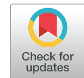
Science of the Total Environment 743 (2020) 140472

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Evaluation of groundwater bacterial community composition to inform waterborne pathogen vulnerability assessments



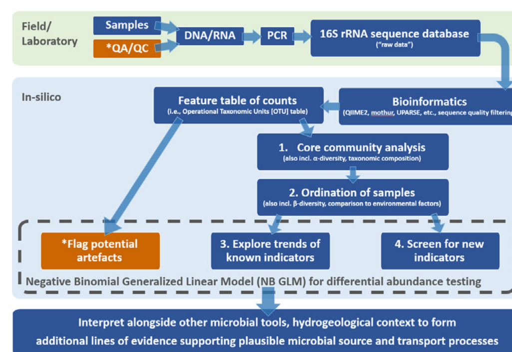
Alex H.S. Chik [a,b,c,*], Monica B. Emelko [c], William B. Anderson [c], Kaitlyn E. O'Sullivan [c], Domenico Savio [d,e], Andreas H. Farnleitner [d,e], Alfred Paul Blaschke [b], Jack F. Schijven [a]

[a] Utrecht University, Domplein 29, 3512 JE Utrecht, Netherlands
[b] TU Wien, Karlsplatz 13, 1040 Vienna, Austria
[c] University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada
[d] Karl Landsteiner University of Health Sciences, Dr.-Karl-Dorrek-Straße 30, 3500 Krems an der Donau, Austria
[e] TU Wien, Gumpendorfer Straße 1a, 1060 Vienna, Austria

HIGHLIGHTS

- Groundwater vulnerability to pathogen intrusion was informed by 16S rRNA sequencing.
- Observed sequences suggestive of fecal source not confirmed by culture-based methods
- Seasonality of *Betaproteobacteria* sequences may reflect surface connectivity.
- Bacterial taxa reflecting episodic water quality changes and adequate purging identified
- Sequences common to surface waters absent or present at low levels

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 February 2020
Received in revised form 5 June 2020
Accepted 22 June 2020
Available online 27 June 2020

Editor: Ewa Korzeniewska

Keywords:
16S rRNA gene amplicon sequencing
Bacterial community analysis
Groundwater surface water interactions
Microbial water quality
Pathogen vulnerability
Purge water sampling

ABSTRACT

Microbial water quality evaluations are essential for determining the vulnerability of subsurface drinking water sources to fecal pathogen intrusion. Rather than directly monitor waterborne pathogens using culture- or enumeration-based techniques, the potential of assessing bacterial community using 16S rRNA gene amplicon sequencing to support these evaluations was investigated. A framework for analyzing 16S rRNA gene amplicon sequencing results featuring negative-binomial generalized linear models is demonstrated, and applied to bacterial taxa sequences in purge water samples collected from a shallow, highly aerobic, unconfined aquifer. Bacterial taxa relevant as indicators of fecal source and surface connectivity were examined using this approach. Observed sequences of *Escherichia*, a genus suggestive of fecal source, were consistently detected but not confirmed by culture-based methods. On the other hand, episodic appearance of anaerobic taxa sequences in this highly aerobic environment, namely *Clostridia* and *Bacteroides*, warrants further investigation as potential indicators of fecal contamination. *Betaproteobacteria* sequences varied significantly on a seasonal basis, and therefore may be linked to understanding surface-water groundwater interactions at this site. However, sequences that are often encountered in surface water bodies (*Cyanobacteria* and *Flavobacteriia*) were notably absent or present at very low levels, suggesting that microbial transport from surface-derived sources may be rather limited. This work demonstrates the utility of 16S rRNA gene amplicon sequencing for contextualizing and complementing conventional microbial techniques, allowing for hypotheses about source and transport processes to be tested and refined.
© 2020 Published by Elsevier B.V.

* Corresponding author at: University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada
E-mail addresses: alex.chik@uwaterloo.ca (A.H.S. Chik), domenico.savio@kl.ac.at (D. Savio), andreas.farnleitner@kl.ac.at (A.H. Farnleitner).

https://doi.org/10.1016/j.scitotenv.2020.140472
0048-9697/© 2020 Published by Elsevier B.V.

1. Introduction

Updated terms of reference in the province of Ontario, Canada for treatment requirements of drinking water derived from Groundwater Under the-Direct-Influence of Surface Water (GUDI/GWUDI) sources will be promulgated imminently. They focus on establishing lines of evidence in support of human/mammalian fecal pathogen contamination or microbial transport from the surface (Ahmed et al., 2013). Although the updated terms of reference will rely heavily on conventional culture- and enumeration-based microbial methods, the potential for emerging methods to support establishing additional lines of evidence has been acknowledged. Due to inherent challenges of microbial concentration estimation through groundwater sampling, non-detects are not uncommon despite their actual presence in the source (Chik et al., 2018). When pathogens or fecal indicators are not observed, additional lines of evidence can be sought to gain a better conceptual understanding of microbial source and transport at a given site.

16S rRNA gene amplicon sequencing offers a broad coverage of microbial water quality, and its costs have been reduced substantially over time. Therefore, it is an attractive tool for complementing conventional microbial methods in assessing the vulnerability of a subsurface drinking water source to fecal pathogen contamination (Griebler and Lueders, 2009; Vierheilig et al., 2015; Savio et al., 2018). The highly conserved regions (e.g., V3-V4 regions) of the 16S rRNA gene are present in every bacterium and archaeon, but are adequately unique to enable their detection and differentiation to the genus level of classification (Yarza et al., 2014). While this technique does not inform whether these sequences originate from viable or culturable microorganisms, it provides a rapid means of characterizing bacterial community composition and has become a staple in many ecological microbiome studies (Vierheilig et al., 2015; Savio et al., 2018).

Microbial communities within subsurface environments suitable for potable water use have been increasingly characterized using 16S rRNA gene sequencing and similar bio-molecular techniques in the past decades (e.g., Griebler and Lueders, 2009; Korbel et al., 2017; Savio et al., 2019). In these contexts, indicator microbial taxa are often sought to formulate hypotheses about relevant processes underpinning microbial water quality, especially those that may indicate source water vulnerability to fecal pathogen intrusion. However, most microbiome studies to date have relied upon proportions (i.e., relative abundance) and rarefied counts (i.e., re-sampling of sequences reads from samples with adequate library sequencing depth) for the identification of indicator taxa (McMurdie and Holmes, 2014). These approaches have been demonstrated to bear a strong risk of bias towards false positive indications of taxa exhibiting differential abundance (McMurdie and Holmes, 2014). Approaches based on relative taxa sequence abundance fail to account for differences in library sequencing depth and amplification biases; rarefied sequence reads can also lead to a loss of statistical power (McMurdie and Holmes, 2014). While statistically rigorous approaches for comparing taxa sequence read abundance across samples have been proposed (McMurdie and Holmes, 2014) and developed (e.g., Robinson et al., 2010; Love et al., 2014), they have not been applied systematically to identify relevant indicator microbial taxa for supporting source water vulnerability assessments to pathogen intrusion. Critically, identification of such indicators requires consideration of variability attributable to the sample collection protocol and other key spatial and temporal gradients spanned by the samples collected.

A groundwater bacterial community was monitored over nine sampling campaigns spanning a year to demonstrate a progression of steps for identifying relevant bacterial indicators. This community was characterized using a focused investigation of two wells installed less than 5 m apart in the same shallow, unconfined aquifer. The bacterial core community (i.e., taxa that are abundant and prevalent across the samples collected) was inspected for obvious trends prior to evaluating sample (dis)similarity based on taxa sequences and environmental parameters concurrently measured. In the absence of obvious trends exhibited by the bacterial core community at high taxonomic levels, differential sequence read abundance testing was performed to examine the behaviour of known bacterial indicators and screen for relevant indicator taxa across spatio-temporal factors. This demonstration provides proof-of-concept for the use, interpretations, and limitations of 16S rRNA gene sequencing for supporting vulnerability evaluations of a subsurface drinking water source to fecal pathogen contamination.

2. Materials and methods

2.1. Site location

The Woodstock groundwater study area is situated near drinking water production wells six kilometers south of Woodstock, Ontario, Canada (Chik et al., 2020). The site comprises of rolling hills and drumlin features. A series of monitoring wells were installed in a shallow unconfined aquifer comprising of mainly sand and gravels interspersed with discontinuous silty till aquitard units (Haslauer, 2005).

A focused investigation of two wells situated 5 m apart in a direction perpendicular to regional groundwater flow was conducted. Both monitoring wells (50.8 millimeter [mm] diameter, polyvinyl chloride [PVC] construction) were screened approximately 5.91 m to 16.57 m below ground surface (mBGS) and vented to the atmosphere. Additional well installation details are provided in Koch (2009) and Critchley (2010). The groundwater table was consistently above the top of the screened interval throughout the duration of this study.

