Pathogens, faecal indicators and human-specific microbial source-tracking markers in sewage

C. García-Aljaro,1,2 A.R. Blanch,1,2 C. Campos3, J. Jofre1,2 and F. Lucena1,2

1 Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain
2 The Water Research Institute, University of Barcelona, Barcelona, Spain
3 Departamento de Microbiología, Pontificia Universidad Javeriana, Bogotá, Colombia

Summary

The objective of this review is to assess the current state of knowledge of pathogens, general faecal indicators and human-specific microbial source tracking markers in sewage. Most of the microbes present in sewage are from the microbiota of the human gut, including pathogens. Bacteria and viruses are the most abundant groups of microbes in the human gut microbiota. Most reports on this topic show that raw sewage microbiological profiles reflect the human gut microbiota. Human and animal faeces share many commensal microbes as well as pathogens. Faecal-orally transmitted pathogens constitute a serious public health problem that can be minimized through sanitation. Assessing both the sanitation processes and the contribution of sewage to the faecal contamination of water bodies requires knowledge of the content of pathogens in sewage, microbes indicating general faecal contamination and microbes that are only present in human faecal remains, which are known as the human-specific microbial source-tracking (MST) markers. Detection of pathogens would be the ideal option for managing sanitation and determining the microbiological quality of waters contaminated by sewage; but at present, this is neither practical nor feasible in routine testing. Traditionally, faecal indicator bacteria have been used as surrogate indicators of general faecal residues. However, in many water management circumstances, it becomes necessary to detect both the origin of faecal contamination, for which MST is paramount, and live micro-organisms, for which molecular methods are not suitable. The presence and concentrations of pathogens, general faecal indicators and human-specific MST markers most frequently reported in different areas of the world are summarized in this review.

Introduction

Sewage is wastewater from homes and other buildings, which flows through municipal sewers. Sewage usually travels from the plumbing installation in the building into a sewer, which will carry it away, either directly to water bodies or into a sewage treatment facility. Sewers may also receive surface run-off water; whether this happens depends on the sewer design and this usually depends on when it was constructed and in which country. Municipal sewage is one of the largest sources of pollution discharged into surface water bodies worldwide. As an example, the microbial pathogens released through sewage are the second leading cause of impairment in US water bodies (USEPA 2004). Sewage mostly consists of grey water (from sinks, baths, showers, dish washers, washing machines, etc.) and black water (from toilets) and, depending on the municipality, it may contain wastewater from factories. It is characterized by its physical characteristics, chemical and toxic constituents, and microbiological status, which is characterized by the diversity and concentrations of micro-organisms, including pathogens.
Among the microbes present in sewage, waterborne pathogens cause illnesses around the world, and constitute one of the major public health concerns worldwide. At the seventieth UN General Assembly held in 2015, 193 countries jointly adopted the Sustainable Development Goals, one of which pledges to provide universal access to adequate sanitation and to halve the proportion of untreated wastewater discharged into nature (United Nations 2015). By recognizing the importance of sanitation, the UN recognized the importance of sewage in the transmission of pathogens that cause waterborne diseases.

Waterborne pathogens enter waterways via a number of routes, including raw sewage, inadequately treated sewage, storm drains, septic systems, run-off from livestock pens and boats that dump sewage. Identifying, quantifying and apportioning the microbial faecal contaminants carried by sewage are essential to assess the impact of sewage discharge into water bodies, to trace the contaminating source and to assess the performance of wastewater treatments.

Due to the usually low and fluctuating numbers of pathogens present in sewage, and methodological difficulties with their detection and enumeration, health and environmental authorities have used indicator micro-organisms since the first quarter of the 20th century to evaluate both the quality of water and the performance of water treatment (WHO 2001; Fujioka 2002). In spite of the countless last-generation molecular methods implemented to detect pathogens, the complexity of sample matrices and the sampling, DNA preparation, processing, sequencing and analysis protocols make studying samples, such as sewage, intricate and barely feasible for practical and routine purposes (Raes and Bork 2008; Raoult and Henrissat 2014). Consequently, the use of indicators is still needed for the assessment of both water quality and water management.

Nevertheless, usually the term ‘indicator’ is used to denote both the index and the indicator function (WHO 2001; Fujioka 2002). The index is related to the occurrence of the surrogate micro-organisms or material (sewage, faecal remnants, pathogens, etc.). In contrast, the indicator function has a much broader definition that includes characteristics regarding resistance to treatments and persistence in water environments similar to those of the surrogate micro-organisms. Traditionally, faecal indicator bacteria (FIB) are considered as indicators that cover both functions; but nowadays, we know that they do not. They do not share either resistance to treatments or persistence in the environment with viruses (AWPRC Study Group on Health Related Water Microbiology 1991; Grabow 2001; WHO 2001) and do not distinguish the source of contamination (Malakoff 2002). This is the reason why there is now growing interest in broadening the spectrum of indicators available.

Assessing both sanitation processes and the contribution of sewage to the faecal contamination of water bodies requires knowledge of the content of pathogens in sewage, microbes indicating general faecal contamination and microbes that are only present in human faecal remains, which are known as human-specific microbial source-tracking (MST) markers.

Origin and diversity of microbes in sewage

Sewage may contain microbes from multiple origins, but in most settings, the vast majority of microbes present in the sewage in a given sewer come from the microbiota of the humans living in the houses from which sewage is collected and, to a lesser extent, nonfaecal micro-organisms, a fraction of which may be residents of the urban sewer infrastructure. Minor faecal contributions can also be attributed to the gut microbiota of animals living in the sewer systems: rats, cockroaches, etc. When combined with surface run-off, the potential origin of the microbiota in the sewage broadens due to the addition of soil, faeces from pets, environmental waters, etc.; but even in this case, sewage in urban areas mostly receives micro-organisms corresponding to the human microbiota.

