Global Distribution of Human-associated Fecal Genetic Markers in Reference Samples from Six Continents

René E. Mayer [§,a,b], Georg H. Reischer [§,a,c], Simone K. Ixenmaier [a,b], Julia Derx [b,d], Alfred Paul Blaschke [b,d], James E. Ebdon [e], Rita Linke [a,b], Lukas Egle [f], Warish Ahmed [g], Anicet R. Blanch [h], Denis Byamukama [i], Marion Savill [j], Douglas Mushi [k], Héctor A. Cristóbal [l], Thomas A. Edge [m], Margit A. Schade [n], Asli Aslan [o], Yolanda M. Brooks [p], Regina Sommer [q,b], Yoshifumi Masago [r], Maria I. Sato [s], Huw D. Taylor [e], Joan B. Rose [p], Stefan Wuertz [t], Orin C. Shanks [u], Harald Piringer [v], Domenico Savio [w], Robert L. Mach [x], Matthias Zessner [f] and Andreas H. Farnleitner [a,b,w,*]

* Corresponding author: Mailing address: Karl Landsteiner University of Health Sciences, Research Division Water Quality and Health, A-3500 Krems, Austria. E-mail: andreas.farnleitner@kl.ac.at. Tel. +43 664 605882244.

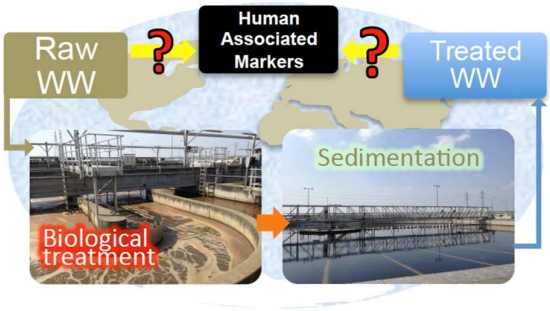
[a] Research Group Environmental Microbiology and Molecular Diagnostics, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, 1060 Vienna, Austria.
[b] InterUniversity Cooperation Centre Water & Health (www.waterandhealth.at), Vienna, Austria.
[c] Molecular Diagnostics Group, IFA-Tulln, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, 3430 Tulln, Austria
[d] Institute of Hydraulic Engineering and Water Resources Management, TU Wien, 1040 Vienna, Austria
[e] Environment & Public Health Research and Enterprise Group, School of Environment and Technology, University of Brighton, BN2 4GJ Brighton, UK.
[f] Institute for Water Quality and Resource Management, TU Wien, 1040 Vienna, Austria
[g] CSIRO Land and Water, 4067 Brisbane, Australia.
[h] Department of Genetics, Microbiology and Statistics. University of Barcelona, 08028 Barcelona, Spain.
[i] Department of Biochemistry, Makerere University, P. O. Box 27755 Kampala, Uganda.
[j] Affordable Water Limited, 1011 Auckland, New Zealand.
[k] Department of Biosciences, Sokoine University of Agriculture, P.O.BOX 3038 Tanzania
[l] Laboratorio de Aguas y Suelos, Instituto de Investigaciones para la Industria Química (INIQUI), Consejo Nacional de Investigaciones Científicas y Técnicas and Universidad Nacional de Salta, CP 4400 Salta, Argentina.
[m] Canada Centre for Inland Waters, Environment Canada, Burlington, L7R 4A6 Ontario, Canada
[n] Bavarian Environment Agency, 86179 Augsburg, Germany.
[o] Department of Epidemiology and Environmental Health Sciences, Georgia Southern University, Statesboro, 30460 Georgia, USA
[p] Department of Microbiology and Molecular Genetics, Michigan State University East Lansing, 48824 Michigan, USA
[q] Institute for Hygiene and Applied Immunology, Water Hygiene, Medical University of Vienna, 1090 Vienna, Austria.
[r] New Industry Creation Hatchery Center, Tohoku University, 980-8577 Sendai, Japan
[s] CETESB - Cia. Ambiental do Estado de São Paulo, Departamento de Análises Ambientais, 05459-900 São Paulo, Brasil
[t] Singapore Centre for Environmental Life Sciences Engineering and School of Civil and Environmental Engineering, Nanyang Technological University, 637551 Singapore
[u] U.S. Environmental Protection Agency, Office of Research and Development, 45268 Cincinnati, Ohio, USA
[v] VRVis Research Center, 1220 Vienna, Austria
[w] Research Division Water Quality and Health, Karl Landsteiner University of Health Sciences, 3500 Krems, Austria
[x] Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, 1060 Vienna, Austria

[§] These authors contributed equally to the publication

53

54 Keywords: microbial source tracking, pollution microbiology, fecal pollution, Bacteroidetes,

55 quantitative PCR, genetic markers, water quality, QMRA, QMRAcatch.


Raw WW
Human Associated Markers
Treated WW
Sedimentation
Biological treatment

TOC Art

Abstract

Numerous bacterial genetic markers are available for the molecular detection of human sources of fecal pollution in environmental waters. However, widespread application is hindered by a lack of knowledge regarding geographical stability, limiting implementation to a small number of well-characterized regions. This study investigates the geographic distribution of five human-associated genetic markers (HF183/BFDrev, HF183/BacR287, BacHum-UCD, BacH, and Lachno2) in municipal wastewaters (raw and treated) from 29 urban and rural wastewater treatment plants (750 – 4,400,000 population equivalents) from 13 countries spanning six continents. In addition, genetic markers were tested against 280 human and non-human fecal samples from domesticated, agricultural and wild animal sources. Findings revealed that all genetic markers are present in consistently high concentrations in raw (median $\log_{10}$ 7.2 - 8.0 marker equivalents (ME) 100 ml^{-1}) and biologically treated wastewater samples (median $\log_{10}$ 4.6 - 6.0 ME 100 ml^{-1}) regardless of location and population. The false positive rates of the various markers in non-human fecal samples ranged from 5% to 47%. Results suggest that several genetic markers have considerable potential for measuring human-associated contamination in polluted environmental waters. This will be helpful in water quality management and pollution modeling and health risk assessment across the globe (as demonstrated by QMRAcatch).

1 INTRODUCTION

The disease burden from poor water quality, sanitation, and hygiene is estimated to be responsible for up to 4% of all deaths worldwide [1, 2]. Limited access to safe drinking water has led the World Health Organization to develop strategies for managing water quality with the goal to protect and promote human health [3]. To provide information about microbiological water quality, cultivation of standard fecal indicator bacteria, such as *E. coli* and enterococci, are typically used according to certified standard procedures [4, 5]. However, standard fecal indicator bacteria measurements do not provide information about the origin of fecal pollution, because these organisms are present in the feces of most warm blooded animals [6]. The field of microbial source tracking (MST) seeks to develop methods allowing for the discrimination between different animal sources of fecal pollution to improve water quality management [7].

