Implementation and integration of microbial source tracking in a river watershed monitoring plan

Elisenda Ballesté [a,1], Katalin Demeter [a,2], Bartholomew Masterson [a], Natàlia Timoneda [b], Laura Sala-Comorera [a], Wim G. Meijer [a,*]

[a] UCD School of Biomolecular and Biomedical Science, UCD Earth Institute, UCD Conway Institute, University College Dublin, Dublin, Ireland
[b] Computational Genomics Laboratory, Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Catalonia, Spain

HIGHLIGHTS

- Sensitivity and specificity of MST markers must be validated in the area of interest.
- The human marker is diluted after rainfall events suggesting human point pollution.
- The ruminant marker was more abundant in summer and autumn in agriculture areas
- The ruminant marker correlated with rainfall and turbidity suggesting diffuse pollution from run-off.
- MST markers are useful tools to improve water management in river catchments.

GRAPHICAL ABSTRACT
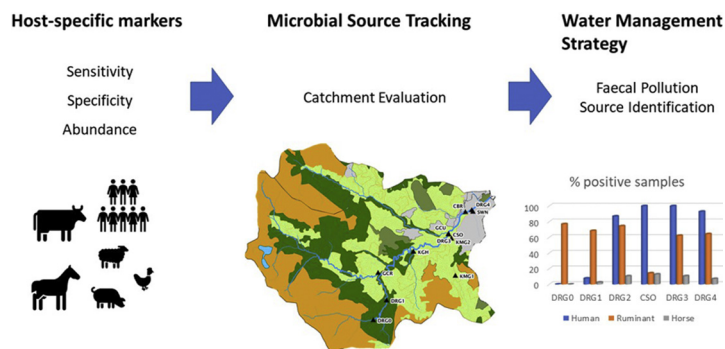
ABSTRACT

Fecal pollution of water bodies poses a serious threat for public health and ecosystems. Microbial source tracking (MST) is used to track the source of this pollution facilitating better management of pollution at the source. In this study we tested 12 MST markers to track human, ruminant, sheep, horse, pig and gull pollution to assess their usefulness as an effective management tool of water quality. First, the potential of the selected markers to track the source was evaluated using fresh fecal samples. Subsequently, we evaluated their performance in a catchment with different impacts, considering land use and environmental conditions. All MST markers showed high sensitivity and specificity, although none achieved 100% for both. Although some of the MST markers were detected in hosts other than the intended ones, their abundance in the target group was always several orders of magnitude higher than in the non-target hosts, demonstrating their suitability to distinguish between sources of pollution. The MST analysis matched the land use in the watershed allowing an accurate assessment of the main sources of pollution, in this case mainly human and ruminant pollution. Correlating environmental parameters including temperature and rainfall with MST markers provided insight into the dynamics of the pollution in the catchment. The levels of the human marker showed a significant negative correlation with rainfall in human polluted areas suggesting a dilution of the pollution, whereas at agricultural areas the ruminant marker increased with rainfall. There were no seasonal differences in the levels of human marker, indicating human

---

* Corresponding author.
   E-mail address: wim.meijer@ucd.ie (W.G. Meijer).
[1] Current address: Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Catalonia, Spain.
[2] Current address: Centre for Water Resource Systems, TU Wien, Vienna, Austria.

pollution as a constant pressure throughout the year, whereas the levels of the ruminant marker was influenced by the seasons, being more abundant in summer and autumn. MST analysis integrated with land use and environmental data can improve the management of fecal polluted areas and set up best practice.

1. Introduction

Fecal contamination of water bodies poses a serious threat as it introduces organic matter, nitrogen, phosphorous as well as potential pathogens causing health risks, environmental degradation, and economic losses, in particular when drinking water supplies and recreational and shellfish harvesting waters are affected (Kay et al., 1994). A description of the catchment characteristics, including possible sources of fecal pollution, is an essential prerequisite to facilitate efficient assessment, management and remediation of the affected region. The identification and assessment of causes of pollution that might affect bathing waters is enshrined in the European Bathing Water Directive 2006/7/EC (European Comission, 2006).

A large number of library independent molecular markers to track the source of fecal pollution in water have been developed (Roslev and Bukh, 2011), which in general are based on the detection of host specific intestinal bacterial species using endpoint or quantitative PCR (Bernhard and Field, 2000; Caldwell and Levine, 2009; García-Aljaro et al., 2017; Kildare et al., 2007; Layton et al., 2006; Lu et al., 2008; Reischer et al., 2006). In comparison to other MST approaches, these library-independent techniques are cost effective, relatively easy to implement and provide fast results. DNA extracted from a single water sample can be used to analyze for the presence of multiple MST markers.

The composition of the animal and human intestinal microbiome is influenced by a number of factors, including age, diet and health status, as well as the geographical location of the host (Eckburg et al., 2005; Ley et al., 2008; Shanks et al., 2011). Host specific MST markers may therefore show distinct geographical and temporal distributions (Mayer et al., 2018; Reischer et al., 2013; Yahya et al., 2017). It is therefore essential that MST markers are validated for the geographical area where they will be used. This includes determining the: i) abundance of MST markers at the source, since low initial concentration of these markers might become a problem when fecal matter is dispersed following its release in the environment, ii) the specificity of the marker, as it may be present in other hosts albeit at low concentrations and, iii) the temporal stability in different host groups (Roslev and Bukh, 2011). It is therefore essential to establish these basic characteristics of the markers prior to their use in the field to assess water quality in a catchment.

The deployment of MST markers to assess water quality in a catchment is challenging, and a range of factors influencing the interpretation of the results of an MST analysis need to be considered. For example, fecal pollution in a catchment is frequently the result of a contamination derived from several different sources, which increases the microbial diversity and complexity of a water sample. The variation in marker composition and abundance in a water body may introduce a bias in expected patterns. In addition, fecal pollution may arise from point or diffuse sources, which may be immediately and intermittently deposited in the waterbody as is often the case for wild life, settling on fields before it eventually enters a waterbody as runoff after rainfall, in for example agricultural pollution, or be continuously discharged from waste water treatment plants or through sewer misconnections, or sporadically from combined sewer overflows (Rogers et al., 2011; Unc and Goss, 2004; Verhougstraete et al., 2015). Spatial analysis using Geographical Information Systems (GIS) are becoming a very useful tool for water assessment (Peed et al., 2011). Combining GIS with fecal indicator bacteria and MST, environmental data such as rainfall, flow and

temperature may improve water management and assessment of specific areas (Heaney et al., 2015; Nnane et al., 2011; Pascual-Benito et al., 2020; Sowah et al., 2017; Verhougstraete et al., 2015).

In this study we have gone through several steps for the further implementation of MST markers as a toolbox for water routine monitoring. 1) We have evaluated the performance of different host specific markers detecting human, cow, sheep, horse, pig, and gull pollution on fecal samples collected within the study area. Sensitivity, specificity and abundance of the markers were addressed using PCR and real time quantitative PCR in these samples. 2) The MST markers showing a better performance were selected and tested in the study area, a river catchment with a potential source of animal and human fecal pollution addressed according to the land uses during a three-year sampling period. 3) Finally, the performance of the MST markers was validated using land cover and environmental (precipitation and temperature) data for the given catchment.

## 2. Material and methods

2.1. Study site

The Dargle River and its tributary rivers drain a catchment of approximately 133 km^2 and flows into the Irish Sea in Bray, 27 km south of Dublin. The 20 km long river has a slope of approximately 2.7%, and a dry weather flow of 3 m^3 s^{-1} that may increase 100-fold during severe weather events. There are extreme variations in climate within the catchment. In the eastern, coastal part of the catchment, annual rainfall is ca. 800 mm, there are approximately 185 days with rain per year, and monthly mean rainfall figures range between 50 mm in summer and 70 mm in winter. Mean daily temperatures lie between 5 °C in winter and 15 °C in summer; mean daily sunshine duration lies between 1.6 h per day in winter and 5.6 h in summer. In the mountains, rainfall is 1500–2000 mm per year, with snow-cover lasting a month; daily temperatures are much lower and sunshine hours fewer.

Although the Dargle catchment is relatively small, it is characterized by a very diverse land use (Fig. 1, Table 1). The lower catchment is mainly urban in nature, including the towns of Kilmacanogue, Enniskerry and Bray with 900, 2700, and 32,000 inhabitants, respectively, while the upper catchment is characterized by tillage, cattle and sheep farming, forestry and peat bogs. The catchment is home to a few riding stables.

2.2. Sample collection

Water samples ($n = 354$) were collected over a period of 3 years at 10 sampling sites, located along the Dargle river (DRG1, DRG2 and DRG3) and in the main tributaries: Swan (SWN), County Brook (CBR), Kilmacanogue (KMG1 and KMG2), Killough (KGH), Glencullen (GCU) and Glencree (GCR) rivers (Fig. 1). These sampling sites were selected to be representative of the different land uses and tributaries. Sewage effluent samples ($n = 25$) were collected from the Enniskerry wastewater treatment plant (6000 p.e.) (secondary treatment with nutrient reduction) before discharging into the Dargle River. Samples were collected in sterile containers, transported on ice and stored at 4 °C until analysis.

