

# SCIENTIFIC REPORTS



OPEN

## Exploring the under-investigated “microbial dark matter” of drinking water treatment plants

Antonia Bruno<sup>1</sup>, Anna Sandionigi<sup>1</sup>, Ermanno Rizzi<sup>2,3</sup>, Marzia Bernasconi<sup>4</sup>, Saverio Vicario<sup>5,6</sup>, Andrea Galimberti<sup>1</sup>, Clementina Cocuzza<sup>7</sup>, Massimo Labra<sup>1</sup> & Maurizio Casiraghi<sup>1</sup>

Received: 15 March 2016  
Accepted: 09 February 2017  
Published: 14 March 2017

Scientists recently reported the unexpected detection of unknown or poorly studied bacterial diversity in groundwater. The ability to uncover this neglected biodiversity mainly derives from technical improvements, and the term “microbial dark matter” was used to group taxa poorly investigated and not necessarily monophyletic. We focused on such under-investigated microbial dark matter of drinking water treatment plant from groundwater, across carbon filters, to post-chlorination. We tackled this topic using an integrated approach where the efficacy of stringent water filtration (10000 MWC0) in recovering even the smallest environmental microorganisms was coupled with high-throughput DNA sequencing to depict an informative spectrum of the neglected microbial diversity. Our results revealed that the composition of bacterial communities varies across the plant system: Parcubacteria (OD1) superphylum is found mainly in treated water, while groundwater has the highest heterogeneity, encompassing non-OD1 candidate phyla (Microgenomates, Saccharibacteria, Dependientiae, OP3, OP1, BRC1, WS3). Carbon filters probably act as substrate for microorganism growth and contribute to seeding water downstream, since chlorination does not modify the incoming bacterial community. New questions arise about the role of microbial dark matter in drinking water. Indeed, our results suggest that these bacteria might play a central role in the microbial dynamics of drinking water.

In October 2015, NASA announced the indirect evidence of liquid water on Mars<sup>1</sup> raising hopes about the existence of the essential medium for life as was hypothesized more than twenty years ago<sup>2</sup>. The possible presence of putative tiny microbial cells in a meteorite from the red planet originated a never-ending debate about the existence of cells that were considered too small to be organisms by part of the scientific community. The quest for extraterrestrial life is fascinating and still unsolved, but the existence of ultra-small biodiversity and/or unknown biodiversity in aquatic environments is much closer to us than previously expected. Adopting the definition of Solden and colleagues<sup>3</sup>, those microorganisms accounting for a large proportion of life and biodiversity but whose basic metabolic and ecological properties are not known are called *microbial dark matter*. Candidate Phyla (CP), that are composed of uncultured organisms, constitute this under-investigated portion of biodiversity and their description represents a key challenge for the scientific community<sup>3</sup>.

For example, ultra-small bacteria (i.e. median cell volume:  $0.009 \pm 0.002 \mu\text{m}^3$ , genome size: less than 1 Mb<sup>4,5</sup>) have been surprisingly detected in groundwater on Earth<sup>4,6</sup> and show dimensions under the minimal predicted sizes<sup>7</sup>. They are currently defined as a candidate taxon, including at least 35 phyla with still unsolved phylogenetic relationships and without representatives isolated in culture<sup>6</sup> with only few exceptions<sup>8–10</sup>.

These findings led researchers to consider drinking water treatment plants (DWTs) as a source of unexpected biodiversity in terms of environmental microorganisms whose interactions at the community level are still poorly known. The occurrence of Candidate Phyla Radiation (CPR) bacteria, some of which have been shown

