Tap water as a natural vehicle for microorganisms shaping the human gut microbiome

Gabriele Andrea Lugli, Giulia Longhi, Leonardo Mancabelli, Giulia Alessandri, Chiara Tarracchini, Federico Fontana, Christian Milani, and Marco Ventura

Laboratory of Probiogenomics, Department of Chemistry, Life Sciences, and Environmental Sustainability, University of Parma, Parma, Italy.
GenProbio Srl, Parma, Italy.
Microbiome Research Hub, University of Parma, Parma, Italy.
APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland.

Summary

Fresh potable water is an indispensable drink which humans consume daily in substantial amounts. Nonetheless, very little is known about the composition of the microbial community inhabiting drinking water or its impact on our gut microbiota. In the current study, an exhaustive shotgun metagenomics analysis of the tap water microbiome highlighted the occurrence of a highly genetic biodiversity of the microbial communities residing in fresh water and the existence of a conserved core tap water microbiota largely represented by novel microbial species, representing microbial dark matter. Furthermore, genome reconstruction of this microbial dark matter from water samples unveiled homologous sequences present in the faecal microbiome of humans from various geographical locations. Accordingly, investigation of the faecal microbiota content of a subject that daily consumed tap water for 3 years provides proof for horizontal transmission and colonization of water bacteria in the human gut.

Introduction

Freshwater is estimated to represent about 2.5% of all water on Earth, while the remainder constitutes saltwater from seas and oceans. In developed countries, potable water is readily accessible as tap water and bottled natural mineral water, both of which are subject to strict safety regulations and very regular inspections (Eichler et al., 2006). Conventional drinking water treatment plants perform filtration, sedimentation, disinfection and flocculation, thereby allowing assessment of the microbial load (Dodd, 2012; Chao et al., 2013; Loubet et al., 2016). Nonetheless, some microorganisms may persist and proliferate in drinking water, with bacterial concentrations estimated to be around $10^6–10^8$ cells per litre (Hammes et al., 2008; Hong et al., 2010). Thus, like other consumed items, water may represent a natural vehicle of microorganisms able to interact with the human gut and its microbiota (Dimidi et al., 2019; Milani et al., 2019).

Some studies have observed significant associations between tap water composition and human health (Bouchard et al., 2011). In this context, differences were observed in the composition of the gut microbiota of mice that drank water from different sources, including tap water, highlighting an increase of clinically important taxa such as *Acinetobacter* and *Staphylococcus* in the faeces and mucosa-adhered samples of animals (Dias et al., 2018). Similar examples have also been reported in a human population context, given the recent evidence of bacterial spread from microbial biofilms present in drinking water distribution systems (Chan et al., 2019) and investigations regarding the potential for tap water to influence human health mediated through the gut microbiota (Bowyer et al., 2020).

Microbial populations in drinking water are challenging to assess because most of the bacteria present appear to be non-culturable (Loy et al., 2005; França et al., 2015) or are present in a viable-but-non-culturable state (Szewzyk et al., 2000). For this reason, metagenomic sequencing approaches represent a well-established method to study the culturable and unculturable parts of the drinking water microbiota (Brumfield et al., 2020; Sala-Comorera et al., 2020). Furthermore, recent studies provide exciting insights into the
composition of tap water microbiota by revealing that increased urban development causes shifts in bacterial community composition of water (Simonin et al., 2019). Thus, microbial community analyses can be helpful to diagnose environmental conditions as indicators of ecosystem health and water quality condition. In this context, the predominant bacteria that have been detected in drinking water are members of the phyla Actinobacteria and Proteobacteria, with the genera *Afilia*, *Bradyrhizobium* and *Mycobacterium* being prevalent among tap water and drinking fountain samples (Brumfield et al., 2020).

In the current study, we investigated the microbiota composition of tap water from the city pipeline of Parma, Italy, using a shallow metagenomics approach, which allows accurate taxonomic profiling of microbial communities from water samples down to species level. In addition, the metagenomics data were used in conjunction with data sets retrieved from the NCBI repository to describe the occurrence of a core tap water microbiota as well as the existence of specific microbial groups that are typical of the various geographical regions investigated. Within this context, we mapped the presence of a microbial taxon shared between the tap water microbiota and the gut microbiota of an individual daily drinking that water, suggesting horizontal transmission by these bacteria, which in turn appears to impact other resident members of the human gut microbiota.

**Results and discussion**

*Uncovering tap water microbial biodiversity*

Sixteen water samples, named W001 to W016, were collected from public fountains (*n* = 12) and household taps (*n* = 4) throughout the Parma district in Italy to explore the microbial biodiversity of water samples across the city (Table 1). Shallow shotgun metagenomic sequencing was performed to identify the microorganisms that populate water samples at species level. Sequencing output constituted about one million paired-end reads, with an average of 64 thousand reads per sample (Table S1), thus allowing accurate assessment of the microbial species inhabiting each water sample (Hillmann et al., 2018). This taxonomic survey revealed that just five species were present in each water sample at a relative abundance higher than 10%, i.e. *Acidovorax delafieldii* (55% in W012), *Aqua bacterium commune* (26% in W005), *Sphingomonas ursincola* (23% in W011), *Sphingobium fluviale* (14% in W015) and *Sphingomonas aerolata* (12% in W009) (Fig. 1). As reported in the literature, these species had previously been identified in biofilms of drinking water samples (Kalmbach et al., 1999; Busse et al., 2003; Morohoshi et al., 2017). Moreover, unclassified members of 13 additional bacterial genera were identified with relative abundances ranging from 10% to 39%, highlighting the marked presence of as yet to be isolated and characterized bacterial species (Fig. 1). Interestingly, the major difference identified between samples collected from public fountains and household taps was represented by the presence of *Sphingomonas*, found as the most abundant taxa in nine out of 12 public fountains water samples (Table 1).