Human microbiota

The number of microbial cells in the human body is estimated to average from $3.5 \times 10^{13}$ to $4.0 \times 10^{13}$ (Sender et al. 2016). The main contributor to the human microbiome is the gut. The low microbiological levels of grey waters (which include human surface microbes carried away when washing) confirm this assessment (Winward et al. 2009). Moreover, among the gut sections, the colon, with an estimated concentration exceeding $10^{11}$ bacteria per ml and an average volume of 400 ml, is by far the main contributor to the human microbiota (Sender et al. 2016). However, only about one third of these bacteria can be cultivated by current methods (Simon and Gorbach 1984), although culturomics has made great progress recently with regard to the diversity of gut microbes detected by culture (Lagier et al. 2016). Thus, colon content, released into the sewage through faeces (100–200 g wet weight of faeces per day and person on average), represents by far the greatest input of microbes into sewage, with some $1 \times 10^{12}$–$2 \times 10^{13}$ microbial cells per person per day in healthy individuals. Of course, these numbers vary in some pathological conditions, such as diarrhoea.

The human colon microbiota is a complex population that consists of bacteria, archaea, microbial eukaryotes and viruses. Methods based on quantitative PCR and new generation methods such as high-throughput sequencing and bioinformatic analysis of metagenomic studies have recently increased the information available on both the
diversity and the relative abundances estimated in the past using cultural methods. For healthy individuals, the relative abundance and diversity is as follows. Bacteria represent the most numerous and diverse domain. Members of many bacterial phyla reside in the human gastrointestinal tract, and there is a wide consensus that Bacteroidetes (i.e. Bacteroides and Prevotella) and Firmicutes (i.e. Clostridium, Lachnospiraceae; Ruminococcus and Enterococcus) are the dominant ones, accounting for more than 90% in most studies. Members of Proteobacteria (i.e. Escherichia coli, Acinetobacter, Pseudomonas, Aeromonas, etc.) are also common and diverse, but they represent only some 0.1% of the total value at most (Bäckhed et al. 2005; Eckburg et al. 2005; Qin et al. 2010; Arumugam et al. 2011; The Human Microbiome Project Consortium 2012). The diversity, as estimated by phylogenetic analysis, is colossal, with roughly 1000 species present. Several clones (biotypes) of each species are present in a given individual, and hence there are believed to be between 7000 and 10 000 different bacterial clones in the gut microbiota of a healthy person (Bäckhed et al. 2005; Lozupone et al. 2012). Archaea are characterized by low diversity and concentrations not greater than 10^6 cells per ml: well below the bacterial figures (Eckburg et al. 2005; Abell et al. 2006). Yeasts, filamentous fungi and, to a lesser extent, protozoa are the dominant microbial eukaryotes in the human gut of healthy individuals (Parfrey et al. 2011; Hallen-Adams and Suhr 2017). Reported values range from 10^6 to 10^7 cells per ml (Wang et al. 2014) and the reported diversity is very low (Parfrey et al. 2011).

Other than cellular microbes, viruses, mostly bacteriophages, represent another important portion of the human gut microbiota (Breitbart et al. 2003; Letarov and Kulikov 2009; Reyes et al. 2012). Their numbers in the gut or in faecal samples are quite high, with reported values up to 1 × 10^{10} virus-like particles (VLP) per ml (Letarov and Kulikov 2009). Regarding diversity, more than 1200 viral genotypes have been identified in human faeces (Breitbart et al. 2003). Among the RNA viruses found in the gut, a prevalence of transient plant viruses has been reported in certain human faeces samples (Zhang et al. 2006).

In spite of the general trends indicated in the precedent paragraphs, the gut microbiota is immensely diverse, varies between individuals and can fluctuate over time in a given person (Minot et al. 2011; Lozupone et al. 2012). Age, genetics, environment, diet and socioeconomic and cultural factors seem to be causes of this interindividual variation (Ley et al. 2006; Lozupone et al. 2012). However, in spite of interindividual variations, the combined contribution of the individuals in a certain population seems to give similar sewage compositions in a given geographical or cultural area (Western countries, Africa, South America, etc.) that are stable over time. In contrast, pronounced differences in bacterial assemblages between residents of different cultural and geographical areas have been reported (Arumugam et al. 2011; Wu et al. 2011; Yatsunenko et al. 2012).

Furthermore, a fraction of individuals contribute pathogens; the amounts and diversity of which vary over time (seasonality, occurrence of outbreaks or epidemics) and geography (the prevalence of infectious diseases in different areas of the globe is highly variable). Indeed, sick people at the acute stage of a disease contribute high amounts of pathogens over short periods (days). Additionally, asymptomatic carriers contribute low amounts, but over extended periods. As in the case of commensals, there is also geographical variation in pathogen incidence; although in this case, it seems that the major causes are socio-economic (poor sanitation, no access to safe drinking water, inadequate hygiene, etc.) rather than age, genetics, environment and diet. Among the pathogens in faeces, we mostly find those transmitted via the faecal-oral route.

Animal microbiota

Another aspect that is crucial for water resource management is the composition of the microbiota found in the caecum/colon of healthy animal specimens that is excreted as faeces, and that reach the same water bodies as sewage through run-off. The general trends for different animals are similar to those found in humans regarding both the concentrations (10^{11} bacteria per ml colon content); proportions of bacteria (more than 98%), archaea (about 1%) and microbial eukaryotes (<0.5%); and diversity and interindividual variation. Regarding bacteria, Firmicutes and Bacteroidales are the most abundant for different animals; and numbers of Proteobacteria are far smaller, as in humans (De Menezes et al. 2011; Isaacson and Kim 2012; Oakley et al. 2014). However, detailed phylogenetic analysis reveals that in spite of the dominance of the same phyla, at the genus, species or biotypes level, there are differences between different animal species. Regarding the virome, the results show the same trends as in humans; there is predominance of bacteriophages and animal viruses more or less specific to the animal species, great diversity and concentrations one to two log units higher than for bacteria (Sachsenröder et al. 2014; Zhang et al. 2014).