<sup>1</sup>University of Milan-Bicocca, ZooPlantLab, Biotechnology and Biosciences Department, Piazza della Scienza 2, 20126, Milan, Italy. <sup>2</sup>National Research Council (CNR), Institute of Biomedical Technologies (ITB), Via Fratelli Cervi, 93, 20090 Segrate (MI), Italy. <sup>3</sup>Fondazione Telethon Piazza Cavour, 1, 20121, Milan, Italy. <sup>4</sup>Metropolitana Milanese S.p.A., Via Giuseppe Meda 44, 20141, Milan, Italy. <sup>5</sup>Institute of Atmospheric Pollution Research, National Research Council, C/O Physics Department, University of Bari “Aldo Moro”, Via Giovanni Amendola, 173 70126, Bari, Italy. <sup>6</sup>National Research Council (CNR), Institute of Biomedical and Technologies (ITB), via Giovanni Amendola, 122/D, 70126, Bari, Italy. <sup>7</sup>University of Milan-Bicocca, Medicine and Surgery Department, Via Cadore 48, 20126, Monza, Italy. Correspondence and requests for materials should be addressed to M.C. (email: maurizio.casiraghi@unimib.it)

to have ultrasmall cell sizes, and other Candidate Phyla (CP) of uncultured bacteria is underestimated in many environments, including DWTPs.

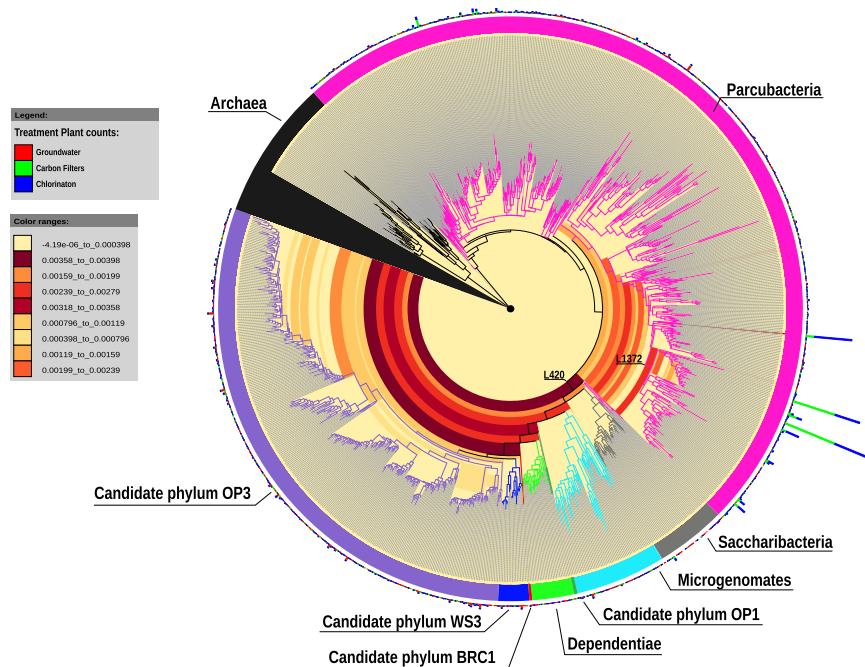
Understanding the microbial diversity and ecology of DWTPs is necessary for designing innovative and effective control strategies that will ensure safe and high quality drinking water. Previous studies of microbial communities in drinking water utilized various molecular techniques, such as DGGE, FISH, T-RFLP, clone libraries, and microarrays<sup>11</sup>. Only in a few recent studies have bacterial communities in this particular environment been analysed using High-Throughput DNA Sequencing (HTS) techniques<sup>12–14</sup>. In our opinion, the DWTP's intrinsic complexity requires an innovative combined strategy, in which new and previously known tools (e.g. a stringent water filtration, HTS and bioinformatics) are used in an integrated analytic environment. In this study, we coupled the stringent filtration conditions with the deep resolution of HTS in order to cope with the significantly lower concentrations of organisms in drinking water when compared to other known cases (such as the Human Microbiome Project<sup>15</sup>). To our knowledge, this is the only option in a case where organisms are scattered, uncultivable, and very small. Moreover, as a post-sequencing analysis we used a phylogenetic entropy approach implemented in PhyloH<sup>16</sup> to provide a comprehensive view of the microbial diversity in water samples. Differently from conventional phylogenetic methods, this is still an innovative approach in environmental microbial ecology that allows the identification of which lineage, or groups of lineages, give the most significant contribution to the diversity. Compared to conventional phylogenetic methods, the permutation process on all sequences, implemented in PhyloH, prevents any subsampling procedure (i.e. rarefaction): subsampling entails the ability to detect only highly abundant variants, reducing the resolution power, and hiding the signal coming from rare OTUs<sup>17</sup>.