A useful MST assay should have high source-specificity (low number of false positives) and excellent source-sensitivity (low number of false negatives) [7]. MST assay specificity and sensitivity are typically evaluated based on repeated testing of reference fecal and wastewater samples often collected in close proximity to the research laboratory[8-11]. For example, Boehm et al. (2013) evaluated specificity and sensitivity of 41 MST methods with more than 100 reference samples collected from the California area [12]. To date, the performance of many MST assays described in the literature have not been tested for source-specificity and -sensitivity beyond the regional level [8, 13-15]. For this reason, it is often difficult to identify the most appropriate methods when planning a MST application in a new geographical area. To improve this situation Reischer et al. (2013)[16] compared five quantitative real-time PCR (qPCR) methods targeting human and ruminant pollution sources by testing a large collection of reference fecal samples from 16 countries demonstrating that tested genetic markers were broadly distributed regardless of the location from which the fecal samples originated. The

103 study also emphasized the investigation of the quantitative distribution of genetic marker
104 concentrations in the target and non-target fecal samples in order to fully assess performance.
105 This present study seeks to build upon previous research by providing insights into the
106 occurrence and concentration of human-associated bacterial genetic markers in raw and
107 biologically treated municipal wastewater from multiple geographical locations around the
108 globe. Three widely applied MST bacterial qPCR assays (BacH, BacHum-UCD and
109 HF183/BFDrev), and two recently modified or developed qPCR assays (HF183/BacR287 and
110 Lachno2, respectively) were challenged using wastewater samples collected from 29 facilities
111 spanning 13 countries across six continents. Wastewater samples included both urban and
112 rural plants serving a wide range of population sizes. Considerable effort was made to ensure
113 standardization in sampling collection, handling, and processing. Marker concentration data
114 found in wastewater were used in an exemplary modeling application using the QMRAcatch
115 tool. In addition, the new human-associated marker HF183/BFDrev, HF183/BacR287 and
116 Lachno2 were challenged against a previously established collection of reference fecal
117 samples from six continents [16] to compare their ability to correctly differentiate fecal sources.
118

119 2 EXPERIMENTAL METHODS

120 2.1 Sample collection and wastewater selection criteria

121 The requirements and guidelines for cooperation partners in this MST evaluation project were
122 defined in 2013. In brief, detailed standard operating procedures for sampling and filtration
123 were distributed to all cooperating partners including a demonstrational video showing
124 important filtration and filter packing steps (cf. Supporting video file) to ensure that sample
125 processing was standardized. Partners were also required to use an online sampling protocol
126 to collect metadata. To improve comparability, polycarbonate membrane filters (0.2 µm
127 Millipore, Isopore Membrane Filter – GTTP, Cork, Ireland), preprinted labels, and vials for

sample processing were provided to all partners by the lead laboratory (TU Wien). Partners were requested to select one urban municipal wastewater treatment plant (WWTP) with a pollution load greater than 500,000 population equivalents (PE) and one rural municipal WWTP with less than 50,000 PE (Table 1). An effort was made to select WWTPs receiving minimal levels of industrial waste. All WWTP facilities consisted of mechanical treatment followed by either activated sludge or fixed film treatment. Data on treatment capacity (PE), sewage system (separate, combined) and details of possible inputs from industry or livestock, were provided by the respective WWTP operators.

2.2 Sampling, shipment and quality control

Five hundred milliliter grab samples were taken at all WWTP sites during the morning hours (before 09:00 local time) under dry weather conditions (no rain in the preceding 36 hours). Raw wastewater samples were collected at the post mechanical screening stage, while the biologically-treated wastewater samples were taken post-secondary sedimentation/clarification prior to any advanced (tertiary) treatment (e.g. ultraviolet irradiation, chlorination, or coagulation). Tertiary-treated wastewater was not included, because tertiary treatment methods may vary by WWTP and a large proportion of the investigated plants did not have any tertiary treatment steps. Sampling points were chosen in turbulent zones to promote good mixing and samples were collected about 20 cm below the water surface. Samples were stored in 500ml sterile glass bottles in the dark at < 4°C and immediately transported to the respective collaboration partner laboratory for filtration (holding time < 6 hours), prior to shipping to TU Wien. For each sampling event, four 50 ml replicate sub-samples were filtered and filters were immediately frozen at -20°C (two filters were shipped to TU Wien; the other two were kept by the local cooperation partner as backups). On each sampling occasion an unused filter was put directly into a 2 ml extraction

vial as a blank filter control. Shipment services were carried out in accordance with international law by qualified logistics companies and under controlled frozen conditions on dry ice. Sample filters were only used for DNA extraction if dry ice was still present upon arrival at TU Wien in Austria.

2.3 Additional animal and human DNA sample collection

The fecal DNA samples used to compare false-positive and false-negative rates in the different human-associated markers were collected and extracted during a previous study [16]. In brief, reference sample collection was collected during the period 2007 to 2008 and consisted of 280 fecal samples from six continents including 61 human and 219 non-human fecal samples from various sources such as agricultural and wild animals (for details see Supporting Information SI.) [16].

2.4 DNA extraction

DNA from the filters was recovered at the TU Wien laboratory by phenol/chloroform extraction as previously described [17, 18]. Cells were lysed with CTAB buffer solution, glass beads and a FastPrepR-24 Instrument (MP Biomedicals Inc., Irvine, USA) at a speed setting of 6 m s^{-1} for 30 s. Polycarbonate membrane filters were completely dissolved at this step and the DNA was purified with follow-up washing procedures. The extracted DNA was dissolved in 10mM TRIS HCl, pH=8, and stored at -80°C for no longer than 21 days prior to qPCR analysis. Every extraction event was accompanied by a blank extraction control. The concentration of extracted DNA was measured with Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific, UK) on an Anthos Zenyth fluorometric plate reader (Beckman Coulter, Wien, Austria) to check for loss of DNA during extraction. For one sample the DNA concentration was below the detection limit. This sample was removed from further analyses.

178 **Table 1**: Characteristics of investigated disposal systems and wastewater treatment plants

179 (WWTP)

Country (site location)	Sewerage System	Influence		Population Equivalent (PE)
		Industry	Livestock	
Argentina (rural)	separated	slight	strong	350,000
Argentina (urban)	combined	strong	strong	600,000
Australia (rural)	separated	n.a.	n.a.	50,000
Australia (urban)	separated	slight	no	500,000
Brazil (rural)	separated	no	no	19,100
Brazil (urban)	separated	slight	n.a.	4,400,000
Canada (rural)	separated	slight	slight	20,000
Canada (urban)	combined	no	no	500,000
Germany (rural)	combined	slight	n.a.	16,800
Germany (urban)	combined	moderate	slight	1,000,000
Japan (rural)	separated	no	no	10,200
Japan (urban)	separated	slight	slight	300,000
N. Zealand (rural)	n.a.	n.a.	n.a.	n.a.
N. Zealand (urban)	n.a.	n.a.	n.a.	n.a.
Singapore (urban)	separated	moderate	slight	1,700,000
Spain (rural)	separated	slight	no	45,100
Spain (urban)	separated	slight	no	384,000
Tanzania (rural)	combined	no	no	3,000
Tanzania (urban)	combined	no	no	10,000
Uganda (rural)	separated	no	no	750
Uganda (urban)	separated	no	no	320,000
UK (rural)	combined	no	slight	14,600
UK (urban)	combined	strong	slight	3,500,000
USA (rural)	separated	no	no	3,500
USA (rural)	combined	no	no	16,000
USA (rural)	combined	moderate	slight	29,800
USA (urban)	combined	strong	no	142,000
USA (urban)	combined	moderate	slight	3,000,000
USA (urban)	combined	moderate	no	480,000

180

181 Abbreviations: n.a.: not available; influence of industrial and agricultural pollution sources was assessed based

182 on expert knowledge by local partners after consultation with plant operators.

