

OPEN

# A comparative assessment of conventional and molecular methods, including MinION nanopore sequencing, for surveying water quality

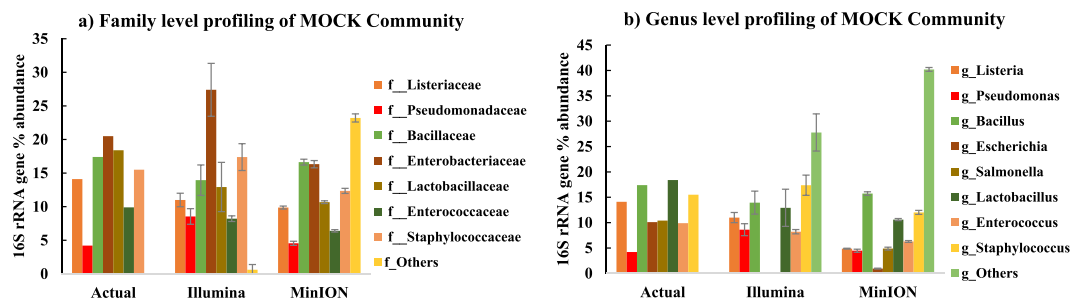
Kishor Acharya<sup>1</sup>, Santosh Khanal<sup>2</sup>, Kalyan Pantha<sup>3,4</sup>, Niroj Amatya<sup>4,5</sup>, Russell J. Davenport<sup>1</sup> & David Werner<sup>1\*</sup>

Nucleic acid based techniques, such as quantitative PCR (qPCR) and next generation sequencing (NGS), provide new insights into microbial water quality, but considerable uncertainty remains around their correct interpretation. We demonstrate, for different water sources in informal settlements in the Kathmandu Valley, Nepal, significant Spearman rank correlations between conventional and molecular microbiology methods that indicate faecal contamination. At family and genera level, 16S rRNA amplicon sequencing results obtained with the low-cost, portable next generation sequencer MinION from Oxford Nanopore Technologies had significant Spearman rank correlations with Illumina MiSeq sequencing results. However, method validation by amplicon sequencing of a MOCK microbial community revealed the need to ascertain MinION sequencing results for putative pathogens at species level with complementary qPCR assays. *Vibrio cholerae* hazards were poorly associated with plate count faecal coliforms, but flagged up by the MinION screening method, and confirmed by a qPCR assay. Plate counting methods remain important to assess viability of faecal coliforms in disinfected water sources. We outline a systematic approach for data collection and interpretation of such complementary results. In the Kathmandu Valley, there is high variability of water quality from different sources, including for treated water samples, illustrating the importance of disinfection at the point of use.

An estimated 1.8 billion people are still exposed to drinking water sources contaminated with faecal matter<sup>1</sup>. In Nepal waterborne diseases account for 15% of all illness and 8% of total deaths<sup>2</sup>. With waterborne diseases accounting for 38% of deaths in children under the age of five<sup>3</sup>, water quality in Nepal is a major public health concern<sup>2,4-6</sup>. According to the Central Bureau of Statistics, Government of Nepal, one out of five families in the Kathmandu Valley does not have access to municipal drinking water, and in most areas availability is <4–7 h/week<sup>3,7</sup>. Many people depend on alternative sources such as groundwater from boreholes and wells, stone spouts, bottled drinking water or water supplied by delivery truck<sup>2,6,8</sup>. The quality of these water sources is not routinely monitored.

The current standard method for microbial examination of both drinking and bathing water requires the isolation and enumeration of organisms that indicate the presence of faecal contamination (i.e. faecal indicator organisms, FIO)<sup>9</sup>. *Escherichia coli* are World Health Organization (WHO) recommended faecal indicator bacteria for drinking water<sup>10</sup>, while *Enterococci* are indicator bacteria for bathing water<sup>11</sup>. FIO are used because there is good epidemiological evidence that they correlate with disease outcomes<sup>12,13</sup>. However, FIO are a poor proxy for organisms with quite different physiologies (e.g. viruses and protozoa) and it is difficult to distinguish FIO from human or animal sources<sup>14</sup>. The routine isolation of all pathogens is impractical<sup>15</sup>, as each requires a unique

<sup>1</sup>School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom. <sup>2</sup>Department of Pharmacology, School of Medicine, University of Colorado, Aurora, Colorado, 80045, USA. <sup>3</sup>Group for Rural Infrastructure Development, Wise use House, Jwagal, Lalitpur, Nepal. <sup>4</sup>Faculty of Chemistry, University Duisburg-Essen, Universitätsstr. 5, D-45141, Essen, Germany. <sup>5</sup>Department of Medical Microbiology, Nobel College, Pokhara University, Kathmandu, Nepal. \*email: [david.werner@ncl.ac.uk](mailto:david.werner@ncl.ac.uk)



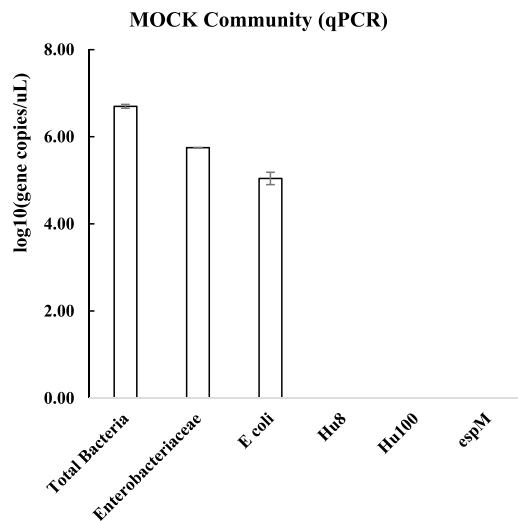
**Figure 1.** Results of the experimentally determined composition of the MOCK Community (MC) at (a) family and (b) genus level with MinION and Illumina NGS techniques. The results are based on 16S rRNA amplicon sequencing. Experimentally determined 16S rRNA gene percentage abundance are compared against the actual composition of the MC as reported by the supplier (i.e. Zymo Research). Data points are an average of duplicate samples for Illumina and triplicate samples for MinION sequencing. Error bars indicate the standard deviation.

