



OPEN

Distribution of host-specific Bacteroidales marker genes in water sources of selected rural areas of Vhembe District, South Africa

Barbara Mogane¹, Leonard Owino Kachienga¹, Ilunga Kamika², Renay Ngobeni-Nyambi³ & Maggy Ndombo Benteke Momba^{1✉}

Access to safe drinking water sources and appropriate sanitation facilities remains a dream in low and middle-income countries including South Africa. This study identified the origin of faecal pollution by using quantitative polymerase chain reaction (qPCR) targeting host-specific Bacteroidales genetic markers to track the distribution of human-specific (*BacHum*) and animal-specific (cattle—*BacCow*, chicken—*Cytb*, pig—*Pig-2-Bac*, dog—*BacCan*) markers in water sources used by rural communities of the Vhembe District Municipality (VDM). Results revealed the prevalence of *BacHum*, *BacCow*, and *BacCan* in all surface water sources in Thulamela Local Municipality (TLM) and Collins Chabane Local Municipality (CLM) during wet (100%) and dry seasons (50–75%). *Cytb* was not detected in untreated spring water in TLM and CLM, and *Pig-2-Bac* was not detected in untreated hand-dug well water in TLM during both seasons. Household-level analysis detected *Cytb* (28.8% wet, 17.5% dry), *BacHum* (34.4% wet, 25% dry for *Pig-2-Bac*) in stored untreated spring water in CLM, and *Cytb* (42.9% wet, 28.5% dry) in untreated hand-dug well water in TLM. Despite differences in detection frequencies of host-specific Bacteroidales, the study highlights the public health concern of faecal pollution in rural VDM households.

Keywords Drinking water, Faecal pollution, Host specific Bacteroidales, Rural communities, Water sources

One of the challenges facing the world today, particularly for the poorest people in low- and middle-income nations, is the contamination of water sources coupled with a lack of sanitation¹. Untreated polluted water not only poses a major threat to human health but also to the ecosystem health. According to the Global Burden of Disease study, 1.2 million people died in 2017 as a result of contaminated water, which is three times the number of homicides in 2017² and was equal to the number of people who died in road accidents, globally³. It is also well known that one of the main causes of death in children under the age of five years is diarrhoeal diseases caused by contaminated water. Furthermore, the lack of sanitation or poor sanitation infrastructure and hygiene practices is responsible for the contamination of water sources in communities. Therefore, common waterborne illnesses are caused by ingesting water contaminated with faecal matter.

Faecal pollution of water sources caused by humans and animals impairs the quality of water and introduces pathogens into these water sources and poses a significant public health risk in communities with a lack of or inadequate sanitation or poor excreta management⁴. The situation is of particular concern in rural communities of most developing countries which depend on untreated surface water and groundwater for domestic purposes^{5,6}. A study conducted by Traoré et al.⁷ has reported surface runoff containing faecal matter, which leads to water source pollution. Identification of faecal pollution is paramount to safeguarding public health and protecting water sources.

¹Department of Environmental, Water and Earth Sciences, Arcadia Campus, Tshwane University of Technology, 175 Nelson Mandela Avenue, Arcadia, Pretoria 0001, South Africa. ²Department of Environmental Sciences, School of Environmental Science, College of Agriculture and Environmental Sciences, Faculty of Sciences, Florida Campus, University of South Africa, Florida, Roodepoort 1710, South Africa. ³Department of Microbiology, Stellenbosch University, Private Bag, X1, Mitieland 7602, South Africa. ✉email: momba1958@gmail.com

There are various pathways by which water sources become contaminated; these include sewage discharge into water sources, agricultural runoff, leaking sewers, which when located deep underground can cause sewage to enter groundwater, and open defecation⁸. In South Africa, however, water scarcity and dysfunctional sewerage infrastructure have exacerbated the pollution of water sources, especially in rural communities⁹. Vhembe District Municipality in Limpopo is one of the rural areas characterised by scarce water sources as a result of arid climate, unfavourable topography, and sandy rivers. However, some studies have confirmed that most water sources such as rivers, springs, and household container-stored water in the municipality are contaminated with microbial pathogens such as *Escherichia coli* and total coliforms. These indicated faecal pollution, which compromised the quality of water often lead to significant health risks^{10,11}. A study by Ngweya & Kgopa¹², has demonstrated that most communities in the Vhembe District were dependent on untreated surface water, which was highly subjected to pollution from various sources, including human and animal waste. In addition, the already scarce water sources in this region are under threat due to faecal pollution¹³. Detecting the sources of faecal contamination, whether human or animal, can help to determine which routes pose risks to human health.

Traditionally, faecal indicator bacteria (FIB), such as *Escherichia coli* and enterococci are used to determine the faecal pollution in water. However, these indicators have been linked to health risks for swimmers¹⁴ and are ineffective at identifying the sources of faecal pollution because of their lack of host specificity, non-specific relationship with human pathogens, and capacity for natural reproduction^{15–17}. As a result, microbial source tracking (MST) techniques, which use host-associated genetic markers to identify different sources of faecal contamination, such as cattle, humans, gulls, and dogs, have been widely utilised¹⁸. These markers, which represent around 25–30% of the human microbiome, have been used as a potential host-associated genetic marker because Bacteroidales are prevalent in faeces and contain these genetic markers¹⁹. Additionally, Bacteroidales adapt to their hosts in different ways, enabling the identification of host sources of faecal pollution²⁰.

Numerous research studies have discussed the applicability of Bacteroidales genetic markers for the identification of faecal contamination origins in underdeveloped nations like Kenya²¹, Tanzania²², and Bangladesh²³. Most MST techniques employ quantitative polymerase chain reaction (qPCR) to detect host-associated markers. When compared to conventional culture-based approaches, qPCR provides rapid quantification of the sources of pollution, delivering information more quickly. Similar concepts are used in end-point PCR. However, the majority of fluorescently labelled DNA probes are used to continually monitor the production of double-stranded DNA (dsDNA)²⁴. The use of these techniques in rural areas of Vhembe District Municipality will assist the Water and Sanitation Services Authorities in addressing the problem of faecal pollution of water sources. In this study, we applied Bacteroidales MST using qPCR assays to identify the sources (origin) of faecal contamination, whether human or animal and to gain an understanding of the routes of pathogen transmission from various water sources (surface water and groundwater sources) used by the communities to container-stored drinking water in homes of Collins Chabane Local Municipality (CLM) and Thulamela Local Municipality (TLM) of the VDM.

Materials and methods

The study was designed by considering four scenarios based on the water sources used by communities. The first scenario focused on households depending only on surface water sources (rivers and dams) that are subjected to treatment at the level of water treatment plants before distribution to the households. The second scenario focused on households that only use untreated water sources from the river. The third scenario was considered for households that only use untreated groundwater sources (spring, hand-dug well) such as spring water for domestic purposes. The emphasis of the fourth scenario was on households that only use untreated groundwater such as hand-dug well water. During the study period, the research team worked hand-in-hand with water service and sanitation officials in the Vhembe District Municipality as well as with the community leaders and their respective households.

Ethical approval and informed consent

Prior to the execution of the study, ethical approval was obtained from the Faculty of Science Research Ethics Committee at the Tshwane University of Technology's (TUT) (FCRE 2019/09/017 (FCPS 03) (SCI) and 20 March 2020, and the study was conducted by taking into consideration all the requirements issued by this committee. Furthermore, permission was obtained from local chiefs, community leaders and the Vhembe District municipality. All the residents were made aware of the study, and for each participating family, an informed consent form outlining its scope was in English and translated into Tshivenda Xitsonga as the majority of settlements in this District are Tshivenda and Tsonga.

