

OPEN

Faecal pollution affects abundance and diversity of aquatic microbial community in anthropo-zoogenically influenced lotic ecosystems

Lisa Paruch, Adam M. Paruch ^{*}, Hans Geir Eiken & Roald Sørheim

The aquatic microbiota is known to be an important factor in the sustainability of the natural water ecosystems. However, the microbial community also might include pathogens, which result in very serious waterborne diseases in humans and animals. Faecal pollution is the major cause of these diseases. Therefore, it is of immense importance to assess the potential impact of faecal pollution, originating from both anthropogenic and zoogenic sources, on the profile of microbial communities in natural water environments. To this end, the microbial taxonomic diversity of lotic ecosystems in different regions of Norway, representing urban and rural areas, exposed to various levels of faecal pollution, was investigated over the course of a 1-year period. The highest microbial diversity was found in rural water that was the least faecally polluted, while the lowest was found in urban water with the highest faecal contamination. The overall diversity of the aquatic microbial community was significantly reduced in severely polluted water. In addition, the community compositions diverged between waters where the dominant pollution sources were of anthropogenic or zoogenic origin. The results provide new insight into the understanding of how faecal water contamination, specifically that of different origins, influences the microbial diversity of natural waters.

Surface water bodies constitute important sources of drinking water, estimated to cover approximately half of the worldwide need¹. Yet, this overall coverage varies regionally, and in Norway, for instance, 85% of drinking water is supplied by surface fresh water sources. Unlike groundwater, which is protected by the overlying soils trapping and attenuating pollutants², surface water bodies are more vulnerable to various contamination sites^{1,2}. The polluted waters pose potential health risks to humans and the environment; thus, quality issues become the core concern for public and regulatory authorities. Water quality assessment considers physical, chemical, and biological aspects. However, considering the main health threats, the microbial profile represents the most important indicator of the biological quality of water environments^{3–5}.

The aquatic microbiota contributes substantially to the sustainability (in terms of balanced physical, chemical and biological conditions, harmonised biodiversity and long-term functionality) of the natural (organic by nature) water ecosystems, playing key roles in organic and inorganic nutrient cycling, respiration, and biological and chemical contaminant removal in aquatic environments⁶. However, it can also pose a negative impact by carrying pathogens, which result in very serious waterborne diseases in both humans and animals^{1,3}. Faecal pollution is the major cause of waterborne disease, since most of the pathogens associated with transmission reside in human and warm-blooded animal faeces. It is therefore unquestionable whether water contamination with enteric pathogens affects the overall quality of aquatic systems. The genuine question is how significant and to what extent it poses a threat to human and environmental health.

Faecal water contamination refers to a dual origin (anthropogenic and zoogenic) of various point and non-point/diffuse pollution sources (comprising discharge of industrial/municipal/domestic wastewater and storm/urban/agricultural water runoffs). It is normally determined by the detection and quantitation of viable bacterial

Norwegian Institute of Bioeconomy Research (NIBIO), Division of Environment and Natural Resources, Fredrik A. Dahls vei 20, Oslo, 1433, Aas, Norway. *email: adam.paruch@nibio.no

indicators, so-called faecal indicator bacteria (FIB), represented primarily by *Escherichia coli* (*E. coli*) and certain *Enterococcus* species such as *E. faecalis* and *E. faecium*⁷. With respect to these FIB and the associated health aspects, various water quality standards and legislations have been established on both global and regional scales, e.g., drinking water guidelines^{3,8}, bathing and recreational water criteria^{9,10}, and management of alternative water resources¹¹. These bacteria; however, cannot define either the origins or sources of the contamination, and therefore, a step forward in research efforts has been taken to identify the actual causes of faecal pollution^{12,13}. This became accessible by implementing microbial source tracking (MST) methods using the most widely applied faecal bacterial source identifier – *Bacteroidales*¹⁴. They include bacteria that are one of the most abundant types in the intestine of humans and other warm-blooded animals. In particular, species of the genus *Bacteroides* comprise the major population of gastrointestinal microbes, and normally constitute approximately 30% of total faecal bacteria; although, this percentage can extend to over 50% of human faecal flora, with concentrations up to 10¹¹ cells per gram of faeces¹⁵. Host-associated molecular markers derived from *Bacteroidales* 16S rRNA genes are by far the most tested/optimised via multiple molecular studies and have exhibited wide geographical stability^{16–20}.

In recent decades, various molecular approaches have been developed to characterise microbial community structure in aquatic ecosystems⁶. Among these, DNA fingerprinting techniques have been one of the most frequently used; however, they are inadequate for taxonomic identification due to weak reproducibility of data and low resolution of community profile^{21,22}. More recently, the advance of next-generation sequencing (NGS) technologies has enabled a deeper and better resolution of taxonomic characterisation; thus, both 454 pyrosequencing and Illumina MiSeq have been employed successfully for aquatic microbiota screening^{23,24}. These new approaches using 16S rRNA and NGS provide significant insight into specific microbial taxonomic characterisation that otherwise is not apparent by traditional microbiological methods^{22,25}.

Although the molecular techniques have proven powerful as tools for studying microbial activities in various water ecosystems, the composition of microbiota remains sparsely addressed in aquatic research⁶. Furthermore, very limited information is available regarding the diversity of water microbial communities impacted by faecal pollution. There are a few studies emphasising these issues, yet only the overall response of bacterial (no other microbes) diversity to anthropogenic (no other origin) faecal pollution is generally addressed^{25,26}. In addition, most of the relevant studies refer to benthic microbiota in marine and lentic ecosystems, which is usually limited to a single domain, primarily *Bacteria*^{22,23}. Hence, the real composition of microbial communities in lotic environments remains largely unknown^{6,27}, and therefore a better picture of microbial diversity in such ecosystems is needed²².

In response to the aforementioned limitations, knowledge gaps, and needs, we applied Illumina MiSeq NGS of the 16S rRNA genes to characterise the microbial taxonomic diversity of faecally polluted lotic ecosystems in different regions of Norway. Faecal pollution is expected to influence the aquatic microbial community structure. Some might argue that faecal microbial communities are much more diverse than freshwater communities, thus higher diversity occurs in faecally polluted waters^{28,29}. Others would argue that faecal contamination has a negative impact, thus reduce the overall aquatic microbial diversity^{26,30,31}. These inconsistent opinions did motivate us to conceive and conduct this work in order to contribute with new insight into the understanding of how faecal pollution affects the microbial diversity of lotic waters. To the best of our knowledge, this is one of the pioneering studies dedicated solely to the application of NGS for the characterisation of aquatic microbial diversity associated with various levels and sources of faecal water contamination in Norway.

Results and Discussion

Faecal water contamination and its origin. The selected study sites (a rural creek – AG, an urban stream – SB, and tributaries of drinking water reservoirs – BJ, OM and TJ, Fig. 1) revealed constant exposure to faecal pollution. Therefore, following the proposed screening criteria (described in “Materials and Methods” section), we investigated samples with an extreme variation in *Escherichia coli* (*E. coli*) concentrations (lowest vs. highest) and those with divergent origins of faecal contamination. The origin was expressed as the percentage contribution profile of host-specific *Bacteroidales* 16S rRNA genetic markers (AllBac, BacH, BacR and Hor-Bac) in the measured faecal contamination. These particular outcomes will not be addressed herein, as they have been described in greater detail elsewhere (detailed results of these examinations are available through open access reports published online at <http://hdl.handle.net/11250/2392445>, <http://hdl.handle.net/11250/2392629>, and <http://hdl.handle.net/11250/2392631>).

The highest *E. coli* numbers were found in the urban stream (SB), while the lowest were detected in the rural creek (AG). Furthermore, the contribution profile of the tested *Bacteroidales* DNA markers exposed the dominant anthropogenic origin in SB vs. zoogenic origin in AG (Table 1). This responded accurately to the divergent type of these two watercourses. The tributaries of drinking water reservoirs (OM, TJ, and BJ) also revealed relatively high *E. coli* numbers. The contribution profile of the tested genetic markers in faecal water contamination exposed the dominant zoogenic origin at the BJ and TJ sites, and exclusive zoogenic origin (100%) at the OM site (Table 1).

