Biological Indicators for Fecal Pollution Detection and Source Tracking: A Review

Enze Li, Faizan Saleem, Thomas A. Edge and Herb E. Schellhorn *

Process of Biology, McMaster University, Hamilton, ON L8S 4L8, Canada; lie40@mcmaster.ca (E.L.);
saleef4@mcmaster.ca (F.S.); edget2@mcmaster.ca (T.A.E.)
* Correspondence: schell@mcmaster.ca

Abstract: Fecal pollution, commonly detected in untreated or less treated sewage, is associated with health risks (e.g., waterborne diseases and antibiotic resistance dissemination), ecological issues (e.g., release of harmful gases in fecal sludge composting, proliferative bacterial/algal growth due to high nutrient loads) and economy losses (e.g., reduced aqua farm harvesting). Therefore, the discharge of untreated domestic sewage to the environment and its agricultural reuse are growing concerns. The goals of fecal pollution detection include fecal waste source tracking and identifying the presence of pathogens, therefore assessing potential health risks. This review summarizes available biological fecal indicators focusing on host specificity, degree of association with fecal pollution, environmental persistence, and quantification methods in fecal pollution assessment. The development of practical tools is a crucial requirement for the implementation of mitigation strategies that may help confine the types of host-specific pathogens and determine the source control point, such as sourcing fecal wastes from point sources and nonpoint sources. Emerging multidisciplinary bacterial enumeration platforms are also discussed, including individual working mechanisms, applications, advantages, and limitations.

Keywords: fecal pollution; fecal source tracking; fecal indicator; water disinfection; bioengineering; bacterial enumeration

1. Introduction

Fecal pollution may cause public health and ecological issues, as it may be associated with the release of pathogens, antibiotic-resistant organisms, high nutrient loads, and the emission of harmful gases [1–4]. Commonly affected areas include soil, sediment, farmland, wastewater treatment plants (WWTPs), and many water bodies such as drinking water sources, recreational water, and groundwater [2,4–7]. Fecal pollution can occur from untreated sewage. In developing and undeveloped countries, untreated sewage is sometimes discharged directly into surface water bodies [8,9]. Polluted water may infiltrate into the ground, rendering groundwater no longer safe for drinking and agricultural purposes [8–10]. According to the WHO, at least 2 billion people worldwide used an unsecured drinking water source polluted with feces in 2017 [11], causing diseases like diarrhea, typhoid, cholera, and dysentery. In 2016, diarrheal diseases ranked as the 8th leading cause of mortality among all ages (more than 1.6 million deaths) and the 5th leading cause of death among children under 5 years old (around 0.45 million deaths, mainly in South Asia and Africa) across 195 countries [12].

Pathogens of different types and infectious doses may differ based on their existing environments, exposure conditions, host susceptibility, and seasonality [13–15]. For example, swimmers exposed to fecal-polluted recreational waters usually develop mild to no symptoms. In contrast, people who cook with fecal-polluted drinking water tend to have acute and severe symptoms. Children under 5 years old and seniors over 70 years old are statistically more vulnerable to waterborne diseases due to reduced immunity [11,16]. Viral pathogens and protozoan parasites may have relatively lower infectious doses than
bacterial pathogens, making them a more serious threat [17]. After seasonal rainfalls, short periods of elevated fecal pollution are frequently reported, suggesting a seasonality of waterborne disease outbreaks [14,15]. Table 1 summarizes recent epidemics of waterborne and foodborne diseases that occurred within the last decade. These diseases, however, often have mild or even no early symptoms. Even when symptoms are severe, it remains challenging to attribute diseases to specific fecal pollution exposure [3]. These frequently reported outbreaks signify the importance of the need for fecal pollution monitoring.

Table 1. Examples of outbreaks of waterborne and foodborne diseases from 2011 to 2021.

<table>
<thead>
<tr>
<th>Leading Pathogen</th>
<th>Diseases</th>
<th>Year</th>
<th>Country</th>
<th>No. of Cases</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Salmonella</em> spp.</td>
<td>Salmonellosis</td>
<td>2010</td>
<td>China</td>
<td>324</td>
<td>[18]</td>
</tr>
<tr>
<td></td>
<td>Salmonellosis</td>
<td>2010</td>
<td>USA</td>
<td>695</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td>Gastroenteritis</td>
<td>2013</td>
<td>USA</td>
<td>634</td>
<td>[20]</td>
</tr>
<tr>
<td><em>Cryptosporidium</em></td>
<td>Cryptosporidiosis</td>
<td>2010</td>
<td>Canada</td>
<td>12</td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>Cryptosporidiosis</td>
<td>2011</td>
<td>Australia</td>
<td>1141</td>
<td>[22]</td>
</tr>
<tr>
<td><em>Shigella sonnei</em></td>
<td>Shigellosis</td>
<td>2012</td>
<td>China</td>
<td>134</td>
<td>[23]</td>
</tr>
<tr>
<td></td>
<td>Shigellosis</td>
<td>2012</td>
<td>Turkey</td>
<td>4239</td>
<td>[24]</td>
</tr>
<tr>
<td><em>E. coli O157:H7</em></td>
<td>Diarrhea/Vomiting</td>
<td>2012</td>
<td>Canada</td>
<td>5</td>
<td>[25]</td>
</tr>
<tr>
<td></td>
<td>Diarrhea/Vomiting</td>
<td>2012</td>
<td>Japan</td>
<td>115</td>
<td>[26]</td>
</tr>
<tr>
<td><em>Campylobacter</em></td>
<td>Campylobacteriosis</td>
<td>2016</td>
<td>UK</td>
<td>69</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>Campylobacteriosis</td>
<td>2016</td>
<td>New Zealand</td>
<td>967</td>
<td>[28]</td>
</tr>
<tr>
<td><em>Hepatitis A virus</em></td>
<td>Hepatitis A</td>
<td>2016</td>
<td>USA</td>
<td>292</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td>Jaundice</td>
<td>2016</td>
<td>India</td>
<td>233</td>
<td>[30]</td>
</tr>
<tr>
<td><em>Rotavirus G2P</em></td>
<td>Gastroenteritis</td>
<td>2017</td>
<td>Germany</td>
<td>32</td>
<td>[16]</td>
</tr>
<tr>
<td><em>Rotavirus G3P</em></td>
<td>Gastroenteritis</td>
<td>2017</td>
<td>Australia</td>
<td>2319</td>
<td>[31]</td>
</tr>
<tr>
<td><em>Adenovirus</em></td>
<td>Respiratory illness</td>
<td>2018</td>
<td>USA</td>
<td>168</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td>Fever</td>
<td>2019</td>
<td>China</td>
<td>27</td>
<td>[33]</td>
</tr>
<tr>
<td><em>Norovirus</em></td>
<td>Diarrhea/Vomiting</td>
<td>2018</td>
<td>South Korea</td>
<td>199</td>
<td>[34]</td>
</tr>
<tr>
<td></td>
<td>Gastroenteritis</td>
<td>2020</td>
<td>Brazil</td>
<td>176</td>
<td>[35]</td>
</tr>
</tbody>
</table>

Another health concern is the undesired dissemination of antibiotic-resistant bacteria (ARB) and antibiotic resistance genes (ARGs) to the environment. Antibiotic residues reach the sewage systems via feces, urine, and body remains [36], which explains why WWTPs are often a hotspot for mixing antibiotics, ARB and ARGs and for their spread into the receiving environment [36]. Elevated ARB and ARGs are frequently measured in areas with increased exposure to municipal wastewater [2,37–39], resulting in a correlation between fecal pollution and antibiotic resistance dissemination. Although antibiotic concentrations in sewage are much lower than their therapeutic dosage in humans or animals, they may still be high enough to affect susceptible environmental bacteria and pose a selective pressure for the evolution of resistant bacteria [2,40]. Multidrug-resistant bacterial pathogens may cause difficult-to-treat or even untreatable infections because conventional antibiotics may not be effective. Moreover, ARGs can spread among non-resistant bacteria by horizontal gene transfer.

Ecological issues associated with fecal pollution are noticeable as well. Fecal wastes are rich in organic matters and nutrients such as ammonium, sulfate, phosphate and potassium [41]. Therefore, their discharge into the environment can be seen as nutrient loads, which can alleviate nutrient limitation and stimulate nutrient recycling, but sometimes causes eutrophication and the emission of harmful gases such as NH₃, H₂S, N₂O and CH₄ [42,43]. Birds can also heavily impact water systems. For example, cormorant defecation in shallow-water ecosystems can lead to algal bloom formation [44]. Defecation by a small cormorant population may not have a significant impact, but a group living in large colonies makes their feces quantitatively relevant as a nutrient input. Similarly, fish can also contribute to the hyper-bloom of cyanobacteria through intensive consumption.
of herbivorous zooplankton, although some fish such as tilapia can digest cyanobacteria completely [45].

