

## **A Citizen Science Approach for Mapping Total Microbial Communities and Waterborne Pathogens in Household Drinking Water in China**

1 **Xinyi Wen<sup>1,2,3</sup>, Chutong Fang<sup>1,4</sup>, Lihan Huang<sup>1,5</sup>, Jiazheng Miao<sup>1,6</sup>, Yajuan Lin<sup>1,7\*</sup>**

2 <sup>1</sup>Division of Natural and Applied Science, Duke Kunshan University, Kunshan, Jiangsu, China

3 <sup>2</sup>Nicholas School of the Environment, Duke University, Durham, NC, United States

4 <sup>3</sup>Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University,  
5 Durham, NC, United States

6 <sup>4</sup>Blum Center for Developing Economies, University of California - Berkeley, Berkeley, CA, United  
7 States

8 <sup>5</sup>Department of Civil and Environmental Engineering, School of Engineering, Stanford University,  
9 Stanford, CA, United States

10 <sup>6</sup>Department of Biomedical Informatics, Harvard Medical School, Harvard University, Boston, MA,  
11 United States

12 <sup>7</sup>Department of Life Sciences, College of Science, Texas A&M University - Corpus Christi, Corpus  
13 Christi, TX, United States

14

15 **\* Correspondence:**

16 Yajuan Lin

17 [yajuan.lin@tamucc.edu](mailto:yajuan.lin@tamucc.edu)

18 **Keywords: Citizen Science<sub>1</sub>, Household drinking water<sub>2</sub>, Microbial communities<sub>3</sub>, Waterborne  
19 pathogens<sub>4</sub>, Metabarcoding<sub>5</sub>.**

## 20 Abstract

21 Ensuring safe drinking water is one of the top priorities in public health as waterborne diseases  
22 remain a global challenge. In China, microbial contamination in drinking water is a particular  
23 concern and comprehensive survey/monitoring of the drinking water microbiome is necessary.  
24 However, traditional culture-based microbial monitoring methods have significant limitations, and  
25 nationwide tap water survey/monitoring in China would require significant resources. Here, a cost-  
26 effective citizen science approach was developed to collect household drinking water samples ( $n =$   
27 50) from 19 provinces in China during December 2020 to August 2021. Using a protocol optimized  
28 for low-biomass samples, 22 out of 50 tap water samples were tested positive for microbial DNA.  
29 The PCR products were pooled for 16S rRNA genes metabarcoding to elucidate the tap water  
30 microbiome and detect waterborne pathogens, yielding 7,635 Amplicon Sequence Variants (ASVs).  
31 Outdoor temperature was found to be the first-order driver of total microbial community structure,  
32 validating our citizen science approach against previous studies. Alarming, pathogenic bacteria  
33 including *Mycobacterium* spp., *Acinetobacter* spp., and *Legionella* spp. were detected in all PCR  
34 positive samples. More importantly, elevated proportions or new appearance of toxin-producing  
35 cyanobacteria (e.g., *Microcystis* spp.) and pathogenic species (e.g., *Salmonella enterica*) were evident  
36 in local tap water samples after the extreme rainstorm event in Zhengzhou on July 20, 2021, and  
37 Typhoon In-Fa Landfall. High pathogen relative abundances were found to be significantly correlated  
38 with high outdoor temperatures. This underscores the need for enhanced drinking water treatment  
39 protocols during and following extreme rainfall events and/or periods of high temperatures,  
40 particularly relevant in the face of global climate change.

41

## 42 1 Introduction

43 With the growing demand for safe drinking water, microbial contamination in water resources and its  
44 related diseases is a focus of the world's water quality control today. Numerous studies show that  
45 drinking water with microbial contamination can cause both acute and chronic damage to human  
46 health, with waterborne diseases posing a significant global health burden (Ramirez-Castillo et al.,  
47 2015, Motlagh and Yang, 2019). There has been a rising incidence of waterborne human diseases  
48 such as gastrointestinal illness (e.g., diarrhea) and liver cancers which could be fatal (Zhang et al.,  
49 2010, Yu et al., 2012, Wen et al., 2020).

50 The prevalence of human pathogens and toxin producing bacteria in ambient and drinking water is a  
51 severe problem recognized globally (Pandey et al., 2014). Therefore, the scarcity of water resources  
52 that many countries including China have been facing is further exacerbated by water contamination  
53 (Wang et al., 2021, Wen et al., 2020, Yu et al., 2012). The situation has deteriorated because of the  
54 increasing impacts of warmer temperatures and extreme precipitation events (e.g., rainstorms, floods,  
55 and typhoons) on microbial communities within source and drinking water due to global climate  
56 change (Zou et al., 2023, Cissé, 2019, Yu et al., 2018, Levy et al., 2018, Cann et al., 2013, Brettar  
57 and Hofle, 2008). This year (2023) we are experiencing record high temperatures and heatwaves  
58 globally, including China, USA, and EU (Paddison, 2023, Zachariah et al., 2023). At the same time,

59 we would expect more extreme precipitation events in the future associated with climate change (Zou  
60 et al., 2023). For example, during mid-April 2023, an intense heatwave affected numerous countries  
61 across Asia (China, India, Thailand, and more), shattering temperature records in over twelve  
62 countries – this heatwave led to temperatures reaching between 38 and 45 degrees Celsius (Cai,  
63 2023). In addition, the heavy rainfall during the winter of 2022 set new records for a 22-day period  
64 between October 1 and January 16 in California. Following a succession of destructive winter storms,  
65 data indicates that certain regions of California have experienced rainfall levels that are 200% higher  
66 than the historical average for the year up to that point (Paddison, 2023).