Just as in humans, a percentage of animals contribute pathogens; some being animal species specific, but others infecting more than one species, including humans: those that cause zoonotic infections. The percentage of illnesses that are caused by zoonotic pathogens is difficult to determine due to a lack of data, but it is thought to be
very significant (Cotruvo et al. 2004), and more difficult to control than those caused by human-specific pathogens.

Run-off water
Run-off water or storm water contributes microbial pollutants to urban wastewater in combined sewer systems. Rain certainly washes microbes from exposed surfaces within the urban constructed (e.g. buildings, streets, roads) and natural (e.g. animals, plants, soil) environments into combined sewers. Among the microbes carried by storm run-off water, urban wildlife and domestic pet waste are the primary sources of faecal microbes. However, other faecal contaminants are present originating from the exposed surfaces in urbanized areas, which constitute ecosystems with their own microbiota. The scarce information available obtained by metagenomics indicates a predominance of Proteobacteria (Acinetobacter, Pseudomonas, Aeromonas, Arcobacter, etc.) (Newton et al. 2013). However, it is to be presumed that the composition of run-off water will be strongly influenced by the characteristics of a given urban area.

Sewage microbiota
Reported differences between raw sewage and activated sludge microbiotas (Liu et al. 2007; Ye and Zhang 2013; Cai et al. 2014) reflect shifts in microbial community composition from original influents to sewage effluents.

Most available information on the microbial content of raw sewage corresponds to influent sewage in wastewater treatment plants. Different metagenomic studies have shown that the great majority of faecal taxa, as well as the population structures of human faecal samples, are preserved in the microbial profile of influent sewage in wastewater treatment plants (McEllan et al. 2010; Newton et al. 2015). However, those studies have also shown that taxa belonging to the phylum Proteobacteria found in influent sewage are associated with the sewer infrastructure or run-off water (Shanks et al. 2013; Newton et al. 2015). These increases in the prevalence of Proteobacteria, mostly of the genera Acinetobacter, Aeromonas, Pseudomonas, Tricococcus and Arcobacter, among others, which occur at much lower relative levels in faeces (Vandewalle et al. 2012), suggest that these organisms proliferate within the sewer system where they are residents. Detailed concentrations of human pathogenic bacteria will be reviewed in a later section.

Concentrations of VLP ranging from $10^8$ to $10^{10}$ per millilitre have been detected in raw sewage (Wu and Liu 2009; Tamaki et al. 2012). Metagenomic analysis of viruses in raw sewage shows great diversity and a predominance of bacteriophages (Cantalupo et al. 2011; Tamaki et al. 2012); although human viruses have also been detected (Cantalupo et al. 2011). Detailed concentrations of human pathogenic viruses will be reviewed later.

Regarding indicators and potential source markers, the numbers of cultivable E. coli, enterococci, Bacteroides and Bifidobacterium in human colon content have been reported to slightly exceed $10^6$, approach $10^8$, and slightly exceed $10^{11}$ and $10^{10}$ per ml on average respectively (Simon and Gorbach 1984). Their most frequently reported colony-forming unit (CFU) and genome copy (GC) values in sewage are shown in Tables 1 and 2. Comparing CFU estimates of the four groups in the gut and the values reported in sewage leads to the conclusion that values found in raw sewage fit with the numbers reported in the colon, bearing in mind that defecation per person and day is 100–200 g, and the production of wastewater is 200–400 l per person and day. Furthermore, the relative numbers indicate that neither E. coli nor enterococci, both facultative anaerobes, suffer significant decay in CFUs during their transit through the sewerage system. In contrast, the two strict anaerobic markers, Bacteroides and Bifidobacterium, undergo an important reduction in viability, but not in GCs, during their sewerage system transit.

The fate of the typical faecal-orally transmitted pathogens reported in Table 3 in the sewerage network has not been assessed, but we can presume that it is similar to the facultative anaerobic bacterial indicators.

Effluent from wastewater treatment facilities
Most municipal wastewater treatment plants are bioreactors designed to treat domestic sewage. Use of activated sludge is the most common procedure. Treated effluents from activated sludge wastewater treatment plants have lower densities of most incoming contaminant microbes; but they are still a continuous source of urban-derived microbes for the receiving waters (Ye and Zhang 2013). Hence, one task for wastewater treatment plant managers is to monitor and track the human sewage microbiota, in order to document its fate and removal.

The microbial profile of influent sewage does not remain constant through the depuration process; although bacteria clearly predominate over the other domains throughout treatment processes (Yu and Zhang 2012). The selection of the predominant families of bacteria during the depuration process is mostly determined by the chemical composition of the incoming waters and type of treatment, and they are very different from one report to another (Sanapareddy et al. 2009; Yu and Zhang 2012; Cai et al. 2014; Shchegolkova et al. 2016). The composition of effluents, which are released into the receiving waters, usually reflects the composition of the activated sludge (Cai et al. 2014).
Table 1  Concentration of human sewage markers

