FEMS Microbiology Letters, 366, 2019, fnz179

doi: 10.1093/femsle/fnz179
Advance Access Publication Date: 16 August 2019
Research Letter

RESEARCH LETTER – Environmental Microbiology

Investigating microbial indicators of anthropogenic marine pollution by 16S and 18S High-Throughput Sequencing (HTS) library analysis

Maria Antonietta Buccheri[1,*,†], Eliana Salvo[2,†], Manuela Coci[3,‡], Grazia M. Quero[4], Luca Zoccarato[5], Vittorio Privitera[1] and Giancarlo Rappazzo[2]

[1]CNR-IMM, Sede secondaria di Catania (Università), 95123 Catania, Italy, [2]Laboratorio di Genetica Molecolare, Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Università di Catania, 95124 Catania, Italy, [3]Microb&Co, 95128 Catania, Italy, [4]Stazione Zoologica Anton Dohrn - Sezione di Ecologia Marina Integrata -, 80121 Naples, Italy and [5]Department of Limnology of Stratified Lakes, Leibniz Institute of Freshwater Ecology and Inland Fisheries, 16775 Stechlin, Germany

*Corresponding author: Laboratorio di Genetica Molecolare, Sezione Biologia Animale, Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Università di Catania, Via Androne, 81-95124 Catania, Italy. Tel: +39-0957306055; E-mail: mariaantonietta.buccheri@cnr.it

One sentence summary: The analysis of marine microbial communities, through High-Throughput Sequencing, allows identification of alternative indicators that unveils human activities impacting a marine protected area within the Mediterranean coastal sea.

†These authors equally contributed to this work.

Editor: Rustam Aminov

‡Manuela Coci, http://orcid.org/0000-0002-1228-1722

ABSTRACT

High-Throughput Sequencing technologies are providing unprecedented inventories of microbial communities in aquatic samples, offering an invaluable tool to estimate the impact of anthropogenic pressure on marine communities. In this case study, the Mediterranean touristic site of Aci Castello (Italy) was investigated by High-Throughput Sequencing of 16S and 18S rRNA genes. The sampling area falls within a Marine Protected Area and, notwithstanding, features an untreated urban wastewater discharge. Seawater samples were collected close to the wastewater output (COL) and at a second station about 400 m further off (PAN), before and after a summer increase in population. Prokaryotic communities clustered according to stations, rather than to seasons. While PAN showed a typical, not impacted, marine microbial composition, COL was consistently enriched in Epsilonproteobacteria and Firmicutes. Protist communities showed a peculiar clustering, as COL at springtime stood alone and was dominated by Ciliophora, while the other samples were enriched in Dinophyta. Analysis of alternative, detectable by High-Throughput Sequencing, microbial indicators, including both faecal- and sewage-associated, allowed uncovering the different sources of pollution in coastal and anthropogenically impacted marine ecosystems, underpinning the relevance of High-Throughput Sequencing-based screening as rapid and precise method for water quality management.

Keywords: anthropogenic impact; Mediterranean Sea; marine microbiome; High-Throughput Sequencing; faecal pollution; faecal alternative indicators

Received: 3 June 2019; **Accepted:** 14 August 2019

INTRODUCTION

Coastal areas of the Mediterranean Sea are subjected to intense anthropogenic pressures, including, among others, fishing, maritime transport, wastewater discharges, industrial activities and massive tourism. These activities can significantly modify the marine ecosystems (Borja *et al.* 2013), and potentially introduce a number of microbes of faecal origin, including pathogenic species. This in turn represents a source of infection and spreading of diseases to human and aquatic populations, with important economic, sanitary and environmental consequences (Stewart *et al.* 2008; Luna, Quero and Perini 2016).

The assessment of bathing water quality is performed worldwide mainly by culture-based methods of detection of faecal indicator bacteria (i.e. *Escherichia coli* and enterococci), following standardised protocols (e.g. Directive 2006/7/EC). However, since these microorganisms can sometimes survive, grow and even adapt to the aquatic environment (Luo *et al.* 2011; Byappanahalli *et al.* 2012), additional sources of faecal contamination other than human (i.e. birds, pets, livestock, etc.) have been also kept into consideration (Byappanahalli *et al.* 2012). Besides, new studies about host-specific indicators are indicating that not a single marker organism can provide a unique specificity for a host source (McLellan and Eren 2014).

Recent studies identified alternative, more reliable and effective indicators of faecal pollution (Stewart *et al.* 2008; reviewed in McLellan and Eren 2014). Bacterial and unicellular eukaryotic taxa indicating faecal 'signatures' (human, animal, sewage and livestock) have been proposed in the last few years (Alonso-Sáez *et al.* 2007; Newton *et al.* 2013; Korajkic *et al.* 2015; de Sousa *et al.* 2017), providing new, fast and interesting tools useful in recognising faecal pollution sources.

High-Throughput Sequencing (HTS) technologies are providing unprecedented inventories of microbial communities in aquatic environments, offering new opportunities for the development of novel microbial indicators of faecal contamination in aquatic environments (Newton *et al.* 2013; Tan *et al.* 2015; Vierheilig *et al.* 2015; Luna, Quero and Perini 2016) and allowing to retrieve source-specific taxa in both bacterial and protist communities (Alonso-Sáez *et al.* 2007; Newton *et al.* 2013; Fisher *et al.* 2015; Korajkic *et al.* 2015; Luna, Quero and Perini 2016; de Sousa *et al.* 2017).These findings make HTS-based approach a powerful tool in assessing bathing and coastal water quality.

