Letters in Applied Microbiology

Letters in Applied Microbiology ISSN 0266-8254

ORIGINAL ARTICLE

Assessing the faecal source sensitivity and specificity of ruminant and human genetic microbial source tracking markers in the central Ethiopian highlands

R.B. Linke[1], G. Kebede[2,3], D. Mushi[4], A. Lakew[5], D.S. Hayes[3,6], W. Graf[3] and A.H. Farnleitner[1,7]

1 Research Group of Environmental Microbiology and Molecular Diagnostics, Institute for Chemical, Biological and Environmental Engineering, Technical University Vienna, Vienna, Austria
2 Department of Biological Sciences, Ambo University, Ambo, Ethiopia
3 Institute of Hydrobiology and Aquatic Ecosystem Management (IHG), University of Natural Resources and Life Sciences, Vienna, Austria
4 Department of Biosciences, Solomon Mahlangu College of Science and Education, Sokoine University of Agriculture, Morogoro, Tanzania
5 National Fishery and Aquatic Life Research Centre, Ethiopian Institute of Agricultural Research (EIAR), Sebeta, Ethiopia
6 Centro de Estudos Florestais (CEF), Instituto Superior de Agronomia, University of Lisbon, Lisbon, Portugal
7 Research Division Water Quality and Health, Karl Landsteiner University for Health Sciences, Krems, Austria

Significance and Impact of the Study: Faecal water pollution poses a health risk to its users worldwide, especially in developing countries where safe household water supplies and sanitary facilities are often scarce or non-existent. This study tested for the first-time genetic microbial source tracking (MST) methods for identifying ruminant- and human-associated bacterial faecal contaminants in Ethiopia in an extensive regional faecal sample bank. The results indicated the marker's suitability for diverse MST applications in the Ethiopian area. Their application will aid policy makers to make proper risk assessment, take corrective action in the management of surface waters and the implementation of remediation measures supporting public health.

Keywords
assay evaluation, Ethiopia, faeces, human, marker, microbial source tracking, ruminant, soil.

Correspondence
Rita B. Linke, Research Group of Environmental Microbiology and Molecular Diagnostics, Institute for Chemical, Biological and Environmental Engineering, Technical University Vienna, Gumpendorferstrasse 1A, 1060 Vienna, Austria.
E-mail: rita.linke@tuwien.ac.at

Rita B. Linke and Geda Kebede contributed equally to this article.

2020/1684: received 4 September 2020, revised 4 December 2020 and accepted 4 December 2020

doi:10.1111/lam.13436

Abstract

This study tested genetic microbial source tracking (MST) methods for identifying ruminant- (BacR) and human-associated (HF183/BacR287, BacHum) bacterial faecal contaminants in Ethiopia in a newly created regional faecal sample bank (n = 173). BacR performed well, and its marker abundance was high (100% sensitivity (Sens), 95% specificity (Spec), median $\log_{10}$ 8·1 marker equivalents (ME) g^{-1} ruminant faeces). Human-associated markers tested were less abundant in individual human samples (median: $\log_{10}$ 5·4 and 4·2 (ME + 1) g^{-1}) and were not continuously detected (81% Sens, 91% Spec for BacHum; 77% Sens, 91% Spec for HF183/BacR287). Furthermore, the pig-associated Pig2Bac assay was included and performed excellent (100% Sens, 100% Spec). To evaluate the presence of MST targets in the soil microbiome, representative soil samples were tested during a whole seasonal cycle (n = 60). Only BacR could be detected, but was limited to the dry season and to sites of higher anthropogenic influence ($\log_{10}$ 3·0 to 4·9 (ME + 1) g^{-1} soil). In conclusion, the large differences in marker abundances between target and non-target faecal samples (median distances between distributions $\geq\log_{10}$ 3 to $\geq\log_{10}$ 7) and their absence in pristine soil indicate that all tested assays are suitable candidates for diverse MST applications in the Ethiopian area.

Introduction

Microbiological water quality is routinely determined by enumerating faecal indicator bacteria (FIB), a common approach in temperate areas (Farnleitner *et al.* 2010; Bradshaw *et al.* 2016) and also in tropical regions (Byamukama *et al.* 2005; Mushi *et al.* 2010). However, due to their ubiquity in human and animal faeces (Farnleitner *et al.* 2010), source determination without any further analysis steps is not possible. Currently, molecular techniques, known as microbial source tracking (MST) methods, most commonly targeting the 16S rRNA gene of host-associated bacteria, are available and frequently used to identify faecal sources. A plethora of methods to identify faecal contaminants of human and animal origin (e.g., Bernhard and Field 2000; Wuertz *et al.* 2011; Reischer *et al.* 2013; Boehm *et al.* 2013) based on host-associated markers have been published. However, most of these methods are mainly applied in temperate regions and it has been argued that these methods cannot be easily transferred from the region for which they were originally developed to new geographical regions (Reischer *et al.* 2013; Boehm *et al.* 2013; Odagiri *et al.* 2015; Mayer *et al.* 2018).

