Elucidation of fecal inputs into the River Tagus catchment (Portugal) using source-specific mitochondrial DNA, HAdV, and phage markers

S. Monteiro\textsuperscript{a,b,*}, J. Ebdon\textsuperscript{a}, R. Santos\textsuperscript{b}, H. Taylor\textsuperscript{a}

\textsuperscript{a} School of Environment and Technology, University of Brighton, Brighton, UK
\textsuperscript{b} Laboratorio Analises, Instituto Superior Tecnico, Lisbon, Portugal

Keywords: Water quality; remediation; fecal indicators; source tracking; mitochondrial markers; viruses

* Corresponding author at: Laboratorio Analises, Instituto Superior Tecnico, Avenida Rovisco Pais, 1049-001 Lisbon, Portugal. E-mail address: silvia.monteiro@tecnico.ulisboa.pt
ABSTRACT

Determining the source of fecal contamination in a water body is important for the application of appropriate remediation measures. However, it has been suggested in the extant literature that this can best be achieved using a ‘toolbox’ of molecular- and culture-based methods. In response, this study deployed three indicators (Escherichia coli (EC), intestinal enterococci (IE) and somatic coliphages (SC)), one culture-dependent human marker (Bacteroides (GB-124) bacteriophage) and five culture-independent markers (human adenovirus (HAdV), human (HMMit), cattle (CWMit), pig (PGMit) and poultry (PLMit) mitochondrial DNA markers (mtDNA)) within the River Tagus catchment (n = 105). Water samples were collected monthly over a 13-month sampling campaign at four sites (impacted by significant specific human and non-human inputs and influenced by differing degrees of marine and freshwater mixing) to determine the dominant fecal inputs and assess geographical, temporal, and meteorological (precipitation, UV, temperature) fluctuations. Our results revealed that all sampling sites were not only highly impacted by fecal contamination but that this contamination originated from human and from a range of agricultural animal sources. HMMit was present in a higher percentage (83%) and concentration (4.20 log GC/100 mL) than HAdV (32%, 2.23 log GC/100 mL) and GB-124 bacteriophage with the latter being detected once. Animal mtDNA markers were detected, with CWMit found in 73% of samples with mean concentration of 3.74 log GC/100 mL. Correlation was found between concentrations of fecal indicators (EC, IE and SC), CWMit and season. Levels of CWMit were found to be related to physico-chemical parameters, such as temperature and UV radiation, possibly as a result of the increasing presence of livestock outside in warmer months. This study provides the first evaluation of such a source-associated ‘toolbox’ for monitoring surface water in Portugal, and the
conclusions may inform future implementation of surveillance and remediation strategies for improving water quality.

1. Introduction

Whilst water is undoubtedly one of the most important resources on the planet, it is also a major vehicle for disease transmission, continuing to cause 1.7 billion cases of childhood diarrheal disease and 525,000 preventable deaths/year (mainly in infants under 5 years of age) (WHO, 2017).

In addition, according to the existing climate change scenario, by 2030, water scarcity in arid and semi-arid places will force the reallocation of 24 to 700 million people (World Water Assessment Programme, 2009; Hameeteman, 2013).

It is, therefore, crucial to ensure freshwater sources are kept free of fecal contamination, especially human sources as they are more likely to contain microorganisms adapted to infect human hosts. As such, water contaminated with human fecal pollution may present significant risks to human health as they may contain pathogens, such as enteric viruses (Noroviruses, Hepatitis A viruses, and Enteroviruses), enteric bacteria (Salmonella enterica, Shigella spp., Campylobacter spp.) and/or protozoa (Cryptosporidium spp. and Giardia lamblia) (Seidel et al., 2016; Haramoto et al., 2018; Holcomb et al., 2020). Nonetheless, non-human animal sources may also act as reservoirs for certain ‘zoonotic’ pathogens, including Escherichia coli O157 H7, Hepatitis E virus and Cryptosporidium (Cotruvo et al., 2004).
Traditional fecal indicator bacteria (FIB), e.g. *Escherichia coli* (EC) and intestinal enterococci (IE), have been used during the last century to monitor microbial water quality (Ashbolt *et al.*, 2001). However, it is well known that most of these microorganisms are not host-associated (i.e. not limited to humans) but also exist in the intestines of other warm-blooded animals (Scott *et al.*, 2002). Therefore, such organisms cannot be used to predict the source of fecal contamination and to accurately understand which pathogens may be present, or where they originate from. Epidemiological studies, when contamination is mainly from non-point sources, have also failed to correlate risks to human health and FIB (Dwight *et al.*, 2004; Colford *et al.*, 2007). This may, in part, be attributed to the fact that whilst EC and IE are able to inhabit different warm-blooded animals, many pathogens are limited to human hosts, especially enteric viruses (Harwood *et al.*, 1999). Additionally, several strains of FIB have been shown to be able to become saprophytes, persisting in certain habitats, including soils, aquatic sediments, and vegetation (Ishii *et al.*, 2006; Ksoll *et al.*, 2007; Badgley *et al.*, 2011). Therefore, the role of FIB in determining effective management of water quality can significantly improve by employing new methodologies that allow for the discrimination of the sources of these microorganisms. Source tracking (ST) methods have been shown to provide rapid and efficient fecal source discrimination in order to expedite cost-effective remediation actions (Ahmed *et al.*, 2019; Kongprajug *et al.*, 2019). Several ST methods have been tested previously including anaerobic bacteria (*Bifidobacterium* and *Bacteroides*), enteric viruses and bacteriophages and mitochondrial DNA (mtDNA). For instance, human adenovirus (HAdV), a member of the enteric viruses family, has been previously shown to have the potential to determine human sources of fecal contamination in different settings. HAdV is highly specific to human fecal contamination and is detected throughout the year at high
concentrations. On the other hand, bacteriophages infecting a specific host strain (such as *Bacteroides fragilis* GB-124) have been described as potential markers to evaluate fecal contamination from human origin (Payan *et al*., 2005; Ebdon *et al*., 2012; McMinn *et al*., 2014). Such bacteriophages have been detected throughout the world in relatively high concentrations in municipal wastewaters worldwide and their temporal and geographic stability has already been addressed in several studies (Ebdon *et al*., 2012; Jofre *et al*., 2014; McMinn *et al*., 2014). Mitochondrial DNA markers have been used to a less extent to determine the sources of fecal contamination although their potential as markers of fecal contamination in environmental waters is considerable. Feces contain substantial amounts of exfoliated epithelial cells and mtDNA is present in many copies per cell (Iyengar *et al*., 1991; Gerber *et al*., 2001; Andreasson *et al*., 2002). Such feature ensures a robust PCR signal in accordance with 16S rRNA genes. Targeting directly mtDNA implies identifying the animal species directly rather than microbial species they host. Nonetheless, further studies are necessary to understand the relevance of carry-over on the environmental presence of a certain mtDNA marker. Carry-over was determined by Caldwell *et al*. (2007) in half of the participants of the study who consumed beef, the concentration was 1-2 orders below than that of the human target, Martellini *et al*. (2005) were unable to detected animal signals from the feces of the human volunteer who had eaten meat products the previous day. Therefore, such results seem to indicate that the impact of carry-over in environmental waters where the dilution effect is an important factor may be negligible. Information on the environmental persistence of mtDNA markers is also sparse. The persistence of these markers in was determined previously but in a microcosmos setting, not evaluating directly in environmental waters where interactions with the
different physico-chemical and biological processes found in the environment are not evaluated (Martellini et al., 2005; He et al., 2016).