The aim of this work was to describe the emergent microbiological water contaminants that are undetectable using classical approaches (e.g. culture based methods). The critical interpretation of the results will then improve the DWTP monitoring capacities and, consequently, their management strategies. To reach these goals, we analysed the microbial community of water at different steps in the potabilization process: i) raw water from the ground (GW), ii) water after passage across granular activated carbon filters (CF), and iii) water after chlorination (CHL), during an extended monitoring campaign of drinking water treatment plants (DWTPs) located in Milan (Northern Italy). As stated above, groundwater is characterised by low concentrated and often uncultivable microorganisms<sup>11</sup> and, as has recently been discovered, very small bacteria<sup>4</sup>. To uncover the whole biodiversity of drinking water, we concentrated our samples using a tangential flow filtration system with a nominal pore rating of 10000 MWCO, which is more stringent than the traditional 0.2  $\mu\text{m}$  pore filter. The concentration was then followed by extraction of the environmental DNA and sequencing of the 16S rDNA V3-V4 regions.

## Results and Discussion

The core of our results is the analysis of the occurrence and diversity of *under-investigated microbial dark matter* along different steps of the DWTP. Unexpectedly, we recorded the presence of bacterial 16S rDNA sequences even after the end of the potabilization process.

Specifically, we found 36% of sequences belonging to Candidate Phyla and Candidate Phyla Radiation and, among these, sequences of ultra-small bacteria<sup>4</sup>. In total 1123 OTUs have been assigned to Parcubacteria (OD1) superphylum, Microgenomates (OP11) superphylum, Saccharibacteria (TM7), Dependistia (TM6), OP3, OP1, BRC1, and WS3 candidate phyla (See Fig. 1 and Supplementary Table S2). The Parcubacteria superphylum was the most represented bacteria group in the DWTP (31% of the entire bacterial community, see Table S2 in Supplementary Information for the complete list), whereas all the other CPR and CP contributed only 4% of the total bacterial diversity. The tree of identified bacteria (Fig. 1) would be intended as an entropy-based “map” to estimate the total lineage diversity in our study, and it is not aimed at resolving the bacterial phylogeny. Our results showed that the investigated bacterial community varied among GW, CF, and CHL in DWTP, where groundwater was characterised by the highest diversity ( $\alpha$ -diversity values: GW = 2.22, CF = 1.54, CHL = 1.57, for more details see html file in Supplementary Information). The GW samples shared a similar composition throughout the whole survey even when considering the different sampling sites within the whole sampling area (phylogenetic turnover mean across samples of the same group: 5%, Fig. 2 and Supplementary Information). Water samples deriving from CF and CHL shared the same diversity ( $\beta$ -diversity expressed as phylogenetic turnover, CF-CHL: 0.57%) which is also significantly different from GW (CF-GW: 14% and CHL-GW: 12.6%,  $p < 0.001$ ). Interestingly, there was no evidence for seasonality affecting microbial composition, suggesting that groundwater could be a stable and resilient ecosystem that is not easily affected by external conditions. Noteworthy, all the classical culture methods applied to these samples highlighted the absence of pathogens commonly screened in drinking water (according to the European Directives). All the water samples included in the analysis were labelled as potable after the chlorination step (see Table S3 and S4 in Supplementary Information). Although all the drinking water around the world is treated before human consumption to remove chemical and biological contaminants, relatively little is known about the changes in microorganism composition during the potabilization processes. It is noteworthy that our data suggests that carbon filters could act as a substrate for microorganism growth and could also contribute to seeding water downstream, since chlorination did not greatly modify the incoming bacterial community (Fig. 2). Pinto and co-authors<sup>12</sup> observed a similar pattern, but they did not focus on the not culturable and poorly known bacteria diversity that is a large fraction of the total microbial diversity recovered in our results. Our data further support the seeding role of carbon filters, since samples that came from the renewed (i.e. sterile, see Supplementary Information) carbon filters were more similar to GW than to CF or CHL samples in operating conditions (CF renewed - GW: 3.72%; CF renewed - CF before filters renewing: 15.3%; CF renewed - CHL before filters renewing: 9.93%) (Fig. 2 and details in Fig. S1 in Supplementary Information). Water samples at the CHL step (after flowing through renewed filters) showed increased diversity (i.e. mean phylogenetic turnover from 5% to 16.5%) when comparing samples deriving from the CHL basin before and after filter replacement. When we excluded the samples collected during the carbon filter renewal, variation across compartment CF and CHL were similar, and variation within the three compartments across time were comparable to variation across