Fresh fecal samples ($n = 181$) were collected from pigs ($n = 20$), horses ($n = 27$), sheep (n = 25), cattle ($n = 22$) and deer ($n = 15$). Pig and cow fecal samples were obtained from UCD Lyons Research
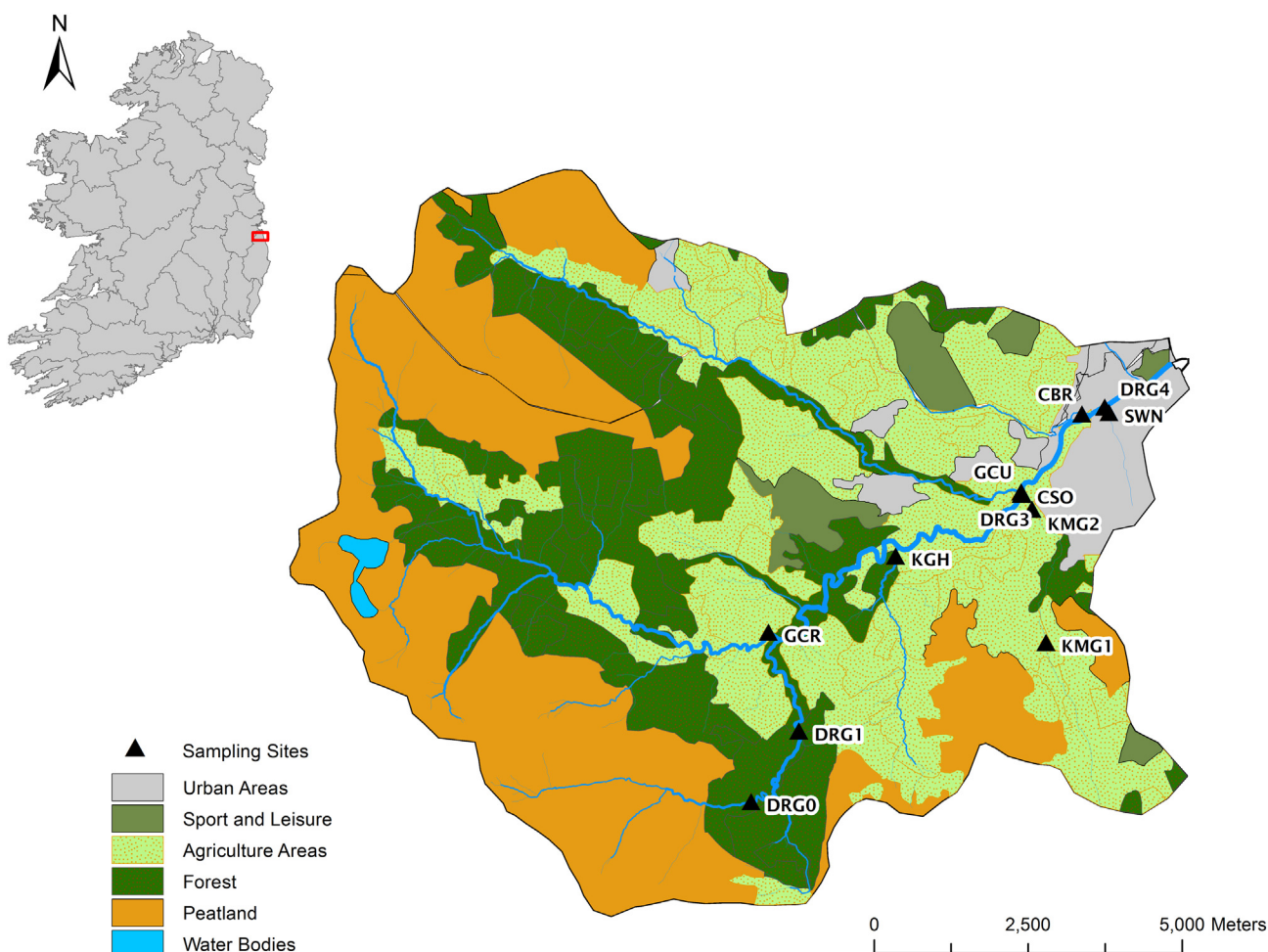
Fig. 1. Dargle catchment land uses and sampling sites. Location of the sampling sites and land uses of the Dargle river catchment.

Farm, horse fecal samples were obtained from an Irish stud farm. Deer samples were obtained from an Irish deer farm. Samples from gulls ($n = 32$) and swans ($n = 8$) were collected in Bray harbor and Irelands' Eye island. The main gull species in these environments are great black-backed gull (*Larus marinus*), herring (*Larus argentatus*) and common gulls (*Larus canus*). Samples of each animal source were collected from at least two different locations. Human samples ($n = 26$) were obtained from Crumlin Hospital (Dublin, Ireland) from individuals not receiving antibiotic treatment. Ethical permission obtained from the Ethics (Medical Research) Committee of Our Lady's Children Hospital Crumlin, Dublin (GEN/169/10).

2.3. Enumeration of FIB

The levels of total coliforms, *E. coli* and intestinal enterococci were determined using Colilert-18 and Enterolert with a Quanti-Tray/2000

Table 1
Percentage of the land uses surface affecting each sampling site in Dargle catchment. Locations and abbreviations of the sampling stations as in Fig. 1.

	DRG1	DRG2/3	KGH	KMG1	KMG2	CBR	SWN	GCU	GCR
Urban	0	0	0	0	1	2.4	96.2	4	0
Sport	0	10.1	0	3.1	2.1	26.1	0	2.3	0
Agriculture	3	52.5	70.6	55.1	59.1	60.1	2.2	36.8	14.4
Forest	31.6	33.1	5.3	1.7	6.8	11.4	1.3	26.4	35.5
Peatland	65.4	4.3	24.1	40.1	31	0	0.3	30.5	48.8
Lakes	0	0	0	0	0	0	0	0	1.3

system (IDEXX Laboratories, UK) according to manufacturer's instructions.

2.4. DNA extraction

DNA was extracted from water samples following filtering of 100 ml water using a filter with 0.2 μm pore-size (Supor 200 PES, Pall Corporation, NY). Filters were placed in 0.5 ml of GITC buffer (5 M guanidine thiocyanate, 100 mM EDTA [pH 8.0], 0.5% Sarkosyl) and frozen at −20 °C in GITC buffer until DNA extraction. DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) with some modifications as reported previously (Gourmelon et al., 2007).

DNA was extracted from 180 to 250 mg of fecal sample using the QIAamp DNA Stool MiniKit (Qiagen, UK) according to the manufacturer's instructions.

2.5. Detection of MST markers by end-point PCR and real time quantitative PCR

2.5.1. End-point PCR

Detection of the MST markers by PCR was performed using primers described previously (Table S.1). Five host-specific markers matched the 16S rRNA gene of *Bacteroidales* species: the human marker HF183, the ruminant marker CF128 (Bernhard and Field, 2000), the horse marker HoF597 and the pig markers PF163 (Dick et al., 2005) and Pig-2-Bac (Mieszkin et al., 2009).

The MST markers were amplified in a 25-μl PCR mixture containing Green GoTaq Flexi Buffer (Promega), 0.2 μM of each primer, 0.625 U

GoTaq DNA polymerase (Promega), 2 mM MgCl$_2$, 200 µM dNTP and 0.1 mg/ml nonacetylated BSA (Applied Biosystems). The reaction mixture was incubated at 95 °C for 2 min, followed by 35 (fecal sample) or 45 (water sample) cycles of 94 °C for 30s, an appropriate annealing temperature for each marker (Table S.1) for 30 s, and an extension step at 72 °C for 20 s (HoF597 and Pig-2-Bac marker) or 40 s (HF183, CF128 and PF163); followed by a final 5 min extension at 72 °C.

The 16S rRNA gene of *Catellicoccus marimammalium* was amplified using primers GullF and GullR (Table S.1) (Lu et al., 2008). The reaction mixture was incubated at 95 °C for 2 min followed by 35 cycles of 94 °C for 30 s, 63 °C for 30 s, and 72 °C for 35 s, which were followed by a final incubation at 72 °C for 5 min.

Nested-PCR was used to detect porcine and ovine mitochondrial DNA (Martellini et al., 2005) (Table S.1). The reaction mixtures were incubated at 95 °C for 2 min, followed by 35 cycles consisting of 94 °C for 40 s, 55 °C for 50 s, and at 72 °C for 45 s; followed by a final 5 min extension.

The universal primers F63 and 1389R (Table S.1) (Marchesi et al., 1998; Osborn et al., 2000) were used (0.4 µM) to amplify the 16S rRNA gene, and was to infer potential PCR inhibitors in samples causing false negative results. The reaction mixture was incubated at 95 °C for 2 min, followed by 30 cycles of 94 °C for 45 s, 55 °C for 45 s, and 72 °C for 90 s, followed by 72 °C for 5 min. Controls included filtration blanks, DNA extraction blanks, no-template PCR control and PCR positive control.

2.5.2. Real time quantitative PCR

The forward primers HF183F and CF128F were paired with the general *Bacteroidales* primer 265R and HoF597F with Bac708R (Bernhard and Field, 2000; Dick et al., 2005; Seurinck et al., 2005) (Table S.1). In addition, a CowM3 marker matching a specific gene (sialic acid-specific 9-*O*-acetylesterase secretory protein homologue) from cow-specific *Bacteroidales* was tested (Shanks et al., 2008). The levels of general *Bacteroidales* were quantified using AllBac-296R and AllBac-412R primers (Layton et al., 2006) (Table S.1). The assays were performed using a Light Cycler 480 Instrument (Roche Diagnostics, Mannheim, Germany). Reaction mixtures (10 µl) consisted of DNA Master SYBR Green I Buffer (Roche) and 0.5 µM of each primer. The amplification conditions were: an initial denaturation for 10 min at 95 °C, followed by 45 cycles at 95 °C for 15 s, 60 °C for 15 s and an extension step at 72 °C for 15 s. A melting curve analysis was carried out by heating the PCR products to 95 °C, cooled to 60 °C for 15 s and slowly reheated to 95 °C at a rate of 0.1 °C/s. PCR conditions to amplify and quantify CowM3 marker were as described (Shanks et al., 2008). Samples were run in triplicate with 1:10 fold dilution of the samples in order to detect potential inhibition effects. The results were expressed as gene copies (gc) 100 mg^{-1} or 100 ml^{-1}. Controls included no-template PCR control and PCR positive control. The detection limit was established through the analysis of serial dilution of fecal and water samples.

The limit of quantification and detection was calculated following Blanchard et al. (Blanchard et al., 2012) and was established at 3, 4, 8, 8 and 10 gc per µl of reaction for qHF183, AllBac, qHoF597, CowM3 and CF128 marker.

2.6. Quantification of land use

ArcGIS 9.3 was used to visualize and analyze CORINE land cover data (CLC2006, http://gis.epa.ie) in order to assess different land use in the Dargle catchment. Additional GIS data was obtained from the Ordnance Survey Ireland (OSi). These layers included river segments, catchment and subcatchments areas and digital elevation models.