In the present study, we performed HTS sequencing of the 16S rRNA and 18S rRNA genes, in order to describe the prokaryotic and unicellular eukaryotic community composition and to explore the presence and variability of bacterial and protist faecal indicators, both traditional and alternative, at a Mediterranean coastal site (Aci Castello, Southern Mediterranean Sea), which is characterised by massive anthropogenic impact and sanitary risk because of the presence of a wastewater output within a marine protected area. The study was carried out comparing two contrasting sampling stations, a human-impacted versus a control one, before and after the bathing summer season.

MATERIALS AND METHODS

Sampling stations

Within the Marine Protected Area 'Isole Ciclopi' at Aci Castello (Catania, Italy) (Fig. 1), two stations, representatives of opposite conditions, were selected for sampling. The first station was located close to the urban wastewater output (COL—37°33.916′N 015°09.890′E) and the second was located about 400 m further off (PAN—37°34.019′N 015°10.144′E). Sampling was performed in April (hereafter defined COL1 and PAN1, respectively) and September (COL2 and PAN2, respectively), corresponding to a minimum and a maximum of anthropogenic pressure, which includes both touristic and wastewater loads. Five litres of seawater were collected from surface water using sterilised Pyrex bottles. Water temperature was measured at each sampling time and other environmental parameters (Table S1, Supporting Information) were retrieved from the datasets available at the Regional Agency of Environmental Protection website (ARPA Sicilia). These parameters were chosen as widely used indicators of nutrient enrichment (phosphorus, nitrogen) (Andersen, Conley and Hedal 2004) and of eutrophication (chlorophyll *a*). After collection, samples were filtered immediately through 0.22-µm Sterivex filters (Millipore Corp., Billerica, MA, USA) using a peristaltic pump. After removing the excess of water, Sterivex filters were stored at −20°C, until DNA extraction.

DNA extraction and Illumina sequencing

DNA was extracted from Sterivex filters using the PowerWater® Sterivex DNA Isolation Kit (MoBio Laboratories Inc., California, USA), according to the manufacturer's instructions. DNA concentrations were determined fluorometrically with Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). DNA samples were stored at −20°C until sequencing.

Illumina Miseq V3 sequencing was carried out at LGC Genomics (Berlin, Germany) on the V3 and V4 regions of the 16S rRNA gene using a modified 341F (5′-CCTAYGGGRBGCASCAG-3′) and a modified 806R (5′-GGACTACNNGGGTATCTAAT-3′) universal bacterial primer, to cover both Archaea and Bacteria domains (Sundberg *et al.* 2013). 18S rRNA gene sequencing was carried out on the V4 region using the primer pair: Eu565F (5′-CCAGCASCYGCGGTAATTCC-3′) and Eu981R (5′-ACTTTCGTTCTTGATYRATGA-3′) (Stoeck *et al.* 2010).

Sequence analysis of prokaryotic 16S rDNA gene libraries

Sequencing adaptors and primers were removed from raw reads, and reads shorter than 100 bp were excluded from downstream analysis. Paired-end sequences were joined with BBMerge 34.48. Read quality scores of each sample were checked with FastQC and the quality filtration step was carried out with split_libraries_fastq.py script by QIIME (Quantitative Insights Into Microbial Ecology, software package version v1.9.1) (Caporaso *et al.* 2010). Chimeras were detected with the QIIME's script identify_chimeric_seqs.py by using usearch61 algorithm (Edgar *et al.* 2011). Reads were clustered into operational taxonomic units (OTUs) using UCLUST v1.2.22 (Edgar 2010) with a >97% similarity threshold with an open-reference OTU picking strategy and default settings. Chimera checking and taxonomy assignment were performed using Greengenes 13.8 as reference database (DeSantis *et al.* 2006). Sequencing summary is reported in the Supporting Information. Sequences were submitted to the Sequence Read Archive (accession number SRP134209).

Sequence analysis of eukaryotic 18S gene libraries

The 18S libraries were pre-cleaned as described earlier for the 16S reads. Chimera checking was run *de novo* with the QIIME

Figure 1. Map of the Marine Protected Area 'Isole Ciclopi' (Acitrezza, hamlet of Aci Castello, Catania, Italy). White line delimitates the full protection area, red line the medium level one and yellow line includes the buffer zone. Sampling stations are indicated.

script identify_chimeric_seqs.py and the reads were dereplicated and clustered into OTUs with UCHIME algorithm (Edgar *et al.* 2011) using 97% of threshold for similarity; this clustering procedure also involves an additional check for chimeras. Taxonomic identity was assigned to each OTU blasting the most abundant read against the PR2 SSU reference database (Guillou *et al.* 2013) with an e-value threshold of 10^{-20}. The best 20 hits were kept for each OTU and, in case of equality, the taxonomic assignation was kept at the level of the last common ancestor. The OTU table was obtained by means of the make_otu_table.py script (QIIME). Sequencing summary is reported in the Supporting Information. Sequences have been submitted to the Sequence Read Archive (accession number SRP134209).

