

1

2

Establishing a marine monitoring programme to assess antibiotic resistance: a case study from the Gulf Cooperation Council (GCC) region

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

Edel Light^{1,2*}, Craig Baker-Austin¹, Roderick M. Card^{2,3}, David Ryder¹, Mickael Teixeira Alves¹,
Hanan A. Al-Sarawi⁴, Khalil Hasan Abdulla⁵, Henrik Stahl⁶, Aliya AL-Ghabshi⁷, Majed Al Ghoribi⁸,
Hanan H. Balkhy⁹, Andrew Joseph¹, Alexandra Hughes¹, David W. Verner-Jeffreys^{1,2}, Brett P. Lyons¹
and Will J.F. Le Quesne¹⁰.

¹Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Weymouth, Dorset DT4 8UB, United Kingdom.

²UK FAO Reference Centre for Antimicrobial Resistance, Veterinary Medicines Directorate, Woodham Lane, New Haw, Addlestone, Surrey, UK.

³Department of Bacteriology, Animal and Plant Health Agency, Woodham Lane, New Haw, Addlestone, Surrey, UK.

⁴Kuwait Environment Public Authority (KEPA) P.O. Box: 24395 Safat 13104, Kuwait.

⁵Supreme Council for Environment, Manama, Bahrain.

⁶Zayed University, Academic City, 192 82 Dubai, United Arab Emirates.

⁷Al-Khabourah Vocational College for Marine Sciences, Ministry of Higher Education, Scientific Research and Innovation, Oman.

⁸King Abdullah International Medical Research Centre, KAIMRC, Saudi Arabia.

⁹World Health Organisation, Geneva, Switzerland. 20, avenue Appia-1211-Switzerland

¹⁰Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Lowestoft, Suffolk NR33 0HT, United Kingdom.

Key words: AMR, bacteria, GCC, emergence, marine pollution

*Author for correspondence

Edel Light
Centre for Environment, Fisheries and Aquaculture Science (CEFAS),
Weymouth,
Dorset,
DT4 8UB,
United Kingdom
edel.light@cefasc.co.uk
Tel : +44 (0)1305 206728
Fax : +44 (0)1305 206600

41 **Abstract**

42 The World Health Organization considers antimicrobial resistance as one of the most pressing
43 global issues which poses a fundamental threat to human health, development, and security.
44 Due to demographic and environmental factors, the marine environment of the Gulf
45 Cooperation Council (GCC) region may be particularly susceptible to the threat of
46 antimicrobial resistance. However, there is currently little information on the presence of AMR
47 in the GCC marine environment to inform the design of appropriate targeted surveillance
48 activities. The objective of this study was to develop, implement and conduct a rapid regional
49 baseline monitoring survey of the presence of AMR in the GCC marine environment, through
50 the analysis of seawater collected from high-risk areas across four GCC states: (Bahrain,
51 Oman, Kuwait, and the United Arab Emirates). 560 *Escherichia coli* strains were analysed as
52 part of this monitoring programme between December 2018 and May 2019. Multi-drug
53 resistance (resistance to three or more structural classes of antimicrobials) was observed in
54 32.5% of tested isolates. High levels of reduced susceptibility to ampicillin (29.6%), nalidixic
55 acid (27.9%), tetracycline (27.5%), sulfamethoxazole (22.5%) and trimethoprim (22.5%) were
56 observed. Reduced susceptibility to the high priority critically important antimicrobials:
57 azithromycin (9.3%), ceftazidime (12.7%), cefotaxime (12.7%), ciprofloxacin (44.6%),
58 gentamicin (2.7%) and tigecycline (0.5%), was also noted. A subset of 173 isolates was whole
59 genome sequenced, and high carriage rates of *qnrS1* (60/173) and *bla_{CTX-M-15}* (45/173) were
60 observed, correlating with reduced susceptibility to the fluoroquinolones and third generation
61 cephalosporins, respectively. This study is important because of the resistance patterns
62 observed, the demonstrated utility in applying genomic-based approaches to routine
63 microbiological monitoring, and the overall establishment of a transnational AMR surveillance
64 framework focussed on coastal and marine environments.

65

66

67

68

69

70

71

72

73 **Introduction**

74 Globally, antimicrobial resistance (AMR) has emerged as a critically important threat to both
75 animal and human health. The reduction in effectiveness of antimicrobials used to treat
76 pathogenic bacterial infections represents a significant global healthcare problem. Drug
77 resistant infections are rising with estimates suggesting up to 10 million fatalities each year by
78 2050 due to antibiotic-resistant infections in Europe and the US alone (O'Neill, 2016), with an
79 annual global shortfall in GDP predicted of between \$1 trillion and \$3.4 trillion by 2030
80 (World Bank, 2017). While many studies have addressed the role of AMR in clinical settings,
81 there is growing interest in the role of the natural environment in modulating risks associated
82 with AMR. Indeed, the growing concern over the clinical threat posed by antibiotic resistant
83 bacteria to human health has finally turned attention to the environmental dimensions of the
84 problem (Karkman *et al.*, 2019, Wellington *et al.*, 2013). Aquatic and marine environments,
85 which receive significant inputs of pharmaceuticals, sewage waste, AMR bacteria, their
86 associated resistance genes as well as mobile genetic elements harbouring AMR have become
87 the focus of scientific interest (Taylor *et al.*, 2011). Discharge of faecal sources into the aquatic
88 environment, via wastewater treatment works, illegal discharges, or agricultural run-off, are
89 key pathways for the input of antimicrobial products, resistance genes and bacteria into marine
90 and aquatic environments (Baquero *et al.*, 2008). In addition, treated or untreated sewage
91 discharges are known to contain a diverse array of anthropogenic pollutants that have been
92 shown to co-select for AMR (Amos *et al.*, 2014, Baker-Austin *et al.*, 2006). The analysis of
93 environmental waters is being adopted as an effective method to monitor the dynamics of
94 antibiotic-resistant pathogens (Amos *et al.*, 2014), particularly given the clear route back into
95 human populations that this environmental niche represents. It has become increasingly
96 recognized that the ability to scrutinize the prevalence, as well as the characteristics, of
97 resistance in environmental bacteria is critical, as part of a One Health approach, to be able to
98 fully understand the AMR situation in clinical and community settings (Walsh, 2018).

99 To date, much of the data underpinning AMR risk from environmental studies has been
100 obtained in the USA and Europe, with few published studies regarding health risks gathered
101 elsewhere. Due to specific demographic and environmental factors the Gulf Cooperation
102 Council (GCC) region may be particularly susceptible to the threat of AMR, with the marine
103 and aquatic environment potentially playing a specific role in its development and propagation
104 (Le Quesne *et al.*, 2018). The Gulf is a shallow sea naturally exposed to extreme conditions of

105 temperature and salinity due to its location, semi-enclosed nature, bathymetry and restricted
106 circulation (Sheppard *et al.*, 2010, Khatir *et al.*, 2020) (Figure 1). As such, pollutants such as
107 sewage contamination, pharmaceuticals and heavy metals entering this system may exert
108 additional selection pressures. The region has recently undergone rapid change; the population
109 of the GCC has increased in size by 17 million over the last 2 decades, largely attributed to an
110 influx of foreign labor, which in many Gulf states out number nationals (Abdul Salam *et al.*,
111 2015). This rapidly increasing and geographically diverse mixing of the GCC population,
112 provides the potential for AMR pressures to have intensified. Furthermore, studies
113 investigating the drivers for AMR development in the GCC suggest that there is a high rate of
114 inappropriate antibiotic use (Al-Yamani *et al.*, 2016; Balkhy *et al.*, 2016; Butt *et al.*, 2017)
115 leading to recommendations calling for the implementation of regional guidelines for the
116 management of common bacterial infections (Balkhair *et al.*, 2014; Al-Yamani *et al.*, 2016).
117 Because the Gulf is also of societal and economic importance to countries bordering its waters,
118 providing food, and drinking water (via desalination) along with supporting tourism and
119 recreational activities, there is a clear need for a greater understanding of AMR risk in the
120 region, particularly from an environmental context.

121 In 2018 we published a discussion article outlining a conceptual framework to more fully
122 understand the risks associated with AMR in the marine and aquatic environment in the GCC
123 region (Le Quesne *et al.*, 2018). As part of this discussion paper, we proposed a potential
124 methodological approach to more fully understand how such environments interact with the
125 wider development and propagation of resistance, including a collaborative multi-national
126 study plan for the region. We focused our investigations on *Escherichia coli*. This species is
127 regularly used as a marker of faecal contamination from both humans and animals and has
128 been recommended for use as an indicator for the surveillance of antimicrobial resistance
129 (AMR) in the environment (Anjum *et al.*, 2021). It is already part of established human and
130 livestock surveillance programmes, with verified methods for its isolation and characterisation.
131 Reference laboratories can easily implement *E. coli* in AMR environmental monitoring
132 programmes (Anjum *et al.*, 2021). As regards marine monitoring, Alves *et al.*, 2014, have
133 shown that *E. coli* can survive in coastal waters. It is relatively robust and easy to grow and
134 there are existing antimicrobial sensitivity test (AST) protocols along with interpretive criteria
135 for many antibiotics, including high priority critically important antimicrobials (HP-CIA,
136 WHO 2019). Here we present the findings of a coordinated multinational monitoring program
137 to assess the prevalence of AMR across the marine waters of the Gulf. Results from this