2.7. Statistical analysis

Sensitivity and specificity for the eight microbial source tracking markers were calculated as described previously (Gawler et al., 2007).

Sensitivity (r) and specificity (s) are defined as $r = a/(a + c)$ and $s = d/(b + d)$, where a is when a fecal DNA sample is positive for the PCR marker of its own target (true positive); b is when a fecal DNA sample is positive for a PCR marker of another target (false positive); c is when a fecal DNA sample is negative for a PCR marker of its own target (false negative); and d is when a fecal DNA sample is negative for a PCR marker of another target (true negative).

The positive predictive value or conditional probability that a particular source of fecal contamination was present when a water sample tested positive for the corresponding MST marker was calculated using Bayes' Theorem (Kildare et al., 2007).

The quantitative values of the qPCR markers were log_{10} converted to achieve normality. Despite of this, not all the variables were normally distributed, thus, Spearman correlation was used to detect relationships among variables. Data was analyzed using R, in the RStudio interface, version 0.97.314, R version 3.1.2 (2014-10-31)—"Pumpkin Helmet" (RStudio, 2015). Other primary packages, in addition to base packages, used in this analysis included *car, agricolae* and *ggplot2* (de Mendiburu, 2017; Fox and Weisberg, 2011; Wickham, 2016).

2.8. Ethics statement

Fully anonymized stool samples were obtained from Our Lady's Children's Hospital, Dublin following approval by the Ethics (Medical Research) Committee of this hospital. The ethics committee waived the requirement for informed consent. REC Reference GEN/169/70. Permission to obtain water samples was not required and field studies did not involve endangered or protected species. Fecal samples from Irish deer and stud farms were collected with permission of and under supervision by the owners.

3. Results

3.1. Sensitivity and specificity of host-specific markers in feces and sewage

It was shown previously that significant geographical variability exists in the sensitivity and specificity of MST markers (Gawler et al., 2007; Reischer et al., 2013; Yahya et al., 2017). The performance of these MST markers was therefore evaluated using feces from local animals and humans, prior to deploying these in the field. DNA was extracted from fresh fecal samples from mammals: cattle, deer, human, horse, pig, sheep, and from birds: gull and swan. Initially the general 16S rRNA gene was amplified, showing that most samples contained amplifiable DNA, except for DNA extracted from 6 out of 32 gull samples that could not be amplified and were discarded.

The performance of the MST markers was very good, achieving a sensitivity higher than 85% for most of the markers. In the case of the human marker HF183, a sensitivity of 73% was obtained when individual fecal samples were analyzed, however, the sensitivity of the human marker was 100% using raw and treated sewage samples. The specificities of all markers were 87% or higher (Table 2).

3.2. MST levels of host-specific markers in feces and sewage by qPCR

The levels of human (HF183), ruminant (CF128), cow (CowM3) and horse (HoF597) markers and total *Bacteroidales* (AllBac) were measured by qPCR in fecal samples from different sources (Fig. 2, Table 3). The number of positive samples, as determined by qPCR, was higher than by end-point PCR. The samples that were positive for qPCR but negative for endpoint PCR contained very low target DNA concentrations and in many cases were below the quantification limit.

The HF183 marker displayed considerable variability in individual human samples, ranging from 3.7 to 8.4 log_{10} gc/100 mg (Fig. 2A). As expected, there was less variability and higher average levels of human marker in raw and treated sewage. The concentration of the horse (HoF597) marker was an order of magnitude (1.3 to 2.5 log_{10}) higher

Table 2
Number of positive samples, sensitivity and specificity for different MST markers evaluated by end-point PCR in fecal samples.

| | Human marker | Ruminant marker | Sheep marker | Horse marker | Pig marker | | | Gull marker |
	HF183	CF128	Ovmito	HoF597	PF163	Pig-2-Bac	Pomito	Gull
Human	**19/26**	1/26	1/26	2/26	0/26	0/26	7/26	0/26
Cow[a]	2/22	**22/22**	0/22	1/22	0/22	0/22	0/22	0/22
Sheep[a]	5/25	**25/25**	**25/25**	1/25	4/25	0/25	0/25	0/25
Deer[a]	0/15	**15/15**	4/15	0/15	0/15	0/15	0/15	0/15
Horse	3/27	5/27	0/27	**26/27**	6/27	0/27	0/27	0/27
Pig	6/20	8/20	0/20	2/20	**20/20**	**15/20**	**20/20**	0/20
Gull	1/26	0/26	0/26	0/26	1/26	0/26	0/26	**22/26**
Swan	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8
% Sensitivity	**73**	**100**	**100**	**96**	**100**	**75**	**100**	**85**
% Specificity	**89**	**87**	**97**	**96**	**92**	**100**	**95**	**100**
Raw sewage[b]	4/4	0/4	2/4	0/4	0/3	0/3	3/3	0/3
Secondary treated effluent[b]	25/25	6/25	19/25	3/22	8/25	0/25	13/25	0/25

Values in bold indicate the targeted samples for the tested MST marker.
 [a] Ruminant species.
 [b] Raw sewage and the secondary treated effluent from two wastewaters treatments plants were not considered for sensitivity and specificity calculations since we cannot assure that human is the only input source.

in horse feces than in feces from other animal sources (Fig. 2B). The ruminant (CF128) marker levels were around 2.5–4.0 $\log_{10}$ higher in ruminant samples compared to feces from humans and other animal sources (Fig. 2C). In contrast to the human marker, the ruminant marker displayed low variability among individuals of the same group. The levels of the cow marker (CowM3) were 2.3–3.2 $\log_{10}$ higher in feces from cows and sheep than from other sources. Surprisingly, half of the sheep samples were positive for this marker. The levels of general *Bacteroidales* were high and similar for all the tested mammalian fecal samples. In contrast, the fecal samples from swans and gulls showed lower levels of the general *Bacteroidales* marker. The negative controls

for sample processing, nucleic DNA extraction, and PCR and qPCR reactions were in all cases negative.

To verify whether significant differences existed when comparing qualitative end-point PCR and the levels of host-marker levels among the corresponding target and non-target samples, as determined by qPCR, a Chi-squared and independent sample Kruskal-Wallis Test analysis was used. The differences within the target and non-targeted samples were statistically significant for all the markers evaluated in the study ($P < 0.05$). Thus, the MST markers tested are potentially useful to discern among different fecal samples in Ireland.
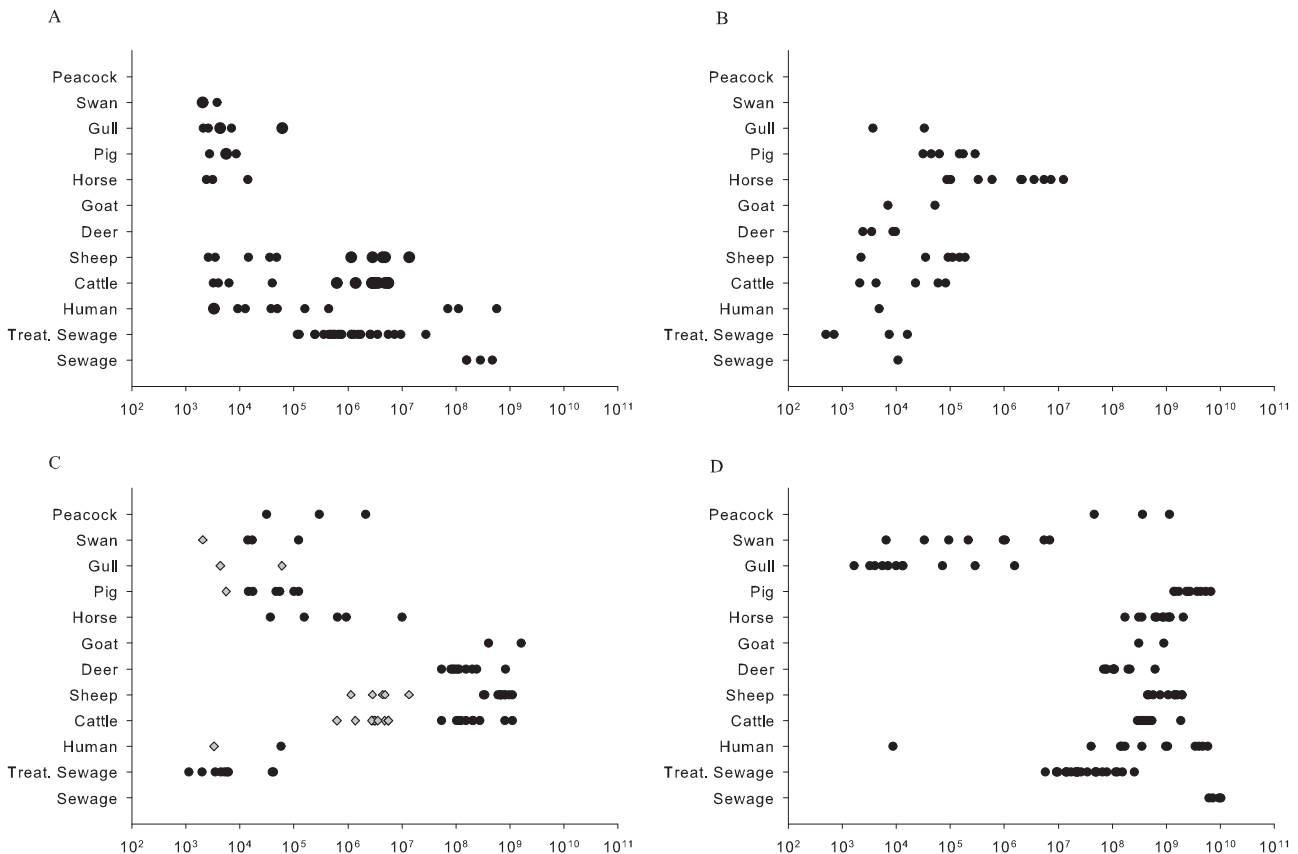
Fig. 2. Abundance of the different MST markers in fecal samples from different animal sources. (A) Human marker (HF183), (B) Horse Marker (HoF597), (C) Ruminant Marker (CF138) (in black) and Cow Marker (CowM3) (in grey) and (D) General *Bacteroidales* (AllBac). Values are expressed in gene copies 100 mg^{-1} or gc ml^{-1}.