183

2.5 qPCR measurement and quality control

The following host-associated fecal genetic 16S-rRNA-gene markers were quantified by qPCR: HF183/BFDrev [11], HF183/BacR287 [8], BacH [19], and BacHum-UCD [9], all of which target human-associated Bacteroidetes, and Lachno2 [20] targeting a human-associated Firmicutes clade. In addition, the general Bacteroidetes marker, AllBac was used as a quality control to assess the ability to amplify DNA extracted from wastewater and rule out the presence of PCR inhibition in the sample extract dilutions [21]. Samples with two matching concentrations (i.e. the ratio [concentration 1:100*10] / [concentration1:10] was between 0.5 to 2) in the 1:10 and 1:100 dilutions were judged free of PCR inhibitor in the 1:10 dilution. qPCR measurements were performed on a Rotorgene Q Cycler (Qiagen, Hilden, Germany). A QIAgility liquid handling robot (Qiagen) was used to prepare qPCR reactions in a total volume of 15 µl, with 2.5 µl of sample DNA, 7.5 µl of Rotor-Gene Multiplex PCR Kit (Qiagen) and 400 mg L^{-1} bovine serum albumin (Roche Diagnostics, Mannheim, Germany). For the AllBac qPCR assay 600 nmol L^{-1} primer AllBac296f, 600 nmol L^{-1} primer AllBac412r, and 25 nmol L^{-1} TaqMan MGB probe AllBac375Bhqr were used [21]. Additionally, as an internal amplification control (IAC) 500 nmol L^{-1} primer IPC-ntb2-fw, 500 nmol L^{-1} primer IPC-ntb2-re, 200 nmol L^{-1} ROX probe IPC-ntb2-probe and 10^3 copies of IAC Template IPC-ntb2 plasmid DNA [22] were added to each AllBac qPCR reaction. For the BacHum-UCD assay 400 nmol L^{-1} primer BacHum-160f, 400 nmol L^{-1} primer BacHum-241r and 80 nmol L^{-1} TaqMan MGB probe BacHum-193p were used [9]. For the HF183/BFDrev assay 1000 nmol L^{-1} primer HF183, 1000 nmol L^{-1} primer BFDREV and 80 nmol L^{-1} TaqMan MGB probe BFDFAM were used [11]. For the HF183/ BacR287 assay 1000 nmol L^{-1} primer HF183, 1000 nmol L^{-1} primer BacR287, and 80 nmol L^{-1} TaqMan MGB probe BacP234MGB were used [8]. For the BacH assay 200 nmol L^{-1} primer BacH_f, 200 nmol L^{-1} primer BacH_r, 100 nmol L^{-1} each of TaqMan MGB probes BacH_pC and BacH_pT were used [19]. For the

209 Lachno2 assay 1000 nmol L^{-1} primer Lachno2F, 1000 nmol L^{-1} primer Lachno2R and
210 80 nmol L^{-1} TaqMan MGB probe Lachno2P were used [20]. (cf. Supporting Information SI.)
211 Quantification was based on plasmid standard dilutions. The respective plasmid stock for
212 each assay was diluted in an unspecific background of 500 µg L^{-1} poly(dI-dC) (Roche
213 Diagnostics, Mannheim, Germany) to avoid adsorption of plasmid DNA to reaction vials at
214 low plasmid concentrations (cf. SI.). A total of at least eight ten-fold serial dilutions of
215 plasmid standard (10^0 to 10^7 gene copies) were performed in each qPCR run. Every run also
216 included several no-template and DNA extraction controls.
217 Each wastewater DNA sample was analyzed in two dilution steps of the original extract (10-
218 and 100-fold dilution) and each dilution in duplicate reactions, in order to check for a possible
219 qPCR inhibition[23]. Additionally, an IAC was run in duplex with the AllBac assay to monitor
220 for qPCR amplification inhibition[23]. Inhibition was assumed to be present if the threshold
221 cycle (Ct) value of the IAC was shifted to higher Ct values by more than one cycle. All qPCR
222 runs in this study revealed a calculated PCR efficiency between 90% and 105% and no-
223 template and extraction controls were consistently negative (i.e. fluorescence never exceeded
224 threshold). The qPCR standard dilutions ranging from 10^0 to 10^7 targets per reaction were
225 used in a linear regression model for calculation of the qPCR calibration curve. Results for
226 wastewater investigations were reported as marker equivalents per filtered wastewater volume
227 (ME vol^{-1}) as previously described [18]. Samples with replicate standard deviations of the ct-
228 value >1 in the 10fold DNA extract dilutions were considered to be not quantifiable and were
229 not considered for further analysis.
230 Results for the fecal DNA setup were measured in the 1:4 dilution of the fecal DNA extracts.
231 They are reported directly as genetic marker copies per qPCR reaction in the same manner as
232 previously published data on the same samples[16]. DNA extracts were reanalyzed using the

AllBac assay and the results indicated that no DNA degradation had occurred during storage (data not shown).

2.6 Data and statistical analysis

All qPCR data were expressed as $\log_{10}(x+1)$, where x is the calculated concentration before applying the logarithm to it. To estimate $\log_{10}$ reductions of the MST markers during wastewater treatment Monte Carlo simulations of the marker concentrations [i.e. $\log_{10}$ (influent) - $\log_{10}$ (effluent)] were performed. As stop criterion a maximum of 100000 simulated cases or a confidence level of 95% with a threshold of 1 % for the mean was set. Visual and statistical data analyses were done with Visplore 2.0 [24] (VRVis GmbH, Austria, Vienna) and Sigma Plot 13.0 (SPSS Inc., Chicago, USA). For multiple comparison of groups One-way ANOVA was used and if significant differences between two or more groups were detected a Tukey Post-Hoc test was performed. To account for multiple statistical testing, statistical significance levels were corrected according to Bonferroni [25]. To support correct comparisons of the variability of the log-normally distributed data, the multiplicative standard deviation (s*) was calculated for the measured results[26 27]. s* is a measure for the variation of $\log_{10}$ normal distributed data and describes the shape of the distribution. When the geometric mean is multiplied with or divided by s*, the resulting values are the higher and lower limits of an interval which covers 68.3% of the mid-range of the distribution [27].

In an exemplary model application the collected human-associated HF183/BacR287 concentrations measured in this study were used for recalibration of the quantitative microbial risk assessment tool QMRAcatch[28]. QMRAcatch is a catchment-based generic, easy-to-use, interactive computational tool to simulate concentrations of fecal indicators and intestinal pathogens at a point of interest (e.g., recreational water uses or drinking water production) and to assess associated microbial infection risks. Host-associated genetic fecal markers are

used to calibrate the model for the specific situation of fecal emissions at the considered habitat. QMRAcatch (free download at www.waterandhealth.at) consists of the following model components: i) a hydrological process model including fate and transport of health-related microbes/viruses in rivers and river/floodplain systems, and ii) QMRA for drinking water safety management or during recreation/bathing activities for the investigated environment. The necessary input data consist of measured MST-marker concentrations and measured or assumed pathogen data in the fecal pollution sources (raw and treated wastewater). The model output consists of simulated concentrations of health-related microbes/viruses in the wastewater and the receiving water, and the treatment requirements (log-reductions) for health-related water safety management[28].

For the exemplary QMRAcatch model application the case study of Derx et al. (2016)[29] at the Danube River in Austria was used. The collected HF183/BacR287 MST marker concentrations in raw and treated wastewater from rural areas (data set from this paper, n=18) were used as input data set for the five selected wastewater treatment plants emitting into the Danube River (details on the methods are provided in the SI) in order to evaluate the general applicability of the recovered data set from around the globe as a surrogate for raw and treated waste emission concentrations of human-associated genetic markers.