Description of the study area and population

The Vhembe District Municipality is located in the Northern part of the Limpopo Province and shares borders with Capricorn and Mopani District Municipalities in the east and west, respectively. The District covers 21,407 km² of land with a total population of 1,393,949 people, according to Stats SA, Community Survey in 2016. Vhembe District Municipality is composed of four (4) local municipalities (Thulamela, Collins Chabane, Musina, and Makhado) and Fig. 1 illustrates the study area and sampling sites. The TLM and CLM were the focus of the current investigation. The former local municipality has a land area of 2,893.936 km² and is located at 22° 57' S and 30° 29' E. There is a total population of 497,237 and 130,321 households. The latter local municipality has a total population of 347,974 people and 91,936 houses spread across 5,467.216 km² (22° 35' S 30° 40' E). In terms of gender, females make up the majority of the population (757,501), while males account for (645,278). The current study covered five villages from the two selected local municipalities based on the water sources and open defecation practices. Table S1 (Supplementary material) illustrates the villages and their population sizes.

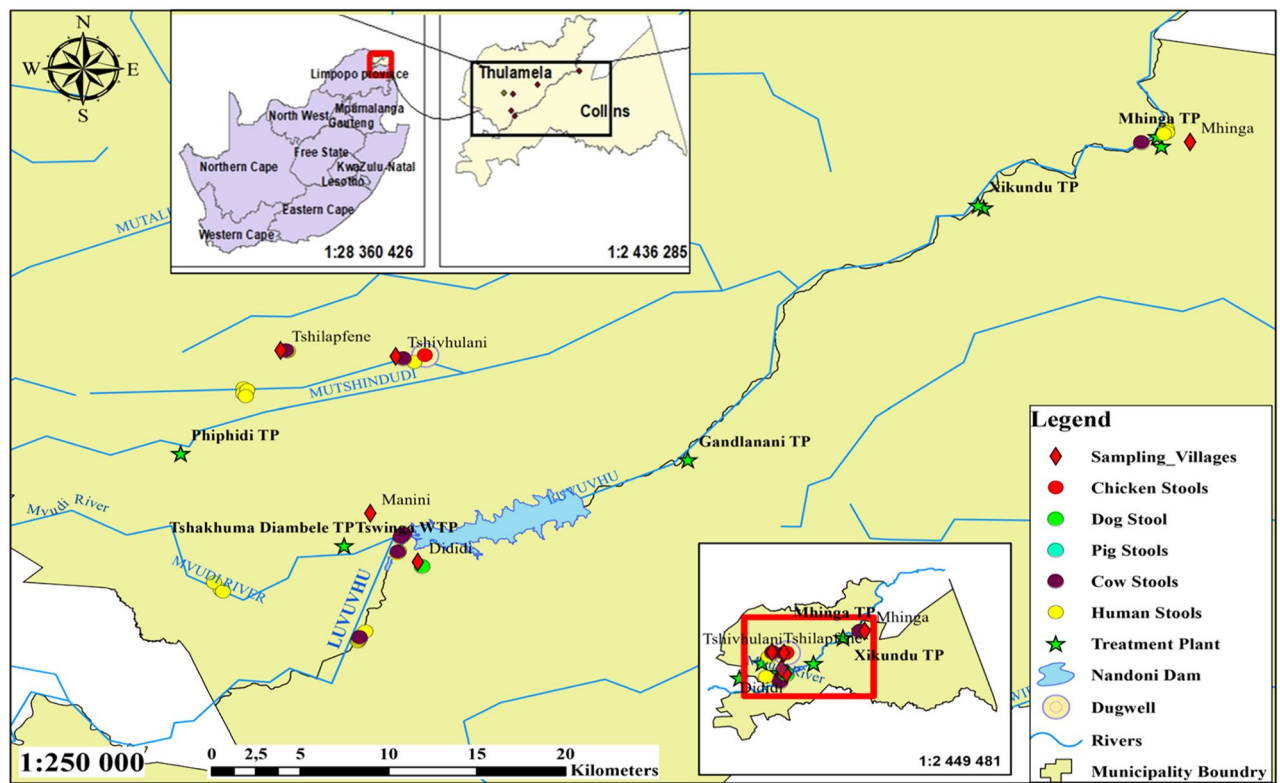


Figure 1. A map showing the geographical location of the study area (ArcMap ArcGis version 10.8).

Study survey and household selection

A survey questionnaire was structured to obtain the following information: (i) demographic data of the household members, such as age, gender, education status, and employment status; (ii) type of water supply system: water sources, frequency of collection, type of water treatment before use; (iii) sanitation and whether open defecation is being practised; and (iv) health and hygiene practices. The Faculty of Science Research Ethics Committee at the Tshwane University of Technology has reviewed the questionnaire to ensure that it covers all relevant aspects of the study objectives and ensured the ethical treatment of participants. The questionnaire was pilot-tested with a small group of participants. Mature and elderly (between 30 and 65 years) were selected during the data collection period from March 2020 to March 2021, the respondents were selected because they are the decision-makers regarding water safety practices and possess extensive knowledge about the water treatment technologies used. An informed consent form was presented first for the participant to agree and sign, and then followed by a structured questionnaire and the information was obtained only from the households who signed the consent form. Where respondents were unable to provide pertinent information, statistical reports published between 2011 and 2016 were used. The decision was made using a standard random sampling approach and a random sampling procedure. In total, 1388 surveys were conducted using structured questionnaires in Thulamela, Makhado, and Collins Chabane Local Municipalities. For villages with the highest number of households (1000), a total of 50 households were randomly selected, representing around 5% of all the households (Murei et al., 2022). Based on the main criteria, which focused on villages practising open defecation in the vicinity of water sources used by the communities, the following villages were selected: 20 households in Manini, 24 in Tshivulani, and 10 in Tshilapfene located in TLM; as for CLM, 20 households in Mhinga, and 8 households in Dididi were selected.

Collection of samples

Collection of faecal samples

In the selected villages, human and animal faecal samples were aseptically collected using sterile stool collection tubes near water sources (rivers, springs, hand-dug wells, and dams) from March 2021 to April 2021 (wet season) and from June 2021 to July 2021 (dry season). Of these, 55 stool specimens were collected during the wet season and 56 during the dry season (dry season). Altogether 111 composite samples collected consisting of human ($n=49$), cow ($n=3$), pig ($n=3$), chicken ($n=3$) and dog ($n=7$) faecal samples. The samples were transported on ice in a cooler bag to the University of Venda, where they were processed in less than 24 h.

Collection of water samples

During March 2021 to April 2021 (wet season) and June 2021 to July 2021 (dry season), a total of 1032 water samples were aseptically collected in 2 L sterile bottles, from the villages' water sources, water treatment facilities,

and homes. The water samples were collected 4 times from each sampling points during wet season and 4 times during dry season to ensure reliability and accuracy of the results, thereafter, they were transferred to TUT Water Research Unit laboratory at Tshwane University of Technology on ice in a cooler box, and then immediately analysed²⁵.