Data and statistical analyses

Alpha diversity indices of species richness (i.e. number of OTUs), exponential Shannon diversity and Pielou's evenness were calculated using iNEXT package (Hsieh, Ma and Chao 2016) on both prokaryotic and protist OTU tables. Estimates of true diversity of the aforementioned indices were extrapolated for each sample at the value of double the mean sample completeness (i.e. sequences' coverage). Pielou's evenness was computed according to the formula $J = H/\log(S)$, where J = Pielou's evenness, H = exponential Shannon diversity and S = species richness as calculated in iNEXT. Cluster analyses were performed on a resemblance matrix based on the Bray–Curtis index. All these analyses were performed in R (version 3.4.0) using the *vegan* package (Oksanen *et al.* 2014) and the *ggplot* and *ggdendro* packages for plotting (Wickham 2006). The analysis of similarity (ANOSIM) was computed in R operating directly on dissimilarity matrix (Anderson and Walsh 2013). Within prokaryotic communities' data, we searched for OTUs identified as belonging to the traditional faecal indicator taxa (i.e. the family Enterobacteriaceae, also including *Escherichia* and *Enterococcus* genera) and alternative faecal indicator taxa (i.e. faeces-associated bacterial families: Bacteroidaceae, Porphyromonadaceae, Clostridiaceae, Lachnospiraceae and Ruminococcaceae; sewage-associated bacterial genera: *Acinetobacter*, *Arcobacter* and *Trichococcus*), according to the approach proposed by Newton *et al.* (2013) and Luna *et al.* (2016). Within eukaryotic communities, we searched for OTUs considered to be good candidates of water quality assessment (i.e. Oligohymenophorea, Saccharomycetales and Apicomplexa) as suggested by recent literature (Korajkic *et al.* 2015; de Sousa *et al.* 2017). OTU richness for these indicator taxa was calculated on rarefied OTU tables to minimise the bias due to the different sequence coverage among samples. Rarefied tables were computed by randomly sub-sampling at the lowest number of reads found among samples.

RESULTS AND DISCUSSION

Environmental parameters

Environmental conditions at the two sampling times are reported in Table S1 (Supporting Information). Overall, September samples were characterised, as expected, by higher temperature than April, as well as by higher values of phosphorus,

nitrogen and chlorophyll *a*. Lower values of dissolved oxygen in September than in April were observed.

Prokaryotic community composition

The highest true species richness was detected in COL2 station, while PAN and COL1 stations had the highest values of Shannon diversity and Pielou equitability indices (Table S2, Supporting Information). These patterns are likely linked to the urban wastewater output close to the COL station, which, especially in the post-tourist season, represents a source of new microorganisms and organic and inorganic matter, possibly leading to an increase of copiotrophic organisms and less evenly distributed prokaryotic communities (Ho, Lonardo and Bodelier 2017).

Cluster analyses (Fig. 2A, left panel) showed that the communities clustered according to station rather than by season. ANOSIM confirmed the presence of a significant difference in community composition at the OTU level between the two stations ($P < 0.01$).

Prokaryotic community composition in PAN and COL at the two different sampling times is shown in Fig. 2A. Prokaryotic community in PAN1 was mainly represented by members of the phyla Alphaproteobacteria (50.9%), Bacteroidetes (21%), Gammaproteobacteria (13.6%) and Cyanobacteria (7.7%). Pelagibacteraceae accounted for the majority of Alphaproteobacteria, being 35.8% of the total, while Rhodobacteraceae contributed to a further 6% of the total Alphaproteobacteria composition. In PAN2, a strong dominance of Bacteroidetes (51.6%) in the total prokaryotic community was observed. Alphaproteobacteria (17.2%) were also abundant, but appreciably less than those in PAN1, followed by Cyanobacteria (17%) and Gammaproteobacteria (9.8%).

Prokaryotic community of sample COL1 was mainly composed of Epsilonproteobacteria (41.3%), Bacteroidetes (17.2%), Gammaproteobacteria (12.7%), Betaproteobacteria (12.3%), Firmicutes (5.4%) and Fusobacteria (4.6%). Alphaproteobacteria and other less represented phyla only accounted for 2.4% and 3% of the total community, respectively. At the genus level, the most abundant taxon in COL1 was represented by *Arcobacter* (36.7%). At the COL2 station, the prokaryotic community was dominated by Bacteroidetes (41.5%), followed by Epsilonproteobacteria (19.6%), Alphaproteobacteria (16.6%), Gammaproteobacteria (9.9%), Betaproteobacteria (6.5%) and Firmicutes (4.2%) (Fig. 2A). In both PAN2 and COL2, Bacteroidetes increased in percentage. However, at each of the stations, different families contributed to Bacteroidetes fraction (Table 1). In PAN2, Flavobacteriaceae accounted for the majority of the phylum, being 35.4% of the total sample composition; conversely, in COL2, Chryomorphaceae was the most represented family accounting for 32.8% of Bacteroidetes. Furthermore, we observed the presence of the *Bacteroides* genus only in the COL station, nearby the wastewater output (Table 1).