3 RESULTS

3.1 Occurrence of human-associated genetic markers in wastewater

Selected human-associated genetic fecal markers were measured in raw and biologically treated wastewater from 29 municipal and rural wastewater treatment plants (Table 1). Genetic markers were detected in all raw and treated wastewater samples (100%). IAC testing and measurement at different sample dilutions confirmed the absence of qPCR amplification inhibition.

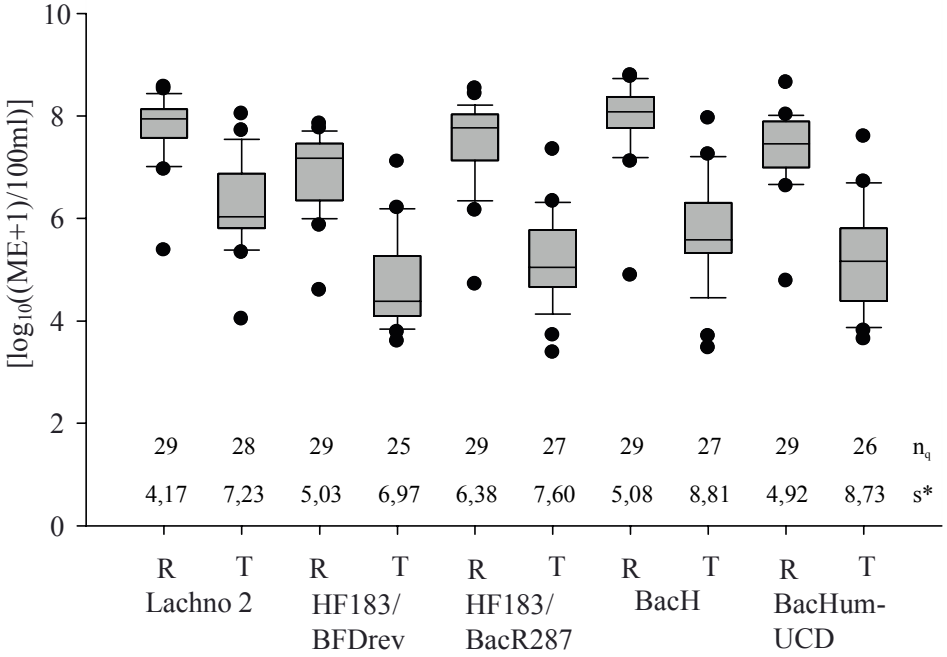


Figure 1: Concentration of human-associated MST markers in raw (R) and biologically treated (T) wastewater. ME: marker equivalents, n_q: number of quantifiable samples out of total of 29 samples each, s*: multiplicative standard deviation, boxes cover the 25th to 75th percentile; line within the boxes, median; whiskers, 10^{th} to 90^{th} percentile, solid circles represent outliers, respectively.

3.2 Concentrations of human-associated genetic markers in wastewater

The concentration of all human-associated genetic markers in raw and treated wastewater samples is shown in Figure 1. HF183/BFDrev showed the lowest concentration of all markers in raw wastewater with a median of $\log_{10}$ 7.2 ME 100 ml^{-1}, whereas, HF183/BacR287 and BacH genetic markers concentrations were slightly higher with medians of $\log_{10}$ 7.8 and $\log_{10}$ 7.8 ME 100 ml^{-1}, respectively. BacHum-UCD was detected with a median of $\log_{10}$ 7.5 ME 100 ml^{-1} and Lachno2 with a median of $\log_{10}$ 8.0 ME 100 ml^{-1}. Only HF183/BDFrev was significantly different from other markers (Lachno2 and BacH) in raw wastewater (Tukey test, Bonferroni corrected significance $p<0.001$, See Table S4).

HF183/BFDrev also showed the lowest concentrations of all assays in treated wastewater with a median of $\log_{10}$ 4.6 ME 100 ml^{-1}, whilst the BacHum-UCD, HF183/BacR287 and BacH genetic markers were higher with medians of $\log_{10}$ 5.2, $\log_{10}$ 5.3 and $\log_{10}$ 5.3 ME 100 ml^{-1}, respectively. Lachno2 had a median concentration of $\log_{10}$ 6.0 ME 100 ml^{-1} in treated wastewater. In treated wastewater only Lachno2 results were significantly different from all other markers except for BacH (Tukey test, Bonferroni corrected significance $p<0.001$, See Table S5).

No statistical differences were observed between wastewater collected in rural and urban areas (Mann-Whitney Rank Sum Test, $p <0.05$, Bonferroni corrected, anonymized raw data is presented in Tables S2 and S3). Hence, data from rural and urban locations were pooled for all subsequent analyses. The authors also refrained from comparing the dataset on a WWTP to WWTP or country-to-country basis, because the sample numbers in each separate country were too low to allow meaningful conclusions to be drawn. Correlation analysis of the pooled dataset revealed a statistically significant association between all five genetic markers ($p <$

313 0.001) with corresponding Spearman rank coefficients ranging from 0.83 to 0.91 in raw
314 sewage and from 0.86 to 0.93 in treated wastewater (Table S6).
315 To investigate variability between datasets, multiplicative standard deviation s* analysis was
316 used. In raw wastewater, the s* for BacHum-UCD (s* = 4.9), HF183/BFDrev (s* = 5.0), and
317 BacH (s*=4.2) were very similar, except HF183/BacR287 with somewhat higher variability
318 (s* = 6.4). In contrast, variability in biologically treated wastewater was much higher with s*
319 values increasing by an average factor of 1.5 (range 1.2 – 1.8) during treatment (Figure 1).
320

321 3.3 Reductions in marker concentrations during wastewater
322 treatment

323 Monte Carlo simulation was used to estimate genetic marker reduction during wastewater
324 treatment. The median $\log_{10}$ genetic marker reductions achieved by secondary wastewater
325 treatment (without disinfection) were 2.1 for BacHum-UCD, 2.2 for HF183/BFDrev, 2.3 for
326 HF183/BacR287 and 2.2 for BacH. Lachno2 showed a lower reduction compared to
327 Bacteroidetes genetic markers with a median of $\log_{10}$ 1,7 (Figure 2 and Figure S1 – S5).
328
329