microbiological isolation technique<sup>16,17</sup>. In addition, culture dependent approaches may require long incubation periods, and there is a demand for more rapid and comprehensive screening methods to detect FIO and/or their markers<sup>14</sup> and putative pathogens in water samples<sup>15,18–21</sup>. Culture independent methods such as quantitative real-time PCR (qPCR)<sup>14</sup>, next generation sequencing (NGS)<sup>22</sup>, DNA hybridisation platforms and immunoassays<sup>23</sup> allow direct measurement of cellular properties that may identify pathogens (e.g., DNA, RNA, cellular proteins) without incubation. These methods may reduce the detection and quantification time to a few hours<sup>19,20,24</sup>. NGS methods potentially allow simultaneous detection of gene fragments from different types of faecal indicators and putative pathogens, such as thermo-tolerant coliforms, faecal coliforms, *Vibrio cholerae*, *Streptococci*, *Bacteroides*, etc., especially through 16S rRNA gene amplicon sequencing (from here on referred to as amplicon sequencing). However, NGS techniques are typically more expensive and require more sophisticated equipment and reagents than culturing methods<sup>18,24–26</sup>. Also, NGS techniques may have limited taxonomic resolution; due to an inability to capture the near-complete genomes of rare taxa in shot-gun sequencing or an inability to reliably classify taxa down to the species level with amplicon sequencing due to the relatively short fragment sizes that can be sequenced on some NGS platforms, or the presence of highly conserved 16S rRNA genes in some families and genera<sup>22,27,28</sup>. This compromises their ability to reliably detect waterborne hazards. To circumvent some of these limitations, different approaches can be combined. For instance, *Cui et al.* combined Illumina-based amplicon sequencing and qPCR to evaluate the pathogen diversity in urban recreational water<sup>29</sup>. *Ahmed et al.* used host associated molecular markers with qPCR and Illumina-based amplicon sequencing to identify faecal pollution sources in environmental waters in Brisbane, Australia<sup>30</sup>. An exciting development for NGS is the MinION, a low cost, memory-stick sized real time sequencer from Oxford Nanopore Technologies Ltd. Its portability opens up the possibilities for sequencing to be done in the field. Recently, *Hu et al.* compared culturing, MinION shotgun sequencing and Illumina MiSeq amplicon sequencing methods to trace faecal contamination from wastewater in urban stormwater systems<sup>24</sup>. They demonstrated high correlations between *E. coli* culturing counts, the relative abundance of human gut microbiome related amplicon sequences, and the frequencies of human gut microbiome genes from shotgun sequencing data. However, they did not assess their method validity using samples of known composition. When applied to drinking water quality monitoring, qPCR and NGS data were not always found to be well-correlated with culture based methods<sup>18,31</sup>. Reliable identification of putative pathogens at species level remains a challenge, especially for water samples containing low amounts of DNA.

To address these uncertainties, we first evaluated various molecular microbiology methods (qPCR, Illumina and MinION NGS) using samples of known composition (i.e. MOCK communities containing DNA of FIO and putative pathogens). Since our samples of interest would include groundwater and treated drinking water samples with low amounts of DNA, we focused on the evaluation of qPCR and NGS 16S rRNA amplicon sequencing methods. We combined these molecular with conventional methods to assess water quality for different types of household water sources in the Kathmandu Valley, Nepal. To our knowledge, this is the first study using amplicon sequencing with the MinION for microbial water quality analysis, and comparing its results with those obtained with other methods and known sample compositions. Based on the results, we propose a tool-box approach to make the most of complementary technologies for water quality monitoring in areas with a significant waterborne disease burden.

## Results

**MOCK microbial community analysis.** For method validation, we used a MOCK community consisting of genomic DNA from eight bacterial and two fungal species. The MOCK community included Gram-negative and Gram-positive faecal indicator bacteria (*Escherichia coli*, and *Enterococcus faecalis*, respectively) and other putative pathogens (*Salmonella enterica* and *Pseudomonas aeruginosa*) (Table S1 in supporting information and Fig. 1). The percentage abundance of 16S rRNA genes from each species provided by the supplier of the MOCK community (i.e. Zymo Research) is considered as the true or actual abundance in this study. The MOCK community contained an equal ratio of genomic DNA for each species. However, different 16S rRNA gene copy numbers for each species leads to a slightly uneven relative abundance for this gene (Table S1). Figure 1 and Table S2 in supporting information compare the actual composition of the MOCK community with that measured by two different NGS methods at family, genus and species level. We sequenced 16S rRNA gene amplicons from the

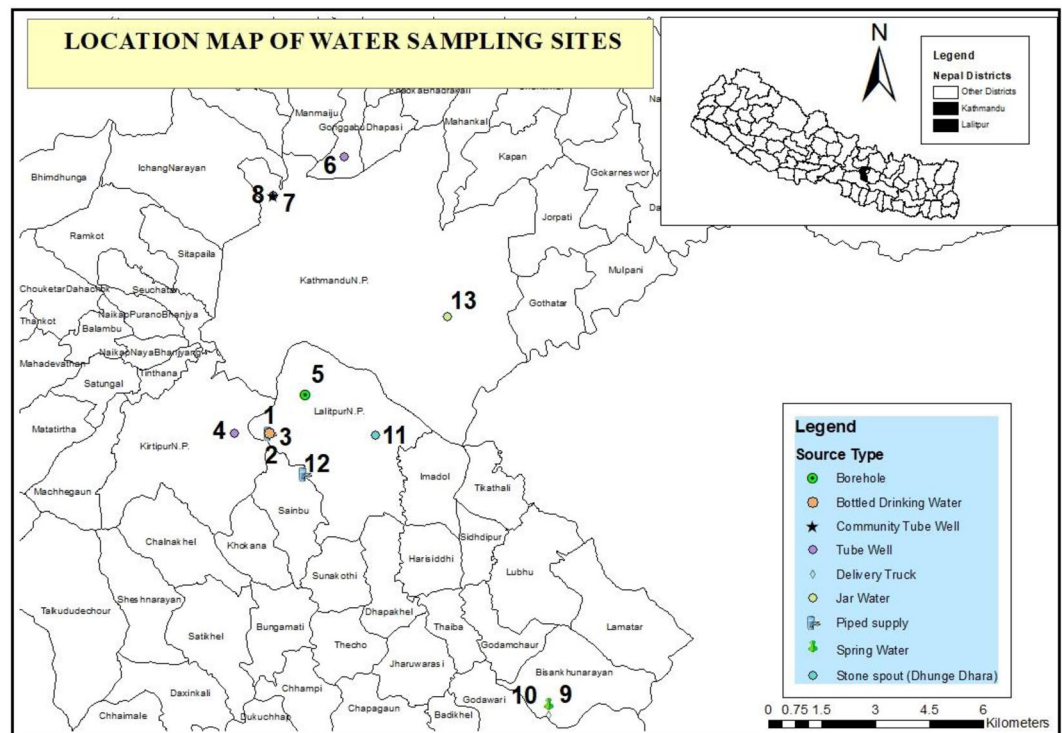


**Figure 2.** Molecular microbiology analysis of the MOCK community. Data points are an average of technical triplicate samples and error bars indicate the standard deviation. Genes associated with human *E. coli* and *Vibrio cholerae* were not detected.

MOCK community using two NGS methods, the portable, memory-stick sized MinION of Oxford Nanopore Technologies, and the MiSeq platform from Illumina, which is currently the most commonly used for microbial community analyses<sup>22,25,32</sup>. Illumina sequencing data showed better taxonomic resolution as compared to MinION data at the family level, since 99% of Illumina reads were correctly classified as those bacterial families present in the MOCK community compared to 76.79% for MinION reads. However, at genus and species level, MinION sequencing data showed better taxonomic resolution as compared to Illumina data (Fig. 1b for genus level and Table S2 for species level). In this study, the presence of the genera *Escherichia* and *Salmonella* in the MOCK community was not identified from the Illumina data, and none of the sequencing reads from Illumina were classified to species level. In contrast, almost 59.41% of the MinION sequencing data were classified to species level. However, at species level, 64.28% of MinION reads were matched to species that were not present in the MOCK community (Table S2). In particular, reads falsely classified as *Escherichia fergusonii* had a higher relative abundance than *Escherichia coli*, 0.65% versus 0.17%, respectively, which compares to their true relative abundances of 0% and 10.1%, respectively in the MOCK community (Tables S1 and S2). In comparison with *E. coli*, MinION sequencing more successfully identified the other faecal pollution indicator, *Enterococcus faecalis*, in the MOCK community. *Enterococcus faecalis* was the most frequently detected *Enterococcus* species, with a measured 16S rRNA read abundance of 5.39% compared to the actual or true 16S rRNA gene abundance of 9.9% (Tables S1 and S2). For the other putative pathogens, *Salmonella enterica* was the most frequently detected *Salmonella* species with a measured 16S rRNA reads abundance of 4.89% compared to a true 16S rRNA gene abundance of 10.4%, and *Pseudomonas aeruginosa* was the most frequently detected *Pseudomonas* species with 3.66% measured abundance compared to a true abundance of 4.2% (Tables S1 and S2). Cross-validation of sequencing results, such as targeting and quantifying marker genes with PCR primers can help in interpreting and affirming sequencing outcomes<sup>30,33</sup>. When using total 16S rRNA, total coliforms, total *E. coli*, human *E. coli* and *Vibrio cholerae* marker gene primers to analyse the MOCK community (Fig. 2), we only identified genes associated with *Enterobacteriaceae* (i.e. total coliform) and *E. coli* (i.e. rodA), but not *Vibrio cholerae* and human *E. coli* genes. These results are consistent with the true or actual composition of the MOCK community, since the *E. coli* DNA in the MOCK community was not from a strain associated with the human gut microbiome.