In this study, we describe a new ST ‘toolbox’ of methods comprising both culture- and molecular-dependent approaches and its application to identify sources within the R. Tagus catchment. The objectives of the study were: (i) to compare the performance (for the first time) of animal-associated mitochondrial DNA markers (cattle (CWMit), pig (PGMit) and poultry (PLMit) fecal contamination) and human-associated markers, including a mitochondrial marker (HMMit), HAdV and GB-124 bacteriophage; (ii) to evaluate the seasonality of each marker; (iii) to study, in a real catchment, the relationship of each MST marker to several environmental conditions, including water and environmental temperature, salinity, and UV radiation, to have a better understanding of the persistence of these markers in environmental freshwaters, particularly those impacted by marine waters, and (iv) to determine the influence of the occurrence of rain and precipitation levels on the presence of these markers in the R. Tagus catchment. This study was design to provide new knowledge on a set of less common MST markers and their potential usefulness in a “toolbox”-like approach for pollution surveillance and aiding in the development of strategies to enhance the water quality of the R. Tagus.

2. Materials and methods

2.1. Sample location

Water samples were collected at four sampling sites in the area of the R. Tagus catchment close to the city of Lisbon (population of 2.27 x 10^6 inhabitants) (Fig. 1; INE, 2019). All samples were collected between February 2015 and March 2016 (13 months) at approximately one-month intervals (n= 105). The four sampling points
selected were: site (1) Vila Franca de Xira (VFX), site (2) Marina Parque das Nações (MNdP), site (3) Alcântara (Alc), and site (4) Belém (Bel) (Fig. 1). Sampling site 1 was selected because of its location at the upper most extent of the tidal limit. Water quality at this sampling site in particular is not impacted by waters from the Atlantic Ocean during high tide, a situation quite distinct from the other sampling sites. Moreover, site 1 is a well-known location for the rearing of livestock (mainly cattle, but also pigs) and for bulls used for bull fighting. Sampling site 2 was chosen because of its proximity to the discharge of two significant wastewater treatment plants (WWTP 3 and 4, Fig. 1). WWTP 4 receives wastewater from livestock farms and discharges into one of the R. Tagus tributaries (Rio Trancão). It is important to understand whether the Rio Trancão increases the contamination in R. Tagus and if so, which sources are the most important to control. This area is also highly urbanized and used for recreational purposes such as dog walking. Sampling site 3 was chosen because of its proximity to the discharge point of the largest WWTP in Lisbon (WWTP 6 Pop. Equiv = 800,000). This WWTP receives and treats effluents from the majority of the Lisbon metropolitan area and is therefore a main source of human fecal contamination within the catchment. This sampling site is also highly influenced by oceanic mixing with water from the Atlantic Ocean during high tide. Finally, sampling site 4 was selected because of its popularity as a recreational site and the fact that it is heavily influenced by the Atlantic Ocean. Sampling was conducted between February 2015 to March 2016, in order to understand the influence of seasonality on the presence and distribution of source-associated markers.
Fig. 1. The R. Tagus (Rio Tejo) catchment, sampling sites and location of wastewater treatment plants (WWTPs) (Google, 2020). Balloon symbols represent the six WWTPs impacting R. Tagus and the squares represent the four sampling locations.