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Method</th>
<th>Concentration (100 ml⁻¹)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathogens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAdV</td>
<td>qPCR</td>
<td>Undetectable—10⁶ GC</td>
<td>Bofill-Mas et al. (2006); Rusiño et al. (2014); Wong et al. (2012)</td>
</tr>
<tr>
<td>HPyV</td>
<td>qPCR</td>
<td>Undetectable—10⁶ GC</td>
<td>Bofill-Mas et al. (2006); Rusiño et al. (2014); Wong et al. (2012)</td>
</tr>
<tr>
<td>Bacteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteroidales (HF183, BacHum and other)</td>
<td>qPCR</td>
<td>10⁸ CFU</td>
<td>Ahmed et al. (2016); Harwood et al. (2018); Silkie and Nelson (2009)</td>
</tr>
<tr>
<td>Bifidobacterium spp.</td>
<td>Culture</td>
<td>10⁵–10⁷ CFU</td>
<td>Harwood et al. (2018); Gourmelon et al. (2010)</td>
</tr>
<tr>
<td>B. adolescentis</td>
<td>qPCR</td>
<td>10⁶–10⁷ GC</td>
<td></td>
</tr>
<tr>
<td>Enterococci</td>
<td>Culture</td>
<td>10⁵–10⁷ CFU</td>
<td>Akiba et al. (2015); WHO (2001); Rose et al. (2004); WHO (2006)</td>
</tr>
<tr>
<td></td>
<td>qPCR</td>
<td>10⁶–10⁷ GC</td>
<td></td>
</tr>
<tr>
<td>Bacteriophages</td>
<td>Plaque hybridization</td>
<td>&gt;85%</td>
<td>Blanch et al. (2006); Schaper et al. (2002)</td>
</tr>
<tr>
<td>Genotypes II and III of F-RNA-specific bacteriophages</td>
<td>Plaques+RTqPCR</td>
<td>&gt;61%</td>
<td>Haramoto et al. (2015)</td>
</tr>
<tr>
<td>Bacteriophages infecting Bacteroides spp.</td>
<td>Bacteroides</td>
<td>1.5 × 10⁴ PFU</td>
<td>Blanch et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Bifidobacterium thetaetaomicron GA17</td>
<td>50–70 × 10⁵ (EU)</td>
<td>Ebdon et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>Bacteroides fragilis GB124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>crAssphage</td>
<td>qPCR</td>
<td>Up to 10⁶ GC</td>
<td>García-Aljaro et al. (2017)</td>
</tr>
</tbody>
</table>

The virus composition also changes in both numbers and prevalent groups, through wastewater treatments (Otawa et al. 2007; Wu and Liu 2009); probably because the predominant viruses are bacteriophages that replicate in the bacteria living in the water treatment environments.

Pathogens are erratically detected in activated sludge and seldom in effluents by metagenomic analysis, and even by more species-specific molecular methods such as real-time qPCR. This is highly probable due to their concentrations being several orders of magnitude lower than those of the predominant bacteria and to the analytical complexity of sewage matrices. However, the reported reductions due to wastewater treatment of human sewage are: bacteria (Cai et al. 2014); cultivable indicator bacteria such as coliform bacteria, E. coli and enterococci (Lucena et al. 2004; Hata et al. 2012); cultivable viral indicator bacteriophages (Lucena et al. 2004; Hata et al. 2012); and the most common bacterial and viral pathogens determined by molecular methods (Hata et al. 2012; Cai et al. 2014; Li et al. 2015) range from 1.0 to 3.0 log₈ units, depending on the wastewater treatment. Consequently, the pathogen content of effluents from common wastewater treatment facilities depends on the pathogen content of the sewage, which can vary enormously as explained below, and the characteristics of the wastewater treatment plant.

Pathogens in raw sewage

Many pathogens, particularly viruses, can be found in high concentrations in the urine and faeces of infected individuals, and hence they are detected in sewage (Sinclair et al. 2008). However, only those whose transmission via the faecal–oral route has been unequivocally established are considered in this review.

The relevant pathogenic viruses, bacteria, protozoa and multicellular parasites transmitted by the faecal–oral route are listed in Table 3, which gives the diseases they produce, their zoonotic condition and the concentrations reported in raw sewage.

Pathogens tend to be scarce in sewage, compared to gut commensal micro-organisms. The detection of pathogens typically requires a previous concentration phase (adsorption, elution or centrifugation for viruses), enrichment (culture in enrichment media for cultivable bacteria; immune-magnetic separation for bacteria and protozoa), or flotation or phase separation (for multicellular parasites).

Culture methods usually based on enrichment and selective culture media are available for all the bacteria mentioned in Table 3. Some of them can be detected using simple approaches, as in the case of E. coli O157: H7 with combined detection using immunomagnetic Sorbitol-MacConkey Medium (March and Ratnam 1986). Others, such as Salmonella, for example, require a more complex approach including pre-enrichment, enrichment and selection, followed by identification of presumptive colonies (Haley et al. 2009). Quantification is achieved either by counting CFUs in agar media, or by quantal methods such as the Most Probable Number in liquid media. Some methods that are well adapted to clean
water samples present difficulties when applied to complex matrices such as sewage. With the exception of quantifying cultivable enteroviruses in BGM cells (EPA 2010), feasible and standardized culture methods are not available for most viruses. In the case of sewage, problems of the toxicity of the sample for the cell culture frequently appear.

Following concentration and enrichment, multicellular parasites and protozoa are routinely detected by microscopic observation or particle detection. Regular microscopy is the preferred method for detecting multicellular parasites for which a digital imaging system to identify and quantify several species of helminth eggs has recently been developed (Maya et al. 2006; Jiménez et al. 2016). Advanced methods such as immunofluorescence microscopy (Rose et al. 1989), flow cytometry (Vesey et al. 1994) or laser scanning cytometry (Montemayor et al. 2007) are preferred for protozoa.

All four groups of pathogens can nowadays be detected by molecular methods based on qPCR, RT-qPCR or ICC-qPCR and ICC-RTqPCR. Nonetheless, in complex matrices such as sewage, there is considerable uncertainty regarding both the concentrations detected and the ratios between noninfectious and infectious units. Quantities detected by molecular methods are greater than those obtained by culture methods (Johnson et al. 1995; Kong et al. 2002; Girones et al. 2010; Hamza et al. 2011; Rocha et al. 2016). Nucleic acid amplification techniques based on membrane integrity have been developed for preferential detection of viable bacteria, such as ethidium monoazide PCR (EMA-PCR) and propidium monoazide PCR (PMA-PCR). However, these methods also have several drawbacks; particularly, the fact that they do not allow for monitoring of the efficacy of UV treatment (Nocker et al. 2007) or other inactivation mechanisms that do not directly target cell membranes (Kim et al. 2008).