#### Method comparison for the characterisation of microbial water from different sources in the Kathmandu Valley, Nepal.

NGS (Illumina and MinION), qPCR and plate count methods were combined to investigate microbial water quality for thirteen water sources in the Kathmandu Valley, Nepal (Fig. 3 and Table 1). The data is shown in Fig. 4, and Tables S3–S7 in supporting information. First, the agreement between the two NGS methods was evaluated for several relevant groups of bacteria that include faecal indicators (*Bacteroides*, *Prevotella*, *Enterobacteriaceae*), and other groups containing putative pathogens (i.e. *Vibrio*, *Pseudomonas*, *Legionella*, *Clostridium* and *Streptococcus*), distinguished at family (*Enterobacteriaceae*) or genus level (*all others*). A significant rank correlation was observed (Spearman rank correlation coefficient,  $p < 0.05$ ) between the relative abundances of *Bacteroides*, *Prevotella*, *Enterobacteriaceae* and all other putative pathogenic genera determined by MinION and Illumina NGS (Table S6 in supporting information). The extent of rank correlation between the NGS and other approaches for microbial water quality assessment is illustrated in Fig. 5 separately for NGS data from Illumina [A] and the MinION [B]. To enable this correlation analysis between NGS data (which is relative abundance) and qPCR and plate count methods (which provide absolute abundance), NGS results were first converted to estimated absolute abundances using the 16S rRNA gene copies number determined by qPCR for each sample (Fig. 4d)<sup>21</sup>. When looking at the agreement between NGS and qPCR methods, *Enterobacteriaceae* Illumina data (Fig. 5A) had a significant correlation with the total coliform qPCR (Spearman rank correlation coefficient 0.54,  $p < 0.05$ ), total *E. coli* qPCR (Spearman rank correlation coefficient 0.59,  $p < 0.05$ ), but not with human *E. coli* qPCR (Spearman rank correlation coefficient 0.19,  $p > 0.05$ ) gene



**Figure 3.** Map showing the location of the 13 sampling sites in the Kathmandu Valley and types of water sources sampled.

copy numbers. Likewise, *Enterobacteriaceae* MinION data (Fig. 5B) had a significant correlation with total coliform qPCR (Spearman rank correlation coefficient 0.56,  $p < 0.05$ ) and total *E. coli* qPCR (Spearman rank correlation coefficient 0.71,  $p < 0.05$ ), but not with human *E. coli* qPCR (Spearman rank correlation coefficient 0.27,  $p > 0.05$ ) gene copy numbers. As expected from the MOCK community results, only the NGS results from the MinION contained several reads aligned with putative pathogens at species level (Table S7 in supporting information). For the water samples from the Kathmandu Valley, *Vibrio cholerae* MinION data (Fig. 4c) were well aligned with qPCR results (Fig. 4h and Table S7), which were targeting copies of the gene encoding for the cholera toxin secretion protein *epsM* as an indicator for the abundance of pathogenic *Vibrio cholerae* strains. These genes were generally detected only in the samples from tube wells (location 6, 7 and 8), which were also the only samples for which NGS with the MinION matched some reads to *Vibrio cholerae* at species level. With regards to the agreement between molecular and plate count methods, all MinION NGS data had a significant and positive correlation with the total coliform plate count data (Spearman rank correlation coefficients  $> 0.66$ ,  $p < 0.05$ ). Positive correlations were also observed between faecal *E. coli* plate count and total coliform qPCR, total *E. coli* qPCR, and human *E. coli* qPCR data, respectively (Spearman rank correlation coefficients  $> 0.57$ ,  $p < 0.05$ ). The correlation between faecal *E. coli* plate count and total *E. coli* qPCR data was particularly strong (Spearman rank correlation coefficient 0.86,  $p < 0.05$ ). It is noteworthy that *Vibrio cholerae* MinION and *Vibrio cholerae* qPCR data were negatively correlated with the faecal *E. coli* plate count (Fig. 5B). Consequently, the *Vibrio cholerae* hazard is not readily detected with the classic fecal pollution indicator method. While the overall agreement between the traditional (i.e. plate count) and molecular microbial water quality assessment methods was thus generally good, there were noteworthy discrepancies for disinfected water samples. Water sampled from location 2, 3 and 13 were disinfected waters (2 was disinfected by the household at the point of use while 3 and 13 were commercially sold bottled water), and showed no evidence of thermo-tolerant coliform bacteria by the plate count method (Table S3 in supporting information). However, these groups of bacteria were detected in samples 3 and 13 in significant numbers when analysed with qPCR (Fig. 4e).

### Chemical water quality assessment outcomes for different water sources in the Kathmandu Valley.

Exceedances of the Nepalese guidance values were observed for the following chemical water quality parameters and locations: Nitrate in water from the stone spout (location 11) and a tube well (location 8), ammonia in one tube well (location 7), iron in location 5 (deep borehole) and location 7 (tube well), manganese in location 5 (deep borehole) and location 7 (tube well), and aluminum in location 4 (tube well). More details are provided in Tables S8–10 in supporting information.