2.2. Sample collection

In this study, one human-targeted culture-dependent assay (GB-124) phage, two human (HMMit and HAdV) and three animal qPCR-based assays (cattle - CWMit; pig – PGMit; and poultry – PLMit) were tested (Ebdon et al., 2007; Schill and Mathes, 2008; Rusiñol et al., 2014). Moreover, indicators of fecal contamination, EC, IE and SC were also determined. At each site, 2 L of river water was collected in 2 x 1 L sterile polyethylene bottles (Vidrolab 2, Portugal) for fecal indicator bacteria and bacteriophage detection and enumeration. On each occasion another 2 L sample was collected for the concentration of mtDNA markers, along with a 10 L grab sample collected in a 10 L sterile polyethylene bottle (Vidrolab 2, Portugal) for the analysis of
HAdV. All samples were then transported to the laboratory facilities at (5 ± 3) °C within a period of 4 hours after collection and analyzed or treated immediately upon arrival to the laboratory.

Stool samples from animals were obtained from the Biblioteca/Bedeteca dos Olivais. One human volunteer at the laboratory provided stool samples. All fecal samples were frozen at -30 °C. Stool samples were used to obtain the genomic material for the construction of standard curves for the quantification of each mtDNA marker and for the creation of positive controls for the qPCR reactions.

2.3. Fecal indicator bacteria

EC and IE were enumerated in river water samples by the Most Probable Number (MPN) method using the Colilert and Enterolert systems (IDEXX Laboratories, USA), according to standard methods (ASTM, 1999; UK Environment Agency, 2009). Briefly, each sample, or dilution thereof was poured into a 100 mL sterile plastic bottle and mixed with sterile H2O up to a final volume of 100 mL. The appropriate substrate was added to the 100 mL bottle and the mixture was allowed to settle until dissolved. The contents were then poured into a Quanti-Tray® and sealed, then incubated at (37 ± 1) °C for 18-22 h and at (41.0 ± 0.5) °C for 18-22 h, for EC and IE, respectively. Positive and negative controls were run for all tests and samples were analyzed in duplicate. Following the incubation period, the trays were observed under UV light and wells showing fluorescence were counted.

2.4. Bacteriophages

Somatic coliphages (SC) were chosen for their potential as viral indicators of fecal pollution, whereas B. fragilis (strain GB-124) bacteriophages were tested as potential
human-associated markers, as they have shown to be limited to human sources in the UK and US (Payan et al., 2005; Ebdon et al., 2007; McMinn et al., 2014, 2017). The detection and enumeration of bacteriophages was performed in duplicate using the standardized double-layer methods specific for the detection of SC and GB-124 phages (ISO 2000, 2001).

Water samples were analyzed directly for SC (as they are more abundant than source-associated markers) whereas 2 L filtrations were performed for GB-124 phages (Mendez et al., 2004). For the determination of GB-124 phages, river samples were amended to a final concentration of 0.05M MgCl₂, filtered through a 47 mm 0.22 μm mixed cellulose esters membrane filter (Whatman, GE Healthcare, US), after which the membranes were cut into eight fragments and placed in a flask containing 5 mL of elution buffer (3% w/v beef extract, 3% v/v Tween 80, and 0.5M NaCl, final pH 9.0). The eluate was placed in an ultrasonic bath for 5 min. The membrane was removed and the eluate was filtered through a 0.22 μm polyvinylidene fluoride (PVDF) filter (PALL, UK) with low protein binding (Tartera et al., 1992). This constitutes the concentrated sample. One mL of the Bacteroides GB-124 inoculum culture was added to 2.5 mL of semi-solid agar medium (Bacteroides Phage Recovery Medium agar (BPRM); 1% agar). One mL of concentrated sample was added subsequently to the mixture. The latter was slowly vortexed in order to avoid the formation of air bubbles, poured over BPRM with 2% agar, and left to set. The plates were then inverted and placed in anaerobic jars containing an anaerobic sachet (Oxoid, UK).

SC were determined by adding 1 mL of EC inoculum culture to 2.5 mL of semi-solid agar medium (Modified Scholtan’s Agar (MSA)) and 1 mL of samples pre-filtered using a 0.22 μm PVDF filter, vortexed, and poured over a layer of solid MSA and left to set. For both phages the plates were incubated at (36 ± 2) °C for 18 (± 2) h.
2.5. Molecular-based assays

For the concentration of mtDNA, 2 L river water samples were centrifuged at 9000 xg for 15 min and the supernatant carefully removed until all the volume was centrifuged. The pellet was then resuspended in 10 mL of sample (Martellini et al., 2005).

To concentrate HAdV from 10 L river water, a method based on direct organic flocculation was used, with the adhesion of viral particles to pre-flocculated skimmed milk (Calgua et al., 2008). In brief, the pH of the samples was pre-regulated to pH 3.5 by adding 1 N HCl. Five mL of pre-flocculated 1 % (w/v) skimmed milk was then added to the sample, so that the final concentration of skimmed milk was 0.01 % (w/v). The samples were stirred at room temperature for 8 h. The material in suspension was allowed to settle for a further 8 h and the supernatant was carefully removed. The flocculated sample was then centrifuged at 12 ºC for 30 min at 7 000 xg. Following centrifugation, the supernatant was carefully removed and the sediment was resuspended in 8 mL of 0.2 M phosphate buffer (0.2 M of sodium monohydrogen phosphate (Na₂HPO₄) and a solution of sodium phosphate (NaH₂PO₄) at a proportion of 1:1; pH 7.5). The volume of the resuspended viral particles was finally adjusted to 10 mL. The eluates for mtDNA and HAdV analysis were kept at -80 ºC until analysis.