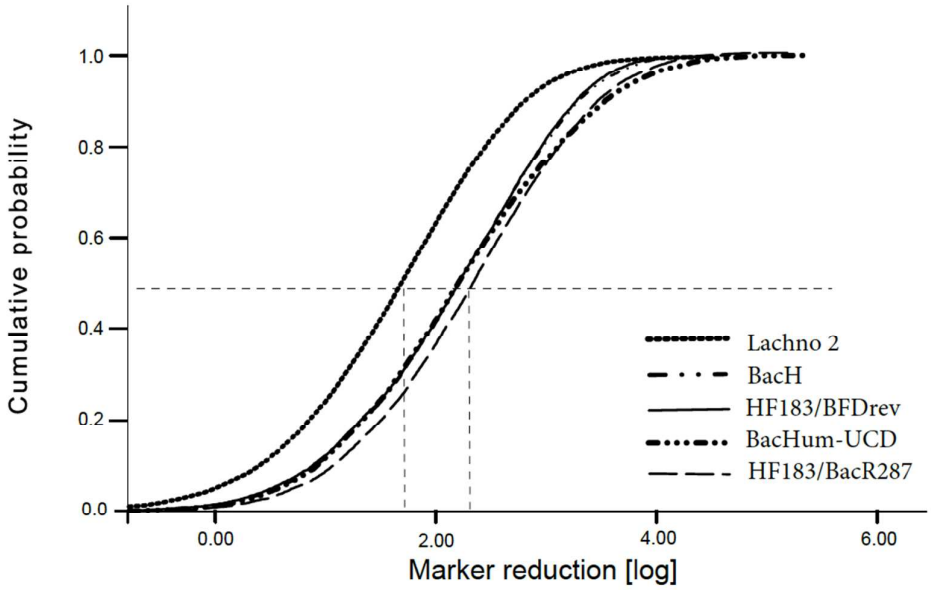
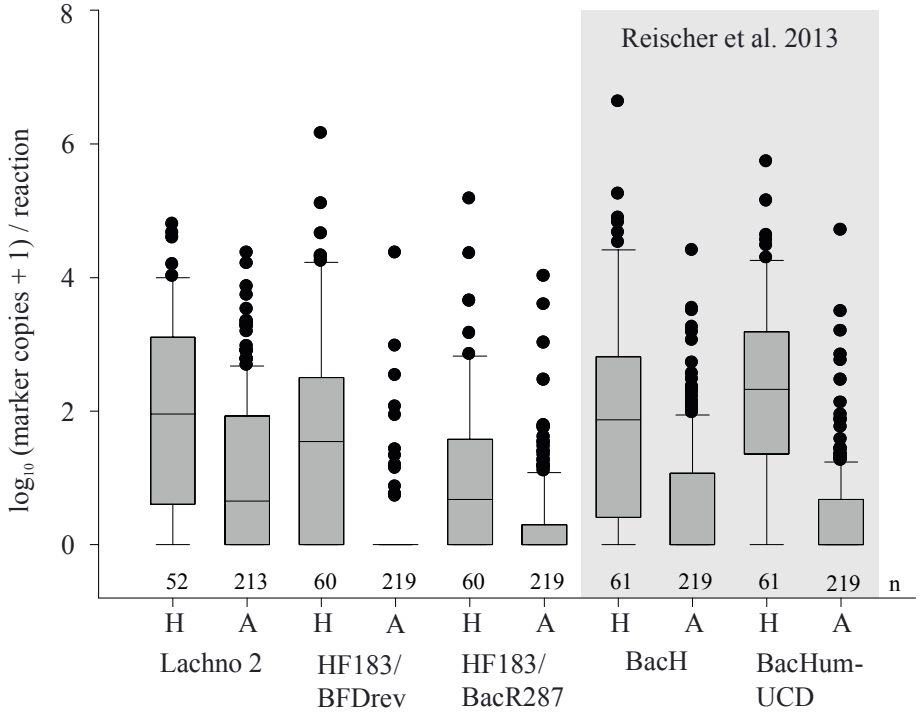


Figure 2: Cumulative distribution function of the Monte Carlo simulated marker reduction values. Dashed horizontal line denotes the 0.5 cumulative probability, corresponding reduction values represents median values (exemplarily highlighted with dashed vertical lines for Lachno2 and HF183/BacR287 respectively).

3.4 Performance trends based on reference fecal sample testing

Relative distributions of false-negative and false positive results were estimated for HF183/BFDrev, HF183/BacR287 and Lachno2 using a previously reported reference fecal sample collection[16]. The detection frequency of genetic markers in reference human samples (target source) was 83% for Lachno2, 58% for HF183/BFDrev, and 62% for HF183/BacR287 compared to previously published values of 77% for BacH and 87% for BacHum-UCD in the same DNA extracts[16]. In the human reference samples, Lachno2 showed the highest median concentration ($\log_{10}$ 2.0 copies per reaction). The HF183/BFDrev marker was detected with a median concentration of $\log_{10}$ 1.6 copies per reaction and HF183/BacR287 with the lowest median concentration of $\log_{10}$ 0.7 copies per reaction (Fig. 3).

346 False-positive rates in animal fecal DNA extracts for the investigated assays were 52% for
347 Lachno2, 5% for HF183/BFDrev, and 27% for HF183/BacR287 (compared to previously
348 published values of 47% for BacH and 32% for BacHum-UCD). Among the tested assays,
349 Lachno2 showed the highest incidence of 'false positives', with a median $\log_{10}$ 1.0 copies per
350 reaction in non-human reference samples. In contrast, the newly reported HF183/BacR287
351 and HF183/BFDrev did not reveal any detectable signals in most non-target samples (Figure
352 3). Marker concntrations were also related to DNA concentrations in the DNA extracts. The
353 distributions of marker concentrations g^{-1} fecal DNA are presented in Figure S6 displaying
354 the same relative distributions as Figure 2. Furthermore, correlation analysis of the
355 concentrations of all the markers in the fecal samples was performed to investigate differences
356 between the human-associated markers. The corresponding Spearman's Rank coefficients
357 ranged from 0.25 to 0.76 (see Table S7, Figure S7).
358



359

360 Figure 3: Genetic marker copies per reaction measured in human (H) and other animal (A)
361 fecal DNA extracts for human-associated genetic markers (grey box previously published
362 data [16]). Results were measured in the 1:4 dilution of the DNA samples and transformed into
363 logarithmic format after addition of 1 to each value. Boxes, 25th and 75th percentile; lines
364 within the boxes, median; whiskers, 10th and 90th percentile, solid circles represent outliers,
365 respectively; n, number of samples in each category.

366

367 4 DISCUSSION

368 4.1 Human-associated genetic markers are widely distributed across 369 the world

370 Human-associated MST genetic markers investigated in this study were ubiquitous in raw
371 (untreated) and biologically treated wastewater samples collected across the world. Genetic
372 markers were detected in 100% of wastewater samples irrespective of the wastewater type

(raw or treated), provenance (all countries), site location (urban or rural), or connected population size. This in itself is an interesting and noteworthy observation, particularly considering the variety of sampling sites from rural wastewater in developing countries such as Tanzania to urban wastewater in highly industrialized countries such as the United Kingdom or Singapore. Other local or regional studies have also reported a high detection frequency of commonly used human-associated molecular genetic markers in wastewater [30-32], but a worldwide distribution has not been previously demonstrated on such a broad geographic level. Pervasive detection of these human-associated genetic markers in wastewater is consistent with the broad occurrence of these markers in fecal samples from around the globe[16] This supports the hypothesis that the target cells belong to the human core intestinal microbiome across populations[33] and underlines the potential for implementation of these methods on a global scale.

4.2 Human-associated genetic markers are highly concentrated in raw and treated sewage

High genetic marker concentrations (10^6 to 10^8 ME 100 ml^{-1}) were found in raw municipal wastewater in all sampled locations suggesting that these markers allow for the detection of raw sewage in environmental waters in water quality management applications [34].

Considering the diverse sample set analyzed in this study, our findings indicate low variability in human-associated genetic marker concentrations. This low variability across sewage samples is highly relevant for the future application of genetic MST modeling approaches such as source apportionment [23, 34] or the support of quantitative microbial risk assessment (QMRA) [28, 35, 36].

An example of such an application is the recently developed 'QMRAcatch' tool, which integrates QMRA with catchment-based hydrological process modelling to predict fecal

pollution levels as well as the associated infection risk for bathing or drinking water. It employs MST markers for source-specific calibration and verification of the hydrological water quality model and uses reference pathogens to simulate pollution and infection risk scenarios using QMRA [28]. QMRAcatch has been used to simulate human-associated fecal pollution in a complex river/floodplain area and for estimating the required reductions of microorganisms and viruses to ensure safe water supply [29].