Many water-transmitted bacteria, protozoa and multicellular parasites are zoonotic (Table 3; Cotruvo et al. 2004). In the case of viruses, the zoonotic condition seems less frequent and to date only HEV has been unequivocally classified as such. Zoonotic pathogens are much more difficult to control through sanitation than those specific to humans. Some pathogens (Salmonella typhi, Shigella dysenteriae and Vibrio cholerae) which were the most worrying pathogens transmitted by the faecal–oral route through water in the past, have now almost been eradicated in high-income countries, but not so in low-income countries (Kotloff et al. 1999; Crump et al. 2004; Ali et al. 2012). Furthermore, multicellular parasites such as Ascaris lumbricoides, Trichuris trichiura, Ancylostoma duodenale and Necator americanus, which are usually zoonotic, have almost been eradicated in high-income countries, but still affect some 30% of the population in low-income nations (WHO 2006; Jiménez et al. 2007).

Most of the pathogens in Table 3 cause gastroenteritis and it has been estimated that globally, 1.45 million people, mostly children, die yearly as a result of diarrhoeal illnesses, 58% of which are caused by inadequate water sanitation and hygiene (WHO 2006). There are very important quantitative and qualitative regional differences, mainly driven by socioeconomic factors. As an example, in the WHO region (http://www.who.int/about_regions/en/) with the lowest income, Africa, the burden of diarrhoeal disease measured in Disability-Adjusted Life Years is 8.5% and by mortality it is 9.0%; whereas in high-income countries these values are <0.1% in both cases (Prüss-Üstün et al. 2008). Rotavirus, E. coli, Shigella and Cryptosporidium are the most prevalent aetiological agents of moderate-to-severe diarrhoea in low-income countries (http://www.who.int/mediacentre/factsheets/fs330/en/, accessed 30 April 2018), whereas norovirus, rotavirus and Campylobacter are the most common in high-income ones (Feigin 2016).

### Table 2
Concentrations of faecal indicators in sewage

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Method</th>
<th>Concentration (100 ml⁻¹)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faecal coliforms</td>
<td>Culture</td>
<td>10⁶–10⁷ CFU</td>
<td>Akiba et al. (2015); WHO (2001); Rose et al. (2004); WHO (2006)</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>Culture</td>
<td>10⁵–10⁶ CFU</td>
<td>Akiba et al. (2015); WHO (2001); Rose et al. (2004); WHO (2006)</td>
</tr>
<tr>
<td></td>
<td>qPCR</td>
<td>10⁶–10⁷ GC</td>
<td>Silkie and Nelson (2009); Srinivasan et al. (2011)</td>
</tr>
<tr>
<td>Enterococci</td>
<td>Culture</td>
<td>10⁶–10⁷ CFU</td>
<td>Akiba et al. (2015); WHO (2001); Rose et al. (2004); WHO (2006)</td>
</tr>
<tr>
<td></td>
<td>qPCR</td>
<td>10⁶–10⁷ GC</td>
<td>Silkie and Nelson (2009); Srinivasan et al. (2011)</td>
</tr>
<tr>
<td>SRCS*</td>
<td>Culture</td>
<td>10⁶–10⁷ CFU</td>
<td>Akiba et al. (2015); WHO (2001); Rose et al. (2004); WHO (2006)</td>
</tr>
<tr>
<td>Somatic coliphages</td>
<td>Culture</td>
<td>5 × 10⁵–5 × 10⁷ PFU</td>
<td>Contreras-Coll et al. (2002); Teklehaimanot et al. (2014); Yahya et al. (2015)</td>
</tr>
<tr>
<td>F-specific coliphages</td>
<td>Culture</td>
<td>10⁵–10⁶ PFU</td>
<td>Contreras-Coll et al. (2002); Hata et al. (2015); Yahya et al. (2015)</td>
</tr>
<tr>
<td>Helminth eggs</td>
<td>Microscopic observation</td>
<td>&lt;0.1–10² PFU</td>
<td>WHO (2006)</td>
</tr>
</tbody>
</table>