Microbial source tracking (MST) markers

DNA extraction and quantification

The membrane filtration technique was used to concentrate human- and non-human-associated bacterial genetic markers from water samples. Briefly, 300 mL of water sample was filtered through sterile nitrocellulose acetate or polycarbonate membrane filter (47 mm diameter and 0.22 µm pore size) (Merck Millipore). The membrane filters were transferred into a sterile 50 mL screw-cap tube containing 15 mL of sterile PBS buffer. The cells were removed from these membranes for 5 min using a benchtop shaker (OrbiCult™ AS1, Esco Lifesciences, Singapore) to obtain a homogeneous mixture. This was followed by centrifugation at 12 000 rpm for 30 min. The supernatant was discarded, and the pellet was then placed in a – 20 °C fridge and frozen until further use.

A 200 µL aliquot of the water pellet samples was transferred into ZR BashingBead™ Lysis Tubes and subjected to DNA extraction using the Quick-DNA™ Faecal/Soil Microbe Microprep kit (Zymo Research, Irvine, CA, USA) according to manufacturers' instructions, the pellet was resuspended in 750 µL of BashingBead™ buffer. A similar procedure was followed for faecal samples by transferring ≤ 150 mg of faecal samples into ZR BashingBead™ Lysis Tubes. The quantity of the extracted DNA was then determined using the NanoDrop™ 2000 spectrophotometer (Thermo Scientific, Johannesburg, South Africa). Negative control was included during each extraction cycle to exclude contamination from the DNA extraction reagents and buffers. All the DNA samples were stored at – 80 °C until further analyses.

Quantitative polymerase chain reaction analyses

Validation of quantitative polymerase Chain

For the validation of each qPCR assay, tenfold serial dilutions were prepared from a DNA plasmid containing known sequences to generate the standard curve (range 10¹ to 10⁵ copies/reaction) using the following formula:

$$y = mx + b \quad (1)$$

where: y is the Ct value, m is the slope, b is the y -intercept, and x is the log(quantity).

The efficiency of quantitative PCR and limit of detection (the lowest concentration at which we obtain 95% detection) of the host-specific Bacteroidales primers were then calculated using the following formula as described elsewhere²⁶:

$$\text{Efficiency} = 10^{(-1/\text{slope})} - 1; \text{ if } n = -3.322, \text{ then } E = 1, \text{ i.e., } 100\% \text{ efficiency} \quad (2)$$

$$\text{Limit of detection (LOD)} = 3.3 * [\text{SD (standard deviation intercept) / Slope}] \quad (3)$$

The primers used to detect host-specific Bacteroidales genetic markers in humans and animals (cattle, chicken, pigs and dogs) in this study were previously utilised in other regions and published^{27,28}. In order to validate these Bacteroidales for the Vhembe District Municipality, a total of 111 faecal samples from humans, cattle, chicken, pigs, and dogs were subjected to qPCR to determine the diagnostic sensitivity, the diagnostic specificity and the relative accuracy of the markers using the following formulae, as described by previous investigators^{29,30}:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (6)$$

where: true positive (TP) is the number of host samples correctly identified as positive for the assayed marker, false positive (FP) is the number of non-target samples incorrectly identified as positive for the assayed marker, true negative (TN) is the number of non-target host samples correctly identified as negative for the assayed marker, and false negative (FN) is the number of non-target samples incorrectly identified as negative.

Detection of host-specific Bacteroidales in water sources and household drinking water

The extracted DNA was subjected to quantitative PCR (qPCR) to determine the source of faecal contamination (human or non-human) by using genetic markers. There were five different types of particular markers: *BacHum* for humans, *BacCow* for cattle, *Pig-2-Bac* for pigs, *Cytb* for chickens, and *BacCan* for dogs. The markers, primers, probes, and cycle conditions for the host-specific markers employed in qPCR are shown in Table S2 (Supplementary material). To provide 20 µL of the PCR mixture, 2 µL of template DNA, 10 µL of Luna™ Universal Probe qPCR Master Mix (1X), 0.4 µL of *BacHum* specific probe (0.2 µM), 0.8 µL of each *BacHum* specific primer (0.4 µM) and 6 µL of nuclease-free water. A negative control containing nuclease-free water, master mix, and extraction blank (from DNA extraction procedure) was also used.

Statistical analysis

Microsoft Excel 365 was used to analyse the coded questionnaire data; frequencies of the demographic and socio-economic data were obtained from the questionnaire and presented in Table 1. A Statistical Package for the Social Sciences tool (SPSS) (version 27) was used to carry out a regression model test to determine an association between the survey results (animals grazing around water sources) and the qPCR results (source of faecal pollution detected) where R^2 was used to measure the strength of the relationship between the model and the dependent variable (source of faecal pollution) on a convenient 0–100% scale.

Results and discussion

Study survey

Population demographic information and socio-economic status during the study period

As can be seen in Table 1, overall, in terms of population demographics and socio-economic status of the selected households of both local municipalities, a similar observation was noted regarding the age and gender: fewer children in the age group of 0–5 years and higher number of individuals in the age group of 26–55 years of which females occupied the highest rate compared to males. Findings of the survey also revealed more than 50% of rural dwellers of the selected households in TLM completed their basic education compared to those in CLM, where the rate of non-educated individuals was up to 51.7%. The unemployment rate was also found to be high in both local municipalities, which also implied high rates of self-employed individuals (focusing on small activities to supply the needs to their families) or old people totally depending on retirement funds. The socio-economic situation has been recurrent in most rural areas of South Africa for many decades. According to Statistics South Africa³¹, 7.1 million people were unemployed in the first quarter of 2020, representing 30.1% of the country's overall unemployment rate.

Water source and sanitation status of the household study cohort

Lack of access to safe drinking water is among other challenges facing the target rural dwellers of the TLM and Collins CLM. More than 30% of households depend on yard taps (39.3%), and community standpipes (32.1%) as their main water supply systems, while others still rely on untreated spring water [CLM (28.6%), and TLM (33.3%)], untreated river water [TLM (16.7%)] and hand-dug well water [TLM (13.0%)] (Table 2), especially during the shortage of treated water supplied by municipalities. This shows that an effort is still required to improve water supply systems for rural communities as most rural areas in the VDM are facing challenges of intermittent water supply, which might be one of the reasons for the direct use of untreated water sources. Several studies have reaffirmed that non-functional water pumps, theft, and vandalism are some of the challenges that affect the supply of potable water in rural areas^{32,33}. As can be seen in Figs. 2a and Fig. 2b, the storage of water in plastic containers is a common practice frequently used by the rural communities of TLM and CLM to counteract shortage of water for drinking and cooking in dwellings.

Findings of the present study also revealed several human and animal activities that contribute to the deterioration of water sources (Fig. 2e–i). Furthermore, grazing animals have access to these water sources, as

Household respondents		Local municipalities			
		CLM		TLM	
		Number	Frequency	Number	Frequency
Age range	0–5	18	12.9%	28	11.6%
	6–15	34	24.3%	46	19.1%
	16–25	23	16.4%	42	17.4%
	26–55	45	32.1%	78	32.4%
	56 +	20	14.3%	47	19.5%
		n = 140		n = 241	
Gender	Males	58	41.4%	108	44.8%
	Females	82	58.6%	133	55.2%
		n = 140		n = 241	
Highest level of education	Basic ¹	82	58.6%	89	36.9%
	Advanced ²	26	18.6%	28	11.6%
	No schooling	32	22.9%	124	51.5%
		n = 140		n = 241	
Employment rate	Employed	13	9.3%	34	14.1%
	Unemployed	38	27.1%	61	25.3%
	*Other	89	63.6%	146	60.6%
		n = 140		n = 241	

Table 1. Population demographic and socio-economic status in Collins Chabane and Thulamela local municipalities among the participating households. ¹Basic education: Primary and Secondary education. ²Advanced education: Higher certificates, Diplomas, and degrees. *self-employed or retired.