A flow cytometry (FC) assay was employed to enumerate microbial cells present in each water sample, thereby providing information on the absolute number of microbes. Based on the predicted number of colony-forming units (CFU), analysed samples ranged from $2 \times 10^5$ CFU ml$^{-1}$ in W009 to $5.3 \times 10^6$ CFU ml$^{-1}$ in W001 (Fig. S1). Interestingly, when excluding sample W001, the average CFU ml$^{-1}$ between water samples was $2.8 \times 10^5$, highlighting W001 as an outlier of the analysis with a high level of bacterial cells probably due to downstream effects of the water distribution system (Fig. S1). Then, FC data obtained for each water sample were used to normalize the taxonomically classified reads obtained through shallow shotgun metagenomics according to a previously described method (Lugli et al., 2020), thereby allowing an estimation of absolute abundance (Fig. S1). The findings showed a high CFU ml$^{-1}$ number of *Bradyrhizobium* spp. and *Afilia* spp. present in sample W001, and *Rothisia* spp. representing a microorganism present at high absolute abundance in sample W003, highlighting bacterial taxa which had also previously been identified in drinking water systems (Brumfield et al., 2020).

Furthermore, taxonomic profiling at species level allowed us to identify the core taxonomic elements of the water microbiome, i.e. taxa that occur with the highest prevalence. Notably, in addition to *Sphingomonas ursincola* and *Blastomonas fulva*, which were determined to be the most prevalent species being detected in 13 out of 16 samples, taxonomic profiling also revealed DNA of unknown bacterial species attributable to 30 genera distributed between samples at a prevalence of >70%. Remarkably, these later findings underscore that many members of the microbial drinking water population have as yet not been subject to any characterization (Fig. 1). Among the latter microbial dark matter, unknown members of six genera were identified in all 16 samples, i.e. *Bradyrhizobium*, *Novosphingobium*, *Pseudomonas*, *Sphingobium*, *Sphingomonas* and *Sphingopyxis*. Intriguingly, while *Pseudomonas* is a ubiquitous environmental bacterial genus, members of the other five genera have previously been isolated from groundwater and drinking water (McAlister et al., 2002; Yoon et al., 2005; Sheu et al., 2013; Singh et al., 2015; Gulati and Ghosh, 2017). Altogether, these metagenomic data sets highlight the substantial knowledge gap pertaining to drinking water-associated bacteria.
To validate the quality of our metagenomic analysis and compare the taxonomic profiles of water samples identified in the current study with those of other geographical locations, sequencing data of 119 drinking water samples were retrieved from nine metagenomic Table 1.

| Sample name | Sample type                          | Sampling day | Most abundant species                     | SRA            | Bioproject      |
|-------------|--------------------------------------|--------------|------------------------------------------|----------------|-----------------|
| W001        | Tap water                            | 22/06/2021   | Bradyrhizobium sp.                       | SRR18015330    | PRJNA086724     |
| W002        | Public drinking fountain water       | 05/05/2021   | Sphingomonas sp.                         | SRR18015329    | PRJNA086724     |
| W003        | Public drinking fountain water       | 05/05/2021   | Rothia sp.                               | SRR18015322    | PRJNA086724     |
| W004        | Tap water                            | 28/05/2021   | Novosphingobium sp.                     | SRR18015321    | PRJNA086724     |
| W005        | Tap water                            | 06/05/2021   | Aquabacterium commune                   | SRR18015320    | PRJNA086724     |
| W006        | Tap water                            | 23/04/2021   | Nitrospira sp.                           | SRR18015318    | PRJNA086724     |
| W007        | Public drinking fountain water       | 10/05/2021   | Sphingomonas sp.                         | SRR18015317    | PRJNA086724     |
| W008        | Public drinking fountain water       | 07/05/2021   | Sphingomonas sp.                         | SRR18015316    | PRJNA086724     |
| W009        | Public drinking fountain water       | 17/05/2021   | Sphingomonas aerolata                   | SRR18015315    | PRJNA086724     |
| W010        | Public drinking fountain water       | 23/04/2021   | Sphingomonas sp.                         | SRR18015314    | PRJNA086724     |
| W011        | Public drinking fountain water       | 17/05/2021   | Erityrobacter sp.                        | SRR18015313    | PRJNA086724     |
| W012        | Public drinking fountain water       | 06/05/2021   | Acidovorax delafeldii                   | SRR18015312    | PRJNA086724     |
| W013        | Public drinking fountain water       | 04/05/2021   | Sphingomonas sp.                         | SRR18015311    | PRJNA086724     |
| W014        | Public drinking fountain water       | 04/05/2021   | Sphingomonas sp.                         | SRR18015310    | PRJNA086724     |
| W015        | Public drinking fountain water       | 23/04/2021   | Sphingomonas sp.                         | SRR18015309    | PRJNA086724     |
| W016        | Public drinking fountain water       | 07/05/2021   | Sphingomonas ursincola                  | SRR18015308    | PRJNA086724     |

Fig. 1. Microbial composition of 16 drinking water samples collected from different locations of the Parma District and delivered by the city water supply system.