2.6. Nucleic acid extraction

DNA from the concentrated river water samples for analysis of HAdV was extracted using QIAamp Viral RNA Mini Kit (Qiagen, Germany), DNA from river samples for the detection of mtDNA was extracted using QIAamp DNA Blood Mini kit (Qiagen, Germany), and DNA to generate the positive controls and standard curves for mtDNA were extracted from stools using QIAamp Fast DNA Stool mini kit (Qiagen, Germany)
according to the manufacturer’s instructions. Final eluted volume was 200 µL for river
for mtDNA extraction (in the river and stool samples) and 80 µL for HAdV extraction.
The samples were kept at (-30 ± 5) ºC until further analysis, within 6 months.

2.7. Real-time PCR

Primers and probe sequences of the human, bovine, pig and poultry markers used are
shown in Table S1. The amplifications were performed in 25 µL of reaction mixture
using the TaqMan® Universal Master Mix II (Applied Biosystems, US). A volume of
12.5 µL of master mix was mixed with the respective concentration of primers and
probes for each assay (forward and reverse primers – 800 nM, probes – 200 nM). The
volume was adjusted to 20 µL with sterile DNA and RNA-free water. In addition, 10-
fold and 100-fold dilutions of every DNA extraction were also assayed. The PCR
reactions were carried out in a 7300 Real-Time PCR System (Applied Biosystems,
US) according to the original manuscript (Hernroth et al., 2002; Schill and Mathes,
2008). For HAdV, the first step occurs at 50 ºC for 2 min, followed by an activation
step at 95 ºC for 10 min, 40 cycles of denaturation at 95 ºC for 15 sec and annealing
at 60 ºC for 1 min. For all mtDNA assays, the first step consisted of an activation step
at 94 ºC for 3 min, followed by 40 cycles of denaturation at 94 ºC for 3 sec and
annealing at 60 ºC for the human, bovine and poultry assays and 57 ºC for pig assay.
Quantification was performed using specific standard curves built of seven points and
in duplicate for each MST marker. A standard curve was performed with each qPCR
run. Quantification for HAdV was constructed using the Amplirun® Adenovirus DNA
control (Vircell, Spain) with values ranging between 1.36 and 1.36 x 10⁴ gene copies
(GC)/μL. For the quantification of mtDNA markers, the standard curves were obtained
by amplification of each target using conventional PCR from extracted stool samples.
After amplification, the amplicons were run in a 4% electrophoresis agarose gel. Upon visualization, each band was excised from the gel and DNA was purified using the illus
tra GFX PCR DNA and Gel band purification kit according to the manufacturer’s instructions (GE Healthcare Life Sciences, US). Following purification of a DNA stock for each quantitative parameter, the concentration of each stock was quantified using the NanoDrop ND-100 spectrophotometer (ThermoScientifics, DE). Serial dilutions of each stock were prepared in DNA/RNA-free water in order to prepare the standard curve for qPCR. Standard curve for HMMit ranged between 5.87 and 5.87 x 10^6 GC/uL, for CWMit ranged from 5.5 and 5.5 x 10^6 GC/uL, for PGmit the standard curve varied between 5.71 and 5.71 x 10^6 GC/uL and for PLMit between 4.1 and 4.1 x 10^6 GC/uL. To ensure that no cross-contamination existed during the complete protocol several quality control steps were added. A process control accompanying the process of concentration, extraction and qPCR was added each time samples were analyzed. A negative extraction control followed the process of extraction. In addition, positive and a non-template controls were tested also with each qPCR run. All results were expressed in terms of GC/100 mL of sample.

2.8. Physico-chemical parameters

Different physico-chemical parameters may influence the behavior of microorganisms in the environment, including temperature (atmospheric and water), UV radiation, salinity, and rainfall (during 24 h previous to sampling). Data for physico-chemical parameters such as atmospheric temperature, rainfall, and UV radiation were obtained from the Instituto Português do Mar e da Atmosfera (IPMA, 2016). Salinity and water temperature data were obtained from the Marine, Environment and Technology Center (MARETEC, 2016).
2.9. Data analysis

All data analysis was performed with either Microsoft Excel 2016 or IBM SPSS Statistics 25 (IBM, NY). The limit of detection (LoD) was calculated by making serial dilutions of the standards used for each marker and determining the lowest quantity of the marker with a confidence level of 99%. All data were converted into a logarithmic format (log\(_{10}\)). Samples with negative results for each parameter were transformed and used as follows for the statistical analysis:

\[
\text{Method Limit of detection} = \frac{\text{LoD}}{2}
\]

To determine the influence of sampling location and season in the occurrence of each marker, the analysis of variances of two-way factors without replication was used. The evaluation of the impact of each physico-chemical parameters in the occurrence of each marker was carried out using Spearman's rank order correlation. One-way ANOVA was performed to determine the influence of the occurrence of precipitation on each marker and Spearman's rank order correlation was used for calculation of correlation coefficients between the levels of precipitation (mm), water temperature and salinity values and the concentration of each marker. Spearman’s rank order correlation was used for calculation of correlation coefficients between parameters.

3. Results

3.1. Indicators of fecal contamination

A total of 105 samples were collected from the four different sampling points within the R. Tagus catchment. EC and IE were present in the majority of river water samples (99% and 98%, respectively) with median concentrations of 2.68 log EC MPN/100mL.
and 2.02 log IE MPN/100mL (Table 1). SC were also present in the vast majority of the river water samples tested (99%) with a median concentration of 2.60 log PFU/100mL. These results show a high level of fecal contamination at all the sampling points tested.