67 A systematic review of China’s drinking water sanitation from 2007 to 2018 shows that microbial  
68 contamination in drinking water is a particular concern in China (Wang et al., 2021). To assess the  
69 potential health risks and obtain necessary data for microbiological water safety management in  
70 China, water monitoring, which supplies a comprehensive understanding of spatiotemporal patterns  
71 in drinking water microbiome, is urgently needed, especially at the point of use (Altenburger et al.,  
72 2019). China CDC (Centers for Disease Control and Prevention) at all levels sample drinking water  
73 twice a year to obtain copious water quality data (Wang et al., 2021). Due to the vastness of China,  
74 this nationwide water monitoring requires considerable investments of capital, time, personnel, and  
75 technology (Wang et al., 2021). Fortunately, previous research has shown that citizen science can be  
76 an effective tool to increase spatial and temporal coverage of data (Pocock et al., 2017). In the  
77 context of China, citizen science could be a cost-effective approach to supplement China’s  
78 professional water monitoring and even dwarfs professional monitoring capabilities (Albus et al.,  
79 2019, Commission, 2020).

80 Citizen science can be broadly defined as a scientific approach in which the public (i.e., people who  
81 have limited knowledge and skill in the targeted field) participate in the generation of scientific  
82 knowledge (Brouwer et al., 2018, Roy et al., 2012, Commission, 2020, Harper, 2018), commonly in  
83 data collection (Brouwer et al., 2018, Roy et al., 2012, Harper, 2018, Commission, 2020). Citizen  
84 science has a history of several centuries in western societies, particularly in the environmental  
85 domain of which the breadth is immense (Albus et al., 2019, Brouwer et al., 2018, Commission,  
86 2020). Among citizen-based environmental monitoring programs, water resources monitoring is one  
87 of the major emerging fields (Albus et al., 2019). It is especially active in Western countries (Roy et  
88 al., 2012, Baalbaki et al., 2019, Buxton et al., 2018, Ho et al., 2020) since the provision of safe  
89 drinking water is a defining aspect of a developed country (Ashbolt, 2015). The National Water  
90 Quality Monitoring Council (NWQMC) website, for example, has over 350 volunteer monitoring  
91 groups registered across the US in 2018 (Baalbaki et al., 2019). A more recent research pointed out  
92 that citizen science will play an increasing important role in promoting freshwater research,  
93 improving public understanding of the necessity to protect aquatic ecosystems, and engaging local  
94 communities and stakeholders in freshwater resource management (Metcalf et al., 2022).

95 However, several research gaps persist. First, compared with the long history and prosperity of  
96 citizen science development in Europe and the US, little research has been done in developing  
97 countries (Baalbaki et al., 2019). This can be attributed to multiple barriers, such as the late  
98 commencement of citizen science initiatives in China, low participation levels, and issues concerning

99 data quality control, etc. Consequently, the cooperation between Chinese scientists and the public are  
100 limited to a few citizen-based environmental projects mainly focusing on birds and plants  
101 monitoring. However, with escalating concerns over environmental issue and the growing prevalence  
102 of big data and social media in China, a new era of citizen science in China is emerging (Zhang et al.,  
103 2013). For example, a recent study conducted by Wu et al. (2022) revealed that most of the citizen  
104 science projects in China aiming to improving water quality are still ongoing, indicating great  
105 potential of the citizen science approach for water monitoring in the country.

106 Second, a global review of citizen science projects related to water quality measurement in the past  
107 20 years (Baalbaki et al., 2019) shows a significant focus on chemical-physical parameters such as  
108 nutrients, water transparency, and temperature, with very few detecting waterborne pathogens.  
109 Among these pathogen assessments, only *E. coli*, an indicator organism of water quality used  
110 worldwide including China (Pandey et al., 2014), was targeted. However, the efficacy of current  
111 indicator organisms in representing the potential presence of pathogens in water resources is still a  
112 subject of ongoing debate (Pandey et al., 2014, Motlagh and Yang, 2019). Additionally, a study  
113 suggests that China should incorporate other microorganisms as alternative indicators to improve its  
114 water quality management (Wen et al., 2020). Therefore, it is necessary to holistically assess  
115 microbial community compositions and pathogens in drinking water.

116 Third, traditional microbiological monitoring of drinking water generally relies on culture-based  
117 methods, such as the heterotrophic plate counts (HPC) of certain microbes (Allen et al., 2004, Garner  
118 et al., 2021). However, these methods can only account for a very small fraction (< 1%) of the  
119 drinking water microbiome (Roeselers et al., 2015, van der Wielen and van der Kooij, 2010, Garner  
120 et al., 2021). To overcome this limitation, this study employed a culture independent approach:  
121 microbial 16S rRNA gene metabarcoding. This method enabled a holistic analysis of the microbial  
122 communities in the water samples (Ramirez et al., 2018, Thompson et al., 2017, Garner et al., 2021).

123 To the best of our knowledge, this study is among the first to collect household drinking water  
124 samples (i.e., tap water) via a citizen science approach across China. To profile the total microbial  
125 communities and waterborne pathogens in 50 tap water samples collected from households in 32  
126 administrative regions spanning 18 provinces/regions of China, we efficiently extracted microbial  
127 DNA from low-biomass tap water and performed high-throughput sequencing of 16S rRNA genes.  
128 The main research questions that this study sought to address are:

- 129 1. Can microorganisms be detected in the sampled tap water? If so, what are the corresponding  
130 spatiotemporal patterns and the driving environmental factors?
- 131 2. More specifically, are waterborne bacterial pathogens detectable within the sampled tap  
132 water? If confirmed, what spatiotemporal patterns do they exhibit, and what are the associated  
133 environmental drivers?
- 134 3. How do extreme precipitation events affect the compositions of total microbial communities  
135 and waterborne pathogens in the sampled tap water?