Panel A displays a histogram with the relative abundance of each microbial species identified in the analysed samples. From left to right, the colour-coding order of the legend reflects average abundances from largest to smallest, i.e. 55% Acidovorax delafeldii in sample W012, 38.9% Sphingomonas spp. in sample W013, and so on.

Panel B shows the percentage of unknown and known bacterial species profiled, while Panel C depicts the prevalent bacterial taxa identified among all processed samples.

Meta-analysis of the water microbiome across the world

To validate the quality of our metagenomic analysis and compare the taxonomic profiles of water samples identified in the current study with those of other geographical locations, sequencing data of 119 drinking water samples were retrieved from nine metagenomic...
projects, named P02 to P10, which cover seven different countries, (Table S2). Taxonomic profiling at species level was performed applying the same pipeline and parameters used for samples collected as part of our own study, as reported above. While three samples were discarded due to low sequencing data (<20,000 DNA sequence reads) (Table S1), along with the P04 sample that was not suitable for statistical purposes, the remaining 115 water samples were analysed together with the microbiome data of 16 samples retrieved in the Parma district.

Beta-diversity investigation represented through principal coordinate analysis (PCoA) based on Bray–Curtis dissimilarity index allowed exploration of the water biome biodiversity as based on different studies (Fig. 2). Marked biodiversity was encountered between almost all projects (PERMANOVA p-value of <0.05) except for P6 (Table S3). These data highlighted unique signatures among water sample microbiomes associated with the same country, possibly due to geographic and environmental factors such as temperature and pH. As similarly reported in P1 (Parma district project), an average of 16% of the microbial DNA of water was classified at species level, while the remaining 84% was attributable to unknown microbial taxa, thus representing a sizable part of the water microbial dark matter. Investigating the prevalence of bacterial taxa between samples of the same project, a limited number of taxa (between 13 and zero) were observed with a prevalence >80%, indicating a high degree of microbial biodiversity and inter-site variability (Fig. 2). Notably, we included two longitudinal studies (P3 and P8) and observed a higher number of high prevalence taxa (13 and 11 respectively), indicating lower inter-sample biodiversity, when considering samples collected longitudinally from the same site (Dai et al., 2018; Vosloo et al., 2021). For example, only Reyranella soli was identified with a high prevalence (83%) in P5 samples, while a similar observation was made for Blastomonas fulva and Sphingomonas ursincola in the case of P1 samples, which showed a prevalence of 81% for these species (Fig. 2).

Interestingly, 84% of unclassified microbial DNA identified among water samples, previously referred as microbial dark matter (Rinke et al., 2013), represented the actual core water microbiota (Table 2). In this context, following DNA filtering steps (see Experimental procedures), the microbial portion of a sample that was not classified at species level revealed unknown bacterial species highlighting the absence of a reference genome deposited in the NCBI repository. The five most prevalent microorganisms identified in the 115 drinking water samples have already been identified as major players in the 16 samples collected in the tap water of Parma district, i.e. members of the genera Bradyrhizobium, Sphingomonas, Pseudomonas, Novosphingobium and Sphingobium (Table 2). Additionally, the DNA of species belonging to Paraburkholderia, Burkholderia and Mesorhizobium was also identified in more than 90% of the profiled waters (Table 2).

Members of the Sphingomonas genera are Gram-negative bacteria isolated from many different land and water habitats thanks to their ability to survive at low nutrient concentrations. More recently, the genus has been subdivided into different genera, in which we can find two other prevalent microbial groups reported above, i.e. Novosphingobium and Sphingobium (Takeuchi et al., 2001). Conversely, members of the Bradyrhizobium and Mesorhizobium genera are Gram-negative nitrogen-fixing bacteria that occur either as free-living soil bacteria or that are found in symbiotic interaction within the roots of leguminous plants (Lorite et al., 2018). Similarly, members of the Paraburkholderia genus are also nitrogen-fixing bacteria correlated with plant growth promotion, while members of the related genus Burkholderia can also be pathogens for humans, being the case for several species of the Pseudomonas genus. However, since these latter microorganisms are ubiquitously distributed in drinking water, they are probably not harmful to humans. Nevertheless, the presence of opportunistic pathogens in rainwater and tap water storage systems has already been discussed, showing the natural occurrences of Pseudomonas aeruginosa, Legionella spp. and Mycobacterium spp. (Zhang et al., 2021). Thus, it would be of great interest to unveil the identified taxa genomic capability so as to assess and understand possible interactions with humans upon ingesting water containing such microbes.