Fecal sludge is widely used as a fertilizer with or without treatment in agriculture due to its high nutrient loads. Fecal sludge application may increase soil fertility, especially regarding carbon retention, but retention of bioavailable phosphorus and nitrogen can be low [46]. Furthermore, untreated fecal sludge poses risks to the crops because of pathogens, heavy metals, and toxic levels of ammonia and nitrite; additionally, the intensive fermentation activities within untreated fecal sludge can result in hypoxia in the rhizosphere [47,48]. To avoid these risks, treatment of fecal sludge or wastewater is required before its agricultural reuse. Composting, the aerobic and thermophilic mineralization of organic matters, is a standard fecal sludge treatment [49]. It effectively eliminates enteric pathogens and parasite eggs due to increased internal temperatures within the fecal sludge. Moreover, mineralized nutrients have improved plant uptake. However, the emission of biogas contaminants produced during composting, including sulfide, ammonia and greenhouse gases (e.g., methane, carbon dioxide and nitrous oxides), is inevitable [41,43,50]. Meanwhile, manure compost can potentially disseminate AGRs and ARB from livestock to soil/crops and finally to humans [51].

This review summarizes five fecal indicator bacteria and four fecal indicator viruses, focusing on their host specificity, degree of association with fecal pollution, environmental persistence, and quantification methods in fecal pollution assessment, together with individual enumeration methods (Figure 1), aiming to serve as a resource for researchers and engineers to evaluate indicators and methodologies for a specific application.

![Figure 1. Relationship between the use of selected microbial markers, potential monitoring sites, and diagnostic information by process operators and regulators. Use of these biological markers can facilitate actionable remedial measures by public health officials.](image)

### 2. Fecal Pollution Detection Using Biological Markers

Fecal pollutants causing water quality problems may come from point sources and/or nonpoint sources. Point source pollution is fecal pollution events resulting from a single identifiable source, such as discharge from a WWTP or a leak from a septic tank. In contrast, nonpoint source pollution results from diffuse sources, such as runoff from residential areas and farmlands. Fecal waste from diffuse sources can be a persistent and cryptic input to the environment that may be present at low concentrations and difficult to trace compared to fecal wastes from point sources because diffuse sources can be a mixture of...
fresh and aged fecal wastes, and therefore require more complicated methods for source tracking [52]. According to the United States Environmental Protection Agency (U.S. EPA), fecal wastes from diffuse sources have become the leading cause of water quality problems in the U.S. [53]. Accurate identification of the source not only helps confine the types of host-specific pathogens but also determines the source control point and remediation strategies. There are several available indicators for fecal pollution detection or fecal source tracking purposes. Here we provide updates on five bacterial and four viral indicators recommended by the WHO, U.S. EPA and other studies.

2.1. Bacterial Markers

One major purpose of fecal pollution detection is to identify the presence of pathogens to identify the fecal waste source and potential health risks. With the realization that waterborne transmission of bacterial diseases was a public health concern in the 18th century, governments established sanitary systems to monitor drinking water safety and build water treatment facilities [54]. A question arises—should one monitor waterborne pathogens directly or employ proxy indicators? There are many bacterial, protozoan, and viral enteric pathogens that can cause diseases. Their methods of detection are often laborious, time-consuming, and require specialized expertise and biosafe laboratory settings. Pathogens in environmental water are sporadic and sparse, making it difficult or costly to measure these pathogens [3]. In addition, waterborne pathogens in the environment are less dangerous at low concentrations compared to those in GI tracts, but their laboratory enumeration often involves concentration from large volumes of water samples; therefore, these concentrated pathogens may cause severe human diseases during their measurement [1,13]. Another reason we should monitor fecal indicators is that there may be unknown waterborne pathogens that will be overlooked if we only monitor known pathogens. Therefore, detection of fecal pollution-associated pathogens is a multifaceted challenge.

Water quality monitoring to detect fecal pollution can employ biological fecal indicators to proxy many potential pathogens. A typical fecal indicator should fulfill the following criteria [1,3,55]:

1. It should be detected only in feces-polluted environments and should not be detected in non-fecal-polluted environments; i.e., it should have a high association with fecal pollution with little or no background noise;
2. It should be detectable in the same host as the pathogens, and its abundance should reflect pathogen abundance both inside and outside the host; i.e., it should have tight correlations with pathogens/diseases;
3. It should be detectable in the host intestine, feces, and also fecal-polluted environmental waters so that it can be assessed using inexpensive and simple methods; i.e., it should have good detection sensitivity;
4. It should not multiply outside the host intestine but should be viable or in a detectable form for some time comparable to fecal pathogens;
5. It should be safe to measure for field and lab workers, not causing severe human diseases.

The value of common fecal indicator bacteria (FIB) is that they occur widely in the feces of birds, mammals, and humans, posing a robust way to detect fecal pollution. However, lack of host specificity is also a problem when determining the specific source of fecal pollution, i.e., fecal pollution source tracking. Advances in the field of microbial source tracking have now provided new fecal indicators that can be used to identify specific fecal pollution sources (e.g., human sewage). If a proxy is used to detect human fecal pollution, ideally it should be exclusively of human sources and not be found in animals. Here we introduce well-validated bacterial, viral and molecular indicators, emphasizing their association with pathogens, their presence in fecal wastes, their host specificity, their persistence in the environment, and their pathogenicity.

Host specificity is one of the most critical criteria for any fecal indicator because it determines whether a given fecal indicator can be used to identify the specific sources of
fecal pollution, i.e., fecal source tracking [56]. Accurate source tracking helps water quality regulators to locate pollution sources, such as fecal pollution from a nearby WWTP or an animal farm. Host specificity can be readily assessed by the proportion of the false-positive results detected in nontarget fecal pollution sources [57]. Fecal indicators with absolute human specificity must be detected only in humans and must not be detectable in any animals, whereas those of poor human specificity may also be detected in animals, plants, or environments without association with fecal pollution. Livestock in a farm/slaughterhouse [58,59] and wild animals in forests/natural reserves [60] can also be sources of fecal pollution and pathogens. Thus, fecal indicators specific to one animal host are worth noticing because they help identify the species of animal hosts and the possible source locations.

Sensitivity is another critical criterion which is evaluated by the positive proportions in target fecal pollution sources. Fecal indicators of perfect sensitivity should be detectable in all target fecal pollution sources, whereas those of poor sensitivity may be detectable in a minor proportion of target fecal pollution sources in situations where pollution is high. Sensitivity measurement is, of course, affected by fecal indicator concentrations. Even fecal indicators of high sensitivity may be undetectable in target samples if they are diluted or have undergone significant losses through sample processing steps, causing false negatives.

2.1.1. Coliforms

Coliform bacteria, or total coliforms or simply coliforms, are a group of Gram-negative, facultatively anaerobic, non-spore-forming and rod-shaped bacteria that can ferment lactose to acid and gas at \( \sim 37 \, ^{\circ}C \) within 48 h, constituting \( \sim 10\% \) of gut microorganisms [61]. Traditionally, total coliforms included four closely related genera within the family Enterobacteriaceae: *Escherichia*, *Enterobacter*, *Klebsiella* and *Citrobacter* [62]. But today, some species of over 19 bacterial genera (e.g., *Hafnia*, *Serratia* and *Raoultella*) can be classified as coliforms based on the phenotypic characteristics above [63]. Total coliforms are used to evaluate water treatment efficacy in fecal wastes removal and assess the integrity of water distribution systems [64] because they mainly originate from the GI tracts of humans and warm-blooded animals (i.e., mammals and birds), and thus, their presence in water can indicate possible fecal pollution. In 1975, the U.S. EPA issued the National Primary Drinking Water Regulations, which allowed a maximum of 5% drinking water samples positive for total coliforms tested in a month [65]. Then, in 1989, the U.S. EPA published the first Total Coliform Rule, which supplemented a Maximum Contaminant Level Goal of zero total coliforms in nationwide drinking waters [66]. To date, total coliforms are still routinely monitored in drinking water.