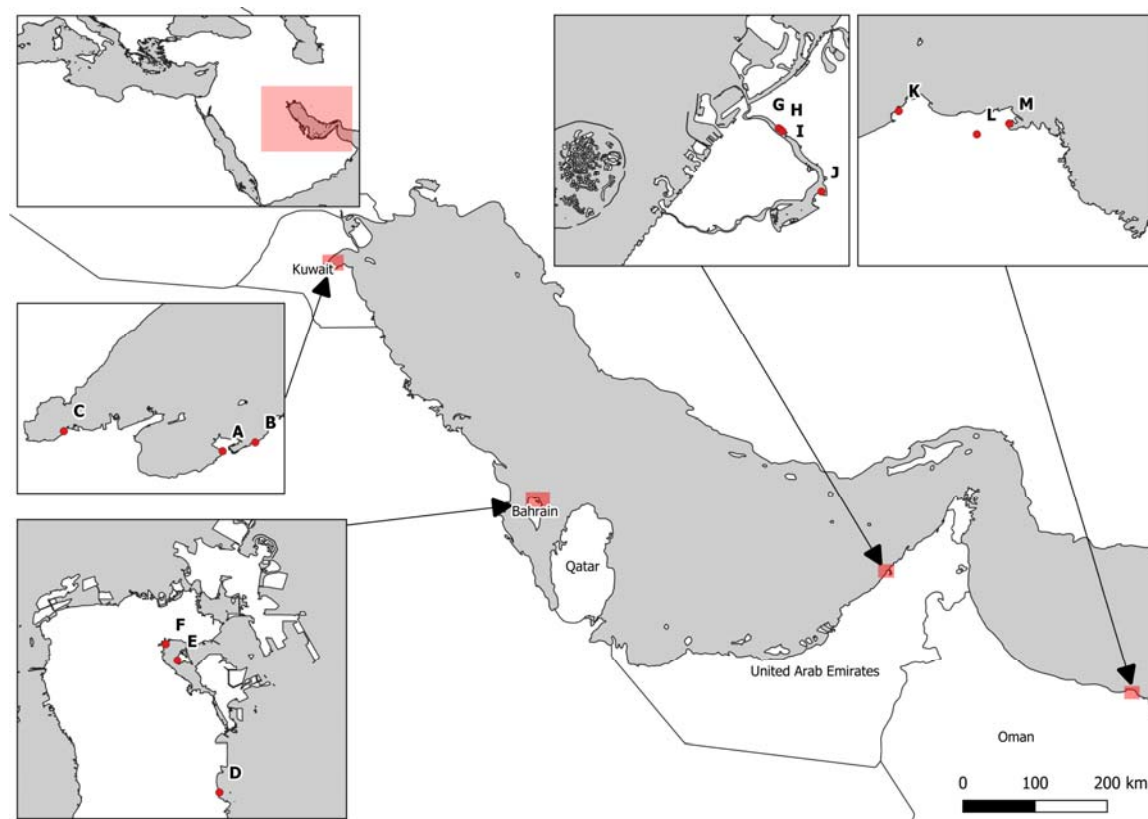
138 regional baseline survey of the presence of AMR in the GCC marine environment, through the
139 analysis of *E. coli* isolated from seawater collected from high-risk areas across four GCC
140 states, are presented along with potential next steps for long-term implementation.

141

142 **Materials and methods**

143 *Sample site choice and sampling plan.* Representatives from the six GCC states of Bahrain,
144 Oman, Kuwait, United Arab Emirates, Saudi Arabia, and Qatar were contacted in the Autumn
145 of 2018 for inclusion in the study. It was agreed from the outset that National partners would
146 collect water samples from at least three separate coastal locations, two highly impacted
147 locations near to (any) known sewage inputs, and one sampling location in an area used for
148 recreational activity, but close to potential sewage inputs (Figure 1). To assist in regional
149 coordination and consistency between sampling campaigns a study plan was circulated to all
150 partners detailing sampling methodology and initial sample processing (Supplementary File 1).
151 Water filtration and bacterial isolation took place at laboratories in each GCC state, and
152 isolates returned to Cefas for downstream analysis. Where project partners did not have the
153 capacity to process samples beyond filtration, the original culture media plates were sent to
154 Cefas laboratories for processing and analysis.

155



156

157 *Figure 1.* Map showing sampling locations from the different Gulf States involved in the study.

158

159 *Water sampling.* At least 200 mL of water was collected from three sites in Bahrain and
160 Kuwait and four sites in Oman and the United Arab Emirates (UAE) between December 2018
161 and May 2019. Two samples were taken from Al Qurum, Oman. Four samples were taken
162 from Dubai Creek, UAE. Due to small sample sizes the data from these sites was combined
163 giving a total of three sample sites from Oman and one from UAE. It was not possible to
164 include samples from either Saudi Arabia or Qatar due to time constraints on the study. Water
165 samples were homogenized and kept at < 10 °C between collection and processing. Data
166 regarding water temperature, salinity and other pertinent observations were collected during
167 sampling, (Supplementary File 2).

168

169 *Microbiological analysis.* For the detection and isolation of *E. coli* through membrane
170 filtration, the following procedure was adopted: 10 mL of sample was serially diluted in ¼
171 strength Ringers Solution (10^0 , 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , and 10^{-5}) and filtered through 47 mm 0.45
172 µm nitrocellulose membrane; the membrane was placed on a plate of Tryptone Bile X-

173 Glucuronide Medium, TBX agar (Oxoid), and incubated at 30 °C for 4 hours followed by 44
174 °C for between 17 and 21.5 hours, one laboratory omitted the 30 °C incubation step; finally,
175 the number of positive (blue-green) colonies on each plate was counted (Vergine *et al.*, 2017)
176 and total *E. coli* enumerated. Blue-green colonies were picked off the plates using sterile loops,
177 inoculated onto nutrient agar slants (Oxoid), and incubated overnight at 44 °C. The slants were
178 stored at 2 - 8 °C before shipping to Cefas (Weymouth, UK) for further microbiological and
179 AMR analyses. The combination of culture medium and incubation temperature was
180 deliberately chosen, for health and safety reasons, to exclude the isolation of toxigenic *E. coli*.
181 The study plan recommended 50 *E. coli* isolates from each location. However, that was not
182 always possible. In locations with a low *E. coli* count all blue-green colonies were selected. In
183 areas with a high *E. coli* count discreet colonies were picked *ad hoc* from the plates. Total
184 numbers were made up by utilising more colonies from areas with a high *E. coli* count.
185 Differences in number between locations were accounted for during statistical analysis.

186

187 *Antimicrobial susceptibility testing.*

188 Upon receipt at Cefas, isolates were removed from slopes onto TBX plates and incubated
189 overnight at 44 °C. Cultures were purified using further subculture steps, before storage on
190 Protect™ beads at -80 °C. Some cultures could not be revived. Numbers of cultures refer to
191 those which were viable and not total isolates received. The control strain *E. coli* ATCC 25922
192 was used throughout. All *E. coli* isolates, were tested for phenotypic resistance by broth
193 microdilution using the Trek Sensititre system (Thermofisher), following manufacturer's
194 instructions. All isolates were tested on the EU surveillance EUVSEC plate for susceptibility
195 towards 14 antibiotics: Ampicillin, Azithromycin, Cefotaxime, Ceftazidime, Chloramphenicol,
196 Ciprofloxacin, Colistin, Gentamicin, Meropenem, Nalidixic acid, Sulfamethoxazole,
197 Tetracycline, Tigecycline and Trimethoprim. The European Committee on Antimicrobial
198 Susceptibility Testing (EUCAST) epidemiological cut-off values (ECOFF) were used for
199 interpretation of wild type vs. non-wild type (WT/NWT) where wild type bacteria possess no
200 acquired resistance mechanisms, and non-wild type bacteria have an acquired resistance
201 mechanism (Schwarz *et al.*, 2010). All isolates showing reduced susceptibility to Cefotaxime
202 and/or Ceftazidime (3rd generation cephalosporins) were tested on the EU surveillance
203 extended spectrum beta lactamase (ESBL) EUVSEC2 plate to determine susceptibility to:
204 Cefoxitin, Ertapenem, Imipenem, Meropenem, Cefotaxime, Ceftazidime, Cefepime,
205 Cefotaxime/Clavulanic acid, Ceftazidime/Clavulanic acid, and Temocillin. This plate is used to

206 elicit antibiotic resistance phenotypes such as: Extended spectrum beta lactamase ESBL,
207 ampicillinase C AmpC, ESBL+AmpC and putative carbapenamase producers, using the criteria
208 published in the technical specifications on harmonised monitoring of antimicrobial resistance
209 in zoonotic and indicator bacteria from food-producing animals and food (EFSA, 2019).
210 Multidrug resistance was defined as any isolate showing the NWT phenotype to three or more
211 classes of antibiotic, EFSA 2020. Throughout the paper any mention of ‘resistance’ as in
212 multidrug resistance MDR, phenotypic resistance etc. refers to the non-wild type phenotype,
213 which is not necessarily synonymous with clinical resistance.

214

215 *Statistics*

216 Pearson’s correlation coefficients were calculated for identifying associations between
217 countries and locations. To assess differences between locations within countries a linear
218 mixed model was conducted considering paired measurements between individual antibiotics,
219 MDR, ESBL and AmpC. Normality of the residuals was verified. A pairwise posthoc test was
220 conducted for estimating locations that differed from each other.