Recently, high-throughput molecular analyses of microbiomes have been used as a tool to monitor the wellbeing of aquatic environments, involving cultural-independent analyses such as metagenomics, metatranscriptomics, metaproteomics and metabolomics (Michán et al., 2021). In the frame of this work, water microbiome profiling revealed a conserved microbial core represented by mostly unclassified and uncharacterized bacteria, which highlights a dearth of knowledge from a genomic and functionality perspective. Thus, microbiomes of waters should be investigated through culturomics experiments to gain access to such novel microbial species. In addition, since many microorganisms cannot be cultivated using standard procedures, deep metagenome sequencing can be performed to reconstruct unknown bacterial genomes.

Investigating the impact of tap water on the human microbiome

Recent literature has supported the notion that ingestion of foods populated by a specific microbiota facilitates

@2022 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology, 24, 3912–3923
transmission and subsequent colonization of these microorganisms in the human gut (Hehemann et al., 2010; Makki et al., 2018; Milani et al., 2019). Since every person consumes about 2 L of water per day, we explored the influence of water consumption on shaping the composition of the human gut microbiota. Therefore, bacterial DNA sequences collected from the 115 drinking water samples belonging to P02–P10 were subjected to genome reconstruction to gather chromosomal fragments of unknown bacterial taxa. In total, 2.9 gb of DNA sequences belonging to unknown bacteria were assembled in this manner. Nonetheless, we decided to investigate only those bacteria identified with the highest prevalence in water samples of P1 and the other projects (Table 2) represented by 622 mb of as yet unclassified members of the genera Bradyrhizobium, Burkholderia, Mesorhizobium, Novosphingobium, Paraburkholderia, Pseudomonas, Reyranella, Sphingobium, Sphingomonas and Sphingopyxis.

To trace the presence of the assembled microbial genomes, we included in this study shotgun metagenomic data of 196 human faecal samples retrieved from the NCBI repository (Table S2). Selection of faecal samples was performed so as to equally cover countries from which the drinking water was collected (Table S2). Then, the total amount of microbial DNA retrieved from the latter samples was used to investigate the presence of bacterial dark matter reconstructed through shotgun metagenomic assemblies. DNA mapping was performed with high sensitivity and high specificity (see Experimental procedures), identifying DNA of the reconstructed bacteria from the water microbiome in 46 of the analysed human faecal samples (Fig. 3). DNA sequences corresponding to unclassified bacteria were distributed across the three analysed continents, with a higher predominance of sequences belonging to Sphingomonas and Reyranella, both identified in 43 faecal samples, followed by Novosphingobium in 37 faecal samples. Following the latter microorganisms, the DNA of
Sphingopyxis, Pseudomonas and Bradyrhizobium was identified in 27–25 samples, while the remaining four genera were identified in 16–14 faecal samples (Fig. 3). Altogether, these data indicate that the DNA of microorganisms belonging to the core microbiota of drinking water samples can also be detected in human faecal samples. Overall, these findings suggest that such tap water-associated microorganisms contribute to the human gut microbiota composition. A further interesting aspect awaiting to be investigated will be how many of these water-associated microorganisms found in the gut of humans were transmitted through direct ingestion of water or to the extensive use of tap water in watering plants for consumption.

Transmission of microorganisms from tap water to the human gut

The in silico-based findings indicate the occurrence of microbial DNA belonging to the core tap water microbiota...
in various human faecal microbiomes. This prompted us to investigate this potential novel route for shaping the human gut microbiota. Thus, we examined the gut microbiota composition, together with the corresponding tap water microbiota, of a subject who daily consumed tap water for the past 3 years. Specifically, tap water consumed by the subject had been processed in the context of this work (corresponding to W001) and subjected to metagenome assembly of its microbial DNA content. Assembled bacterial DNA resulted in 104 kb distributed in 40 contigs, reflecting portions of chromosomes of those microorganisms present in relatively high abundance within the sample. Notably, assembled contigs, which taxonomically were predicted to belong to putative unknown species of the genus *Bradyrhizobiurn*, *Curvibacter* and *Sphingobium*, were used to design

© 2022 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd., *Environmental Microbiology*, 24, 3912–3923
primer pairs for quantitative real-time PCR (qPCR) investigation.

Faecal sample collection was accomplished through three time points aimed at investigating the microbiota of the subject during the consumption of W001 (T1 and T3) and after a period of washout of 2 weeks, during which the subject drank only bottled water (T2) (Fig. 3). Notably, bottled water microbiota was analysed, revealing a negligible amount of microbial DNA, thus corroborating the microbiological sterility of the administered water between T1 and T2. Due to the high CFU ml\(^{-1}\) identified in sample W001 at T1 (5.3 \(\times\) 10\(^6\)), an additional (FC) assay was performed at T3 of the same tap water highlighting a consistent quantification of microbial cells (4.8 \(\times\) 10\(^6\)). Remarkably, employing the *Bradyrhizobium* and *Sphingobium* strain-specific primers, qPCR assays on the faecal samples resulted in quantifying DNA below the detection limit. Instead, *Curvibacter* DNA was identified at each time point, indicating apparent colonization of this *Curvibacter* species in the subject’s gut as it was detected even after 2 weeks following the start of the washout period (Fig. 3). Notably, even if these data have been obtained from a single individual, they suggest that some microbial species residing in tap water can survive to the human colonic tract and colonize and persist in the gut of their human host. A clinical trial involving a substantial number of subjects will need to be executed to validate these findings.