Household respondents		Local Municipalities			
		CLM		TLM	
		Number	Frequency	Number	Frequency
Water sources	Yard taps	11	39.3.0%	20	37.0%
	Community standpoints	9	32.1%	0	0%
	Roof harvested rainwater	0	0%	0	0%
	River/stream	0	0%	9	16.7%
	Spring	8	28.6%	18	33.3%
	Borehole	0	0%	0	0%
	Dug well	0	0%	7	13.0%
		n = 28		n = 54	
Animals grazing around water sources	Cattle	13	46.4%	28	51.9%
	Chickens	6	21.4%	20	37.0%
	Pigs	4	14.3%	0	0
	Dogs	5	17.9%	6	11.1%
		n = 28		n = 54	
Sanitation	Pit latrine (unimproved)	22	78.6%	50	92.6%
	Pit latrine with concrete slab	5	17.9%	2	3.7%
	Septic/Flush	0	0%	2	3.7%
	VIP ¹	0	0%	0	0%
	Open defecation	1	3.6%	0	0%
		n = 28		n = 54	

Table 2. Water sources and sanitation in Collins Chabane and Thulamela local municipalities. ¹VIP (Ventilated improved pit latrine): Improved.

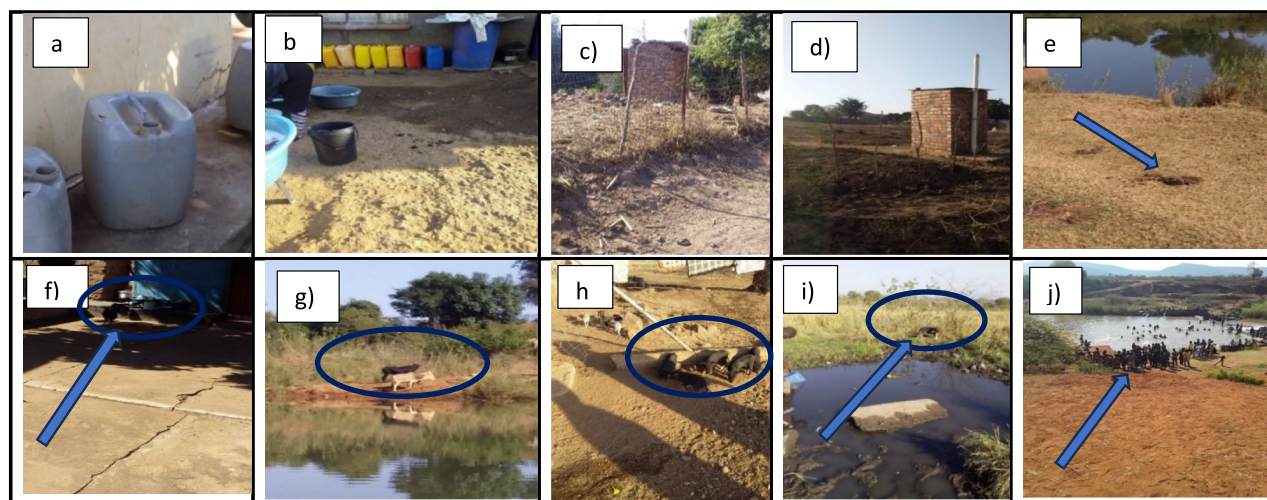


Figure 2. Water storage container in TLM and CLM (a, b), Sanitation (c, d), cattle dung (e), Chicken around household (f), cattle grazing (g), pig around household (h), dog around household (i) recreation (j).

unanimously reported by the majority of the dwellers (82.1–98.2%) in both local municipalities and thus water sources are not protected from contamination. Basic sanitation is considered as the cost-effective technology that enables hygienic excreta disposal³⁴. It includes sanitation facilities such as pour-flush latrines, simple pit latrines, ventilated improved pit latrines and septic/flush facilities. In South Africa, a total of 18 million people lack access to improved sanitation facilities³⁵. This corroborates the findings by the current study, which revealed that most of the selected households had access to unimproved pit latrines (CLM—78.6%;TLM—92.6%) (Table 2, Fig. 2c and Fig. 2d) and none of the dwellers had the VIP latrine; and also, open defecation practice was reported in CLM (3.6%).

Hygiene practices and perception of health during the study period

Hygiene practice plays a significant role in the prevention of diseases associated with water and sanitation; and if thoroughly practised, it can reduce 65% of deaths caused by diarrhoea³. In this study, a hygiene practice such as

washing hands most of the time before handling food, after visiting the toilet and after changing a baby diaper was reported in both municipalities at 47.9% and 37.3% by dwellers in CLM and in TLM (Table S3) (Supplementary material), respectively. Compared to rural areas in other countries such as Bangladesh where the frequencies of washing hands are up to 87.34% before handling food and 95.34% after defecation³⁶, this good hygiene practice remains very low among the dwellers in TLM and in CLM.

Prevalence of faecal markers from water sources to the point of use (households)

Amplification efficiency, lower limit of quantification and performance of the host-specific Bacteroidales markers genes

Previous investigators have pointed out that geographic variations influence the performance of host-specific markers^{29,37}. As a result, the amplification efficiency, lower limit of quantification and performance of the host-specific Bacteroidales marker genes were determined (Table 3). Results revealed that all the host-specific Bacteroidales marker genes used had high qPCR efficiencies ranging from 85 to 98% (Table S4 and Figure S1(Supplementary material)). The chicken *Cytb* and the pig (*Pig-2-Bac*) used in this study had a higher specificity of 100%, these results have indicated that the markers were effective in detecting their specific hosts. The highest specificity of marker gene *Cytb* (100%) in this study, was analogous to that reported in a study conducted at Kathmandu Valley, Nepal³⁸, furthermore, the high specificity of marker gene *Pig-2-Bac* in this study (Table 3) was also similar to what was reported in a study (*Pig-2-bac* (86%) by Odagiri et al.³⁹.

The dog (*BacCan*) had a high specificity of 94% in this study compared to that (*BacCan*, 47%) reported in a study conducted in the Peruvian Amazon⁴⁰. A study by Ahmed et al.⁴¹ has shown that *BacCow* had a high specificity of 96.6% for detection of cow fecal contamination in water samples which was in contrast with the lower specificity of the marker found in this study (56%) (Table 3). Lower specificity was also found for *BacHum* (56%); which was opposite to what was reported elsewhere (*BacHum*, 66%)^{39,40}. The low specificity of *BacCow* and *BacHum* may be improved by increasing sample size and diversity as Green et al.⁴² has proven that larger datasets can improve marker specificity. Using a combination of markers can enhance specificity. For instance, combination of different markers targeting the same host may lower false positives and improve specificity⁴¹. The sensitivity of the human marker (*BacHum*) in this study was 78%, which was similar with the (80%) of sensitivity of the marker stated in a study by⁴⁰. All the remaining marker genes (*BacCan*, *Cytb*, *Pig-2-Bac*, and *BacCow*) had sensitivities as high as 100%, the same results were observed in a study by Malla et al.³⁸ for markers *BacCow* and *Pig-2-bac* (100%).