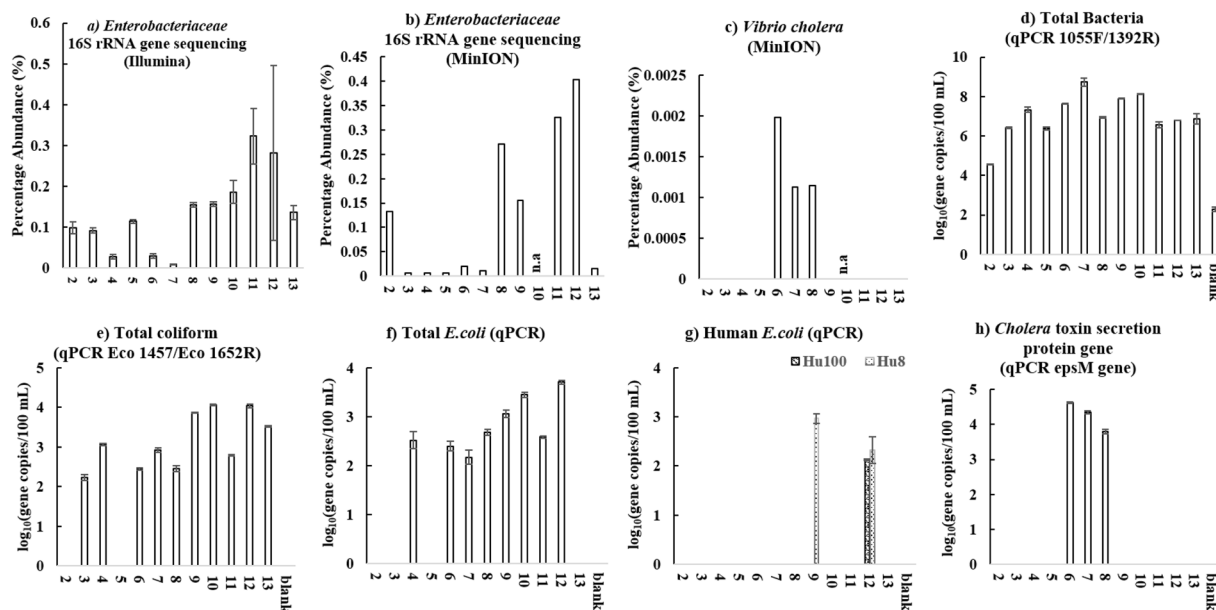
### Discussion

Analysis of a MOCK community revealed that Illumina sequencing data showed better taxonomic resolution as compared to MinION data at the family level, while MinION sequencing data showed better taxonomic resolution at genus and species level. The amplicon size of the targeted 16S rRNA gene sequences was shorter with Illumina than those targeted using the MinION, and short amplicon size is known to compromise the achieved taxonomic

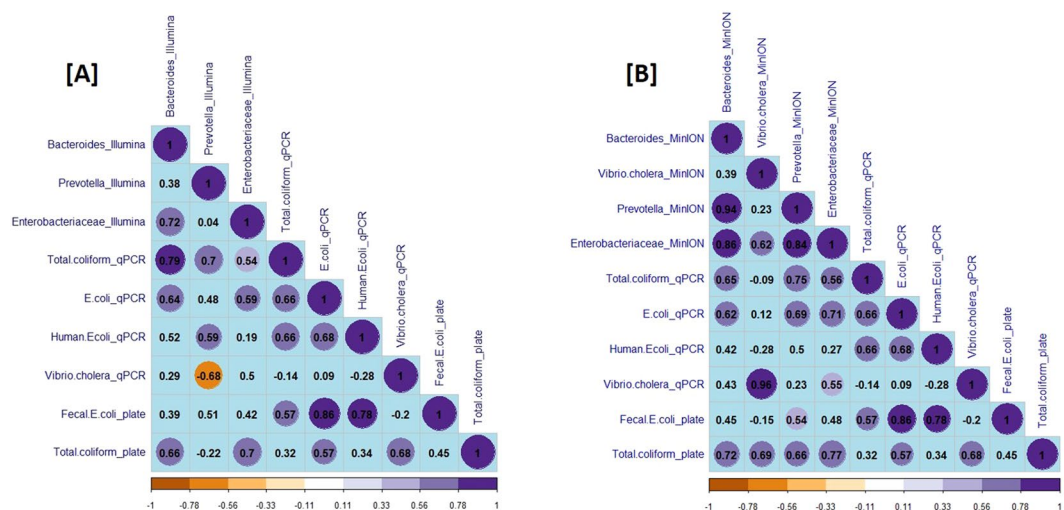
Sample Site ID	Location	Source Type	Latitude	Longitude	Depth to water (m)	Treatment	Additional Protection	Sample collected date	Water Usage	Molecular analysis
1	Dhobighat, Lalitpur	Piped supply	27.673904	85.29678	Ground level	Treated	None	30.10.2018	Washing, Bathing, Gardening	No
2	Dhobighat, Lalitpur	Piped supply	27.673904	85.29678	Ground level	RO/UV treated by houseowner	None	30.10.2018	Drinking	Yes
3	Dhobighat, Lalitpur	Bottled Drinking Water	27.674173	85.297901	Ground level	RO/UV treated	Sealed bottle	30.10.2018	Drinking	Yes
4	Dhobighat, Lalitpur	Tube Well	27.673718	85.294829	10	Untreated	None	30.10.2018	Washing, Bathing, Gardening	Yes
5	Sanepa Height, Lalitpur	Borehole	27.683564	85.306661	91.5	Untreated	None	30.10.2018	Drinking	Yes
6	Basundhara, Kathmandu	Tube Well	27.74304	85.321844	9	Untreated	None	31.10.2018	Washing, Bathing, Gardening	Yes
7	Banasthali, Kathmandu	Tube Well	27.728475	85.295282	9	Untreated	None	31.10.2018	Washing, Bathing, Gardening	Yes
8	Banasthali, Kathmandu	Community Tube Well	27.728501	85.295739	12	Untreated	None	31.10.2018	Washing, Bathing, Gardening, Drinking	Yes
9	Godabari, Lalitpur	Spring Water	27.597466	85.384897	Ground level	Untreated	Mesh Covered	31.10.2018	Washing, Bathing, Gardening, Drinking	Yes
10	Godabari, Lalitpur	Delivery Truck	27.597315	85.384829	Ground level	Untreated	Locked Container	31.10.2018	Washing, Bathing, Gardening, Drinking	Yes
11	Mangalbar, Lalitpur	Stone spout (Dhunge Dhara)	27.673638	85.32535	Ground level	Untreated	None	31.10.2018	Washing, Bathing, Gardening, Drinking	Yes
12	Kusunti, Lalitpur	Piped supply	27.664901	85.308355	Ground level	Treated	None	31.10.2018	Washing, Bathing, Gardening, Drinking	Yes
13	Sinamangal, Kathmandu	Jar Water	27.698192	85.346953	Ground level	Treated	Sealed Jar	31.10.2018	Drinking	Yes

**Table 1.** Water sampling locations, water types, and additional details about the water sources.

resolution when classifying sequences to taxa<sup>22,27</sup>. In this study, the presence of the genera *Escherichia* and *Salmonella* in the MOCK community was not identified from the Illumina data, and none of the sequencing reads from Illumina were classified to species level. In the context of water quality monitoring, this may then result in false negative outcomes at genus and species level for these putative pathogens. MinION sequencing has the advantage of a longer amplicon size, but the disadvantage of a reported higher error rate/lower accuracy than Illumina sequencing<sup>27</sup>. These MinION sequencing errors caused significant underreporting of the relative abundance of actual MOCK community members such as *Escherichia coli*. At the same time, assignment of partially erroneous reads to species by the MinION 16S bioinformatics workflow resulted in mismatched identities and taxonomic classification at species level, i.e. false positive results. While MinION sequencing detected all the species present in the MOCK community, these species were not always those that were most frequently detected within a certain genus. In particular, reads falsely classified as *Escherichia fergusonii* had a higher relative abundance than *Escherichia coli*. Phylogenetic analyses revealed that the 16S rRNA sequences for different strains of *E. coli*, *E. fergusonii*, and other strains from the genera *Shigella*, *Citrobacter*, and *Salmonella*, are closely related (Fig. S1a,b in supporting information). The *E. fergusonii* strains used in the NCBI reference database have 99.73% sequence similarity with an *E. coli* strain. Furthermore, studies have also shown that there is a lack of monophyly between *Shigella* and *Escherichia coli* and among *Shigella* taxonomic groups<sup>34,35</sup>. This makes it clear why the average MinION read accuracy (89% in our study) would make it difficult to reliably distinguish related species, such as those within the *Enterobacteriaceae*, that share a high sequence similarity. Recently developed methods such as Metagenomic phylogenetic analysis (MetaPhlAn) for species-level profiling in large scale microbial community studies<sup>36</sup> or Electronic probe Diagnostic for Nucleic acid Analysis (EDNA) to detect pathogens in metagenomics database<sup>37</sup> might overcome these challenges and contribute in reliably detecting putative pathogens at species level. In addition, both of these techniques use species-specific markers, and can decrease the probability of both false positive and false negative profiling of microorganisms. When considering the implications of these findings for water quality analysis, Illumina sequencing with its shorter read length may result in false negative results, i.e. *E. coli* is present in the sample, but not detected at genus and species level. MinION sequencing with its longer read lengths, but lower accuracy, may result in underreported abundances (i.e. *E. coli*), and also false positives (i.e. *E. fergusonii* and other closely related species). Although both NGS methods for 16S rRNA amplicon sequencing, with their associated bioinformatics platforms, have short-comings when it comes to reliably establishing