The recognition of free-living coliforms in non-enteric environments, such as freshwater, soil and vegetation, that do not have a history associated with fecal pollution [67,68] has cast doubt on the exclusive association between total coliforms and fecal pollution. For example, *Klebsiella varicola* is a member of total coliforms but is typically a plant endosymbiont [68]. Thus, not all members of total coliforms are suitable indicators of fecal pollution, and presence of total coliforms in drinking water does not necessarily indicate fecal pollution [64]. To overcome this issue, fecal coliforms, which are defined as coliforms originating from the intestines of warm-blooded animals capable of fermenting lactose at 44.5–45.5 \( ^{\circ}C \), are proposed as a monitoring target [69]. As a subgroup of total coliforms, fecal coliforms have better associations with fecal pollution because the non-fecal-origin part is excluded [70]. Fecal coliforms detected in aquafarm harvesting areas such as an oyster bed are indicative of fecal pollution, causing ecological issues and economic losses [71]. Note that neither total coliform nor fecal coliform is a natural taxon but a method-based bacteria category for convenience. We now understand that even within the category of fecal coliforms, some are not exclusively of fecal origin, hence the transition to *Escherichia coli* as a better fecal indicator. Among fecal coliforms, *E. coli* is the dominant species isolated.
from water (>95%) and is the most frequently used indicator [72]. Figure 2 describes the relationships among total coliforms, fecal coliforms and E. coli.

![Diagram showing relationships of coliforms and Escherichia coli](image)

**Figure 2. Relationships of coliforms and Escherichia coli.**

Since coliforms can be cultured with ease, there are culture-based and culture-independent methods to detect and enumerate coliforms in water. Frequently used culture-based enumeration methods include the (i) classical plate count on violet-red-bile-lactose agar medium [73] or chromogenic media based on the activities of β-D-galactosidase (of coliforms) and β-glucuronidase (of E. coli) [74], (ii) most probable number procedure in lauryl tryptose broth [75] and IDEXX systems, (iii) membrane filtration in selective M-Endo medium [76] and (iv) lateral-flow immune-chromatographic assay using antibodies recognizing coliform genera [77]. Culture-independent methods include (i) real-time PCR targeting the lacZ gene of coliforms [78,79], (ii) fluorescence-based DNA microarray probing E. coli [80], (iii) flow cytometry [81] and (iv) next-generation sequencing or metagenomics that profile all microorganisms, including total/fecal coliforms [82].

Except for a few strains such as E. coli O157:H7, fecal coliforms are generally not considered to be pathogenic. However, drinking water free of coliforms or not exceeding coliform regulatory thresholds is not necessarily safe to drink. Seven waterborne disease outbreaks, affecting more than 3000 people across 5 American states between 1991 and 1992, occurred in drinking water not exceeding the 1989 Maximum Contaminant Levels [83]. Protozoal parasites Giardia and Cryptosporidium were identified as the etiological agents. Therefore, coliforms should not be the sole indicator for successful drinking water quality assessment.

2.1.2. *Escherichia coli*

*Escherichia coli* is used as a fecal indicator as part of drinking water regulations [54]. The U.S. EPA added *E. coli* to the water monitoring and assessment list in 1985 [84]. According to an early investigation by the U.S. EPA, *E. coli* has the highest correlation to swimming-associated diseases at freshwater beaches on the east coast of the U.S. [85]. First, it shows a strong association with fecal pollution, as it is found in all feces of warm-blooded animals and a few reptiles at high concentrations [54,86], and remains at detectable levels in waters subject to animal fecal pollution, such as sewage, rivers, lakes, groundwater and coastal areas [87,88]. In addition, *E. coli* is a facultative anaerobe and is easy to culture, so it can be readily measured using both culture-based and culture-independent methods. Second, *E. coli* is highly associated with a fecal origin and is less likely to be present in natural environments free of animal feces. It was believed to be of exclusively fecal origin when selected as a fecal indicator, which, however, is no longer accurate with the recognition of “naturalized” *E. coli*, which persists and may replicate in non-enteric/non-fecal environments such as soil, sediment, beaches and aquatic plants [89,90]. “Naturalized”
E. coli may not come from a recent fecal input, though it is still believed to have diverged from E. coli of fecal origins.

It is important to distinguish between fecal pollution caused by human wastes and that caused by animals, because human fecal pollution is generally associated with high health risks from human-specific pathogens and high loads of antibiotics and ARBs. One major limitation of E. coli as a fecal indicator lies in its poor specificity to human feces. Despite the substantial genetic diversity of E. coli fecal isolates, epidemiological studies have identified weak associations between a particular E. coli clone and a given host [91,92]. For example, a B1 subgroup clone that exhibits the hly virulence gene is unique to animals [91], and a B2 subgroup VIII, O81 serotype clone is specific to humans [93]. E. coli strains showing stringent host specificity may be specialized candidates for fecal source tracking indicators. Still, most E. coli strains discovered in human fecal samples are not specific to humans, causing difficulties in fecal pollution source tracking.

E. coli was initially believed to have poor survival and to be unlikely to multiply after being released into the environment through deposition of fecal wastes because of the environment being so different from animal gastrointestinal tracts and because of environmental stresses such as insufficient organic matter [94], sunlight inactivation and high salinity [95]. However, prolonged E. coli survival can be found at low temperatures [96], and enhanced E. coli replication is seen at higher temperatures, where relatively sufficient nutrients are likely to support its regrowth [97]. Lake sediments rich in organic contents also favor the survival of E. coli at lower temperatures [98]. When grown as biofilms, E. coli has a significantly higher tolerance against chlorine than planktonic E. coli [99]. In conclusion, the persistence of E. coli is greatly affected by environmental conditions.

Changes in climate, such as an increase in temperature and rainfall, may cause fluctuations of E. coli counts in waters [100–102]. A rainfall event can result in up to a three-fold increase of E. coli counts in recreational waters compared to before rainfall, and this increase can persist for up to 12 h [102,103]. After a rainfall event, increases in E. coli counts may occur due to the “first flush” from stormwater drainage or agricultural waste runoff coming along with the flow [103,104]. Furthermore, bacteria including E. coli tend to form flocs, adhere to rocks and other particles, then settle down to the sediment of freshwater ecosystems. A heavy rain or storm event may resuspend these bacterial cells from sediments, leading to a 30–37 times higher bacterial load in water [105,106]. Thus, climate and meteorological parameters should be monitored before and during water sample collection for E. coli testing to help interpret data.

Most strains of E. coli are commensal or symbiotic to their animal hosts in the gastrointestinal tracks [107]. For example, indigenous gut E. coli strains can be highly competitive for proline, and their depletion of proline helps inhibit the colonization of pathogenic E. coli strain O157:H7 [108]. These non-pathogenic strains are considered safe fecal indicators (categorized as biosafety Level I organisms in the U.S.) and can be enumerated by similar methods, as described for coliforms in environmental waters. In addition, there are specialized methods that only detect certain E. coli strains, such as ATCC 47,076, in drinking water [109]. Pathogenic E. coli strains, including the aforementioned O157, a Shiga toxin-producing E. coli (STEC) serotype categorized as a biosafety Level II organism, caused 5441 infections across the United States in 1996 [110]. Our previous study [111] identified a nine base-pair conserved DNA sequence insertion located in the ybiX gene that is unique to O157 strains and can be used for O157 quantification by qPCR. Pathogenic E. coli strains may have better correlations with diseases, and therefore, their enumeration can be a means to assess potential health risks to the public.

2.1.3. Enterococcus

Enterococci share traits with E. coli that are useful for environmental monitoring. First, they both are facultative anaerobic bacteria detected abundantly in the GI tracts of humans and many animals and are shed in feces at high concentrations. Furthermore, both are easy to culture, both have generally low pathogenicity (at Biosafety Level 1, except for
a few species), both have abundances highly associated with GI diseases, and both are subsequently common indicators for water quality monitoring, fecal pollution detection and health risk assessment by culture-based and culture-independent methods [54]. Apart from in humans, enterococci have been detected in insects (e.g., drosophila), domestic animals (e.g., cattle and cats), plants (e.g., forage crops) [112–114], as well as non-enteric environments (e.g., soil and beach sand) [54]. These discoveries indicated a limited host specificity of enterococci and possible origins other than feces.

Both *E. coli* and enterococci have long been used as FIB because of their association with waterborne diseases [115]. Two epidemiological studies recognized enterococci as good indicators of GI diseases among swimmers at both freshwater and marine beaches [116,117]. Enterococci have been the sole recommended FIB by U.S. EPA for marine waters since 1986 [118] because they have higher survival in salt water than coliforms (including *E. coli*) and are thought to be more tightly associated with human pathogens [119,120]. However, using enterococci as fecal indicators has been criticized as well because enterococci are not exclusively of fecal origin and have limited host specificity, as aforementioned. These characteristics may hinder accurate fecal pollution source tracking because non-fecal-origin enterococci may not be associated with health risks, and animal-specific enterococci-associated pathogens may have variable survival strategies in the environment and health risks from human-specific pathogens [121].