136

## 137 2 Materials and Methods

## 138 **2.1 Data Collection**

### 139 **2.1.1 Sites and Participants**

140 The sampling sites (Figure 1A) were determined based on the coverage area (i.e., to cover as large an  
141 area as possible) and the possibility to recruit undergraduate student volunteers from the Duke  
142 Kunshan University (DKU) community. Due to the distribution of volunteers, samples mostly came  
143 from central and eastern China (latitude: ~22°N - 40°N; longitude: ~100°E - 122°E), including  
144 Beijing, Shandong Province, Jiangsu Province, Guangdong Province, etc.

145 All student volunteers were recruited after a simple screening process. First, their relevant  
146 experiences (e.g., majors) were considered and those with natural science major and lab experience  
147 were preferred. Besides, ideal volunteers would go home or go on a trip to any place(s) in China  
148 during the sampling period. Volunteers who met those two criteria or at least one of them were  
149 recruited through one-to-one conversation on WeChat (a Chinese social media app). Communication  
150 between volunteers and investigators is a key component of this study to ensure the quality of the  
151 samples as much as possible. This is the main reason why volunteers were recruited from the DKU  
152 community – they can easily reach out to us either in person or via WeChat once any problem  
153 emerges.

154

### 155 **2.1.2 Sample Collection**

#### 156 **2.1.2.1 Preparation and Tool Kit**

157 A sampling protocol was developed based on Buxton et al.'s (Buxton et al., 2018) research on citizen  
158 science methods because volunteers in this study performed highly similar tasks (i.e., water sample  
159 collection and filtration) as their research. A detailed version of the protocol is provided in the  
160 “Citizen science sampling protocol and materials” (CS 1). Two innovative aspects of this protocol  
161 are (1) the easy-to-use and low-cost Corning syringe filters (instead of pumps and Sterivex filter  
162 cartridge) were adopted for water sampling, and (2) the disinfection procedure was emphasized by  
163 listing all the possible exposed object and surface during the entire sampling process. Briefly, it is  
164 recommended to conduct the experiment on the day or at most one day before shipping the sample.  
165 Volunteers first put on the gloves and disinfect their hands as well as everything that they may touch  
166 during the operation using the disinfectant wipes. Then, a sterilized 1 L stand-up bag with sodium  
167 thiosulfate was unsealed and filled with 1 L tap water. A disposable and sterilized 50 mL syringe was  
168 used to pass the sample water across a sterilized Corning syringe filter unit (0.20 µm pore size, 28  
169 mm diameter) and refilled until 1 L of water has been filtered or the filter unit has become blocked.  
170 Afterwards, a syringe of air was pushed through the filter unit to reduce the amount of residual water  
171 in the sealed unit. The filter unit was then sealed in a Ziplock bag and kept frozen in a household  
172 freezer prior to transportation to the laboratory (DKU Environmental Research Lab in Kunshan). For  
173 transportation, the protocol requires burying the filter unit sample among four to five ice packs in a  
174 Styrofoam box. Depending on the distance between the sampling site and Kunshan, along with the  
175 student's mode of travel (i.e., same-day flights or high-speed train), samples were either delivered via

176 express delivery service or personally carried to the laboratory by volunteers.

177 To further clarify the procedures and reduce variability in sample collection, an 11-minute video  
178 tutorial (720p resolution) was created, providing volunteers a visual guide on the tools and  
179 procedures for sample collection, filtration, storage, and shipping (Figure 1B & C and CS 3).  
180 Furthermore, in-person demonstrations were provided at DKU Environmental Research Lab for  
181 available volunteers following the demonstration model of Willis et al. (Willis et al., 2018). A  
182 compact sampling kit was distributed to each volunteer before their trip (Figure 1B). Each kit (in a  
183 78.7×51.2×49.2 inch Styrofoam box) contained one set of all tools mentioned in the protocol and a  
184 sampling information form (CS 1 & 2) adapted from Buxton et al.'s (Buxton et al., 2018). This form  
185 collected information such as volunteer names, sampling time, geographic coordinate of the sampling  
186 site obtained from cell phone, etc. Alongside the form, volunteers were requested to label the Ziploc  
187 bag containing the sample with their names for convenient sample tracking in the stage of data  
188 collection. Upon recipient in lab, each sample was associated with an anonymous ID number and any  
189 data with identifiers was securely eliminated, ensuring no specific identity-related information was  
190 present during data processing. The ID number of each sample consists of two parts separated by an  
191 underscore – a two-letter abbreviation of the sampling site (city) followed by a four-digit sampling  
192 date (Table S1). For example, the sample collected in Beijing on June 20 was named as “BJ\_0620”  
193 and the full names of other mentioned samples are provided in the “Non-standard abbreviations”  
194 section.

195 An online sharing document (Tencent Sharing Document, with English translation provided in CS 3)  
196 was sent to each volunteer via WeChat, including the brief introduction to the research project,  
197 sampling kit, and sampling protocol as well as the links to the video tutorial and the electronic  
198 sampling information form (a Qualtrics survey was created to supplement the printed form). Notably,  
199 key elements in the document were highlighted in red and a larger font size for clarity. All the  
200 materials were provided in Chinese to facilitate understanding among the recruited student volunteers  
201 who were all Chinese. The online sharing document enabled them to easily see any modifications  
202 that were made later.

203 Before the recruitment of volunteers, several trial runs of the preliminary experiment were carried out  
204 by the research team, including sampling sites in Kunshan, Hangzhou, Changzhou, and Beijing.  
205 Along with the sampling protocol and kit, methods and conditions for sample shipping (i.e., express  
206 delivery vs. same-day bullet trains or planes) were tested and compared to identify the most effective  
207 procedures for DNA preservation.