*Sulphite-reducing Clostridia spores.
Table 3  Pathogens in sewage

<table>
<thead>
<tr>
<th>Agent</th>
<th>Disease</th>
<th>Concentrations in sewage (100 ml⁻¹)</th>
<th>Reference†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter</em></td>
<td>Gastroenteritis</td>
<td><strong>10²</strong>—<strong>10⁵ MPN</strong></td>
<td>Jones (2001); WHO (2006)</td>
</tr>
<tr>
<td>Enteropathogenic</td>
<td>Gastroenteritis, HUS</td>
<td><strong>10¹</strong>—<strong>10² MPN</strong></td>
<td>García-Aljaro et al. (2005)</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E. coli O157, STEC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Salmonella spp</em></td>
<td>Gastroenteritis</td>
<td><strong>1</strong>—<strong>10⁴ MPN</strong></td>
<td>Koivunen et al. (2003); WHO (2006)</td>
</tr>
<tr>
<td><em>Shigella spp</em></td>
<td>Gastroenteritis</td>
<td><strong>10¹</strong>—<strong>10² MPN</strong></td>
<td>Pant and Mittal (2007); Teklehaimanot et al. (2014, 2015)</td>
</tr>
<tr>
<td><em>Vibrio cholerae</em></td>
<td>Gastroenteritis</td>
<td>&lt;1—10⁴ MPN</td>
<td>Pant and Mittal (2007); WHO (2006)</td>
</tr>
<tr>
<td><strong>Helminths</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascaris lumbricoides*</td>
<td>Helminthiasis</td>
<td>&lt;0—1—10²</td>
<td>WHO (2006)</td>
</tr>
<tr>
<td>Trichuris trichurus*</td>
<td>Helminthiasis</td>
<td>&lt;0—1—10</td>
<td>WHO (2006)</td>
</tr>
<tr>
<td>Anclylostoma duodenalis*</td>
<td>Helminthiasis</td>
<td>&lt;0—1—10²</td>
<td>WHO (2006)</td>
</tr>
<tr>
<td>Necator americanus*</td>
<td>Helminthiasis</td>
<td>&lt;0—1—10²</td>
<td>WHO (2006)</td>
</tr>
<tr>
<td><strong>Protozoa</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giardia lamblia*</td>
<td>Giardiasis</td>
<td>3—10⁴</td>
<td>Hachich et al. (2013); WHO (2006)</td>
</tr>
<tr>
<td>Cryptosporidium parvum*</td>
<td>Cryptosporidiosis</td>
<td>&lt;1—10³</td>
<td>Montemayor et al. (2005); WHO (2006)</td>
</tr>
<tr>
<td><em>Cyclospora cayetanensi</em></td>
<td>Persistent diarrhoea</td>
<td>&lt;20% positive samples</td>
<td>Ayed et al. (2012); Galván et al. (2013)</td>
</tr>
<tr>
<td><strong>Viruses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noroviruses</td>
<td>Gastroenteritis</td>
<td>Undetectable—10⁸ GC</td>
<td>Miura et al. (2015); Pérez-Saut et al. (2012)</td>
</tr>
<tr>
<td>Rotaviruses</td>
<td>Gastroenteritis</td>
<td>Undetectable—10⁸ GC</td>
<td>Barnil et al. (2015); Prado et al. (2011)</td>
</tr>
<tr>
<td>Adenoviruses</td>
<td>Gastroenteritis</td>
<td>Undetectable—10⁸ GC</td>
<td>Bofil-Mas et al. (2006); La Rosa et al. (2010); Williams and Hurst (1988)</td>
</tr>
<tr>
<td><em>Astroviruses</em></td>
<td>Gastroenteritis</td>
<td>10⁻⁵—10⁻⁴ GC</td>
<td>El-Senousy et al. (2007); Hata et al. (2015)</td>
</tr>
<tr>
<td>Enteroviruses</td>
<td>Wide range of diseases</td>
<td>Undetectable—10⁶ FFU†</td>
<td>Costán-Longo et al. (2008); Shulman et al. (2014)</td>
</tr>
<tr>
<td>Hepatitis A virus</td>
<td>Hepatitis</td>
<td>Undetectable—5 × 10⁶ FFU†</td>
<td>Pintó et al. (2007); Villar et al. (2007)</td>
</tr>
<tr>
<td>Hepatitis E virus*</td>
<td>Hepatitis</td>
<td>Undetectable—5 × 10⁶ FFU†</td>
<td>Masclaux et al. (2013)</td>
</tr>
</tbody>
</table>

Those agents marked with ‘**’ are zoonotic; ‘†’ literature refers to concentrations in raw sewage; ‘‡’ Fluorescent foci units.

Most infections are seasonal, and consequently the occurrence of circulating pathogens varies throughout the year (Altizer et al. 2006; Fisman 2007). One very illustrative example is *Campylobacter*, which is ubiquitous worldwide and seasonal according to the climate of the area (Kovats et al. 2005). Consequently, its abundance in sewage is very variable. *Campylobacter* concentrations in sewage in Europe approach 10⁵ MPN 100 ml⁻¹ in the peak season but are less than 10³ MPN 100 ml⁻¹ in the low-incidence season (Jones 2001).

So, because of seasonality and the major geographical differences in prevalence, the presence of pathogens in sewage is highly unpredictable, and depends on the geographical location and the time of year. This, added to the difficulty and uncertainty in the enumeration methods, most likely explains the great disparity of values found in the literature. Additionally, reliable data in low-income countries are very scarce.

### Faecal indicators in sewage

Faecal indicators are members of the microbiota of the gut of humans and animals, which are widely distributed in human and animal faeces. Therefore, they are indicators with an index function of nonspecific faecal pollution detection.

#### Faecal indicator bacteria

Since the first quarter of the 20th century, FIB have been used to assess both water quality and the performance of water treatments, and they have been included in water quality regulations in countries all over the world. FIB are a heterogeneous group of bacterial taxa defined by phenotypic characteristics that allow for their selective detection and quantification. They are total coliforms, which are a heterogeneous group of bacterial taxa defined by the presence of thermostable (also known as faecal) coliforms, *Escherichia coli*, enterococci (also known as faecal streptococci) and spores of sulphite-reducing clostridia (Fujio 2002; NRC.
2004; WHO 2006). Even though they have been assumed not to replicate outside the gut, it has recently been reported that even the most restrictive, such as *E. coli* and enterococci, can replicate in a few natural niches when nutrients are available, the temperature is adequate (Ishii and Sadowsky 2008; Staley et al. 2014) and competitors with better fitness in these niches are absent. However, these conditions will not apply to the vast majority of situations where FIB are used as indicators.

Methods to detect FIB are operational procedures. The presence/absence and quantitative methods standardized by regulatory agencies (International Standard Organization (ISO), US Environmental Protection Agency (USEPA), European Committee for Standardization (CEN)), national agencies and associations (i.e. Standard Methods), as well as equivalent methods developed by private companies as user-friendly kits (i.e. Aquagenx, Colilert, Colifast, Enterolert, Hach Kit, Readycult, etc.) are available worldwide (WHO 2001; Fujioka 2002; NRC 2004; Harwood et al. 2018). Most of these methods are culture dependent. However, methods based on qPCR are emerging for *E. coli*, enterococcus and clostridia (Wang et al. 2007; Silkie and Nelson 2009; Srinivasan et al. 2011). All these molecular methods have the handicap that they do not distinguish dead from alive organisms, thus hindering their use as indicators for many purposes.

*Escherichia coli* and species of enterococci and clostridia include a variety of different clonal lineages that are disseminated worldwide (Collins et al. 1994; Tenaillon et al. 2010; Staley et al. 2014). Those residing in the gut of a given person or animal can vary over time. Consequently, they vary within a population over time and between different populations in space. However, with few exceptions, the great majority of these lineages are detected by the standardized methods and their concentrations in sewage are steady over time and geographical location. The faecal load of a given sewage (production of wastewater per person), effects of rain on water run-off and sewer sediment suspension, time of sampling, etc., shape the concentrations of faecal indicators in sewage. The typical concentrations of the indicators reported worldwide are summarized in Table 2. Faecal coliforms/*E. coli* is the most abundant, followed by enterococci and spores of sulphite-reducing clostridia, with concentrations roughly one and two log₁₀ units lower respectively.