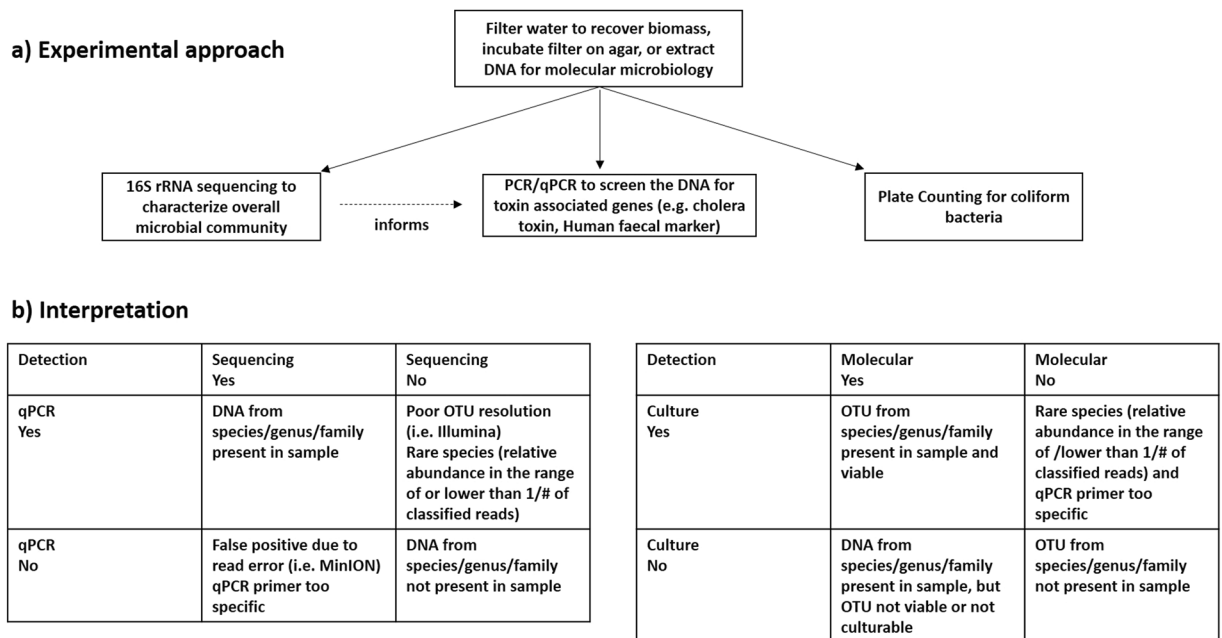


**Figure 4.** Molecular microbiology analysis of water samples from different water sources in the Kathmandu Valley. Numbers on the x-axis indicate the ID of the different sampling sites as described in Table 1. Data points for qPCR results are an average of technical triplicate samples and error bars indicate the standard deviation. Data points for Illumina NGS sequencing are an average of duplicate samples and error bars indicate the standard deviation. Library preparation for MinION NGS was unsuccessful for the sample from location 10 and thus data are not available (n.a), while the DNA extraction yield for the blank was too low for NGS, and for the sample from location 1, the yield was too low for all molecular microbiology methods.



**Figure 5.** Result of correlation analysis between different microbial water quality indicators determined with different methods: plate count, qPCR and NGS (A: Illumina and B: MinION). The colour intensities of circles are proportional to the correlation coefficients (Spearman rank correlation,  $p < 0.05$ ,  $n = 12$ ), and the circled numbers represents statistically significant coefficients.

identities at species level, they can provide initial insight into the likely microbial community composition, which is not readily available from other analytical methods. While species level identities need to be interpreted with great caution, more reliable information is obtained at genus and family level, and can guide further investigations of microbial water quality. A combination of NGS for screening, and qPCR methods for validation, can thus help identify false negative or false positive results. When using total 16S rRNA, total coliforms, total *E. coli*, human *E. coli* and *Vibrio cholerae* marker gene primers to analyse the MOCK community, the results were consistent with the true composition of the MOCK community. For example, detection of *E. coli* by qPCR would help identify one of the *Enterobacteriaceae* species in the MOCK community, which Illumina NGS had classified only to family level. The absence of the cholera toxin excretion protein gene by qPCR would flag up a few MinION NGS results



**Figure 6.** Outline of the experimental approach and methods (both traditional and DNA based) used for microbial water quality analysis, and the interpretation of water quality data obtained with those methods. Sample contamination due to non-sterile working practice may also explain discrepancies between different methodological approaches.

mis-classified as *Vibrio cholerae* as potential false positives. For the water samples from the Kathmandu Valley, the Spearman rank correlations between various methods to target FIO were generally positive and significant in this study, in line with previous findings<sup>38</sup>. However, there were discrepancies between thermo-tolerant coliform bacteria by the plate count method, and those analysed with qPCR for several of the disinfected water samples. These discrepancies suggest that coliform bacteria were inactivated by the disinfection, but inactivated cells and/or their DNA were still present in the water. For example, DNA based methods such as qPCR and NGS (both Illumina and MinION) will not differentiate between viable and dead bacterial cells, or extracellular DNA, and therefore may have poor comparability with culture based techniques<sup>18,31</sup>, especially in disinfected waters. In such instances, the combined information of DNA based and culture based methods adds meaningful value to the water quality survey, because it suggests that some water samples were fit for consumption due to treatment (i.e. disinfection by the bottling company), but that the water sources were influenced by faecal contamination. Nucleic acid extraction methods incorporating PMA (propidium monoazide) can distinguish between active/live and dead bacteria, and may provide an avenue for the reliable detection of viable pathogens in future work<sup>39</sup>. Other minor discrepancies between various methods may also be due to method limitations (Fig. 6). The detection of bacteria from water with culture independent methods requires a sufficient amount of target DNA, and effective DNA extraction. However, rare species may not always be present in 100 mL of sample, and DNA extraction efficiency is not generally 100% and therefore can cause loss of target bacteria which are present in low amounts in the sampled water<sup>19,40</sup>. In addition, primers used for molecular methods can be too specific, and therefore some of the bacteria that can grow on the plate may still be missed with such methods. While a qPCR method can in theory amplify and detect a single target gene copy, samples with a low amount of target can show high variability and fall outside the linear range of the qPCR standard curve<sup>41</sup>. Despite of the several caveats discussed above, the methods used in this study are complementary, and if used in combination, they enable a fuller understanding the potential risks associated with different water sources than any method on its own (Fig. 6). For example, *Vibrio cholerae* MinION data from the Kathmandu Valley were well aligned with qPCR results targeting a marker gene for pathogenic *Vibrio cholerae* strains. These genes were generally detected only in the samples from tube wells (location 6, 7 and 8), which were also the only samples for which NGS with the MinION matched some reads to *Vibrio cholerae* at species level. The combined data gives a much higher level of confidence in the presence of potentially pathogenic *Vibrio cholerae* bacteria in these water samples, than NGS data on its own. The *Vibrio cholerae* hazard was not detected with Illumina NGS, and *Vibrio cholerae* MinION and *Vibrio cholerae* qPCR data were negatively correlated with the faecal *E. coli* plate count. Screening with MinION NGS can flag up such hazards and guide the choice of subsequent testing protocols, such as qPCR, to verify the hazard. In addition to the method development and cross-comparison we also wanted to demonstrate their applicability to comprehensive water quality monitoring in informal settlements. A discussion of chemical water quality parameters is provided as supporting information. WHO and Nepalese drinking water quality guidelines state that water intended for drinking or in the distribution system for drinking purposes should not contain faecal coliforms in a 100 mL sample<sup>10,42</sup>. The water from four locations in the Kathmandu Valley (one community tube well (location 8), spring water (location 9), stone spout (location 11) and one piped supply (location 12)), out of thirteen locations, was contaminated with culturable faecal coliforms, and