Microbial community composition and cross-comparison. In the microbial diversity study, 13,167,241 sequence reads were extracted and obtained following pair-end reads assembly. They were evenly distributed among all the examined samples. After quality filtering, 615,567 unique reads, with an average length of 265 bp, were used in downstream analysis. A total of 49,107 operational taxonomic units (OTUs) were predicted, with 32,272 assigned to bacteria, 1,722 to archaea, and the remaining 15,113 were uncharacterised. The taxonomic classification after assigning the clustered OTUs against the Greengenes Database (at 97% identity level), 74 phyla, consisting of 71 bacteria (96%) and 3 archaea (4%), were characterised, and a further 1,354 genera were identified.

Bacterial population distribution over the studied lotic ecosystems in different regions of Norway is presented in Figs. 2 and 3. The sequencing data analyses revealed that *Proteobacteria* and *Bacteroidetes* were two

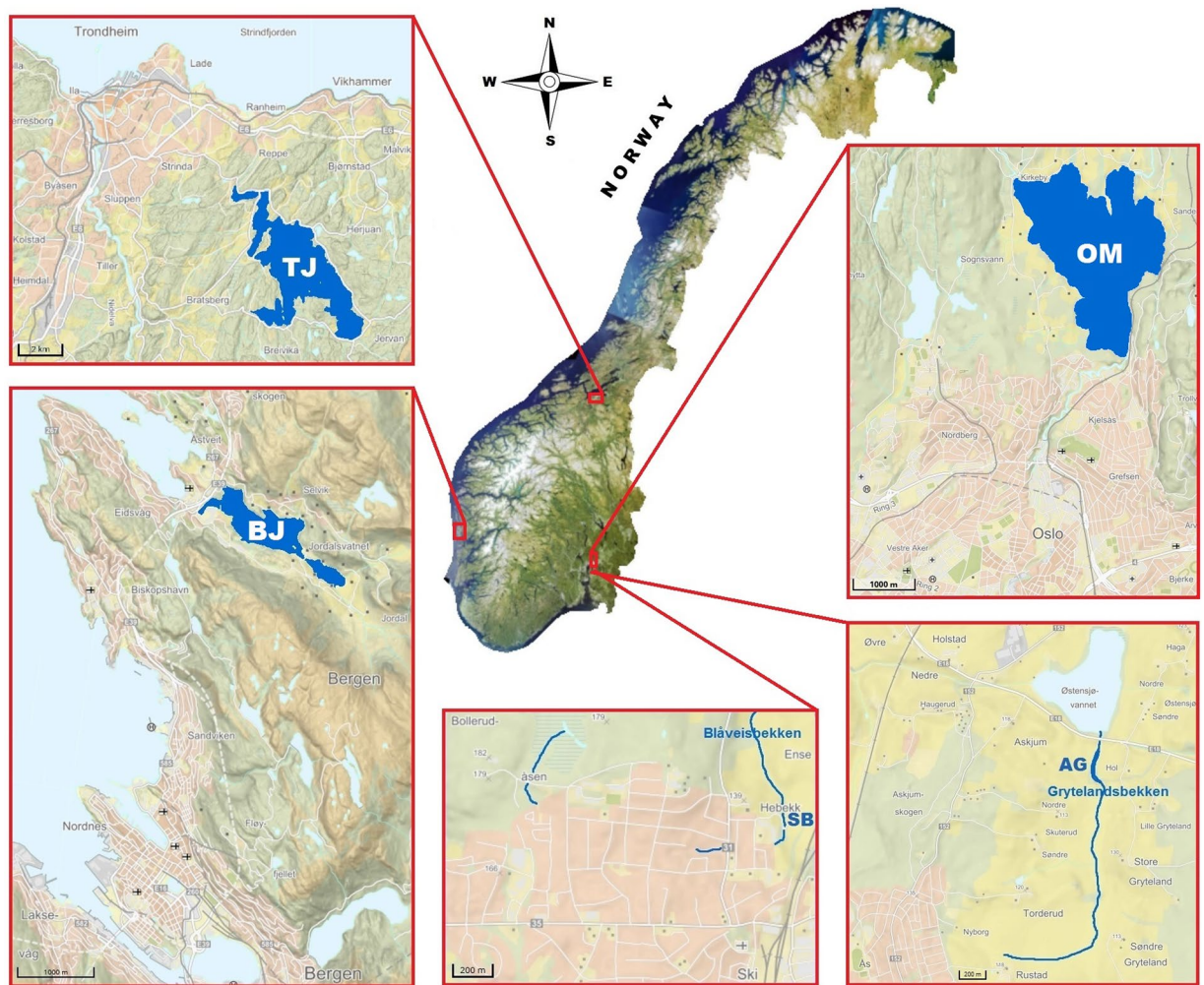


Figure 1. Location of the study sites: TJ – Trondheim, Jonsvannet drinking water reservoir; BJ – Bergen, Jordalsvatnet drinking water reservoir; OM – Oslo, Maridalsvannet drinking water reservoir; AG – Aas, Grytelandsbekken rural creek; and SB – Ski, Blåveisbekken urban stream. The satellite image and maps were obtained from NIBIO’s primary map service the Source/ Kilden (https://kilden.nibio.no/?lang=nb&topic=arealinformasjon&bgLayer=norgebilder_cache2&X=7334000.00&Y=400000.00&zoom=0, https://kilden.nibio.no/?lang=nb&topic=arealinformasjon&bgLayer=farger_cache&X=7037946.92&Y=275666.55&zoom=7#).

Study sites	<i>E. coli</i> counts min – max mean	Dominant origin of faecal contamination in water samples		Total number of tested samples
		anthropogenic	zoogenic	
AG	$\frac{1-144.5}{9}$	1	9	10
SB	$\frac{560-20\,050}{3\,897}$	8	2	10
BJ	$\frac{1-200.5}{23}$	2	13	15
OM	$\frac{2-200.5}{38}$	0	10	10
TJ	$\frac{2-200.5}{58}$	2	13	15

Table 1. Concentrations of *Escherichia coli* (*E. coli*) expressed as the most probable number (MPN) in 100 ml of water sampled in the study sites (AG – Aas, Grytelandsbekken rural creek; SB – Ski, Blåveisbekken urban stream; BJ – Bergen, Jordalsvatnet drinking water reservoir; OM – Oslo, Maridalsvannet drinking water reservoir; and TJ – Trondheim, Jonsvannet drinking water reservoir). Number of samples with the dominant anthropogenic or zoogenic faecal contamination in water.