The persistence of enterococci in the non-enteric environment is robust, as enterococci can survive across a wide range of temperatures (from 10 to 45 °C) [54], are intrinsically resistant to multiple antibiotics (e.g., ampicillin, cephalosporin and kanamycin) [122], and can survive high salinity (6.5% w/w NaCl) and alkaline (up to pH 10) [123]. *Enterococcus faecium* and *Enterococcus faecalis* are generally the two most abundant species of *Enterococcus* found in human fecal microbiota [124]. They generally have a longer persistence in marine water than *E. coli* [125,126]. They tend to adhere to biotic or abiotic surfaces to form biofilms, protecting them from environmental stresses such as temperature fluctuations, UV radiation, predation, and wave action [127]. Heavy rain or a storm can disrupt these biofilms and release the enterococci into the water column [128]. In environmental waters, adverse conditions, such as solar radiation, starvation, and low temperatures, may decrease enterococci culturability and induce them to enter the VBNC state [129,130]. VBNC cells are unlikely to be detected via culture-based methods.

Epidemiological studies support an association between enterococci from human fecal sources (e.g., urban sewage) and waterborne diseases in recreational waters [116,131,132]. Higher enterococci densities are associated with an increasing occurrence of GI illnesses among beachgoers at beaches impacted by treated sewage effluent [116]. In addition, human pathogens such as *Giardia* and *Salmonella* are found in higher densities and positively correlated with enterococci concentrations in recreational waters impacted by sewage effluent [131,132]. Along with human fecal pollution, exposure to animal feces poses health threats. Globally, one-third of children’s deaths due to GI illnesses in 2015 were caused by pathogens that are found in animal feces [133]. It is predicted that enterococci densities as low as 35 CFUs/100 mL from cattle manure pose health risks [121]. Thus, enterococci levels are indicative of general health risks from both human and animal fecal pollution. However, without associated source tracking information, it is difficult to unambiguously identify the fecal hosts/sources of enterococci.

### 2.1.4. *Clostridium perfringens*

*Clostridium perfringens* is a Gram-positive, spore-forming, biofilm-forming, anaerobic and pathogenic species (at biosafety Level II) associated with feces [134]. It is among the first fecal indicators in aquatic environments [135]. *C. perfringens* can be found in the gut of fish and warm-blooded animals, such as pigs and ducks [136–138], as well as in some non-enteric environments, such as soil, streams and marine sediments [139,140]. Despite having limited host specificity, *C. perfringens* can be present in higher abundances in human
feces than animal feces. As a pathogenic species, its presence itself indicates possible health risks to humans and animals.

Species of the genus *Clostridium* are obligate anaerobes and cannot multiply in environmental waters. Spores of *C. perfringens* are coated by a thick peptidoglycan cortex structure that provides resistance against environmental stresses such as chlorination, heat and UV radiation and resistance against toxins in industrial wastes [141]; thus, *C. perfringens* spores can survive longer in polluted waters than other FIB, especially in freshwaters [141,142].

The inability to multiply in non-enteric environments and the prolonged environmental viability of its spores make *C. perfringens* a better indicator of long-term or accumulative fecal pollution. Criticisms against the usage of *C. perfringens* also focus on its prolonged viability and wide distribution in aquatic environments. Its spores can be detected in places far from the pollution input site, indicative of either remote or old fecal pollution [143]. Besides, its abundance does not always correlate with abundances of other FIB [139]. Hence, combined usage of *C. perfringens* and other FIB may better predict pathogens.

Detection and quantification of *C. perfringens* can be conducted by culture-based and culture-independent methods. It can be isolated from water samples by the membrane filtration and Rapid Fung double tube methods [144], then cultivated in anaerobic chambers at 44–46 °C [145]. Available media include mCP medium, tryptose-sulfite-cycloserine agar, and tryptone-sulfite-neomycin agar [146,147]. Rapid and culture-independent enumeration methods include qPCR targeting the *cpn60* gene [148], DNA sequencing targeting 16S rDNA genes or the whole genome, and the newly developed Saltatory Rolling Circle Amplification targeting the *cpa* gene, which is conducted at a constant temperature and does not require a thermocycler or gel electrophoresis [149].

2.1.5. Bacteroides

Species of genus *Bacteroides* are frequently used fecal indicators for source tracking purposes. They are low-pathogenic indicators (biosafety Level I) with promising host specificity compared to facultative anaerobic FIB [150]. *Bacteroides* are obligate anaerobes of gut origins, unlikely to multiply in a non-enteric environment, and are present at significantly higher densities in the GI tracts and feces of humans than coliforms or enterococci [57]. *Bacteroides* can be found in warm-blooded animal hosts such as humans, dogs, pigs, and also fish, which are cold-blooded [151–153]. Unlike coliforms or enterococci, *Bacteroides* spp. isolated from humans and some animal hosts are unique to one or a group of hosts [154]. Table 2 shows examples of host-specific 16S rDNA sequences of these *Bacteroides* spp. that can be promising molecular markers to identify fecal pollution from humans and animal hosts [155,156] by culture-independent methods, such as qPCR.

**Table 2.** Host-specific 16S rDNA fragments of *Bacteroides* spp. used to discriminate fecal pollution from human and animal hosts.

<table>
<thead>
<tr>
<th>Host</th>
<th>Primer Set</th>
<th>Sequence (5′ → 3′)</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>HF183F/Bac708R</td>
<td>F: ATCATGAGTTCACTATGCGG&lt;br&gt;R: CAATCGGAGTTCTTCGTTG</td>
<td>[157]</td>
</tr>
<tr>
<td></td>
<td>HF134F/HF654R</td>
<td>F: GCCGTCTACTCTTGCC&lt;br&gt;R: CCTGCCTCTACTGTACTC</td>
<td>[157]</td>
</tr>
<tr>
<td></td>
<td>HuBac566F/HuBac 692R</td>
<td>F: GGGTTTAAAGGGAGCGTAGG&lt;br&gt;R: CTACACCACGAATTCGCGCT</td>
<td>[158]</td>
</tr>
<tr>
<td></td>
<td>BacHF/BacHR</td>
<td>F: CTTGGCCAAGCTTTCGAAAG&lt;br&gt;R: CCCCATGTCTTACCCGAAATAC</td>
<td>[159]</td>
</tr>
<tr>
<td>Cattle</td>
<td>CF128F/Bac708R</td>
<td>F: CCAACTTTCCCWWTTACTC&lt;br&gt;R: See above</td>
<td>[157]</td>
</tr>
<tr>
<td></td>
<td>CS406F/Bac581R</td>
<td>F: GAAGGATGAAAGTTCTATGGATTG&lt;br&gt;R: CGCTCCTTTTAACCCCAATAAA</td>
<td>[160]</td>
</tr>
<tr>
<td></td>
<td>CS621F/Bac725R</td>
<td>F: AACCACAGCCCGGATT&lt;br&gt;R: CAATCGGAGTTCTTCGATATC</td>
<td>[160]</td>
</tr>
</tbody>
</table>
Table 2. Cont.

<table>
<thead>
<tr>
<th>Host</th>
<th>Primer Set</th>
<th>Sequence (5′ → 3′)</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swine</td>
<td>Bac41F/PS183R</td>
<td>F: TACAGGCTTAACACATGCAAGTCG R: CTCTACCGTATTTAAATCCGCTTT</td>
<td>[160]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: CGGGTTGAAAATCGGCTTTATAGAAG R: CGCTCCCTTTAAACCCAATAAAA</td>
<td>[160]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: CC GGATTAATACCGTATGA R: See above</td>
<td>[161]</td>
</tr>
<tr>
<td>Dog</td>
<td>BacCan-545F/Bac725R</td>
<td>F: GGAGGCGACAGACGGGTITT R: See above</td>
<td>[155]</td>
</tr>
<tr>
<td>Chicken</td>
<td>C160F-HU/Bac265R-HU</td>
<td>F: AAGGGAGATTAATACCGATGATG R: CCGTTACCCCGCTACTAC</td>
<td>[156]</td>
</tr>
<tr>
<td>Duck</td>
<td>Bac366F-HU/Duck474R-HU</td>
<td>F: TTGGTCAATGCGGCGGAAG R: GCACATCCCAACACGTAGA</td>
<td>[156]</td>
</tr>
</tbody>
</table>

Unlike facultative anaerobic FIB, the isolation and cultivation of Bacteroides cells require anoxic chambers, which is costly, time-consuming and labor-intensive [162]. Bacteroides can grow at 37 °C in many selective media, such as Brain Heart Infusion medium, Bacteroides Bile Esulin agar, and Blood Hemin Vitamin K medium containing antibiotics [162–164]. In most applications, Bacteroides grows into visible colonies on plates within 24 h, and colonies will be large enough to enumerate in 36 and 48 h. Some Bacteroides species can be identified to the species level within four hours by a few biochemical tests, such as the API ZYM system that detects activities of Bacteroides-specific enzymes [165].