208

### 209 **2.1.2.2 Sample Collection and Filtration**

210 The formal sample collection by student volunteers was conducted from December 2020 to August  
211 2021. At each sampling site, one to four samples were collected. For sites with more than one  
212 sample, all valid samples were included to represent the site's microbiome. With the help of 25

213 volunteers, 50 household drinking water samples were collected from 32 administrative regions  
214 spanning 19 provinces/regions in China.

215 To explore the effects of extreme weather events on water quality, close attention was paid to local  
216 weather forecasts when rainy and typhoon seasons were approaching. Consequently, two extreme  
217 weather events were captured during this study. Tap water samples were collected by volunteers from  
218 Changzhou, Jiangsu Province and Hangzhou, Zhejiang Province before and after the landfall of  
219 Typhoon In-Fa (July 22 – 31, 2021 in China), and from Zhengzhou, Henan Province following an  
220 extremely destructive flood event (2021 Henan Floods).

221 Typhoon In-Fa (number 2106) was a Category 2 typhoon (SSHWS) which has been the second-  
222 wettest tropical cyclone ever recorded in China. As a tropical storm, it consecutively hit Putuo  
223 District of Zhoushan and Pinghu in Zhejiang Province on July 25 and 26 respectively (Wikipedia,  
224 2021). Typhoon In-Fa passed nearby Hangzhou from July 25 to 26 as a typhoon and Changzhou  
225 from July 26 to 27 as a tropical storm (NMC-Typhoon, 2021). On the other hand, 2021 Henan  
226 Floods, as a major part of 2021 Henan Floods (July 17 – 23, 2021), was indirectly influenced by  
227 Typhoon In-Fa. From July 19 to 21, Zhengzhou suddenly encountered the most severe rainstorm of  
228 the last 50 years which has led to extremely severe urban inland inundation, floods, and landslide. In  
229 particular, the 24-hour precipitation in Zhengzhou from 8:00 on July 20 to 8:00 on July 21 was 624.1  
230 millimeters (2.05 ft), which is only about 14.4 mm (0.57 in) lower than the average annual  
231 precipitation in Zhengzhou in the past 22 years (638.5 mm) (Chen, 2021). According to Investigation  
232 report on 2021 Henan Floods issued by Ministry of Emergency Management of the People's  
233 Republic of China (2022), the flood impacted over 14 million individuals and killed 398 people in  
234 Henan, with number of death in Zhengzhou accounting for 95.5% of the total death. Moreover, it  
235 resulted in 16 million hectares of submerged agricultural land and direct economic damages  
236 amounting to \$20.69 billion, accompanied by significantly higher indirect expenses.

237 For the areas affected by the typhoon landfall, student volunteers were instructed to sample tap water  
238 once pre- and post-typhoon respectively within a week. As for Zhengzhou, since the sudden  
239 occurrence of the extreme rainfalls and flooding was not anticipated and predicted by the China  
240 Meteorological Administration (NMC, 2021), pre-flooding sampling was not arranged. However, we  
241 successfully coordinated with and supplied the sampling kits to two volunteers residing in  
242 Zhengzhou - one living in a seriously affected region and the other in a moderately affected region -  
243 during the flood. They were instructed to collect three to four tap water samples on different days  
244 throughout the week following the flood.

245

### 246 **2.1.2.3 Sample Storage Test**

247 Due to the unexpected floods in Zhengzhou, the delivery service in Henan Province was suspended,  
248 and therefore the volunteers were unable to conduct the filtration immediately after they collected the  
249 water samples without the sampling kit. Nevertheless, we asked the volunteers to collect duplicate  
250 tap water samples in unused mineral water bottles with one stored at 4°C and the other at -20°C until  
251 the arrival of the sampling kit. The water samples were stored for 10 days before filtration. To the

252 best of our knowledge, little research has encountered the situation where water samples are filtered  
253 after a long period of storage. Therefore, to understand the impacts of such sample storage under  
254 different temperatures and duration on microbe concentrations in water samples, we conducted a  
255 series of sample storage test in the lab.

256 We first conducted a preliminary test with tap water samples collected from the tap in our lab and  
257 surface water samples from an artificial water feature at DKU. Some of the surface water samples  
258 showed positive PCR signal, but no positive PCR signals existed for any tap water samples in  
259 repeated trials, indicating that environmental DNA is undetectable in DKU tap water. Therefore, to  
260 simulate the potential contamination in tap water, we prepared 2x diluted samples with the tap water  
261 and the surface water collected from DKU and used them for the sample storage test.

262 Three tests have been conducted and could be classified into two sets based on the PCR signal of the  
263 samples with immediate DNA extraction, i.e., positive and negative. In the first test, the surface water  
264 samples collected from DKU water feature were divided into three groups of two samples, and were  
265 stored in the refrigerator for 3 days, 7 days, and 10 days, respectively. The two samples in each group  
266 were stored at 4°C and -20°C respectively. In the second test, new surface water samples were  
267 divided into two groups of two samples with one group stored in the refrigerator for 3 days and the  
268 other stored for 7 days. The two samples in each group were stored at 4°C and -20°C respectively. In  
269 the third test, a water sample collected from Yangcheng Lake was stored at -20°C for 3 days. The  
270 sample filtration, DNA extraction, and PCR amplification were then conducted to analyze the storage  
271 impacts on microbe detection.