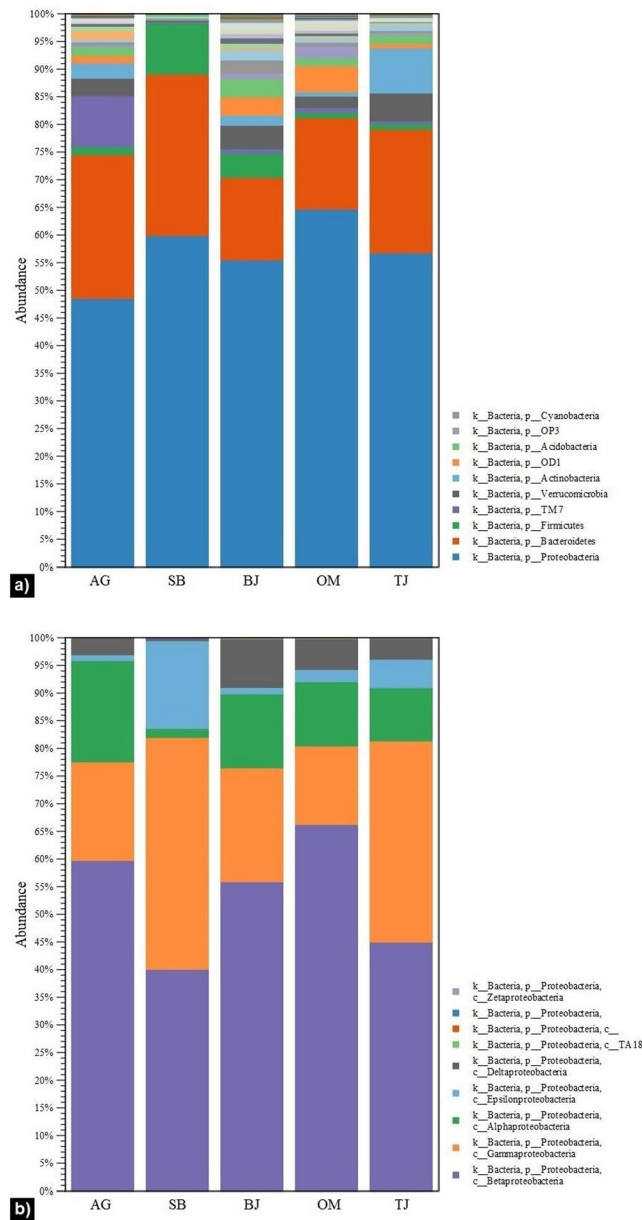


Figure 2. Relative abundance of bacterial populations among all studied sites (AG – Aas, Grytelandsbekken; SB – Ski, Blåveisbekken; BJ – Bergen, Jordalsvatnet; OM – Oslo, Maridalsvannet; and TJ – Trondheim, Jonsvannet): **(a)** all bacterial phyla, **(b)** *Proteobacteria* class distribution. The stacked bar charts were generated using the tools in CLC Microbial Genomics Module version 2.5.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/clc-microbial-genomics-module-latest-improvements>).

predominant phyla, representing 70–90% relative abundance (Fig. 2a). Published findings have reported that these phyla are “the core” bacterial population in lotic waters^{20,26}. Furthermore, our outcomes indicate that *Proteobacteria* contributed mostly (from over 48% to nearly 65%) to the total OTU numbers. They were represented by nine classes (Fig. 2b), but primarily by *Betaproteobacteria*, *Gammaproteobacteria*, and *Alphaproteobacteria*, which is consistent with discoveries from previous relevant studies^{32,33}. The class distribution pattern was very similar in all tested waters, with the exception of site SB (Fig. 2b), which harboured relatively low amounts of *Alphaproteobacteria* (2%), but higher amounts of *Epsilonproteobacteria* (16%). After filtering out these two predominant phyla, we also detected prevalence of other phyla, such as *Firmicutes*, *TM7*, *Verrucomicrobia*, *Actinobacteria*, and *Acidobacteria* (Fig. 3a). For instance, *TM7*, *Verrucomicrobia*, and *Actinobacteria* were three top phyla at site AG; *Firmicutes*, *Verrucomicrobia*, and *Acidobacteria* at site BJ; *OD1*, *OP3*, and *Verrucomicrobia* at site OM; and *Actinobacteria*, *Verrucomicrobia*, and *Acidobacteria* at site TJ. A notable observation was that *Firmicutes* bacteria were found at a remarkably high proportion (85% relative abundance) in the urban stream (site SB with the dominant anthropogenic faecal contamination in water). Furthermore, 52% of this phylum was assigned to the genus *Faecalibacterium* (Fig. 3b). While *Firmicutes* constitute the largest portion of the human

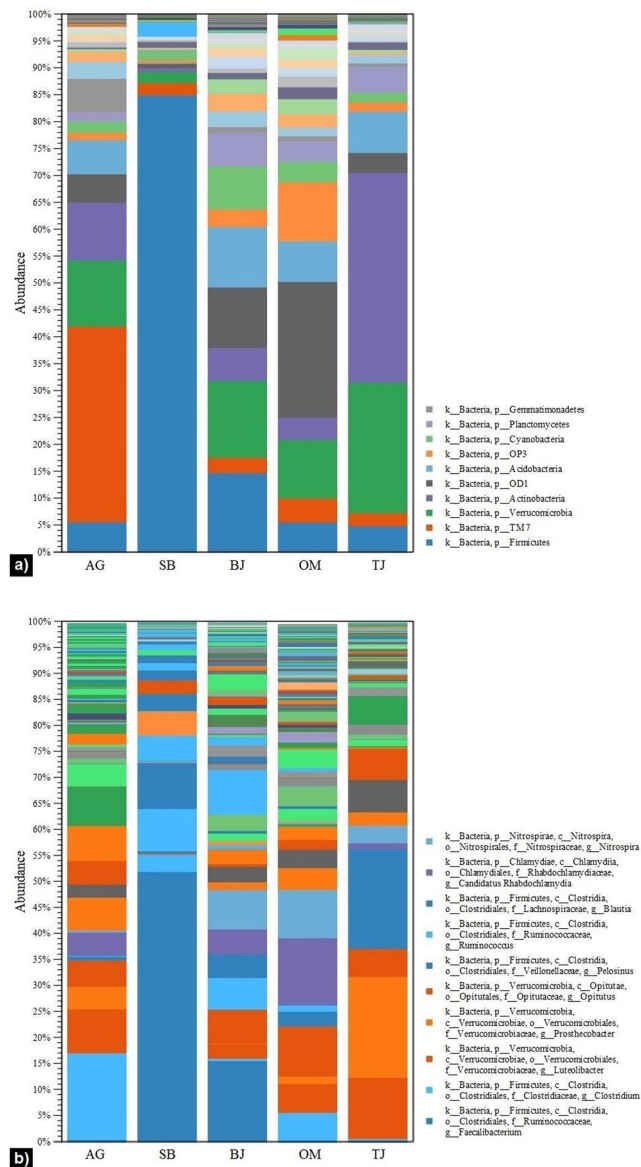


Figure 3. Relative abundance of bacterial populations among all studied sites (AG – Aas, Grytlandsbekken; SB – Ski, Blåveisbekken; BJ – Bergen, Jordalsvatnet; OM – Oslo, Maridalsvannet; and TJ – Trondheim, Jonsvannet): (a) bacterial phyla other than *Proteobacteria* and *Bacteroidetes*; (b) genera distribution of the other bacterial phyla. The stacked bar charts were generated using the tools in CLC Microbial Genomics Module version 2.5.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/clc-microbial-genomics-module-latest-improvements>).

gastrointestinal microbiome^{34,35}, *Faecalibacterium* is one of the most abundant and important commensal bacteria in the human gut microbiota³⁶. The importance of this observation is that it demonstrates a strong relationship and agreement with the results from faecal source tracking using the genetic markers in quantitative PCR, which showed a high dominance of human faecal contamination in the urban stream (Table 1). Very relevant to our outcomes are the findings by Unno *et al.*³⁷ reporting *Firmicutes* as one of the major human faecal bacteria found in watersheds. Similar discoveries were recently revealed in a study by Sun *et al.*²⁹, who detected high levels of *Firmicutes* in the Yangtze River and attributed them to the faecal bacteria carried in the effluent discharged from a nearby wastewater treatment plant.

With regards to the archaeal population, all sequences were assigned to three phyla: *Euryarchaeota*, *Crenarchaeota*, and *Parvarchaeota*, with predominance of the latter (Fig. 4a). The absolute dominance of *Parvarchaeota* (65–86% relative abundance) was revealed in the tributaries of drinking water reservoirs (BJ, OM, and TJ). *Crenarchaeota* composed the second largest phylum, especially in water from sites AG (44%) and SB (30%), dominated by two particular genera, *Candidatus Nitrososphaera* and *Nitrosopumilus*, respectively (Fig. 4b). The latter represents an active archaeon that oxidises ammonia to nitrite, while the former was firstly sequenced from soil and was further proven to be positively correlated with agricultural practices and