221 Cohen’s Kappa test was used to compare MIC with the presence of resistance genes for 8 of
222 the antimicrobials. We used the interpretation as described by Stubberfield *et al.*, 2019 where a
223 k value of 0 indicates a test that agrees as well as could be expected, compared to a k of 1
224 which indicates complete agreement. A result of $k > 0.900$ was interpreted as almost perfect
225 agreement, $k 0.800 - 0.899$ as strong agreement, $k 0.600 - 0.799$ as moderate agreement, k
226 $0.400 - 0.599$ as weak agreement and $k 0.200 - 0.399$ as minimal agreement.

227

228 *Whole genome sequencing.*

229 A subset of 173 isolates was chosen for whole genome sequencing, to provide insight into the
230 genetic diversity of the isolates and the AMR resistance determinants they harboured. The
231 selection was made to encompass isolates with resistance to HP CIAs and/or having an MDR
232 phenotype. Isolates from all four countries were included in the selection. Briefly, genomic
233 DNA was extracted from 173 strains following the QIAamp DNA Mini kit protocol. Libraries
234 were prepared with the Nextera XT DNA sample preparation kit from Illumina and sequenced
235 using two lanes on a NovaSeq 6000, with 150 cycles of paired-end sequencing.

236

237 *Quality Control.* FastQC version 0.11.8 and MultiQC version 1.7 were used for assessing the
238 quality of raw sequencing data (Andrews *et al.*, 2016). KAT version 2.4.1 was used to compare
239 27bp kmers from sequences which were assembled, as described below, and raw sequencing
240 data (Mapleson *et al.*, 2017). Any samples with two or more coverage peaks or which had a
241 large number of distinct, high coverage kmers which aligned against multiple, different parts of
242 the assembly were identified as having a large degree of heterogeneity, presumed due to
243 inclusion of multiple bacterial isolates during library preparation of a single sample, and were
244 therefore excluded. Isolates were also excluded due to poor coverage or assembly of
245 sequencing data, as assessed using the Nullarbor pipeline.

246

247 *Analysis of Bacterial Isolates using Nullarbor.* Version 2 of the Nullarbor pipeline was used to
248 generate a report on each bacterial isolate (Seemann *et al.*, 2020), which includes steps for
249 cleaning reads, species identification, *de novo* assembly, multi-locus sequence typing and
250 calling Single Nucleotide Polymorphisms (SNPs). Removal of adaptors and low-quality bases
251 or reads was carried out using version 0.39 of Trimmomatic (Bolger *et al.*, 2014). Each
252 bacterial isolate was assigned to the closest matching species using version 2.0.8 of the kraken2
253 software (Wood *et al.*, 2019), which used version 1 of the minikraken2 database, as released in
254 April 2019, as a reference. Reads were assembled into a draft genome using SKESA version
255 2.3.0 (Souvorov *et al.*, 2018). The sequence type of each bacterial isolate was identified using
256 version 2.16.1 of the MLST software (Seemann, 2020a), parameterised to search against the *E.*
257 *coli* sequence typing scheme. The MLST software in turn makes use of the PubMLST website
258 (<https://pubmlst.org/>) developed for microbial MLST designations (Jolley and Maiden, 2010).
259 Snippy version 4.3.6 was used to call SNPs (Seemann, 2020b), with the published genome of
260 *E. coli* strain K-12 sub-strain MG1655 being used as a reference (accession number
261 GCA_000005845.2). The core SNPs, as identified by snippy, were used to infer phylogeny,
262 using version 1.6.10 of IQ-tree (Nguyen *et al.*, 2015), which was parameterised to use the
263 maximum-likelihood general time reversible model with unequal rates and base frequencies as
264 well as four rate categories and the fast tree search mode.

265

266 *Identification of resistance genes using ResFinder.* Version 4.0.0 of the ResFinder software
267 was downloaded and installed on the 11th of November 2020, along with the latest version of
268 the resfinder and pointfinder database (Bortolaia *et al.*, 2020; Zankari *et al.*, 2013). Results

269 where coverage and identity scores were equal to one were reported as normal. Genes
270 identified in isolates with a more fragmented genome assembly, where there was incomplete
271 coverage of the gene coding region, were only reported as belonging to a specific gene family,
272 rather than being reported as a specific allele, and were described as '*partial sequences*'.
273 Complete gene sequences, which aligned against a reference, but with a lower identity score of
274 0.95 and 0.99, were described as '*novel variants*'. In either case, sequences were aligned
275 against reference sequences from the ResFinder database to confirm the result, using MAFFT
276 version 7.471 and UGENE v37.0 (Kato *et al.*, 2013; Okonechnikov *et al.*, 2012).

277 *Visualisation of Results.* The R statistical programming language (v3.5.1) was used for
278 processing results from antibiotic testing (R Core Team, 2018), along with the readxl (v1.3.0)
279 for importing Excel spreadsheets (Wickham *et al.*, 2019), ggplot (v3.1.0) for plotting data
280 (Wickham, 2009), dplyr (v0.8.0.1) for creating summary statistics (Wickham *et al.*, 2019),
281 stringr (v1.3.1) for running various regular expressions (Wickham, 2018), tidyr (v0.8.3) for
282 reshaping tables from the wide to long convention (Wickham and Lionel, 2019), nlme for
283 fitting the linear mixed model to the data (Pinheiro *et al.*, 2020) and multcomp for conducting
284 pairwise comparisons (Hothorn *et al.*, 2008). Phylogenetic inferences, along with phylogenetic
285 data and information on presence and absence of specific resistance genes was then plotted
286 using version 5.5 of the ITOL webservice (Letunic and Bork, 2019).

287

288 **Results**

289 *Microbiological*

290 A total of 560 *E. coli* isolates was analysed for this monitoring programme, sampled between
291 December 2018 and May 2019, and obtained from Kuwait (n=223), Bahrain (n=156), Oman
292 (n=152) and UAE (n=29) (Supplementary File 2). Isolates fully susceptible to all 14
293 antimicrobials tested comprised 47.5% (n=266), whereas 30.5% (n=171) of isolates were
294 multidrug resistant. The most commonly observed resistance phenotypes were: ciprofloxacin
295 (44.6%), ampicillin (29.6%), nalidixic acid (27.9%), tetracycline (27.5%), trimethoprim
296 (25.5%) and sulfamethoxazole (22.5%). Resistance to chloramphenicol was 6.6%. Resistance
297 to the third generation cephalosporins cefotaxime and ceftazidime was at 12.7%; 8.8% of
298 which were classed as having the ESBL phenotype, 2.0% with having the AmpC phenotype
299 and 1.4% expressing both ESBL and AmpC phenotypes. The proportion of isolates showing
300 resistance to other HP-CIAs was as follows: azithromycin (9.3%), gentamicin (2.7%), colistin

301 (0.5%) and tigecycline (0.5%). No isolates showed resistance to meropenem. The overall
 302 phenotypic results, and those for the individual sampling sites, are indicated in Table 1.

	% of isolates showing reduced susceptibility, MDR, ESBL, AmpC, ESBL+AmpC										Total %
	Kuwait			Bahrain			Oman			UAE	
	AlGazali n=100	AlSalam n=100	Doha Bay n=23	Askar n=52	Nabih Saleh n=52	Tubli Bay n=52	Al Qurum n=71	Muttrah n=34	Darsait n=17	Dubai Creek n=29	
Sensitive	56.0	46.0	21.7	46.2	34.6	43.4	50.7	35.3	61.7	72.4	42.5
Ampicillin	31.0	34.0	56.5	34.6	34.6	42.3	4.2	47.1	14.9	13.8	29.6
Azithromycin	15.0	7.0	17.4	9.6	13.5	7.7	4.2	5.9	2.1	13.8	9.3
Cefixime	12.0	12.0	30.4	13.5	19.2	21.6	1.4	8.8	10.6	10.3	12.7
Cefuroxime	12.0	12.0	30.4	13.5	19.2	17.3	1.4	14.7	10.6	10.3	12.7
Chloramphenicol	9.0	4.0	8.7	5.8	5.8	3.8	0.0	23.5	10.6	3.4	6.6
Ciprofloxacin	39.0	42.0	69.6	42.3	63.5	53.8	45.1	55.9	29.8	17.2	44.6
Colistin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.3	0.5
Gentamicin	6.0	2.0	0.0	3.8	0.0	0.0	0.0	11.8	2.1	0.0	2.7
Meropenem	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nalidixic acid	25.0	24.0	34.8	11.5	40.4	28.8	45.1	47.1	14.9	6.9	27.9
Sulamethoxazole	24.0	21.0	30.4	21.2	28.8	40.4	4.2	41.2	14.9	10.3	22.5
Tetracycline	32.0	34.0	39.1	28.8	26.9	34.6	5.6	41.7	7.3	13.8	27.5
Tigecycline	1.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	0.0	3.4	0.5
Trimethoprim	32.0	28.0	39.1	25.9	29.4	34.6	4.2	32.4	14.9	23.7	25.5
MDR	33.0	35.0	56.5	32.7	35.5	40.4	4.2	47.1	19.1	17.2	30.5
ESBL	6.0	9.0	26.1	7.7	11.5	11.5	1.4	8.8	10.6	10.3	8.8
AmpC	5.0	1.0	4.3	1.9	5.8	0.0	0.0	0.0	0.0	0.0	2.0
ESBL+AmpC	0.0	1.0	0.0	3.8	1.9	7.7	0.0	0.0	0.0	0.0	1.4
Possible carbapenemase	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

MDR multi drug resistant. Resistant to 3 or more classes of antibiotic
 ESBL Extended spectrum beta-lactamase. Plasmid mediated resistance to penicillins and cephalosporins
 AmpC Chromosomal beta-lactamase. Resistance to penicillins, 2nd and 3rd gen cephalosporins
 Carbapenemase Phenotypic reduced sensitivity to carbapenem antibiotics. Mechanism not detected

303

304

305 Table 1 Heatmap depicting percentage resistance, MDR, and ESBL, AmpC phenotypes of
 306 analysed *E. coli* isolates. Raw MIC data can be seen in Supplementary file 3.