This study aimed to evaluate for the first time the faecal source sensitivity and specificity of MST methods for the most prevalent faecal contamination sources in Ethiopia (ruminants and humans). BacR (ruminant-associated) was chosen since it has been proven to be one of the best performing and most stable assays to identify ruminant-associated faecal pollution in several studies worldwide (Reischer *et al.* 2013; Boehm *et al.* 2013). BacHum and HF183/BacR287 (both human-associated) were selected since they are among the most cited and best performing assays to identify human-associated pollution (Mayer *et al.* 2018). In addition, HF183/BacR287 has a standardized procedure recognized by the U.S. Environmental Protection Agency (EPA Method 1696; USEPA, 2019). The Pig2Bac assay (pig-associated; Mieszkin *et al.* 2009) was included since it has been shown to be very robust in several regions (e.g., Reischer *et al.* 2013).

Assay evaluation was based on an extensive regional faecal sample bank ($n = 173$) established especially for this study. Soil samples, collected during the dry and wet seasons, were included to evaluate the presence of MST targets in the soil microbiome since their presence could potentially interfere with faecal source tracking results (e.g., Vierheilig *et al.* 2012).

Results and discussion

Qualitative aspects of assay faecal sensitivity and specificity

Binary assay sensitivity and specificity achieved for the ruminant BacR assay were 100% and 95%, respectively. Only 4 of 75 samples (5%) gave false-positive results (1 dog, 2 pigs and 1 pigeon, Table 1). The assay performance observed in this study is consistent with the findings of other studies (Reischer *et al.* 2013; Boehm *et al.* 2013; Malla *et al.* 2018). This might be attributed to the BacR target being strongly correlated with the total intestinal *Bacteroidetes* population, suggesting that they are part of the intestinal core microbiome of ruminants worldwide (Reischer *et al.* 2013). A Spearman rank-order correlation analysis further supported this result by giving a significant correlation between the AllBac and BacR marker ($\rho = 0{\cdot}91$, $p = 0{\cdot}000$). The ruminant BacR assay therefore proved to be a robust ruminant-targeted candidate method to be used in the study area.

Binary assay sensitivity and specificity for the human-associated assays were somewhat lower. For BacHum they were 81% and 91%, while for HF183/BacR287 they were 77% and 91%, respectively (Table 1). While several studies using these assays in various geographic regions are available (e.g., Boehm *et al.* 2013; Reischer *et al.* 2013; Yahya *et al.* 2017; Zhang *et al.* 2020) reports from the

Table 1 Numbers of qPCR positives with the tested assays in source species or source groups

		qPCR positive/total sample number tested			
Source	*n*	BacR	BacHum	HF183/BacR287	Pig2Bac
Ruminants*	57	57/57	0/29	0/29	0/29
Cattle	*33*	*33/33*	*0/33*	*0/33*	*0/33*
Goat	*12*	*12/12*	*0/12*	*0/12*	*0/12*
Sheep	*12*	*12/12*	*0/12*	*0/12*	*0/12*
Human	52	0/14	42/52	40/52	0/13
Herbivores†	21	0/21	0/21	0/21	0/21
Horse	*7*	*0/7*	*0/7*	*0/7*	*0/7*
Donkey	*7*	*0/7*	*0/7*	*0/7*	*0/7*
Mule	*7*	*0/7*	*0/7*	*0/7*	*0/7*
Chicken	8	0/8	0/8	0/8	0/8
Pig	13	2/10	3/10	3/10	13/13
Dog	9	1/9	4/9	4/9	0/9
Wildlife‡	13	1/13	1/13	1/13	0/13
Pigeon	*4*	*1/4*	*0/4*	*0/4*	*0/4*
Monkey	*9*	*0/9*	*1/9*	*1/9*	*0/9*
Specificity (%)§		95	91	91	100
Sensitivity (%)¶		100	81	77	100

n = number of samples tested.

*Ruminant samples: cattle, sheep and goat.

†Herbivore samples: horse, donkey and mule.

‡Wildlife: pigeon and monkey.

§Source-specificity (%) = 100 × true-negatives/(true-negatives + false-positives).

¶Source-sensitivity (%) = 100 × true-positives/(true-positives + false-negatives).

The animals from within these groups (like ruminants, herbivores) are given in italic to seperate them somehow from the main groups.

African continent are rare (Pickering *et al.* 2012; Yahya *et al.* 2017; Bauza *et al.* 2018). Jenkins *et al.* (2009) assessed the performance of BacHum in Kenya, which yielded a binary assay sensitivity and specificity of 18% and 100%, respectively, which is notably lower than the results observed here. However, it has to be noted that the number of samples and sources used for evaluation was limited ($n = 42$) compared to the present study. More recently, Holcomb *et al.* (2020) assessed the assay performance of BacHum and HF183/BacR287 in Mozambique and observed assay sensitivities and specificities of 50 and 64% for BacHum and 81 and 67% for HF183/BacR287.

In the present study, false-positive results (9%) for BacHum and HF183/BacR287 were found in samples from carnivores (dogs; 4/9) and omnivores (pigs; 3/10). No false positives were found in samples from ruminants (cattle, sheep and goats), herbivores (horses, donkeys and mules) and chickens. Of the wildlife samples (pigeons and monkeys), only one was false-positive (a monkey). In contrast with BacR, a Spearman rank-order test did not show any correlation between AllBac and the human-associated markers (BacHum: $\rho = 0{\cdot}02$, $p = 0{\cdot}919$, HF183/BacR287: $\rho = 0{\cdot}11$, $p = 0{\cdot}47$).