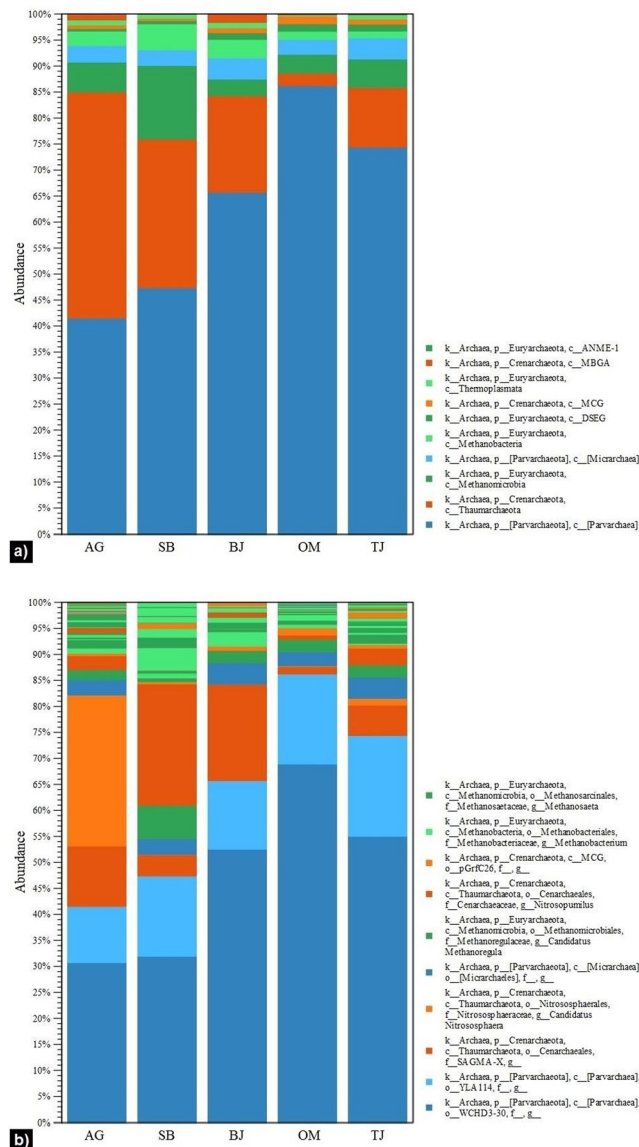


Figure 4. Relative abundance of archaeal populations among all studied sites (AG – Aas, Grytlandsbekken; SB – Ski, Blåveisbekken; BJ – Bergen, Jordalsvatnet; OM – Oslo, Maridalsvannet; and TJ – Trondheim, Jonsvannet): (a) phylum and class level; (b) genus level. The stacked bar charts were generated using the tools in CLC Microbial Genomics Module version 2.5.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/clc-microbial-genomics-module-latest-improvements>).

management such as fertilization with nitrogen compounds and farmyard manure³⁸. This greatly links to the type of our study site, the rural creek (AG) in an agriculturally dominated catchment, which consists largely of farmlands (60%) and forest/marshlands (31%). Furthermore, the seasonal quantitative microbial source tracking (QMST) investigations conducted at this study site exposed agricultural run-off to be the major zoogenic source contributing to faecal water pollution¹⁷.

The alpha diversity depicted in the plot as a function of rarefaction estimation (Fig. 5) indicates that the diversity and richness profiles were relatively variable from site to site. As a result, the complete alpha diversity had the following order: AG > OM > TJ > BJ > SB. The highest diversity was found in the least polluted rural creek (site AG with the dominant zoogenic faecal contamination in water), while the lowest was exposed in the urban stream (site SB), which suffered from an extreme anthropogenic faecal contamination (Table 1). This indicates that faecal pollution has a negative impact on the aquatic microbial community structure. It is presumably attributed to the fact that faecal input constitutes an extra selective pressure on aquatic microbial community, in a way to selectively promote the enrichment of some favourable microbes and thus result in the suppression of other microbial populations. This responds to the fact that any discharge of faecal matter contributes to the increased load of chemical, biological, and organic contaminants in water environments. These further create various stress factors on the basic ecological processes of ecosystems, such as disturbance in water and energy cycle, nutrient flow and availability, and biodegradation. In response to such stresses, the microbial community will be expectedly

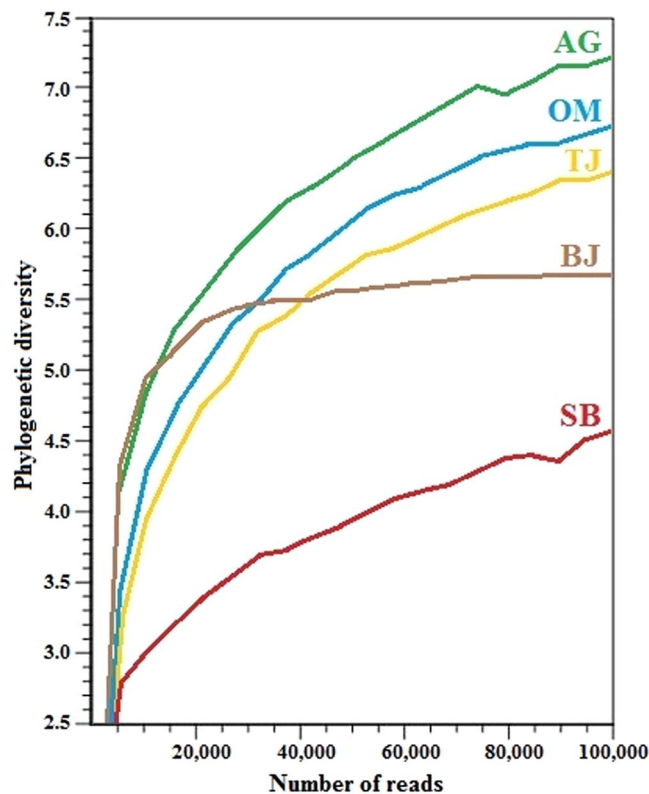


Figure 5. Phylogenetic alpha diversity of all studied sites (AG – Aas, Grytlandsbekken; SB – Ski, Blåveisbekken; BJ – Bergen, Jordalsvatnet; OM – Oslo, Maridalsvannet; and TJ – Trondheim, Jonsvannet). The alpha rarefaction plot was generated using the tools in CLC Microbial Genomics Module version 2.5.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/clc-microbial-genomics-module-latest-improvements>).

re-shaped. Such changes have been reported in association with nutrient and organic matter variabilities^{39,40} as well as variations in temperature, oxygen, salinity, and acidification^{26,41}. Significant impacts of combined various organic and inorganic contaminants on the microbial community structure in aquatic ecosystems have also been observed^{42,43}.

The beta diversity heat map (Fig. 6) visualises the hierarchical clustering of the sites based on their similarities of microbial compositions. PERMANOVA analysis reveals that the different clusters based on the groups were statistically significant (p -value = 0.00097). In this hierarchy, the two most-related clusters were disclosed, AG–TJ and BJ–OM. These were further merged to compose a separate cluster with the single site SB. This analysis clearly reveals that microbiota in the urban stream (SB) exposed to extremely high anthropogenic faecal pollution was greatly deviated from that found at the other study sites with the dominant zoogenic faecal contamination in water. Furthermore, the LEfSe algorithm (Fig. 7) carried out on the characterised microbial phyla in the divergent types of watercourses (rural/zoogenic – AG and urban/anthropogenic – SB) determined two distinct groups of microbes, which were apparently driving the microbial discrepancy between these two lotic ecosystems. As many as 184 distinct genera were considered to be differentially altered between the zoogenically and anthropogenically polluted waters. Moreover, the Wilcoxon rank-sum test reveals that the differences in microbial diversity between these two ecosystems were extremely significant (p -value < 0.0001). Typical human gut microbiota, mostly *Proteobacteria* (e.g. *E. coli*), *Firmicutes* (e.g. *E. faecalis* and *E. faecium*) and *Bacteroidetes* (e.g. *Bacteroidales*)^{34,44}, was found more dominantly in the urban stream (site SB with the dominant anthropogenic faecal contamination in water), while in the rural creek (site AG exposed to pollution from zoogenic sources), more diverse microbes co-existed (Fig. 7), and many of them were often detected in natural waters, for instance, *Verrucomicrobia*, *SR1*, *OD1*, *OP11*, *OP3*, and *Planctomycetes*^{45,46}.