Distribution of host-specific Bacteroidales genetic markers in surface water prior to and after treatment

In terms of source of faecal contamination from water sources to end-user (Table S5-supplementary material), *BacHum* and *BacCow* were prevalent in the water sources (Luvuvhu River upstream and downstream) during both seasons in CLM (100% for both marker genes during wet season and 75% for *BacHum* and 50% for *BacCow* during the dry season). These markers were distributed equally (100%) at the point of abstraction of the water treatment plant (WTP) during the wet season, while it was down to 50% for *BacHum* and 100% for *BacCow* during the dry season. None of the target host-specific Bacteroidales genetic markers was detected in the treated water of the WTP during the study period. However, at the household level, *BacCan* and *BacCow* were detected during the wet season (3.5% and 8.8%, respectively) and during the dry season (5% for both these marker genes) in the treated water prior to storage. Their prevalence increased in household water storage containers with the detection of all target host-specific Bacteroidales marker genes (*BacHum*, *Cytb*, *BacCan* and *BacCow*) except for *Pig-2-Bac* during both seasons. In TLM, the presence of host-specific Bacteroidales marker genes in downstream and upstream water samples of the Mvudi River, Luvuvhu River and Nandoni Dam was also evidenced, with *BacHum* and *BacCow* predominating both at a rate of 100% during the wet season, and at rates of 75% and 50% during the dry season, respectively. While the same host-specific Bacteroidales markers genes could be detected (100%) at the abstraction site of the WTP during both seasons, none of them were detected in the final treated water of the WTP. In contrast, at the household level, *BacCan* and *Cytb* were detected, but at a low detection rate during the wet season (both 5%) compared to *BacCow* (15%) in treated water prior to storage, and during

		<i>BacHum</i>	<i>Bac-Cow</i>	<i>Pic2Bac</i>	<i>CytB</i>	<i>Bac-Can</i>
Source	No of sample	Positive sample	Positive sample	Positive sample	Positive sample	Positive sample
Amplification of <i>Bacteroidales</i> host-specific markers						
Human	49	38 (77.6%)	28 (57.1%)	0 (0%)	0 (0%)	0 (0%)
Cow	49	37 (75.5%)	49 (100%)	0 (0%)	0 (0%)	0 (0%)
Pig	3	0 (0%)	0 (0%)	3 (100%)	3 (100%)	0 (0%)
Chicken	3	0 (0%)	0 (0%)	1 (33.3%)	3 (100%)	2 (66.7%)
Dog	7	2 (28.6%)	1 (14.3%)	0(0%)	0 (0%)	7 (100%)
Performance of the host-specific genetic for specificity, sensitivity and accuracy (%)						
Specificity		56	56	100	100	94
Sensitivity		78	100	100	100	100

Table 3. Performance of the host-specific Bacteroidales marker genes.

the dry season, the *BacCan* marker was detected at a very low rate (3.6%), compared to that of *BacCow* (11.3%). The *BacHum* and *BacCan* marker genes were present during the wet season at higher rates (both 15%) in HH water storage containers, and during the dry season, these two marker genes were apparent with low detection rates of 5% (Table S5-supplementary material). In rural South Africa, cows and humans commonly share water sources like rivers for drinking, recreation, fishing, washing clothes, and bathing. Similar practices were observed in CLM and TLM surface water during the study period. Previous studies^{6,7} indicate that cattle and humans are major sources of fecal contamination in Limpopo Province. Detection frequencies for *BacHum*, *BacCow*, and *BacCan* were higher in surface water during the wet season compared to the dry season, similar to findings in other studies^{43,44}, including markers detected at WTP abstraction points⁴⁵. *Pig-2-Bac* was not detected in surface water in this study but was found in other studies^{12,46} however, it was absent in WTP treated water in all studies. The high prevalence of animal markers in household water storage containers aligns with studies in Bangladesh and Tanzania, which reported high ruminant pollution in urban informal settlements in household drinking water of urban informal settlements^{47,48}.

Distribution of host-specific Bacteroidales genetic markers in untreated surface water stored in household containers Results in Table 4 revealed that, the TLM residents depend on the Mutshindudi River water. Regardless of the seasons and the sampling points, the water from this river was shown to be contaminated by both faeces of human and animal origin, as evidenced by the prevalence of their corresponding host-specific Bacteroidales genetic markers such as *BacHum*, *BacCan* and *BacCow* during both seasons. The detection rates of these faecal genetic markers ranged between 50 and 100% with the highest rates during the wet season. Variations (50–75%) in their detection rate were noted during the dry season. Despite the low detection rates of faecal marker genes at household level, the prevalence of *BacHum*, *BacCan* and *Cytb* should not be underestimated during the wet season (22.2%, 44.4% and 30.6%, respectively), and during the dry season (19.4% for both *BacHum* and *BacCan* marker genes and 25% for *Cytb*). The results obtained in this scenario have shown that human open defecation and animal faeces could be considered as the polluters of this water source and thus corroborate the findings of the study by⁴⁹, who also detected human, cattle and dog faecal pollution in surface water (rivers), a study by Harwood et al.⁵⁰ have detected *BacHum* in environmental water samples with different frequencies (60–80%) on areas affected by human activities. *BacCan* was also detected in 50–75% of samples in urban settings impacted by high dog population, this was similar to what was detected in this study³⁰. Furthermore, the *BacCow* was detected in areas associated with cattle farming activities, this corroborates with the results in this study⁵¹. The detection rates for markers *BacHum*, *BacCow*, *BacCan*, *Cytb* and *Pig-2Bac* in this study vary from markers detected in other studies at household level^{41,52,53}, this could be as a result of regional variations or differences in sample collection.

Distribution of host-specific Bacteroidales genetic markers in untreated spring water stored in household containers Protection of groundwater sources from faecal contamination in rural areas is vital, as they are regarded as the most important sources of water used for household domestic purposes⁵⁴. As stipulated in Table 5, more than 28.6% of the household study cohort in CLM totally depends on spring water. However, this water source was characterised by the prevalence of animal-specific Bacteroidales marker genes such as *Pig-2-Bac*, *BacCan* and *BacCow* during both seasons with a detection rate of 100% during wet season and 50–75% during the dry season. With the exception of human-specific Bacteroidales marker genes, the prevalence of all the animal marker genes was also observed at the level of household with *BacCow* exhibiting the highest rate compared to other Bacteroidales marker genes (Table 5). In TLM, only *BacHum* and *BacCow* were detected during both seasons with a detection rate of 50% and 100%, respectively, in the rainy season, and 25% and 62.5%, respectively, in the dry seasons. At household level, a similar observation as in CLM was observed in terms of the prevalence of all the markers during the wet season (16.7% for *BacHum*, *Cytb*, *Pig-2-Bac*, and *BacCow*, and 5.6% for *BacCan*) and during the dry season (13.9% for *BacHum*, 11.1% for *Cytb*, and 5.6% for *BacCan* marker genes), with the exception of *Pig-2-Bac*. In rural Kisumu, Kenya⁵⁵, detected markers *BacHum*, *BacCan*, and *BacCow* in spring and household water samples with low frequencies: *BacHum* (10% in spring, 20% in household), *BacCan* (15% in spring, 25% in household), and *BacCow* (5% in spring, 10% in household). Similarly⁵⁴, reported low detection frequencies of these markers in rural India and Nepal. In contrast, this study found higher detection frequencies