3.2. Source Tracking markers

The LoD for assays HAdV, HMMit, PGMit, CWMit, and PLMit were 20.0, 8.0, 8.0, 5.0, and 5.0 GC per reaction, respectively. The human-associated markers of fecal contamination (GB-124 phages, HAdV, and HMMit) displayed notably different distributions (Table 1). HMMit was the most frequently detected marker (present in 83% of samples) and was also present at the highest concentration (4.20 log GC/100mL). HAdV were found in 32% of samples with a median concentration of 2.23 log GC/100mL. Conversely GB-124 phages were detected in just a single river water sample and at low concentration (1 PFU/100mL).

From the non-human markers (cattle (CWMit), pig (PGMit) and poultry (PLMit)), CWMit was the most frequently detected marker (present in 73% of samples) and was also found at the highest concentration (median 3.74 log GC/100mL) (Table 1). PGMit was detected in over 50% of the samples at a median of 2.99 log GC/100mL.

Table 1. Quantitative results and % of positive samples for FIB, non-specific phages (SC), and human (GB124, HMMit, HAdV), cattle (CWMit), pig (PGMit) and poultry (PLMit) fecal markers in R. Tagus catchment

<table>
<thead>
<tr>
<th>Parameter</th>
<th>n</th>
<th>Median (log *units/100mL)</th>
<th>% positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>105</td>
<td>2.68</td>
<td>99</td>
</tr>
<tr>
<td>IE</td>
<td>105</td>
<td>2.02</td>
<td>98</td>
</tr>
<tr>
<td>SC</td>
<td>105</td>
<td>2.60</td>
<td>99</td>
</tr>
<tr>
<td>GB-124</td>
<td>105</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>HAdV</td>
<td>105</td>
<td>2.23</td>
<td>32</td>
</tr>
<tr>
<td>HMMit</td>
<td>105</td>
<td>4.20</td>
<td>83</td>
</tr>
<tr>
<td>CWMit</td>
<td>105</td>
<td>3.74</td>
<td>73</td>
</tr>
<tr>
<td>PGMit</td>
<td>105</td>
<td>2.99</td>
<td>56</td>
</tr>
</tbody>
</table>
Finally, PLMit was the least frequently detected mtDNA marker, present in 39% of samples and at the lowest concentration (2.33 log GC/100mL).

### 3.3. Geographical and temporal distribution

EC and IE were detected in all of the samples tested during winter and autumn, whereas during spring and summer, they were detected in approximately 90% of samples from the sampling points. SC were detected in all samples tested, with the exception of two samples collected during the summer at sampling site 4 (Bel). The median concentration and frequency of detection of each parameter are summarized in Table 2. Human fecal contamination (as determined by the presence of HAdV and HMMit) was detected all year around. HAdV was less prevalent during summer months (22% of samples positive), whereas HMMit was found to be less prevalent during autumn with lower concentrations during summer and autumn months. CWMit marker was also detected frequently (regardless of time of year) but was more abundant during the spring and summer months. PGMit was also consistently detected at similar concentrations, irrespective of season. PLMit was more common during the summer with positive results found at all sampling sites (percentage of positive between 40% and 75% positive).
Table 2. Quantitative results and percentage of positive samples for FIB, non-specific phages (SC), and human (GB124, HMMit, HAdV), cattle (CWMit), pig (PGMit) and poultry (PLMit) fecal markers in R. Tagus catchment by location and season

<table>
<thead>
<tr>
<th>Parameter</th>
<th>1. VFX (*units/100mL)</th>
<th>2. MPdN (*units/100mL)</th>
<th>3. Alc (*units/100mL)</th>
<th>4. Bel (*units/100mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>2.96</td>
<td>2.68</td>
<td>2.69</td>
<td>2.36</td>
</tr>
<tr>
<td>IE</td>
<td>1.79</td>
<td>2.34</td>
<td>2.02</td>
<td>1.91</td>
</tr>
<tr>
<td>SC</td>
<td>2.65</td>
<td>2.54</td>
<td>2.70</td>
<td>2.30</td>
</tr>
<tr>
<td>GB-124 phage</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>HAdV</td>
<td>2.23</td>
<td>2.23</td>
<td>2.56</td>
<td>2.23</td>
</tr>
<tr>
<td>HMMit</td>
<td>4.22</td>
<td>4.20</td>
<td>4.04</td>
<td>4.30</td>
</tr>
<tr>
<td>CWMit</td>
<td>3.73</td>
<td>3.92</td>
<td>3.49</td>
<td>3.68</td>
</tr>
<tr>
<td>PGMit</td>
<td>3.06</td>
<td>3.21</td>
<td>2.54</td>
<td>3.09</td>
</tr>
<tr>
<td>PLMit</td>
<td>2.33</td>
<td>3.02</td>
<td>2.33</td>
<td>2.33</td>
</tr>
<tr>
<td>winter</td>
<td>3.15</td>
<td>2.60</td>
<td>2.24</td>
<td>2.92</td>
</tr>
<tr>
<td>spring</td>
<td>2.46</td>
<td>1.91</td>
<td>1.60</td>
<td>2.20</td>
</tr>
<tr>
<td>summer</td>
<td>3.22</td>
<td>2.45</td>
<td>2.12</td>
<td>2.70</td>
</tr>
<tr>
<td>autumn</td>
<td>2.65</td>
<td>1.94</td>
<td>1.60</td>
<td>2.20</td>
</tr>
</tbody>
</table>

*units = CFU (EC, IE); PFU (SC, GB-124); GC (HAdV, HMMit, CWMit, PGMit, PLMit); Bold denotes the highest reading for each parameter (with respect to sampling site and season)
The concentration of fecal indicators and ST markers differed considerably throughout the year, with the HMMit and CWMit markers showing greater variance in their seasonal values compared to the remaining markers (Fig. 2). The results showed that the concentration of EC, IE and SC generally decreased during summer and spring (Table 2, Fig. 2A).