Information on cross-reactivity in the literature is diverse. For BacHum, if reported to occur, it was mainly found for samples from deer, dogs, pigs and cows (e.g., Boehm *et al.* 2013; Odagiri *et al.* 2015; Zhang *et al.* 2020) whereas for the HF183/BacR287 assay, it was most commonly observed for samples from dogs, cats, chickens, turkeys and deer (e.g., Green *et al.* 2014; Feng *et al.* 2018; Ahmed *et al.* 2019). The frequently and herein observed cross-reaction with dog and pig samples might be due to the close proximity in which people live with these animals. Feng *et al.* (2020), for example, demonstrated that a cross-reaction was not simply a lack of specificity of the marker gene itself, but that co-habitation also influenced the microbial community structure. Similarly, Zhang *et al.* (2020) reported that frequent contact between humans and pets in urbanized areas leads to a higher proportion of false-positive signals from pets in human-associated markers. These markers therefore can be regarded as good MST candidates for indicating the presence of anthropogenic activities in the area, rather than showing human faecal pollution alone.

Albeit pigs are underrepresented in the study area, the Pig2Bac assay was included as it is considered a globally very robust procedure (Reischer *et al.* 2013). The binary assay sensitivity and specificity were both 100%, as the marker was detected in all samples, and none were false positive. Elsewhere, its reported assay sensitivity and specificity were 95–100% and 88–100%, respectively (e.g., Mieszkin *et al.* 2009; Malla *et al.* 2018; Zhang *et al.* 2020).

Quantitative aspects of assay faecal sensitivity and specificity

Results for quantitative assay evaluation are given as $\log_{10}$ (marker equivalents + 1) per g faeces ($\log_{10}$ (ME + 1) g^{-1}; Fig. 1) in the following. However, since data on quantitative assay evaluation is sometimes also referenced to ng^{-1} DNA or copies $reaction^{-1}$, we also gave our results in those units as Supporting Information to support comparability between studies. Correlation analysis (Table S3) revealed a very high correlation between the expression (ME + 1) g^{-1} faeces and (ME + 1) ng^{-1} DNA ($\rho = 0{\cdot}95$, $p = 0{\cdot}000$, $n = 122$). The different units of expression are, therefore, comparable in their quantitative relationships for our study (Figs S1 and S2).

The ruminant BacR marker concentration was very high (median: $\log_{10}$ 8·1 (ME + 1) g^{-1}, Fig. 1) and sample-to-sample variation was low (range: $\log_{10}$ 5·6–8·9 (ME + 1) g^{-1}). Median marker concentrations of the cattle, goat and sheep samples showed similar values ($\log_{10}$ 8·3, $\log_{10}$ 7·8 and $\log_{10}$ 7·8 (ME + 1) g^{-1}, respectively). Such high marker abundance of BacR in ruminant samples is in accordance with the results published in the original study (Reischer *et al.* 2006) and others (e.g., Malla *et al.* 2018). However, in addition to the marker concentration in target and non-target samples being decisive for assay applicability, their relative concentration differences need to be taken into account. The observed median distance in the marker concentration of target and non-target samples was ≥ 7 $\log_{10}$ levels. This, in contrast with other MST methods, shows the very high marker abundance in ruminant samples and their much lower abundance in non-targets, make BacR an ideal candidate to trace ruminant-associated faecal pollution in the study area.

Human-associated BacHum and HF183/BacR287 targets were observed in somewhat lower concentrations (median: $\log_{10}$ 5·4 and 4·2 (ME + 1) g^{-1}, respectively; Fig. 1) than of ruminants. The detection was independent of the gender and age of the donors (Table S4). Sample-to-sample variability was considerably higher than for the ruminant marker, as also observed elsewhere (range: $\log_{10}$ 4–9 copies g^{-1} faeces; e.g., Nshimyimana *et al.* 2017, Ahmed *et al.* 2019, Zhang *et al.* 2020). The marker concentration of the false-positive tested samples ranged from $\log_{10}$ 2·9 to 6·5 (ME + 1) g^{-1} (BacHum) and $\log_{10}$ 3·6 to 6·1 (ME + 1) g^{-1} (HF183/BacR287). The median distance of correct-positive tested target samples and non-target samples was ≥ 4 and ≥ 3 $\log_{10}$ orders for BacHum and HF183/BacR287, respectively (Fig. 1). Both assays can, therefore, also be regarded as very valuable candidates in the region as long as the possibility of cross-contamination is considered, especially if animals from the groups