307

308 The statistical analysis looked for significant differences at the 5% level between levels of
 309 phenotypic resistance to antibiotics, ESBL and AmpC phenotypes, and MDR. These were
 310 compared firstly between countries, and secondly between the different locations within each
 311 country. When looking at resistance there was a strong positive association between Kuwait,
 312 Bahrain and Muttrah in Oman ($r>0.85$). No statistical difference in levels of MDR, ESBL or
 313 AmpC was found between countries. Within Kuwait, Doha Bay had significantly higher levels
 314 of resistance to antibiotics than either AlGazali ($p=0.0002$) or AlSalam ($p=1.10^{-4}$). Doha Bay
 315 also had significantly higher numbers of ESBL producers than AlGazali ($p=0.007$). In Bahrain
 316 both Nabih Saleh and Tubli Bay had significantly higher levels of resistance than Askar
 317 ($p=0.014$ and $p=0.044$; respectively). In Oman, Muttrah had significantly higher levels of
 318 resistance than either AlQurum ($p=1.10^{-4}$) or Darsait ($p=1.10^{-4}$). All three locations, Muttrah,

319 Darsait and AlQurum were significantly different from each other as regards numbers of
320 multidrug resistant isolates with Muttrah having the highest level followed by Darsait. There
321 were very low numbers of isolates returned from UAE. These were sampled from 4 areas sited
322 geographically close together. They were regarded as a single sample and as such there is no
323 within country comparison for UAE.

324

325 ***Whole Genome Sequencing***

326 *Genomic analyses:* 173 *E. coli* were sequenced. They were selected according to their
327 phenotypic resistance profiles, with a focus on isolates with multidrug resistance and/or
328 resistance to HP CIAs. The median depth of sequencing coverage was 306x and isolates
329 aligned against at least eighty percent of the reference genome *Escherichia coli* K-12
330 (GCA_000005845.2). Individual assembly data is available in Supplementary file 4. The raw
331 and assembled sequencing data have been submitted to NCBI and archived under BioProject
332 accession number PRJNA691754.

333

334 *Genotype v phenotype:* Cohen's kappa test was used to compare MIC with the presence of
335 resistance genes for ampicillin, cefotaxime, ceftazidime, gentamicin, azithromycin,
336 sulfamethoxazole-trimethoprim, ciprofloxacin and tetracycline, and for the ESBL and AmpC
337 phenotypes. There was almost perfect agreement ($k > 0.900$) between MIC and associated
338 resistance genes for ampicillin, tetracycline, azithromycin, ciprofloxacin, and the AmpC
339 phenotype. There was strong agreement (k 0.800 – 0.899) for gentamicin, ceftazidime,
340 cefotaxime, and sulfamethoxazole-trimethoprim, with moderate agreement (k 0.600 – 0.799)
341 for the ESBL phenotype. The frequency gene observations used to make these predictions are
342 listed in Supplementary file 5 and the positive and negative predictive values and kappa values
343 are listed in Supplementary File 6.

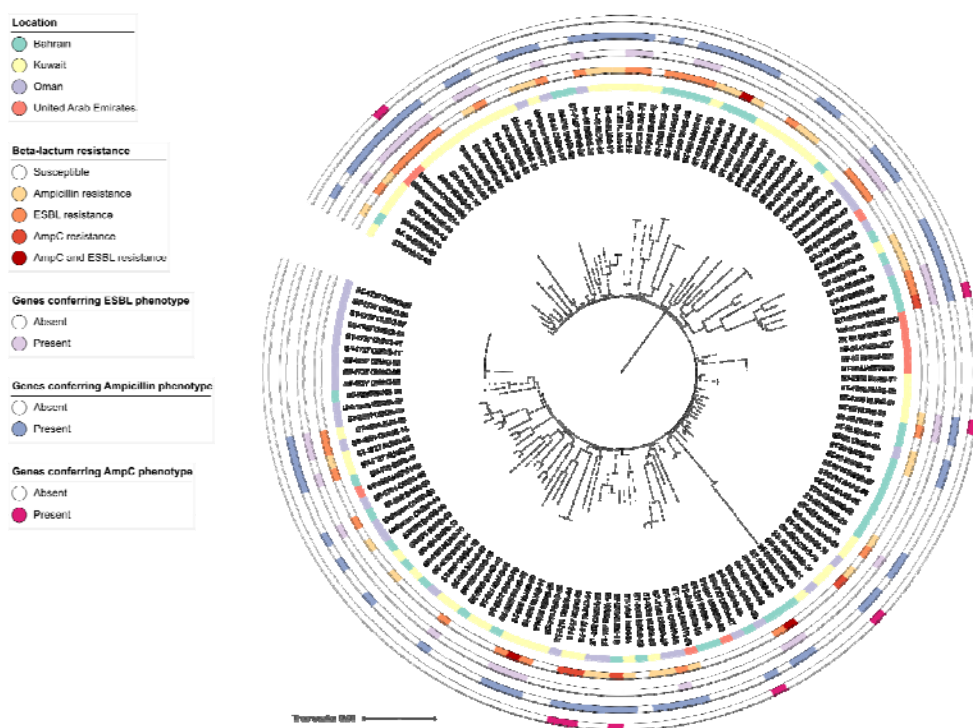
344

345 *Gene prevalence:* Of the 173 sequenced isolates 39.9% did not have any resistance genes
346 detected, while 45.1% carried resistance genes to three or more classes of antibiotic. The most
347 prevalent genes found were *sul* (47.4%), *tet* (45.1%), *qnr* (41.6%), *dfr* (38.7%), *bla*_{TEM-1}
348 (37.0%), *bla*_{CTXM} (26.6%), and *mph* (15.6%). No resistance genes were detected for the HP-
349 CIA antibiotics colistin, meropenem or tigecycline (Supplementary File 5).

350

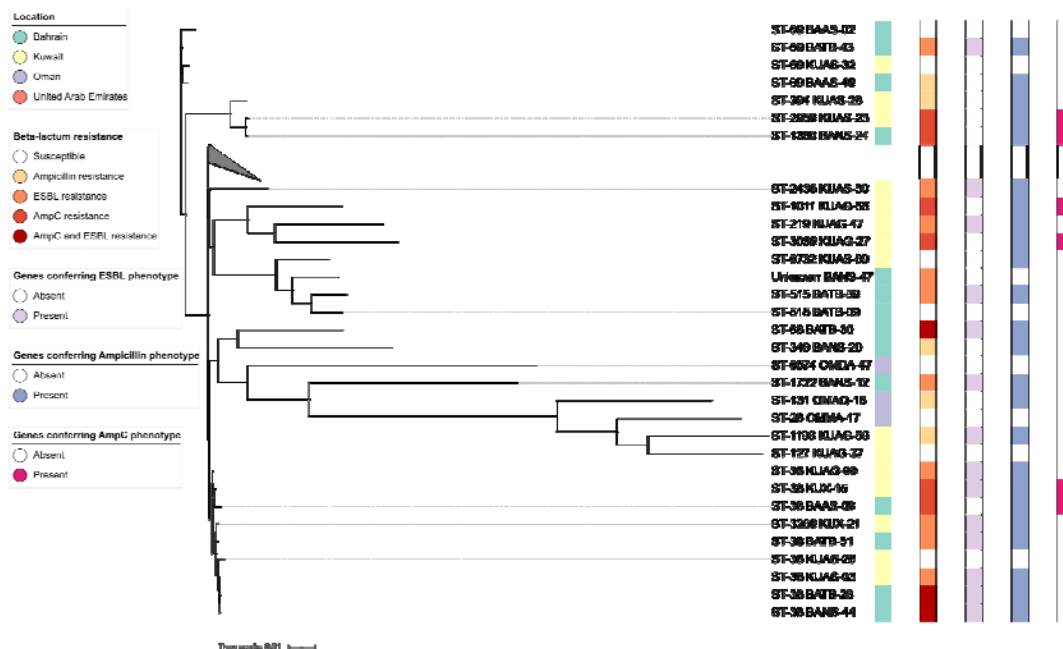
351 *Resistance to the β -lactam antibiotics:* Resistance to the β -lactam antibiotics including
352 ampicillin and the third generation cephalosporins, cefotaxime and ceftazidime, was
353 widespread involving several different genes. The *bla*_{TEM-1} gene conferring resistance to
354 ampicillin was present in 37.1% of the sequenced isolates. The *bla*_{DHA-1}, and *bla*_{CMY} variants,
355 consistent with the AmpC phenotype, were found at 6.4% and 1.8% respectively. There was a
356 very high prevalence of the ESBL gene *bla*_{CTXM-15} (25.4%), and to a lesser extent the other
357 ESBL genes *bla*_{CTXM-14b} (0.6%), *bla*_{CTXM-55} (0.6%), and *bla*_{SHV-12} (1.2%). Figures 2(a) and 2(b)
358 show the relationship between isolate location, sequence type, phenotypic resistance to the β -
359 lactam antibiotics and the presence of genes associated with resistance to ampicillin and the
360 ESBL and AmpC phenotypes.