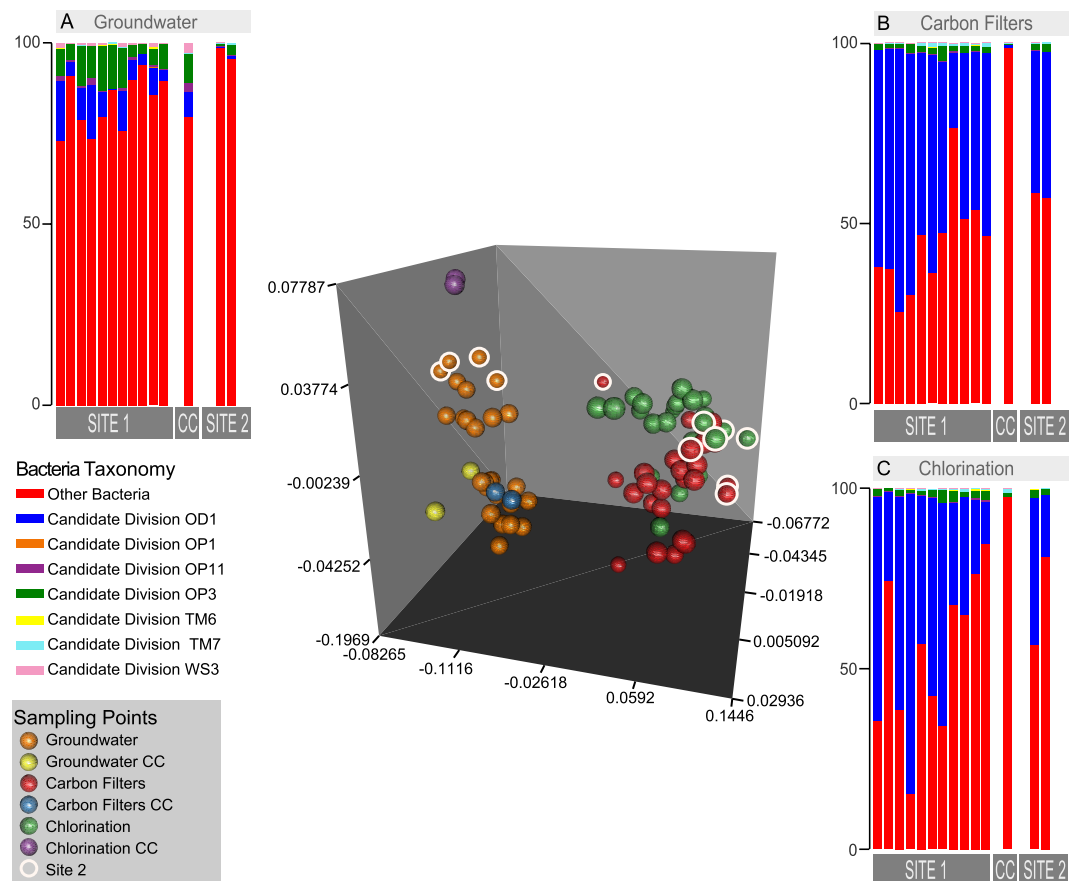


**Figure 1. Hairy pacman graphical output from PhyloH analysis.** The output couples the phylogenetic information from the RAxML tree and the contribution of the different branches/lineages to the mutual information (i.e. the information shared between each lineage and the grouping at which a given observation belongs expressed as a gradient of colours, where yellow is the null contribution and dark red the maximum). Multiple bars represent the proportion of counts associated with each lineage with respect to the three different sampling points.