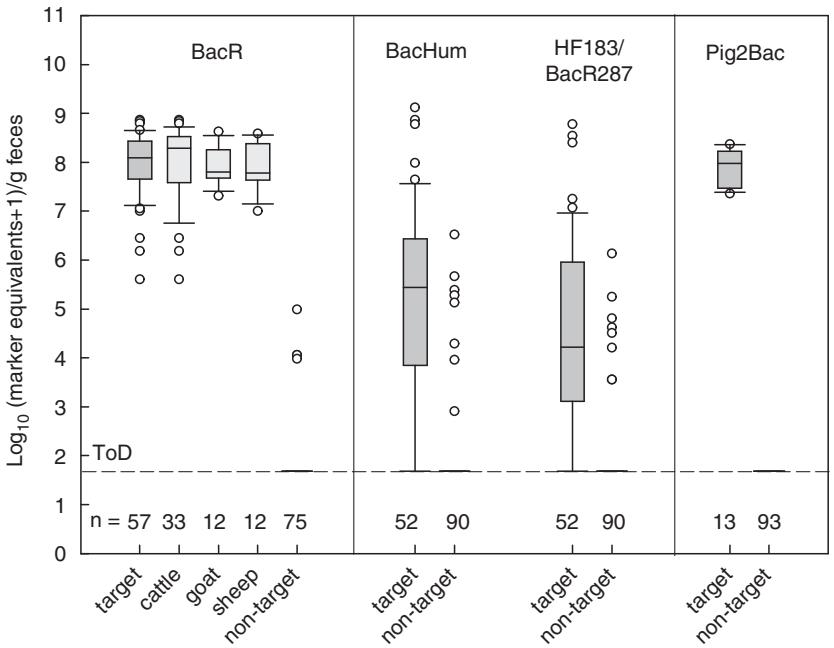

Figure 1 Marker equivalents per g faeces for target and non-target samples of all assays tested – BacR, BacHum, HF183/BacR287 and Pig2Bac. The results were obtained from the 1:4 dilution of the DNA samples and transformed into logarithmic format after the addition of 1 to each value. Boxes, 25th and 75th percentile; lines within the boxes, median; whiskers, 10th and 90th percentile, respectively; open circles represent outliers; dashed line, Threshold of Detection (ToD) which is $\log_{10}$ 1·81 (marker equovalents + 1) per 100 mg fresh weight of faeces; n, total number of samples in each category.

that tested false-positive (dogs and pigs) are present in considerable numbers.

The marker concentrations in the faecal target samples recorded with the Pig2Bac assay were in a range similar to that for BacR ($\log_{10}$ 7·4 to 8·4 (ME + 1) g^{-1}). Variability among the individual samples was the lowest of all markers, which is in line with the results from other studies (Mieszkin *et al*. 2009; Malla *et al*. 2018; Kongprajug *et al*. 2019). The Pig2Bac assay, therefore, seems to also represent an MST procedure targeting a highly abundant bacterial species in the intestinal core microbiome of pigs, which results in the robustness of this assay.

Marker occurrence in soil samples

Soils may represent a possible natural reservoir of phylogenetically related microbial targets and interference with MST. Therefore, cross-checking is a prerequisite for the safe application of molecular MST techniques (Vierheilig *et al*. 2012). In the present study, human-associated markers could not be detected in the uppermost soil layer, and BacR detection was restricted to more densely populated sites. Furthermore, its occurrence was subject to fluctuations over time (Table S6). During the wet season, BacR could not be detected at any site. During the dry season (September to January/February), BacR was detected in 83% of the samples collected at sites of higher anthropogenic influence (AW3–AW5), whereas it was only found in 25% of the samples collected from less densely populated and agriculturally used sites (AW1-2). This observation is further supported by a significant negative correlation between the monthly rainfall sum and the BacR marker concentration at sites AW3–AW5 (Table S7). BacR marker concentrations, when detected, ranged from $\log_{10}$ 3·0 to 4·9 (ME + 1) g^{-1} soil (Table S6).

Although open defecation is practised in the region, cattle outnumber humans. Additionally, while cattle produce approximately 23·6 kg wet weight d^{-1}, humans produce only 0·15 kg wet weight d^{-1} (Geldreich 1978). It is therefore reasonable to detect the BacR marker in the uppermost soil layer during the dry season when no leaching takes place and the soil does not present a natural reservoir. This result is also in agreement with the observed agricultural management practice. During the non-cropping season (dry season), all community livestock are freely grazing (e.g. riparian areas, hillsides, communal grazing areas and other fragile areas; December – June). During the cropping season, however, there are restrictions on the mobility of livestock and cattle are kept away from fields and graze on riparian areas,

hillsides, communal grazing areas and other fragile areas. Moreover, the results are further supported by the prevalence of ruminant faecal pellets, remarkably reflecting rainfall patterns in the study area (data not shown; see Mushi et al., submitted). These results clearly indicate that neither ruminant BacR nor human HF183/BacR287 false positive PCR signals will arise from pristine or non-faecal influenced soil compartments. Positive PCR signals from the ruminant BacR assay may only be a result from soils (and associated soil erosion processes) undergoing recent faecal contamination by ruminant animal populations.

Implications of these results for the Ethiopian region

Faecal water pollution poses a health risk, especially in developing countries where safe household water supplies are often scarce or non-existent. In these parts of the world, people still rely partly or entirely on surface water for their daily water needs (drinking water, household purposes, watering livestock, irrigating fields).