Table 3

Positive samples, percentage of detection, median and 95% confidence interval (C·I.) of the $\log_{10}$ transformed levels of the MST markers (gc 100 mg^{-1} or gc 100 ml^{-1}) evaluated by qPCR in fecal and wastewater samples. Median values are calculated using positive samples only.

	Human marker qHF183				Ruminant marker qCF128				Cow marker CowM3
	No. positive samples/total samples	% of detection	Median	95% C·I.	No. positive samples/total samples	% of detection	Median	95% C·I.	No. positive samples/total samples
Human	**10/12**	**83%**	**4.95**	**3.72–8.43**	1/12	8%	4.76	n.a.	1/10
Cow[a]	4/10	40%	3.70	3.52–4.48	**10/10**	**100%**	**8.25**	**7.86–8.98**	10/10
Sheep[a]	5/10	50%	4.15	3.44–4.66	**10/10**	**100%**	**8.83**	**8.52–9.02**	5/10
Deer[a]	0/9	0%	n.a.	n.a.	**10/10**	**100%**	**8.04**	**7.81–8.68**	0/10
Horse	3/11	27%	3.5	3.39–4.08	5/10	50%	5.81	4.69–6.79	1/10
Pig	2/11	18%	3.68	3.46–3.91	6/11	55%	4.70	4.18–5.07	1/10
Gull	3/11	27%	3.41	3.33–3.80	0/11	0%	n.a.	n.a.	2/10
Swan	1/8	12%	3.58	n.a.	3/8	38%	4.23	4.15–5.00	1/8
Raw sewage	**4/4**	**100%**	**8.33**	**8.20–8.64**	n.a.	n.a.	n.a.	n.a.	n.a.
Secondary treated effluent	**25/25**	**100%**	**5.89**	**5.07–7.34**	8/25	32%	3.71	3.15–4.61	n.a

n.a. not applicable; all samples are negative.
Values in bold indicate the targeted samples for the tested MST marker.

3.3. Extreme variation in fecal contamination is related to land use

The Dargle catchment is characterized by a highly varied land use, including pristine, agricultural, forested and urban areas (Fig. 1), which was analyzed using Arc Hydro tools of ArcGIS using CORINE land cover data. This informed the localization of sampling stations within the catchment, with a view to capture the impact of land use on water quality. The catchment was sampled at 10 stations (Fig. 1). In addition to MST markers, the levels of fecal indicator bacteria (FIB), i.e. *E. coli* and Enterococci, were determined to evaluate the level of fecal pollution in the sub-catchments of the Dargle.

The highest levels of FIB (Figs. 3A, B, S1) in the Dargle catchment over a 3 year period were obtained from sampling stations surrounded by urban areas: The Swan (SWN) and Kilmacanogue (KMG2) rivers, and the Dargle River after the confluence of the outfall (CSO) of the Enniskerry WWTP (DRG3). The DRG3 station was located just 30 m downstream DRG2, however the secondary treatment effluent of the WWTP discharged between the two points explaining an increase of one order of magnitude in the *E. coli* and Enterococci levels (Fig. 3A, B). FIB levels were lower upstream in the catchment. Interestingly, there was a 10-fold increase in FIB levels in the Kilmacanogue River when comparing the sampling sites KMG1 and KMG2, which are separated by approximately 2 km. Lower levels were detected in County Brook (CBR), Killough River (KGH), Kilmacanogue River upstream of Kilmacanogue town (KMG1), and the Dargle before the Enniskerry WWTP outfall (DRG2). Pristine waters are found in the Glencullen (GCU) and Glencree (GCR) Rivers, and the upstream reaches of the Dargle River (DRG1).

3.4. Land use and the biological source of fecal water pollution

The Dargle catchment is very diverse in the levels of fecal contamination (Fig. 3, Table 4), which is to be expected based on its land use (Fig. 1). In order to analyze the biological sources of pollution, water samples were initially analyzed using end-point PCR. The presence/absence of the human (HF183), ruminant (CF128) and horse (HoF597) MST markers were evaluated by end-point PCR, since these target the main pollution pressures in the areas of interest (human, ruminant and equine). Positive end-point PCR samples were evaluated by qPCR to quantify the concentration of the markers.

The human marker was detected in all samples collected at the DRG3, KMG2 and SWN sampling stations (Fig. 3C). The DRG3 site is impacted by the effluent from the Enniskerry WWTP, which is located 30 m upstream of the sampling site. The KMG2 site is located in Kilmacanogue, downstream of Kilmacanogue town. The Swan River traverses several housing estates in Bray. The human marker was also detected with high frequency (>80% of positive samples) in DRG2, KGH

and GCU, but was virtually absent in samples taken at the DRG1 station (<10%), which is located in the pristine areas of the Dargle catchment (Fig. 3C, Table 4). The ruminant marker was less prevalent than the human marker in water samples. CF128 was detected in >60% of the sampling stations surrounded by agricultural areas: DRG1, DRG2, DRG3, KGH, CBR, GCU and GCR, but was virtually absent in the SWN (~20%), which flows through an almost exclusive urban area (Fig. 3F, Table 4). While the horse marker (HoF597) was rarely observed in general, its presence stood out at the KMG2 and CBR sampling stations (>20% of samples tested positive). There are several horse-riding stables located upstream from these sampling sites.

Bayesian statistics was used to determine the probability of detecting feces from the target-host within a given station using the end-point PCR results. Given the prevalence of the three MST markers in the catchment, as well as the sensitivity and specificity of these markers as determined using fecal samples (Table 2), the conditional probability of detecting the biological source of pollution was evaluated (Kildare et al., 2007). The conditional probability of having human pollution when the HF183 marker is detected was 0.95. The conditional probabilities for the ruminant marker (CF128) and horse marker (HoF597) were 0.93 and 0.75 respectively.

The levels of the human MST marker (qHF183) were high (10^5–10^6 gc 100 ml^{-1}) in sampling sites with anthropogenic impact: SWN and KMG2, DRG3 and the CSO. Intermediate levels of the human marker, between 10^3 and 10^4 gc 100 ml^{-1}, were detected in sampling sites with medium anthropogenic impact: KGH, CBR, GCU, KMG1 and DRG2. The levels of the human marker were below or at the detection limit at DRG1 and GCR (Fig. 3E and Table 4).

There was a fairly even distribution of the levels of the ruminant marker along the different sampling points of the Dargle catchment (Figs. 3F, S1 and Table 4). However, in sampling points with an agricultural impact: KGH, KMG1, KMG2 and CBR, the levels of the ruminant marker soared sporadically, often related to rainfall or unknown events.

3.5. Correlation between MST marker and FIB levels

The level of general *Bacteroidales* marker (using AllBac marker) was assessed in order to evaluate the potential use of total *Bacteroidales* as a fecal indicator. The levels of AllBac were higher than 10^4 gc 100 ml^{-1} in all samples (Fig. 3D) and they correlated with the levels of *E. coli* (r_s: 0.797; n: 354) and Enterococci (r_s: 0.687; n: 354). With respect to individual sampling stations, AllBac correlated with the *E. coli* levels at all stations, expect for the upstream sampling sites: DRG1 and GCR, which displayed none or very low levels of *E. coli* and Enterococci. Only five of the sampling sites showed a significant correlation between the AllBac marker and Enterococcus levels (Table S2).

Cow marker CowM3			Horse Marker qHoF597				Bacteroidales marker AllBac			
% of detection	Median	95% C.I	No. positive samples/total samples	% of detection	Median	95% C.I	No. positive samples/total samples	% of detection	Median	95% C.I
10%	3.52	n.a.	1/11	9%	3.68	n.a.	12/12	100%	8.76	5.95–9.71
100%	6.50	5. 95–6.75	5/10	50%	4.36	3.39–4.89	10/10	100%	8.62	8.48–9.02
50%	6.64	6.14–7.04	6/10	60%	5.00	3.64–5.24	10/10	100%	9.09	8.66–9.29
0%	n.a.	n.a.	4/10	40%	3.74	3.41–3.97	10/10	100%	8.03	7.85–8.58
0%	n.a.	n.a.	10/10	100%	6.32	4.94–6.99	11/11	100%	8.92	8.36–9.19
10%	3.75	n.a.	6/11	55%	4.98	4.53–5.40	10/10	100%	9.42	9.14–9.78
20%	4.21	3.69–4.72	2/11	18%	4.04	n.a.	11/11	100%	4.00	3.36–5.83
12%	3.31	n.a.	0/8	0%	n.a.	n.a.	8/8	100%	5.65	4.06–6.80
n.a	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	4/4	100%	9.92	9.80–10
n.a.	n.a.	n.a.	4/20	20%	3.67	3.22–4.63	23/23	100%	7.35	6.97–8.17

The human marker (qHF183) levels were strongly correlated with those of *E. coli* and Enterococci and AllBac in KMG2 ($r_s = 0.666$; $r_s = 0.653$ and $r_s = 0.765$; $n = 42$), while there was significant correlation between qHF183 and *E. coli* and AllBac, but not Enterococci ($r_s = 0.549$, $r_s = 0.734$; $n = 37$) in DRG3 (Table S2). Although high levels of the human marker were detected in SWN, moderate correlation coefficients were observed between this marker and *E. coli* and AllBac ($r_s = 0.423$, $r_s = 0.525$; $n = 42$) suggesting that other factors determine the fecal pollution pattern at this site. Although low levels of qHF183 were observed in KMG1 there is a moderate relationship with *E. coli*, Enterococci and AllBac ($r_s = 0.569$; $r_s = 0.547$ and $r_s = 0.430$; $n = 24$) (Table S2).

The ruminant marker was moderately correlated with *E. coli*, Enterococci and AllBac in DRG2 ($r_s = 0.600$; $r_s = 0.469$, and $r_s = 0.393$, $n = 40$) and KGH ($r_s = 0.552$; $r_s = 0.493$ and $r_s = 0.356$; $n = 42$). In CBR, SWN and GCR a correlation with just AllBac was detected ($r_s = 0.415$, $n = 42$; $r_s = 0.503$, $n = 42$; $r_s = 0.552$, $n = 18$) (Table S2).