272

### 273 **2.1.3 DNA Extraction, PCR Amplification, and Bacterial 16S Metabarcoding**

274 At the DKU Environmental Research Lab, all samples were stored in a -80°C freezer prior to  
275 processing. For each sample, the filter membrane was removed from the sealed syringe filter unit and  
276 cut into eight strips using a sterile razor blade on a disposable petri dish. DNA extraction was then  
277 conducted using the Qiagen DNeasy Plant Mini Kit following the manufacturer's instructions, with  
278 the exception of an additional bead-beating step implemented to enhance extraction efficiency. In  
279 brief, the filter strips were placed into a 2 mL microcentrifuge tube where 0.2 g of 0.1 mm Zr bead  
280 and 400 microliters of lysis buffer AP-1 were added for beat beating at 2000 rpm for 5 minutes. The  
281 final extracted DNA of each sample from the kit was dissolved in 40 µL Qiagen elution buffer and  
282 stored in a -80°C freezer.

283 For PCR amplification, the V4 region of 16S rRNA gene was amplified by the universal primer pairs  
284 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 805R (5'-GACTACNVGGGTATCTAAT-3')  
285 with dual barcode index and heterogenous spacers (Lin et al., 2019, Kozich et al., 2013). KAPA HiFi  
286 PCR Kit and the manufacturer's protocol was adopted. All PCR reactions were performed in  
287 triplicates with 25 µL of each reaction mixture. Agarose gel electrophoresis was then performed to  
288 visualize amplicon fragments and PCR products were purified using QIAquick PCR Purification Kit



289 (Qiagen) following the manufacturer's instructions. The purified PCR amplicons of each sample was  
290 dissolved in 30  $\mu$ L elution buffer and stored in a -80°C freezer.

291 Finally, the concentration of the purified PCR amplicons was measured by Qubit™ 4 Fluorometer.  
292 Equimolar amounts of purified amplicons were pooled together and sent to *Genewiz* in Suzhou for an  
293 Illumina Miseq (250 PE) sequencing run.

294

#### 295 **2.1.4 Auxiliary Data Collection**

296 Previous studies found that drinking water microbiome exhibited seasonality which was correlated  
297 with temperature (Ley et al., 2020, Pinto et al., 2014). Given the difficulties in measuring water and  
298 indoor temperature with the citizen science approach, outdoor temperature was collected as the  
299 indicator of seasonal environmental temperature change.

300 The outdoor temperatures when collecting valid samples (n = 40) were retrieved from Weather  
301 Underground (2022), an online portal that provides local weather data on an hourly basis.

302 In addition, to assess the potential risks of harmful bacteria detected in sampled tap water,  
303 information such as drinking water standards and microbial fact sheets was retrieved from  
304 environmental and health authorities including The United States Environmental Protection Agency  
305 (EPA) (EPA, 2023, EPA, 2022), and World Health Organization (WHO) (WHO, 2017).

306

## 307 **2.2 Data Analysis**

### 308 **2.2.1 Sequence Processing**

309 To profile the microbiome in sampled tap water, 16S rRNA gene sequences were processed by a  
310 series of bioinformatics tools. First, paired-end sequencing reads with dual indices were  
311 demultiplexed and then trimmed to remove barcodes and primers using *Cutadapt* (Martin, 2011). The  
312 resulting reads were then further processed following the *DADA2*(v1.16) pipeline (Callahan et al.,  
313 2016). Specifically, using the embedded functions in *DADA2*, quality filtering was performed before  
314 merging paired-end sequencing reads. After chimera checking, Amplicon Sequence Variant (ASV)  
315 was identified, and the abundance table for each sample was constructed. Finally, the  
316 “assignTaxonomy” function in *DADA2* was used to assign taxonomy to each ASV based on Silva  
317 r138 reference database (DOI 10.5281/zenodo.4587955).

318

### 319 **2.2.2 Assessment of Total Microbial Communities**

320 The relative abundance (RA) of all ASV in each sample was calculated and the following analyses of  
321 microbial compositions were based on RA.

322 Apart for the samples collected after Typhoon In-Fa and 2021 Henan Floods (“post-weather  
323 samples”), the remaining samples (“normal samples”) were categorized in two ways to examine the  
324 spatial and temporal patterns of microbial community compositions. Firstly, due to the vast  
325 geographical scope of the sampling sites, the “normal samples” were divided into seven regions for  
326 region-based analysis (Yang et al., 2021, Xie et al., 2022). These regions, designated as Regions 1-7,  
327 were determined based on the geolocation feature of China, which are Beijing-Tianjin region, North,  
328 Northwest, Central, Central Coast, South Coast, and Southwest respectively (Figure 1). Secondly, to  
329 account for potential impacts of climate and seasonality, the “normal samples” were divided into  
330 three nearly equal groups based on the outdoor temperatures: (a) High (H):  $T > 25\text{ }^{\circ}\text{C}$ ; (b) Medium  
331 (M):  $15\text{ }^{\circ}\text{C} < T \leq 25\text{ }^{\circ}\text{C}$ ; (c) Low (L):  $T \leq 15\text{ }^{\circ}\text{C}$  (Table 1). The mean RA of each ASV in the seven  
332 geographic regions and three temperature categories was calculated for the class- and ASV-level  
333 analyses.

334 To investigate the alpha diversity of tap water microbial communities, each library was resampled in  
335 equal depth, and Chao1, Fisher, Shannon, and Simpson diversity indices were then calculated from  
336 observed read counts of ASVs using the “Phyloseq” package (version 1.38.0) in R (McMurdie and  
337 Holmes, 2013). The Shannon and Simpson indices including both ASV richness and evenness were  
338 computed because of their reduced sensitivity to differences in sample depth (Haegeman et al., 2013,  
339 Preheim et al., 2013).