To assess their usefulness for modeling purposes, HF183/BacR287 concentrations found in raw and treated wastewater at rural WWTPs in this current study were used to recalibrate the QMRAcatch model applied in the previous study on human-associated fecal pollution[29] (details on the method are provided in the SI). The new data could successfully replace the original calibration data from Austrian rural WWTP [26] as model input. Figure S8 and Table S8 show that using the novel global dataset to simulate values for marker concentrations in wastewater sources and receiving waters in the study area resulted in an equally tight fit of the simulated with observed concentrations at the sampling sites in the catchment. Thus, the results were highly compatible with the original outcomes based solely on Austrian data [29]. Beyond being a demonstration for the applicability of the MST marker data in modeling approaches, this result also indicates that the data collected in this study might serve as a best available approximation of marker levels in areas where no data on human-associated marker concentrations in wastewater currently exists. Other applications of MST data include source-specific fecal contaminant transport modeling[37] and epidemiological investigations [38].

In general, the measured marker concentration levels correspond to those found in a recent study [26] which investigated the occurrence of MST genetic markers in Austrian and German WWTPs ranging from small, household-sized plants, to facilities serving large populations over the course of a 12-month period. The results of both recent and the current studies have important implications for wastewater treatment efficacy testing. Most human-associated

423 genetic markers exhibited a two order of magnitude reduction after wastewater treatment with
424 the exception of Lachno2. Lachnospiraceae are Gram-positive bacteria and may be more
425 resilient to treatment processes or even capable of growing in specific niches within sewerage
426 systems and treatment plants [39, 40]. It should be emphasized that wastewater investigated in
427 this study went through primary and secondary (biological) treatment, but no advanced
428 tertiary treatment such as ultraviolet radiation disinfection or chlorination. The tertiary
429 treatment stage was omitted from this investigation because methods are very diverse and
430 many participating countries have not implemented any wastewater disinfection technologies.
431 Nevertheless, reduction values found in this study were similar to other studies investigating
432 reduction values of bacterial and viral genetic markers in wastewater treatment, both with and
433 without disinfection [26, 41, 42]. However, future studies investigating the influence of
434 disinfection on human-associated genetic markers are warranted.
435 Municipal wastewater plays an important role in the pathway of human fecal pollution and
436 associated pathogenic agents entering the environment and ultimately affecting public health
437 [7]. The high concentration of genetic markers found in wastewater samples during this study
438 provides further evidence to demonstrate that these MST approaches serve as useful
439 indicators for the detection of sewage pollution in impacted surface waters.
440

441 ## 4.3 Comparison of human-associated genetic marker trends

442 Careful examination of human-associated genetic marker occurrence in wastewater and fecal
443 samples suggests that some markers may be more suitable than others for water quality
444 management. While correlations between human-associated genetic marker concentrations
445 were strong among wastewater samples, a different trend was observed with human fecal
446 DNA tests. This is likely due to the composition of wastewater representing fecal waste from
447 a group of individuals resulting in a homogenized mixture. In contrast, individual fecal

448 samples contain genetic markers from a single gut system potentially reducing the sensitivity
449 of a human-associated MST method. It should also be noted that the fecal DNA reference
450 samples used in this study contained relatively low DNA concentration due to extensive DNA
451 purification. This led to higher limits of detection for these samples and correspondingly to a
452 generally higher false-negative rate and a lower false positive rate than might have been
453 observed in more concentrated samples. The HF183/BFDrev assay in particular has been
454 suspected to be unable to detect very low marker copy numbers[8]. Also the fecal DNA extracts
455 had been stored at -80°C for several years between the two studies which might also affect
456 marker concentrations. Another important factor to consider is the MST genetic marker itself.
457 Results indicate strong correlations in genetic marker concentrations between HF183/BFDrev,
458 HF183/BacR287, and BacHum-UCD, while correlations to BacH were much weaker. In fact,
459 HF183 genetic markers share the same forward primer while the BacHum-UCD forward
460 primer has a 16-base overlap with the HF183 forward primer [8, 9, 11, 43]. Therefore, these three
461 genetic markers likely detect the same human-associated Bacteroidetes clade. However, there
462 are differences in the performance of these Bacteroidetes genetic markers, with slightly higher
463 concentrations and correspondingly lower false-negative rate for BacHum-UCD, contrasted
464 by lower false-positive rate for HF183/BFDrev and HF183/BacR287. This "trade-off"
465 between source-sensitivity and -specificity is often encountered in MST approaches [44]. In
466 contrast to the Bacteroidetes genetic markers, Lachno2 targets Firmicutes contributing to a
467 different performance pattern with slightly higher concentrations in wastewater and human
468 feces, but high concentrations in animal fecal samples. This difference in performance could
469 have important ramifications for future water quality applications.
470

471 ## 4.4 Implications for Water Quality Management

Our findings demonstrate that human-associated genetic markers tested in this study are highly sensitive tools for the detection and quantification of sewage contamination across six continents. However, the observed lower sensitivity with individual human fecal samples suggests that these genetic markers may not be as useful in scenarios where few individuals are contributing to the human fecal pollution load. In addition, no genetic marker achieved 100% specificity indicating that a single MST method may not be suitable across all geographic locations and the importance of verifying sensitivity and specificity with local reference samples prior to initiating a MST water quality study. Other strategies such as source profiling[16], the use of conditional probabilities [9, 45], or machine learning approaches [46] could also help to evaluate the utility of a particular MST genetic marker or group of markers to correctly identify human fecal contamination. Limitations in source-specificity might also be compensated by combining bacterial MST genetic markers with promising viral methods[47] [26] or human mitochondrial DNA approaches [48]. In addition, study findings may have important implications for calibrating future microbial fecal pollution and QMRA models using novel genetic marker occurrence information from reference samples[29].

Supporting Information (SI) is available and contains additional Experimental Methods, 8 additional tables, 8 additional figures.

ACKNOWLEDGEMENTS

This paper was realized by the support of the Austrian Science Fund (FWF) as part of the research projects P23900 granted to Andreas H. Farnleitner. Further support came also from the Science Call 2015 "Ressource und Lebensgrundlage Wasser“ Project SC15-016 funded by the Niederösterreichische Forschungs- und Bildungsgesellschaft (NFB), the FWF research project P22309 (both granted to AHF), and the FWF research project P22032 (granted to

497 Georg Reischer). This study is a joint publication of the Interuniversity Cooperation Center
498 Water & Health (www.waterandhealth.at). Any opinions, findings, and conclusions or
499 recommendations expressed in this material are those of the author(s) and do not necessarily
500 reflect the views of the United States Environmental Protection Agency.

501 We would like to dedicate this manuscript to our fellow author, collaborator and dear friend
502 Professor Huw Taylor, who was sadly never able to see its publication. Huw was a Professor
503 of Microbial Ecology at the University of Brighton (UK) and will be much missed by his
504 fellow scientists from around the World.