The use of biological molecular markers bypasses the difficulties of anaerobic cultivation and enables the routine detection of Bacteroides spp. The aforementioned host-specific 16S rDNA markers can be used to detect and quantify Bacteroides in environmental samples and thus enable source tracking of Bacteroides spp. Among human-specific Bacteroides 16S rDNA sequences, the region to be amplified by the primer set HF183F (Human Fecal Bac183 forward primer) and Bac708R (reverse primer) ranks among the most recognized markers to assess human fecal pollution and to discriminate it from animal fecal pollution because of its high sensitivity and specificity to humans [166]. Although HF183 fluorescent probe-based assays may sometimes cross-react with animal feces, its concentration in human feces can be 1000-fold more, or higher, than in animal feces [167]. In 2012, the U.S. EPA recommended an illness-based recreational water quality threshold of 32 illnesses per 1000 primary contact recreators using enterococci or E. coli [118]. In 2020, as Bacteroides HF183 becomes widely used, a threshold of 525 HF183 copies/100 mL recreational water was proposed by Boehm and Soller [168]. Apart from 16S rDNA sequences, there are other Bacteroides human molecular markers, such as genes encoding a hypothetical protein of Bacteroidales-like cell surfaces [169], which showed a >97% human specificity.

Obligate anaerobes such as Bacteroides cannot grow or multiply in environmental waters and have shorter survival periods than aerobes, facultative anaerobes, or viruses [57]. This property can provide a time progression of fecal pollution in environmental waters after Bacteroides spp. are released [170]. Molecular markers of both viable and non-viable cells can be detected using qPCR and DNA sequencing, so that they are less affected by cell states [171]. For example, the portion of culturable enterococci and E. coli in unfiltered freshwater can be reduced by 90% within 1.2 h and 1.4 h in sunlight, whereas it takes more than 6 h for the same reduction for HF183 and human polyomavirus [167].

2.2. Viral Markers

Fecal pollution detection using conventional FIB has several limitations. For general FIB of poor host specificity (e.g., E. coli and enterococci), they do not discriminate between human and animal fecal sources and therefore are inappropriate for source tracking [172]. Human-specific fecal indicators, mainly genetic markers of Bacteroides spp., are capable of fecal pollution source tracking from both point and non-point sources [150] but are not
always representative of enteric viral pathogens or protozoa. This is because viruses and parasite protozoa react to environmental stresses and wastewater treatments differently from bacteria [173]. There are outbreaks caused by waterborne viral pathogens or protozoa when FIB are not present or are below the regulated levels in water samples, as aforementioned [83]. One reason could be that FIB quantification, especially by culture-based methods, can provide false negative results if cells are in the VBNC state [174]. Therefore, it is important to employ fecal indicator viruses in water quality assessment for their higher correlation with enteric viral pathogens. These viral indicators can be either of fecal origin, such as coliphage and crAssphage, whose hosts are gut bacteria, or highly associated with human activities, such as pepper mild mottle virus, a plant virus which is commonly found in dietary pepper products [175].

2.2.1. Coliphage

In human guts, the viral community (termed human gut virome) is dominated by bacteriophages [176]. Coliphages are bacteriophages that specifically infect coliforms and are not pathogenic to humans. They are generally found in the gut and share similar morphologies, sizes, structures and resistance against disinfection methods with enteric viruses [177,178]. For example, coliphages can be resistant to environmental stresses, such as hydrostatic pressure (less than $1 - \log_{10}$ reduction at 600 MPa for 1 h), thermal inactivation ($4 - \log_{10}$ reduction at 72 °C for 20 min) and high or low pHs (no significant $\log_{10}$ reductions between pH 4 to 11 for 3 h) [179,180]. Therefore, coliphages can act as indicators for enteric viruses and human fecal pollution in rivers, drinking water and underground water [181]. Based on the infection site, coliphages are classified as somatic coliphages, which are DNA phages that infect coliform cells through their attachment to the outer cell wall, and male-specific (or F+) coliphages, which can be DNA or RNA phages that infect male (F+) host cells through the fertility (F) pili [61]. Apart from \textit{E. coli}, somatic coliphages can infect other pathogenic gut bacteria, such as \textit{Shigella} spp. and the multi-drug resistant \textit{Klebsiella pneumoniae} [182].

Two culture-based methods recommended by the U.S. EPA for enumerating somatic coliphages in freshwater and groundwater are the spot-test assay for their presence/absence and the single-agar-layer plate assay to quantify the plaque-forming units (PFU) [183,184]. The former requires 24 h, including 18 h for incubation and 6 h for the spot test, whereas the latter requires 16 h from sample processing to plaque enumeration. Hence, both assays do not allow for rapid detection. Another Bluephage method relies on the potential of \textit{E. coli} strains overexpressing $\beta$-glucuronidase when facing coliphage-mediated bacterial lysis. The substrate resulting from the overexpressed $\beta$-glucuronidase is chromogenic and produces a blue color that indicates coliphage density [185]. This method features an excellent sensitivity (as low as 1 PFU/100 mL), an excellent specificity ensured by the coliphage, and a rapid procedure within 6.5 h (2.5 h for pre-growing host \textit{E. coli} strains, 4 h for the assay) [186]. Thus, the Bluephage method can provide a rapid and sensitive estimation of fecal pollution in water samples.

F+ RNA coliphages are divided into four genogroups (I–IV) based on genomic organization and serological cross-reactivity. Higher abundances of genogroups II and III are detected in rivers, streams and groundwater affected by municipal sewage [187], whereas higher abundances of genogroups I and IV are detected in waters contaminated by animal farms or bovine wastewaters [188]. Therefore, genogroups II and III are identified as specific to human fecal pollution, whereas I and V are identified as specific to animal fecal pollution. F+ RNA coliphages can be enumerated by a modified Bluephage method, which selects F+ (male) \textit{E. coli} (e.g., strains CB14 and CB16) as the host [189]. Furthermore, \textit{E. coli} strain CB-390 allows for simultaneous detection of somatic and F+ RNA coliphages [190]. Since somatic and F+ coliphages have varying environmental persistence, the simultaneous detection of both can result in a more accurate estimation of total fecal pollution. For source tracking purposes, however, separate enumeration of somatic and F+ coliphages is more suitable, because somatic coliphages may be $4 - \log_{10}$
more abundant than F+ coliphages due to enhanced environmental persistence, which can mask the population of F+ coliphages [191].

2.2.2. CrAssphage

CrAssphage is a double-stranded DNA bacteriophage that infects *Bacteroides intestinalis*. Unlike most bacteriophages, the discovery of crAssphage was not based on biological experiments but was based on bioinformatics. In 2014, its genome was assembled for the first time by the de novo cross-assembly method from 12 previously published human fecal viromes [176]. It is named after the software *crAss*. This name is unconventional because it does not indicate any phylogeny, taxonomy or host information. So far, crAssphage shows no homology to any known virus. In 2018, crAssphage was isolated from human feces for the first time, and its host, *Bacteroides intestinalis*, a human gut symbiont, was identified [192].

The presence of crAssphage is common among human fecal viromes and fecal metagenomes, as it is present in 342 out of 466 datasets surveyed (73%) [176]. In addition, it is highly abundant in the human gut, accounting for 90% of human fecal viromes and 1.68% of total human fecal metagenomes, making it the most abundant phage in the human gut. crAssphage was thought to be of exclusively human fecal origin, but recently it has been found in cats [56] and monkeys [193]. The crAssphage relatives detected in monkeys are distantly related to the published human crAssphage genome, suggesting a long evolutionarily divergent history in humans and primates. Nevertheless, crAssphage is highly specific to humans, showing a specificity up to 98% [56].

A known host is required for any virus to be cultured. When the crAssphage genome was first identified, its host was predicted by in silico approaches, including OFR homology searches (blastn), CRISPR spacer identification, and co-occurrence profiling of the crAssphage genome and 151 fecal bacterial metagenomes [176]. Among the predicted hosts, *Bacteroides* was believed to be the most likely host. In 2018, a phage that dominated 96% of total viral reads of the *Bacteroides intestinalis* strain APC919/174 plate was isolated and found to be related to a known but previously uncultured crAssphage genome. That was the first successful isolation of crAssphage.

Quantification of crAssphage in environmental waters is mainly by culture-independent methods, such as metagenomics and qPCR. Primers in PCR/qPCR targeting predicted coding regions of the crAssphage genome have been designed and validated [172]. The two most frequently used primer sets CPQ_056 and CPQ_064 (Table 3) exhibit the highest sensitivity and specificity, meaning that PCR assays using these primer sets can obtain (1) PCR products of correct sizes among human feces/sewage samples; (2) no PCR product among animal fecal samples; (3) PCR products of correct sizes among diluted human feces/sewage samples; or (4) no spurious PCR products, such as primer-primer dimers. A number of subsequent studies have validated the amplification capability of the two primer sets [194–198].