Overall, PAN community composition showed a typical, unimpacted surface marine prokaryotic community, with a dominance of Alphaproteobacteria, which included the order Rhodobacterales (largely *Roseobacter*) and SAR11 clade, the most abundant bacterial taxon in the epipelagic Mediterranean layer in both coastal and open sea settings (Luna 2015). Gammaproteobacteria and Cyanobacteria were also found among the most abundant bacterial groups in PAN samples, at percentages comparable to those previously observed in the Mediterranean Sea (Alonso-Sáez *et al.* 2007; Quero and Luna 2014; Luna 2015). Members of the phylum Bacteroidetes were also relatively abundant, and it has been shown to represent the second most abundant bacterial phylum in sea surface, especially in coastal areas (Alonso-Sáez *et al.* 2007; Díez-Vives, Gasol and Acinas 2014). Bacteroidetes being important players in the degradation of complex and polymeric organic matter (Kirchman 2002), we assume that their increase after summer results from the higher load of organic matter into the environment, likely due to the increased human activities and density in the area.

Conversely, COL communities, regardless of the sampling time, were enriched in Epsilonproteobacteria and Betaproteobacteria, and relatively high abundance of Firmicutes and Fusobacteria was also observed. Furthermore, at OTU level, the most abundant OTU in the COL station belonged to the family Cryomorphaceae (class Flavobacteria, phylum Bacteroidetes), while the second most abundant OTU was identified in the genus *Arcobacter* (class Epsilonproteobacteria).

Eukaryotic community composition

A possible effect of the urban wastewater output was also detectable in the protist alpha diversity results (Table S3, Supporting Information) as COL2 had the highest values of Shannon diversity and Pielou's evenness. Conversely, the highest species richness was observed in COL1 and PAN1. Cluster analysis (Fig. 2B, left panel) also showed that the eukaryotic communities clustered by season. However, in spring, COL1 composition was a stand-alone (Fig. 2B, left panel).

Eukaryotic community composition of PAN and COL station is shown in Fig. 2B. Protist community at PAN1 was mainly composed of Dinophyta (56.69%) and Archaeplastida (22.41%), followed by Ciliophora (8.75%), Stramenopiles (3.12%) and Hacrobia (1.48%). After summertime, at PAN, Dinophyta (65.90%) and Archaeplastida (27.11%) still represented the most abundant phyla; however, Ciliophora and Hacrobia were no more represented among the most abundant groups, and Stramenopiles decreased (1.31%). At the COL1 station, we observed a dominance of Ciliophora (48.36%), followed by Archaeplastida (15.2%), Stramenopiles (7.34%) and Dinophyta (3.06%). At the COL2 station, a strong increase in Dinophyta (52.63%) was observed, and Archaeplastida (21.84%), Fungi (7.75%) and Ciliophora (3.77%) were among the other most represented phyla in September. As reported in detail in Table 2, the phylum Dinophyta was mainly composed of the class Dinophyceae. The phylum Ciliophora was represented by Oligohymenophorea in COL1 and Spirotrichea in PAN1. Archaeplastida were mainly composed of pico- and microalgae belonging to the classes Chlorodendrophyceae and Mamiellophyceae.

Our results showed that protist community at COL1 was remarkably different from PAN and COL2 assemblages. In fact, while Dinophyta prevailed in PAN and COL2, heterotrophic Ciliophora predominated in COL1. Ciliophora are important components of aquatic environments for several reasons, mainly because they act as 'intermediate' component in food webs that link nutrient flow between smaller microorganisms and larger metazoans (Rossi *et al.* 2016). COL2 had lower abundance of Syndiniales, which are typical endosymbiont of microeukaryotes, crustaceans, fish and algae (van den Hoek, Mann and Jahns 1995); this might suggest a depletion/reduction of such organisms in COL2. Relevant is the detection of fungal sequences almost exclusively in COL2; indeed, Fungi have recently been acknowledged to play a relevant role in the degradation of organic matter (Grossart *et al.* 2019) that is constantly released by the nearby