361



362

363 Figure 2a



364

365 *Figure 2b*

366 *Figures 2a and Figure 2b* Phylogenetic trees showing isolates which are separated by an
 367 average branch distance of (a) less and (b) greater than 0.01. The beta-lactam resistance
 368 phenotype observed during phenotypic testing is included, along with any genes thought to
 369 contribute to the ESBL (*bla_{SHV-12}*, *bla_{CTX-M-14b}*, *bla_{CTX-M-15}*, *bla_{CTX-M-55}*), AmpC (*bla_{CMY-2}*,
 370 *bla_{DHA-1}*) and Ampicillin (*bla_{TEM-1A}*, *bla_{TEM-1B}*, *bla_{TEM-1C}*) resistance phenotypes.

371

372 *Resistance to the quinolones:* There were very high levels of resistance to the quinolones
 373 spread throughout the region. Of 246 isolates showing phenotypic resistance 103 were
 374 sequenced. There were two distinct MIC peaks for ciprofloxacin corresponding to 0.25 mg L⁻¹
 375 and >8 mg L⁻¹. There were corresponding high levels of the *qnrS* genes: *qnrS1* (56), *qnrS13*
 376 (2), *qnrB4* (6), *qnrS1* with *qnrB4* (1), *qnrB7* (1). These tended to correlate to the lower MIC
 377 peak of 0.06 <Cip ≤ 2 mg L⁻¹. Chromosomal mutations in the *gyrA*, *parC* and *parE* genes were
 378 noted more with the higher MIC peak 8 <Cip ≥ 32, particularly in the absence of the *qnr* genes.

379

380

381

382 *Sequence type analysis:*

383 A diverse range of sequence types was identified throughout the region. We have highlighted
384 those associated with clinical disease, also clones within the different sequence types which
385 were MDR and both expressed and contained genes conferring reduced susceptibility to the
386 third generation cephalosporins. The relatedness criteria for clonality in *E. coli* was a SNP
387 distance of ≤ 10 (Schürch *et al.*, 2018). The most prevalent sequence types (STs) were; ST1737
388 (9), ST10 (8), ST38 (8) and ST155 (8). All ST1737 isolates came from the same location in
389 Oman and had a SNP distance of 0. We believe that this was method related and the colonies
390 stemmed from an original *E. coli* isolate. ST34, ST131, ST295, and ST746 were present in
391 small numbers. These sequence types, including ST10, ST38 and ST155, are frequently
392 recovered from food and human intestinal samples and have strong associations with
393 uropathogenic and diarrhoeal disease. In particular, ST10, ST38 and ST155 isolates are found
394 commonly in the gut of humans and food animals. They have an intrinsic ability to acquire
395 AMR genes which may lead them to play a crucial role as a reservoir of these genes
396 (Matamoros *et al.*, 2017). Seven of the ST10 isolates were MDR and 5 carried the *bla*_{CTXM-15}
397 gene. Seven of the ST38 isolates were MDR and 6 carried the *bla*_{CTXM-15} gene. Six of the
398 ST155 isolates were MDR with 5 carrying either *bla*_{CTXM-15} or *bla*_{CMY-2} genes. A clone
399 consisting of 3 isolates from Oman were MDR and carried the *bla*_{CTXM-15} gene. This clone did
400 not have an identified ST. Fourteen clonal groups were found. Eleven involved samples taken
401 from the same locations at the same time. The three clonal groups which contained isolates
402 from different locations all came from Bahrain. One location was close to a wastewater outlet
403 whilst the other was about 2km away at the other end of a bay.

404

405 **Discussion**

406 The WHO has clearly outlined in its Global Action Plan that it is crucial that countries adopt
407 integrated surveillance programs for monitoring antibiotic resistance. As such many countries
408 are now implementing national surveillance programs, which cover both clinical and veterinary
409 settings. However, few such schemes appear to exist for monitoring AMR in environmental or
410 aquatic settings (Marano *et al.*, 2020). The WHO have developed and recommended the
411 Tricycle Programme as a mechanism to collect harmonised data on prevalence of ESBL
412 producing *E. coli* from different sectors, including the environment (Matheu *et al.*, 2017). Our
413 study builds on this initiative to provide further phenotypic and genetic information on *E coli*

414 recovered from marine waters in the Gulf, including whether they are ESBL producers. This
415 data can be used to better assess transmission pathways, likely risks posed by new and
416 emerging resistances, and better monitor the effectiveness of intervention strategies.

417 There are challenges associated with setting up and running transnational surveillance
418 programs. For example, in the study presented here there were some logistical problems in
419 isolating a representative library of *E. coli* strains from all sites and across all sampling periods.
420 The isolates screened here probably represent different industrial, domestic, and agricultural
421 sources encompassing diverse mammalian and avian hosts, as such it is not possible to make
422 definitive assertions regarding the sources of observed resistances. Also, it should be noted that
423 the study presented here is limited in terms of size and scope and represents a proof of concept
424 regarding the application of environmental surveillance to study AMR. Irrespective, the
425 success of such a short “snapshot” surveillance effort cannot be understated and is undoubtedly
426 important in the global effort to tackle AMR.

427 The AMR surveillance programme outlined here is both important and notable for several
428 reasons. Firstly, the initial results generated from this study show that a high proportion of the
429 *E. coli* isolates recovered from marine waters across the region were multi-drug resistant and
430 many showed reduced susceptibility to HP-CIAs. In this regard, the data presented here mirrors
431 findings that have been reported from both clinical and environmental studies in the region (Al-
432 Zarouni *et al.*, 2008, Balkhy *et al.*, 2016, Khan *et al.*, 2020), outlining high levels of resistance
433 in analysed strains. Of note, multi-drug resistance (resistance to 3 or more structural classes of
434 antimicrobials) was observed in 30.5% of tested *E. coli* strains isolated in this study. High
435 levels of reduced susceptibility to a range of structurally diverse agents, including ampicillin
436 (29.6%), nalidixic acid (27.9%), tetracycline (27.5%), sulfamethoxazole (22.5%) and
437 trimethoprim (25.5%) were observed (Table 1). Reduced susceptibility to the beta lactam
438 antibiotics was widespread in the strains analysed, and across all countries. Worryingly,
439 reduced susceptibility to HP-CIAs, including ciprofloxacin (44.6%), the third generation
440 cephalosporins cefotaxime (12.7%) and ceftazidime (12.7%), tigecycline (0.5%) and colistin
441 (0.5%) was also observed. Perhaps most significantly, reduced susceptibility to ciprofloxacin
442 was detected in over 40% of tested *E. coli* isolates (Table 1). This matches previous data
443 observed for marine isolates reported from Kuwait (Al-Sarawi *et al.*, 2018). Previous meta-
444 analysis studies have shown high level resistance to ciprofloxacin in environmental and food-
445 derived *E. coli* strains, but at generally lower levels than observed here (14.4%, range 5.4-
446 33.4%). However, high ciprofloxacin resistance (~52%) has previously been described in

447 studies of *E. coli* isolated from urine samples and poultry carcasses in Iran (Pormohammad *et*
448 *al.*, 2019). Identification of reduced susceptibility to colistin and tigecycline in environmental
449 *E. coli* strains is of some concern, given that these are HP-CIA (WHO 2019). Although the
450 levels were low (0.5%) there remains the potential for these resistances to be present at higher
451 frequencies given the limited scope of this monitoring study. Due to the labile nature of
452 tigecycline it is not always easy to generate an accurate MIC. All isolates with an MIC above
453 the ECOFF were repeated. Two of these were one dilution above the ECOFF and the other two
454 dilutions above. As tigecycline is considered a drug of last resort to treat carbapenem-resistant
455 infections, the detection of such resistances in environmental strains is particularly noteworthy.

456 Secondly, as part of this surveillance programme we chose to whole genome sequence a subset
457 of isolates (173/560 strains, ~30% of the total library). This sequencing effort revealed an
458 expected diversity of resistance genes as well as helped identify potentially clinically relevant
459 sequence types (STs). Of note, high carriage rates of *qnrS1* (60/173 sequenced isolates),
460 *bla_{CTX-M-15}* (45/173) and *bla_{DHA-1}* (11/173) were observed, correlating with reduced
461 susceptibility to the fluoroquinolones and third generation cephalosporins, respectively. These
462 latter genes are typically carried on plasmids and are both associated with extended spectrum
463 cephalosporin resistance in a range of clinically important Gram-negative bacteria, including *E.*
464 *coli*. Plasmid mediated resistance is particularly relevant to the persistence and spread of AMR
465 both within and between species in the aquatic environment. Previous studies have shown that
466 both genes are circulating in hospital and community (Peterson and Bonomo, 2005). They are
467 also evident in pathogens in the region (Khan *et al.*, 2020). Their presence at such high levels
468 in the aquatic environment is concerning. One possible explanation could be the level of
469 sewage pollution. Many of our samples were taken close to sewage outlets with ‘obvious
470 pollution’ described as present at the time of sampling. Despite finding no statistical difference
471 in the levels of resistance and MDR between countries there were some differences between
472 locations within countries, with significantly higher levels found in busy harbours and close to
473 sewage outlets compared to recreational beaches. Interestingly, while the rates of both ESBL
474 and carbapenamase producing *E. coli* are increasing in clinical settings and in the community
475 in the GCC (Zowawi *et al.*, 2020) we only detected ESBL *E. coli*. We did not detect resistance
476 to meropenem or the presence of *bla_{KPC}*, *bla_{IMP}*, *bla_{NDM}*, *bla_{SPM}*, *bla_{VIM}*, or *bla_{OXA-48}*. Despite
477 seeing low level phenotypic resistance to both colistin and tigecycline we did not detect either
478 the *mcr* or *tetX* genes associated with plasmid mediated resistance to these CIAs.