**Conclusions**

Tap water is considered a food, providing essential elements to our body, which are vital for our lives. However, tap water can also be a reservoir of microorganisms that, once ingested, may colonize our intestine, influence the gut microbiota and be responsible for different metabolic activities associated with human health. In the current study, we were interested in assessing the notion that water is not only crucial for our nutritional and physiological requirements but may also be important as a delivery vehicle of microorganisms to the gut. Notably, and in contrast to most consumed foods, the microbial community composition of tap water is not very well studied (Sala-Comorera et al., 2020). Here, we clearly show that a large part of the microorganisms present in water is represented by as yet to be characterized bacteria, thus representing constituting a lot of microbial dark matter. These findings should prompt dedicated investigations on these bacteria, aimed at isolation, cultivation and subsequent dissection of their biological features. In fact, as outlined in our study, a large number of these putative novel bacterial taxa, which make up part of the core tap water microbiota, are also identified as part of the human faecal microbiota representing various different metagenomes and geographical regions. Thus, this co-sharing scenario of members of the core tap water microbiota and the human gut microbiota may impact human health through modulation of the gut microbiota. This may therefore represent an intriguing and novel scenario that warrants further careful exploration. Here, our findings indicate that horizontal transmission and subsequent colonization of microorganisms from tap water to the human gut is possible. However, a clinical trial encompassing a more extensive set of individuals drinking tap water encompassing different microbiota and involving the isolation of the microorganisms using culturomics approaches needs to be performed in order to corroborate our data.

**Experimental procedures**

**Tap water samples and sampling conditions**

Sixteen tap water samples, including public fountains and household taps, were randomly selected from different locations and distribution systems encompassing various parts of Parma town and its territory. To ensure that the collected samples are representative of consumed water, at least 5 L of water was directly collected from the tap, keeping at a safe distance from the faucet and letting some flow down before directly flushing the water into sterile bottles to minimize any contaminations. Water samples were transported to the laboratory and kept at 4°C for further analysis.

**Microbial DNA extraction**

For bacterial DNA extraction, 5 L of a given water sample was filtered through 0.45 μm pore size hydrophilic mixed cellulose esters (Pall Corporation, Port Washington, NY, USA). Filters were placed in standard Petri dishes and were cut into small pieces to ensure total sterility. DNA was extracted from the filters using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research, D4300) following the manufacturer’s instructions. Then, each tap water sample’s DNA concentration and purity were investigated by employing a Picodrop microtiter Spectrophotometer (Picodrop, Hinxton, UK).

**Shallow shotgun sequencing**

According to the manufacturer’s instructions, DNA library preparation was performed using the Nextera XT DNA sample preparation kit (Illumina, San Diego, CA, USA). First, 1 ng input DNA from each sample was used for the library preparation, which underwent fragmentation, adapter ligation and amplification. Then, Illumina libraries were pooled equimolarly, denatured and diluted to a
concentration of 1.5 pM. Next, DNA sequencing was performed on a MiSeq instrument (Illumina) using a 2× 250 bp Output sequencing Kit together with a deliberate spike-in of 1% PhiX control library.

**Short read taxonomic classification**

Sequenced paired-end reads of each water sample were subjected to a filtering step removing low-quality reads (minimum mean quality score 20, window size 5, quality threshold 25 and minimum length 100) using the fastq-mcf script (https://github.com/ExpressionAnalysis/ea-utils/blob/wiki/FastqMcf.md) to analyse high-quality sequenced data only. Then, an additional filtering step was performed to remove possible contaminating human DNA sequences from each sample through reads mapping employing the BWA aligner (Li and Durbin, 2009). Filtered reads were then collected and taxonomically classified through the METAnnotatorX2 pipeline (Milani et al., 2021), using a set of databases of reference genomes whose taxonomy was previously validated to maximize the accuracy of homology-based taxonomic classification of reads (Milani et al., 2021).

**Metagenome assembly**

Filtered reads were subjected to whole metagenome assembly using Spades v3.15 (Wedemeyer et al., 2017) with default parameters and the metagenomic flag option (-meta) together with k-mer sizes of 21, 33, 55 and 77. As mentioned above, for the short reads, reconstructed contig sequences were taxonomically classified based on their sequence identity using megablast against the same RefSeq database (Chen et al., 2021). ORFs of each assembled genome were predicted with Prodigal (Hyatt et al., 2010) and annotated utilizing the MEGAnnotator pipeline (Lugli et al., 2016). In all, the METAnnotatorX2 pipeline was employed for various purposes, from read filtering to taxonomic classification of the assembled contigs (Milani et al., 2018, 2021).