TLM											
			Wet season				Dry season				
			Frequency (no. of positive samples/no. of tested samples)								
			Host-specific Bacteroidales								
			n = 44 n = 44								
			Total n = 88								
Sampling site	Number of samples	BacHum	Cytb	Pig-2-Bac	BacCan	BacCow	BacHum	Cytb	Pig-2-Bac	BacCan	BacCow
Mutshindudi river downstream	4	4 (100%)	0 (0%)	0 (0%)	4 (100%)	4 (100%)	2 (50 %)	0 (0%)	0 (0%)	4 (100%)	3 (75%)
Mutshindudi river upstream	4	4 (100%)	0 (0%)	0 (0%)	4 (100%)	4 (100%)	3 (75%)	0 (0%)	0 (0%)	2 (50%)	3 (75%)
HH* (container-stored water)	36	8 (22.2%)	16 (44.4%)	0 (0%)	11 (30.6%)	0 (0%)	7 (19.4%)	9 (25%)	0 (0%)	7 (19.4%)	0 (0%)

Table 4. Prevalence of host-specific Bacteroidales marker genes from untreated surface water sources to household (HH) stored water. *HH: Household.

CLM											
		Wet season					Dry season				
		Frequency (no. of positive samples/no. of tested samples)									
		Host-associated Bacteroidales									
Sampling site	Number of samples	BacHum	Cytb	Pig-2-Bac	BacCan	BacCow	BacHum	Cytb	Pig-2-Bac	BacCan	BacCow
	n = 36						n = 36				
Total n = 72											
Spring	4	0 (0%)	0 (0%)	4 (100%)	4 (100%)	4 (100%)	0 (0%)	0 (0%)	2 (50%)	2 (50%)	3 (75%)
HH container-stored water	32	0(0%)	12 (37.5%)	20 (62.5%)	12 (37.5%)	12 (37.5%)	0 (0%)	9 (28.1%)	11 (34.4%)	8 (25%)	8 (25%)
TLM											
	n=40						n=40				
Total n = 80											
Spring	4	2 (50%)	0 (0%)	0 (0%)	0 (0%)	4 (100%)	1 (25%)	0 (0%)	0 (0%)	0 (0%)	2 (50%)
HH* (Container-stored water)	36	6 (16.7%)	6 (16.7%)	6 (16.7%)	2 (5.6%)	6 (16.7%)	5(13.9%)	24 (11.1%)	0 (0%)	2 (5.6%)	0 (0%)

Table 5. Prevalence of host-specific Bacteroidales marker genes from untreated spring water source to household stored water. *HH: Household.

of *BacHum*, *Pig-2-Bac*, *BacCan*, and *BacCow*, due to variations in environmental conditions, population density, and pollution sources.

Distribution of host-specific Bacteroidales genetic markers in untreated hand-dug well water stored in household containers

In this scenario, where some of the household study cohorts in TLM depend on hand-dug wells, the overall water samples tested positive for animal-specific Bacteroidales marker genes (*Cytb*, *BacCow* and *BacCan*) during both seasons. The detection rate of these marker genes was 100% during the wet season, while during the dry season, it ranged between 50% (for both *Cytb* and *BacCow*) and 75% for *BacCan*. The storage of the hand-dug well water at household level resulted in the appearance of *BacCan* and *Cytb* at lower detection rates during the wet season (28.9%, and 42.9%, respectively) and during the dry season (14.3% for *BacCan* and 28.5% for *Cytb* marker genes) compared to the detection frequencies for these same marker genes in the non-stored well water (100% detection rate for *BacCan* and *Cytb* in the wet season, and detection rate of 75% for *BacCan* and 50% for *Cytb* during the dry season) (Table S6-supplementary material). These findings demonstrated the prevalence of animal faecal pollution in the hand-dug well water source and during storage. They are in agreement with the findings reported in remote rural areas of Pueblo Nuevo, Nicaragua and Bangladesh, where animal faecal matter was prevalent in household water storage containers and in hand-dug wells^{56,57}. In this study, *BacHum* and *BacCow* were not detected in household container-stored water samples in both seasons, similar to findings by Johnson et al.¹². However, *Pig-2-Bac* was detected in 100% of these samples, higher than the 80–95% range reported in other studies^{12,46}.

Association between animal grazing around water sources and source of faecal contamination

An association was established between animals grazing around water sources and the source of faecal pollution in water used by the communities; however, the association ranged from moderate (R^2 of 0.534—cattle and *BacCow*) to very weak (R^2 of 0.1062—chicken and *Cytb*) as presented in Table 6 for both municipalities during the wet season and during the dry season. The association also ranged from weak (R^2 of 0.320—cattle and *BacCow*) to very weak (R^2 of 0.138—chicken and *Cytb*). Furthermore, the association determined between animals grazing around water sources and the specific Bacteroidales marker genes in both municipalities was shown to have a statistical significance with p-values < 0.05 in most of the cases while no statistical significance was observed between dog faecal pollution and the marker gene *BacCan*. Raw data (Table S7-supplementary material).

Conclusion

Microbial source tracking techniques using host- specific Bacteroidales marker genes have been used in the USA, European countries, and other developing countries such as Bangladesh and Kenya. The application of these techniques in rural areas of CLM and TLM allowed an understanding of the origin of faecal pollution of water sources and the distribution of these marker genes in water sources used by the communities. The findings of this study revealed the predominance of *BacHum* and *BacCow*, which demonstrates that cows and humans are the main sources of faecal pollution of the water sources in both municipalities with *BacHum* and *BacCow* detection frequencies ranging from 50 to 100%. Chicken, dog, and pig faecal pollution was the most prevalent in household drinking water. The absence of all the target marker genes in treated water of WTP indicates the effectiveness of the water treatment processes, which result in the production of safe drinking water. However, when stored at the households, this treated water revealed faecal contamination from different origins, which could be due to poor hygiene and lack of improved sanitation facilities. The findings in TLM have displayed the predominance of animal faecal pollution in hand-dug well water during both seasons, *BacCow*, *BacCan*, and *Cytb* were the main

Animal host and host- specific Bacteroidales markers	Standardised coefficient beta	Wet season	R ²	Adjusted R ²	P value	Mean square	t
		Standard error					
Cattle	0.043	0.016	0.5349	0.126	0.009	575.172	4.211
<i>BacCow</i>							
Chicken	0.053	0.0193	0.1062	0.197	0.015	575.142	4.349
<i>Cytb</i>							
Pig	0.589	0.018	0.1425	0.150	0.002	621.150	3.21
<i>Pig-2-Bac</i>							
Dog	0.050	0.014	0.3138	0.033	0.001	231.402	1.341
<i>BacCan</i>							
Dry season							
Cattle	0.264	1.807	0.320	− 0.011	0.023	16.535	0.480
<i>BacCow</i>							
Chicken	0.241	1.547	0.138	0.045	0.039	502.802	2.106
<i>Cytb</i>							
Pig	0.402	0.973	0.141	− 0.010	0.003	20.917	0.527
<i>Pig-2-Bac</i>							
Dog	0.062	1.261	0.204	0.120	0.600	423.134	3.436
<i>BacCan</i>							

Table 6. Association between animals grazing around water sources and faecal pollution detected in water sources and household drinking water.

markers. In household container-stored water, chicken and dog fecal pollution were most prevalent. In CLM, the spring water source displayed animal fecal pollution (*BacCow*, *BacCan*, and *Pig2-Bac*). Household drinking water exhibited high levels of cow, chicken, dog, and pig fecal contamination. In TLM spring water, only *BacHum* and *BacCow* were predominant. However, at the household level, drinking water quality was compromised by a wide range of animal fecal pollutants (chicken, dog, cow, pig) and human fecal pollution. These findings have reflected that animal fecal pollution is a significant concern for both hand-dug wells and spring water sources, more especially at the household level, affecting the quality of drinking water. This study, therefore, suggests the implementation of a robust integrated water and sanitation management plan for the protection of various water sources from the point of treatment or point of collection to the point of use at household level.