The aquifer in which the wells are developed is the shallowest of the regional aquifer system and features extremely high groundwater velocities (aquifer hydraulic conductivity in the well screened interval estimated to be 4.8×10^{-4} to 1.9×10^{-2} m/s (Devlin et al., 2012; Critchley, 2010)) and highly aerobic conditions (mean dissolved oxygen during study period = 7.9 mg/L). The highly aerobic conditions are suggestive of a relatively high degree of groundwater-surface connectivity. Regional groundwater flow is generally in a southeasterly direction. Annual recharge at this location is estimated to be 396 mm/year (Koch, 2009). Local infiltration travel time through the unsaturated zone to the water table is estimated to be on the order of 2.8 to 5.6 years (Sousa, 2013).

2.2. Experimental design

Given that the atmosphere is the primary source of dissolved oxygen in groundwater (Rose and Long, 1988), the extremely aerobic nature of this unconfined aquifer suggests a high degree of groundwater-surface connectivity. Accordingly, bacterial community composition was expected to exhibit seasonality. However, spatial heterogeneity and the extent of purging performed prior to sample collection are also known to confound bacterial community characterization. Accordingly, three primary factors that contribute to the variability of the 16S rRNA gene amplicon sequences (herein referred to as 'sequence reads') were investigated: the season during which sampling was performed (three levels), the well sampled (two levels), and the purging extent (three levels). Seasonality was represented by three sampling campaigns per season as guided by surficial climatic conditions and groundwater level measurements (Supplementary Information S1). At this site, the respective periods are: summer ("baseline" period characterized by relatively stable water table conditions, typically June–September), winter (transitional period typically exhibiting the lowest but variable annual water table levels due to occasional surface melt events, typically October–January), and spring freshet (period exhibiting highest annual water table levels due to snowmelt-dominated recharge, typically February–May) (Pasha, 2018). Two wells (WO77, WO78) within the same hydrogeological formation were sampled to characterize spatial heterogeneity attributable to the individual wells. WO77 was sampled for only one event during the summer and phased in such that both wells were consistently monitored in all sampling campaigns for the

following seasons. As groundwater samples are expected to be increasingly representative of the suspended microorganisms in the aquifer pore water without artefacts from well-related biofilms with extended purging (Cullimore, 2007; Chik et al., 2020), a minimum of two groundwater samples were collected in each well during early (<20 min), intermediate (>20 min, <1 h) and late stages of purging (>1 h, <4 h) that loosely correspond with major zones of interrogation projections as described by Cullimore (2007). These projections delineate zones for describing the influence of subsurface biofilms in and surrounding wells on abstracted water quality.

A total of 83 samples were collected. All factors were handled as fixed effects as levels associated with each factor are not random; inferences extrapolated based on these data would not be very precise. A greater number of years of sampling from additional wells would allow for these factors to be treated as random effects and broader extrapolation of this work. However, this level of characterization was beyond the scope of this proof-of-concept demonstration and does not contribute to supporting the key objectives of this work.

2.3. Sample collection

Samples were collected in general accordance with purge-water sampling protocols described by Cullimore (2007). Briefly, wells during each sampling campaign were pumped from quiescence using dedicated pumps. The pumps (Flojet model No. 4105 Series diaphragm pump, Irvine, California, USA) and attached PVC tubing were pre-sterilized in the laboratory using a dilute bleach solution (0.6% sodium hypochlorite) followed by a sterile deionized water rinse. Sterile deionized water was also used on site to prime the pumps as necessary. Groundwater samples were collected in sterile Wheaton™ bottles throughout well purging. Physical and select chemical water quality parameters (turbidity, pH, dissolved oxygen, electrical conductivity, and temperature) were monitored using a portable turbidimeter (Hach®, Colorado, USA) and a portable multiparameter meter (YSI Quattro Professional Plus, YSI Inc./Xylem Inc., Ohio, USA). These parameters generally did not fluctuate substantially after the removal of 3–5 well purge volumes (approximately 10 to 15 min of purging) (Chik et al., 2020). All probes and meters were calibrated using standards as per the manufacturers' instructions. Additional chemical water quality parameters, including one biochemical parameter (i.e., adenosine triphosphate [ATP]), were also monitored in a related study (Chik et al., 2020). Summary statistics are presented in Supplementary Information S2.

Two fecal indicators were enumerated in 100 mL aliquots of the groundwater samples. *Escherichia coli* was enumerated by membrane filtration (Standard Method 9222) using Trypton-Bile-X-Glucuronide (TBX) medium (Oxoid, Hampshire, UK) and incubation at 44±0.5°C for 44±4 h, and male-specific (F+) coliphage was enumerated in accordance with US EPA Method 1601 (US EPA, 2001). All fecal indicator samples yielded non-detects. Additional parameters relevant for microbial water quality were also evaluated in selected samples. Microbial cell densities estimated using the FACSCalibur™ flow cytometer (BD Biosciences, New Jersey, USA) yielded results below reliable quantification limits of the method (<1000 particles/mL) (Chik et al., 2020). A qualitative culture-based tool, Biological Activity Reaction Tests (ALGE-BART™, Droycon Bioconcepts Inc., Regina, Canada), was also deployed during selected sampling campaigns to test for the presence of grass-green algae, blue-green algae, desmids, diatoms and euglenoids. These tests consistently yielded non-detects.

2.4. 16S rRNA gene amplicon sequencing

750 mL aliquots of the groundwater samples were filtered through EMD Millipore Sterivex polyethersulfone 0.22 μm syringe filters (Millipore, Massachusetts, USA), stored in dedicated Whirl-Pak™ (Nasco©, Fort Atkinson, Wisconsin, USA) and kept in a −80°C freezer until DNA extraction. A commercial kit (DNeasy PowerWater Sterivex Kit, Qiagen, MO BIO Laboratories, USA) was then used to extract the DNA (e.g., Fiedler et al., 2018). 16S rRNA gene fragments were amplified and labelled with a unique, sample-specific multiplex-identifier ("barcode") in a PCR-based one step barcoding procedure. Pro341F-Pro805R (5′-CCT ACG GGN BGC ASC AG-3′,5′-GAC TAC NVG GGT ATC TAA TCC-3′) primers were used to target 16S rRNA V3-V4 regions (Takahashi et al., 2014). Briefly, PCR was set in triplicate for each sample (25 microlitres (μL)). The reaction mixture contained 2.5 μL of 10× standard Taq buffer, 0.5 μL of 10 millimolar (mM) dNTP, 1.0 μL of bovine serum albumin (20 mg/mL), 5.0 μL of 1 micromolar (μM) forward primer, 5.0 μL of 1 μM reverse primer, 1.0 μL DNA, 0.125 μL of Taq DNA polymerase (5u/μL) and 9.875 μL of PCR water, DNA was denatured at 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, 50°C for 30 s and 72°C for 50 s and followed by a final extension step at 72°C for 10 min. The triplicate PCR products were pooled. Two μL of PCR amplicons were loaded onto a 2% TAE agarose gel to verify the amount and size of PCR products. PCR products with equivalent amounts of correct amplicons were pooled, gel purified, and quantified using the Qubit dsDNA HS assay kit. Library DNA was sequenced with MiSeq Reagent Kit v2 (2 × 250 cycles); raw sequencing data have been deposited in the NCBI SRA database (accession number PRJNA625549). A laboratory blank comprising of sterile DI water was performed for every 30 samples as negative controls. The microbiome data analyses described below were first conducted with these quality control samples included to evaluate the extent of the reagent contamination. While these results are not central to the goals of the present investigation, they are integral to the broader interpretation of results emanating from the field purge water samples. Accordingly, results of the statistical analyses performed to identify bacterial taxa that require judicious interpretation are summarized and presented in Supplementary Information S3.

2.5. Microbiome data analysis

Taxa have routinely been described as Operational Taxonomic Units (OTUs) generated from the *de-novo* clustering of sequences into bins using a threshold of 97% sequence similarity. Clustering-independent approaches for determining amplicon sequence variants (ASVs, also known as sub-OTUs (Knight et al., 2018)) have been more recently advocated to replace OTUs given their ability to distinguish subtle but real biological sequence variants (Callahan et al., 2017). While ASVs undoubtedly contribute to increased differentiation of taxa at the species and sub-species levels (which may be fundamentally important in cladistics-based studies), appreciable differences in bacterial community composition related to changing source water quality conditions appear to be captured at higher taxonomic classifications (e.g., class/order) across many studies (Lin et al., 2012; Flynn et al., 2013; Ben Maamar et al., 2015; Gülay et al., 2016; Graham et al., 2017; Pogoda, 2017; Lee et al., 2018; Fiedler et al., 2018). Accordingly, a phenetics-based OTU approach was applied.

The UPARSE amplicon analysis algorithm (Edgar, 2013) from the USEARCH8 (32-bit) package (Edgar, 2010) implemented in IMNGS (Lagkouvardos et al., 2016) using default quality settings was used to generate OTUs. A minimum abundance cutoff of 0.25% for each OTU in a sample was set for its inclusion in the final OTU feature table. OTU sequences were aligned and classified as taxa using the SILVA reference database (v. 132, Quast et al., 2013; Glöckner et al., 2017); OTUs classified as chloroplast (2965 sequence reads, 0.1% of sequence reads), mitochondria (602 sequence reads, 0.02% of sequence reads), or archaea (79,582 sequence reads, 3% of sequence reads) were also filtered prior to data analysis.