Intriguingly, the statistical tests strengthen the analytical investigations by demonstrating practical examples of how different molecular methods could be used to interpret results, and thus validate the final conclusions. MST assays based on host-specific genetic markers have been widely implemented as a rapid and robust molecular method to identify the faecal origins in water. The assays are normally designed to target a specific group of faecal bacteria, for instance *Bacteroides*⁴⁷, *Bifidobacterium*⁴⁸, and *Faecalibacterium*⁴⁹, which are genetically discriminated between the habitat hosts. Based on the generated quantitative data from QMST, a pollution-contributing profile can be further achieved⁵⁰. In comparison, high-throughput sequence data acquired by NGS, the MiSeq technique in the present study, provide a comprehensive picture of the entire microbial community in water, and thus vary radically from the QMST by presenting total aquatic microbiota vs. target faecal species.

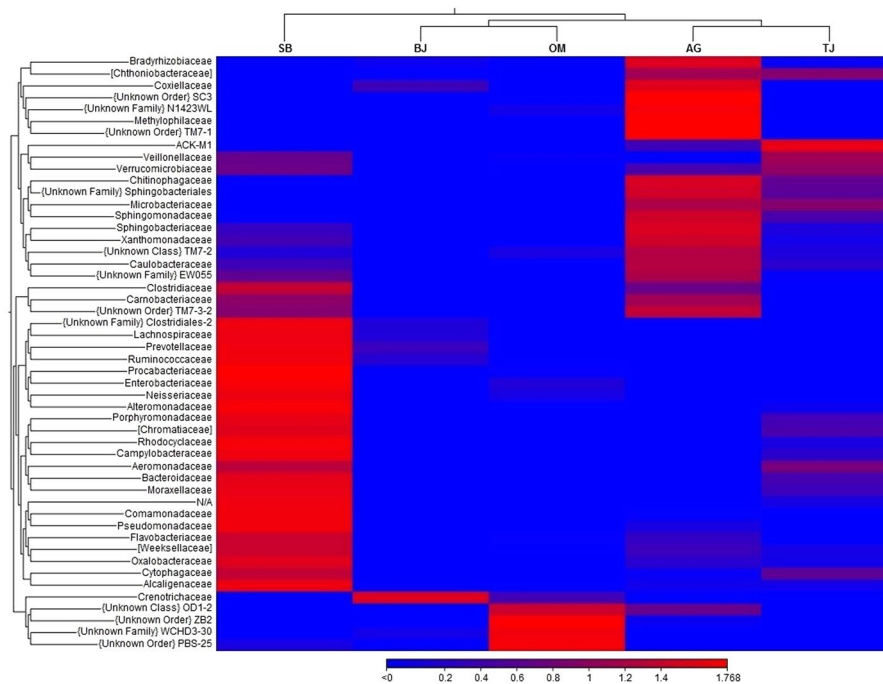


Figure 6. Beta diversity heat map with hierarchical clustering of the study sites (AG – Aas, Grytelandsbekken; SB – Ski, Blåveisbekken; BJ – Bergen, Jordalsvatnet; OM – Oslo, Maridalsvannet; and TJ – Trondheim, Jonsvannet). The hierarchical clustering heat map was generated using the tools in CLC Microbial Genomics Module version 2.5.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/clc-microbial-genomics-module-latest-improvements>).

Concluding Remarks

The joint application of QMST and NGS technologies in water faecal pollution research presents relatively comprehensive and mutually supportive approaches. The sequencing data provide extremely valuable input to better interpret the microbial community structure in the studied aquatic ecosystems, in particular, enabling closer monitoring of the alteration in microbial composition under various stresses, e.g. faecal pollution pressures. Findings derived from these technologies applied in our studies clearly reveal (analytically and statistically) that faecal pollution has a significant impact on the overall aquatic microbial community. The diversity of microbiota was substantially reduced in severely polluted water. Furthermore, the community compositions and richness profiles diverged greatly between lotic ecosystems, where the dominant faecal contamination sources were of anthropogenic or zoogenic origin. Consequently, water with high anthropogenic faecal pollution demonstrated low diversity of microbiota, while more distinct aquatic microbial profile characterised ecosystems exposed to zoogenic faecal sources.

Materials and Methods

Data screening for sampling site selection. To select the most distinctive sites characterising the microbial taxonomic diversity of faecally polluted lotic ecosystems in Norway, we focused on choosing those that accomplished most of the following criteria: (1) continuously exposed to varying pressures of faecal contamination, but not from direct input sources (e.g. direct discharges of sewage or effluent from wastewater treatment plant); (2) characterise sensitive spots (e.g., drinking water sources); (3) represent different geographical regions of the country; and (4) include dissimilar types of watercourses. For these purposes, we conducted site screening based on the original data from our recent studies assessing the rates of faecal pollution in tributaries of drinking water reservoirs supplying water to Norway's three largest cities: Oslo (680 000 citizens), Bergen (280 000 citizens), and Trondheim (190 000 citizens) located in different regions of the country (Eastern, Western, and Central Norway; Fig. 1). In addition, we utilised data of microbial quality measures carried out in two different types of watercourses (an urban stream partially culverted and a rural creek) located in the neighbouring municipalities, Ski (30 000 people) and Aas (20 000 people), located 20 and 30 km south-east of Oslo, respectively (Fig. 1).

Consequently, all the analysed data were derived from our latest national projects conducted in cooperation with local authorities/municipalities of Oslo, Bergen, and Trondheim, and partly from an international project, *AQUARIUS* (<https://eeagrants.org/project-portal/project/CZ09-0013>). Based on the preliminary screening of the numerous data, we selected the most representative sampling sites exposed to the constant pressure of faecal pollution originating from various anthropogenic and zoogenic sources. These were confirmed by standard *Escherichia coli* (*E. coli*) examinations and quantitative microbial source tracking (QMST) using host-associated genetic markers, i.e. AllBac (universal marker), BacH (human), BacR (ruminant) and Hor-Bac (horse). Detailed results of these examinations are available through open access reports published online at <http://hdl.handle>.

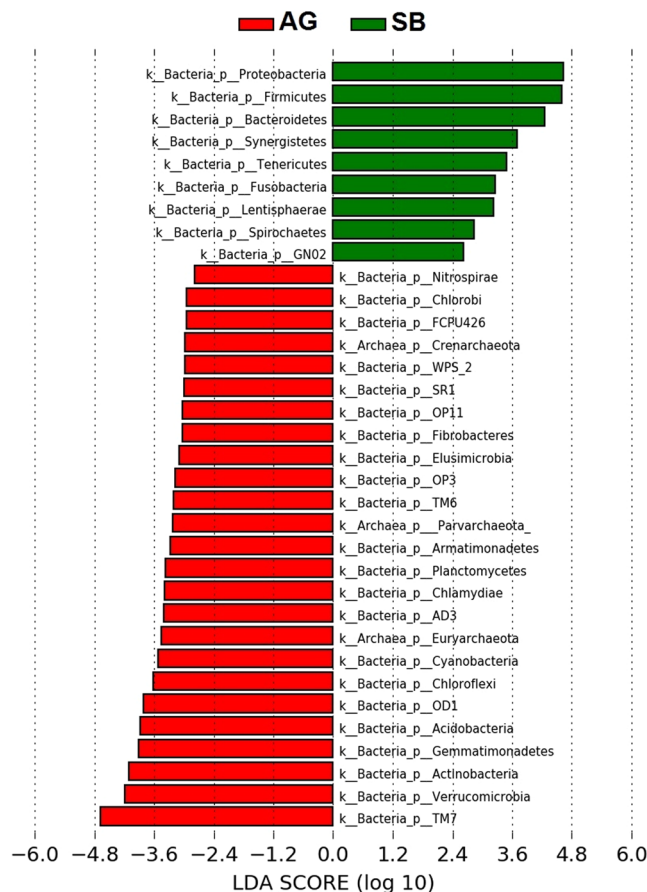


Figure 7. Linear discriminant analysis Effect Size (LEfSe) rank plot of differentially abundant microbial taxa in the least and most faecally polluted sites (AG – Aas, Grytlandsbekken and SB – Ski, Blåveisbekken, respectively) expressed as Linear Discriminant Analysis (LDA) logarithmic scores. The LEfSe plot was generated using the tools in Galaxy Version 1.0 (the Huttenhower Lab, Harvard School of Public Health, Boston, MA, USA, <http://huttenhower.sph.harvard.edu/galaxy>).