Table 3. Available primers for qPCR detection of crAssphage, HAdV and HPyV.

<table>
<thead>
<tr>
<th>Target Virus</th>
<th>qPCR Assay</th>
<th>Primer or Probe Sequence (5′ → 3′)</th>
<th>Target Gene</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrAssphage</td>
<td>CPQ_056</td>
<td>F: CAGAAGTACAAACTCTTAAAAACGTAGAG</td>
<td>Predicted coding region</td>
<td>[172]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GATGACCAAATAAACAGCCATTAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probe: [FAM]-AATAACGATTTACGTGATGTAAC-[MGB]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CPQ_064</td>
<td>F: TGTATAGATGCTGCTGCAACTGTACTG</td>
<td>Predicted coding region</td>
<td>[172]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: CGTTGTTTTCATCTTTATCTTGCCAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probe: [FAM]-CTGAAATTGTTCATAAGCAA-[MGB]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAdV</td>
<td>VTB-1-HAdVF</td>
<td>F: GCCTGGGAACAAGGTCACA</td>
<td>Hexon</td>
<td>[199]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GCGTAAAGGCACCTTGGTAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probe: [Quasar]-CAGTGGCTGACCTTGATGATGTT-[BHQ2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VTB-2-HAdVC</td>
<td>F: GAGAOGTAGTCACTCTCCAGCGGTGAAT</td>
<td>Hexon</td>
<td>[199]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GAGAOGTAGTCACTCTCCAGCGGTGAAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probe: [FAM]-CCATCGCAAGACCGGCACACA-[BHQ1]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Cont.

<table>
<thead>
<tr>
<th>Target Virus</th>
<th>qPCR Assay</th>
<th>Primer or Probe Sequence (5′ → 3′)</th>
<th>Target Gene</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BKPyV and JCPyV</td>
<td>SM2: AGTCCTTTAGGTCTTCTACCTTT</td>
<td>P6: GGTGCCAACCTATGGAACAG</td>
<td>Conserved T-antigen</td>
<td>[200]</td>
</tr>
<tr>
<td>KGJ3 probe: [FAM]-TCATCACTGGCAAACAT-[MGB]</td>
<td>F: ATGTGTCGACGTAGATGAAAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPyV-2</td>
<td>R: GGAAGTCCTTTAGGTCTTCTACCTTT</td>
<td>Probe: [FAM]-AGGATCCCAACACTCTACCCCACCTAAAAAGA-[MGB]</td>
<td>JVGFP5</td>
<td>[201]</td>
</tr>
</tbody>
</table>

So far, the following features make crAssphage a fecal indicator virus suitable for human fecal pollution detection and source tracking. First, it is ubiquitous and abundant in human feces and sewage (5.28–7.28 log_{10} copies/100 mL) [194]; it is rarely detected in environments free of human fecal pollution [56]; and it has concentrations strongly correlated with sewage ARGs [2], pathogenic viruses [198,202] and conventional FIB [197]. In addition, it is not pathogenic to humans [193] and has molecular markers that can be detected by rapid, culture-independent and cost-effective methods [172], although it is unclear whether crAssphage can replicate itself outside human intestines but within Bacteroides cells.

2.2.3. Human Adenovirus (HAdV)

Adenovirus was first isolated in 1953 from human adenoids, hence the name [203]. Adenovirus is not a taxonomic unit and represents a group of viruses from six genera of the family Adenoviridae: Atadenovirus, Aviadenovirus, Ichtadenovirus, Mastadenovirus, Siadenovirus and Testadenovirus [60]. Although adenoviruses share a similar morphology, their hosts and genomic organizations differ from genus to genus. For example, hosts of mastadenoviruses are exclusively mammals, whereas hosts of aviatadenoviruses are exclusively birds. In addition, mastadenoviruses can synthesize unique proteins V and IX; the former is responsible for redistributing viral components into or out of the host cell nucleus, while the latter is a transcriptional activator [204,205]. So far, adenovirus has been detected in species of almost every vertebrate class, such as birds, mammals, reptiles, amphibians and fish, but has not been detected in invertebrates [60]. All human-specific adenoviruses (HAdV) belong to the genus Mastadenovirus.

HAdV is highly specific to humans. As it targets a conserved region of the hexon gene, HAdV is not detected by qPCR in samples that do not contain human feces, meaning there are no false-positive cases, whereas somatic coliphages, Bacteroides HF183 and crAssphage CPQ_056 are detected in the feces of cows, chickens and cats [56,206]. Except for HAdV, animal-specific adenoviruses can be used to identify animal fecal pollutions as well. Bovine adenovirus is found to be specific to cows [58], and porcine adenovirus is found to be specific to pigs [207]. Hence, despite the broad host spectrum of the adenovirus group, specific adenoviruses have excellent host specificity, enabling source tracking of human and animal feces.

HAdV in groundwater has a stable persistence and may survive for long periods. For example, the low decay rates in groundwater (0.0076 log_{10}/d at 4 °C and 0.0279 log_{10}/d at 20 °C) result in a long natural T_{90} (time required for 1−log_{10} reduction) of 131 d at 4 °C and 36 d at 20 °C [208]. In WWTPs, HAdV is highly resistant to UV disinfection and is frequently detected in the effluent and receiving environmental waters [209,210]. Among four enteric viral pathogens, echovirus, coxsackievirus, poliovirus and HAdV type 2, a UV dose of 27 mW/cm^2 sufficiently brought 3−log reductions to the first three viruses, whereas it is 119 mW/cm^2 for HAdV type 2 [211]. One possible reason is that adenoviruses have a double-stranded DNA genome, which is more stable than RNA or single-stranded DNA genomes. Additionally, HAdV uses human cellular enzymes to repair its DNA damage [212]. In this case, HAdV is not truly resistant to UV radiation but is actively able to repair the damage to DNA brought by UV.

HAdV is typically found at much lower concentrations than E. coli, crAssphage, or Bacteroides, often at levels close to or lower than its detection limit. For example, HAdV can...
be detected in sewage samples at ~10^5 gene copies/L, and only 10^2–10^3 gene copies/L in rivers, urban streams and estuarine waters [209]. Such low concentrations in environmental waters can be a challenge for HAdV plaque assays, because it is estimated that 1–10^3 virions are required for 1 PFU to form depending on the virus [213]. Hence, environmental water samples should be concentrated for HAdV plaque assay by membrane filtration or other methods [214]. In culture, HAdV may take up to 1 week to infect, reproduce within and lyse human cells to generate visible PFUs [215]. The required time for a HAdV plaque assay is not practical for routine water quality monitoring or rapid decision making. Additionally, HAdV is pathogenic to humans, causing upper respiratory symptoms and fever, though it causes mild or asymptomatic infection in most cases [216].

Culture-independent methods to enumerate HAdV from environmental waters include regular PCR, nested PCR, qPCR, integrated cell culture qPCR, and flow cytometry [210,217]. qPCR assays using a nested primer set and a degenerate primer set targeting the hexon gene (Table 3) indicate excellent the host specificity of HAdV and bovine adenovirus (no false positives), with an overall sensitivity of 78% and 73%, respectively [58]. Additionally, fluorescent antibodies targeting two HAdV proteins, Hexon and E1A, can be used in fluorescence-activated cell sorting for HAdV rapid detection with excellent linearity between fluorescence intensity and HAdV concentration (R^2 > 0.9) and sensitivity in the range of 1–10^4 PFU per assay [210].

2.2.4. Human Polyomavirus (HPyV)

HPyVs are a group of human enteric viruses that belong to the genera Alpha-, Beta- and Delta-Polyomavirus within the family Polyomaviridae, with circular double-stranded DNA genomes [218]. First isolated in 1971 from human brain tissue, there are 14 distinct HPyV species identified so far [218,219]: some are human feces-specific (e.g., HPyV-6 and HPyV-12) and some are human urine-specific (e.g., HPyV-9, BKPyV and JCPyV). These HPyVs are potential source tracking indicators due to their human specificity, environmental prevalence and resilience. Animal-specific PyVs, such as bovine, porcine, and ovine PyVs, are specific to cattle, pigs, or sheep, so they can be used for the identification and source tracking of animal fecal pollution [220].

HPyVs are more specific to humans when compared to FIB and are less likely to be detected in non-human fecal samples. For example, qPCR assays targeting HPyV (the T-antigen gene), Bacteroides HF183 and crAssphage (CPQ_56) among 359 animal feces samples exhibit a human specificity of 99.72%, 96.65% and 98.61%, respectively [56]. Despite the excellent human specificity, HPyV has variable sensitivity based on sample types. For example, urine-specific HPyV is detected in 6 out of 26 urine samples from healthy individuals, ranging from 6.61×10^2 to 1.2×10^7 copies/mL, showing a limited sensitivity of 23.1% [200]. In WWTPs, however, HPyV is detected in 36/37 influent and effluent wastewater samples, showing a high sensitivity of 97.3% [209].