After sequence quality filtering, a total of 2,541,261 sequence reads across all 83 purge water samples were assignable to 288 bacterial OTUs. The mean relative abundance of unclassified bacterial sequence reads accounted for 26% at the class level, and increased to 55% at the genus level. Twenty bacterial OTUs were unclassified at the phylum rank (mean relative abundance 9.5%) and were taxonomically filtered

as these sequences are likely associated with DNA that were unintended targets of the primers used, or are novel chimeras (Haas et al., 2011). The mean sequencing library depth after quality and taxonomic filtering was 27,891 sequence reads. The sequences affiliated with 183 of 259 bacterial OTUs (for which model fitting was possible) were deemed to be either significantly greater or present exclusively in the environmental samples (Supplementary Information S3).

All analyses were performed within the "R" environment using the packages "phyloseq" (McMurdie and Holmes, 2013) "vegan" (Oksanen et al., 2019) and "DESeq2" (Love et al., 2014) on the raw sequence read data. Alpha diversity was evaluated using the number of observed species and the *Chao1* and *Shannon* indices (Chao, 1984; Shannon, 1948). Non-metric multidimensional scaling (NMDS) was conducted using Bray-Curtis dissimilarities calculated from log-transformed sequence counts (with a pseudo-count of one for non-detects) to facilitate a two-dimensional representation of the sample relationships. This log-transformation down weights the influence of the most abundant OTUs given that most differences between samples are anticipated to be observed with the transient/rare bacterial community. The pseudo-count adjustment was used only for the visualization of samples rather than for statistical testing. Permutational Multivariate Analysis Of Variance (PERMANOVA) using the ordinated distance matrices against key factors was performed using the *adonis* function in the package "vegan" (Oksanen et al., 2019). Constrained Analysis of Principal Coordinates (CAP) was also performed to explore relationships between sample 16S rRNA bacterial composition and physical and chemical parameters measured concurrently.

Key taxa contributing to the observed sample relationships were identified by statistical testing of the original raw sequence counts. Rather than omitting samples due to low library sequence depth, the negative binomial generalized linear model (NB GLM) framework as implemented through DESeq2 (Love et al., 2014) was used to explore taxa differential abundance. A negative binomial model allows for overdispersion in the sequence count distribution. Briefly, NB GLMs were applied to sequence counts n_{ij} for each taxon i in sample j with fitted mean μ_{ij} and a taxon-specific dispersion parameter α_i. The fitted mean is a function of a sample-specific size factor s_j and a parameter q_{ij} proportional to the expected true concentration of that taxon for sample j. The coefficients β_i provide the $\log_2$-fold-difference estimates for each taxon. Log_2-fold-difference estimates were re-expressed as $\log_{10}$-fold-differences for more intuitive order-of-magnitude interpretations. The default setting for shrinkage estimates was utilized; this allows for imprecise $\log_{10}$-fold-difference estimates associated with OTUs comprising of low counts to be shrunk and facilitate the comparison of these estimates across factors (Love et al., 2014). Saturated models comprising of two- and three-way interactions between all factors were first investigated. Overall model significance was evaluated by determining the difference in likelihood values between a fitted model and a reduced model. All models were reduced to simple additive GLMs of the three main factors after a lack of significant (p>0.05) interactions was observed for the majority of taxa. The significance of the differential abundance estimated for a taxon between factor levels was tested using a Wald chi-square test (Love et al., 2014). All p-values were adjusted for multiple comparisons controlling for the false discovery rate (Benjamini and Hochberg, 1995).

3. Results

3.1. The core bacterial community: dominant taxonomic groups observed in groundwater samples

Two major discontinuities in the empirical taxa abundance distribution comprising all groundwater samples were observed (Supplementary Information S4). These continuities were used to distinguish the bacterial core community (OTUs observed in greater than 91% of samples), the transient bacterial community (OTUs observed between 56% and 91% of samples), and the rare bacterial community (OTUs observed in less than 56% of samples). The discontinuities are marked by OTU sequence reads affiliated with the genera *Novosphingobium* (present in 76 of 83 samples, Fig. 1 dashed-line) and *Brevibacillus* (present in 46 of 83 samples, Fig. 1 dash-dotted line). Additionally, 107 bacterial OTUs were consistently observed in all groundwater samples collected; their reads collectively comprised 78% of reads within each sample, on average. Sequence reads affiliated with OTUs from the phyla *Proteobacteria*

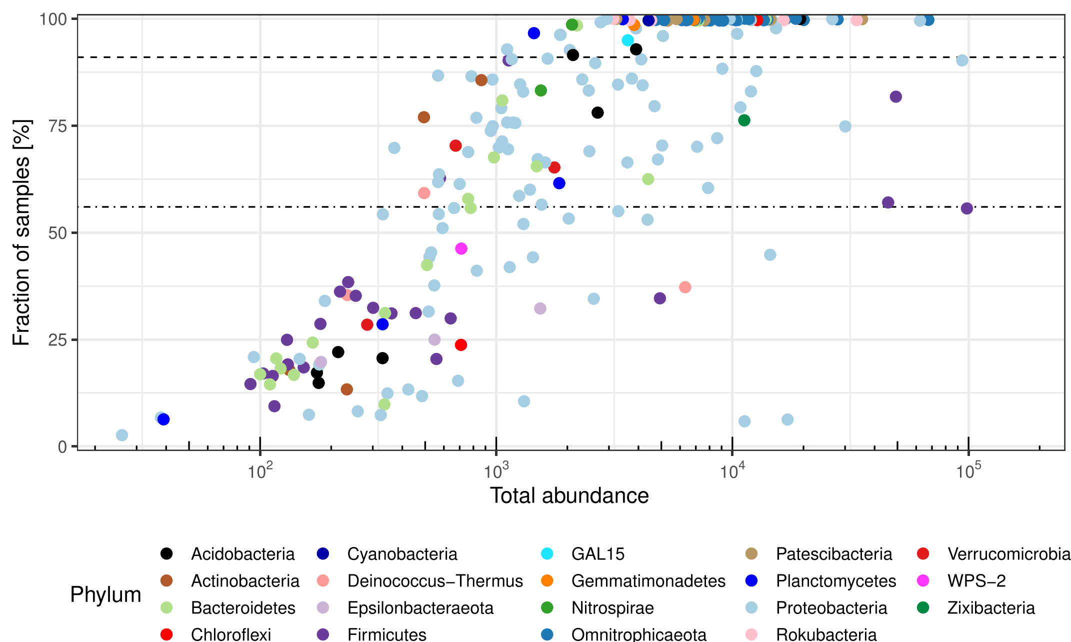


Fig. 1. Total abundance of OTU sequence reads across all samples plotted against their prevalence (fraction of samples across which an OTU was detected). Dashed (91%) and dot-dashed (56%) lines denote the prevalence thresholds determined based on the empirical taxa abundance distribution of all samples (Supplementary Information S4).

(372,762 reads across 27 OTUs), *Omnitrophia* (646,620 reads across 56 OTUs), and *Patescibacteria* (167,169 reads across 8 OTUs) dominated the core bacterial community. *Rokubacteria* (115,352 sequence reads across 9 OTUs), *Nitrospirae* (38,080 sequence reads across 4 OTUs), *Cyanobacteria* (48,081 sequence reads across 3 OTUs) can also be considered part of this environment's core community. Notably, sequence reads belonging to an OTU of the *Escherichia* genus was the most abundant of the *Proteobacteria* (93,794 reads) and ubiquitous across all 83 purge water samples. The same sequence reads were also present in the negative controls, albeit at significantly lower levels than in the purge water samples ($p<0.001$). *Bacteroidetes*, *Firmicutes* and *Proteobacteria* were the only phyla that had OTU sequences represented as part of the core, transient and rare bacterial communities. A relative abundance plot of bacterial taxa on phylum level across all purge water samples was also generated (Fig. 2); additional relative abundance plots comprising of only core community taxa, and a focused examination of *Proteobacteria* composition, are presented in Supplementary Information S4. These plots complement the empirical taxa distribution plot and highlight the stability of the core bacterial community and the dominance of members affiliated with the *Proteobacteria*, *Omnitrophia*, and *Patescibacteria* phyla. Sequence reads affiliated with members of the *Firmicutes* phylum are notably excluded from the core community, but were generally present at higher relative abundances during the intermediate stages of purging. Sequence reads of *Proteobacteria* OTUs appear more prominent in the summer than the following seasons. Over the same seasons, the relative proportions of *Betaproteobacteria* to other *Proteobacteria* classes changed substantially.