**Flow cytometry analysis**

The samples for FC were collected in sterilized screw tap tubes (Sarstedt) and were transported to the laboratory within 1 h of collection, temporarily stored at 4°C, and measured within a few hours after collection. Then, 1 ml of water sample was stained with 1 μl ml⁻¹ SYBR Green I (1:100 dilution in DMSO; Molecular Probes, Eugene, OR, USA) and incubated in the dark for 15 min before measurement. Count experiments were performed using an Attune NxT flow cytometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a blue laser set at 50 mW and tuned to an excitation wavelength of 488 nm. Multiparametric analyses were performed on scattering signals, i.e. forward scatter and side scatter, and SYBR Green I fluorescence was detected on the FL1 channel. The detection limit was determined experimentally by filtering one aliquot of water sample and one of Attune Focusing Fluid 1× through 0.20 μm pore size hydrophilic mixed cellulose esters (Pall Corporation). Then 1 ml of each sample was stained with 1 μl ml⁻¹ SYBR Green I as mentioned above. Cell debris was excluded from the acquisition analysis by a sample-specific FL1 threshold, and collected data were statistically analysed with Attune NxT flow cytometer software.

**DNA mapping**

Microbial DNA retrieved from 197 human faecal samples was aligned to the reconstructed chromosomal portions of unknown water bacteria to evaluate the presence of water microorganisms in the gut of humans. The Bowtie2 program was used to align the DNA sequences through multiple-hit mapping and a ‘very sensitive’ policy (Langdon, 2015). The mapping was performed using a minimum score threshold function (–score-min C,-13,0) to limit reads of arbitrary length to one mismatch and retain those matches with at least 99% full-length identity. The SAMtools software package (Danecek et al., 2021) was then used to count the mapped reads among each bacterial taxon, rejecting hits with less than 10 reads to achieve a consistent output.

**Experimental design**

The experiment involves a healthy adult male who daily drank tap water corresponding to sample W001 for the last three years. The objective was to collect faecal samples before and after the two weeks of the washout period. The first faecal sample collection was performed before the washout (F01-T1) to identify bacteria introduced by the consumption of tap water. Then, we collected faecal samples at T2 and T3 to cover the end of the washout and the restoration of W001 administration. Faecal samples were stored at –80°C until use. Comitantly, W001 was collected at T1 and T3 using the procedure reported above.

**Quantitative real-time PCR**

The abundance of microorganisms identified in W001 was evaluated through qPCR in faecal sample F01. The DNA of F01 was extracted and diluted at a concentration of 10 ng. The presence of *Bradyrhizobium* spp., *Sphingobium* spp. and *Curvibacter* spp. DNA was evaluated using q-PCR with primer pair (5’-TGCCGT-CACTCATTAGCT-3’/5’-GAGAACGCAGATCACCT-
TC-3’), (5’-CTGAACTGTTCGATCGGCTG-3’/5’-GCCATCGACCTCCTTATCCA-3’) and (5’-AGACCGTACA-GATCGTCG-3’/5’-TACACATGCAAGTCAAGTCAAGTCAAGTG-3’) respectively. In detail, qPCR was performed using SoFast EvaGreen Supermix (Bio-Rad) on a CFX96 system (BioRad, CA, USA) following previously described protocols (Milani et al., 2015). Each PCR reaction mix contained the following: 12.5 μl 2× SYBR SuperMix Green (BioRad), 5 μl of DNA at a concentration of 10 ng μl⁻¹, each of the forward and reverse primers at 0.5 μM, and nuclease-free water was added to obtain a final volume of 20 μl.

Statistical analysis

Bacterial abundance at the species level was validated by ANOVA analysis. Furthermore, PERMANOVA analysis was performed using 1000 permutations to estimate p-values of differences among water samples in PCoA analyses. Statistical analyses were performed by using OriginPro graphing and analysis 2021.

Acknowledgements

We thank GenProbio Srl for the financial support of the Laboratory of Probiogenomics. D.v.S. is a member of The APC Microbiome Institute funded by Science Foundation Ireland (SFI), through the Irish Government’s National Development Plan (Grant numbers SFI/12/RC/2273a and SFI/12/RC/2273b). This work was financially supported by a Post-Doc fellowship (Bando Ricerca Finalizzata). F.T. is funded by the Italian Ministry of Health through the Bando Ricerca Finalizzata (Grant Number GR-2018-12365988). Open Access Funding provided by Universita degli Studi di Parma within the CRUI-CARE Agreement.

Author Contributions

G.A.L. performed bioinformatics analyses and wrote the manuscript; G.L. performed the in vitro analyses and edited the manuscript; G.A. and C.T. validated the in vitro analyses; L.M. and F.F. validated the bioinformatics analyses; F.T., C.M., and D.v.S. supervised the project and edited the manuscript; M.V. designed the study and edited the manuscript.

Data Availability

Shotgun metagenomics data are accessible through SRA study accession number PRJNA806724.

Ethics Statement

Faecal experimental procedures and protocols involving human faecal samples were approved by the Ethics Committee of Parma University. Signed informed consents were obtained from the individuals enrolled in this study.

References

Bouchard, M.F., Sauvé, S., Barbeau, B., Legrand, M., Brodeur, M.E., Bouffard, T., et al. (2011) Intellectual impairment in school-age children exposed to manganese from drinking water. Environ Health Perspect 119: 138–143.