Ethiopia is no exception. Particularly in rural areas, people are heavily dependent on the availability of surface water. Land use here consists mainly of agricultural land used for rain-fed crops, irrigation and pasture farming (Tessema 2011). The climatic situation is characterized by a bimodal precipitation pattern with a short rainy season from March to May and a second one from July to September (Tessema 2011). Whereas during the dry season the risk for faecal input into surface water systems is mainly from increased anthropogenic activities (watering livestock, bathing), during the rainy season the input is mainly indirectly caused by flooding from nearby grazing areas. This study therefore focused on the detection of faecal contamination from ruminants, as these occur in large numbers and are of existential importance (Table S2), and human sources.

The assay evaluation confirmed the robustness of the ruminant BacR assay. The marker was found in all ruminant faecal samples at very high concentrations, making the assay an ideal candidate for detecting ruminant-associated faecal pollution in the study area. The human-associated markers were present in human stool samples at high, but still considerably lower, concentrations (3–4 $\log_{10}$ orders) than BacR in ruminant samples. From these results, it appears that the two human-associated assays are less sensitive in their environmental detection compared to BacR. For areas with only sporadic housings where open defecation is practised, these assays may therefore not be sensitive enough. In larger settlement areas, however, where open defecation is also practised, their application appears very reasonable. Interestingly, both assays also suggest a general anthropogenic impact indication, since cross-detection was mainly found in samples from dogs and pigs, closely living together with human individuals.

In summary, all tested MST procedures are very promising candidates for the determination of important faecal sources in the study area. However, further studies will be required to also test their applicability in the field. Studies on marker persistence under the given climatic conditions as well as marker detection in samples with a probably difficult sample matrix will be needed. The latter is mainly due to the region being strongly affected by soil erosion, increasing the risk of high inorganic turbidity that is likely to affect the efficiency of DNA extraction (Lever *et al.* 2015).

Materials and methods

Study area

The study area, the headwaters of the Awash River, is located in central Ethiopia at an altitude of approximately 3000 m above sea level. The catchment area of the river covers approximately 116 000 km^2 (AWBA 2017) and the climate in the river basin varies from humid subtropical in the highlands to arid in the Afar lowlands (Dinar *et al.* 2008). Annual average temperature in the Awash River basin ranges from 16·7 to 29°C and relative humidity varies from 49·7 to 60·2% (Hailu *et al.* 2017). The Awash River basin is the most important and most heavily used river basin in the country. The area is of great importance for agriculture and especially for livestock breeding. The livestock population, mainly cattle, sheep and goats (FAO 2019), is estimated at 5·96 million TLU (Tropical Livestock Units, FAO 2013). Approximately 18·3 million people live in the Awash basin (AWBA 2017). The Awash River water is crucial for irrigation and serves as an important water source for domestic consumption (Degefu *et al.* 2013).

The river also receives considerable amounts of industrial and domestic waste that strongly affect the water quality (Hailu *et al.* 2017). Despite intensive efforts to limit open defecation in order to improve the sanitary situation, only a few households in the study area have their own latrine. The majority continue to practice open defecation. Wastewater treatment is rather unusual, apart from larger cities such as Addis Abeba and some universities that discharge their wastewater into stabilization ponds. Household sewage is usually disposed into backyards or ditches specifically constructed for this purpose.

Sample collection and DNA extraction

Faecal samples were collected in the Awash basin from March 2017 to February 2018. For a detailed description

of the sampling sites the reader is referred to Kebede *et al.* (2020). A total of 173 faecal samples from individual specimens of livestock, humans and wildlife species were collected. The final data set included faecal samples from the following sources: ruminants ($n = 57$), composed of samples from cattle ($n = 33$), goats ($n = 12$) and sheep ($n = 12$); humans ($n = 52$); herbivores ($n = 21$) composed of samples from horses ($n = 7$), donkeys ($n = 7$) and mules ($n = 7$); chickens ($n = 8$), pigs ($n = 13$), and dogs ($n = 9$); and wildlife ($n = 13$) composed of samples from pigeons ($n = 4$) and monkeys ($n = 9$).

The most frequently kept livestock animals in the region are cattle, goats and sheep. Donkeys, mules and horses are kept as working animals. Wildlife is underrepresented in the sample set. In the forest areas there is game, which is hunted. At the time of sampling, however, a massive protest was raised against the government in the region and suspicions were expressed that armed groups had entrenched themselves in the forests, making sampling there impossible. Analysis was thus focused on livestock and human pollution sources. In addition, a systemic door-to-door survey of residents about the prevalence of different livestock present in the upper catchment area of Awash River was carried out as part of this study before selection of specific MST procedures. The results verified the assumption that ruminants and humans must be considered the most prevalent sources of faecal contamination in the upper Awash basin. Poultry and other non-ruminant livestock were less prevalent (Table S2). Sensitivity and specificity testing of the candidate MST methods, therefore, focused on ruminant- and human-associated faecal sources.

All faecal samples were collected in sterile sampling vials (Greiner, Austria) and transported in the dark on ice to the laboratory. Samples were stored at −20°C within 6 h after sampling, and DNA extraction was performed within 2 months after sampling. Human stool samples were anonymously donated by healthy persons of varying ages (2–41 years old) from the districts Ambo and Ginchi (Table S1).