<http://hdl.handle.net/11250/2392445>, <http://hdl.handle.net/11250/2392629>, and <http://hdl.handle.net/11250/2392631>. Scientific backgrounds and methodological procedures have been described in greater detail elsewhere⁵⁰.

In total, 60 water samples collected over the course of a 1-year period were examined (Table 1). The samples reflect an assortment of the site screening criteria, and thus represent lotic ecosystems with two extreme levels of faecal water pollution: the least contaminated site – a rural creek in Aas, named Grytlandsbekken (AG), and the most contaminated site – an urban stream in Ski, named Blåveisbekken (SB); as well as three sensitive sites located in different regions of Norway, i.e., tributaries of drinking water reservoirs for Oslo, Bergen, and Trondheim, named Maridalsvannet (OM), Jordalsvatnet (BJ), and Jonsvannet (TJ), respectively (Fig. 1). No specific background data on the sampled sites were collected, except the most divergent sites, i.e. AG (the least faecally polluted rural creek with the dominant zoogenic contamination in water) and SB (the site exposed to extremely high anthropogenic faecal pollution). In these two lotic ecosystems, physico-chemical parameters were measured at each (the same) sampling occasion (over the course of a 1-year period) and the data are presented in Table 2. Neither temporal and spatial features nor climatic conditions differentiated these two sites (Fig. 1), and therefore other external influences than the defined faecal water contamination were not substantial. For that reason, site AG and SB were deliberately selected for exploration of zoogenic vs. anthropogenic influences as well as the highest vs. lowest faecal contamination effects on the lotic microbial diversity, which is the new and compelling aspect of our research.

Faecal contamination, source tracking, and origin profiling. Detection of faecal water contamination and definition of its origins, namely anthropogenic and zoogenic, was performed according to the methodological toolbox implemented and verified in Norway⁵⁰. The completed procedure consists of three independent subsequent steps: (1) faecal pollution detection – examination of *E. coli* using the Colilert 18/Quanti-Tray[®]2000 method (IDEXX Laboratories Incorporated, Westbrook, Maine, USA) and bacteria enumeration as the most probable number (MPN)/100 mL; (2) faecal source tracking – detection and quantitation of host-specific *Bacteroidales* 16S rRNA genetic markers through real-time quantitative polymerase chain reaction (RT-qPCR); and (3) faecal origin profiling – characterisation of faecal source contribution in the actual sample tested. The key focus is given to the confirmed faecal contamination (step 1); thus, only samples positive for *E. coli* were

Parameter	Values min – max mean	
	AG	SB
Chemical oxygen demand (COD)	$\frac{9-38}{24}$	$\frac{6-129}{28}$
Total suspended solids (TSS)	$\frac{5.6-92.3}{17.7}$	$\frac{6.1-1350}{168}$
Total dissolved solids (TDS)	$\frac{107-225}{150}$	$\frac{120-316}{177}$
Total organic carbon (TOC)	$\frac{3.1-12.4}{7.6}$	$\frac{3-9}{5.9}$
Dissolved organic carbon (DOC)	$\frac{3-12}{7}$	$\frac{3-8.4}{5.6}$
Total phosphorus (TP)	$\frac{0.017-0.228}{0.083}$	$\frac{0.028-0.71}{0.114}$
Phosphate phosphorus (PO ₄ -P)	$\frac{0.013-0.094}{0.032}$	$\frac{0.004-0.112}{0.036}$
Total nitrogen (TN)	$\frac{2.56-9.66}{5.08}$	$\frac{1.5-6.36}{3.13}$
Ammonium nitrogen (NH ₄ -N)	$\frac{0.004-0.899}{0.123}$	$\frac{0.026-2.17}{0.307}$
Nitrite nitrogen (NO ₂ -N)	$\frac{0.008-0.068}{0.03}$	$\frac{0.005-0.183}{0.041}$
Nitrate nitrogen (NO ₃ -N)	$\frac{1.84-4.37}{2.72}$	$\frac{1.21-2.02}{1.54}$
Electrical conductivity (EC)	$\frac{11.8-31.6}{19.6}$	$\frac{18-40.7}{27.1}$
pH	6.9–7.7	7.4–8.4

Table 2. Physico-chemical parameters measured in the least faecally contaminated site (AG – Aas, Grytelandsbekken rural creek) and the most faecally contaminated site (SB – Ski, Blåveisbekken urban stream). Values in mg/l, and mS/m for EC.

processed in further molecular investigations (step 2). Since there are no significant correlations between *E. coli* bacteria and the *Bacteroidales* DNA markers, the faecal origin profile (step 3) was assessed as a percentage contribution of the markers defined in the contaminated water. The scientific background and techniques of this toolbox have been described in greater detail elsewhere⁵⁰.

DNA extraction, PCR, and MiSeq sequencing. From each faecally polluted sample, 200 mL water was concentrated in an ultrafiltration unit. Microbial genomic DNA was extracted from each yielded filter using a PowerSoil[®] DNA Isolation Kit (MO BIO Laboratories, Carlsbad, California, USA), following the protocol applied in earlier studies⁵⁰. Purified DNA materials from the same sampling site were pooled together in equal amount and proceeded to library preparation of 16S rRNA amplicons. Such DNA pooling strategy has been revised and implemented in other published studies covering various ecosystems (soil and water) and different macro-/microorganism communities^{33,51–53}. Five nanograms of each pooled genomic DNA was used in PCR to generate an amplicon library for each sample, according to the procedure instructions of the NEXTflex[™] 16S V4 Amplicon-Seq Kit 2.0 (Bioo Scientific Corporation, Austin, TX, USA). Briefly, the method consists of two-step PCR procedures. In the first PCR, the 16S rRNA V4 region was amplified using V4-specific primer sets; following PCR purification using AMPure XP Magnetic Beads (Agencourt Bioscience Corporation, Beverly, MA, USA), the clean-up product was subjected to the second PCR, where an Illumina adaptor overhang and unique 12-bp index were attached to the final product. Following purification using AMPure beads, the DNA concentration was measured by a Qubit[™] Fluorometer (Life Technologies, Eugene, OR, USA) using a Quant-IT[™] dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA). Amplicon products were normalised and equimolar pooled. The pooled library sample was analysed by 1% agarose gel electrophoresis and subsequently sequenced by an Illumina MiSeq sequencer at the Norwegian Sequencing Centre (Oslo, Norway) using a MiSeq Reagent Kit V3 (Illumina Inc., San Diego, CA, USA).

Sequencing data analyses. The sequencing reads were analysed using the Microbial Genomics Module 2.0 added onto the CLC Genomic Workbench 10.1.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark <http://www.qiagenbioinformatics.com/products/clc-genomics-workbench>). The entire analysis workflow was composed of a number of sequential and interconnected steps, consisting of quality filtration of the sequence data, operational taxonomic unit (OTU) clustering, and alpha and beta diversity studies basically using default parameter settings as configured in the workflow of the Microbial Genomics Module. Sequence trimming involved adapter trimming, quality trimming, and length trimming. The primer sequence was removed and the reads with a quality score lower than 20 were discarded. The maximum accepted number of ambiguous nucleotides was 2, and the length of the reads was fixed at 200–500 bp. Furthermore, chimeric sequences including singletons were detected and trashed. The remaining qualified unique reads were used for OTU clustering, which was performed by alignment to the Greengenes Database (Greengenes v_13_5) at 97% sequence similarity. A phylogenetic tree was subsequently constructed using a maximum likelihood approach based on a multiple sequence alignment (MSA)

of the OTU sequences by the MUSCLE tool. This was further applied for alpha and beta diversity measures. The former, expressed as species richness in an ecosystem, was estimated from the rarefaction analysis using the resulting phylogenetic tree with a maximum sampling depth to 100,000 reads. The latter, which examines changes in species diversity between ecosystems, was measured using the Euclidean distance criterion. All sequence data were deposited at the NCBI Sequence Read Archive under accession number SRP133171, as part of BioProject PRJNA434659.