HPyV is primarily excreted in human urine and feces; hence, its presence in environmental waters reflects anthropogenic pollution. BKPyV and JCPyV, specific to human urine, are the first HPyV identified in environmental samples (from sewage) [221]. Since then, increasing HPyV species have been detected in many environmental waters. HPyV is prevalent in rivers, urban streams, recreational beaches and drinking water [209,222,223]. WWTPs are believed to be the key point for HPyV releasing to open water bodies. Though unable to multiply, JCPyV was reported to have a T90 as long as 64 d in sewage at 20 °C [224]. Viruses adsorbed onto particle surfaces may have prolonged survival relative to free viruses, possibly because particles reduce contact between viruses and disinfectants [225]. In that case, HPyVs excreted in human feces are expected to have longer environmental persistence than HPyVs being excreted in the urine.

Like HAdVs, HPyVs are present at low concentrations in environmental waters, and therefore HPyV enumeration often involves the concentration of large volumes of water samples. Simple membrane filtration, using a 0.22 or 0.45 µm filter, fails to concentrate HAdV or HPyV particles, as their size is much smaller than the pore size of regular
filters. Salt and acidic pH may increase viral adsorption to the filter because viruses in ambient waters are negatively charged [226]. Hence, current membrane filtration for virus concentration often requires pH adjustment of the water sample, for example, acidification using HCl if a nitrocellulose membrane is used, to obtain a higher HPyV recovery without introducing PCR inhibitors [200,227].

Concentrated HPyVs can be quantified by plaque assay using human fetal glial cells or tumor cells, immunofluorescence, microarray and PCR/qPCR [56,228,229]. qPCR targeting the conserved T-antigen of BKPyV and JCPyV is the most extensively used molecular method for HPyV detection in environmental waters (Table 3), with a sensitivity (i.e., limit of detection) as low as 10 copies per reaction [200]. HPyVs detected by immunofluorescence assay are generally $2^{-\log_{10}}$ less than HPyVs detected by qPCR, because only infectious HPyVs can be detected by immunofluorescence assay or plaque assay, whereas both infections and non-infectious HPyVs can be detected by qPCR [228]. For some samples, it is important to analyze the infectivity of viruses, and culture-based methods remain the gold standard. Infectious HPyVs do not necessarily cause diseases, as HPyVs are generally not pathogenic and are categorized as Biosafety Level II.

Compared to bacterial indicators, HPyV better correlates with viral pathogens in the environmental waters. For example, norovirus, a human enteric pathogen, is generally not routinely monitored partially because of its pathogenicity and RNA genome. Norovirus cannot be efficiently predicted by FIB because it has longer environmental persistence and may be present in high concentrations in non-disinfected wastewater effluent and receiving waters, such as rivers, urban streams and estuarine water, without causing outbreaks [209,230]. Among these waters, HPyV is not only prevalent but also significantly correlates ($p < 0.05$) with norovirus (determined by reverse transcription qPCR). In recreational coastal beaches impacted by non-point source fecal pollution, HPyV levels may not be significantly associated with levels of Bacteroides, enterococci, total coliforms or fecal coliforms, but can be significantly associated with levels of HAdV [222]. The poor associations between fecal indicator viruses and FIB strengthen the need to employ fecal indicator viruses for water quality monitoring, as FIB are often poorly correlated with viral pathogens, whereas HAdVs and HPyVs themselves are human enteric pathogens, sharing similar environmental surviving strategies with other human enteric viral pathogens, and are therefore of greater potential to indicate health risks.

3. Emerging Bioengineering Enumeration Methods

All current biological methods have limitations. Culture-based methods (e.g., plate counting and membrane filtration) are labor- and time-consuming and often lead to delayed results that are insufficient to catch sudden bacterial outbreaks or to support proactive actions. It is possible that by the time fecal pollution is detected in drinking water using culture-based methods, the investigated water may have already been distributed and consumed. In addition, culture-based methods provide information on a fraction of the total bacterial population, because non-culturable bacteria or VBNC cells are not included. A limitation for molecular methods such as qPCR and metagenomics is that they cannot distinguish DNA of living and dead cells, i.e., they cannot exclude the interference of relic DNA. It is estimated that relic DNA can persist for weeks to years in soil, account for 40% of total DNA being sequenced, and have inflated the observed richness of prokaryotes and fungi by 55% [231]. Sample types (e.g., water, soil, plant) and DNA extraction methods also impact the DNA recovery. Multi-disciplinary platforms integrated with cutting-edge techniques in the fields of biology, engineering, chemistry and computer science may be powerful alternatives or supplements to current methods to overcome (at least some of) the existing limitations.

3.1. A Time-Lapse Coherent Imaging Platform

A lens-free imaging platform capable of early detection and species classification of living FIB on agar plates is a promising new technology [232], though manual membrane
filtration of water samples and inoculation of concentrated indicator cells are required. When an inoculated agar plate is placed in the incubator, the imaging system scans the plate every 30 min, simultaneously with the bacterial incubation. Captured plate images are sent to the integrated software, where periodically captured plate images are digitally stitched together in chronological order to form time-lapse image stacks. These stacks are reconstructed to reveal static impurities (e.g., dust, air bubble and light speckle) and growing bacterial colonies on the agar surface.

Fifteen pure culture agar plates of three total coliform bacteria (E. coli, Klebsiella aerogenes, and Klebsiella pneumoniae) are blindly tested on this platform. For validation of colony recognition sensitivity and precision, colonies on these plates are also manually enumerated. A colony recognition sensitivity of 80% is achieved when E. coli, K. aerogenes and K. pneumoniae are incubated for ~6.8 h, ~8.8 h, and ~6.0 h, respectively. When incubation continues to 12 h, a sensitivity higher than 95% is achieved for all three coliforms. After 6 h incubation, precision for all three coliforms is close to 100%. A species classification accuracy of ~80% is achieved when E. coli, K. aerogenes and K. pneumoniae are incubated for ~8.0 h, ~12.0 h, and ~7.6 h, respectively. For E. coli and K. pneumoniae, higher species classification accuracies (~97.2% and ~98.5%) are achieved when incubated for 12 h. Colony recognition sensitivity, colony recognition precision, and species classification accuracy are positively correlated with the incubation period.

This platform enables a rapid (87 s to scan a 60-mm diameter petri dish), high-resolution (with a linewidth scanning resolution of ~3.5 µm) and automatic plate image capture for every 30 min (interval adjustable). Such a high scanning resolution enables early recognition of emerging colonies that are invisible to naked eyes. Compared to EPA-approved culture-based methods (e.g., 1103.1 and 1604 methods), this platform is more time- and labor-saving. In the blind testing, a whole test (bacterial incubation, colony recognition and species classification) can be completed within 12 h, with an excellent limit of detection as low as 1 CFU/L, at a low cost of only $0.6, including the cost of preparing an agar plate. However, this platform is not fully automatic because manual sample processing is still required.

3.2. An Online Culture-Independent Imaging Platform

Drinking water requires faster and more reliable quality monitoring than water for other purposes because it is directly consumed by humans. Though water utilities monitor water quality regularly, if culture-based methods are applied, results come out 1–3 d delayed and do not provide information of unculturable bacteria or VBNC cells. To resolve these problems, an automated, online and culture-independent platform capable of enumerating total bacteria within 10 min has been designed [233]. This platform is installed within drinking water pipeline systems, where drinking water flows through a flow cell under the pressure of the water source or an inlet pump. To initiate a measurement, the flow cell becomes sealed to trap 6 µL of water sample inside. The stagnant trapped water sample is scanned by a lens-based imaging system at different focuses, i.e., it undergoes a 3D scanning. Captured images are sent to the platform-integrated software for analysis. Then the flow cell reopens and gets flushed, preparing for the next measurement. By combining images of different focuses (i.e., depths), a 3D model of the water sample can be established from which particles will be recognized, together with their 59 optical parameters, such as area, length, contrast and granularity. Then, the pre-trained software classifies bacteria and abiotic particles based on the 59 optical parameters.

In blind testing using monotype suspensions containing either bacterial or abiotic particles, a classification accuracy of >80% is achieved for all bacterial suspensions, and less than 8% of abiotic particles are mistakenly classified as cells. The average classification accuracy for monotype suspensions is 90%. In the blind testing using mixed suspensions containing both bacteria and abiotic particles, the overall classification accuracy is 78%. It is worth noticing that for mixed suspensions, due to lack of manual curation, false negative (i.e., bacteria being classified as abiotic particles) and false-positive cases (abiotic
particles being classified as bacteria) can both occur but cannot be distinguished [234]. Based on the small sampling volume (6 µL), this platform has a theoretical lower limit of detection of $1.6 \times 10^2$ particles/mL. Water samples with lower particle concentrations may be associated with high uncertainty due to sample inhomogeneity. On the other hand, an upper limit of $5 \times 10^9$ particles/mL is also determined, because water samples with higher particle concentrations may not be detected due to strong background light scattering, and heavy water pollution will significantly shorten the lifespan of a flow cell.