3.2. Seasonal/event-based influences and spatial heterogeneity as key drivers of bacterial community diversity

All alpha diversity metrics calculated (observed diversity richness [OTUs], estimated OTU richness [Chao1], and Shannon diversity, Supplementary Information S5) were consistently higher in well WO78 than WO77 as supported by a Wilcoxon rank sum test ($p<0.001$). Alpha diversity was not linked to the stage of purging ($p>0.20$). Seasonality appeared to be influential only for some alpha diversity metrics (i.e., number of OTUs and Chao1 indices, $p<0.10$), which indicated that more diversity was exhibited in samples collected during the summer period.

Bray-Curtis distances were calculated to characterize the beta-diversity between samples; NMDS performed based on these distances (Fig. 3) revealed a clear separation of samples between the well (PERMANOVA, $R^2 = 0.189$, $p<0.001$) and the seasons (PERMANOVA, $R^2 = 0.142$, $p<0.001$) from which the samples were collected. The purging stage appeared least influential of the factors investigated but remained a significant factor within the ordination (PERMANOVA, $R^2 = 0.054$, $p<0.001$). When PERMANOVA was repeated using individual sampling campaigns as levels in place of the seasonal factor, substantially more of the variation was explained by sampling campaigns ($R^2 = 0.395$, $p<0.001$) than the well membership ($R^2 = 0.160$, $p<0.001$) and purging stage ($R^2 = 0.044$, $p<0.001$). The implications from this analysis are two-fold. First, spatial heterogeneity is demonstrated to exert considerable influence on bacterial community composition in the samples, even when collected from wells spaced merely 5 m apart in the same aquifer. Second, sampling campaign-based changes explained more of the variability in bacterial community composition than the seasonal factor (i.e., summer, winter, spring as specified) while accounting for spatial heterogeneity and the purging extent. This observation suggests that 16S rRNA gene amplicon sequencing is sufficiently sensitive—and therefore suitable for—denoting episodic microbial water quality changes occurring at event-based scales. However, a full site characterization (including the re-definition of seasonality at this site) is beyond the scope of this proof-of-concept demonstration. Consequently, subsequent analyses of individual OTUs in this study were focused on the factors originally specified. CAP yielded similar inferences as the NMDS ordination, although the variability in bacterial community composition attributable to seasonal/sampling campaign influences was made more prominent (Supplementary Information S6). This result was to be expected as the measured environmental variables (e.g., groundwater electrical conductivity, temperature, pH) differed more between sampling campaigns than between wells or throughout the purging process. As expected, these seasonally-influenced parameters were weakly aligned with the first two ordination axes (Supplementary Information S6).

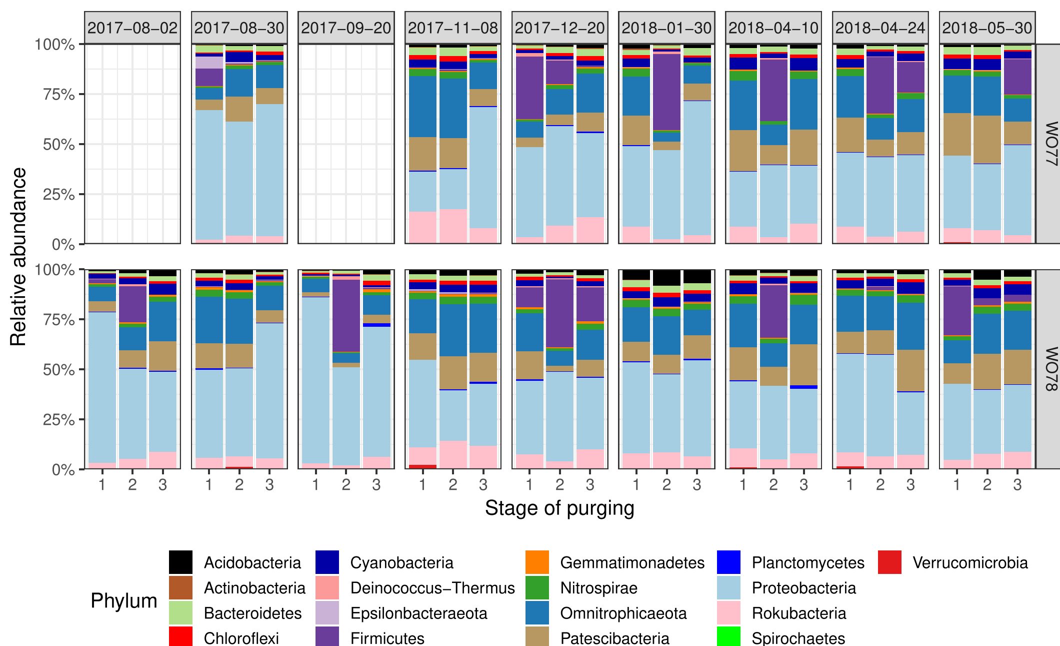


Fig. 2. Summary of relative abundance of bacterial taxa on phylum level, by well membership (WO77 vs. WO78) organized by sampling campaign.

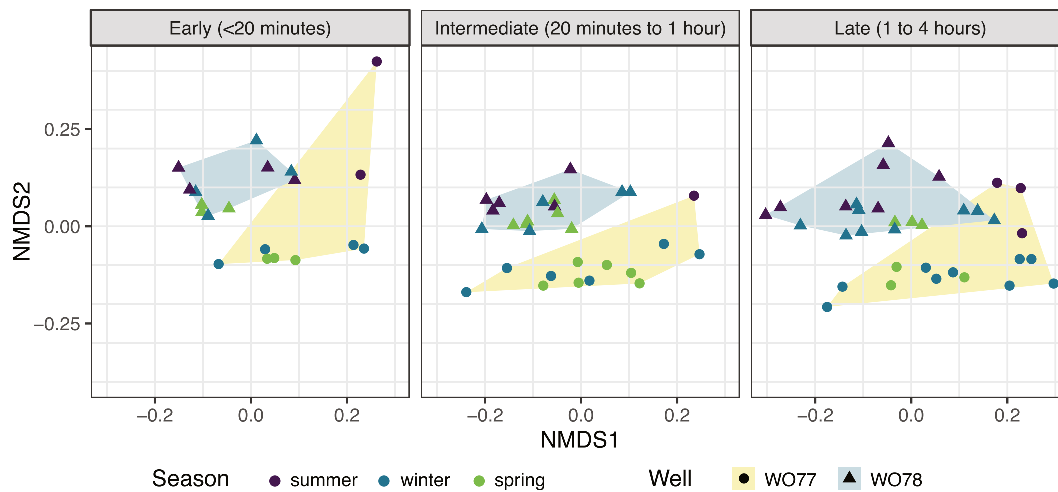


Fig. 3. Non-metric multidimensional scaling (NMDS) ordination of purge water samples using Bray-Curtis dissimilarities (Stress = 0.138). Three facets were used to portray the purging stage achieved at the time of sample collection. Circles and triangles represent the wells from which the samples were collected (WO77 and WO78, respectively); their colours are indicative of the season during which the samples were collected.

3.3. Identification of relevant OTUs through differential abundance testing

Useful indicator taxa were identified by ranking the significance of factor coefficients within the three-factor NB GLM fitted for each OTU (Supplementary Information S7). OTUs comprising of more than an average of 5 sequence reads per sample are represented in Fig. 4 and summarized by class, in order of increasing $\log_{10}$-fold differences in comparison to the base level of each factor. Rather than blanket removal of all OTUs detected in negative control samples, the same statistical framework was used to flag OTUs in the purge water samples that were also detected in the negative controls to account for the possibility of background/reagent contamination.

Despite a generally higher alpha-diversity in WO78 than WO77, the sequences of the vast majority of OTUs comprising the core bacterial community were not significantly different between the two wells. The higher alpha-diversity is suspected to be linked with a higher abundance of bacteria present in WO78; ATP measurements in the same wells indicated higher levels of microbial activity in WO78 that is suggestive of higher microbial densities (Chik et al., 2020). Many OTUs of the same bacterial class responded similarly to each factor investigated. For the purging stage factor coefficients, several OTUs affiliated with the *Bacilli* and *Alphaproteobacteria* classes were significantly higher during the intermediate purging stage. No OTUs between the intermediate and the late stages of purging exhibited a significant increase in sequence read abundance. The shifts of *Proteobacteria* composition between seasons noted by means of relative abundance was confirmed; sequence reads of OTUs from the bacterial class *Betaproteobacteria* were most significantly different between seasons. Sequence reads affiliated with classes of *Bacilli*, *Clostridia*, and *Bacteroidetes* OTUs were also noted to be significantly different between seasons.