3.6. Seasonality of MST markers and fecal contamination

To assess whether water quality was subject to seasonal variations, the dataset was grouped according to seasons: spring (March to May), summer (June to August), autumn (September to November) and winter (December to February) (Nnane et al., 2011). Using One Way Anova analysis, seasonal influence was detected for *E. coli* ($P < 0.001$), Enterococci ($P < 0.001$) and general *Bacteroidales* ($P = 0.044$), with the highest levels occurring during summer and autumn. The presence of the human marker (qHF183) was not influenced by the season during which the samples were taken ($P = 0.719$). In contrast, in general there was a seasonal effect on the levels of the ruminant marker (qCF128; $P < 0.001$). Most samples testing positive for the ruminant marker in the upstream reaches of the Dargle (DRG1), and in the Kilmacanogue and Killough rivers, were taken in summer and autumn, with only a small percentage testing positive in winter (Table S3 and Fig. S2).

3.7. The relationship between turbidity, FIB and MST markers

In general, there was a significant correlation between the turbidity of water samples and rainfall occurring on the day of sampling $r_s = 0.410$, the day before sampling $r_s = 0.418$ and both days $r_s = 0.492$), with a significant correlation at each sampling station (Table S4). With exception of CBR, all river sampling sites showed a significant correlation between rainfall and the levels of either *E. coli*, Enterococci, or both (Table S5). For the MST markers a positive correlation between the rainfall levels was observed with qCF128 in DRG2 and CBR. In contrast, a negative correlation of qHF183 with rainfall was observed in DRG3 and SWN.

A strong correlation between turbidity and FIB and MST was not observed. *E. coli* displayed a positive relation with turbidity in KMG1, qCF128 in DRG2 and CBR, whereas a negative correlation was observed with qHF183 in DRG1, DRG3 and SWN (Table S5).

4. Discussion

Several microbial source tracking markers have been tested successfully around the world (Ahmed et al., 2008a; Boehm et al., 2013; Gawler et al., 2007; Heaney et al., 2015; Peed et al., 2011; Santiago-Rodriguez et al., 2012; Yahya et al., 2017). However selecting suitable microbial source tracking markers may be difficult since considerable variability may exist in the composition of the microbiome of individuals due to age, genetics, diet, health status or geographical location (Biagi et al., 2011; Shanks et al., 2011; Yatsunenko et al., 2012). In this study we tested eight MST markers using end-point PCR and four markers using qPCR, with a view to determine their usefulness in their application for an effective management of water quality. The selected MST markers have been described elsewhere (Bernhard and Field, 2000; Dick et al., 2005; Lu et al., 2008; Martellini et al., 2005; Mieszkin et al., 2009; Seurinck et al., 2005).

The variability of genetic marker abundance and prevalence in populations from other geographical locations suggests that the use of MST markers developed in a geographical area requires a priori characterization of the assay performance at each watershed of interest before being implemented (Ahmed et al., 2008b; Shanks et al., 2010b, 2010a; Yahya et al., 2017). Thus, in the first instance the selected MST markers were tested in fecal samples from known sources. All the PCR markers showed a high sensitivity and specificity, although none of them achieved 100% for both parameters. Although some of the MST markers were detected in hosts other than the intended ones, their abundance in the target group was always significantly higher than in the non-target hosts, demonstrating their suitability to distinguish between sources of pollution. These MST markers are diluted once the fecal matter is introduced in the environment. The lower levels reported in non-targeted samples are therefore likely to be below the detection limit once the fecal matter has entered the watershed.

The levels of the human marker (HF183) variate in individual human stool samples. However, when the marker was tested in raw and treated sewage samples, 100% sensitivity and fairly constant marker levels were observed, which is in agreement with earlier observations (Seurinck et al., 2005). These authors report levels of 10^5 to 10^9 gc gr^{-1} of wet human feces and 10^9 to 10^{10} gc L^{-1} of sewage, similarly to the results of this study. The human gut harbors three predominant enterotypes dominated either by *Bacteroides*, *Prevotella* or *Ruminococcus* which could be related to for example diet (Arumugam et al., 2011; Wu et al., 2011; Xu and Knight, 2015), which might explain the high degree of variability in marker levels among individuals. However, the main
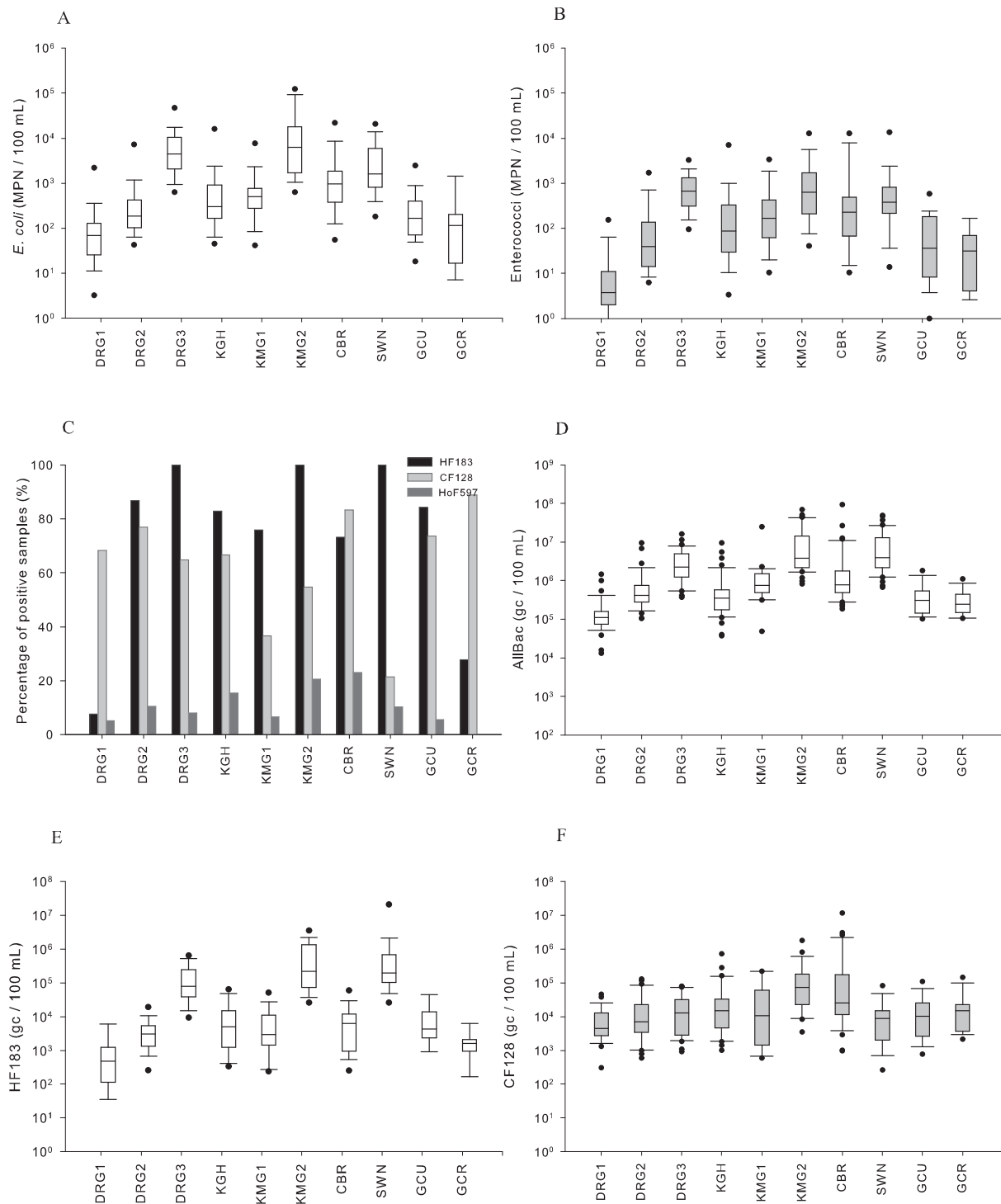
Fig. 3. Levels of the fecal indicators and MST markers along the Dargle catchment. (A) E. coli (MPN 100 ml^{-1}), (B) Enterococci (MPN 100 ml^{-1}), (C) Percentage of positive samples detected by PCR for the HF183, CF128 and HoF597 markers, (D) levels of AllBac, (E) HF183 and (F) CF128 by qPCR (gc 100 ml^{-1}). Box of the box plots represents the 25th, 50th and 75th percentile, the whiskers the 10% and 90th percentile and the dots the 5th and 95th percentile. Locations and abbreviations of the sampling stations as in Fig. 1.

sources of human pollution are sewer misconnection, effluents from WWTP or CSO, as well as septic tanks, rather than from individuals. The two human markers tested in this work: HF183 using end-point PCR and with SYBRGreen for qPCR were classified as the most sensitive and specific using binary analysis in a multi laboratory method evaluation study (Raith et al., 2013).

The effect of diet on the composition of the microbiota has also been reported for cattle, suggesting a strong correlation between bacterial microbiota structures and feeding practices. *Prevotella* spp. were the

dominant group when animals were fed with unprocessed grain, while Ruminococcacea were the main group when other feeds were used (Shanks et al., 2011). In this study we tested the ruminant marker CF128 by end-point PCR (Bernhard and Field, 2000) and since it showed a good specificity, a SYBRGreen assay with this primer was developed. The probe-labeled assay targeting the cow specific marker CowM3 (Shanks et al., 2008) was used to be able to discern among cattle and other ruminants (especially sheep and deer). The CF128 marker was detected in cows, sheep and deer by PCR and qPCR and showed a cross

Table 4
Presence and abundance of the FIB and MST markers in freshwater samples.
Median and 95% confidence interval (C.I.) of the $\log_{10}$ transformed data from the MST markers evaluated by qPCR. Positive samples/Total samples (PS/TS), percentage of positive samples evaluated by PCR. Locations and abbreviations of the sampling stations as in Fig. 1.