479 The genetic basis of resistance to front-line clinically relevant drugs is the focus of current
480 work on several isolates, especially those demonstrating reduced susceptibility to tigecycline
481 and colistin. A range of resistance profiles and associated genes were identified here,
482 highlighting the applicability of combining genome sequencing into a routine surveillance
483 framework to study environmental AMR as part of a wider surveillance system. Numerous
484 potentially pathogenic strains were identified during this study. In particular, ST10, ST38,
485 ST155, ST34, ST131, ST295, ST746 were all identified here (Supplementary File 3). These
486 sequence types encompass both environmental and clinically associated strains. They are
487 frequently recovered from food and human intestinal samples and have strong associations
488 with uropathogenic and diarrhoeal disease. Recently a clinical molecular epidemiological
489 review of ST prevalence in extra-intestinal pathogenic *E. coli* (ExPEC) isolates from Saudi
490 Arabia found ST131 to be the dominant form observed (Alqasim, 2020). ST10, ST38 and
491 ST155 isolates are also found commonly in the gut of humans and animals (Matamoros *et al.*,
492 2017, Chattaway *et al.*, 2014, Castellanos *et al.*, 2017). The utility of WGS to detect and
493 further characterise strains displaying unusual or clinically relevant resistances (e.g. ESBL
494 producing strains, isolates resistant to colistin and/or tigecycline) as demonstrated here, and in
495 a rapid and cost-effective manner may allow genomic approaches to be further implemented as
496 part of AMR surveillance initiatives in the future (EFSA, 2019).

497 To our knowledge this study represents the first attempt to design and implement a regional
498 transnational surveillance programme for studying AMR in marine and coastal environments.
499 It follows on from discussion/opinion articles that highlighted the need for regional
500 coordination to tackle the emergence of AMR in the GCC region (Le Quesne *et al.*, 2018,
501 Balkhy *et al.*, 2016). Both papers specifically outlined the urgent need to ascertain AMR data
502 from environmental sources and to develop rapid baseline field surveys for the presence of
503 AMR bacteria and associated genes in marine and aquatic environments across the GCC as part
504 of a wider and coordinated monitoring. One of the main sources of AMR bacteria into the
505 wider marine environment is via sewage discharge either from sewage treatment plants (STPs)
506 or direct inputs (Aarestrup and Woolhouse 2020). In regions such as South-East Asia it is still
507 considered that transmission via contact with contaminated environments (e.g. water and soil),
508 and from companion or livestock animals is regarded as low risk when compared to direct
509 human to human transmission or contact in a clinical setting (Ng & Gin 2019). However, in
510 water scarce regions such as the GCC there is a greater reliance on water re-use for
511 environmental enrichment (recreational parks etc), and this route as a source of exposure back

512 to humans is often over-looked. Evidence suggests that while water treatment (including UV
513 and chlorination) significantly reduces the concentration of bacteria in processed compared to
514 raw effluent, the AMR of surviving bacteria in the final effluent may actually be higher than in
515 pre or partially treated effluent (Al Jasim *et al.*, 2015; Aslan *et al.*, 2018). More information is
516 needed on removal efficiencies of technology employed at STPs across the GCC when
517 developing new AMR surveillance programmes, especially in effluents intended for discharge
518 into the environment or reuse for irrigation (Al Jasim *et al.*, 2015). Further information is also
519 needed to better understand the potential risks to humans who use coastal waters for
520 recreational purposes (e.g. swimming, fishing and boating). There is obviously a direct
521 exposure risk via this route and the threat of exposure to a wide range of illness when in
522 contact with bacterially contaminated recreation waters is clearly established, even in high-
523 income countries with well-developed waste-water treatment infrastructure (Leonard *et al.*,
524 2018a). Data from European coastal waters has also shown that members of the public who
525 regularly engage in water sports are at greater risk of being exposed to, and colonised by,
526 clinically important antibiotic-resistant *E. coli* (Leonard *et al.*, 2018b). Therefore, it will be
527 essential to further address these pathways to fully understand the risk posed by antibiotic
528 resistant bacteria and antibiotic resistant genes on AMR prevalence in these receiving
529 environments.

530 This work highlighted the need for establishing standardised methods for collection, analysis
531 and interpretation of results. In partnership with GCC regional partners in Bahrain, Oman,
532 Kuwait, and the United Arab Emirates, we were able to imbed and implement a coordinated
533 environmental study that provides AMR data - albeit in a temporally and spatially constrained
534 context - that can be utilised for various purposes, such as studying the emergence of resistant
535 and clinically-relevant resistances, to identify new and emergent resistance genes, characterise
536 potentially pathogenic strains present in the environment, and to suggest a comparison with
537 AMR data generated from clinical strains to known prescription/antibiotic usage data, among
538 others. A more structured sampling programme in future environmental surveillance designs
539 may be better able to attribute source of *E. coli* and clinical comparison. Many AMR national
540 action plans have yet to fully adopt a One Health approach to include environmental
541 components of AMR. Intervention strategies can only be devised to prevent and control the
542 spread of AMR when all possible sources and sinks have been identified. It is anticipated that
543 significant benefits will be seen across the Gulf States if they also implement comprehensive
544 One Health based AMR and AMU management processes, particularly if these are harmonised

545 at a regional level. Taken together, this study demonstrates that a monitoring framework can be
546 achieved in both a relatively short space of time and with limited infrastructure or established
547 laboratory networks in place. In addition, this study has demonstrated greater consistency and
548 standardization of the data collected across the region, forged transnational and international
549 collaborations, and helped encourage improved data linkage in the GCC region.

550 The overall objective of the study – to design and carry out a baseline study to get an estimate
551 of AMR in the marine environment of the Gulf – has been achieved. This has highlighted the
552 feasibility of developing and implementing harmonised marine environmental AMR
553 surveillance and monitoring programmes, that could directly support integrated surveillance
554 activities across these different sectors. Key knowledge gaps remain, but the data provided here
555 will help facilitate the development of more comprehensive action plans, and engagement with
556 stakeholders that typically do not form the focus of existing regional and national working
557 groups. As such it will enable a strategic overview of the coverage of different surveillance and
558 research activities in relation to the marine and aquatic environments to help guide the
559 development of future monitoring and research programmes.

560 **Data availability statement**

561 The assembled genomes, as well as the raw sequencing data, have been deposited at
562 DDBJ/ENA/GenBank under the BioProject accession number PRJNA691754.

563

564 **References**

565 Aarestrup, F.M., Woolhouse, M.E.J. (2020). Using sewage for surveillance of antimicrobial
566 resistance. *Science* 367, 6478, 630-632. DOI: 10.1126/science.aba3432

567

568 Abdul Salam, A., I. Elsegaey, R. Khraif, A. Al Mutairi, A. Aldosari. Components and public
569 health impact of population growth in the Arab world. *PLoS One*, 10 (5) (2015), Article
570 e0124944, 10.1371/journal.pone.0124944

571

572 Alqasim, A. (2020). Extraintestinal pathogenic *Escherichia coli* in Saudi Arabia: A review of
573 antimicrobial resistance and molecular epidemiology. *Tropical Journal of Pharmaceutical
574 Research* 19 (2): 447-453. DOI: 10.4314/tjpr.v19i2.30

575

576 Alves, M.S., Pereira, A, Araújo, S. M, Castro, B. B, Correia, A. C, and I Henriques. Seawater
577 is a reservoir of multi-resistant *Escherichia coli*, including strains hosting plasmid-mediated
578 quinolones resistance and extended-spectrum beta-lactamases genes. *Frontiers in
579 Microbiology*. 2014 Aug 20;5:426. doi: 10.3389/fmicb.2014.00426.