therefore was not fit for consumption according to standard methods. The water from a spring (location 9) had the highest faecal *E. coli* concentration. These observations raise concerns, because springs are commonly used as a water source, also for the production of bottled or jar water and even tap water. Molecular methods also identified DNA from faecal bacteria in treated water samples (e.g., bacteria from coliform groups were detected in the water from locations 3, 12 and 13 when quantified with qPCR). 16S rRNA gene copy numbers were high in piped water from location 12, indicating a high amount of bacterial DNA in the tap water. Indicator gene copy numbers for the total coliforms were relatively higher as compared to other locations in the water sampled from location 9 (spring water), 10 (delivery truck) and 12 (piped water). Indicator genes for human *E. coli* pollution were only detected in water sampled from location 9 (spring water) and 12 (piped water). The molecular data raises strong concerns about the quality of spring water as a drinking water source, while total coliforms detected by plate count in one sample from the piped supply (location 12, also positive for faecal coliforms by plate count), and delivery truck water (location 10) raises strong concerns about the effectiveness of water treatment, as previously reported by other studies<sup>2,6,8</sup>. Clearly, water treatment needs to be more robust, if water sources such as springs have very poor microbial quality. The MinION NGS and qPCR *Vibrio cholerae* data also raises serious concerns about the safety of water from shallow tube wells (location 6, 7 and 8). Knowledge of common waterborne disease-causing agents prevailing in the sampling region (Table S11 in supporting information) can provide additional context for the interpretation of microbial water quality data. None of the MinION NGS reads from locations 2, 3 and 4 matched pathogens known to cause disease in the Kathmandu Valley, while in all the other locations, at least one such pathogen was present in the water samples according to MinION NGS results. In the water sample from location 6 (tube well), putative pathogens like *Campylobacter*, *Clostridium*, *Salmonella*, *Shigella* and *Vibrio cholerae* were detected, while *Clostridium botulinum*, *Escherichia coli*, *Salmonella enterica* and *Shigella* were detected in the piped water sample from location 12. A significant number of reads were matched by FASTQ 16S workflow of ONT at species level to bacterial agents mentioned in the approved list of biological agents by Health and Safety Executive (HSE), UK<sup>43</sup>. While the majority of OTUs in Illumina data were not resolved at species level, the pattern for the relative abundance of *Pseudomonas* and *Erysipelothrix* across sampling locations was consistent with the MinION NGS data for putative pathogen species within these genera, showing an unusually high number of MinION reads matched to *Pseudomonas aeruginosa* in location 4, and *Erysipelothrix rhusiopathiae* in location 6. The relative abundance of MinION reads matched to putative pathogens on the HSE list was on average higher in water from tube wells compared to other water types. As mentioned earlier, due to the risk of false positive identifications at species level with the MinION NGS method, additional tests using qPCR and culturing techniques (to check for viability) would be required to confirm the presence of these putative pathogens in the water samples. MinION NGS data is helpful in providing direction for such future work. With careful interpretation and cross-validation of results, NGS data adds significant value to water quality assessments, as demonstrated in this survey. As a low-cost, portable NGS tool generating long sequencing reads, the MinION of ONT is especially promising. The significant rank correlations between the relative abundances of *Bacteroides*, *Prevotella*, *Enterobacteriaceae* and all other putative pathogenic genera determined by MinION and Illumina NGS suggest that the portable, memory-stick sized MinION provides a valid alternative to the Illumina platform, which is that currently most widely used for NGS community-based microbial water quality monitoring and environmental surveillance<sup>22,25</sup>. However, further improvements of protocols, nanopore technology, and bioinformatics workflows are needed to improve read accuracy and avoid false assignment of reads to species. For instance, *Calus et al.* have recently developed a workflow for MinION sequencing to estimate the diversity of MOCK communities with average sequence accuracy of 99.5%. While their library preparation protocol using rolling circle amplification and enhanced bioinformatics dramatically improved the accuracy of MinION NGS, the total number of accepted reads generated per sequencing run was reduced accordingly<sup>44</sup>. The MinION NGS has already been applied in the field of public and animal health, and water security to identify specific pathogenic strains<sup>24,45–48</sup>. *Theuns et al.* successfully used MinION NGS as a diagnostic tool for porcine viral enteric disease and revealed porcine kobuvirus as the main enteric disease causing virus in swine<sup>45</sup>. *Rames and Macdonald* were able to detect Enteroviruses (EV) in wastewater (WW) samples with MinION sequencing, after spiking EV into the WW prior to sequencing<sup>46</sup>.

## Methods

**Water sampling locations.** Water samples in this study were collected from the most common water sources used by residents in the Kathmandu Valley. These comprised shallow tube wells, deep boreholes, stone spouts (dhunge dhara, a traditional ornamental spring), piped water, water supplied by delivery truck, bottled drinking water, and jar water (i.e. commercially available large containers of water used for drinking and cooking). Water samples were collected from 13 sites across the Kathmandu Valley. The locations of these water sources are presented in Fig. 3 and described in Table 1.

**Collection of water samples for analysis.** From each of the sources, 3 litres of water were collected in sterile 1 litre bottles. During sampling, the lid of the sterile bottle was opened and closed aseptically, and bottles were rinsed thoroughly with water from the same source before sample collection. From the piped water supply and stone spout, water was allowed to flow directly into the sterile bottles. For tube wells and boreholes, the water was pumped using a vacuum pump and polyethylene pipe. Sufficient purging was performed to prevent the collection of standing water already present in the pipe. Prior to collecting the water, the outlet of the tube was sterilized using 70% ethanol and flamed with a burning cotton swab, and water was then allowed to flow directly into the sterile bottle. The bottled drinking water and jar water were bought from local vendors. The sample bottles were stored in an insulated cold-box with ice packs inside and transported to the laboratory to be processed within two hours.