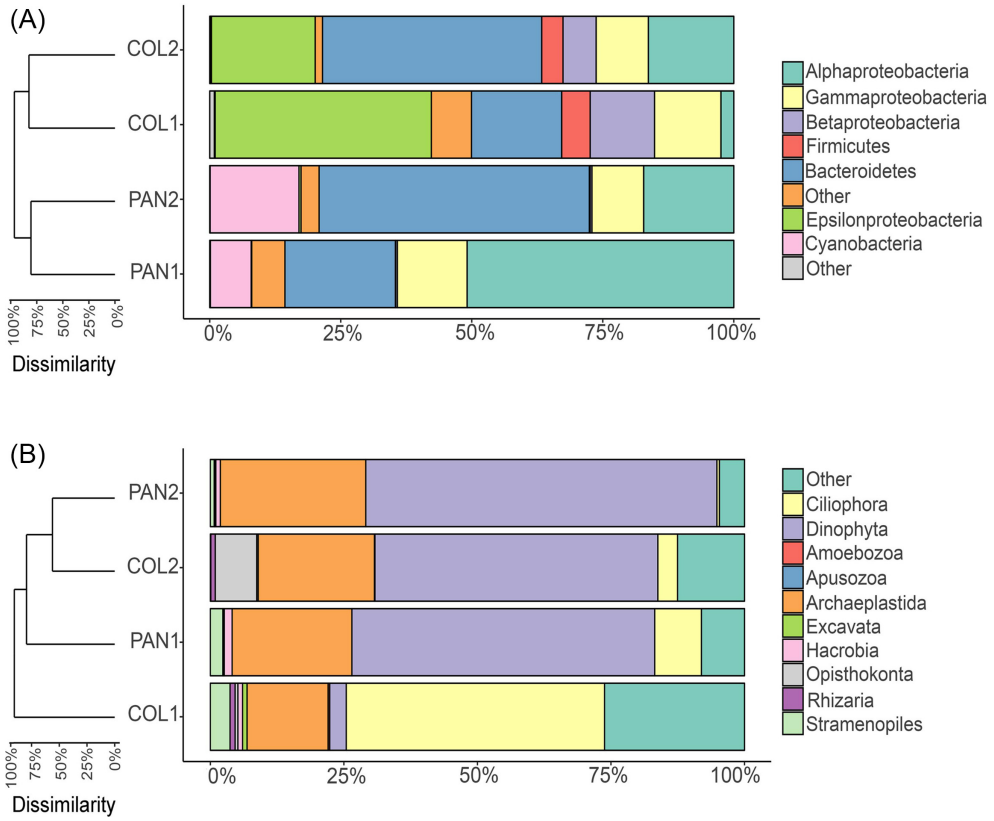

Figure 2. (A) Prokaryotic communities' composition of the two stations at both sampling times at phylum and class (for Proteobacteria) levels. 'Others' cumulatively indicates taxa with relative abundances <1%. **(B)** Protist communities' composition at phylum level. On the left of each panel shown are the clusters obtained based on the Bray–Curtis dissimilarity matrix of the 16S (A) and 18S (B) OTU tables.

Table 1. Prokaryotic community composition at the phylum/class and family level.

Phylum/class	Family	PAN1	PAN2	COL1	COL2
Cyanobacteria		7.6	16.9	0.10	0.20
	Synechoccaceae	7.0	16.3		0.20
Bacteroidetes		21.0	51.6	17.20	41.50
	Flavobacteriaceae	13.7	35.4		2.70
	Cryomorphaceae	2.6	6.6		32.80
	Bacteroidaceae	0.01	0.09	6.80	2.38
	Porphyromonadaceae	0.01	0.08	4.30	1.57
Alphaproteobacteria		50.9	17.2	2.40	16.60
	Pelagibacteraceae	35.8	1.7	0.74	0.02
	Rhodobacteraceae	6.3	8.5	0.78	16.90
Fusobacteria				4.50	0.10
Firmicutes		0.10	0.20	5.40	4.20
	Lachnospiraceae		0.02	2.00	1.26
Epsilonproteobacteria		0.10	0.40	41.30	19.60
	Campylobacteraceae	0.08	0.38	38.57	19.47
	Helicobacteraceae	0.01	0.01	2.69	0.16

Values are reported as percentage.

wastewater outlet. Overall, eukaryotic communities' composition described earlier reflected the structure of eukaryotic communities detected in the photic zone of similar stations in the Mediterranean Sea as previously reported by de Vargas *et al.* (2015).

Indicators of faecal pollution

HTS data were used to identify and relatively quantify faecal indicator taxa, either traditional or alternative (Newton *et al.* 2013; Luna, Quero and Perini 2016).

Table 2. Composition of protist communities at the phylum and order/class level.

Phylum	Order/class	PAN1	PAN2	COL1	COL2
Dinophyta		56.69	65.90	3.06	52.63
	Dinophyceae	48.92	55.83	2.55	50.07
	Syndiniales	7.68	10.05	0.50	2.56
Ciliophora		8.75	0.49	48.36	3.77
	Oligohymenophorea			46.87	3.09
	Spirotrichea	8.67	0.49	0.61	0.58
Archaeplastida		22.41	27.11	15.20	21.84
	Chlorodendrophyceae	3.46	26.51	0.09	14.74
	Chlorophyceae			10.15	
	Mamiellophyceae	13.32	0.48	0.50	0.20
	Florideophyceae	5.56	0.00	3.85	6.46
Stramenopiles		3.12	1.31	7.34	0.85
	Bacillariophyta	1.85	0.65	3.45	
	Labyrinthulea	0.58	0.13	2.85	0.42
Fungi		0.17	0.11	0.51	7.75

Values are reported as percentage.

Table 3. Relative abundance of prokaryotic and eukaryotic indicators at both sites.

	Relative abundance	
	PAN	COL
Prokaryotic indicators		
Traditional		
Enterobacteriaceae	0.00	0.30
Faeces-associated		
Bacteroidaceae	0.01	1.63
Porphyromonadaceae	0.00	0.16
Lachnospiraceae	0.04	2.94
Clostridiaceae	0.05	4.62
Ruminococcaceae	0.01	1.19
Sewage-associated		
Acinetobacter	0.03	2.01
Arcobacter	0.22	27.97
Eukaryotic indicators		
Oligohymenophorea	0.00	9.51
Saccharomycetales	0.00	0.09
Apicomplexa	0.00	0.21

Data are obtained by calculating the mean value between relative read percentages at each sampling time. Values are reported as percentage.

