9)	AS (7)	KH (3)
	Reef							
Abudefduf		0.3	7.3	0.3		2.6	0.4	
Acanthurus	2.3	1.7	2.5		0.7	6.1	2.4	5.3
Arothron							1.1	
Balistoides								1.3
Caesio							0.1	
Caranx	0.1							
Cephalopholis		0.4						
Chaetodon				0.5	0.1	0.1		
Chlorurus				21.3			0.1	
Ctenochaetus	0.3	0.4				0.1		
Dascyllus							0.1	
Decapterus	0.1							
Diodon		0.1						
Epinephelus	6.9						0.1	
Labroides					1.2			
Lethrinus					0.1	0.9		
Lutjanus		0.1	0.2	0.5	0.1	7.7	0.1	4.7
Macolor		0.6	1.7	5.5			0.4	
Megalaspis					0.1			
Naso	0.6	2.1			0.2	6.9	0.1	
Pomacanthus						1.1		
Rhinecanthus		0.1				0.2		
Scarus	1.4	4.3	12.0	185.0	1.2	19.4		
Siganus			1.2		0.9	0.3		13.7
Thalassoma		0.1	3.8		0.6	3.8		
Tylosurus				1.5	0.4		0.7	
Total	11.7	10.2	28.7	214.6	5.6	49.2	5.6	25

Mean normalized unique sequences assigned 0.1 – 264

355

356 **Table 1** Summary of reef fish detected (presence/absence) at the species-level by seawater sampling at seven coral reefs and a boat harbor along
357 the central Red Sea coastline near Thuwal, Saudi Arabia. Matches were obtained using BLASTn searches in the NCBI mitochondrial DNA 16S
358 database or using a custom 16S database (see Materials and methods). A similarity of 98% was accepted for a species-level match; to be
359 conservative, matches at the species-level were not accepted if that taxon was not known from the Red Sea and matches were not accepted if 16S
360 sequences from all known Red Sea species within that genus were not available in each database, respectively.
361

Family	Species	NCBI 16S database	Custom 16S database
Acanthuridae	*Ctenochaetus striatus* (Quoy & Gaimard 1825)		√
Acanthuridae	*Naso brevirostris* (Cuvier 1829)		√
Acanthuridae	*Naso unicornis* (Forsskål 1775)	√	
Balistidae	*Balistoides viridescens* (Bloch & Schneider 1801)		√
Balistidae	*Balistapus undulatus* (Park 1797)	√	
Balistidae	*Rhinecanthus rectangulus* (Bloch & Schneider 1801)		√
Belonidae	*Ablennes hians* (Valenciennes 1846)	√	
Belonidae	*Tylosurus crocodilus* (Lesueur 1821)	√	√
Blenniidae	*Salarias fasciatus* (Bloch 1786)	√	
Clupeidae	*Amblygaster sirm* (Walbaum 1792)	√	
Diodontidae	*Diodon hystrix* Linnaeus 1758	√	
Diodontidae	*Diodon liturosus* Shaw 1804		√
Labridae	*Labroides dimidiatus* (Valenciennes 1839)	√	√
Labridae	*Scarus collana* Rüppell 1835		√
Labridae	*Thalassoma lunare* (Linnaeus 1758)		√
Labridae	*Thalassoma rueppellii* (Klunzinger 1828)		√
Lutjanidae	*Macolor niger* (Forsskål 1775)		√
Pomacanthidae	*Pomacanthus maculosus* (Forsskål 1775)		√
Pomacentridae	*Abudefduf vaigiensis* (Quoy & Gaimard 1825)		√
Pomacentridae	*Dascyllus trimaculatus* (Rüppell 1829)		√
Siganidae	*Siganus luridus* (Rüppell 1829)		√
Siganidae	*Siganus rivulatus* (Forsskål 1775)		√
Sparidae	*Acanthopagrus bifasciatus* (Forsskål 1775)	√	

362
363
364