	DRG1		DRG2		DRG3		KGH		KMG1	
	Median	95% C.I.	Median	95% C.I.	Median	95% C.I.	Median	95% C.I.	Median	95% C.I.
E. coli	1.84	0.64–2.71	2.27	1.72–3.36	3.65	2.84–4.36	2.48	1.71–3.85	2.70	1.84–3.47
Ent	0.57	<1–1.94	1.59	0.80–3.16	2.83	2.05–3.40	1.94	0.73–3.55	2.22	1.18–3.29
qHF183	2.68	1.61–3.67	3.48	2.48–4.15	4.89	4.07–5.79	3.70	2.56–4.71	3.47	2.41–4.57
qCF128	3.65	3.16–4.52	3.84	2.96–5.01	4.11	3.16–4.89	4.18	3.20–5.36	4.02	2.84–5.33
AllBac	5.05	4.53–5.80	5.62	5.13–6.55	6.35	5.70–6.97	5.55	4.90–6.58	5.88	5.50–6.34
	PS/TS	% of positive	PS/TS	% of positive	PS/TS	% of positive	PS/TS	% of positive	PS/TS	% of positive
HF	3/40	7.5%	33/38	86.8%	36/36	100%	34/41	82.9%	22/29	75.9%
CF	28/41	68.3%	30/39	76.9%	24/37	64.9%	28/42	66.7%	11/30	36.7%
HoF	2/39	5.1%	4/38	10.5%	3/37	8.1%	6/39	15.4%	2/30	6.7%

	KMG2		CBR		SWN		GCU		GCR	
	Median	95% C.I.	Median	95% C.I.	Median	95% C.I.	Median	95% C.I.	Median	95% C.I.
E. coli	3.80	2.81–5.03	2.98	1.97–3.99	3.20	2.26–4.31	2.22	1.67–3.02	2.06	0.85–3.17
Ent	2.81	1.72–4.00	2.34	1.04–4.04	2.59	1.13–4.14	1.56	0.54–2.44	1.49	0.41–2.23
qHF183	5.35	4.44–6.35	3.81	2.64–4.62	5.29	4.42–7.32	3.63	2.96–4.66	3.20	2.38–3.67
qCF128	4.87	3.93–5.88	4.41	3.37–6.44	3.94	2.42–4.92	4.01	3.13–4.79	4.18	3.48–4.98
AllBac	6.59	6.07–7.66	5.89	5.38–7.13	6.60	5.88–7.66	5.49	5.06–6.14	5.38	5.03–5.93
	PS/TS	% of positive	PS/TS	% of positive	PS/TS	% of positive	PS/TS	% of positive	PS/TS	% of positive
HF	41/41	100%	30/41	73.2%	41/41	100%	16/19	84.2%	5/18	27.8%
CF	23/42	54.8%	35/42	83.3%	9/42	21.4%	14/19	73.7%	16/18	88.9%
HoF	8/39	20.5%	9/39	23.1%	3/39	10.3%	1/18	0%	0/18	0%

reaction mainly with horse and pig. However, it was four orders of magnitude more abundant in ruminant than non-ruminant samples. The concentration of CowM3 marker was around two orders of magnitude lower than CF128. This marker targets the gene encoding sialic acid-specific 9-O-acetylesterase secretory protein of some species of *Bacteroides* spp. (Shanks et al., 2010b). Whereas a single copy of this gene is present in the genome, up to 7 copies of the 16S rRNA gene per genome have been reported in several *Bacteroidales* species (Lee et al., 2009); the latter is therefore easier to detect following dilution of fecal matter in the water column. The CowM3 and CF128 markers target different genes with different copy numbers and in most likely different bacterial species. In contrast to previous reports, 50% of the sheep samples in the current study tested positive for the CowM3 marker, with similar levels of the marker in cow and sheep feces. In contrast, all deer fecal samples tested negative for this marker. Dorai-Raj et al. (2012) showed that bovine and ovine microbiotas are very similar, preventing the development of a bacterial MST marker to differentiate between these species. Sheep and cows are kept in intensive farming in many countries, whereas in Ireland these animals are mainly kept on pastures, which may account for the similarities in microbiota and poor specificity of the CowM3 marker. This problem may be solved using markers targeting mitochondrial DNA (Caldwell et al., 2007; Martellini et al., 2005; Schill and Mathes, 2008).

In this study, markers targeting mitochondrial DNA from pig and sheep showed a high specificity and sensitivity. Just a few positives were observed in human fecal samples, raw and treated sewage, which might be due to a recent ingestion of pork and sheep products. Other studies analyzing mitochondrial DNA detected the bovine marker in individuals who had eaten beef 24 h before the sample collection; however, the levels were two orders of magnitude lower than those of the human marker (Caldwell et al., 2007). In this study we did not perform a quantitative analysis for these markers which might have shown much lower concentration of pig or sheep mitochondrial DNA in human feces. Therefore, mitochondrial DNA may prove to be a good option to be used with a combination of bacterial markers to help to discriminate among animals with similar microbiota such as cow, sheep and deer (Casanovas-Massana et al., 2015; Schill and Mathes, 2008).

Two *Bacteroidales* pig markers were also evaluated in this work, whereas PF163 showed a higher sensitivity than Pig-2-Bac, a higher specificity was achieved by the latter, suggesting the use of a

combination of markers to increase source tracking resolution (Ballesté et al., 2010). Finally, the gull marker showed a very high specificity, and was not detected in other avian samples like swans.

The HF183 and CF128 markers were evaluated in several European countries, including Ireland (Gawler et al., 2007), which showed 88% sensitivity and 100% specificity for HF183 and 100% sensitivity and 96% specificity for CF128, which is slightly higher than what was observed in this study. Both markers were also evaluated in Galway (West of Ireland) what allows us to compare regional geographic variability (Dorai-Raj et al., 2009). For HF183 a sensitivity of 12% for individual fecal samples and 70% for sewage was reported with 100% of specificity, whereas 94% sensitivity and 95% specificity was observed for CF128. In the latter cross-reaction was observed with mainly pig fecal samples, as is the case in this study. These markers therefore perform in a consistent manner in different areas, and at different times.

The levels of the host specific *Bacteroidales* markers (qCF128) in the target host were several orders of magnitude higher, ranging from 10^7 to 10^8 gc 100 mg wet wt^{-1}, than in other hosts, like what was observed previously using samples taken in California (Silkie and Nelson, 2009). The very high levels of these markers in their target hosts mean that remain above the detection limit following release in a waterbody.

The next step was to evaluate the performance of the selected MST markers in a relatively small catchment with well-defined land uses. The analysis of land uses in the watershed, the visual inspection of the area, together with the microbial source tracking analysis allowed a very accurate assessment of the main hazards and sources of pollution in the catchment. The Dargle catchment has a population of 58,745 (census of 2006). This study showed that human fecal pollution is the main impact in the catchment. It has been observed that human pollution is mainly coming from the effluent of a WWTP (between DRG2 and DRG3), a non-detected sewer misconnection/s in the Kilmacongue River (KMG1 and KMG2) and the Swan River crossing the biggest urban area in the catchment showing even higher levels than those observed in treated sewage. Human fecal pollution was also prevalent with lower levels in agricultural areas probably due to leaks of septic tanks from the farms and houses spread around. A study combined land use information and small stream sampling reported a positive significant correlation between abundance of human associated markers and septic systems following a wet weather event (Peed et al., 2011). Unfortunately, in this study we did not have information about septic

systems and sanitary sewer lines locations which may strongly support the success of the MST strategies. The levels of the human associated maker in this study showed a negative significant correlation in DRG3 and SWN with rainfall suggesting a dilution of the pollution after rainfall events. There were no differences among seasons between the levels of human marker, indicating human pollution as a constant pressure along the year.

The presence of the CF128 marker was higher in areas surrounded by agriculture and the levels were fairly even along the catchment being less prevalent in urban areas (Swan River) showing a basal ruminant pollution in the river. Data from the number of animals was obtained from the Census of Agriculture 2000 from the Central Statistics Office Ireland and reported around 3280 cattle and 24,115 sheep heads in the catchment. Unfortunately, this data is reported for electoral districts, instead of the catchment's areas used in this study, therefore animal density could not be used to match the land uses with GIS. However, these numbers together with wild deer in the area might explain this basal fecal contamination. Significant correlation between the levels of the ruminant associated marker with *E. coli*, Enterococci in DRG2 and KGH indicates ruminant pollution as the main pressure in these points. The high levels of ruminant associated marker in CBR and its relationship with AllBac, but not with FIB, suggests an additional source of pollution in this point such as bird pollution holding reduced numbers of *Bacteroidales* such it has been observed in gulls and swans. The ruminant marker showed seasonal patterns being more abundant in summer and autumn and a positive correlation with maximum temperature in DRG1, DRG3, KGH and GCU, four sampling sites surrounded by forest and agriculture. Since lower persistence of *Bacteroidales* markers have been reported with higher temperatures (Schulz and Childers, 2011), the higher levels of the ruminant marker may be related to a higher contribution in warm seasons. Additionally, in two points (DRG2 and CBR) surrounded by agriculture the ruminant marker correlated with rainfall and turbidity, suggesting an important input coming from run-off. Identifying these inputs to the system can improve catchment management establishing agricultural best management practices (BMP) like cattle exclusion fencing or controlled tile drained management (Wilkes et al., 2014, 2013).

At most sites, the *Bacteroidales* marker showed a high correlation with *E. coli* and Enterococci, while a low correlation was observed high in upstream waters. Here, none or very low concentrations of *E. coli* but high levels of the *Bacteroidales* marker was detected. This suggests that the marker may be detecting non-intestinal environmental bacteria, in accordance with other studies detecting *Bacteroidales* markers in pristine and drinking waters (van der Wielen and Medema, 2010; Vierheilig et al., 2012). Its use as fecal indicator may depend of the sampling site and the levels of pollution (Mulugeta et al., 2012). *E. coli* and Enterococci in general showed a high seasonal pattern and high correlation with maximum temperature detecting higher levels in warmer periods. A shorter persistence of these markers with higher temperature has also extensively described (Ballesté et al., 2018; Blaustein et al., 2013; Jamieson et al., 2004) suggesting that the high levels are mainly related to major pollution inputs during such periods. Moreover, there was a positive correlation with FIB and rainfall, except for DRG3 where a negative correlation was found due to the dilution of the source of the pollution (the WWTP effluent) and CBR where no correlation was found. In the case of the MST markers analyzed in this study, such a correlation was only observed at a few sites, suggesting a potential of other fecal input in the catchment with rainfall events. These results confirm that BMP of run-off could improve water quality.