Fig. 2. Concentration of the various fecal indicators and source tracking markers with respect to season A) culture-based indicators/markers and B) molecular-based markers in the R. Tagus catchment.
In general, the presence of markers did not appear to be significantly influenced by the sampling site location (Table 2, Fig. S1). The influence of sampling site and season upon each parameter was conducted using the analysis of variance (two-way ANOVA).

The concentration of the fecal indicators (EC, IE, SC) was highly affected by seasonality ($p < 0.05$) and for EC the sampling site also played an important role ($p < 0.05$). The variance of the concentration of most source-associated markers was not clearly attributed to either sampling site, or seasonal factors ($p > 0.05$; Fig. 2B), with the exception of the CWMit that showed a statistically significant difference with respect to season ($p < 0.05$; Fig. 2B). Further discrimination was performed to understand the effect of season on the distribution of the indicators and markers (Table S3). For EC, IE and SC there was a significant difference between winter concentrations and concentrations in spring and summer ($p < 0.001$ and $< 0.05$), whereas CWMit concentrations were lowest and highest during autumn and summer, respectively. Nonetheless, CWMit concentrations also varied with respect to the other seasons ($p < 0.05$). EC concentrations were furthermore affected by the sampling site, with Site 4 (Belém) primarily responsible for the observed differences (Table S4).

### 3.4. Correlation between physico-chemical parameters and non- and source-associated markers

Different physico-chemical parameters are known to influence the behavior of microorganisms in the environment, including temperature (atmospheric and water), UV radiation, salinity, and rainfall (IPMA, 2016; MARETEC, 2016). To understand whether temperature and/or UV influenced the concentrations of each parameter, Spearman’s rank order correlation was carried out. Analysis of the effect of UV and temperature on the indicators and markers (Table S5) showed that UV radiation and
atmospheric temperature appeared to have a combined impact on EC, IE, SC, and CWMit. The remaining markers were not significantly affected by atmospheric temperature or UV radiation levels. Another important factor potentially influencing environmental water quality is precipitation, which was evaluated in two ways. Firstly, a binary analysis was conducted (one-way ANOVA) on the influence of rainfall vs. no rainfall (on day of sampling or 24 hrs before) on the presence of each fecal indicator and marker. Secondly, the relationship between rainfall levels (mm) and the concentration at which fecal indicators and source-specific markers were detected was also determined using Spearman’s rank order correlation. One-way ANOVA showed that for the majority of the fecal indicators and markers analyzed, occurrence of rainfall on the day of sampling (or 24 hrs prior), had little, or no appreciable influence on which markers were present (Table S6). However, the one-way analysis of variance identified significant correlation between IE, SC, and PLMit marker and presence of rainfall on the day of sampling (or 24 hrs prior) ($p < 0.05$). Additionally, the influence of rainfall was noticeable (though to a lesser extent) for SC and PLMit in samples from site 3 (Alc) and site 4 (Bel), respectively (Table S6; Fig. S2). Although some parameters were related to the occurrence of rainfall, the correlation between the levels of rainfall and the concentration of other parameters showed a different pattern (Table S7). Interestingly, the presence of rainfall appeared to be moderately correlated with HAdV concentration (Spearman rank order correlation; $r = 0.417$, $p$-value $< 0.05$). The level of rainfall was also negatively associated with the concentrations of IE and SC but showed no correlation with the remaining parameters. The relationship between water temperature/salinity and the concentration of the different parameters was also evaluated at each site using Spearman’s correlation (Table S8). The results revealed that water temperature negatively affected concentrations of EC, IE, and SC
(r = -0.379, p < 0.01; r = -0.404, p < 0.01; r = -0.389, p < 0.01). Weak but positive correlation was found between water temperature and CWMit (r = 0.293, p < 0.01). The remaining parameters were not shown to correlate with water temperature. Water salinity only weakly correlated with the concentration of EC (r = -0.268, p < 0.01) and no correlation was found with IE. Although no correlation was found between other tested parameters and salinity, this parameter influenced most of them negatively (r below zero).

3.5. Relationship between source-associated markers and fecal indicators

To determine whether the presence and concentration of source-associated markers (HMMit, HAdV, GB-124, CWMit, PGMit, and PLMit) were quantitatively correlated to the presence and concentration of fecal indicators (FIB and SC), Spearman’s correlation was performed. Among fecal indicators, the strongest correlation was found (Table 3) between EC and IE (r = 0.673, p < 0.05) and the weakest between IE and SC (r = 0.469, p < 0.05). Significant but weak correlations were also observed between EC and HAdV (p < 0.05), EC and PGMit (p < 0.05), IE and PGMit (p < 0.05), HAdV and CWMit (p < 0.05) and CWMit and PGMit (p < 0.05) (Table 3). The concentration of human- and cattle-associated markers tended to decrease with respect to the concentration of fecal indicators (Table 3; Table S9). PGMit always presented the highest correlation coefficients with EC and SC.
Table 3. Spearman’s correlation between the Log10 concentration of non-specific fecal markers and the different source-specific markers.