replicates (Fig. S1 in Supplementary Information). A strong point of our approach is the capacity to identify a critical taxon with no or few previous taxonomic information due to the application of the phylogenetic entropy as implemented in PhyloH<sup>16</sup>. Specifically, in our study the use of PhyloH allowed us to investigate the contribution to the total diversity of different lineages instead of summarising the results as a simple check-list of predefined taxa as typically seen in many published works<sup>16</sup>. For instance, in our analysis, the lineage named L1372 belonging to Parcubacteria (OD1) superphylum (Fig. 1 and html in Supplementary Information) characterized treated waters (4%, 68%, and 58% of sequences of GW, CF, and CHL samples, respectively). In carbon filters, the proportion of L1372 increased dramatically reaching about 22% of the total bacterial sequences. On the contrary, lineage L420, that includes all the non-OD1 under-investigated bacteria phyla observed, was typically found in GW samples (58%, 7%, and 10% of GW, CF, and CHL samples, respectively). The Parcubacteria members not belonging to lineage L1372 were spread across the compartments at a low percentage (Fig. 1). L1372 and L420 lineages explained 6.4% of the total turnover across the three compartments. The Parcubacteria (OD1) superphylum was spread across the entire DWTP, but our results indicate that the group L420 (including all the non-OD1 under-investigated bacteria phyla observed) was typical in groundwater. Thus less abundant phyla strongly contributed to  $\alpha$ -diversity in groundwater.

## Conclusions

Our results highlight the presence of under-investigated microorganisms across the entire DWTP, even after the potabilization process. But what are the implications of “having a drink” of these mostly unknown microorganisms? According to the parameters provided by international directives (e.g. the European 98/83/CE), drinking water analysed during this survey was clearly “labelled” as potable. Nowadays, these poorly studied bacteria are not (directly or indirectly) linked to any pathogenic condition and are not considered good markers for particular biological activities/status. Consequently, they are not routinely screened. Nevertheless, the detection of this group of uncultivable bacteria in drinking water and their incredible persistence in DWTP open new scenarios. For instance, according to their extremely small genomes, it is likely that ultra-small bacteria depend on other bacteria to survive<sup>4,6,18</sup>. The common theory is that they are extracellular obligate symbionts. If this is true, a simple but key question is: what are the real interactions in the microbial network characterising drinking water? Is it possible that ultra-small bacteria occurrences and concentrations are indirectly linked to the peculiarities of drinking water through the bacteria symbioses? And lastly, but with important practical consequences, could the removal or controlled maintenance of microbial dark matter affect water plant management? We have more questions than answers, but it is clear that DWTPs should be treated as complex ecosystems rather than inert systems where a tangled network of microbial interactions take place from the source (groundwater, river, lake, and so on) to the tap in our house. Better knowledge of these networks is crucial for improving the management of drinking water facilities.



**Figure 2. PCoA using phylogenetic turnover as a distance metric.** Groundwater (GW) samples are separated by PC1 from Carbon filters (CF) and Chlorination (CHL) samples. Samples belonging to CF and CHL characterised by new (i.e. sterile) carbon filters (CC) are more similar to GW than to CF and CHL samples (see the text for further details). Samples deriving from a different DWTP (Site 2) are circled white. A-B-C barplot describes the phyla distribution of ultra-small bacteria recovered in the different sampling points.

## Methods

**Sampling.** Samples were obtained from two drinking water treatment plants located in Milan, Italy. We collected water samples from different steps of the potabilization processes: i) from groundwater (GW), ii) after passage through granular activated carbon filters (CF), and iii) after chlorination (CHL). The sampling campaign lasted one year, from December 2013 to November 2014. In total we collected 42 samples, listed in Table S1.

**Sample concentration.** In order to reduce the volume of the samples and therefore concentrate the bacteria, we used a tangential flow filtration (TFF) system. The system involves a peristaltic pump (Masterflex L/S Economy Drive), Tygon tubing, sterile reservoirs, and filtration modules. The tangential flow filter used was a VivaFlow 200 cassette (Sartorius) composed of polyethersulfone (PES) with a nominal pore rating of 10000 MWCO and a surface area of 200 cm<sup>2</sup>. The system was scaled up with an additional unit connected in parallel to increase the filtration surface area and the flow speed. All tubing, tubing connections, and containers were sterilized with sodium hypochlorite or autoclaved prior to each experiment. Every step was conducted in the laminar flow cabinet. The TFF system was run at a transmembrane pressure of 1.5 bar. TFF experiments were carried out within 24 h after sampling, and samples were always kept at 4 °C. For each sampling point, seven liters of water were concentrated to obtain 100 mL of operative volume.