4. Discussion

4.1. Bacterial diversity and core community analysis as initial indicators for surface-water groundwater interactions

A handful of studies deploying NGS methods to document microbial communities in subsurface environments suitable for potable water supplies have emerged over the past decade (e.g. Vierheilig et al., 2015; Savio et al., 2018). Although there is no evidence that a community endemic to groundwater sources exists or has been described (Griebler and Lueders, 2009), differences in groundwater and surface water bacterial communities may be used to evaluate the vulnerability of subsurface water sources to fecal pathogen intrusion (e.g. Lin et al., 2012; Ben Maamar et al., 2015; Braun et al., 2016; Graham et al., 2017; Pogoda, 2017; Fiedler et al., 2018). Savio et al. (2019) observed dynamic response of *Flavobacteriia* proportions in karst aquifer water bacterial core community during recharge events. Where the hydrogeological setting is susceptible to rapid changes to microbial water quality (e.g., due to flashy responses attributable to preferential flow paths in mature karst aquifers with high secondary porosity), major bacterial core community shifts are likely indicative of heightened vulnerability to surface-derived microbial transport. In contrast, the hydrogeological setting investigated herein is an unconsolidated sand-and-gravel aquifer where changes to groundwater quality can be expected to be gradual in comparison. In this study, relative abundance and prevalence-based approaches at higher taxonomic levels did not reveal substantial shifts to the core bacterial community observed. Accordingly, a bacterial core community analysis of 16S rRNA gene amplicon sequencing data may serve only as an initial litmus test for obvious signs of vulnerability to surface-derived fecal pathogen intrusion.

Failure to observe substantial shifts to the bacterial core community should not be directly interpreted as a lack of surface water influence or source water quality changes. Both alpha- and beta-diversity metrics indicated that spatial heterogeneity and aspects of seasonality at this site were influential to overall bacterial community composition. In taking spatial heterogeneity into consideration, bacterial community composition was revealed to be sensitive to episodic microbial water quality changes occurring at event-based scales. However, their influences on relevant microbial indicator taxa—especially those comprising the transient/rare community—must be further parsed out considering the inherent limitations of 16S rRNA gene amplicon sequencing (e.g., artefacts arising due to unequal sample sequencing depths). This requires dedicated approaches for differential abundance analysis of sequence read count data. The ensuing discussion focuses on how site-specific source and transport hypotheses can be developed based on the taxa identified from the progression of core community analysis to differential abundance testing. These hypotheses can be refined and further tested to inform groundwater vulnerability assessments to fecal pathogen intrusion.

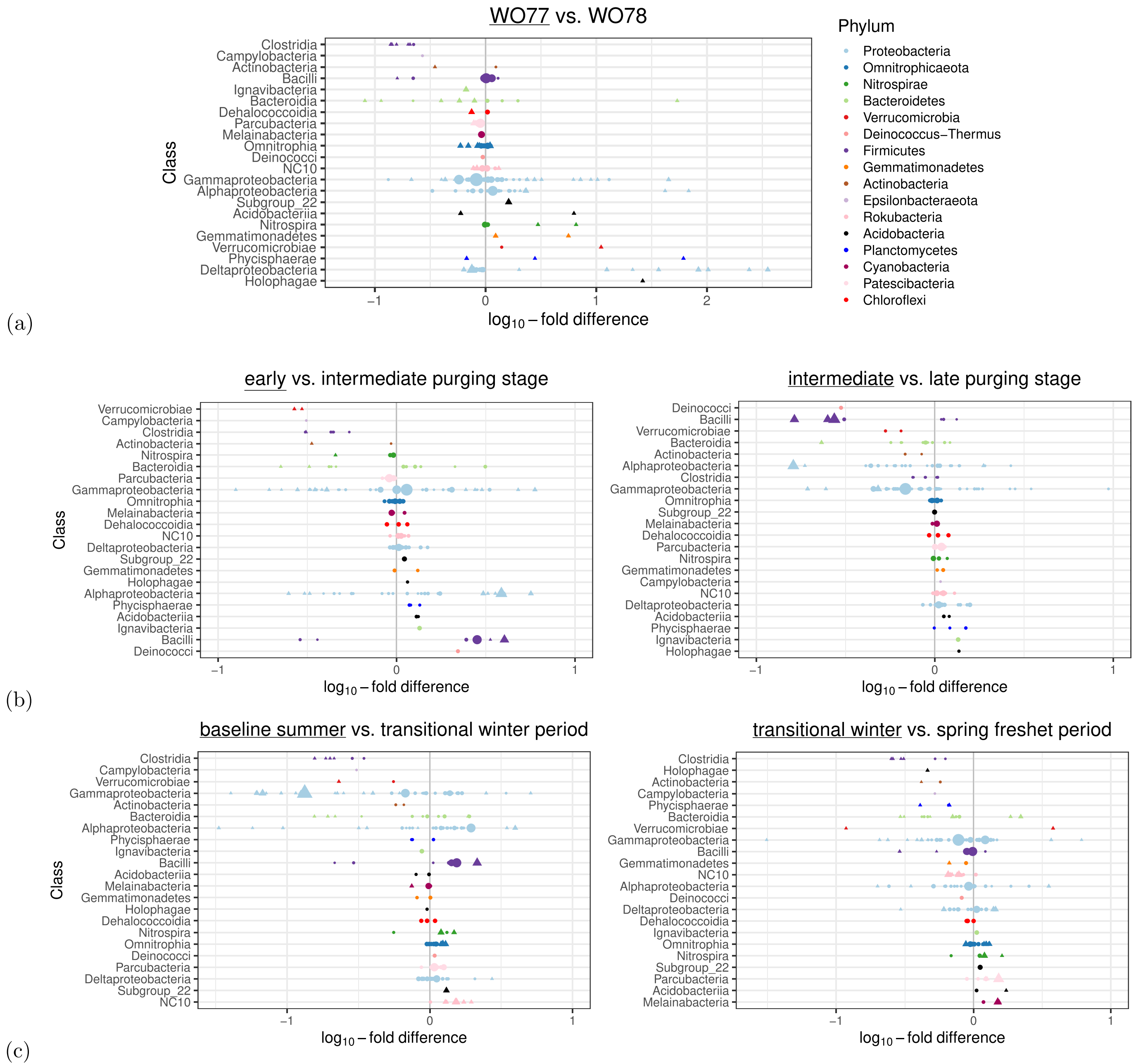


Fig. 4. Differential abundance of various OTUs by class, in order of increasing $\log_{10}$-fold differences in comparison to the base level of each factor across (a) the well factor, (b) the purging stage factor (early vs. intermediate vs. late stages), and (c) the seasonal factor (baseline summer vs. transitional winter vs. spring freshet periods). The base level is underlined in the heading above each plot. Thus, a positive $\log_{10}$ fold difference indicates elevated OTU sequences relative to the base level. Circles represent OTUs that did not exhibit statistically significant different factor coefficients ($p<0.05$), triangles represent those that are significant. The size of the symbols is proportional to the normalized mean number of OTU sequences across all samples.

4.2. Escherichia sequence reads from 16S rRNA gene sequencing require judicious interpretation

Fecal contamination of human/mammalian origin remains the most relevant source of human waterborne diseases globally (Santo Domingo and Ashbolt, 2012). Many members of the genus *Escherichia* are universal gut biota found in humans and warm blooded animals (Madigan et al., 2017); *Escherichia coli* remains the de facto universal indicator of fecal contamination (Santo Domingo and Ashbolt, 2012; Odonkor and Ampofo, 2013). While the negative controls did yield some *Escherichia* sequence reads to indicate possible contamination of the DNA extraction kits used (e.g., Corless et al., 2000; Salter et al., 2014; Pollock et al., 2018), the levels observed in the controls were significantly lower than those typically found in the groundwater samples. Their presence and abundance across all groundwater samples (mean 1165.7 sequence reads per sample) considering the small *effective* volume amplified and sequenced indicate the probable presence of a persistent fecal source likely quantifiable by culture-based methods (only 3.0 μL of undiluted DNA extract from the 100 μL elution volume representing the original 750 mL filtered sample volume was amplified, and only a portion equivalent to 4 nanomolar (nM) of the pure amplicons was sequenced). However, none of the conventional culture-based enumeration procedures (*Escherichia coli* and male-specific coliphage) performed using 100 mL aliquots yielded detections

of the target fecal indicators. This discrepancy underscores the challenges of corroborating results emanating from culture-based and molecular-based methods.

Many factors can ultimately confound the use of *Escherichia* sequences detected using 16S rRNA gene sequencing as fecal indicators. This may include the inherent limitations of measurement error, physiological state of bacteria from which the recovered sequences originate, and the presence of background environmental DNA sequence fragments from deceased microorganisms. Moreover, *Escherichia coli* may be more prevalent in the natural environment than historically anticipated, including those that may not originate from mammalian fecal sources (Scott et al., 2002; van der Wielen and Medema, 2010; Frick et al., 2018). Therefore, extreme caution should be exercised in using these gene sequences directly as replacements of culture-based fecal indicators without enhanced characterization of these sequences. This type of characterization is possible through the design and use of targeted primers and protocols for *Escherichia coli* as in Sabat et al. (2000) and Reischer et al. (2008) or more extensive source monitoring programs designed to track known sources (e.g., septic tanks, agricultural manure, poikilothermic organisms). The use of targeted fecal gene sequences for concentration estimation and comparisons with culture-based results would require evaluation against a standard PCR regression curve (e.g., as applied in quantitative PCR); the method sensitivity limit (MSL, Chik et al., 2018) based on the minimal detectable number of fecal gene sequences per reaction volume (Reischer et al., 2006, 2008) can further facilitate comparisons of method sensitivity using different instruments, protocols, and water matrices.