Other than these numbers, there are some facts that in our opinion need to be mentioned about FIB. Firstly, their concentrations in a given sewage show no seasonality (Muniesa et al. 2012); and secondly, their relative concentrations do not display geographical differences (Lucena et al. 2004).

Their similar resistance to treatments and persistence in the environment to those of bacterial pathogens (WHO 2001) has led to their use as indicators for water management and quality control. However, their value as surrogate indicators of viruses and parasites has been questioned (AWPRC Study Group on Health Related Water Microbiology 1991; Grabow 2001; Staley et al. 2014) and efforts have been made over the last few decades to make faecal indicators available that better reflect the behaviour of viruses and parasites.

Viral indicators
Bacteriophages that infect enteric bacteria have been proposed as indicators of faecal pollution and/or viruses, and are gaining ground to such an extent that coliphages are being included in some water quality guidelines (Jofre et al. 2016; McMinn et al. 2017).

For practical purposes, similarly to FIB, faecal indicator bacteriophages are organized into operational groups rather than in taxa, mostly because of the current methods for their detection and quantification. Somatic, F-specific and the subset of F-specific RNA coliphages (Jofre et al. 2016; McMinn et al. 2017) have been extensively studied for this purpose. Humans and animals excrete both groups. Somatic coliphages are those that infect *E. coli* through the cell wall. F-specific coliphages, also named sexual coliphages, are those that infect *E. coli* through the sexual pili; while F-specific RNA phages can be further differentiated by adding RNAase to the detection medium. The sum of both somatic and F-specific coliphages constitutes the total coliphage group. Host strains have been developed to count total coliphages (Guzmán and Moc-Llivina 2008).

Feasible and cost-effective presence/absence and quantitative (double-layer plaque assay) methods standardized by regulatory agencies (ISO, USEPA) and associations (i.e. Standard Methods) are available (Jofre et al. 2016; McMinn et al. 2017; Harwood et al. 2018). These methods are mostly characterized by the use of certain bacterial host strains. Recently, some fast and friendly methods that can be adapted to ready-to-use kits have become available (Salter et al. 2010; Muniesa et al. 2018). Molecular methods based on counting phage particles are on hand for some specific bacteriophages, mostly F-specific phages, but not for the entire operational groups (Wolf et al. 2010).

Concentrations of somatic and F-specific coliphages in untreated sewage from different areas of the planet are available, as in the case of FIB, their comparative concentrations are similar around the world (Table 2). Somatic coliphages are the most abundant, with values ranging from $5 \times 10^5$ to $5 \times 10^8$ in 100 ml. In raw sewage samples in which all the parameters have been tested, *E. coli*/faecal coliform numbers exceeds the numbers of somatic coliphages by less than one log₁₀ unit; and somatic
coliphages surpass the quantities of F-specific coliphages by roughly 0.5–1.0 log_{10} units (Table 2).

As occurs with FIB, concentrations of both groups of coliphages in raw sewage show no seasonal variations (Muniesa et al. 2012; Dias et al. 2017) or differences in geographical distribution (Lucena et al. 2003, 2004). They are also excreted by animals and their concentrations and ratios in animal sources of faecal pollutions (abattoir wastewater, slurries from farms, etc.) are similar to those in sewage (Blanch et al. 2006).

**Indicators of parasites**

Helminth eggs have been used as an indicator of parasites for some purposes, mostly in low-income countries, where, as reported in the previous section, they are still quite prevalent, with values ranging from <10 to 10^5 l^{-1} depending on the region, weather condition and season (WHO 2006).

**Human-specific microbial source-tracking markers in sewage**

The ubiquitousness of general faecal indicators in animal faecal sources hinders applications that require determination of the origin of the faecal contamination, such as the management of surface waters. Identification of the faecal sources through microbes, known as microbial source tracking (MST), offers a diverse set of markers and methods to identify human and other faecal contamination sources.

Since the 1980s, great effort has been put into MST, and a great number of approaches have been assayed for identifying mostly human, ruminant, porcine and poultry sources. This review focuses exclusively on those approaches for the determination of human faecal sources based on library-independent methods. This means that there is no need for a previous reference database in each region where they are applied. If, in the section on general faecal indicators, having living indicators is considered paramount, in the case of human-specific MST markers in sewage this characteristic is not particularly relevant. Having a marker of general faecal contamination that shares similar resistance to water treatments and environmental persistence together with an appropriate human-specific MST marker or markers is paramount when apportioning human and nonhuman faecal contamination in a given water body.

At present, the human-specific MST markers being used are either pathogens or commensal micro-organisms from the human gut. However, the use of pathogens is conditioned by the epidemiological situation and by pathogen prevalence in each region and seasonality (Fong et al. 2005). The human-specific MST markers that have currently been studied with the methods employed for their detection and reported values are summarized in Table 1.

Among pathogens, some human viruses have been proposed as MST markers of human faecal contamination. Human adenovirus and polyomavirus are generally present in sewage and show high water environmental persistence and significant host specificity (Ahmed and Harwood 2017). Values (Table 1) range from undetectable to 10^6 GC 100 ml^{-1}. The main problems with these markers are the variability in the prevalence associated with pathogens, due to seasonality and geographical disparity (Fong et al. 2005).