<table>
<thead>
<tr>
<th></th>
<th>EC</th>
<th>IE</th>
<th>SC</th>
<th>GB124 phages</th>
<th>HAdV</th>
<th>HMMit</th>
<th>CWMit</th>
<th>PGMit</th>
<th>PLMit</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>-</td>
<td><strong>0.673</strong>(^a)</td>
<td><strong>0.511</strong>(^a)</td>
<td>0.055</td>
<td>0.204(^b)</td>
<td>-0.091</td>
<td>-0.107</td>
<td>0.221(^b)</td>
<td>0.093</td>
</tr>
<tr>
<td>IE</td>
<td>-</td>
<td>-</td>
<td><strong>0.469</strong>(^a)</td>
<td>-0.042</td>
<td>0.083</td>
<td>-0.059</td>
<td>-0.069</td>
<td>0.247(^b)</td>
<td>0.193(^b)</td>
</tr>
<tr>
<td>SC</td>
<td>-</td>
<td>-</td>
<td>-0.097</td>
<td>0.134</td>
<td>-0.079</td>
<td>-0.188</td>
<td>0.080</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>GB124 phages</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.202(^b)</td>
<td>0.019</td>
<td>0.046</td>
<td>0.132</td>
<td>-0.075</td>
<td></td>
</tr>
<tr>
<td>HAdV</td>
<td>-</td>
<td>-0.012</td>
<td>-</td>
<td>0.223(^b)</td>
<td>0.014</td>
<td>0.144</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMMit</td>
<td>-</td>
<td>-</td>
<td>0.061</td>
<td>0.022</td>
<td>-0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CWMit</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.255(^a)</td>
<td>-0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGMit</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLMit</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Two tailed significance, \(p < 0.01\).
\(^b\) Two tailed significance, \(p < 0.05\).
Conversely, the remaining markers showed low, or no significant correlations with the fecal indicators. To understand whether the source-specific MST markers correlated quantitatively to the non-specific fecal markers during rainfall events, correlation analysis using Spearman rank correlation coefficients were performed in this situation (Table S10). PGMit was moderately positively correlated with IE during rain events ($r = 0.444$, $p$-value = 0.05). Although not statistically correlated at a $p$-value of 0.05, PGMit presented the highest correlation coefficients with the remaining indicators of fecal contamination, EC and SC. In stark contrast, the remaining markers showed low, or no significant correlations with the non-specific markers.

4. Discussion

This study was underpinned by a 13-month sampling campaign to determine the presence, concentration and behavior of the fecal indicators and novel source-associated markers, with respect to season and sampling site. This information was lacking from the extant literature for mitochondrial markers, so our findings represent an important gap in current knowledge that is crucial for establishing their suitability as source-associated markers for routine deployment in river catchments. Our study design not only encompassed seasonal fluctuations, but also investigated the influence of precipitation (presence and levels), since it has been demonstrated that more than 90% of the flux of fecal contamination from diffuse sources (as determined by fecal indicator concentrations) can be related to hydrological events (Reischer et al., 2008). The results of this catchment deployment have indicated that human fecal sources are the dominant source of contamination at all four sampling points monitored. Human fecal contamination was more prevalent than the remaining targeted sources of fecal contamination. Fecal contamination of cattle (CWMit) origin
also appeared to heavily influence water quality, as this marker was commonly detected (in over 70% of samples) and at high concentrations (median concentration $3.74 \log \text{GC/100mL}$). Sampling sites 1 (VFX) and 2 (MNdP) were suspected to both contain a significant component of contamination from cattle origins, since site 1 is known for the rearing and exportation of cattle and site 2 receives the influent from two WWTP known to receive waste from cattle farms, along with direct riverine inputs (run-off) from the cattle farms on the south side of the river. Nonetheless, the other animal sources were also frequently detected during this study, with more than half of the samples testing positive for porcine contamination and about 40% positive for poultry. Porcine fecal inputs were thought to be more likely at sampling site 2, due to its proximity to pig farming operations on not only the north bank of the R. Tagus, but also on the southern bank. Although found in higher concentration on sampling site 2, porcine fecal contamination was also detected at all sampling locations, with sampling site 1 having similar prevalence and concentration to sampling site 2. Lisbon metropolitan area consists of 18 municipalities, many of which are predominantly urban, with approximately three million inhabitants (INE, 2014). The city of Lisbon, with a population of approximately 500,000 inhabitants (INE, 2014), is mainly an urban area but many of the municipalities on the north and south bank of the River Tagus are not only densely populated with human inhabitants but also characterized by high numbers of agricultural animals, including cattle farms, pig farms and poultry abattoirs. Despite being regarded as an urban area, in 2018 the Lisbon metropolitan area contained 5% of the total number of cattle in Portugal (75,000 cattle units), 10% of the total number of pigs (220,000 pigs) and more than 180,000 poultry (INE, 2019). In addition to the large number of farm animals on the outskirts of the city, the farms and abattoirs are often not equipped with adequate wastewater treatment processes, or
do not have any form of on-site treatment, meaning that fecal contamination generated by these industries either goes mostly directly to the municipal wastewater, or is discharged into nearby environmental waters. The WWTP featured in this study did not (at the time the study was conducted) have sufficient treatment/disinfection in place to substantially reduce the microbiological parameters (and also mitochondrial cells) in the final effluent. These factors are likely to be the reason why a high percentage of samples collected during this study contained high concentrations of human and agricultural inputs. On the other hand, the human marker GB124 bacteriophage was detected in just one sample, at low concentration, suggesting that this host appeared to be geographically restricted from the catchment and that other Bacteroides hosts (e.g. GA-17) may be more appropriate in this region. The existence of phages in feces and wastewater is highly dependent on the presence in the mammalian gut of Bacteroides host strains, which are homogenous in terms of receptors and modification-restriction enzymes resulting in capability of phage replication (Puig et al., 1999; Ebdon et al., 2012; McMinn et al., 2014). Therefore, the diversity of Bacteroides strains is the main factor responsible for the variations in the levels of infectivity. Bacteroides strains have evolved over time in very well compartmentalized environments and have been segregated according to regional dietary regimens and other factors such as host age and climate (Benno et al., 1989; Moore and Moore, 1995; Puig et al., 1999). B. fragilis strain GB124 was isolated in the UK and the phages infecting this strain have been shown to be almost exclusively of human origin. The diet in Southern Europe is somewhat different from that of the UK, where climate is generally cooler and wetter, two factors that may influence the different evolutionary paths in Bacteroides strains. Nonetheless, phages capable of infecting B. fragilis strain GB124 have been found to be present at detectable levels
in studies carried out across the US and Brazil (McMinn et al., 2014; 2017; Prado et al., 2018).