Table 4. Cumulative indicators' OTU numbers retrieved at both sites and sampling times.

	Number of OTUs	
	PAN	COL
Prokaryotic indicators		
Traditional		
Enterobacteriaceae	3	41
Faeces-associated		
Bacteroidaceae	14	144
Porphyromonadaceae	9	82
Lachnospiraceae	8	251
Clostridiaceae	0	31
Ruminococcaceae	10	171
Sewage-associated		
Acinetobacter	9	103
Arcobacter	29	132
Eukaryotic indicators		
Oligohymenophorea	0	30
Saccharomycetales	2	19
Apicomplexa	3	10

In PAN station, the number of reads of traditional indicators (i.e. Enterobacteriaceae) accounted for a low number of sequences corresponding to 0.003% of the total sequences (Table 3). Faeces-associated indicators such as Lachnospiraceae (0.014%), Clostridiaceae (0.001%), Porphyromonadaceae (0.042%), Bacteroidaceae (0.050%) and Ruminococcaceae (0.013%) were higher than traditional indicators. Sewage-associated indicators mainly contributed to the faecal-associated sequences in PAN samples, with a remarkable presence of *Arcobacter* (0.219%) followed by *Acinetobacter* (0.030%).

At COL station, the cumulative number of sequences linked to faecal contamination accounted for 70–300 times those present at PAN station (Table 3). Enterobacteriaceae accounted for 0.300% of sequences. Faeces-associated alternative indicators were mainly represented by Bacteroidaceae (4.622%) and Porphyromonadaceae (2.937%), even though all the other alternative faeces-associated indicators were detected at high numbers. A striking number of sequences belonging to the genus *Arcobacter* (27.973%) were observed at COL. We could not detect any *Trichococcus* sequence at the PAN station nor at COL.

As expected, being subjected to a significant anthropogenic pressure and likely receiving important loads of faecal bacteria, we observed an overall increase of bacterial signatures of faecal pollution at COL station. Interestingly, as previously reported in other works (Luna, Quero and Perini 2016), the relative abundance of faecal- (Bacteroidaceae, Porphyromonadaceae, Lachnospiraceae, Ruminococcaceae) and sewage-associated (*Arcobacter*, *Acinetobacter*) bacteria always exceeded that of the traditional indicators (Enterobacteriaceae). We also calculated OTU richness for each of the faecal indicator taxa. OTUs' richness varied between the two stations, ranging from 0 to 251 OTUs depending on the faecal indicator (Table 4). At the PAN station, the cumulative number of OTUs (i.e. the sum of all OTUs recorded at the two sampling times) belonging to the traditional indicators was 3 for Enterobacteriaceae. The number of OTUs within the faeces-associated and the sewage-associated indicators was higher, ranging from 0 to 29 (Table 4).

At COL, the cumulative number of OTUs (i.e. the sum of all OTUs recorded at the two sampling times) belonging to the traditional indicators was 41 for Enterobacteriaceae, while, within the faeces-associated and the sewage-associated indicators, it varied from 31 to 251 (Table 4). An increment in the number of OTUs of these indicators could be noticed in COL, and a more relevant increment was recorded for the alternative indicators than for traditional ones.

In our study, both traditional and alternative indicators were recorded. However, we show that traditional indicators can underestimate the actual faecal contamination in recreational water samples, and foster the need for the use of alternative indicators. The alternative indicators, especially found at COL, allowed us to acknowledge the existence of multiple sources of contamination. In fact, Bacteroidales and Firmicutes are a major constituent of the faecal microbiome of humans and other animals (McLellan and Eren 2014). According to McLellan and colleagues, Lachnospiraceae, within Clostridiales, might deserve important functional roles specific to humans, as they were found in both sewage and human faecal samples, but not in cattle or chicken faeces. In contrast, Ruminococcaceae, despite being a good alternative indicator, were shared by humans, cows and, to a lesser extent, chickens. The higher abundance of Firmicutes assessed in COL samples supports a contamination of untreated sewage water with likely human and livestock origins. Besides, we found evidence in COL of high sewage pollution, *Arcobacter* being the most abundant taxon, which is considered as an emergent enteropathogen and a potential zoonotic agent (Collado and Figueras 2011). It has been suggested that sewage may be an important reservoir for these microbes (Collado *et al.* 2008), and it has been also recovered from anthropogenically impacted coastal stations in the Mediterranean Sea. Our data highlight the importance of *Arcobacter* as signature of anthropogenic impact in coastal environments.

Unlike investigations of prokaryotic indicators, there is a long road ahead in the quest for eukaryotic marker species of pollution, even though aquatic eukaryotes, and in particular protists, are also important bioindicators of water quality. Korajkic *et al.* (2015) suggested that ciliates from the Oligohymenophorea may be considered good candidates for water quality assessment. Yet, some Ciliophora (bacterivores) are also able to reduce bacterial load in wastewater treatment plants, thanks to their ability to resist to extreme environmental conditions such as anoxia and relatively high heavy-metal concentrations (Madoni 2011). Furthermore, Korajkic *et al.* (2015) also mentioned that Fungi, such as Saccharomycetales, and Apicomplexa (Alveolata) may help in the recognition of the presence of organic matter and contamination from cattle (Korajkic *et al.* 2015). Fungi have a well-known role in decomposing organic matter (Grossart *et al.* 2019) and they represent a key component during wastewater treatment processing in the activated sludge phase (Matsunaga, Kubota and Harada 2014). The possible presence of Apicomplexa species would suggest livestock contamination. Indeed, these genera of Apicomplexa are described as enteroparasites of cattle, other domestic and wild animals, and even humans (Feleke *et al.* 2008; de Sousa *et al.* 2017).