Bowyer, R.C.E., Schillereff, D.N., Jackson, M.A., Le Roy, C., Wells, p.M., Spector, T.D., and Stevens, C.J. (2020) Associations between UKtap water and gut microbiota composition suggest the gut microbiome as a potential mediator of health differences linked to water quality. Sci Total Environ 739: 139697.

Brumfield, K.D., Hasan, N.A., Leddy, M.B., Cotruvo, J.A., Rashed, S.M., Colwell, R.R., and Huq, A. (2020) A comparative analysis of drinking water employing metagenomics. PLoS One 15: e0231210.

Busse, H.J., Denner, E.B.M., Bucozils, S., Salkinoja-Salonen, M., Bennasar, A., and Kämpfer, P. (2003) Sphingomonas aurantiaca sp. nov., Sphingomonas aerolata sp. nov. and Sphingomonas faeni sp. nov., air- and dustborne and Antarctic, orange-pigmented, psychrotolerant bacteria, and emended description of the genus Sphingomonas. Int J Syst Evol Microbiol 53: 1253–1260.

Chan, S., Pullerits, K., Keucken, A., Persson, K.M., Paul, C.J., and Rådström, P. (2019) Bacterial release from pipe biofilm in a full-scale drinking water distribution system. npj Biofilms Microbiomes 5: 9.

Chao, Y., Ma, L., Yang, Y., Ju, F., Zhang, X.X., Wu, W.M., and Zhang, T. (2013) Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. Sci Rep 3: 3550.

Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015) High speed BLASTN: an accelerated MegaBLAST search tool. Nucleic Acids Res 43: 7762–7768.

Dai, D., Rhoads, W.J., Edwards, M.A., and Pruden, A. (2018) Shotgun metagenomics reveals taxonomic and functional shifts in hot water microbiome due to temperature setting and stagnation. Front Microbiol 9: 2695.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., et al. (2021) Twelve years of SAMtools and BCFtools. Gigascience 10: giab008.

Dodd, M.C. (2012) Potential impacts of disinfection processes on elimination and deactivation of antibiotic resistance genes during water and wastewater treatment. J Environ Monit 14: 1754–1771.