**DNA extraction and sequencing.** Environmental DNA extraction was carried out with an automated nucleic acid extraction (NucliSens EasyMAG system, bioMérieux), based on magnetic beads. Starting from 1 mL of sample, the nucleic acids were eluted into a final volume of 50 µL and stored at -80 °C. Illumina MiSeq 16S (V3-V4 region of 16S rRNA gene) libraries were generated following standard protocol (16S Metagenomic Sequencing Library Preparation, Part # 15044223 Rev. B) with modifications due to the low DNA concentrations. Specifically, DNA extracts were normalized on Ct values of Real Time PCR with the same primer pairs instead of measuring the total amount of microbial DNA with fluorometric/spectrophotometric methods. Amplicon PCR was performed using the primer pairs 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG3'/5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAAT CC3' at an initial concentration of [10 µM] with the aim of increasing the volume of DNA in the reaction. The PCR-clean up step after amplicon PCR was modified in the final resuspension volume with a two-fold increase in the sample



concentration (as described in Bruno and colleagues<sup>14</sup>). Samples were sequenced using  $2 \times 300$  paired-end chemistry (MiSeq Reagent Kit v3). Technical replicates were included in order to verify the sequencing reproducibility (84 samples in total). 16S rRNA gene sequence processing and operational taxonomic unit (OTU) selection Illumina reads were paired and pre-processed using USEARCH script<sup>19</sup>. During the Quality filter step reads were filtered out if: 1) ambiguous bases were detected, 2) lengths were outside the bounds of 250 bp and/or 3) average quality scores over a sliding window of 40 bp dropped below 25. Reads were then processed by VSEARCH 1.1.3 software version (<https://github.com/torognes/vsearch>), which removed noise and chimeras prior to performing de novo clustering into OTUs at 97% sequence identity and discarding those OTUs with <75 sequences.

**Microbial composition and community structure analysis.** A representative sequence was selected randomly for each OTU and classified with the RDP (Ribosomal Database Project) classifier v2.2<sup>20</sup> using the Silva reference set (119 release)<sup>21</sup>. The taxonomic assignment of each sequence was obtained with a confidence score of at least 0.8. Data included in the analyses take into consideration the fraction of the under-investigated microbial community constituted by Candidate Phyla (CP) and Candidate Phyla Radiation (CPR), and among these the occurrence of ultra-small bacteria has been reported<sup>4</sup>. As a consequence, a total of 3,996,876 reads passed the QC step. The OTU representative sequence set was aligned to the Silva dataset using mothur<sup>22</sup>, and OTUs have been assigned to Parcubacteria (OD1) superphylum, Microgenomates (OP11) superphylum, Saccharibacteria (TM7), Dependientiae (TM6), OP3, OP1, BRC1, and WS3 candidate phyla. Based on the alignment of OTU representative sequences, a phylogenetic tree was then built using RAxML version 7.4.2<sup>23</sup> with the GTRGAMMA model, bootstrapping (1,000 replicates), best maximum likelihood tree inference, and displayed with iTol<sup>24</sup> (<http://itol.embl.de/>) representing the output of PhyloH analysis with the ultra-small bacteria OTUs count as multibarplots. Multibarplots are generated with QIIME<sup>25</sup>. 3D PCoA was built using *rgl* R package<sup>26</sup>.