Soil samples were collected from five independent plots monthly over a 1-year period (March 2017 to February 2018) to cover both the dry and rainy seasons. Upon collection, the samples were kept at − 20°C until DNA extraction. Details about the sampling sites and sample collection are available as supporting information (Table S2).

From all faecal and soil samples, DNA was extracted using the MoBio Power Soil DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA) according to the manufacturer's instructions. Per sample either 0·25 g of faeces (fresh weight) or 0·25 g soil were used. For each extraction batch an extraction control using only reagents was prepared.

MST genetic marker detection

Four different MST assays to identify ruminant- (BacR; Reischer *et al.* 2006), human- (BacHum and HF183/BacR287; Kildare *et al.* 2007, Green *et al.* 2014) and pig-associated (Pig2Bac; Mieszkin *et al.* 2009) faecal pollution were evaluated. In addition to the applied host-associated faecal genetic markers, a general *Bacteroidetes* marker, AllBac (Layton *et al.* 2006), was run in duplex with the *ntb2* fragment as an internal amplification control (Anderson *et al.* 2011). All qPCR reactions were run on a Rotor-Gene Q thermocycler (Qiagen, Netherlands) in duplicate and were performed in a 15 µl reaction volume as described previously (Mayer *et al.* 2018). Detailed information on the cycling conditions, data reading and quality assessment of the qPCR performance is given as Supporting Information.

Data analysis and statistics

All data analysis was performed with Microsoft Excel and Sigma Plot 10 (Systat Software Inc., Chicago, IL). All qPCR data are expressed as $\log_{10}(x + 1)$, where x is the concentration calculated from the standard curve before applying the logarithm to it. Quantitative results of the marker concentrations in the faecal and soil samples are given as marker equivalents (ME) g^{-1} instead of copies per g since the extraction efficiency of the MoBio Power Soil Kit and associated with it the exact cell number remains unknown (Reischer *et al.* 2006). The amount of faeces used for DNA extraction, copies in undiluted DNA extract and the minimal theoretically detectable marker concentration per reaction defines the detection threshold (Reischer *et al.* 2008; Bustin *et al.* 2009). The resultant value is referred to as the threshold of detection (ToD, Reischer *et al.* 2006). Non-detects from the qPCR enumeration were set to the ToD level of $\log_{10}$ 1·81 (marker equivalents + 1) per 100 mg faeces for the present data set.

Binary assay specificity was defined as the percentage of non-target samples not detected with the respective qPCR assay. As such, all samples yielding <1 copy/reaction were charged. In contrast, binary assay sensitivity was defined as the percentage of target samples giving a positive signal in the qPCR. As such, all samples yielding >1 copy/reaction were charged.

As a statistical metric to estimate the quantitative faecal source discrimination efficiency of the tested MST marker, we determined the $\log_{10}$ median distance between the target and the non-target faecal samples. This is a

methodical modification after Reischer *et al.* (2013) who used the distance between the 25th percentile marker concentration in target samples and the 75th percentile concentration in the non-target samples.

Spearman's rank-order correlation was used for calculation of the correlation coefficients among the parameters using SPSS Statistics Software ver. 25 (SPSS Inc., Chicago, IL).

Conflict of Interests

The authors have no conflict of interest and no competing financial interest to declare.

References

Ahmed, W., Gyawali, P., Feng, S. and Mclellan, S.L. (2019) Host specificity and sensitivity of established and novel sewage-associated marker genes in human and nonhuman faecal samples. *Appl Environ Microbiol* **85**(14), e00641–19.

Anderson, A., Pietsch, K., Zucker, R., Mayr, A., Müller-Hohe, E., Messelhäusser, U., Sing, A., Busch, U. *et al.* (2011) Validation of a duplex real-time PCR for the detection of *Salmonella* spp. different food products. *Food Anal Method* **4**(3), 259–267.

Awash Basin Authority (AWBA). (2017) *Executive summary of strategic River Basin Plan for Awash Basin*. www.cmpethiopia.org.

Bauza, V., Madadi, V., Ocharo, R.M., Nguyen, T.H. and Guest, J.S. (2019) Microbial source tracking using 16S rRNA amplicon sequencing identifies evidence of widespread contamination from young children's feces in an Urban Slum of Nairobi, Kenya. *Environ Sci Technol* **53**, 8271–8281.

Bernhard, A.E. and Field, K.G. (2000) Identification of nonpoint sources of faecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from faecal anaerobes. *Appl Environ Microbiol* **66**(4), 1587–1594.

Boehm, A.B., Van De Werfhorst, L.C., Griffith, J.F., Holden, P.A., Jay, J.A., Shanks, O.C., Wang, D. and Weisberg, S.B. (2013) Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Res* **47**(18), 6812–6828.

Bradshaw, J.K., Snyder, B.J., Oladeinde, A., Spidle, D., Berrang, M.E., Meinersmann, R.J., Oakley, B., Sidle, R.C. *et al.* (2016) Characterizing relationships among faecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. *Water Res* **101**, 498–509.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Hugett, J., Kubista, M., Mueller, R., Nolan, T. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**, 611–622.