Data availability

The datasets degenerated during this study are available in the supplementary material.

Received: 2 April 2024; Accepted: 29 July 2024

Published online: 26 August 2024

References

- Chirgwin, H. *et al.* Interventions promoting uptake of water, sanitation and hygiene (WASH) technologies in low- and middle-income countries: An evidence and gap map of effectiveness studies. *Campbell Syst. Rev.* **17**, 1194. <https://doi.org/10.1002/cl2.1194> (2021).
- Ritchie, H.M. & Roser, M. "Clean Water and Sanitation". Published online at Our World In Data.org. '<https://ourworldindata.org/clean-water-sanitation>' (Access 03/03/2023) (2021).
- Vazsonyi, A., Wittekind, J., Belliston, L. & Loh, T., Unodc. Global Study on Homicide. *Unodoc* **20**, 1–125. <https://doi.org/10.1023/B:JOQC.0000037731.28786.e3> (2014).
- Bianco, K. *et al.* Possible health impacts due to animal and human fecal pollution in water intended for drinking water supply of Rio de Janeiro, Brazil. *J. Water Supply Res. Technol.*—AQUA **69**, 70–84. <https://doi.org/10.2166/aqua.2019.061> (2020).
- Moropeng, R. C. & Momba, M. N. B. Assessing the sustainability and acceptance rate of cost-effective household water treatment systems in rural communities of makwane village, south africa. *Crystals* **10**, 1–16. <https://doi.org/10.3390/cryst10100872> (2020).
- Obi, C. L., Potgieter, N., Bessong, P. O. & Matsaung, G. Assessment of the microbial quality of river water sources in rural Venda communities in South Africa. *Water SA*. **28**, 287–292. <https://doi.org/10.4314/wsa.v28i3.4896> (2002).
- Traoré, A. N. *et al.* The impact of human activities on microbial quality of rivers in the Vhembe District SouthAfrica. *Int. J. Environ. Res. PublicHealth* <https://doi.org/10.3390/ijerph13080817> (2016).
- Feng, S., Ahmed, W. & McLellan, S. L. Ecological and technical mechanisms for cross-reaction of human fecal indicators with animal hosts. *Appl. Environ. Microbiol.* **86**, 02319–02419. <https://doi.org/10.1128/AEM.02319-19> (2020).
- Van der Merwe-Botha, M (2009) Water Quality: A vital dimension of water security Development Planning Division Working Paper Series No. 14. Dbsa 32 (2016).
- Aduana, T., Maribeng, R. & Ngene, B. Microbial water quality of selected water sources in the Vhembe District, South Africa. *Afri. J. of Microbiol. Res.* **8**, 245–253. <https://doi.org/10.5897/AJMR2013.6003> (2014).
- Tawfik, A., El-Gohary, F. & Amin, M. Assessment of water quality and pollution in the Vhembe District. *Water SA* **41**, 45–50. <https://doi.org/10.4314/wsa.v41i1.5> (2015).
- Johnson, T., Mugisha, E. & Akampulira, E. Seasonal variations of microbial indicators in Sub-Saharan African dug wells. *Water Res.* **170**, 115302 (2020).
- DFED (Department of finance and Economic Development). Limpopo State of the environment report (Phase 1). Limpopo Department of Finance and Economic Development, South Africa (2004).
- Pruss, A. Review of epidemiological studies on health effects from exposure to recreational water. *Int. J. Epidemiol.* **27**, 1–9 (1998).