**Microbial water quality analysis.** Total coliform and faecal *E. coli* bacteria were determined by membrane filtration in duplicate at Nobel College (NC), Kathmandu, Nepal, following Standard Methods for the Examination of Water and Wastewater<sup>49</sup>. Different volumes of water (250 mL up to 2 L) were filtered through 0.22 µm membranes (Sartorius UK Limited, Surrey, UK) depending on the turbidity of the water from each sampling site and were immediately stored at -20 °C to preserve DNA for subsequent molecular microbiology. The molecular work was then conducted at Newcastle University (NU), Newcastle upon Tyne, UK. The total DNA from prokaryotic biomass retained by the membrane was extracted using a PowerWater® DNA Isolation Kit as per the manufacturer's instructions (QIAGEN, Crawley, UK). DNA purity and concentration were determined using a DS-11 FX + Spectrophotometer/Fluorometer (DeNovix, Delaware, USA). 30 ng of DNA was used to build a 16S rRNA prokaryote gene sequencing library for nanopore sequencing using a 16S Barcoding kit (SQK-RAB204 from Oxford Nanopore Technologies (ONT), Oxford, UK) as per the manufacturer's instructions and loaded onto a MinION sequencing apparatus flow cell (R9.4.1, FLO-MIN106). Table S12 in supporting information lists the primers used for 16S rRNA amplicon sequencing with the MinION. The flow cell was placed into the MinION sequencing device and controlled using ONT's MinKNOW software. The sequencing run was performed for 48 hrs. The raw reads (i.e. HDF5 raw signals) were base-called (i.e. converting the electrical signals generated by a DNA or RNA strand passing through the nanopore into the corresponding base sequence of the strand) with Albacore (Version; v2.3.3) software (ONT, Oxford, UK) producing.fastq files. Base-called data were uploaded to the EPI2ME interface, a platform for cloud based analysis of MinION data. Data interpretation was performed with the FASTQ 16S workflow (for quality filtering, a quality score  $\geq 7$  was used). The FASTQ 16S workflow [rev.2.1.1] analysis revealed the taxonomic classification of base-called reads along with their frequency, which ultimately was used to estimate the relative abundance of the putative human pathogens mentioned in the review by the UK Health and Safety Executive (HSE 2013).

Further, the extracted DNA were also sequenced (paired end sequencing; 2 × 250 bp) in duplicate with an Illumina Miseq platform (NU-OMICS, Northumbria University, UK) using the primer set targeting the V4 region of the bacterial 16S rRNA (Table S12) as described elsewhere<sup>50</sup>. The amplicon data from Illumina were processed using an open source software package: Quantitative Insights Into Microbial Ecology, QIIME 2 (<https://qiime2.org/>). Denoising and dereplication of pair end sequencing, including chimera removal and trimming of reads based on positional quality scores, were performed using the Divisive Amplicon Denoising Algorithm 2 (DADA2)<sup>51</sup>. The VSEARCH clustering method was used to cluster the quality-filtered sequences into ASVs (amplicon sequencing variants) that were converted into OTUs (operation taxonomic units), with a threshold of 97% identity<sup>52</sup>. Finally, taxonomy of each OTU was assigned by matching to the GreenGenes database (v13\_8), based on a naïve Bayesian classifier with default parameters.

In order to assess bias and errors in both NGS methods, a MOCK community (i.e. DNA mixture of known bacterial species in fixed proportion, see Table S1) provided by Zymo Research (Catalogue number: D6306), Freiburg, Germany, was also included in the sequencing runs in triplicate, and processed and analysed in the same way as the samples. To assist interpretation of NGS results for the MOCK community, we constructed a phylogenetic tree and identity matrix using the 16S rRNA gene sequences for different strains of *E. coli*, *E. fergusonii*, closely related other strains from the genera *Shigella*, *Citrobacter*, and *Salmonella*, and a more distant strain (*Enterococcus faecalis*) recorded in the NCBI reference database (Fig. S1a,b). It should be noted that the FASTQ 16S workflow in EPI2ME uses the NCBI 16S rRNA database as a reference database for taxonomic classification.

Real time PCR assays (qPCR) were performed to quantify the number of target genes on a BioRad CFX C1000 system (BioRad, Hercules, CA USA) using the primers shown in Table S12. For quantification of the target genes, 2 µl template DNA was used in a reaction mixture containing 5 µL 2 × SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), 500 nmol L<sup>-1</sup> of each forward and reverse primer, and molecular grade H<sub>2</sub>O (Invitrogen, Life Technologies, Paisley, UK) to a final volume of 10 µL. Reaction conditions for quantification of each target gene were 98 °C for 3 min (1x), then 98 °C for 15 s, and the Primer Annealing Temperature (Ta) for 60 s (Table S12) (40 cycles). Standard curves were constructed using the synthesized nucleotide sequence of the target gene (Invitrogen, Life Technologies, Paisley, UK), and generated every time a qPCR analysis was performed, in parallel with the amplification of test samples. Serial dilution (10-fold) of the standards was performed to obtain standard solutions in the range of 10<sup>8</sup>–10<sup>1</sup> target gene copies/µL. All samples were run in triplicate and molecular grade H<sub>2</sub>O replaced template in control reactions. To avoid inhibitor effects, DNA samples were diluted to a working solution of 5 ng/µL.

**Chemical water quality analysis.** Chemical water quality analysis methods are described in supporting information.

**Statistical analysis.** A Spearman rank correlation analysis between different microbial water quality indicators determined by standard plate counting method, qPCR and next generation sequencing (NGS) approaches was performed in R-studio.

### Data availability

Data created during this research are openly available at DOI 10.25405/data.ncl.9693533. Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for access instructions.

Received: 17 April 2019; Accepted: 10 October 2019;