The MST toolbox evaluated in this study shows a good potential to be used for a better management and assessment of fecal pollution. However, it is advisable to evaluate the markers on fecal samples collected in the area where they will be applied in order to know their specificity, sensitivity, prevalence and understand their patterns in the environmental samples (Ballesté et al., 2020). The combined use of FIB, MST markers, environmental data and Geographical Information System to integrate land uses analysis with visual exploration, achieves appreciably enhanced description of the potential main causes of fecal pollution in a river catchment area. This data may be applied to develop hydrological models integrating bacterial data to facilitate the application of measures to eliminate or reduce the levels of fecal pollution at the source.

CRediT authorship contribution statement

Elisenda Ballesté: Investigation, Conceptualization, Formal analysis, Writing - original draft, Methodology. **Katalin Demeter:** Investigation. **Bartholomew Masterson:** Writing - review & editing. **Natàlia Timoneda:** Formal analysis. **Laura Sala-Comorera:** Investigation. **Wim G. Meijer:** Funding acquisition, Project administration, Supervision, Writing - original draft, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.139573.

References

Ahmed, W., Powell, D., Goonetilleke, A., Gardner, T., 2008a. Detection and source identification of faecal pollution in non-sewered catchment by means of host-specific molecular markers. Water Sci.Technol. 58, 579–586.

Ahmed, W., Stewart, J., Powell, D., Gardner, T., 2008b. Evaluation of Bacteroides markers for the detection of human faecal pollution. Lett.Appl.Microbiol. 46, 237–242.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., MetaHIT Consortium, Antolín, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Mérieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S.D., Bork, P., 2011. Enterotypes of the human gut microbiome. Nature 473, 174–180. https://doi.org/10.1038/nature09944.

Ballesté, E., Bonjoch, X., Belanche, L.A.L.A., Blanch, A.R., 2010. Molecular indicators used in the development of predictive models for microbial source tracking. Appl. Environ. Microbiol. 76, 1789–1795. https://doi.org/10.1128/AEM.02350-09.

Ballesté, E., García-Aljaro, C., Blanch, A.R., 2018. Assessment of the decay rates of microbial source tracking molecular markers and faecal indicator bacteria from different sources. J. Appl. Microbiol. 125, 1938–1949. https://doi.org/10.1111/jam.14058.

Ballesté, E., Belanche-Muñoz, L.A., Farnleitner, A.H., Linke, R., Sommer, R., Santos, R., Monteiro, S., Maunula, L., Oristo, S., Tiehm, A.A., Stange, C., Blanch, A.R., 2020. Improving the identification of the source of faecal pollution in water using a modelling approach: from multi-source to aged and diluted samples. Water Res. 171, 115392. https://doi.org/10.1016/j.watres.2019.115392.

Bernhard, A.E., Field, K.G., 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. Appl.Environ.Microbiol. 66, 4571–4574.

Biagi, E., Candela, M., Franceschi, C., Brigidi, P., 2011. The aging gut microbiota: new per-spectives. Ageing Res. Rev. 10, 428–429. https://doi.org/10.1016/j.arr.2011.03.004.

Blanchard, P., Regnault, J., Schurr, F., Dubois, E., Ribière, M., 2012. Intra-laboratory valida-tion of chronic bee paralysis virus quantitation using an accredited standardised real-time quantitative RT-PCR method. J. Virol. Methods 180, 26–31. https://doi.org/10.1016/j.jviromet.2011.12.005.

Blaustein, R.a., Pachepsky, Y., Hill, R.L., Shelton, D.R., Whelan, G., 2013. Escherichia coli sur-vival in waters: temperature dependence. Water Res. 47, 569–578. https://doi.org/10.1016/j.watres.2012.10.027.

Boehm, A.B., Van De Werfhorst, L.C., Griffith, J.F., Holden, P.a., Jay, J.a., Shanks, O.C., Wang, D., Weisberg, S.B., 2013. Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. Water Res. 47, 6812–6828. https://doi.org/10.1016/j.watres.2012.12.046.

Caldwell, J.M., Levine, J.F., 2009. Domestic wastewater influent profiling using mitochon-drial real-time PCR for source tracking animal contamination. J. Microbiol. Methods 77, 17–22. https://doi.org/10.1016/j.mimet.2008.11.007.

Caldwell, J.M., Raley, M.E., Levine, J.F., 2007. Mitochondrial multiplex real-time PCR as a source tracking method in fecal-contaminated effluents. Environ. Sci. Technol. 41, 3277–3283.

Casanovas-Massana, A., Gomez-Donate, M., Sanchez, D., Belanche-Munoz, L.A., Muniesa, M., Blanch, A.R., 2015. Predicting fecal sources in waters with diverse pollution loads using general and molecular host-specific indicators and applying machine learning methods. J Env. Manag. 151, 317–325. https://doi.org/10.1016/j.jenvman.2015.01.002.

Dick, L.K., Bernhard, A.E., Brodeur, T.J., Santo Domingo, J.W., Simpson, J.M., Walters, S.P., Field, K.G., 2005. Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. Appl.Environ.Microbiol. 71, 3184–3191.

Dorai-Raj, S., O', G.J., Colleran, E., 2009. Specificity and sensitivity evaluation of novel and existing Bacteroidales and Bifidobacteria-specific PCR assays on feces and sewage samples and their application for microbial source tracking in Ireland. Water Res. 43, 4980–4988.

Dorai-Raj, S., O'Grady, J., Cormican, M., Colleran, E., 2012. Identification of host-specific Bacteroidales 16S rDNA sequences from human sewage and ruminant feces. J. Basic Microbiol. 52, 277–284. https://doi.org/10.1002/jobm.201100184.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., Relman, D.A., 2005. Diversity of the human intestinal microbial flora. Sci-ence 308, 1635–1638. https://doi.org/10.1126/science.1110591.

European Comission, 2006. Directive 2006/7/EC of the European Parliament and of the council of 15 February 2006 concerning the management of bathing water quality and repealing directive 76/160/EEC. Off. J. Eur. Union 37–51 (doi:L 102/15).

Fox, J., Weisberg, S., 2011. An {R} Companion to Applied Regression.

García-Aljaro, C., Ballesté, E., Muniesa, M., Jofre, J., 2017. Determination of crAssphage in water samples and applicability for tracking human faecal pollution. Microb. Biotechnol. 10, 1775–1780. https://doi.org/10.1111/1751-7915.12841.

Gawler, A.H., Beecher, J.E., Brandao, J., Carroll, N.M., Falcao, L., Gourmelon, M., Masterson, B., Nunes, B., Porter, J., Rince, A., Rodrigues, R., Thorp, M., Walters, J.M., Meijer, W.G., 2007. Validation of host-specific Bacteroidales 16S rRNA genes as markers to deter-mine the origin of faecal pollution in Atlantic Rim countries of the European Union. Water Res. 41, 3780–3784.

Gourmelon, M., Caprais, M.P., Segura, R., Le, M.C., Lozach, S., Piriou, J.Y., Rince, A., 2007. Evaluation of two library-independent microbial source tracking methods to identify sources of fecal contamination in French estuaries. Appl.Environ.Microbiol. 73, 4857–4866.

Heaney, C.D., Myers, K., Wing, S., Hall, D., Baron, D., Stewart, J.R., 2015. Source tracking swine fecal waste in surface water proximal to swine concentrated animal feeding operations. Sci. Total Environ. 511, 676–683. https://doi.org/10.1016/j.scitotenv.2014.12.062.

Jamieson, R., Gordon, R., Joy, D., Lee, H., 2004. Assessing microbial pollution of rural sur-face waters: a review of current watershed scale modeling approaches. Agric. Water Manag. 70, 1–17. https://doi.org/10.1016/j.agwat.2004.05.006.

Kay, D., Fleisher, J.M., Salmon, R.L., Jones, F., Wyer, M.D., Godfree, A.F., Zelenauch-Jacquotte, Z., Shore, R., 1994. Predicting likelihood of gastroenteritis from sea bathing: results from randomised exposure. Lancet (London, England) 344, 905–909.

Kildare, B.J., Leutenegger, C.M., McSwain, B.S., Bambic, D.G., Rajal, V.B., Wuertz, S., 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. Water Res. 41, 3701–3715.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R., Sayler, G., 2006. Development of Bacteroides 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl.Environ.Microbiol. 72, 4214–4224.

Lee, Z.M.-P., Bussema, C., Schmidt, T.M., 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. 37, D489–D493. https://doi.org/10.1093/nar/gkn689.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008. Evolution of mammals and their gut microbes. Science (80-.). 320, 1647–1651.

Lu, J., Santo Domingo, J.W., Lamendella, R., Edge, T., Hill, S., 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. Appl.Environ.Microbiol. 74, 3969–3976.

Marchesi, J.R., Sato, T., Weightman, A.J., Martin, T.A., Fry, J.C., Hiom, S.J., Dymock, D., Wade, W.G., 1998. Design and evaluation of useful bacterium-specific PCR primers that am-plify genes coding for bacterial 16S rRNA. Appl.Environ.Microbiol. 64, 795–799.

Martellini, A., Payment, P., Villemur, R., 2005. Use of eukaryotic mitochondrial DNA to dif-ferentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. Water Res. 39, 541–548. https://doi.org/10.1016/j.watres.2004.11.012.