580

- 581 Al-Jassim, N., Anasari, M.I., Harb, M., Hong, P.Y. (2015). Removal of bacterial contaminants
582 and antibiotic resistance genes by conventional wastewater treatment processes in Saudi
583 Arabia: Is the treated wastewater safe to reuse for agricultural irrigation? *Water Research*, 73
584 (15), 277-290. <https://doi.org/10.1016/j.watres.2015.01.036>
585
- 586 Al Sarawi, H.A., Jha, A.N., Baker-Austin, C., Al-Sarawi, M.A., Lyons, B.P. (2018). Baseline
587 screening for the presence of antimicrobial resistance in *E. coli* isolated from Kuwait's marine
588 environment. *Marine Pollution Bulletin*, 129 (2), 893-898.
589 <https://doi.org/10.1016/j.marpolbul.2017.10.044>
590
- 591 Al-Yamani, A., F. Khamis, I. Al-Zakwani, H. Al-Noomani, J. Al-Noomani, S. Al-Abri.
592 Patterns of antimicrobial prescribing in a tertiary care hospital in Oman. *Oman Medical*
593 *Journal.*, 31 (1) (2016), pp. 35-39, 10.5001/omj.2016.07
594
- 595 Al-Zarouni, M. Senok, A. Rashid, F. Al-Jesmi, S.M. and D. Panigrahi. Prevalence and
596 antimicrobial susceptibility pattern of extended-spectrum beta-lactamase-producing
597 Enterobacteriaceae in the United Arab Emirates. *Med. Princ. Pract.* 2008, 17, 32–36
598
- 599 Amos, G.C., Zhang, L., Hawkey, P.M., Gaze, W.H., and Wellington, E.M. Functional
600 metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes.
601 *Veterinary Microbiology*. (2014) Jul 16;171(3-4):441-7. doi: 10.1016/j.vetmic.2014.02.017.
602 Epub 2014 Feb 16. PMID: 24636906.
603
- 604 Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data.
605 Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> Accessed 16th
606 [April 2021](#).
607
- 608 Anjum, M.F., Schmitt, H., Börjesson, S., Thomas U Berendonk, T.U. et al. (2021)The potential
609 of using *E. coli* as an indicator for the surveillance of antimicrobial resistance (AMR) in the
610 environment. *Current Opinion in Microbiology* 64 152-158,
611 <https://doi.org/10.1016/j.mib.2021.09.011>.
612
- 613 Aslan, A., Cole, Z., Bhattacharya, A., and O. Oyibo, (2018). Presence of antibiotic-resistant
614 *Escherichia coli* in wastewater treatment plant effluents utilized as water reuse for irrigation.
615 *Water*, 10, 805
616
- 617 Balkhy, H.H., Assiri, A.M., Al Mousa, H., Al-Abri, S.S., Al-Katheeri, H., Alansari, H.,
618 Abdulrazzaq, N.M., Aidara-Kane, A., Pittet, D., and E Erlacher-Vindel. The strategic plan for
619 combating antimicrobial resistance in Gulf Cooperation Council states; Elsevier: Amsterdam,
620 The Netherlands, 2016.
621
- 622 Balkhair, A., Y.M. Al-Farsi, Z. Al-Muharrmi, R. Al-Rashdi, M. Al-Jabri, F. Neilson, S.S. Al-
623 Adawi, M. El-Beeli, S. Al-Adawi. Epidemiology of multi-drug resistant organisms in a
624 teaching hospital in Oman: A one-year hospital-based study. *Scientific World Journal*. (2014),
625 p. 157102, 10.1155/2014/157102
626
- 627 Baker-Austin, C, M. Wright, R. Stepanauskas, and J V. McArthur. Co-selection of antibiotic
628 and metal resistance. *Trends in Microbiology*. 14:176-181, 2006.
629

- 630 Baquero, F. *et al.* (2008). Antibiotics and antibiotic resistance in water environments. *Current*
631 *Opinion in Biotechnology*. 19, 260–265
632
- 633 Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
634 sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114-20. doi:
635 10.1093/bioinformatics/btu170.
636
- 637 Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A.,
638 Allesoe, R. L., Rebelo, A. R., Florensa, A. F., Fagelhauer, L., Chakraborty, T., Neumann, B.,
639 Werner, G., Bender, J. K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., Malhotra-
640 Kumar, S., ... Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from
641 genotypes. *The Journal of antimicrobial chemotherapy*, 75(12), 3491–3500.
642 <https://doi.org/10.1093/jac/dkaa345>
643
644
- 645 Butt, A.A., C.S. Navasero, B. Thomas, S.A. Marri, H.A. Katheeri, A.A. Thani, A.A. Khal, T.
646 Khan, A.B. Abou-Samra. Antibiotic prescription patterns for upper respiratory tract infections
647 in the outpatient Qatari population in the private sector. *International Journal of Infectious*
648 *Diseases.*, 55 (2017), pp. 20-23
649
- 650 Castellanos, L. R., Donado-Godoy, P., León, M., Clavijo, V., Arevalo, A., Bernal, J. F.,
651 Timmerman, A. J., Mevius, D. J., Wagenaar, J. A. & Hordijk, J. (2017). High heterogeneity of
652 *Escherichia coli* sequence types harbouring ESBL/AmpC genes on *IncII* plasmids in the
653 Colombian poultry chain. *PloS One*, 12(1).
654
- 655 Chattaway, M. A., Jenkins, C., Ciesielczuk, H., Day, M., DoNascimento, V., Day, M.,
656 Rodríguez, I., van Essen-Zandbergen, A., Schink, A. K., Wu, G., Threlfall, J., Woodward, M.
657 J., Coldham, N., Kadlec, K., Schwarz, S., Dierikx, C., Guerra, B., Helmuth, R., Mevius, D.,
658 Woodford, N., J. Wain. (2014). Evidence of evolving extraintestinal enteroaggregative
659 *Escherichia coli* ST38 clone. *Emerging Infectious Diseases*, 20(11), 1935–1937.
660 <https://doi.org/10.3201/eid2011.131845>
661
- 662 EFSA (2019). Technical specifications on harmonised monitoring of antimicrobial resistance
663 in zoonotic and indicator bacteria from food-producing animals and food. *EFSA Journal* 2019;
664 17(6) 5709 <https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2019.5709>
665
- 666 EFSA (2020). The European Union Summary Report on Antimicrobial Resistance in zoonotic
667 and indicator bacteria from humans, animals and food in 2017/2018. *EFSA Journal*, 18(3),
668 e06007. doi: doi: 10.2903/j.efsa.2020.6007
669
- 670 Ewels, P., M. Magnusson, S. Lundin, and M. Källér. (2016). MultiQC: summarize analysis
671 results for multiple tools and samples in a single report. *Bioinformatics*. 32(19): 3047-3048.
672
- 673 Gunnar, K., Erika, M., Jenny, Å., Derek, B., & John, T. (2020). Minimum Inhibitory
674 Concentrations (MIC) and zone diameter distributions and Epidemiological cut-offs (ECOFFs).
675 Retrieved from http://www.eucast.org/mic_distributions_and_ecoffs/
676
- 677 Hothorn, T., F. Bretz and P. Westfall (2008). Simultaneous inference in general parametric
678 models. *Biometrical Journal* 50(3), 346-363.
679

- 680 Jolley, K. A., and Maiden, M. C. J. (2010). BIGSdb: scalable analysis of bacterial genome
681 variation at the population level. *BMC Bioinformatics* 11:595. doi: 10.1186/1471-2105-11-595
682
- 683 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
684 improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772–780.
685 <https://doi.org/10.1093/molbev/mst010>
686
- 687 Karkman, A., K. Parnanen, D.G.J. Larsson. Fecal pollution can explain antibiotic resistance
688 gene abundances in anthropogenically impacted environments.
689 *Nature Communications*. 10 (2019), Article 80, 10.1038/s41467-018-07992-3
690
- 691 Khan, M.A., Thurgood, N.E., Faheem, S.M., Rais, N., Ansari, M.Z., Kaleem, S.M. and S. T.
692 Khan. (2020). Occurrence of extended spectrum beta-lactamase Gram-negative bacteria from
693 non-clinical sources in Dubai, United Arab Emirates. *Water* 2020, 12, 2562.
694
- 695 Khatir, Z., Leitão, A., and Lyons, B.P. The biological effects of chemical contaminants in the
696 Arabian/Persian Gulf: A review. *Regional Studies in Marine Science*. (2020) Jan 1;33:100930
697
- 698 Le Quesne, W., C. Baker-Austin, D. W. Verner-Jeffreys, H. A. Al-Sarawi, H. H. Balkhy and B.
699 P. Lyons. Antimicrobial Resistance in the Gulf Cooperation Council Region: a proposed
700 framework to assess threats, impacts and mitigation measures associated with AMR in the
701 marine and aquatic environment. *Environment International*. 121, 1003-1010 (2018)
702
- 703 Leonard, A.F.C., Singer, A., Ukoumunne, O.C., Gaze, W.H., Garside, R. (2018a). Is it safe to
704 go back into the water? A systematic review and meta-analysis of the risk of acquiring
705 infections from recreational exposure to seawater. *International Journal of Epidemiology*, 572–
706 586 DOI: 10.1093/ije/dyx281
707
- 708 Leonard, A.F.C., Zhang, L., Balfour, A.J., Garside, R., Hawkey, P.M., Murray, A.,
709 Ukoumunne, O.C., Gaze, W.H. (2018b). Exposure to and colonisation by antibiotic-resistant *E.*
710 *coli* in UK coastal water users: Environmental surveillance, exposure assessment, and
711 epidemiological study (Beach Bum Survey). *Environment International* 114, 326-333.
712 <https://doi.org/10.1016/j.envint.2017.11.003>
713
- 714 Letunic, I and Bork P (2019). (Interactive Tree Of Life) - is a very impressive online tool for
715 the display, manipulation and annotation of phylogenetic and other trees. *Nucleic Acids Res*
716 47(W1): W256-W259.
717
- 718 Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., Clavijo, B.J. (2017). KAT: a
719 K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*.
720 2017;33(4):574-576. doi:10.1093/bioinformatics/btw663.
721
- 722 Marano, R.B.M.; Fernandes, T.; Manaia, C.M.; Nunes, O.; Morrison, D.; Berendonk, T.U.;
723 Kreuzinger, N.; Tenson, T.; Corno, G.; Kassinos, D.F., et al. (2020). A Global Multinational
724 Survey of Cefotaxime-Resistant Coliforms in Urban Wastewater Treatment Plants.
725 *Environment International*, 144, 106035. <https://doi.org/10.1016/j.envint.2020.106035>
726
- 727 Matamoros, S., van Hattem, J.M., Arcilla, M.S. *et al.* Global phylogenetic analysis of
728 *Escherichia coli* and plasmids carrying the *mcr-1* gene indicates bacterial diversity but plasmid
729 restriction. *Scientific Reports* 7, 15364 (2017). <https://doi.org/10.1038/s41598-017-15539-7>

730

731 Matheu, J., Aidara-Kane, A., Andremont, A. (2017). The ESBL tricycle AMR surveillance
732 project: a simple, one health approach to global surveillance. *AMR Control*, Jul 31, 2017.