This platform enables a fully automatic, rapid and high-frequency (10-min time resolution) monitoring of bacteria and abiotic particles, making it possible to detect sudden water quality changes and classify them as biotic or abiotic. This platform does not require chemical supplies or produce any hazardous wastes. Acquired data can be wirelessly transferred to a data server in real time (i.e., online). The only required maintenance is to replace the flow cell every few months. As a culture-independent platform without species preference, it is suitable to estimate the total bacterial biomass. However, dead or inactive bacteria (e.g., those inactivated by UV treatment), which are no longer of health risk, can still be detected, as long as their cells remain intact, with the same or similar optical parameters as living cells. In addition, it does not apply to enumerating virus or protozoa, which are relatively either too small or too big compared to water chemical particles.

### 3.3. An On-Site Microflow Cytometry Free of Fluorescence Staining

Flow cytometry can be used to enumerate bacteria and viruses in water samples, such as coliforms in drinking water and HAdV in sewage effluents [81,210]. These methods have limited on-site use because fluorescence staining is often involved, which requires a bulky cytometer, expensive reagents, and complicated sample preparation. The development of a microfluid-based portable microflow cytometer enables rapid and on-site *E. coli* enumeration without fluorescence staining [235]. A major part of this platform is a four-layer panel; between the middle two layers is a 50-µm-wide microfluidic channel. The water sample and sheath fluid are injected into the channel by a syringe pump. The sheath fluid is injected at a much higher rate than the water sample so that cells in the water sample are forced to pass through the channel one by one unidirectionally towards the outlet. An input light beam is perpendicular to the channel, where light is differentially scattered by cells, particles or water. An on-chip lens system beneath the panel records the side-scattered light signals to separate *E. coli* cells and particles.

This microflow cytometer is tested using two suspensions, one made of pure *E. coli* cells diluted in PBS buffer and the other made of both *E. coli* cells and polystyrene beads with diameters of 2 and 4 µm to mimic particles from environmental waters. For a pure *E. coli* suspension, the sample-to-sheath flow-rate ratio is found to be negatively correlated with detection efficiency. When *E. coli* is injected at 50 µL/h with a constant sheath flow of 313 µL/h, a high detection efficiency (97%) can be achieved. When the sample flow rates increase to 100 and 150 µL/h, the detection efficiencies decrease to 63% and 45%, respectively. Regardless of sample flow rates, the high signal/noise ratios (9.7 to 11) are supposed to reduce the possibility of false counting. For mixed suspension, the presence of 2 and 4 µm diameter polystyrene beads does not affect the detection of *E. coli* cells, because the side-scatter signal intensity is proportional to particle size, and therefore, the beads have significantly higher side-scatter signal intensities than *E. coli* cells. Furthermore, even the 2 and 4 µm beads have distinct intensities and can be distinguished. Therefore, the presence of particles larger than 2 µm is unlikely to affect the detection of *E. coli* by the platform.

This microflow cytometer features a rapid (50 s for a measurement), less costly (no fluorescence staining required) and on-site (portable equipment) enumeration of *E. coli* in water samples, with excellent cell detection efficiencies at low sample-to-sheath flow rate ratios. The high signal/noise ratios decrease false cell counting from the background noise, and particles of similar or larger sizes of *E. coli* do not interfere with cell detection. However, interference may arise when particles smaller than 2 µm are present; in that case, cells require other staining techniques, such as fluorescence staining, to be distinguished.
One major limitation of this platform lies in its limited throughput, as higher sample injection rates would decrease detection efficiencies. Additionally, this platform is not suitable for long-term operation, because the throughput decreases over time due to cell precipitation in the syringe during measurements. In addition, *E. coli* and water have close refractive indexes, so a pre-concentration is required to increase cell density by 3–6 orders of magnitude; otherwise, site-scatter signals produced by *E. coli* cannot be sufficiently separated from the background noise.

4. Conclusions

Traditional water monitoring relies on *E. coli*, enterococci and total coliform testing, which are well-validated and widely used methods. These are routinely monitored in drinking water or recreational water because they are associated with waterborne illnesses and can be enumerated with ease. However, their limited host specificity hampers accurate source tracking of fecal pollution from nonpoint sources. It is particularly important to identify human sources of fecal pollution, as they usually present high health risks. Source tracking indicators, such as certain species/stains of *Bacteroides*, human adenovirus and crAssphage, have better host specificity and therefore can be used for human fecal pollution source tracking. Whether using fecal indicator bacteria or viruses, a single fecal indicator is unlikely to predict all pathogens or diseases; hence, multiple indicators are recommended to be used in combination to assess the water quality and health risks throughout. These additional tools can augment traditional monitoring and serve as a toolbox for researchers and engineers to determine appropriate combinations based on their needs.

The enumeration of fecal indicators by culture-based methods is relatively inexpensive and easy to perform. Among them, plate counting is the most common for bacterial indicators. In addition, cell-culture-based methods remain the gold standard to assess the infectivity of viral indicators/pathogens. However, culture-based methods are time-consuming, often leaving insufficient time for the water quality regulatory agencies to take proactive actions. In addition, unculturable, difficult-to-culture and viable-but-not-culturable fecal indicators are unlikely to be enumerated by culture-based methods. Thus, culture-based methods tend to underestimate the health risks. Molecular methods, such as qPCR and digital PCR, on the other hand, are rapid, of good replicability and have broader applicability to various fecal indicators. One limitation is that molecular methods cannot distinguish detected genetic materials from living or dead cells; therefore, they tend to overestimate the health risks, because dead or inactive indicators/pathogens cannot be excluded.

Emerging multidisciplinary bacterial enumeration platforms often incorporate cutting-edge techniques in biology, engineering and computer science, aiming to simplify procedures, prompt automation, improve detection sensitivity, and reduce the requirements of time, cost and labor. The first platform discussed here enables automatic colony plate counting and species classification; the second platform is suitable for high-frequency and long-term quality monitoring of drinking water, and the third platform introduces an on-site flow cytometry based on microfluidics without fluorescence staining. These platforms currently have individual limitations, but still can be powerful alternatives or supplements to current biological methods.

**Author Contributions:** Writing—original draft preparation, E.L., F.S.; writing—review and editing, H.E.S., T.A.E.; investigation, H.E.S., T.A.E.; supervision, H.E.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** We gratefully acknowledge funding supporting environmental studies provided by the Ontario Research Excellence Fund (Grant # RE09-077, Ontario Ministry of Economic Development, Job Creation and Trade), the Ontario Ministry of Environment, Conservation and Parks (MECP), and the Natural Sciences and Engineering Research Council (NSERC) of Canada (Grant # ALLRP 554507-20).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.
Data Availability Statement: Not applicable.

Acknowledgments: We wish to thank our municipal and provincial partners, including the Ministry of the Environment, Conservation and Parks (MECP) and Toronto Public Health, for helpful discussions in support of our monitoring studies.

Conflicts of Interest: The authors declare no conflict of interest.

References


70. Coliform Bacteria in Drinking Water Supplies; New York State Department of Health, Center for Environmental Health: New York, NY, USA, 2017.


99. Schwering, M.; Song, J.; Louie, M.; Turner, R.J.; Cери, H. Multi-species biofilms defined from drinking water microorganisms provide increased protection against chlorine disinfection. *Biofouling* 2013, 29, 917–928. [CrossRef] [PubMed]

100. Mellefont, L.A.; McMeekin, T.A.; Ross, T. Performance evaluation of a model describing the effects of temperature, water activity, pH and lactic acid concentration on the growth of *Escherichia coli*. *Int. J. Food Microbiol.* 2003, 82, 45–58. [CrossRef]


140. Skanavis, C.; Yanko, W.A. Clostridium perfringens as a potential indicator for the presence of sewage solids in marine sediments.


192. Shkoporov, A.N.; Khokhlova, E.V.; Fitzgerald, C.B.; Stockdale, S.R.; Draper, L.A.; Ross, R.P.; Hill, C. *ΦCrAss001* represents the most abundant bacteriophage family in the human gut and infects *Bacteroides* [CrossRef]


Robbertse, B. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* 2020, 2020, baaa062. [CrossRef] [PubMed]
231. Carini, P.; Marsden, P.J.; Leff, J.W.; Morgan, E.E.; Strickland, M.S.; Fierer, N. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* **2016**, *2*, 16242. [CrossRef]