505

506

507 REFERENCES

508

509 1. Pruss, A.; Kay, D.; Fewtrell, L.; Bartram, J., Estimating the burden of disease from
510 water, sanitation, and hygiene at a global level. *Environmental Health Perspectives* **2002,**
511 *110*, (5), 537-542.
512 2. Troeger, C.; Forouzanfar, M.; Rao, P. C.; Khalil, I.; Brown, A.; Reiner, R. C.;
513 Fullman, N.; Thompson, R. L.; Abajobir, A.; Ahmed, M., et al., Estimates of global, regional,
514 and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis
515 for the Global Burden of Disease Study 2015. *Lancet Infect. Dis.* **2017,** *17*, (9), 909-948.
516 3. WHO, Water Quality and Health Strategy 2013-2020. **2013**.
517 4. ISO, Microbiology of food and animal feeding stuffs -- Horizontal method for the
518 enumeration of beta-glucuronidase-positive *Escherichia coli* Colony-count technique at 44
519 degrees C using membranes and 5-bromo-4-chloro-3-indolyl beta-D-glucoronide (ISO 16649-
520 1:2001 04 15) International Organisation of Standardisation, Geneva, Switzerland. **2001**.
521 5. ISO, Water Quality – Detection and Enumeration of Intestinal Enterococci – Part 2:
522 Membrane Filtration Method (ISO 7899-2: 2000). Geneva, Switzerland: International
523 Organization of Standardization. **2000**.
524 6. Farnleitner, A. H.; Ryzinska-Paier, G.; Reischer, G. H.; Burtscher, M. M.; Knetsch, S.;
525 Kirschner, A. K. T.; Dirnboeck, T.; Kuschnig, G.; Mach, R. L.; Sommer, R., *Escherichia coli*
526 and enterococci are sensitive and reliable indicators for human, livestock and wildlife faecal
527 pollution in alpine mountainous water resources. *J. Appl. Microbiol.* **2010,** *109*, (5), 1599-
528 1608.
529 7. Hagedorn, C.; Harwood, V. J.; Blanch, A., *Microbial Source Tracking: Methods,*
530 *Applications, and Case Studies*. Springer: New York, USA, 2011; p 642.
531 8. Green, H. C.; Haugland, R. A.; Varma, M.; Millen, H. T.; Borchardt, M. A.; Field, K.
532 G.; Walters, W. A.; Knight, R.; Sivaganesan, M.; Kelty, C. A., et al., Improved HF183
533 Quantitative Real-Time PCR Assay for Characterization of Human Fecal Pollution in
534 Ambient Surface Water Samples. *Appl. Environ.Microbiol.* **2014,** *80*, (10), 3086-3094.
535 9. Kildare, B. J.; Leutenegger, C. M.; McSwain, B. S.; Bambic, D. G.; Rajal, V. B.;
536 Wuertz, S., 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and
537 dog-specific fecal Bacteroidales: A Bayesian approach. *Water Res.* **2007,** *41*, (16), 3701-
538 3715.
539 10. Ahmed, W.; Goonetilleke, A.; Powell, D.; Gardner, T., Evaluation of multiple sewage-
540 associated *Bacteroides* PCR markers for sewage pollution tracking. *Water Res.* **2009,** *43*,
541 (19), 4872-4877.
542 11. Haugland, R. A.; Varma, M.; Sivaganesan, M.; Kelty, C.; Peed, L.; Shanks, O. C.,
543 Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative
544 detection of selected Bacteroidales species and human fecal waste by qPCR. *Syst. Appl.*
545 *Microbiol.* **2010,** *33*, (6), 348-357.
546 12. Boehm, A. B.; Van De Werfhorst, L. C.; Griffith, J. F.; Holden, P. A.; Jay, J. A.;
547 Shanks, O. C.; Wang, D.; Weisberg, S. B., Performance of forty-one microbial source
548 tracking methods: A twenty-seven lab evaluation study. *Water Res.* **2013,** *47*, (18), 6812-
549 6828.
550 13. Ahmed, W.; Sritharan, T.; Palmer, A.; Sidhu, J. P. S.; Toze, S., Evaluation of Bovine
551 Feces-Associated Microbial Source Tracking Markers and Their Correlations with Fecal
552 Indicators and Zoonotic Pathogens in a Brisbane, Australia, Reservoir. *Appl.*
553 *Environ.Microbiol.* **2013,** *79*, (8), 2682-2691.
554 14. Shanks, O. C.; White, K.; Kelty, C. A.; Sivaganesan, M.; Blannon, J.; Meckes, M.;
555 Varma, M.; Haugland, R. A., Performance of PCR-Based Assays Targeting Bacteroidales

556 Genetic Markers of Human Fecal Pollution in Sewage and Fecal Samples. *Environ. Sci.*
557 *Technol.* **2010,** *44*, (16), 6281-6288.
558 15. Yahya, M.; Blanch, A. R.; Meijer, W. G.; Antoniou, K.; Hmaied, F.; Balleste, E.,
559 Comparison of the Performance of Different Microbial Source Tracking Markers among
560 European and North African Regions. *J Environ Qual* **2017,** *46*, (4), 760-766.
561 16. Reischer, G. H.; Ebdon, J. E.; Bauer, J. M.; Schuster, N.; Ahmed, W.; Astrom, J.;
562 Blanch, A. R.; Bloschl, G.; Byamukama, D.; Coakley, T., et al., Performance Characteristics
563 of qPCR Assays Targeting Human- and Ruminant-Associated Bacteroidetes for Microbial
564 Source Tracking across Sixteen Countries on Six Continents. *Environ. Sci. Technol.* **2013,** *47*,
565 (15), 8548-8556.
566 17. Griffiths, R. I.; Whiteley, A. S.; O'Donnell, A. G.; Bailey, M. J., Rapid method for
567 coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA-
568 and rRNA-based microbial community composition. *Appl. Environ.Microbiol.* **2000,** *66*, (12),
569 5488-5491.
570 18. Reischer, G. H.; Kasper, D. C.; Steinborn, R.; Mach, R. L.; Farnleitner, A. H.,
571 Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and
572 evaluation of this method in alpine karstic regions. *Appl. Environ.Microbiol.* **2006,** *72*, (8),
573 5610-5614.
574 19. Reischer, G. H.; Kasper, D. C.; Steinborn, R.; Farnleitner, A. H.; Mach, R. L., A
575 quantitative real-time PCR assay for the highly sensitive and specific detection of human
576 faecal influence in spring water from a large alpine catchment area. *Lett. Appl. Microbiol.*
577 **2007,** *44*, (4), 351-356.
578 20. Newton, R. J.; VandeWalle, J. L.; Borchardt, M. A.; Gorelick, M. H.; McLellan, S. L.,
579 Lachnospiraceae and Bacteroidales alternative fecal indicators reveal chronic human sewage
580 contamination in an urban harbor. *Appl. Environ.Microbiol.* **2011,** *77*, (19), 6972-6981.
581 21. Layton, A.; McKay, L.; Williams, D.; Garrett, V.; Gentry, R.; Sayler, G.,
582 Development of Bacteroides 16S rRNA gene TaqMan-based real-time PCR assays for
583 estimation of total, human, and bovine fecal pollution in water. *Appl. Environ.Microbiol.*
584 **2006,** *72*, (6), 4214-4224.
585 22. Anderson, A.; Pietsch, K.; Zucker, R.; Mayr, A.; Müller-Hohe, E.; Messelhäusser, U.;
586 Sing, A.; Busch, U.; Huber, I., Validation of a Duplex Real-Time PCR for the Detection of
587 Salmonella spp. in Different Food Products. *Food Analytical Methods* **2010,** *4*, (3), 259-267.
588 23. Reischer, G. H.; Kollanur, D.; Vierheilig, J.; Wehrspaun, C.; Mach, R. L.; Sommer,
589 R.; Stadler, H.; Farnleitner, A. H., Hypothesis-Driven Approach for the Identification of Fecal
590 Pollution Sources in Water Resources. *Environ. Sci. Technol.* **2011,** *45*, (9), 4038-4045.
591 24. Piringer, H.; Berger, W.; Krasser, J., HyperMoVal: Interactive Visual Validation of
592 Regression Models for Real-Time Simulation. *Computer Graphics Forum* **2010,** *29*, (3), 983-
593 992.
594 25. Bonferroni, C. E., Teoria statistica delle classi e calcolo delle probabilità.
595 *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **1936,**
596 *8*, 3-62.
597 26. Mayer, R. E.; Bofill-Mas, S.; Egle, L.; Reischer, G. H.; Schade, M.; Fernandez-Cassi,
598 X.; Fuchs, W.; Mach, R. L.; Lindner, G.; Kirschner, A., et al., Occurrence of human-
599 associated Bacteroidetes genetic source tracking markers in raw and treated wastewater of
600 municipal and domestic origin and comparison to standard and alternative indicators of faecal
601 pollution. *Water Res.* **2016,** *90*, 265-276.
602 27. Limpert, E.; Stahel, W. A.; Abbt, M., Log-normal distributions across the sciences:
603 Keys and clues. *Bioscience* **2001,** *51*, (5), 341-352.