733

734 Ng, C., G, K, Y-H. (2019). Monitoring Antimicrobial Resistance Dissemination in Aquatic
735 Systems. *Water*, 11 (1), 71, <https://doi.org/10.3390/w11010071>.

736

737 Nguyen, L. T., H.A. Schmidt, A. von Haeseler, B.Q. Minh (2015) IQ-TREE: A fast and
738 effective stochastic algorithm for estimating maximum likelihood phylogenies.. *Mol. Biol.*
739 *Evol.*, 32:268-274. <https://doi.org/10.1093/molbev/msu300>

740

741 Okonechnikov, K., Golosova, O., Fursov, M., & UGENE team (2012). Unipro UGENE: a
742 unified bioinformatics toolkit. *Bioinformatics (Oxford, England)*, 28(8), 1166–1167.
743 <https://doi.org/10.1093/bioinformatics/bts091>

744

745 O'Neill. J. Tackling drug-resistant infections globally: final report and recommendations the
746 review on antimicrobial resistance; (2016). Available from: <https://amr-review.org/> Accessed
747 16th April 2021.

748

749 Paterson, D. L, and R. A. Bonomo. Extended-spectrum beta-lactamases: a clinical update.
750 *Clinical Microbiology Reviews* 2005 Oct;18(4):657-86.

751

752 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team (2020). *nlme: Linear and*
753 *Nonlinear Mixed Effects Models*. R package version 3.1-148. URL:
754 <https://CRAN.R-project.org/package=nlme>>.

755

756 Pormohammad, A., Nasiri, M. J., & Azimi, T. (2019). Prevalence of antibiotic resistance in
757 *Escherichia coli* strains simultaneously isolated from humans, animals, food, and the
758 environment: a systematic review and meta-analysis. *Infection and drug resistance*, 12, 1181–
759 1197. <https://doi.org/10.2147/IDR.S201324>

760

761 R Core Team (2018). R: A language and environment for statistical computing. R Foundation
762 for Statistical Computing, Vienna, Austria. Available online at <https://www.R-project.org/>.
763 Accessed 16th April 2021.

764

765 Schürch, A. C., S. Arredondo-Alonso, R.J.L. Willems. R.V. Goering. (2018). Whole genome
766 sequencing options for bacterial strain typing and epidemiologic analysis based on single
767 nucleotide polymorphism versus gene-by-gene-based approaches. *Clinical Microbiology and*
768 *Infection* 24. 350-354.

769

770 Schwarz, S., Silley, P., Simjee, S., Woodford, N., van Duijkeren, E., Johnson, A.P., and
771 Gaastra W. Assessing the antimicrobial susceptibility of bacteria obtained from animals.
772 *Veterinary Microbiology*, (2010), 141(1):1-4.

773

774 Seemann, T. (2020a). MLST. Retrieved from <https://github.com/tseemann/mlst>

775

776 Seemann, T. (2020b). Snippy. Retrieved from <https://github.com/tseemann/snippy>

777

778 Seemann, T., Goncalves de Silva, A., Bulach, D., Schultz, M., Kwong, J., & Howden, B.
779 (2020). Nullarbor. Retrieved from <https://github.com/tseemann/nullarbor>

780

781 Souvorov, A., Agarwala, R., & Lipman, D. J. (2018). SKESA: Strategic k-mer extension for
782 scrupulous assemblies. *Genome Biology*, 19(1), 1–13. [https://doi.org/10.1186/s13059-018-](https://doi.org/10.1186/s13059-018-1540-z)
783 1540-z

784

785 Sheppard, C., M. Al-Husiani, F. Al-Jamali, F. Al-Yamani, R. Baldwin, J. Bishop, F. Benzoni,
786 E. Dutrieux, N.K. Dulvy, S.R.V. Durvasula, D.A. Jones, R. Loughland, D. Medio, M.
787 Nithyanandan, G.M. Pilling, I. Polikarpov, A.R.G. Price, S. Purkis, B. Riegl, M. Saburova,
788 K.S. Namin, O. Taylor, S. Wilson, K. Zainal. The Gulf: a young sea in decline. *Marine*
789 *Pollution Bulletin.*, 60 (2010), pp. 13-38

790

791 Stubberfield, E., AbuOun M., Sayers E., O'Connor H., Card R., Anjum M. (2019) Use of
792 whole genome sequencing of commensal *Escherichia coli* in pigs for antimicrobial resistance
793 surveillance, United Kingdom, 2018. [https://doi.org/10.2807/1560-7917.](https://doi.org/10.2807/1560-7917.ES.2019.24.50.1900136)
794 ES.2019.24.50.1900136

795

796

797 Taylor, N., D. Verner-Jeffreys and C. Baker-Austin. Aquatic systems: maintaining, mixing
798 and mobilising antimicrobial resistance? *Trends in Ecology and Evolution*, 6:278-294, 2011.

799

800 Walsh T.R. (2018). A one-health approach to antimicrobial resistance. *Nature Microbiology* 3,
801 854–855. <https://doi.org/10.1038/s41564-018-0208-5>

802

803 Wellington, E. M. H., Boxall, Alistair B. A., Cross, Paul, Feil, Edward J., Gaze, William H.,
804 Hawkey, P. M., Johnson-Rollings, Ashley S., Jones, Davey L., Lee, Nick, Otten, Wilfred,
805 Thomas, Christopher M. and Williams, A. Prysor (2013) The role of the natural environment in
806 the emergence of antibiotic resistance in Gram-negative bacteria. *The Lancet Infectious*
807 *Diseases*, Vol.13 (No.2). pp. 155-165. doi:10.1016/S1473-3099(12)70317-1

808

809 Wickham *et al.*, (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43),
810 1686, <https://doi.org/10.21105/joss.01686>

811

812 Wickham and Bryan (2019). Hadley Wickham and Jennifer Bryan (2019). readxl: Read Excel
813 Files. R package version 1.3.1. <https://CRAN.R-project.org/package=readxl>

814

815 Wickham (2016). Hadley Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-
816 Verlag New York, 2016.

817

818 Wickham *et al.* (2019). Hadley Wickham, Romain François, Lionel Henry and Kirill Müller
819 (2019). dplyr: A Grammar of Data Manipulation. <https://CRAN.R-project.org/package=dplyr>

820

821 Wickham (2019). Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common
822 String Operations. <https://CRAN.R-project.org/package=stringr>

823

824 Wickham (2019). Hadley Wickham (2019). tidyr: Tidy Messy Data. [https://CRAN.R-](https://CRAN.R-project.org/package=tidyr)
825 [project.org/package=tidyr](https://CRAN.R-project.org/package=tidyr)

826

827 Wood D, Lu J, Langmead B. (2019). Improved metagenomic analysis with Kraken 2. *Genome*
828 *Biology* 20: 257. doi:10.1186/s13059-019-1891-0

829

830 WHO (2019). Critically important antimicrobials for human medicine, 6th revision. Available
831 from: <https://www.who.int/foodsafety/publications/antimicrobials-sixth/en/> Accessed 16th April
832 2021.

833

834 World Bank. Drug-resistant infections a threat to our economic future; (2017). Available from:
835 www.worldbank.org. Accessed 16th April 2021.

836

837 Zankari E., Hasman H., Cosentino S., Vestergaard M., Rasmussen S., Lund O., *et al.* (2012).
838 Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial*
839 *Chemotherapy*.. 67 2640–2644. 10.1093/jac/dks261

840

841 Zankari, E., Hasman, H., Kaas, R. S., Seyfarth, A. M., Agersø, Y., Lund, O., Larsen, M. V., &
842 Aarestrup, F. M. (2013). Genotyping using whole-genome sequencing is a realistic alternative
843 to surveillance based on phenotypic antimicrobial susceptibility testing. *The Journal of*
844 *antimicrobial chemotherapy*, 68(4), 771–777. <https://doi.org/10.1093/jac/dks496>

845

846 [Zowawi, H., Balkhy, H., Walsh, T., Paterson D. \(2020\). \$\beta\$ -lactamase production in key Gram-](#)
847 [negative pathogen isolates from the Arabian peninsula. *Clinical Microbiological Reviews* 26](#)
848 [\(3\)](#)

849 DOI: <https://doi.org/10.1128/CMR.00096-12>

850

851 **Supplementary Files**

852 **Supplementary file 1.** A copy of the study plan for the collection and initial processing of
853 water samples.

854 **Supplementary file 2.** Data gathered during sample collection and initial sample processing
855 e.g. location, water temperature, comments, *E. coli* concentration cfu/mL.

856 **Supplementary file 3.** Raw data including sample ID, location, ST type, MIC values.

857 **Supplementary file 4.** Statistics describing the genome assemblies of each *Escherichia coli*
858 isolate.

859 **Supplementary file 5.** Frequency gene observations/relevant chromosomal mutations for
860 resistance phenotypes.

861 **Supplementary file 6.** Correlation of whole genome sequencing, ECOFF values and test
862 performances for *Escherichia coli* isolates. G+: gene/SNP present; G-: gene/SNP absent; NPV:
863 negative predictive value; P+: phenotype resistant; P-: phenotype sensitive; PPV: positive
864 predictive value; kappa: Cohen's kappa coefficient.

865