15. Barnes, B. & Gordon, D. M. Coliform dynamics and the implications for source tracking. *Environ. Microbiol.* **6**, 501–509. <https://doi.org/10.1111/j.1462-2920.2004.00591.x> (2004).
16. Bergholz, P. W. *et al.* Environmental patterns are imposed on the population structure of *Escherichia coli* after fecal deposition. *Appl. Environ. Microbiol.* **76**, 211–219. <https://doi.org/10.1128/AEM.01880-10> (2010).
17. Noble, R. T. *et al.* Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination in a microbial source tracking comparison study Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination. <https://doi.org/10.2166/wh.2003.0021> (2004).
18. Staley, Z. R., Grabuski, J., Sverko, E. & Edge, T. A. Comparison of Microbial and Chemical Source Tracking Markers To Identify Fecal Contamination Sources in the Humber River (Toronto, Ontario, Canada) and Associated Storm Water Outfalls *Appl.* **82**, 6357–6366. <https://doi.org/10.1128/AEM.01675-16> (2016).
19. Wexler, H. M. Bacteroides: The good, the Bad, and the Nitty-Gritty. *Clin. Microbiol. Rev.* **20**, 593–621. <https://doi.org/10.1128/CMR.00008-07> (2007).
20. Bernhard, A. E. & Field, K. G. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* **66**, 1587–1594. <https://doi.org/10.1128/AEM.66.4.1587-1594.2000> (2000).
21. Jenkins, M. W., Tiwari, S., Lorente, M., Maina, C. & Wuertz, S. Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays. *Water Res.* **63**, 125–132 (2014).
22. Pickering, A. J., Julian, T. R., Mamuya, S., Boehm, A. B. & Davis, J. Bacterial hand contamination among Tanzanian mothers varies temporally and following household activities 16, 233–239. <https://doi.org/10.1111/j.1365-3156.2010.02677.x> (2011).
23. Ferguson, A. S. *et al.* Comparison of fecal indicator bacteria in ground water from southern Bangladesh and eastern India. *Sci. Total Environ.* **441**, 314–322. <https://doi.org/10.1016/j.scitotenv.2012.05.060> (2012).
24. Cao, Y., Raith, M. R. & Griffith, J. F. Science Direct Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. *Water. Res.* **70**, 337–349. <https://doi.org/10.1016/j.watres.2014.12.008> (2014).
25. APHA. *Standard Methods for the Examination of Water and Wastewater* (20th Edition). Published jointly by the American Public Health Association (APHA), Water Environment Federation (WEF) and American Water Works Association (AWWA), Washington DC, USA. 254–278 (2001).
26. Kralik, P. & Ricchi, M. A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything. *Front. Microbiol.* **8**, 1–9. <https://doi.org/10.3389/fmicb.2017.00108> (2017).
27. Mieszkina, S., Furet, J. P., Corthier, G. & Gourmelon, M. Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. *Appl. Environ. Microbiol.* **75**, 3045–3054. <https://doi.org/10.1128/AEM.02343-08> (2009).
28. Zhuang, F. F. *et al.* Quantitative detection of fecal contamination with domestic poultry feces in environments in China. *AMB Express* <https://doi.org/10.1186/s13568-017-0379-0> (2017).
29. Ballesté, E. & Meijer, W. G. Identifying the biological and geographical origin of faecal contamination. <https://doi.org/10.13140/RG.2.1.3047.3442> (2015).
30. Kildare, B. J. *et al.* 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: A Bayesian approach. *Water Res.* **41**, 3701–3715 (2007).
31. Stats SA (Statistics South Africa). Quarterly Labour Force Survey. Quarter 1: 2020. Statistical release PO211. P02111stQuarter2020. pdf (statssa.gov.za) (Accessed 28/10/2023) (2020).
32. Johannessen, Å., Rosemarin, A. & Thomalla, F. International journal of disaster risk reduction strategies for building resilience to hazards in water, sanitation and hygiene (W bnb bbnASH) systems: The role of public private partnerships. *Int. J. Disaster Risk Reduct.* <https://doi.org/10.1016/j.ijdr.2014.07.002> (2014).
33. Rural Water Supply Network. Handpump data: Selected countries in sub-Saharan Africa. In: Harvey, P. (Ed.), Sustainable Rural Water Supplies (2009).
34. WHO/UNICEF (World Health Organization & United Nations Children's Fund). (Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines. World Health Organization. <https://apps.who.int/iris/handle/10665/258617> (Accessed 27/09/2023) (2017).
35. Akpakli, D. E., Manyeh, A. K., Akpakli, J. K., Kukula, V. & Gyapong, M. Determinants of access to improved sanitation facilities in rural districts of southern Ghana: Evidence from Dodowa health and demographic surveillance site. *BMC Res. Notes* <https://doi.org/10.1186/s13104-018-3572-6> (2018).
36. Imtiaz, K. S. *et al.* Practice of personal hygiene among rural women of a selected community in Bangladesh. *North. Int. Med. Coll. J.* **6**, 29–31 (2014).
37. Gawler, A. H. *et al.* Validation of host-specific Bacteroidales 16S rRNA genes as markers to determine the origin of faecal pollution in Atlantic Rim countries of the European Union. *Water Res.* **41**, 3780–3784. <https://doi.org/10.1016/j.watres.2007.01.028> (2007).
38. Malla, B., Yamamoto, K., Furukawa, K. & Haramoto, E. Performance evaluation and application of host-specific Bacteroidales and mitochondrial DNA markers to identify sources of fecal contamination in river water in Japan. *PLoS Water* **3**, e0000210 (2024).
39. Odagiri, M. *et al.* Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Sci. Total Environ.* **502**, 462–470. <https://doi.org/10.1016/j.scitotenv.2014.09.040> (2015).
40. Schiaffino, F. *et al.* Validation of microbial source tracking markers for the attribution of fecal contamination in indoor-household environments of the Peruvian Amazon. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2020.140531> (2020).
41. Ahmed, W., Harwood, V. J., Gyawali, P., Sidhu, J. P. & Toze, S. Comparison of concentration methods for quantitative detection of sewage-associated viral markers in environmental waters. *Appl. Environ. Microbiol.* **78**, 7770–7779 (2012).
42. Green, H. C. *et al.* Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. *Appl. Environ. Microbiol.* **80**, 3086–3094 (2014).
43. Martinez, L., Johnson, R. & Patel, A. Seasonal dynamics of microbial indicators in rural water sources. *Water Res.* **150**, 360–370. <https://doi.org/10.1016/j.watres.2018.12.025> (2024).
44. Nguyen, V., Brown, L. & Miller, R. Impact of seasonal variation on microbial contamination in surface water. *Sci. Total Environ.* **764**, 144321. <https://doi.org/10.1016/j.scitotenv.2020.144321> (2021).
45. Brown, K., Smith, J. & Nguyen, T. Microbial contamination at water treatment plant abstraction points. *J. Environ. Sci.* **85**, 45–52. <https://doi.org/10.1016/j.jes.2020.04.005> (2020).
46. Smith, J., Patel, R. & Singh, K. Microbial contamination in rural Indian water sources. *J. Water Health* **16**, 547–555 (2018).
47. Boehm, A. B. *et al.* Occurrence of host-associated fecal markers on child hands, household soil, and drinking water in rural Bangladeshi households. *Environ. Sci. Technol. Lett.* **3**, 393–398 (2016).
48. Harris, A. R., Boehm, A. B. & Davis, J. Mechanisms of post-supply contamination of drinking water in Bagamoyo. *Tanzania. J. Water. Health.* **11**, 543554. <https://doi.org/10.2166/wh.2013.023> (2013).
49. Schriewer, A. *et al.* Presence of Bacteroidales as a Predictor of Pathogens in Surface Waters of the Central California Coast 76: 5802–5814. <https://doi.org/10.1128/AEM.00635-10> (2010).
50. Harwood, V. J., Staley, C., Badgley, B. D., Borges, K. & Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* **38**, 1–40 (2014).

51. Shanks, O. C. *et al.* Performance assessment PCR-based assays targeting Bacteroidales genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **76**, 1359–1366 (2010).
52. Harwood, V. J. *et al.* Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. *Water Res.* **47**, 6812–6822 (2013).
53. Layton, A. *et al.* Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Res.* **47**, 6897–6908 (2013).
54. Shrestha, S. *et al.* Dynamics of domestic water consumption in the Urban area of the Kathmandu valley: Situation analysis pre and post 2015 Gorkha Earthquake. *Water (Switzerland)* **9**, 1–17. <https://doi.org/10.3390/w9030222> (2017).
55. Odagiri, M. *et al.* Water quality and human health in the Kisumu region of Kenya. *Environ. Sci. Technol.* **50**, 8260–8269. <https://doi.org/10.1021/acs.est.6b01208> (2016).
56. Ercumen, A. *et al.* Animal feces contribute to domestic fecal contamination: Evidence from *E. coli* measured in water, hands, food, flies, and soil in Bangladesh. *Environ. Sci. Technol.* **51**, 8725–8734. <https://doi.org/10.1021/acs.est.7b01710> (2017).
57. Malla, B. *et al.* Validation of host-specific Bacteroidales quantitative PCR assays and their application to microbial source tracking of drinking water sources in the Kathmandu Valley Nepal. *J. Appl. Microbiol.* **125**, 609–619. <https://doi.org/10.1111/jam.13884> (2018).

Acknowledgements

Special thanks are expressed to the National Research Foundation for funding this research, Vhembe District Municipality and its rural communities for participating in the research and the University of Venda for assisting on the processing of samples.

Author contributions

Maggy Ndombo Benteke Momba: Supervision, conceptualization and conceived the project; Maggy Ndombo Benteke Momba, Barbara Mogane, Renay Ngobeni-Nyambi, Ilunga Kamika and Leonard Owino Kachienga conceived and designed the experiments; Maggy Ndombo Benteke Momba, Barbara Mogane, Renay Ngobeni-Nyambi, Ilunga Kamika and Leonard Owino Kachienga performed the experiments and analysed data. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by the South African Research Chairs Initiatives (SARChI) in Water Quality and Wastewater Management, funded by the Department of Science and Technology, as administrated by the National Research Foundation (UID87310). Additional funding was received from Tshwane University of Technology.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68771-3>.

Correspondence and requests for materials should be addressed to M.N.B.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024, corrected publication 2024