4.3. Anaerobic taxa sequences as potential vulnerability indicators within an aerobic environment

Sequences of *Clostridia* OTUs in this subsurface environment were generally rare (with several OTUs detected in the negative controls). However, the occurrence of sequence reads of these obligate anaerobes coincided with other anaerobic taxa such as *Bacteroides* and *Prevotella* (phylum *Bacteroidetes*) as well as a notable decrease in several aerobic *Bacillales* OTUs during some sampling campaigns. These anaerobic taxa dominate human and mammalian gut microbiota (Embley and Stackebrandt, 1997; Turnbaugh et al., 2009), but are not exclusive to these sources (van der Wielen and Medema, 2010). In similar aerobic freshwater environments, they are expected to be orders of magnitude lower than aerobic bacteria (e.g., Balkwill and Ghiorse, 1985). Coupled with the fact that *Clostridia* are subject to documented biases attributable to DNA/RNA extraction from gram-positive bacteria (which have structures that are difficult to disrupt, as typical of bacteria within the *Firmicutes* phylum (Embley and Stackebrandt, 1997; von Wintzingerode et al., 1997; Frostegård et al., 1999)), their detection from the overwhelmingly aerobic conditions characteristic of this aquifer (mean dissolved oxygen = 7.9 mg/L, historical mean dissolved oxygen = 8.7 mg/L (Critchley, 2010)) was made more remarkable. Accordingly, the divergent behaviour of anaerobic and aerobic genera may be suggestive of episodic microbial source water quality changes and warrants additional investigation as a site-specific vulnerability indicator of fecal contamination.

4.4. Betaproteobacteria sequences as potential indicators of seasonal subsurface processes

Consistent with the majority of past studies, sequences belonging to the phylum *Proteobacteria* were most dominant in this study (Lin et al., 2012; Flynn et al., 2013; Ben Maamar et al., 2015; Gülay et al., 2016; Graham et al., 2017; Pogoda, 2017; Lee et al., 2018; Fiedler et al., 2018). The conditions of the present study are perhaps most closely aligned with bacterial groups observed from nitrate-impacted groundwaters in Brittany, France (Ben Maamar et al., 2015). The nitrate levels were much higher in the groundwater samples (45–57 mg/L) than in the present study (14.0 mg/L) and sequences from the *Betaproteobacteria* class were identified to be the most abundant. Ben Maamar et al. (2015) observed that owing to their prevalence across samples and their distinct relative abundance profiles, *Betaproteobacteria* were useful to distinguish "recent" (<25 years) from "older" (>40 years) groundwaters. OTUs belonging to the *Comamonadaceae*, *Burkholderiaceae*, and *Oxalobacteraceae* bacterial families were observed to dominate over other *Betaproteobacteria* families in "recent" groundwaters compared to "older" groundwaters. In another study of a riverbank filtration site, the increase of *Burkholderiaceae* sequences in wells relative to other *Proteobacteria* has also been suggested to be indicative of allochthonous microbial loads (Fiedler et al., 2018).

In the present study, OTUs from the *Betaproteobacteria* class (classified in SILVA v. 132 as the order *Betaproteobacteriales* within the class of *Gammaproteobacteria*) were identified to exhibit the most significant seasonal responses. Their levels (both in terms of relative and differential abundance) were elevated during the summer when elevated groundwater electrical conductivity measurements may otherwise suggest an "older" average groundwater age. While this observation apparently contradicts the inferences drawn by Ben Maamar et al. for their use as an indicator of groundwater age, we speculate that both observations may be collectively explained by the elevated nitrate levels associated with the "recent" groundwater in Ben Maamar's study and the historically elevated nitrate concentrations during the growing season at this agricultural site (Haslauer, 2005; Koch, 2009). Closer examination of relative abundance at lower levels of bacterial taxonomy (Supplementary Information S8) revealed the dominance of *Pseudomonadaceae* and *Burkholderia* OTUs. Both families are known to contain denitrifying organisms (and generally regarded as copiotrophs) that could catalyze the full series of denitrification reactions or specialize in the reduction of nitrite produced by bacteria affiliated with the *Enterobacteriales* order (Lycus et al., 2017; Grießmeier and Gescher, 2018). Additional microbial and chemical water quality analyses focused on *Betaproteobacteria* taxa and nitrogenous chemical species in wells within the vicinity may be warranted to investigate their use as indicators of surface water-groundwater interaction (as they may be linked to agricultural activities occurring at this site). Bacterial biomass turnover times through analysis of leucine incorporation as in van Driezum et al. (2018) may also be used to discern whether the seasonal response exhibited by *Betaproteobacteria* is a result of autochthonous (i.e., growth of existing bacteria in response to nutrient availability in the groundwater) or allochthonous processes (i.e., introduction of microorganisms from external water source[s]).

4.5. Bacillales and Sphingomonadales sequences as potential indicators of adequate well purging

Subsurface microbial biofilms in and surrounding the well screen are known to contribute microbial water quality artefacts in abstracted well water, especially upon initiation of well purging activities (Cullimore, 2007; Chik et al., 2020). Consequently, extensive purging is often necessary to obtain samples that are representative of suspended microorganisms in the aquifer pore water void of these artefacts (Harter et al., 2014). In this regard, the elevated levels of sequence reads affiliated with *Bacillales* and *Sphingomonadales* OTUs during the intermediate purging stage appear to indicate the relative position of the "focal biomass" (Cullimore, 2007) in the subsurface. Both genera of bacteria are ubiquitous in soil and groundwater environments (Brooks et al., 2015). Some of the *Bacillales* (*Paenibacillus*, *Brevibacillus*) and *Sphingomonadales* (*Sphingomonas*, *Novosphingobium*) sequences belong to genera known to produce extracellular polymeric substances (EPS) and are key taxa found in biofilm reactors mimicking aerobic groundwater conditions (Ross et al., 2001; Rickard et al., 2002). However, as *Bacillales*-affiliated OTUs may be subject to aforementioned biases attributable to DNA/RNA extraction from gram-positive bacteria, sequence reads of *Sphingomonadales* OTUs may be more useful to

indicate potential well-related biofilm artefacts related to purging activities. *Sphingomonas* spp. in particular have been identified to be a "highly co-aggregating genera" that are quantitatively important members of freshwater biofilms due to their ability to provide other community members with a colonization advantage through adhesion mechanisms (Rickard et al., 2002). The significant reduction of sequence reads from both *Bacillales* and *Sphingomonas* OTUs—and the lack of significant increase of all other OTUs during the last purging stage—provides a useful indication of adequate well purging for aquifer-representative microbial water quality samples.

4.6. Scarce Cyanobacteria and Flavobacteriia sequences as potential lines of evidence of limited surface water-groundwater bacterial transport processes

Recently, pigment-bearing microorganisms including those commonly found in the phylum *Cyanobacteria* have been suggested as indicators of recent surface-derived microbial water quality influence (Ahmed et al., 2013). Many of the bacteria within this phylum are obligate photoautotrophs which have been deemed unlikely to persist in the subsurface void of light sources (Minda et al., 2008). Therefore, their increased levels in the subsurface is likely indicative of surface-derived microbial transport. *Flavobacteriia* abundance in the bacterial community has also been suggested to indicate microbial water quality changes associated with surface discharge events in riverbank filtration and karst environments respectively (Fiedler et al., 2018; Savio et al., 2019).

In this study, all *Cyanobacteria* sequences were affiliated with three OTUs belonging to the class *Melainabacteria*. However, this bacterial class has been more recently proposed as a candidate phylum sibling to *Cyanobacteria* as its members can fix nitrogen and does not perform photosynthesis (Di Rienzi et al., 2013). The lack of other *Cyanobacteria* taxa and scarce levels of *Flavobacteriia*-affiliated sequences across all samples—along with consistent non-detects across all culture-based ALGE-BARTs conducted—indicate that the shallow unconfined aquifer in the present study may not be highly susceptible to local/recent surface-derived microbial water quality impacts through transport. This hypothesis is further supported by hydrogeological evidence that average travel times through the unsaturated zone to the water table at this location has been estimated to be on the order of 2.8 to 5.6 years (Sousa, 2013). The confirmation of the abundance and prevalence in higher soil horizons or nearby surface water features (e.g., ephemeral streams and agricultural ditches) may be a promising line of evidence indicating limited surface-derived microbial transport processes.

5. Conclusions

In this work, the utility of bacterial community composition to inform vulnerability assessments of subsurface water sources to fecal pathogen intrusion was investigated. An initial evaluation of the core bacterial community did not yield any obvious indication of a high level of vulnerability. Spatio-temporal factors influencing bacterial community composition and identification of useful indicator taxa was subsequently demonstrated. Spatial heterogeneity exerted appreciable influence on bacterial community composition; its consideration was paramount to parsing out changes attributable to event-specific/seasonal factors. The consideration of these factors are necessary to meaningfully characterize microbial water quality at the aquifer scale. Accordingly, examination of bacterial indicator taxa relevant as indicators of fecal source and surface connectivity was performed using NB GLMs considering these key spatio-temporal factors.