Statistical analyses. A beta diversity hierarchical clustering heat map was created to visualise the relatedness of the aquatic microbial ecosystems by the similarity of their communities. It was generated through the abundance of features in the tested samples and expressed as a binary tree clustering the study sites based on the Euclidean distance at the phylum level using the Trimmed Mean of M-values (TMM) and the Z-score (the number of standard deviations from the population mean) normalisations. In addition, PERMANOVA analysis was carried out to ascertain statistical significance of the clusters. These tests were executed by applying CLC Microbial Genomics Module version 2.5.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/clc-microbial-genomics-module-latest-improvements>). Moreover, a differential abundance analysis was performed using the Linear discriminant analysis Effect Size (LEfSe) algorithm⁵⁴ to unravel/disclose the driving microbial groups, which were mostly responsible for the difference in the least and most faecally polluted ecosystems (AG and SB, respectively). The formatted OTU abundance table was uploaded to the Galaxy/Hutlab application web-based platform (Biostatistics Department, Harvard School of Public Health, Boston, MA, USA <http://huttenhower.sph.harvard.edu/galaxy>) for the pairwise comparison of the microbial taxa. The comparison was tested for statistical significance (the Wilcoxon rank-sum test) and the alpha value for the pairwise Wilcoxon test was set at 0.05. The identified features characterising the differences between the ecosystems were processed using Linear Discriminant Analysis (LDA) with a threshold score set at 2.0.

Artwork. Figure presenting the locations of the study sites (Fig. 1) is based on a satellite image and topographical map of Norway provided by the Norwegian Institute of Bioeconomy Research (NIBIO). The Institute's geographical data is recorded in the NIBIO's primary map service the Source/Kilden (<https://kilden.nibio.no>). All maps and map images collected in the Source/Kilden are open for all and can be saved or printed (<https://www.nibio.no/en/subjects/soil/national-land-resource-map?locationfilter=true>).

Figures containing results and statistics (Figs. 2–7) were generated using the tools in CLC Microbial Genomics Module version 2.5.1 (CLC Bio, QIAGEN Company, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/clc-microbial-genomics-module-latest-improvements>) and Galaxy Version 1.0 (the Huttenhower Lab, Harvard School of Public Health, Boston, MA, USA, <http://huttenhower.sph.harvard.edu/galaxy>).

Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Received: 5 June 2019; Accepted: 15 November 2019;

Published online: 19 December 2019

References

- Rickert, B., Chorus, I. & Schmoll, O. Protecting surface water for health. Identifying, assessing and managing drinking-water quality risks in surface-water catchments (World Health Organisation: Geneva, Switzerland, 2016).
- Pedley, S. *et al.* Pathogens: Health relevance, transport and attenuation. In: Protecting groundwater for health: Managing the quality of drinking-water sources (eds. Schmoll, O., Howard, G. & Chilton, J.) 49–80 (World Health Organisation, IWA Publishing, London, UK, 2006).
- WHO. Guidelines for drinking-water quality: fourth edition incorporating the first addendum (World Health Organization, Geneva, Switzerland, 2017).
- Paerl, H. W. *et al.* Microbial indicators of aquatic ecosystem change: current applications to eutrophication studies. *FEMS Microbiol. Ecol.* **46**(3), 233–246, [https://doi.org/10.1016/S0168-6496\(03\)00200-9](https://doi.org/10.1016/S0168-6496(03)00200-9) (2003).
- Saxena, G., Bharagava, R. N., Kaithwas, G. & Raj, A. Microbial indicators, pathogens and methods for their monitoring in water environment. *J. Water Health* **13**(2), 319–339, <https://doi.org/10.2166/wh.2014.275> (2015).
- Zeglin, L. H. Stream microbial diversity in response to environmental changes: review and synthesis of existing research. *Front. Microbiol.* **6**, 454, <https://doi.org/10.3389/fmicb.2015.00454> (2015).
- Farnleitner, A. H. *et al.* *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**(5), 1599–1608, <https://doi.org/10.1111/j.1365-2672.2010.04788.x> (2010).
- EU. Commission Directive (EU) 2015/1787 of 6 October 2015 amending Annexes II and III to Council Directive 98/83/EC on the quality of water intended for human consumption (The European Commission, Official Journal of the European Union, L260/6, 2015).
- EU. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC (Official Journal of the European Union, L64/37, 2006).
- USEPA. Recreational Water Quality Criteria. Office of Water 820-F-12-058 (United States Environmental Protection Agency, Washington, D.C., USA, 2012).
- WHO. Guidelines for the safe use of wastewater, excreta and greywater in agriculture and aquaculture, Vol. 1–4 (World Health Organization, Geneva, Switzerland, 2006).
- O'Mullan, G. D., Elias Dueker, M. & Juhl, A. R. Challenges to managing microbial fecal pollution in coastal environments: Extra-enteric ecology and microbial exchange among water, sediment, and air. *Curr. Pollution Rep.* **3**(1), 1–16, <https://doi.org/10.1007/s40726-016-0047-z> (2017).
- Naphtali, P., Mohiuddin, M. M., Paschos, A. & Schellhorn, H. E. Application of high-throughput 16S rRNA sequencing to identify fecal contamination sources and to complement the detection of fecal indicator bacteria in rural groundwater. *J. Water Health* **17**(3), 393–403, <https://doi.org/10.2166/wh.2019.295> (2019).