340 Beta diversity between drinking water microbiomes was calculated based on the ASV table. Beta  
341 diversity was assessed by calculating Bray-Curtis dissimilarity with respect to the RA of each ASV  
342 using “Phyloseq”.

343

### 344 **2.2.3 Assessment of Potential Waterborne Bacterial Pathogens**

345 The bacteria genera in the dataset that containing pathogenic species were selected based on *Aquatic*  
346 *Pollution: An Introductory Text* (Laws, 2017) and “Guidelines for Drinking-water Quality (4<sup>th</sup>  
347 Edition)” by WHO (2017). Subsequently, the taxonomy of the resulting ASVs was further validated  
348 using the BLAST+ tool (Camacho et al., 2009) against the NCBI database during February 2022.  
349 Based on our criteria, only BLAST results with percent identity (p-ident)  $> 97\%$  (Johnson et al.,  
350 2019) and expect value (E-value)  $< 10\text{e-}100$  (Vej, 2007) were considered reliable results.

351 Subsequently, all confirmed pathogen genera/species were grouped into two categories based on their  
352 occurrence in the drinking water samples. Those detected in more than 30% of all samples (i.e., 7  
353 samples) were categorized as “common pathogens” while the rest were referred to as “rare  
354 pathogens.”

355

## 356 **2.2.4 Statistical Analyses**

357 All statistical analyses were performed in R and a p-value threshold of 0.05 was considered  
358 significant.

359 Fisher's exact test (Upton, 1992) was conducted to determine whether there is a significant  
360 relationship between PCR signals (i.e., positive or negative) and average annual rainfall (AAR)  
361 (humid regions: > 800 mm, non-humid regions: < 800 mm). The grouping of sampling sites was  
362 based on Ma et al.'s (Ma et al., 2022) study which obtained AAR data from National Meteorological  
363 Information Center.

364 To analyze alpha diversity of total microbial communities, Dunn's test for a post-hoc Kruskal–Wallis  
365 pairwise comparison (McKight and Najab, 2010) was conducted to identify significant differences  
366 between the seven geographical regions and the three temperature categories. For beta diversity  
367 analysis, a principal coordinates analysis (PCoA) was performed based on the distance matrix of  
368 Bray-Curtis dissimilarity to visualize the ordination among the samples by *plot\_ordination*. To test  
369 the significant differences among groups of tap water samples, permutational multivariate analysis of  
370 variance (PERMANOVA) and a post-hoc pairwise Adonis test were performed with *adonis2* function  
371 (Oksanen et al., 2015) and *pairwise.adonis()* function (Arbizu, 2020) respectively. To check if  
372 within-group variation is confused with among-group variation (Anderson, 2001), a permutation test  
373 for homogeneity of multivariate dispersions (PERMDISP) (Anderson, 2006) was performed with the  
374 *betadisper()* function. In PERMANOVA and PERMDISP, the number of permutations was set to  
375 999.

376 To analyze the compositions and influencing factors of waterborne pathogens in the tap water,  
377 several correlation tests based on Spearman's rank correlation coefficients were conducted. First,  
378 Mantel test (Smouse et al., 1986) was conducted to examine correlations between total microbial  
379 communities (top 100 ASVs), potential pathogens, geographic locations, and outdoor temperatures  
380 based on Bray-Curtis dissimilarity, Euclidean Distance, and Haversine Distance matrices  
381 (permutations = 9999). Second, linear regression fits between outdoor temperatures and RAs of total  
382 potential pathogens were examined and visualized with *ggscatter()* in the "ggpubr" package. Third,  
383 linear correlations between RAs of potential pathogens, RAs of total microbial communities (Top10  
384 ASVs), and alpha diversity of total microbial communities were assessed with *cor()*. The  
385 corresponding p values and confidence intervals were computed with *cor.mtest()*. The matrix with  
386 significance level codes was visualized with *corrplot.mixed()*.

387

## 388 **3 Results and Discussion**

### 389 **3.1 Sample Validity and PCR Signals**

390 Out of the 50 drinking water samples, 40 passed our quality control and they were collected from 28  
391 administrative regions across China (Table S1). 10 samples were excluded from the study because of  
392 either (1) lab processing errors, or (2) improper handling during shipping and/or storage, which may

393 have compromised DNA quality. Among the 40 valid samples, 29 samples showed positive PCR  
394 signals which were collected from 16 cities in 10 provinces, 4 municipalities, and Macau (Figure 2).  
395 Fisher's exact test (Table S2) reveals that no significant difference in PCR signals existed between  
396 the two rainfall groups ( $P = 0.69$ ). Interestingly, all the three tap water samples collected in Zhejiang  
397 Province, including the post-typhoon Hangzhou sample showed no PCR signal. This might be due to  
398 the high quality of source water in Zhejiang Province (Han et al., 2020) and effective drinking water  
399 treatment in those two developed cities (Hangzhou and Ningbo). Alternatively, factors such as  
400 new/clean plumbing systems, suitable plumbing materials that do not support the growth of  
401 microorganisms, and shorter water stagnation time in the plumbing could also account for the  
402 negative PCR results (Ley et al., 2020, Ji et al., 2015, Ling et al., 2018).