16S rRNA gene sequences linked to fecal contamination (i.e., *Escherichia* genus) were consistently detected in all samples. However, their use in lieu of culture-based fecal indicators is not recommended due to inherent differences between culture-based and culture-independent methods. On the other hand, the divergent behaviour of anaerobic taxa (specifically *Clostridia* and *Bacteroides*) and aerobic taxa (*Bacilli*) sequences during some events was observed. Their episodic appearance in highly aerobic conditions at this site warrants further investigation as possible indicators of fecal contamination. The sequence reads belonging to the order *Burkholderiales* (class: *Betaproteobacteria*) exhibited dynamics that likely reflect seasonal water quality changes and nutrient dynamics at this site. How these changes are linked to surface-water groundwater interactions can further be elucidated through focused evaluation of this taxon, the use of biochemical techniques such as leucine incorporation, and chemical water quality analyses. Additionally, the abundance of *Bacillales* and *Sphingomonadales* sequences may be useful for informing the adequacy of well purging to obtain groundwater samples representative of suspended microorganisms in the aquifer pore water. Finally, the few *Cyanobacteria* and *Flavobacteriia* gene sequences observed—both of which are commonly reported in microbial community studies of surface water sources—provide a promising line of evidence that microbial transport from surface-derived sources may be rather limited at this site. Additional targeted chemical and microbial sampling informed by this work can be used to refine site-specific hypotheses related to microbial source and transport, thereby informing vulnerability assessments of these subsurface water sources to fecal pathogen intrusion.

CRediT authorship contribution statement

Alex H.S. Chik: Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Monica B. Emelko:** Conceptualization, Methodology, Project administration, Writing - review & editing. **William B. Anderson:** Investigation, Writing - review & editing. **Kaitlyn E. O'Sullivan:** Investigation, Formal analysis, Writing - review & editing. **Domenico Savio:** Investigation, Writing - review & editing. **Andreas H. Farnleitner:** Investigation, Writing - review & editing. **Alfred P. Blaschke:** Project administration, Writing - review & editing. **Jack F. Schijven:** Project administration, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Metagenom Bio Inc., performed the 16S rRNA gene sequencing. We thank Oxford County and Dr. Dave Rudolph for providing site access. We are also grateful to Dr. Kirsten Mueller (UW), Dr. Maria Mesquita (UW) and Dr. Ilias Lagkouvardos (Technische Universität München) for insightful discussions. Terry Ridgway, Mark Sobon, Benjamin J.M. Beelen, and Caitlin Wong are acknowledged for their help throughout the sampling campaigns. The authors also acknowledge the support of this research by the Regional Municipality of Waterloo and Alberta Innovates Grant No AI2385.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.140472.

References

Ahmed, A., Emelko, M.B., Conant, B., Chik, A.H.S., 2013. Development of the new MOE guidance document to determine minimum treatment requirements for municipal residential groundwater systems. OWWA/OMWA Joint Annual Conference. Ottawa, ON.

Balkwill, D.L., Ghiorse, W.C., 1985. Characterization of subsurface bacteria associated with two shallow aquifers in Oklahoma. Appl. Environ. Microbiol. 50, 580–588.

Ben Maamar, S., Aquilina, L., Quaiser, A., Pauwels, H., Michon-Coudouel, S., Vergnaud-Ayraud, V., Labasque, T., Roques, C., Abbott, B.W., Dufresne, A., 2015. Groundwater