In our study, looking at the above-mentioned taxa, i.e. Oligohymenophorea, Saccharomycetales and Apicomplexa, we found that the number of reads accounted for less than 0.0012% of total sequences at PAN station. On the other hand, at COL station a high number of sequences belonging to Oligohymenophorea (9.51%) was recorded. A certain amount of Saccharomycetales (0.09%) and Apicomplexa (0.21%) was also present in COL (Table 3). Similar to prokaryotes, we compared the number of OTUs affiliated with each of these groups. The number of OTUs varied between the two stations, ranging from 0 to 30 OTUs depending on the taxa. At the PAN station, the cumulative number of OTUs (i.e. the sum of all OTUs recorded at the two sampling times) belonging to them was 0 for Oligohymenophorea, 2 for Saccharomycetales and 3 for Apicomplexa. At the COL station, the cumulative number of OTUs was higher: 30 for Oligohymenophorea, 19 for Saccharomycetales and 10 for Apicomplexa (Table 4).

In our study, the presence of Oligohymenophorea, Fungi and Apicomplexa coming together exclusively at the COL station suggests that a possible association may be present that would be helpful as a signature of strong human/livestock contamination. Moreover, the presence of Apicomplexa mutually supports the results of the prokaryotic community analysis, hence underpinning the likelihood of a strong human/livestock contamination within the sewage. The high abundance of Oligohymenophorea in COL community almost exclusively in April may be ascribed to seasonal rain events, especially considering that the months preceding the sampling were unusually rainy. This would support the hypothesis that COL sewage would be a mixture of microorganisms influenced by both anthropogenic origin and freshwater environmental sources, such as rainwater infiltrations, as suggested in other systems by McLellan *et al.* (2013).

CONCLUSIONS

Our results show that COL area represents a large reservoir of faecal bacteria in the Aci Castello sea waters (Eastern Sicily), with potential to contaminate the nearby coasts, including the Marine Protected Area. This scenario poses health consequences due to the presence of bathing beaches and touristic destinations. Nevertheless, we point out here that further studies are needed to investigate the ability of the different bacterial populations, including the faeces- and sewage-associated bacteria, to persist or decay once they reach the marine environment. If so, such bacteria can be potentially transported towards adjacent coastal areas in the presence of specific hydrological conditions.

We showed that the coupled analyses of the three faecal signatures (traditional, faecal- and sewage-associated bacteria) are useful to discern among different pollution sources in coastal areas. Moreover, the parallel analysis of HTS-based studies of 16S and 18S amplicons allowed to achieve a more robust assessment of human-related impact in aquatic environment. International effort is being spent towards standardisation of HTS-driven microbial indicators of aquatic pollution; our study would help laying the foundations of microbial water quality assessment with the potential of being translated into actionable data for water quality managers.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSLE* online.

ACKNOWLEDGEMENTS

The authors thank Dott. Salvatore Casabianca and Dott. Antonino Brancato (Agenzia Regionale Protezione Ambiente, Regione Sicilia, Struttura Territoriale di Catania, U.O. Monitoraggi Ambientali — ARPA Sicilia) for the access to the ARPA database and for useful discussion. The authors also thank Prof.

Stefania Stefani (Laboratory of Medical Microbiology and Antibiotic Resistance, University of Catania, Italy) for her constant support and Dr. Fabio Viglianisi for technical assistance. A special thanks to Dr. Emanuele Mòllica, former Director of the Marine Protected Area 'Isole Ciclopi'. Memory of his passion for the sea coupled with professionalism and kindness will always stay with us. The authors are most grateful to Dr. Gian Marco Luna for critical reading of the manuscript. The graphical abstract was created with Biorender.com.

FUNDING

This research was supported by the FP7 European Project 'Winning Applications of nanoTEchnology for Resolutive hydropurification—WATER' (grant agreement no. 316082) and by the Annual Research Plan 2016-18 of the Department of Biological Geological Environmental Sciences, University of Catania (grant no. 22722132110). The authors acknowledge the PON project Bio-nanotech Research and Innovation Tower (BRIT), financed by the Italian Ministry for Education, University and Research (MIUR) (grant no. PONa3_00136).

Conflict of interest. None declared.

REFERENCES

Alonso-Sáez L, Balagué V, Sà EL *et al.* Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* 2007;**60**:98–112.

Andersen JH, Conley DJ, Hedal S. Palaeoecology, reference conditions and classification of ecological status: the EU Water Framework Directive in practice. *Mar Pollut Bull* 2004;**49**:283–90.

Anderson MJ, Walsh DCI. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecol Monogr* 2013;**83**:557–74.

Borja A, Elliott M, Andersen JH *et al.* Good environmental status of marine ecosystems: what is it and how do we know when we have attained it? *Mar Pollut Bull* 2013;**68**:1–3.