28. Schijven, J.; Derx, J.; Husman, A. M. D.; Blaschke, A. P.; Farnleitner, A. H., QMRAcatch: Microbial Quality Simulation of Water Resources including Infection Risk Assessment. *J. Environ. Qual.* **2015,** *44*, (5), 1491-1502.
29. Derx, J.; Schijven, J.; Sommer, R.; Zoufal-Hruza, C. M.; van Driezum, I. H.; Reischer, G.; Ixenmaier, S.; Kirschner, A.; Frick, C.; Husman, A. M. D., et al., QMRAcatch: Human-Associated Fecal Pollution and Infection Risk Modeling for a River/Floodplain Environment. *J. Environ. Qual.* **2016,** *45*, (4), 1205-1214.
30. Ahmed, W.; Stewart, J.; Powell, D.; Gardner, T., Evaluation of Bacteroides markers for the detection of human faecal pollution. *Lett. Appl. Microbiol.* **2008,** *46*, (2), 237-242.
31. Layton, B. A.; Cao, Y. P.; Ebentier, D. L.; Hanley, K.; Balleste, E.; Brandao, J.; Byappanahalli, M.; Converse, R.; Farnleitner, A. H.; Gentry-Shields, J., et al., Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Res.* **2013,** *47*, (18), 6897-6908.
32. Ahmed, W.; Masters, N.; Toze, S., Consistency in the host specificity and host sensitivity of the Bacteroides HF183 marker for sewage pollution tracking. *Lett. Appl. Microbiol.* **2012,** *55*, (4), 283-289.
33. Shade, A.; Handelsman, J., Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol* **2012,** *14*, (1), 4-12.
34. Wang, D.; Farnleitner, A. H.; Field, K. G.; Green, H. C.; Shanks, O. C.; Boehm, A. B., Enterococcus and Escherichia coli fecal source apportionment with microbial source tracking genetic markers--is it feasible? *Water Res* **2013,** *47*, (18), 6849-61.
35. Staley, C.; Gordon, K. V.; Schoen, M. E.; Harwood, V. J., Performance of Two Quantitative PCR Methods for Microbial Source Tracking of Human Sewage and Implications for Microbial Risk Assessment in Recreational Waters. *Appl. Environ.Microbiol.* **2012,** *78*, (20), 7317-7326.
36. Brown, K. I.; Graham, K. E.; Soller, J. A.; Boehm, A. B., Estimating the probability of illness due to swimming in recreational water with a mixture of human- and gull-associated microbial source tracking markers. *Environ Sci Process Impacts* **2017,** *19*, (12), 1528-1541.
37. Sokolova, E.; Astrom, J.; Pettersson, T. J.; Bergstedt, O.; Hermansson, M., Decay of Bacteroidales genetic markers in relation to traditional fecal indicators for water quality modeling of drinking water sources. *Environ Sci Technol* **2012,** *46*, (2), 892-900.
38. Odagiri, M.; Schriewer, A.; Daniels, M. E.; Wuertz, S.; Smith, W. A.; Clasen, T.; Schmidt, W. P.; Jin, Y.; Torondel, B.; Misra, P. R., et al., Human fecal and pathogen exposure pathways in rural Indian villages and the effect of increased latrine coverage. *Water Res* **2016,** *100*, 232-244.
39. McLellan, S. L.; Newton, R. J.; Vandewalle, J. L.; Shanks, O. C.; Huse, S. M.; Eren, A. M.; Sogin, M. L., Sewage reflects the distribution of human faecal Lachnospiraceae. *Environ. Microbiol.* **2013,** *15*, (8), 2213-2227.
40. McLellan, S. L.; Huse, S. M.; Mueller-Spitz, S. R.; Andreishcheva, E. N.; Sogin, M. L., Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ. Microbiol.* **2010,** *12*, (2), 378-392.
41. Pouillot, R.; Van Doren, J. M.; Woods, J.; Plante, D.; Smith, M.; Goblick, G.; Roberts, C.; Locas, A.; Hajen, W.; Stobo, J., et al., Meta-Analysis of the Reduction of Norovirus and Male-Specific Coliphage Concentrations in Wastewater Treatment Plants. *Appl Environ Microbiol* **2015,** *81*, (14), 4669-81.
42. B. J. Cisneros, J. B. R. Global Water Pathogen Project (GWPP). http://www.waterpathogens.org (2018.01.03),
43. Bernhard, A. E.; Field, K. G., A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. *Appl. Environ.Microbiol.* **2000,** *66*, (10), 4571-4574.

654 44. Harwood, V. J.; Staley, C.; Badgley, B. D.; Borges, K.; Korajkic, A., Microbial source
655 tracking markers for detection of fecal contamination in environmental waters: relationships
656 between pathogens and human health outcomes. *Fems Microbiol. Rev.* **2014,** *38*, (1), 1-40.
657 45. Wuertz, S.; Wang, D.; Reischer, G. H.; Farnleitner, A. H., Library-Independent Source
658 Tracking Methods. In *Microbial Source Tracking: Methods, Applications, and Case Studies*,
659 Hagedorn, C.; Blanch, A. R.; Harwood, V. J., Eds. Springer: New York, USA, 2011; pp 61-
660 113.
661 46. Casanovas-Massana, A.; Gomez-Donate, M.; Sanchez, D.; Belanche-Munoz, L. A.;
662 Muniesa, M.; Blanch, A. R., Predicting fecal sources in waters with diverse pollution loads
663 using general and molecular host-specific indicators and applying machine learning methods.
664 *Journal of Environmental Management* **2015,** *151*, 317-325.
665 47. Ogilvie, L. A.; Bowler, L. D.; Caplin, J.; Dedi, C.; Diston, D.; Cheek, E.; Taylor, H.;
666 Ebdon, J. E.; Jones, B. V., Genome signature-based dissection of human gut metagenomes to
667 extract subliminal viral sequences. *Nat Commun* **2013,** *4*, 2420.
668 48. He, X. W.; Liu, P.; Zheng, G. L.; Chen, H. M.; Shi, W.; Cui, Y. B.; Ren, H. Q.; Zhang,
669 X. X., Evaluation of five microbial and four mitochondrial DNA markers for tracking human
670 and pig fecal pollution in freshwater. *Sci Rep-Uk* **2016,** *6*.
671
672

673