403

### 404 **3.2 Sample Storage Condition Influence**

405 In the first test (Figure S1A), the water sample showed negative PCR signal after the immediate  
406 extraction. This test aimed to investigate whether the storage condition could sustain the original  
407 state of the samples and whether the storage condition would lead to the proliferation of bacteria.  
408 "LC" represented storage at 4°C, and "LD" represented storage at -18°C. 8.9 represented water  
409 samples stored for 3 days before filtration, 8.12 represented water samples stored for 7 days, and 8.16  
410 represented water samples stored for 10 days. The result showed that 8.9 LC, 8.16 LC, and 8.16 LD  
411 showed positive signal in PCR, and no patterns can be detected from this result. If 3-day storage at  
412 4°C could lead to bacteria proliferation, 7-day storage at 4°C was also expected to have positive  
413 signal, which did not happen in this test. Since the storage and DNA extraction was conducted  
414 separately for each storage condition, the result may also be caused by operational mistakes. These  
415 suggested that the result of the first test was unreliable and no conclusions can be made. As a result,  
416 a second test was conducted, and the result was presented in Figure S1B.

417 In the second test, a clear pattern was revealed: if at the beginning there were no bacteria in the  
418 samples, then storage at -18°C would not lead to the growth of bacteria. But storage at 4°C failed to  
419 maintain the original state, and it was clear that after 3 days and 7 days, the original microbial  
420 community composition in the samples have been altered. This indicated that samples stored at 4°C  
421 could not be used due to proliferation. Although the second test could prove that -18°C was sufficient  
422 to prevent bacteria growth when originally there were no bacteria, it is unsure when there were  
423 bacteria at the beginning, whether -18°C storage could influence the normal cell function. Therefore,  
424 a third test was required to examine this relationship. In the third test, after 3-day storage at -18°C, a  
425 much brighter band was observed compared to the band of immediate extraction (Figure S1C). 2  
426 possible explanations can be given to this phenomenon: 1) bacteria proliferated at -18°C; 2) freeze-  
427 thaw operations led to cell fragmentation. The first explanation was less likely since the second test  
428 has proved that -18°C storage could prevent bacterial growth.

429

### 430 **3.3 Sequence Reads, ASVs, and Taxonomy Classification**

431 The DNA of microorganisms in 29 household drinking water samples were successfully extracted,  
432 amplified and submitted for sequencing. However, following the demultiplexing process, five  
433 samples did not yield significant reads (all < 50 per library) and were excluded from the dataset,  
434 leading to a loss of two sampling sites (Shenzhen and Xiamen). In addition, two samples collected  
435 after the 2021 Henan Floods were excluded after the quality filtering and chimera checking because  
436 of relatively low reads (<3000 per library) compared to others. As a result, the 16S rRNA gene  
437 amplicon dataset in this study contained 22 samples after the preliminary processing (Table S1).

438 In DADA2 quality filtering, proportions of output and input reads of 13 (59.1%) samples were lower  
439 than 50%, indicating low quality of raw DNA reads. This may be due to 1) potential degradation of  
440 DNA samples during the shipping or DNA processing; 2) detectable DNA from dead bacteria cells,  
441 which are harmless to humans. After quality filtering and chimera removing, sequencing reads per  
442 sample ranged from 9,794 to 150,656, averaging at 60,884.

443 From the resulted dataset, 7635 ASVs were identified. According to the DADA2 taxonomy  
444 classification based on Silva r138 reference database, 100%, 99.5%, 96.8%, 92.3%, 81.8%, and  
445 59.4% of the sequencing reads could be assigned to kingdom, phylum, class, order, family, and genus  
446 level respectively. At the kingdom level, 99.9% of the reads were assigned to domains of bacteria  
447 (1,294,678 reads) and archaea (43,816 reads), while the other 0.1% of the sequences belonged to  
448 eukaryotes (739 reads).

449

### 450 **3.4 Total Microbial Communities**

#### 451 **3.4.1 Compositions of Total Microbial Communities**

##### 452 **3.4.1.1 Phylum Level Pattern by Sample**

453 52 microbial phyla including 46 bacterial phyla and 6 archaeal phyla were detected in the household  
454 drinking water samples (n = 22). In all samples, 9 bacterial phyla and 1 archaeal phylum accounted  
455 for over 90.7% of total taxonomically assigned reads at phylum level (Table 2). All the ten phyla are  
456 both abundant and prevalent (i.e., occur in 91-100% of all samples), indicating a relatively even  
457 microbial distribution pattern in the samples at the Phylum level. The five most dominant bacterial  
458 phyla were *Proteobacteria* (mean percentage  $\pm$  SD: 55.0%  $\pm$  19.8%), *Planctomycetota* (10.5%  $\pm$   
459 7.9%), *Acidobacteriota* (7.0%  $\pm$  6.0%), *Actinobacteria* (5.9%  $\pm$  6.6%), and *Cyanobacteria* (4.7%  $\pm$   
460 3.5%), comparable to a previous study in China (Han et al., 2020). Three of them were reported  
461 tolerant to drinking water treatment and distribution processes except for *Planctomycetota* and  
462 *Acidobacteriota* (Han et al., 2020). Interestingly, this bacterial composition differs from the primary  
463 bacterial assemblages revealed by several highly-cited studies conducted in other countries including  
464 the U.S. and Portugal, which predominantly feature *Proteobacteria*, *Actinobacteria*, and  
465 *Bacteroidetes* (Hull et al., 2017, Pinto et al., 2012, Ivone et al., 2013). This implies a unique  
466 microbiome intrinsic to China's drinking water systems.

467 The phylum-level taxonomic composition for each sample is detailed in Figure 3 and Table S3. In  
468 this study, *Proteobacteria* (class  $\alpha$ - and  $\gamma$ -*Proteobacteria*) was the most predominant phylum in 20  
469 samples and the second most in others. It accounts for 29.9 – 99.0% of the reads in each sample, with  
470 the highest relative abundance detected in Huizhou (HZ\_0219). Among all the samples,  
471 *Sphingomonas* was the most abundant genus of *Proteobacteria*. This result was consistent with the  
472 previous study (Han et al., 2020) that found *Proteobacteria* to be dominant in tap water collected  
473 mostly from central and eastern China, and the genus *Sphingomonas* grew during chlorination (Jia et  
474 al., 2015) or monochloramine treatment (Chiao et al., 2014).