Fecal indicators generally displayed a marked seasonal pattern, with higher concentrations during the colder months and lower concentrations during the warmer months. This suggests that the decrease in the fecal indicators is correlated with the increase in solar radiation (intensity and duration) and possible biological predation of the indicators/markers within environmental waters (Kay et al., 2005; Schultz-Fademrecht et al., 2008; Byappanahalli et al., 2012). Conversely, the concentration of the source-associated markers was not impacted by seasonality, with the exception of the cattle marker, which was present at higher concentration in warmer months, particularly during summer. The increase in the concentration of this source-associated marker is likely the result of a number of potential factors. First, cattle numbers are far higher in the catchment during the warmer months (as animals housed indoors over winter are moved outside); secondly, the chance of rainfall (storm events) at this time of the year is more likely to mobilize diffuse fecal material into surface waters; and thirdly, low-flow conditions mean that fecal inputs have a greater impact on the water quality and source-associated markers are more abundant due to reduced dilution. These findings are in accordance with the findings in other studies, e.g., by Rusiñol et al., 2014, who detected contamination of animal origin at a higher concentration during the summer in a study performed in five different countries (Brazil, Greece, Hungary, Spain, and Sweden)

Our study also showed that IE, PLMit and HAdV were impacted by rainfall occurrence and levels. Similar results were demonstrated in previous studies (Haramoto et al., 2006; Hata et al., 2014; Rodrigues et al., 2015). Levels of the PLMit marker were significantly lower in samples collected on dry days, compared to those collected
following precipitation (during previous 24hrs) and concur with previous findings (Barros et al., 2007). HMMit displayed no clear seasonality, being unaffected by atmospheric temperature and UV radiation levels. Surprisingly, little information exists on the influence of different environmental physicochemical parameters on the degradation of mitochondrial cells, or on the mitochondrial DNA itself. The detection of mitochondrial DNA from human, cattle, and porcine sources in raw wastewater, partially treated wastewater and disinfected effluent was studied previously (Martellini et al., 2005) and showed that following each treatment stage, only human mitochondrial DNA remained at detectable levels. He et al. (2016) determined that temperature and sunlight negatively correlated with the concentration of human and pig mtDNA markers in a field microcosm using dialysis tubes. Malla and Haramoto, 2020 have compiled the information on mtDNA MST markers and their persistence in the environment. However, the studies presented were performed in microcosms, rather than on samples collected directly from the river channel and were not tested alongside as larger number of physico-chemical parameters as featured in this study. The PGMit marker was weakly correlated to EC and IE during the entire sampling campaign but showed moderate correlation with IE during and following precipitation events. Therefore, this study demonstrated the capacity of certain markers, such as PGMit, to link source-associated markers concentration to fecal indicators. This should facilitate the estimation of the contribution from particular sources to the total load of fecal contamination assessed by conventional fecal indicators. The correlation found between PGMit and IE is even more interesting when considering the US EPA’s revised recreational water quality criteria for IE, that aims to keep the risk of gastrointestinal illness in swimmers below 30 illnesses per 1000 swimmers (USEPA, 2012).
This study demonstrated that the 13-month sampling campaign has contributed to an improved understanding of the effect of seasonality on indicator and source-associated marker dynamics and highlighted the importance of catchment and precipitation event sampling for effective quantitative source tracking. Moreover, the parametric analysis of precipitation events allowed for the quantitative assessment of the porcine fecal contamination component. It must be emphasized here that the potential dominant sources of fecal contamination in study sites on the R. Tagus were either of human or cattle origin.

Whilst it is clear that cattle, pig and poultry rearing are common at certain points within the study catchment, the prevalence and concentration of non-human markers during the sampling campaign, highlight the potential impact that these non-point sources can have on water quality within the R. Tagus catchment. Although human sources are a major contributor to contamination within the R. Tagus (and as such represent a risk to public health), fecal contamination from a range of non-human sources may also present considerable potential risks to public health. For example, several zoonotic pathogens are able to infect or exist in a range of different warm-blooded mammals (including humans), such as Cryptosporidium and Giardia, and Hepatitis E virus. In addition to the application of tailored models to the area of study by combining local samples including also factors such as spatial and seasonal inactivation, it has been suggested (Ballesté et al., 2020) that dilution and aging may help improve the accuracy reducing the number of source-associated markers to be tested. To conclude, this study provides new data concerning the temporal and spatial distribution of fecal indicators and source-associated markers and correlations between different variables, which could be extremely helpful in the management of
water quality and in the application of remediation measures to enhance and restore the R. Tagus catchment.

**Declaration of competing interest**

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


source tracking tools in fresh and marine waters from five different geographical areas.

Water Res. 59, 119-129.