Eichler, S., Christen, R., Hölzl, C., Westphal, P., Bötel, J., Brettar, I., et al. (2006) Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-based 16S rRNA gene fingerprinting. Appl Environ Microbiol 72: 1858–1872.
Federal Food, Drug, and Cosmetic Act, Title 21 Chapter 9 of the United States Code.
França, L., Lopéz-Lopéz, A., Rosselló-Móra, R., and da Costa, M.S. (2015) Microbial diversity and dynamics of a groundwater and a still bottled natural mineral water. Environ Microbiol 17: 577–593.
Gulati, P., and Ghosh, M. (2017) Biofilm forming ability of Sphingomonas paucimobilis isolated from community drinking water systems on plumbing materials used in water distribution. J Water Health 15: 942–954.
Hammes, F., Berney, M., Wang, Y., Vital, M., Köster, O., and Egli, T. (2008) Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. Water Res 42: 269–277.
Hehemann, J.H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature 464: 908–912.
Hillmann, B., Al-Ghalith, G.A., Shields-Cutler, R.R., Zhu, Q., Goh, D.M., Beckman, K.B., et al. (2018) Evaluating the information content of shallow shotgun metagenomics. mSystems 3: e0069-18.
Hong, P.Y., Hwang, C., Ling, F., Andersen, G.L., LeChevallier, M.W., and Liu, W.T. (2010) Pyrosequencing analysis of bacterial biofilm communities in drinking water meters of a drinking water distribution system. Appl Environ Microbiol 76: 5631–5635.
Hyatt, D., Chen, G.L., LoCascio, p.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryote gene recognition and translation initiation site identification. BMC Bioinformatics 11: 119.
Kalmbach, S., Manz, W., Wecke, J., and Szewczyk, U. (1999) Aquabacterium gen. nov., with description of Aquabacterium citratophilum sp. nov., Aquabacterium parvum sp. nov. and Aquabacterium commune sp. nov., three in situ dominant bacterial species from the Berlin drinking water system. Int J Syst Bacteriol 49: 769–777.
Langdon, W.B. (2015) Performance of genomic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Min 8: 1.
Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transformer. Bioinformatics 25: 1754–1760.
Lorite, M.J., Estrella, M.J., Escaray, F.J., Sannazzaro, A., Castro, I.M.V.E., Monza, J., et al. (2018) The rhizobial lotus symbioses: deeply specific and widely diverse. Front Microbiol 9: 2055.
Loubet, P., Roux, P., Guérin-Schneider, L., and Bellon-Maurel, V. (2016) Life cycle assessment of forecasting scenarios for urban water management: a first implementation of the WaLA model on Paris suburban area. Water Res 90: 128–140.
Loy, A., Beisker, W., and Meier, H. (2005) Diversity of bacteria growing in natural mineral water after bottling. Appl Environ Microbiol 71: 3624–3632.
Lucena, F., Blanch, A.R., and García-Aljaro, C. (2020) The rhizobia-ursincola isolated from industrial cooling water systems. J Biosci Bioeng 123: 569–575.
Makki, K., Deehan, E.C., Walter, J., and Bäckhed, F. (2018) The impact of dietary fiber on gut microbiota in host health and disease. Cell Host Microbe 23: 705–715.
McAlister, M.B., Kulakow, L.A., O’Hanlon, J.F., Larkin, M.J., and Ogden, K.L. (2002) Survival and nutritional requirements of three bacteria isolated from ultrapure water. J Ind Microbiol Biotechnol 29: 75–82.
Michan, C., Blasco, J., and Alhama, J. (2021) High-throughput molecular analyses of microbiomes as a tool to monitor the wellbeing of aquatic environments. J Microbial Biotechnol 14: 870–885.
Milani, C., Andrea Lugli, G., Duranti, S., Turroni, F., Mancabelli, L., Ferrario, C., et al. (2015) Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. Sci Rep 5: 15782.
Milani, C., Casey, E., Lugli, G.A., Moore, R., Kaczorowska, J., Feehily, C., et al. (2018) Tracing mother-infant transmission of bacteriophages by means of a novel analytical tool for shotgun metagenomic datasets: METAnnotatorX. Microbiome 6: 145.
Milani, C., Duranti, S., Napoli, S., Alessandri, G., Mancabelli, L., Anzalone, R., et al. (2019) Colonization of the human gut by bovine bacteria present in parmesan cheese. Nat Commun 10: 1286.
Milani, C., Lugli, G.A., Fontana, F., Mancabelli, L., Alessandri, G., Longhi, G., et al. (2021) METAnnotatorX2: a comprehensive tool for deep and shallow metagenomic data set analyses. mSystems 6: e0058321.
Morohoshi, T., Sato, N., Iizumi, T., Tanaka, A., and Ikeda, T. (2017) Identification and characterization of a novel N-acyl-homoserine lactone gene in Sphingomonas ursincola isolated from industrial cooling water systems. J Biosci Bioeng 123: 569–575.
Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499: 431–437.
Sala-Comorera, L., Caudet-Segarra, L., Galofré, B., Lucena, F., Blanch, A.R., and García-Aljaro, C. (2020) Unravelling the composition of tap and mineral water microbiota: divergences between next-generation sequencing techniques and culture-based methods. Int J Food Microbiol 334: 108850.
Sheu, S.Y., Shiau, Y.W., Wei, Y.T., and Chen, W.M. (2013) Sphingobium fontiplumphi sp. nov., isolated from a freshwater spring. Int J Syst Evol Microbiol 63: 1906–1911.
Simonin, M., Voss, K.A., Hassett, B.A., Rocca, J.D., Wang, S.Y., Bier, R.L., et al. (2019) In search of microbial indicator taxa: shifts in stream bacterial communities along an urbanization gradient. Environ Microbiol 21: 3655–3668.
Singh, H., Du, J., Yang, J.E., Yin, C.S., Kook, M.C., and Yi, T.H. (2015) Novosphingobium aquaticum sp. nov., isolated from lake water in Suwon, Republic of Korea. Antonie van Leeuwenhoek 108: 851–858.
Szewczyk, U., Szewczyk, R., Manz, W., and Schleifer, K.H. (2000) Microbiological safety of drinking water. Annu Rev Microbiol 54: 81–127.
Takeuchi, M., Hamana, K., and Hiraishi, A. (2001) Proposal of the genus Sphingomonas sensu stricto and three new genera, Sphingobium, Novosphingobium and Sphingopyxis, on the basis of phylogenetic and chemotaxonomic analyses. *Int J Syst Evol Microbiol* **51**:1405–1417.

Vosloo, S., Huo, L., Anderson, C.L., Dai, Z., Sevillano, M., and Pinto, A. (2021) Evaluating de novo assembly and binning strategies for time series drinking water metagenomes. *Microbiol Spectr* **9**: e0143421.

Wedemeyer, A., Kliemann, L., Srivastav, A., Schielke, C., Reusch, T.B., and Rosenstiel, P. (2017) An improved filtering algorithm for big read datasets and its application to single-cell assembly. *BMC Bioinformatics* **18**: 324.

Yoon, J.H., Lee, C.H., Yeo, S.H., and Oh, T.K. (2005) Sphingopyxis baekryungensis sp. nov., an orange-pigmented bacterium isolated from sea water of the Yellow Sea in Korea. *Int J Syst Evol Microbiol* **55**:1223–1227.

Zhang, X., Xia, S., Ye, Y., and Wang, H. (2021) Opportunistic pathogens exhibit distinct growth dynamics in rainwater and tap water storage systems. *Water Res* **204**: 117581.

**Supporting Information**

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

**Fig. S1.** Absolute abundance of the microbial composition of 16 drinking water samples collected from different locations of the Parma district and delivered by the city water supply system. Due to the high CFU ml⁻¹ of sample W001, its histogram is reported with a different CFU ml⁻¹ scale to appreciate the absolute abundance of the microbial community in the other 15 water samples.

**Table S1.** Shallow shotgun sequencing filtering data
**Table S2.** Public shotgun sequencing metadata
**Table S3.** Biodiversity between water projects.