Among the commensal microbes, bacteria and bacteriophages are at present being studied. Regarding bacteria, genome fragments of commensal bacteria of the human microbiota detected by qPCR are used, particularly genome fragments of Bacteroidales, Lachnospiraceae, Bifidobacterium, Brevibacterium, Catellicoccus, Enterococcus, Helicobacter, methanogens, Rhodococcus and Streptococcus. Most methods based on qPCR target the 16S rRNA gene of these bacteria. However, nonribosomal genes have also been used, mainly targeting genes involved in bacterium–host interactions (Yampara-Iquise et al. 2008; Lee and Lee 2010; Shanks et al. 2010; Tambalo et al. 2012; Eren et al. 2015). Reported values (Table 1) in sewage are high enough to be detected after the dilution of the sewage in water bodies. Methods that apply these molecular markers show differences in sensitivity and specificity. The most sensitive show some cross-identification, being found in faecal contamination from nonhuman species. Moreover, geographical variation in the efficacy of MST markers has also been reported mainly due to host genetics, antibiotic usage, immunological factors and dietary effects (Reischer et al. 2013; Yahya et al. 2017). Consequently, the validation of the selected markers in a given region is essential prior to any MST study.

Two main groups of bacteriophages infecting bacteria of the human colon microbiota have been studied (Table 1). The concentrations of both groups allow them to be detected after dilution. First, one is the distribution of geno-groups of F-specific RNA bacteriophages, of which there are four: geno-groups II and III are mostly of human origin; whereas I and IV prevail in nonhuman faecal remnants (Jofre and Grabow 2011). Well-defined two-step procedures based on plaque formation by the double-layer agar method followed by identifying plaques either by plaque hybridization (Hsu et al. 1995) or by RT-PCR amplification of the phages in the plaques (Ogorzaly and Gantzer 2006) are available. Direct determination of phages in suspension by RT-qPCR is also
possible (Wolf et al. 2010). The main problems with this marker are that there are some exceptions regarding the origin (Schaper et al. 2002) and that the different subgroups differ in resistance to treatments and persistence in water (Muniesa et al. 2009). The second group is that of bacteriophages infecting some particular Bacteroides strains (Jofre et al. 2014) for which a well-standardized method exists (ISO 2001). They show no seasonality, although geographical differences of concentrations in sewage have been reported (Jofre et al. 2014). Recently, a novel bacteriophage predicted by computational analysis to be highly abundant in human faeces has been used in combination with E. coli to differentiate human from animal faecal contamination in sewage samples (García-Aljaro et al. 2017).

All together, this makes it difficult at present to consider that a single marker will resolve the problem of identifying the origin of faecal contamination in each water body; and that the most advisable approach is the use is a combination (tool box) of markers which might increase the rate of success at identifying the main faecal pollution sources in water. Selection of the most convenient markers is a preliminary analysis, which has been suggested to improve decision-taking when developing MST using predictive models (Blanch et al. 2006; Ballesté et al. 2010; Gourmelon et al. 2010). However, other factors affect the combined use of markers when applied in modelling, such as the dilution of markers in the water body, their environmental inactivation and the complexity of several pollution sources (Casanovas-Massana et al. 2015). So, the environmental persistence and water treatment resistance of any applied MST marker needs to be previously evaluated for management strategies (Bae and Wurts 2009; Ballesté and Blanch 2010; Green et al. 2011; Brooks and Field 2017). Then, the integration of inactivation and dilution factors for each marker used is critical to develop predictive models that are suitable for the complexity of ecosystems and water flows. Several computational techniques have been assayed to define predictive models to assess the faecal pollution sources in water. Some mathematical approaches (Ichnaea®) are based on empirical data concerning the phenomena modelled and considering the effects of dilution of the pollution event and the ageing of the selected MST markers once they reach the environment. These provide an initial selection of the most reliable combination of MST markers and later allow us to recognize, classify and predict patterns from existing data (Sánchez et al. 2011). Other computational approaches (SourceTracker) rely on a Bayesian approach using results obtained in marker gene and functional metagenomics studies to identifying sources and proportions of contamination (Knights et al. 2011).

Conclusions

Identifying, quantifying and apportioning the microbial faecal contaminants carried by sewage are essential to assess the impact of sewage discharge into water bodies, to trace the contaminating source and to evaluate the operation of wastewater treatments.

Raw sewage microbiota reflects human gut microbiota, which is dominated by bacteria and viruses. Anaerobic Firmicutes and Bacteroidetes are the predominant bacteria in the human gut microbiota, whereas aerobic and facultative Proteobacteria are far less abundant. Pathogens are even less abundant, and their concentrations fluctuate both seasonally and geographically.

On the way from houses to water bodies or wastewater treatment plants, sewage microbiota incorporates microbes that reside in the sewers as well as microbes linked to surface run-off. Both of these sources provide mostly aerobic and facultative Proteobacteria.

Despite the microbes growing in biological treatment plants that depend on the characteristics of the treatment and the composition of the influent sewage, and which may become dominant, and the decrease in the concentrations of incoming microbes, treatment plant effluents are still a continuous source of urban-derived microbes, including pathogens, which could constitute a serious public health problem in some cases.

In spite of the countless, last-generation molecular methods implemented to detect pathogens, the complexity of sample matrices and the sampling, DNA preparation, processing, sequencing and analytical protocols, make the study of water samples such as sewage intricate and barely feasible for practical and routine purposes.

Changes in the dominant bacterial groups during sewage transport and treatment detected by metagenomic studies still make this approach inappropriate for water management practices. Moreover, pathogen concentrations fluctuate seasonally and geographically, which could become more accentuated as a consequence of climate change in the near future. The distribution and concentration of the current faecal indicators, to the best of our knowledge, is not expected to fluctuate.

Consequently, traditional indicator bacteria are still the preferred choice to assess water quality and water treatments operation. Since the traditional bacterial indicators do not reflect what occurs with some pathogens, particularly viruses, and do not distinguish between human and animal faecal pollution, viral indicators (i.e. bacteriophages) and human-specific MST markers are emerging as tools for water quality management.
Acknowledgements

This work was supported by the Spanish Government (research project AGL2016-75536-P), and the Catalan Government (Research group 2017SGR170 and Xarxa de Referència en Biotecnologia).

Conflict of Interest

The authors have no conflict of interest to declare.

References


Pathogens, faecal indicators and source markers in sewage