14. Hagedorn, C., Blanch, A. R. & Harwood, V. J. *Microbial Source Tracking: Methods, Applications, and Case Studies*. Springer, New York, NY, <https://doi.org/10.1007/978-1-4419-9386-1> (2011).
15. McQuaig, S., Griffith, J. & Harwood, V. J. Association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution. *Appl. Environ. Microbiol.* **78**(18), 6423–6432, <https://doi.org/10.1128/AEM.00024-12> (2012).
16. Harwood, V. J., Staley, C., Badgley, B. D., Borges, K. & Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* **38**(1), 1–40, <https://doi.org/10.1111/1574-6976.12031> (2014).
17. Paruch, L., Paruch, A. M., Blankenberg, A.-G. B., Haarstad, K. & Mæhlum, T. Norwegian study on microbial source tracking for water quality control and pollution removal in constructed wetland treating catchment run-off. *Water Sci. Technol.* **76**(5), 1158–1166, <https://doi.org/10.2166/wst.2017.303> (2017).
18. Reischer, G. H. *et al.* 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**(15), 8548–8556, <https://doi.org/10.1021/es304367t> (2013).
19. Sowah, R. A., Habteselassie, M. Y., Radcliffe, D. E., Bauske, E. & Risse, M. Isolating the impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia, United States. *Water Res.* **108**, 330–338, <https://doi.org/10.1016/j.watres.2016.11.007> (2017).
20. Staley, Z. R., Grabuski, J., Sverko, E. & Edge, T. A. Comparison of microbial and chemical source tracking markers to identify fecal contamination sources in Humber River (Toronto, Ontario, Canada) and associated storm water outfalls. *Appl. Environ. Microbiol.* **82**(21), 6357–6366, <https://doi.org/10.1128/AEM.01675-16> (2016).
21. Kaevska, M., Vidsenska, P., Sedlar, K. & Slana, I. Seasonal changes in microbial community composition in river water studied using 454-pyrosequencing. *SpringerPlus* **5**, 409, <https://doi.org/10.1186/s40064-016-2043-6> (2016).
22. Zinger, L., Gobet, A. & Pommier, T. Two decades of describing the unseen majority of aquatic microbial diversity. *Mol. Ecol.* **21**(8), 1878–1896, <https://doi.org/10.1111/j.1365-294X.2011.05362.x> (2012).
23. Parulekar, N. N. *et al.* Characterization of bacterial community associated with phytoplankton bloom in a eutrophic lake in South Norway using 16S rRNA gene amplicon sequence analysis. *PLoS One* **12**(3), <https://doi.org/10.1371/journal.pone.0173408> (2017).
24. Wang, P., Chen, B., Yuan, R., Li, C. & Li, Y. Characteristics of aquatic bacterial community and the influencing factors in an urban river. *Sci. Total Environ.* **569–570**, 382–389, <https://doi.org/10.1016/j.scitotenv.2016.06.130> (2016).
25. Calderón, O., Porter-Morgan, H., Jacob, J. & Elkins, W. Bacterial diversity impacts as a result of combined sewer overflow in a polluted waterway. *Glob. J. Environ. Sci. Manag.* **3**(4), 437–446, <https://doi.org/10.22034/gjesm.2017.03.04.009> (2017).
26. Ibekwe, A. M., Ma, J. & Murinda, S. E. Bacterial community composition and structure in an Urban River impacted by different pollutant sources. *Sci. Total Environ.* **566–567**, 1176–1185, <https://doi.org/10.1016/j.scitotenv.2016.05.168> (2016).
27. Kaushik, R., Balasubramanian, R. & Dunstan, H. Microbial quality and phylogenetic diversity of fresh rainwater and tropical freshwater reservoir. *PLoS One* **9**(6), <https://doi.org/10.1371/journal.pone.0100737> (2014).
28. Mlejnková, H. & Sovová, K. Impact of pollution and seasonal changes on microbial community structure in surface water. *Water Sci. Technol.* **61**(11), 2787–2795, <https://doi.org/10.2166/wst.2010.080> (2010).
29. Sun, H. *et al.* Diversity, abundance, and possible sources of fecal bacteria in the Yangtze River. *Appl. Microbiol. Biotechnol.* **101**(5), 2143–2152, <https://doi.org/10.1007/s00253-016-7998-2> (2017).
30. Drury, B., Rosi-Marshall, E. & Kelly, J. J. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Appl. Environ. Microbiol.* **79**(6), 1897–905, <https://doi.org/10.1128/AEM.03527-12> (2013).
31. Wang, L. *et al.* Shift in the microbial community composition of surface water and sediment along an urban river. *Sci. Total Environ.* **627**, 600–612, <https://doi.org/10.1016/j.scitotenv.2018.01.203> (2018).
32. Liu, Z., Huang, S., Sun, G., Xu, Z. & Xu, M. Phylogenetic diversity, composition and distribution of bacterioplankton community in the Dongjiang River, China. *FEMS Microbiol. Ecol.* **80**(1), 30–44, <https://doi.org/10.1111/j.1574-6941.2011.01268.x> (2012).
33. Pinto, A. J., Xi, C. & Raskin, L. Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environ. Sci. Technol.* **46**(16), 8851–8859, <https://doi.org/10.1021/es302042t> (2012).
34. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**(4), 789–799, <https://doi.org/10.1016/j.cell.2014.09.053> (2014).
35. Johnson, E. L., Heaver, S. L., Walters, W. A. & Ley, R. E. Microbiome and metabolic disease: revisiting the bacterial phylum Bacteroidetes. *J. Mol. Med.* **95**(1), <https://doi.org/10.1007/s00109-016-1492-2> (2017).
36. Miquel, S. *et al.* *Faecalibacterium prausnitzii* and human intestinal health. *Curr. Opin. Microbiol.* **16**(3), 255–261, <https://doi.org/10.1016/j.mib.2013.06.003> (2013).
37. Unno, T. *et al.* Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environ. Sci. Technol.* **44**(20), 7777–7782, <https://doi.org/10.1021/es101500z> (2010).
38. Zhalnina, K. *et al.* *Ca. Nitrososphaera* and *Bradyrhizobium* are inversely correlated and related to agricultural practices in long-term field experiments. *Front. Microbiol.* **4**, 104, <https://doi.org/10.3389/fmicb.2013.00104> (2013).
39. Mahmoudi, N. *et al.* Microbial community composition and diversity in Caspian Sea sediments. *FEMS Microbiol. Ecol.* **91**(1), 1–11, <https://doi.org/10.1093/femsec/fiu013> (2015).
40. Van Horn, D. J., Sinsabaugh, R. L., Takacs-Vesbach, C. D., Mitchell, K. R. & Dahm, C. N. Response of heterotrophic stream biofilm communities to a gradient of resources. *Aquat. Microb. Ecol.* **64**, 149–161, <https://doi.org/10.3354/ame01515> (2011).
41. Mesa, V. *et al.* Bacterial, archaeal, and eukaryotic diversity across distinct microhabitats in an acid mine drainage. *Front. Microbiol.* **8**, 1756, <https://doi.org/10.3389/fmicb.2017.01756> (2017).
42. Pratt, B., Riesen, R. & Johnston, C. G. PLFA analyses of microbial communities associated with PAH-contaminated riverbank sediment. *Microb. Ecol.* **64**(3), 680–691, <https://doi.org/10.1007/s00248-012-0060-8> (2012).
43. Yao, X.-F., Zhang, J.-M., Tian, L. & Guo, J.-H. The effect of heavy metal contamination on the bacterial community structure at Jiaozhou Bay, China. *Braz. J. Microbiol.* **48**(1), 71–78, <https://doi.org/10.1016/j.bjm.2016.09.007> (2017).
44. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food components. *Eur. J. Nutr.* **57**(1), 1–24, <https://doi.org/10.1007/s00394-017-1445-8> (2018).
45. Balmonte, J. P., Arnosti, C., Underwood, S., McKee, B. A. & Teske, A. Riverine Bacterial Communities Reveal Environmental Disturbance Signatures within the *Betaproteobacteria* and *Verrucomicrobia*. *Front. Microbiol.* **7**, 1441, <https://doi.org/10.3389/fmicb.2016.01441> (2016).
46. Hou, Z. *et al.* Geochemical and Microbial Community Attributes in Relation to Hyporheic Zone Geological Facies. *Sci. Rep.* **7**(1), 12006, <https://doi.org/10.1038/s41598-017-12275-w> (2017).
47. Kabiri, L., Alum, A., Rock, C., McLain, J. E. & Abbaszadegan, M. A tool box strategy using *Bacteroides* genetic markers to differentiate human from non-human sources of fecal contamination in natural water. *Sci. Total Environ.* **572**, 897–905, <https://doi.org/10.1016/j.scitotenv.2016.07.188> (2016).
48. Gómez-Doñate, M., Ballesté, E., Muniesa, M. & Blanch, A. R. New molecular quantitative PCR assay for detection of host-specific *Bifidobacteriaceae* suitable for microbial source tracking. *Appl. Environ. Microbiol.* **78**(16), 5788–5795, <https://doi.org/10.1128/AEM.00895-12> (2012).

49. Sun, D. *et al.* Application of *Faecalibacterium* 16S rDNA genetic marker for accurate identification of duck faeces. *Environ. Sci. Pollut. Res.* **23**(8), 7639–7647, <https://doi.org/10.1007/s11356-015-6024-z> (2016).
50. Paruch, L., Paruch, A. M., Blankenberg, A.-G. B., Bechmann, M. & Mæhlum, T. Application of host-specific genetic markers for microbial source tracking of faecal water contamination in an agricultural catchment. *Acta Agric. Scand.* **65**(S2), 164–172, <https://doi.org/10.1080/09064710.2014.941392> (2015).
51. Castle, S. C. *et al.* Impacts of sampling design on estimates of microbial community diversity and composition in agricultural soils. *Microb. Ecol.*, <https://doi.org/10.1007/s00248-019-01318-6> (2019).
52. Osborne, C. A., Zwart, A. B., Broadhurst, L. M., Young, A. G. & Richardson, A. E. The influence of sampling strategies and spatial variation on the detected soil bacterial communities under three different land-use types. *FEMS Microbiol. Ecol.* **78**(1), 70–9, <https://doi.org/10.1111/j.1574-6941.2011.01105.x> (2011).
53. Sato, H., Sogo, Y., Doi, H. & Yamanaka, H. Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. *Sci. Rep.* **7**(1), 14860, <https://doi.org/10.1038/s41598-017-14978-6> (2017).
54. Segata, N., *et al* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**(6), <https://doi.org/10.1186/gb-2011-12-6-r60> (2011).

Author contributions

L.P., A.M.P. and H.G.E. conceived and designed the studies. L.P. conducted molecular analyses, created database and performed bioinformatics. A.M.P. conducted microbial tests and faecal source tracking. H.G.E. and R.S. supervised samples verification and testing methods. L.P. and A.M.P. drafted text and created all figures with comments, suggestions and revisions from H.G.E. and R.S. The manuscript was written by L.P. and A.M.P. and was reviewed by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.M.P.

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

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



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

© The Author(s) 2019