isolation governs chemistry and microbial community structure along hydrologic flowpaths. Front. Microbiol. 6. https://doi.org/10.3389/fmicb.2015.01457.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
Braun, B., Schröder, J., Knecht, H., Szewzyk, U., 2016. Unraveling the microbial community of a cold groundwater catchment system. Water Res. 107, 113–126. https://doi.org/10.1016/j.watres.2016.10.040.
Brooks, G., Carroll, K.C., Butel, J., Morse, S.A., Mietzner, T.A., 2015. Spore-forming gram-positive Bacilli: Bacillus and Clostridium species. Jawetz, Melnick, & Adelberg's Medical Microbiology, 27th ed. McGraw-Hill Education, New York, NY.
Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11, 2639–2643. https://doi.org/10.1038/ismej.2017.119.
Chao, A., 1984. Nonparametric estimation of the number of classes in a population. Scand. J. Stat. 11, 265–270.
Chik, A.H.S., Schmidt, P.J., Emelko, M.B., 2018. Learning something from nothing: the critical importance of rethinking microbial non-detects. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.02304.
Chik, A.H.S., Emelko, M.B., Blaschke, A.P., Schijven, J.F., 2020. Illuminating subsurface microbial water quality patterns using adenosine triphosphate and dynamic time warping approaches. Ground Water Monitoring & Remediation https://doi.org/10.1111/gwmr.12397.
Corless, C.E., Guiver, M., Borrow, R., Edwards-Jones, V., Kaczmarski, E.B., Fox, A.J., 2000. Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. J. Clin. Microbiol. 38, 1747–1752.
Critchley, C., 2010. Stimulating In Situ Denitrification in an Aerobic, Highly Conductive Municipal Drinking Water Aquifer. Master's thesis. University of Waterloo Waterloo, Ontario, Canada.
Cullimore, D.R., 2007. Practical Manual of Groundwater Microbiology. Second edition. CRC Press.
Devlin, J.F., Schillig, P.C., Bowen, I., Critchley, C.E., Rudolph, D.L., Thomson, N.R., Tsoflias, G.P., Roberts, J.A., 2012. Applications and implications of direct groundwater velocity measurement at the centimetre scale. J. Contam. Hydrol. 127, 3–14. https://doi.org/10.1016/j.jconhyd.2011.06.007.
Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F., Ley, R.E., 2013. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. eLife 2, e01102. https://doi.org/10.7554/eLife.01102.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461.
Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998. https://doi.org/10.1038/nmeth.2604.
Embley, T.M., Stackebrandt, E., 1997. Species in practice: exploring uncultured prokaryote diversity in natural samples. Systematics Association Special 54, 61–82.
Fiedler, C.J., Schönher, C., Proksch, P., Kerschbaumer, D.J., Mayr, E., Zunabovic-Pichler, M., Domig, K.J., Perfler, R., 2018. Assessment of microbial community dynamics in river bank filtrate using high-throughput sequencing and flow cytometry. Front. Microbiol. 9, 2887.
Flynn, T.M., Sanford, R.A., Ryu, H., Bethke, C.M., Levine, A.D., Ashbolt, N.J., Santo Domingo, J.W., 2013. Functional microbial diversity explains groundwater chemistry in a pristine aquifer. BMC Microbiol. 13, 146. https://doi.org/10.1186/1471-2180-13-146.00047.
Frick, C., Vierheilig, J., Linke, R., Savio, D., Zornig, H., Antensteiner, R., Baumgartner, C., Bucher, C., Blaschke, A.P., Derx, J., Kirschner, A.K.T., Ryzinska-Paier, G., Mayer, R., Seidl, D., Nadiotis-Tsaka, T., Sommer, R., Farnleitner, A.H., 2018. Poikilothermic animals as a previously unrecognized source of fecal indicator bacteria in a backwater ecosystem of a large river. Appl. Environ. Microbiol. 84. https://doi.org/10.1128/AEM.00715-18.
Frostegård, Å., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Gall, F.L., Jeannin, P., Nesme, X., Simonet, P., 1999. Quantification of bias related to the extraction of DNA directly from soils. Appl. Environ. Microbiol. 65, 5409–5420.
Glöckner, F.O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., Ludwig, W., 2017. 25 years of serving the community with ribosomal RNA gene reference databases and tools. J. Biotechnol. 261, 169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198.
Graham, E.B., Crump, A.R., Resch, C.T., Fansler, S., Arntzen, E., Kennedy, D.W., Fredrickson, J.K., Stegen, J.C., 2017. Deterministic influences exceed dispersal effects on hydrologically-connected microbiomes. Environ. Microbiol. 19, 1552–1567. https://doi.org/10.1111/1462-2920.13720.
Griebler, C., Lueders, T., 2009. Microbial biodiversity in groundwater ecosystems. Freshw. Biol. 54, 649–677. https://doi.org/10.1111/j.1365-2427.2008.02013.x.
Grießmeier, V., Gescher, J., 2018. Influence of the potential carbon sources for field denitrification beds on their microbial diversity and the fate of carbon and nitrate. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01313.
Gülay, A., Musovic, S., Albrechtsen, H.-J., Al-Soud, W.A., Sørensen, S.J., Smets, B.F., 2016. Ecological patterns, diversity and core taxa of microbial communities in groundwater-fed rapid gravity filters. The ISME Journal 10, 2209–2222. https://doi.org/10.1038/ismej.2016.16.
Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., 2011. Chimeric 16S rRNA sequence formation and detection in sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21, 494–504.
Harter, T., Watanabe, N., Li, X., Atwill, E.R., Samuels, W., 2014. Microbial groundwater sampling protocol for fecal-rich environments. Groundwater 52, 126–136. https://doi.org/10.1111/gwat.12222.
Haslauer, C.P., 2005. Hydrogeologic Analysis of a Complex Aquifer System and Impacts of Changes in Agricultural Practices on Nitrate Concentrations in a Municipal Well Field: Woodstock, Ontario. Master's thesis. University of Waterloo Waterloo, Ontario, Canada.
Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., Melnik, A.V., Morton, J.T., Navas, J., Quinn, R.A., Sanders, J.G., Swafford, A.D., Thompson, L.R., Tripathi, A., Xu, Z.Z., Zaneveld, J.R., Zhu, Q., Caporaso, J.G., Dorrestein, P.C., 2018. Best practices for analysing microbiomes. Nat. Rev. Microbiol. 16, 410. https://doi.org/10.1038/s41579-018-0029-9.
Koch, J., 2009. Evaluating Regional Aquifer Vulnerability and BMP Performance in an Agricultural Environment Using a Multi-scale Data Integration Approach. Master's thesis. University of Waterloo Waterloo, Ontario, Canada.
Korbel, K., Chariton, A., Stephenson, S., Greenfield, P., Hose, G.C., 2017. Wells provide a distorted view of life in the aquifer: implications for sampling, monitoring and assessment of groundwater ecosystems. Sci. Rep. 7, srep40702. https://doi.org/10.1038/srep40702.
Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., Clavel, T., 2016. IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci. Rep. 6, 33721. https://doi.org/10.1038/srep33721.
Lee, J.-H., Lee, B.-J., Unno, T., 2018. Bacterial communities in ground-and surface water mixing zone induced by seasonal heavy extraction of groundwater. Geomicrobiol J. 35, 768–774. https://doi.org/10.1080/01490451.2018.1468834.
Lin, X., McKinley, J., Resch, C.T., Kaluzny, R., Lauber, C.L., Fredrickson, J., Knight, R., Konopka, A., 2012. Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. The ISME Journal 6, 1665–1676. https://doi.org/10.1038/ismej.2012.26.
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.
Lycus, P., Lovise Bøthun, K., Bergaust, L., Peele Shapleigh, J., Reier Bakken, L., Frostegård, Å., 2017. Phenotypic and genotypic richness of denitrifiers revealed by a novel isolation strategy. The ISME Journal 11, 2219–2232. https://doi.org/10.1038/ismej.2017.82.
Madigan, M.T., Bender, K.S., Buckley, D.H., Sattley, W.M., Stahl, D.A., 2017. Brock Biology of Microorganisms. 15th ed. Pearson, New York.
McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217. http://dx.plos.org/10.1371/journal.pone.0061217.
McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput. Biol. 10. https://doi.org/10.1371/journal.pcbi.1003531.
Minda, R., Joshi, V.P., Bhattacharjee, S.K., 2008. The evolutionary significance of 'obligate' Photoautotrophy of Cyanobacteria. Curr. Sci. 94, 850–852.
Odonkor, S.T., Ampofo, J.K., 2013. Escherichia coli as an indicator of bacteriological quality of water: an overview. Microbiol. Res. 4, e2.
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. vegan: community ecology package. URL. https://CRAN.R-project.org/package=vegan.
Pasha, E., 2018. Quantifying Groundwater Recharge During Dynamic Seasonality in Cold Climates. Ph.D. thesis. University of Waterloo Waterloo, Ontario, Canada.
Pogoda, K., 2017. Targeted 16S rRNA Metagenomic Analysis of E. Coli Contaminated Private Drinking Well Waters in Southern Ontario. Master's thesis. Queen's University, Kingston, Ontario, Canada.
Pollock, J., Glendinning, L., Wisedchanwet, T., Watson, M., 2018. The madness of microbiome: attempting to find consensus "best practice" for 16S microbiome studies. Appl. Environ. Microbiol. 84. https://doi.org/10.1128/AEM.02627-17 e02627–17.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.
Reischer, G.H., Kasper, D.C., Steinborn, R., Mach, R.L., Farnleitner, A.H., 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine Karstic regions. Appl. Environ. Microbiol. 72, 5610–5614. https://doi.org/10.1128/AEM.00364-06.
Reischer, G.H., Haider, J.M., Sommer, R., Stadler, H., Keiblinger, K.M., Hornek, R., Zerobin, W., Mach, R.L., Farnleitner, A.H., 2008. Quantitative microbial faecal source tracking with sampling guided by hydrological catchment dynamics. Environ. Microbiol. 10, 2598–2608. https://doi.org/10.1111/j.1462-2920.2008.01682.x.
Rickard, A.H., Leach, S.A., Hall, L.S., Buswell, C.M., High, N.J., Handley, P.S., 2002. Phylogenetic relationships and coaggregation ability of freshwater biofilm bacteria. Appl. Environ. Microbiol. 68, 3644–3650. https://doi.org/10.1128/AEM.68.7.3644-3650.2002.
Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616.
Rose, S., Long, A., 1988. Monitoring dissolved oxygen in ground water: some basic considerations. Ground Water Monitoring & Remediation 8, 93–97.
Ross, N., Villemur, R., Marcandella, E., Deschenes, L., 2001. Assessment of changes in biodiversity when a community of ultramicrobacteria isolated from groundwater is stimulated to form a biofilm. Microb. Ecol. 42, 56–68.
Sabat, G., Rose, P., Hickey, W.J., Harkin, J.M., 2000. Selective and sensitive method for PCR amplification of Escherichia coli 16S rRNA genes in soil. Appl. Environ. Microbiol. 66, 844–849.
Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination

can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87. https://doi.org/10.1186/s12915-014-0087-z.

Santo Domingo, J., Ashbolt, N.J., 2012. Fecal pollution of water. Encyclopedia of Earth. National Council for Science and the Environment, Washington, D.C., USA.

Savio, D., Stadler, P., Reischer, G.H., Kirschner, A.K.T., Demeter, K., Linke, R., Blaschke, A.P., Sommer, R., Szewzyk, U., Wilhartitz, I.C., Mach, R.L., Stadler, H., Farnleitner, A.H., 2018. Opening the black box of spring water microbiology from alpine karst aquifers to support proactive drinking water resource management. WIREs. Water 5, e1282. https://doi.org/10.1002/wat2.1282.

Savio, D., Stadler, P., Reischer, G.H., Demeter, K., Linke, R.B., Blaschke, A.P., Mach, R.L., Kirschner, A.K.T., Stadler, H., Farnleitner, A.H., 2019. Spring water of an alpine karst aquifer is dominated by a taxonomically stable but discharge-responsive bacterial community. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.00028.

Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., Lukasik, J., 2002. Microbial source tracking: current methodology and future directions. Appl. Environ. Microbiol. 68, 5796–5803. https://doi.org/10.1128/AEM.68.12.5796-5803.2002.

Shannon, C.E., 1948. A mathematical theory of communication. Bell system technical journal 27, 379–423.

Sousa, M., 2013. Using Numerical Models for Managing Water Quality in Public Supply Wells. Ph.D. thesis. University of Waterloo Waterloo, Ontario, Canada.

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., Nishijima, M., 2014. Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. PLoS One 9, e105592. https://doi.org/10.1371/journal.pone.0105592.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R., Gordon, J.I., 2009. A core gut microbiome in obese and lean twins. Nature 457, 480–484. https://doi.org/10.1038/nature07540.

US EPA, 2001. Method 1601: Male-specific (F+) and Somatic Coliphage in Water by Two-step Enrichment Procedure. Technical Report EPA 821-R-01-030. Office of Water Washington, D.C.

van der Wielen, P.W.J.J., Medema, G., 2010. Unsuitability of quantitative Bacteroidales 16S rRNA gene assays for discerning fecal contamination of drinking water. Appl. Environ. Microbiol. 76, 4876–4881. https://doi.org/10.1128/AEM.03090-09.

van Driezum, I.H., Chik, A.H.S., Jakwerth, S., Lindner, G., Farnleitner, A.H., Sommer, R., Blaschke, A.P., Kirschner, A.K.T., 2018. Spatiotemporal analysis of bacterial biomass and activity to understand surface and groundwater interactions in a highly dynamic riverbank filtration system. Sci. Total Environ. 627, 450–461. https://doi.org/10.1016/j.scitotenv.2018.01.226.

Vierheilig, J., Savio, D., Ley, R.E., Mach, R.L., Farnleitner, A.H., Reischer, G.H., 2015. Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. Water Sci. Technol. 72, 1962–1972.

von Wintzingerode, F., Göbel, U.B., Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21, 213–229.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R., 2014. Uniting the classification of cultured and uncultured bacteria and Archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol. 12, 635–645. https://doi.org/10.1038/nrmicro3330.