Byamukama, D., Mach, R.L., Kansiime, F., Manafi, M. and Farnleitner, A.H. (2005) Discrimination efficacy of faecal pollution detection in different aquatic habitats of a high-altitude tropical country, using presumptive coliforms, *Escherichia coli*, and *Clostridium perfringens* spores. *Appl Environ Microbiol* **71**, 65–71.

Degefu, F., Lakew, A., Tigabu, Y. and Teshome, K. (2013) The water quality degradation of upper Awash River, Ethiopia. *Ethiopian J Environ Studies Manag* **6**(1). https://doi.org/10.4314/ejesm.v6i1.7

Dinar, A., Hassan, R., Mendelsohn, R. and Benhin, J. (2008) *Climate change and agriculture in Africa. Impact assessment and adaptation strategies*. UK London; Sterling, VA: Earthscan.

FAO (2013) *Coping with water scarcity, the role of agriculture: developing a water audit for Awash river basin, a synthesis report*. Addis Ababa, Ethiopia: FAO. [Available at http://www.fao.org/nr/water/docs/awash/AwashSynthesisReport.pdf].

FAO (2019) *Livestock, health, livelihoods and the environment in Ethiopia. An integrated analysis*. Rome. 108 pp.

Farnleitner, A.H., Ryzinska-Paier, G., Reischer, G.H., Burtscher, M.M., Knetsch, S., Kirschner, A.K., Dirnbock, T., Kuschnig, G. *et al.* (2010) *Escherichia coli* and enterococci are sensitive and reliable indicators for human, livestock and wildlife faecal pollution in alpine mountainous water resources. *J Appl Microbiol* **109**, 1599–1608.

Feng, S., Bootsma, M. and Mclellan, S.L. (2018) Human-associated *Lachnospiraceae* genetic markers improve detection of faecal pollution sources in urban waters. *Appl Environ Microbiol* **84**, e00309–18.

Feng, S., Ahmed, W. and Mclellan, S.L. (2020) Ecological and technical mechanisms for cross-reaction of human fecal indicators with animal hosts. *Appl Environ Microbiol* **86**(5). https://doi.org/10.1128/AEM.02319-19

Geldreich, E.E. (1978) Bacterial populations and indicator concepts in feces, sewage, stormwater and solid wastes. In *Indicators of viruses in water and food* ed. Berg, G. Ann Arbor, MI: Ann Arbor Science Publishing Inc.

Green, H.C., Haugland, R.A., Varma, M., Millen, H.T., Borchardt, M.A., Field, K.G., Walters, W.A., Knight, R. *et al.* (2014) Improved HF183 quantitative real-time PCR assay for characterization of human faecal pollution in ambient surface water samples. *Appl Environ Microbiol* **80**, 3086–3094.

Hailu, R., Tolossa, D. and Alemu, G. (2017) Water security: stakeholders' arena in the Awash River Basin of Ethiopia. *Sustain Water Resour Manag* **5**, 513–531.

Holcomb, D.A., Knee, J., Sumner, T., Adriano, Z., De Bruijn, E., Nala, R., Cumming, O., Brown, J. *et al.* (2020) Human faecal contamination of water, soil, and surfaces in households sharing poor-quality sanitation facilities in

Maputo, Mozambique. *Int J Hyg Environ Health* **226**, 113496.

Jenkins, M.W., Tiwari, S., Lorente, M., Maina, C. and Wuertz, S. (2009) Identifying human and livestock sources of faecal contamination in Kenya with host-specific Bacteroidales assays. *Water Res* **43**(19), 4956–4966.

Kebede, G., Mushi, D., Linke, R.B., Dereje, O., Lakew, A., Hayes, D.S., Farnleitner, A.H. and Graf, W. (2020) Macroinvertebrate indices versus microbial faecal pollution characteristics for water quality monitoring reveals contrasting results for an Ethiopian river. *Ecol Ind* **108**, 105733.

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

Kongprajug, A., Chyerochana, N., Somnark, P., Leelapanang Kampaengthong, P., Mongkolsuk, S. and Sirikanchana, K. (2019) Human and animal microbial source tracking in a tropical river with multiple land use activities. *Int J Hyg Environ Health* **222**, 645–654.

Layton, A., Mckay, L., Williams, D., Garrett, V., Gentry, R. and Sayler, G. (2006) Development of Bacteroides 16S rRNA gene taqman-based real-time PCR assays for estimation of total, human, and bovine faecal pollution in water. *Appl Env Microbiol* **72**(6), 4214–4224.

Lever, M.A., Torti, A., Eickenbusch, P., Michaud, A.B., Santl-Temkiv, T. and Jorgensen, B.B. (2015) A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Front Microbiol* **6**, 476.

Malla, B., Ghaju Shrestha, R., Tandukar, S., Bhandari, D., Inoue, D., Sei, K., Tanaka, Y., Sherchand, J.B. *et al.* (2018) 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.

Mayer, R.E., Reischer, G.H., Ixenmaier, S.K., Derx, J., Blaschke, A.P., Ebdon, J.E., Linke, R., Egle, L. *et al.* (2018) Global distribution of human-associated fecal genetic markers in reference samples from six continents. *Environ Sci Technol* **52**, 5076–5084.