Published online: 31 October 2019

## References

1. UN. *United Nations Sustainable Development Goals*, <https://sustainabledevelopment.un.org/> (2019).
2. Warner, N. R., Levy, J., Harpp, K. & Farruggia, F. Drinking water quality in Nepal's Kathmandu Valley: a survey and assessment of selected controlling site characteristics. *Hydrogeology Journal* **16**, 321–334 (2008).
3. Shrestha, S., Haramoto, E., Malla, R. & Nishida, K. Risk of diarrhoea from shallow groundwater contaminated with enteropathogens in the Kathmandu Valley, Nepal. *Journal of water and health* **13**, 259–269 (2015).
4. Subedi, M. & Aryal, M. Public perception about drinking jar water and its bacteriological analysis. *Nepal Med Coll J* **12**, 110–114 (2010).
5. Silvanus, V., Gupta, R. K. & Shreshta, S. R. Assessment of Water Supply and Microbial Quality of Water Among Schools in The Rural Kathmandu Valley, Nepal. *Nepal Med Coll J* **18**, 44–47 (2016).
6. Karkey, A. *et al.* The ecological dynamics of fecal contamination and Salmonella Typhi and Salmonella Paratyphi A in municipal Kathmandu drinking water. *PLoS neglected tropical diseases* **10**, e0004346 (2016).
7. Udmale, P., Ishidaira, H., Thapa, B. R. & Shakya, N. M. The status of domestic water demand: Supply deficit in the Kathmandu Valley, Nepal. *Water* **8**, 196 (2016).
8. Prasai, T., Lekhak, B., Joshi, D. R. & Baral, M. P. Microbiological analysis of drinking water of Kathmandu Valley. *Scientific World* **5**, 112–114 (2007).
9. APHA. *Standard Methods for the Examination of Water and Wastewater*, American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC (2015).
10. WHO. *Guidelines for drinking-water quality: fourth edition incorporating the first addendum*. Geneva: World Health Organization (2017).
11. EU. Directive 2006/7/EC of the European Parliament and of the Council of 16 February 2006 Concerning the Management of Bathing Water Quality and Repealing Directive 76/160/EEC. *Official Journal of the European Union L64*, pp. 37–51 (2006).
12. WHO. WHO recommendations on scientific, analytical and epidemiological developments relevant to the parameters for bathing water quality in the Bathing Water Directive (2006/7/EC). (2018).
13. Gruber, J. S., Ercumen, A. & Colford, J. M. Jr. Coliform bacteria as indicators of diarrheal risk in household drinking water: systematic review and meta-analysis. *PLoS one* **9**, e107429 (2014).
14. 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 microbiology reviews* **38**, 1–40 (2014).
15. Field, K. G. & Samadpour, M. Fecal source tracking, the indicator paradigm, and managing water quality. *Water research* **41**, 3517–3538 (2007).
16. Ongley, E. *Water Quality Monitoring-A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes*. *United Nations Environment Programme and the World Health Organization* (1996).
17. Bartram, J., Ballance, R. & World Health, O. *Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programs* (1996).
18. Gensberger, E. T. *et al.* Evaluation of quantitative PCR combined with PMA treatment for molecular assessment of microbial water quality. *Water research* **67**, 367–376 (2014).
19. Brettar, I. & Höfle, M. G. Molecular assessment of bacterial pathogens—a contribution to drinking water safety. *Current Opinion in Biotechnology* **19**, 274–280 (2008).
20. Schang, C. *et al.* Evaluation of techniques for measuring microbial hazards in bathing waters: A comparative study. *PLoS one* **11**, e0155848 (2016).
21. Vignola, M., Werner, D., Wade, M. J., Meynet, P. & Davenport, R. J. Medium shapes the microbial community of water filters with implications for effluent quality. *Water research* **129**, 499–508 (2018).
22. Tan, B. *et al.* Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. *Frontiers in microbiology* **6**, 1027 (2015).
23. Noble, R. T. & Weisberg, S. B. A review of technologies for rapid detection of bacteria in recreational waters. *Journal of water and health* **3**, 381–392 (2005).
24. Hu, Y. O. O. *et al.* Stationary and portable sequencing-based approaches for tracing wastewater contamination in urban stormwater systems. *Scientific reports* **8**, 11907 (2018).
25. Cai, L. & Zhang, T. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environmental science & technology* **47**, 5433–5441 (2013).
26. Lu, X. *et al.* Bacterial pathogens and community composition in advanced sewage treatment systems revealed by metagenomics analysis based on high-throughput sequencing. *PLoS One* **10**, e0125549 (2015).
27. Benítez-Páez, A., Portune, K. J. & Sanz, Y. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. *GigaScience* **5**, 4 (2016).
28. Lan, Y., Rosen, G. & Hershberg, R. Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. *Microbiome* **4**, 18 (2016).
29. Cui, Q., Fang, T., Huang, Y., Dong, P. & Wang, H. Evaluation of bacterial pathogen diversity, abundance and health risks in urban recreational water by amplicon next-generation sequencing and quantitative PCR. *Journal of Environmental Sciences* **57**, 137–149 (2017).
30. Ahmed, W. *et al.* Toolbox approaches using molecular markers and 16S rRNA gene amplicon data sets for identification of fecal pollution in surface water. *Appl. Environ. Microbiol.* **81**, 7067–7077 (2015).
31. Batista, A. M. M. *et al.* Microbiological safety of a small water distribution system: evaluating potentially pathogenic bacteria using advanced sequencing techniques. *Water Science and Technology: Water Supply* **18**, 391–398 (2018).
32. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
33. Neave, M. *et al.* Multiple approaches to microbial source tracking in tropical northern Australia. *MicrobiologyOpen* **3**, 860–874 (2014).
34. Pettengill, E. A., Pettengill, J. B. & Binet, R. Phylogenetic analyses of Shigella and enteroinvasive Escherichia coli for the identification of molecular epidemiological markers: whole-genome comparative analysis does not support distinct genera designation. *Frontiers in microbiology* **6**, 1573 (2016).
35. Zuo, G., Xu, Z. & Hao, B. Shigella strains are not clones of Escherichia coli but sister species in the genus Escherichia. *Genomics, proteomics & bioinformatics* **11**, 61–65 (2013).
36. Truong, D. T. *et al.* MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature methods* **12**, 902 (2015).
37. Espindola, A. S. *et al.* Inferring the presence of aflatoxin-producing Aspergillus flavus strains using RNA sequencing and electronic probes as a transcriptomic screening tool. *PLoS one* **13**, e0198575 (2018).
38. Varma, M. *et al.* Quantitative real-time PCR analysis of total and propidium monoazide-resistant fecal indicator bacteria in wastewater. *Water Research* **43**, 4790–4801 (2009).

39. Pang, Y.-C., Xi, J.-Y., Xu, Y., Huo, Z.-Y. & Hu, H.-Y. Shifts of live bacterial community in secondary effluent by chlorine disinfection revealed by Miseq high-throughput sequencing combined with propidium monoazide treatment. *Applied microbiology and biotechnology* **100**, 6435–6446 (2016).
40. Agudel, R. M. *et al.* Monitoring bacterial faecal contamination in waters using multiplex real-time PCR assay for *Bacteroides* spp. and faecal enterococci. *Water SA* **36** (2010).
41. Forootan, A. *et al.* Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomolecular detection and quantification* **12**, 1–6 (2017).
42. DWSS. National Drinking Water Quality Standards. Kathmandu: Department of Water Supply and Sewerage (2005).
43. HSE. The Approved List of biological agents. Third edition. Merseyside: Health and Safety Executive (2013).
44. Calus, S. T., Ijaz, U. Z. & Pinto, A. J. NanoAmpli-Seq: A workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. *GigaScience* **7**, giy140 (2018).
45. Theuns, S. *et al.* Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobuvirus as an important enteric virus. *Scientific reports* **8**, 9830 (2018).
46. Rames, E. & Macdonald, J. Evaluation of MinION nanopore sequencing for rapid enterovirus genotyping. *Virus research* **252**, 8–12 (2018).
47. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *nature protocols* **12**, 1261 (2017).
48. Tyler, A. D. *et al.* Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific reports* **8**, 10931 (2018).
49. APHA. *Standard Methods for the Examination of Water and Wastewater*. 21st edn, (American Public Health Association/American Water Works Association/Water Environment Federation, 2015).
50. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology* **79**, 5112, <https://doi.org/10.1128/AEM.01043-13> (2013).
51. Callahan, B. J. *et al.* DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods* **13**, 581 (2016).
52. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584–e2584, <https://doi.org/10.7717/peerj.2584> (2016).

## Acknowledgements

This research was funded by the Engineering and Physical Sciences Research Council, EPSRC (Grant number: EP/P028527/1). We would like to acknowledge the Department of Medical Microbiology, Nobel College, Pokhara University, Kathmandu, Nepal for providing the laboratory facilities. The authors also thank Mr. Tom Komar from the School of Engineering, Newcastle University, who helped us to prepare an excel template to extract putative pathogens from NGS data.

## Author contributions

K.A. initiated and designed the fieldwork, conducted data collection and analysis, and wrote the manuscript. S.K. participated in the field work and conducted sample analysis. K.P. helped in planning the field work, participated in the field work and conducted sample analysis. N.A. helped in planning the field work, participated in this field work and conducted sample analysis. R.J.D. provided suggestions in preparing the manuscript. D.W. conceived and raised funding for the work, helped with the data analysis, and provided suggestions in preparing the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-51997-x>.

**Correspondence** and requests for materials should be addressed to D.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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