Mayer, R.E., Reischer, G.H., Ixenmaier, S.K., Derx, J., Blaschke, A.P., Ebdon, J.E., Linke, R., Egle, L., Ahmed, W., Blanch, A.R., Byamukama, D., Savill, M., Mushi, D., Cristóbal, H.A., Edge, T.A., Schade, M.A., Aslan, A., Brooks, Y.M., Sommer, R., Masago, Y., Sato, M.I., Taylor, H.D., Rose, J.B., Wuertz, S., Shanks, O.C., Piringer, H., Mach, R.L., Savio, D., Zessner, M., Farnleitner, A.H., 2018. Global distribution of human-associated fecal genetic markers in reference samples from six continents. Environ. Sci. Technol. 52, 5076–5084. https://doi.org/10.1021/acs.est.7b04438.

de Mendiburu, F., 2017. Agricolae: Statistical Procedures for Agricultural Research.

Mieszkin, S., Furet, J.P., Corthier, G., Gourmelon, M., 2009. Estimation of pig fecal contam-ination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. Appl.Environ.Microbiol. 75, 3045–3054.

Mulugeta, S., Hindman, R., Olszewski, A.M., Hoover, K., Greene, K., Lieberman, M., Mauro, S.A., 2012. Contamination level and location of recreational freshwater influence the ability to predict Escherichia coli concentration by qPCR targeting Bacteroides. J Env. Manag. 103, 95–101. https://doi.org/10.1016/j.jenvman.2012.03.012.

Nnane, D.E., Ebdon, J.E., Taylor, H.D., 2011. Integrated analysis of water quality parameters for cost-effective faecal pollution management in river catchments. Water Res. 45, 2235–2246. https://doi.org/10.1016/j.watres.2011.01.018.

Osborn, A.M., Moore, E.R., Timmis, K.N., 2000. An evaluation of terminal-restriction frag-ment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ.Microbiol. 2, 39–50.

Pascual-Benito, M., Nadal-Sala, D., Tobella, M., Ballesté, E., García-Aljaro, C., Sabaté, S., Sabater, F., Martí, E., Gracia, C.A., Blanch, A.R., Lucena, F., 2020. Modelling the seasonal impacts of a wastewater treatment plant on water quality in a Mediterranean stream using microbial indicators. J. Environ. Manag. 261, 110220. https://doi.org/10.1016/j.jenvman.2020.110220.

Peed, L.A., Nietch, C.T., Kelty, C.A., Meckes, M., Mooney, T., Sivaganesan, M., Shanks, O.C., 2011. Combining land use information and small stream sampling with PCR-based methods for better characterization of diffuse sources of human fecal pollution. Env. Sci Technol 45, 5652–5659. https://doi.org/10.1021/es2003167.

Raith, M.R., Kelty, C.A., Griffith, J.F., Schriewer, A., Wuertz, S., Mieszkin, S., Gourmelon, M., Reischer, G.H., Farnleitner, A.H., Ervin, J.S., Holden, P.A., Ebentier, D.L., Jay, J.A., Wang, D., Boehm, A.B., Aw, T.G., Rose, J.B., Balleste, E., Meijer, W.G., Sivaganesan, M., Shanks, O.C., 2013. Comparison of PCR and quantitative real-time PCR methods for the char-acterization of ruminant and cattle fecal pollution sources. Water Res. 47, 6921–6928. https://doi.org/10.1016/j.watres.2013.03.061.

Reischer, G.H., Kasper, D.C., Steinborn, R., Mach, R.L., Farnleitner, A.H., 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. Appl.Environ.Microbiol. 72, 5610–5614.

Reischer, G.H., Ebdon, J.E., Bauer, J.M., Schuster, N., Ahmed, W., Åström, J., Blanch, A.R., Blöschl, G., Byamukama, D., Coakley, T., Ferguson, C., Goshu, G., Ko, G., de Roda Husman, A.M., Mushi, D., Poma, R., Pradhan, B., Rajal, V., Schade, M.A., Sommer, R., Taylor, H., Toth, E.M., Vrajmasu, V., Wuertz, S., Mach, R.L., Farnleitner, A.H., 2013. Per-formance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. Environ. Sci. Technol. 47, 8548–8556. https://doi.org/10.1021/es304367t.

Rogers, S.W., Donnelly, M., Peed, L., Kelty, C.A., Mondal, S., Zhong, Z., Shanks, O.C., 2011. Decay of bacterial pathogens, fecal indicators, and real-time quantitative PCR genetic markers in manure-amended soils. Appl Env. Microbiol 77, 4839–4848. https://doi.org/10.1128/AEM.02427-10.

Roslev, P., Bukh, A.S., 2011. State of the art molecular markers for fecal pollution source tracking in water. Appl.Microbiol.Biotechnol. 89, 1341–1355.

RStudio, 2015. RStudio: Integrated Development Environment for R.

Santiago-Rodriguez, T.M., Tremblay, R.L., Toledo-Hernandez, C., Gonzalez-Nieves, J.E., Ryu, H., Santo Domingo, J.W., Toranzos, G.A., 2012. Microbial quality of tropical inland wa-ters and effects of rainfall events. Appl Env. Microbiol 78, 5160–5169. https://doi.org/10.1128/AEM.07773-11.

Schill, W.B., Mathes, M.V., 2008. Real-time PCR detection and quantification of nine po-tential sources of fecal contamination by analysis of mitochondrial cytochrome b tar-gets. Environ. Sci. Technol. 42, 5229–5234. https://doi.org/10.1021/es800051z.

Schulz, C.J., Childers, G.W., 2011. Fecal bacteroidales diversity and decay in response to variations in temperature and salinity. Appl Env. Microbiol 77, 2563–2572. https://doi.org/10.1128/AEM.01473-10.

Seurinck, S., Defoirdt, T., Verstraete, W., Siciliano, S.D., 2005. Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. Environ.Microbiol. 7, 249–259.

Shanks, O.C., Atikovic, E., Blackwood, A.D., Lu, J., Noble, R.T., Domingo, J.S., Seifring, S., Sivaganesan, M., Haugland, R.A., 2008. Quantitative PCR for detection and enumera-tion of genetic markers of bovine fecal pollution. Appl.Environ.Microbiol. 74, 745–752.

Shanks, O.C., White, K., Kelty, C.A., Hayes, S., Sivaganesan, M., Jenkins, M., Varma, M., Haugland, R.A., 2010a. Performance assessment PCR-based assays targeting bacteroidales genetic markers of bovine fecal pollution. Appl Env. Microbiol 76, 1359–1366. https://doi.org/10.1128/AEM.02033-09.

Shanks, O.C., White, K., Kelty, C.A., Sivaganesan, M., Blannon, J., Meckes, M., Varma, M., Haugland, R.A., 2010b. Performance of PCR-based assays targeting Bacteroidales ge-netic markers of human fecal pollution in sewage and fecal samples. Env. Sci Technol 44, 6281–6288. https://doi.org/10.1021/es100311n.

Shanks, O.C., Kelty, C.A., Archibeque, S., Jenkins, M., Newton, R.J., McLellan, S.L., Huse, S.M., Sogin, M.L., 2011. Community structures of fecal bacteria in cattle from different an-imal feeding operations. Appl.Environ.Microbiol. 77, 2992–3001.

Silkie, S.S., Nelson, K.L., 2009. Concentrations of host-specific and generic fecal markers measured by quantitative PCR in raw sewage and fresh animal feces. Water Res. 43, 4860–4871.

Sowah, R.A., Habteselassie, M.Y., Radcliffe, D.E., Bauske, E., Risse, M., 2017. Isolating the impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia, United States. Water Res. 108, 330–338. https://doi.org/10.1016/j.watres.2016.11.007.

Unc, A., Goss, M.J., 2004. Transport of bacteria from manure and protection of water resources. Appl. Soil Ecol. 25, 1–18. https://doi.org/10.1016/j.apsoil.2003.08.007.

Verhougstraete, M.P., Martin, S.L., Kendall, A.D., Hyndman, D.W., Rose, J.B., 2015. Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale. Proc. Natl. Acad. Sci. 112, 10419–10424. https://doi.org/10.1073/pnas.1415836112.

Vierheilig, J., Farnleitner, A.H., Kollanur, D., Bloschl, G., Reischer, G.H., 2012. High abundance of genetic Bacteroidetes markers for total fecal pollution in pristine alpine soils suggests lack in specificity for feces. J. Microbiol. Methods 88, 433–435. https://doi.org/10.1016/j.mimet.2012.01.009.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

van der Wielen, P.W., Medema, G., 2010. Unsuitability of quantitative Bacteroidales 16S rRNA gene assays for discerning fecal contamination of drinking water. Appl Env. Microbiol 76, 4876–4881. https://doi.org/10.1128/AEM.03090-09.

Wilkes, G., Brassard, J., Edge, T.A., Gannon, V., Jokinen, C.C., Jones, T.H., Marti, R., Neumann, N.F., Ruecker, N.J., Sunohara, M., Topp, E., Lapen, D.R., 2013. Coherence among different microbial source tracking markers in a small agricultural stream with or without livestock exclusion practices. Appl. Environ. Microbiol. 79, 6207–6219. https://doi.org/10.1128/AEM.01626-13.

Wilkes, G., Brassard, J., Edge, T.A., Gannon, V., Gottschall, N., Jokinen, C.C., Jones, T.H., Khan, I.U.H., Marti, R., Sunohara, M.D., Topp, E., Lapen, D.R., 2014. Long-term monitoring of waterborne pathogens and microbial source tracking markers in paired agricultural watersheds under controlled and conventional tile drainage management. Appl. Environ. Microbiol. 80, 3708–3720. https://doi.org/10.1128/AEM.00254-14.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D., Lewis, J.D., 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science (80-.) 334, 105–108. https://doi.org/10.1126/science.1208344.

Xu, Z., Knight, R., 2015. Dietary effects on human gut microbiome diversity. Br. J. Nutr. 113 (S1), S5. https://doi.org/10.1017/S0007114514004127.

Yahya, M., Blanch, A.R., Meijer, W.G., Antoniou, K., Hmaied, F., Ballesté, E., 2017. Comparison of the performance of different microbial source tracking markers among European and North African regions. J. Environ. Qual. 46, 760. https://doi.org/10.2134/jeq2016.11.0432.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R., Gordon, J.I., 2012. Human gut microbiome viewed across age and geography. Nature 486, 222–227. https://doi.org/10.1038/nature11053.