**PhyloH.** The measurement of the variety of sequences found in the different samples was done within the overall analysis framework defined in Sandionigi *et al.*<sup>16</sup>, where the ecological concept of gamma diversity  $D_\gamma$  and  $\alpha$ -diversity  $D_\alpha$  are identified as the exponential of the phylogenetic generalization of Shannon ( $H_p$ ) proposed by Chao *et al.*<sup>27</sup> and the exponential of the mean  $H_p$  per group of observations, respectively (equation (1)):

$$\begin{aligned} H_p &= H_p(T) = H_p\gamma = - \sum_{i \in BT} \frac{Li}{\tau T} p_i \log(p_i) \\ \tau T &= \sum_{i \in BT} Li p_i \\ H_p\alpha &= H_p(T|G) = \sum_{g \in G} p_g \left( - \sum_{i \in BT_g} \frac{Li}{\tau T_g} p_{ig} \log(p_{ig}) \right) \\ \tau T_g &= \sum_{i \in BT_g} Li p_{ig} \end{aligned} \quad (1)$$

where  $p_i$  is the frequency of observations of organisms descendant of node  $i$  and  $Li$  is the length of the branch of the node  $i$  over the phylogenetic tree  $T$ . By subdividing observations in groups, it is possible to define  $p_{ig}$  as the frequency of observations of organisms descendant of node  $i$  and belonging to group  $g$ . The ecological concept of  $\beta$ -diversity is identified with the exponential of mutual information between the species observation and the grouping (I(Obs, G) as proposed by Jost (2007)<sup>28</sup>, and we applied this concept to the phylogenetic generalization of Shannon (equation (2))

$$I_p(T, G) = H_p(T) - H_p(T|G) = H_{p\beta} = H_{p\gamma} - H_{p\alpha} \quad (2)$$

This phylogenetic generalization of mutual information describes the information shared between the lineage and the grouping at which a given observation belongs. By modifying the order of the summation, it is possible to extract the contribution of each branch/lineage to mutual information ( $H_{p\beta i}$ ) (equation (3))

$$\begin{aligned} H_{p\beta i} &= - \frac{Li}{\tau T} p_i \log(p_i) - \sum_{g \in G} - p_g \frac{Li}{\tau T_g} p_{ig} \log(p_{ig}) \\ H_{p\beta} &= \sum_{i \in Bt} (H_{p\beta i}) \end{aligned} \quad (3)$$

Following Chao<sup>27</sup>, in our work we reported the exponential of  $H_{p\gamma}$  such that the unit of measurement is an equivalent number of equi-abundant independent lineages, meaning the number of branches of a star tree in which each terminal taxon is equally abundant and would produce the same level of diversity than in the actually observed sample. As a summary of the differentiation of communities (i.e. the  $\beta$ -diversity), we preferred not to use the exponential of  $H_\beta$  that would produce estimates in equivalent number of samples, a measure quite ambiguous when samples have different numbers of observations. As a result, we normalized the  $H_\beta$  by its maximum possible values given the experimental design. Mutual information shared across two variables cannot be bigger than the entropy of the least entropic variable. Given  $I_p(T, G)$ , the number of groups is fixed, while  $T$  is unknown prior to data observation, so the maximum value that mutual information could take is  $H(G)$ , and therefore, mutual information was normalized between 1 and zero using this value. This measure was defined as *turnover* and, in the case of two groups, is the percentage of observations belonging to a not shared lineage.

Due to the possible effect derived from a different amount of reads assigned to treated water and ground-water, we modified the calculation of  $p_i$  and  $p_{ig}$  such that each level would contribute equally to those estimates

and not proportionally to its number of reads. These changes do not obscure our capacity to correctly estimate if the mutual information is different from zero. Given that, as in Sandionigi *et al.*<sup>16</sup>, significance was obtained by comparing the original dataset with results from a permuted data set in which grouping labels were randomly re-assigned to observations. All output files generated by PhyloH during this study are stored in the “Analysis” folder in the Supplementary Information files.