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

Mushi, D., Kebede, G., Linke, R.B., Lakew, A., Hayes, D.S., Graf, W. and Farnleitner, A.H. (submitted) Microbial faecal pollution of riverine water in Ethiopian highlands is driven by ambient faecal pellet abundance and rainfall-induced catchment runoff. *Int J Hyg Environ Health*.

Mushi, D., Byamukama, D., Kivaisi, A.K., Mach, R.L. and Farnleitner, A.H. (2010) Sorbitol-fermenting *Bifidobacteria* are indicators of very recent human faecal pollution in streams and groundwater habitats in urban tropical lowlands. *J Water Health* **8**, 466–478.

Nshimyimana, J.P., Cruz, M.C., Thompson, R.J. and Wuertz, S. (2017) Bacteroidales markers for microbial source tracking in Southeast Asia. *Water Res* **118**, 239–248.

Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P.R., Panigrahi, P. and Jenkins, M.W. (2015) Validation of Bacteroidales quantitative PCR assays targeting human and animal faecal contamination in the public and domestic domains in India. *Sci Total Environ* **502**, 462–470.

Pickering, A.J., Julian, T.R., Marks, S.J., Mattioli, M.C., Boehm, A.B., Schwab, K.J. and Davis, J. (2012) Fecal contamination and diarrheal pathogens on surfaces and in soils among Tanzanian households with and without improved sanitation. *Environ Sci Technol* **46**, 5736–5743.

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

Reischer, G.H., Haider, J.M., Sommer, R., Stadler, H., Keiblinger, K.M., Hornek, R., Zerobin, W., Mach, R.L. *et al.* (2008) Quantitative microbial faecal source tracking with sampling guided by hydrological catchment dynamics. *Environ Microbiol* **10**(10), 2598–2608.

Reischer, G.H., Ebdon, J.E., Bauer, J.M., Schuster, N., Ahmed, W., Astrom, J., Blanch, A.R., Bloschl, G. *et al.* (2013) Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ Sci Technol* **47**, 8548–8556.

Tessema, S.M. (2011) *Hydrological modeling as a tool for sustainable water resources management: a case study of the Awash River Basin*. Doctoral Dissertation, KTH Royal Institute of Technology.

USEPA (2019) *Method 1696: Characterization of Human Fecal Pollution in Water by HF183/BacR287 TaqMan Quantitative Polymerase Chain Reaction (qPCR) Assay*. Washington, DC: United States Environmental Protection Agency

Vierheilig, J., Farnleitner, A.H., Kollanur, D., Bloschl, G. and Reischer, G.H. (2012) High abundance of genetic Bacteroidetes markers for total faecal pollution in pristine alpine soils suggests lack in specificity for feces. *J Microbiol Methods* **88**, 433–435.

Wuertz, S., Wang, D., Reischer, G.H. and Farnleitner, A.H. (2011). Library-independent source tracking methods. In *Microbial Source Tracking: Methods, Applications, and Case Studies* eds. Hagedorn, C., Blanch, A. R., Harwood, V. J. pp. 61–113. New York: Springer.

Yahya, M., Blanch, A.R., Meijer, W.G., Antoniou, K., Hmaied, F. and Balleste, E. (2017) Comparison of the performance of different microbial source tracking markers among

European and North African regions. *J Environ Qual* **46** (4), 760–766.

Zhang, Y., Wu, R., Lin, K., Wang, Y. and Lu, J. (2020) Performance of host-associated genetic markers for microbial source tracking in China. *Water Res* **175**, 115670.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Marker concentration expressed as marker copies per PCR (2·5 µl sample volume) for the target and non-target samples of all assays tested—BacR, HF183/BacR287, BacHum and Pig2Bac

Figure S2 Marker equivalents per ng DNA for target and non-target samples of all assays tested—BacR, HF183/BacR287, BacHum and Pig2Bac

Table S1 List of human stool samples and additional information on DNA concentration and marker concentration (BacHum and HF183/BacR287) of the individual samples.

Table S2 Additional information on livestock numbers and human settlements in the surrounding area of the soil sampling sites AW1–AW5. Information was obtained from a door-to-door survey and local authorities.

Table S3 Spearman rank-order correlation of the different units of expression of all target samples (ruminant, human and pig) marker concentrations (n = 122).

Table S4 Spearman rank-order correlation of human-associated marker concentrations and sex and age of donators (n = 52).

Table S5 Spearman rank-order correlation of data on DNA concentration and human-associated marker concentrations (n = 52).

Table S6 BacR marker concentrations measured in soil samples from five sites with different land use patterns alongside Awash River.

Table S7 Spearman rank-order correlation of monthly rainfall sums (compare Table S8) at sites AW1 – AW5 and ruminant-associated marker concentrations in soil samples (n = 12).

Table S8 The rainfall data [mm] was obtained from the recording rain gauge of the Holeta Agricultural Research Center under the Ethiopian Agricultural Research Institute (EIAR) which is representative for the whole sampling area covering sites AW1 – AW5. Days on which soil samples were collected are marked with a thicker borders.