Byappanahalli MN, Nevers MB, Korajkic A *et al.* Enterococci in the environment. *Microbiol Mol Biol Rev* 2012;**76**:685–706.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.

Collado L, Figueras MJ. Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clin Microbiol Rev* 2011;**24**:174–92.

Collado L, Inza I, Guarro J *et al.* Presence of *Arcobacter* spp. in environmental waters correlates with high levels of fecal pollution. *Environ Microbiol* 2008;**10**:1635–40.

DeSantis TZ, Hugenholtz P, Larsen N *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.

de Sousa KCM, Fernandes MP, Herrera HM *et al.* Molecular detection of *Hepatozoon* spp. in domestic dogs and wild mammals in southern Pantanal, Brazil with implications in the transmission route. *Vet Parasitol* 2017;**237**:37–46.

de Vargas C De, Audic S, Henry N *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Ocean Plankton* 2015;**348**:1–12.

Díez-Vives C, Gasol JM, Acinas SG. Spatial and temporal variability among marine Bacteroidetes populations in the NW Mediterranean Sea. *Syst Appl Microbiol* 2014;**37**:68–78.

Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–1.

Edgar RC, Haas BJ, Clemente JC *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;**27**:2194–200.

Feleke A, Petros B, Lemecha H *et al.* Study on monthly dynamics of ticks and seroprevalence of *Anaplasma marginale*, *Babesia bigemina* and *Theileria mutans* in four indigenous breeds of cattle in Ghibe Valley, Ethiopia. *SINET Ethiop J Sci* 2008;**31**:11–20.

Fisher JC, Murat Eren A, Green HC *et al.* Comparison of sewage and animal fecal microbiomes by using oligotyping reveals potential human fecal indicators in multiple taxonomic groups. *Appl Environ Microbiol* 2015;**81**:7023–33.

Grossart H-P, Van den Wyngaert S, Kagami M *et al.* Fungi in aquatic ecosystems. *Nat Rev Microbiol* 2019;**6:** 339–54.

Guillou L, Bachar D, Audic S *et al.* The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 2013;**41**:D597–604.

Ho A, Lonardo DP Di, Bodelier PLE. Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol Ecol* 2017;**93**:fix006.

Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol Evol* 2016;**7**:1451–6.

Kirchman DL. The ecology of *Cytophaga– Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* 2002;**39**:91–100.

Korajkic A, Parfrey LW, McMinn BR *et al.* Changes in bacterial and eukaryotic communities during sewage decomposition in Mississippi river water. *Water Res* 2015;**69**:30–9.

Luna GM. Diversity of marine microbes in a changing Mediterranean Sea. *Rend Lincei* 2015;**26**:49–58.

Luna GM, Quero GM, Perini L. Next generation sequencing reveals distinct fecal pollution signatures in aquatic sediments across gradients of anthropogenic influence. *Adv Oceanogr Limnol* 2016;**7**:1–7.

Luo C, Walk ST, Gordon DM *et al.* Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci USA* 2011;**108**:7200–5.

Madoni P. Protozoa in wastewater treatment processes: a minireview. *Ital J Zool* 2011;**78**:3–11.

Matsunaga K, Kubota K, Harada H. Molecular diversity of eukaryotes in municipal wastewater treatment processes as revealed by 18S rRNA gene analysis. *Microbes Environ* 2014;**29**:401–7.

McLellan SL, Eren AM. Discovering new indicators of fecal pollution. *Trends Microbiol* 2014;**12**:697–706.

McLellan SL, Newton RJ, Vandewalle JL *et al.* Sewage reflects the distribution of human faecal Lachnospiraceae. *Environ Microbiol* 2013;**15**:2213–27.

Newton RJ, Bootsma MJ, Morrison HG *et al.* A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. *Microb Ecol* 2013;**65**:1011–23.

Oksanen J, Blanchet FG, Kindt RL *et al.* Vegan: Community Ecology Package. R Package Version 2.2-0. 2014.

Quero GM, Luna GM. Diversity of rare and abundant bacteria in surface waters of the Southern Adriatic Sea. *Mar Genomics* 2014;**17**:9–15.

Rossi A, Boscaro V, Carducci D *et al.* Ciliate communities and hidden biodiversity in freshwater biotopes of the Pistoia province (Tuscany, Italy). *Eur J Protistol* 2016;**53**:11–9.

Stewart JR, Gast RJ, Fujioka RS *et al.* The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. *Environ Health* 2008;**7**:1–14.

Stoeck T, Bass D, Nebel M *et al.* Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 2010;**19**:21–31.

Sundberg C, Al-Soud WA, Larsson M *et al.* 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. *FEMS Microbiol Ecol* 2013;**85**:612–26.

Tan B, Ng C, Nshimyimana JP *et al.* Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. *Front Microbiol* 2015;**6**:1027.

van den Hoek C, Mann DG, Jahns HM. Algae, an introduction to phycology. *J North Am Benthol Soc* 1995;**82**:615–6.

Vierheilig J, Savio D, Ley RE *et al.* Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. *Water Sci Technol* 2015;**72**:1962–72.

Wickham H. ggplot2: elegant graphics for data analysis. *J Stat Software* 2006;**35**:2006.