## References

- Ojha, L. *et al.* Spectral evidence for hydrated salts in recurring slope lineae on Mars. *Nat. Geosci.* **8**, 829–832 (2015).
- McKay, D. S. *et al.* Search for past life on Mars: possible relic biogenic activity in martian meteorite ALH84001. *Science* **273**, 924–930 (1996).
- Solden, L., Lloyd, K. & Wrighton, K. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr. Opin. Microbiol.* **31**, 217–226 (2016).
- Luef, B. *et al.* Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* **6**, 6372 (2015).
- Kantor, R. S. *et al.* Small Genomes and Sparse Metabolisms of Sediment-Associated Bacteria from Four Candidate Phyla. *MBio* **4**, e00708-13-e00708-13 (2013).
- Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
- Size Limits of Very Small Microorganisms. (National Academies Press, doi: 10.17226/9638 1999).
- Luo, F., Devine, C. E. & Edwards, E. A. Cultivating microbial dark matter in benzene-degrading methanogenic consortia. *Environ. Microbiol.* **18**, 2923–2936 (2016).
- Soro, V. *et al.* Axenic culture of a candidate division TM7 bacterium from the human oral cavity and biofilm interactions with other oral bacteria. *Appl. Environ. Microbiol.* **80**, 6480–9 (2014).
- Kindaichi, T. *et al.* Phylogenetic diversity and ecophysiology of Candidate phylum Saccharibacteria in activated sludge. *FEMS Microbiol. Ecol.* **92** (2016).
- Douterelo, I. *et al.* Methodological approaches for studying the microbial ecology of drinking water distribution systems. *Water Res.* **65**, 134–156 (2014).
- Pinto, A. J., Xi, C. & Raskin, L. Bacterial Community Structure in the Drinking Water Microbiome Is Governed by Filtration Processes. *Environ. Sci. Technol.* **46**, 8851–8859 (2012).
- Chao, Y. *et al.* Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. *Sci. Rep.* **3**, 219–234 (2013).
- Bruno, A. *et al.* One step forwards for the routine use of high-throughput DNA sequencing in environmental monitoring. An efficient and standardizable method to maximize the detection of environmental bacteria. *Microbiologyopen*, doi: 10.1002/mbo3.421 (2016).
- Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Sandionigi, A. *et al.* Towards a better understanding of *Apis mellifera* and *Varroa destructor* microbiomes: introducing ‘ phyloh ’ as a novel phylogenetic diversity analysis tool. *Mol. Ecol. Resour.* **15**, 697–710 (2015).
- McMurdie, P. J. *et al.* Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
- Nelson, W. C. & Stegen, J. C. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front. Microbiol.* **6**, 713 (2015).
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–200 (2011).
- Cole, J. R. *et al.* The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33**, D294–6 (2005).
- Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
- Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–41 (2009).
- Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–90 (2006).
- Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–8 (2007).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Nenadi, O., Adler, D. & Zucchini, W. An R An R-library for 3D visualization library for 3D visualization with OpenGL with OpenGL Introduction: Motivation and goals. (2003)
- Chao, A. *et al.* Phylogenetic diversity measures based on Hill numbers. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**, 3599–609 (2010).
- Jost, L. Partitioning Diversity Into Independent Alpha And Beta Components. *Ecology* **88**, 2427–2439 (2007).

## Acknowledgements

We would like to thank the laboratory manager of the Metropolitana Milanese S.p.A. Angela Manenti for her great contribution during the whole survey and the precious help of the technicians of the Metropolitana Milanese S.p.a. for the sample collection (and here a special thank you for Rosolino Azzarello). The work was partially conducted during the Cariplo Foundation (Milan, Italy) project 2015-0275: “Xenogenetics of drinking water: Microbiome and Resistome analyses for the improvement of water safety assessment tools (XenoMicroResist)”.

## Author Contributions

A.B. participated in the experimental design, performed the molecular lab analyses up to the HTS sequencing, and wrote the first draft of the manuscript. A.S. performed the bioinformatics analyses, including the statistical tests and collaborated in writing the manuscript. E.R. performed the HTS runs and read the manuscript. M.B. coordinated the activities at the DWTP, discussed the results, and read the manuscript. S.V. implemented the statistical analyses with A.S., wrote PhyloH, and read the manuscript. A.G. discussed the results and read the manuscript. C.C. helped in the DNA extractions and read the manuscript. M.L. collaborated in the experimental design and discussed the results. M.C. designed the experiment, discussed the progress and results, and revised the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Bruno, A. *et al.* Exploring the under-investigated “microbial dark matter” of drinking water treatment plants. *Sci. Rep.* 7, 44350; doi: 10.1038/srep44350 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017