475 However, two samples CZ\_0220 (Changzhou, Feb. 20) and NC\_0210 (Nanchang, Feb.10) were  
476 dominated by *Crenarchaeota*, a common archaeal phylum. Specifically, *Crenarchaeota* accounted  
477 for 33.3% and 35.0% with the genus *Candidatus Nitrosotenuis* (32.0%) and *Candidatus Nitrosotalea*  
478 (34.9%) being most abundant in samples CZ\_0220 and NC\_0210 respectively. This showed that in  
479 addition to a variety of bacteria, archaea can also grow in tap water, which could be supported by the  
480 studies which detected the archaeal phylum *Crenarchaeota* in drinking water distribution systems or  
481 drinking water-related environments (Roeselers et al., 2015, Inkinen et al., 2021, Dai et al., 2020,  
482 Franca et al., 2015, Bautista-de los Santos et al., 2016). In particular, Inkinen et al. (Inkinen et al.,  
483 2021) found a high abundance of archaeal reads from the genus *Candidatus Nitrosotenuis* and  
484 *Candidatus Nitrosotalea* in drinking water distribution systems supplying non-disinfected waters.  
485 This suggests that the disinfection processes of samples CZ\_0220 and NC\_0210 may be less  
486 effective compared to others.

487 Interestingly, compared to CZ\_0220, the pre- and post-typhoon samples collected from the same city  
488 household in Changzhou in July were much more similar in composition at the phylum level.  
489 However, the post-typhoon sample (CZ\_0728) was richer in *Actinobacteria* (increased from 7.1% to  
490 2.2%) and *Cyanobacteria* (increased from 5.2% to 14.9%) which are the phyla containing potential  
491 waterborne pathogens and the species that produce cyanotoxins respectively. Elevated levels of the  
492 pathogen *Mycobacterium* spp. (more details in Section 3.5 and Supplementary Material: SM1.1) as  
493 well as toxin producing *Cyanobacteria* spp. were identified in CZ\_0728. Specifically for  
494 cyanobacteria, *Microcystis* spp. were higher in RA while *Cylindrospermopsis* sp. and  
495 *Dolichospermum* sp. appeared after the typhoon event. Moreover, other toxic species of  
496 *Cyanobacteria* including *Aphanizomenon* sp. and *Anabaena* sp. were detected in other tap samples  
497 collected from Shanghai, Lanzhou, Xi'an, etc. Many *Cyanobacteria* spp. from those genera can  
498 produce a variety of cyanotoxins such as Microcystins and Cylindrospermopsin which can cause  
499 damage to the liver and kidney with potential carcinogenicity (EPA, 2022). Similarly, a substantial  
500 rise in the RA of *Cyanobacteria* occurred in treated water samples collected from a drinking water  
501 treatment plant in Jiangsu was reported after Typhoon Lekima in August 2019 ( $P < 0.05$ ) (Tang et al.,  
502 2021). In this study, although remained detectable, the proportion of *Cyanobacteria* decreased to the  
503 pre-typhoon level by the third day after the typhoon event.

504

### 505 **3.4.1.2 Class Level Taxonomy**

506 The geographical and temporal distribution of microbial classes in household drinking water in China  
507 is illustrated in Figure 4. The mean RAs of the top 10 microbial classes in all the normal samples,  
508 seven regions, and three temperature categories are summarized in Table S4. The analysis of the top  
509 20 ASVs (mean RA) in each geographical region and temperature group is shown in Figure S2 & SM  
510 1.1 and the detailed taxonomy classification of each ASV is provided in Table S5.

511 Among all the regular samples, the 10 most frequently detected classes were *Alphaproteobacteria*  
512 (mean RA: 37.4%), *Gammaproteobacteria* (16.3%), *Planctomycetes* (8.0%), *Blastocatellia* (5.7%),  
513 *Actinobacteria* (4.2%), *Nitrososphaeria* (3.6%), *Bacteroidia* (2.6%), *Verrucomicrobiae* (2.2%),  
514 *Vampirivibrionia* (2.1%), and *Cyanobacteriia* (1.8%). While many classes were shared across  
515 various regions, their prevalence varied (Figure 4A). Specifically, classes such as  
516 *Alphaproteobacteria*, *Gammaproteobacteria*, *Blastocatellia*, and *Planctomycetes* dominated all the  
517 seven regions, yet their RA differed. Notably, *Alphaproteobacteria* exhibited much higher RA in  
518 samples from R1: Beijing-Tianjin region (59.7%) and R6: South Coast (50.0%) than in those from  
519 R3: Northwest (25.6%) and R5: Central Coast (28.2%). On the other hand, some classes only  
520 predominated a few regions (Figure 4A). For example, *Methylomirabilia* and *Babeliae* were among  
521 the top 10 classes only in R4: Central (1.6%) and R6: Central (1.0%), respectively.

522 Similar to the region-based analysis, the three temperature categories shared many dominant classes  
523 (top 2:  $\alpha$ - &  $\gamma$ -proteobacteria), but a few classes were found to be abundant in only certain categories  
524 (Figure 4B). For instance, *Nitrososphaeria* was the fourth abundant class in Medium T and Low T,  
525 but not dominated High T. Besides that, *Cyanobacteriia* (2.3%) and *Acidimicrobiia* (1.6%) were only  
526 among the top 10 classes of High T whereas *Phycisphaerae* (2.1%) and *OM190* (1.8%) were only  
527 abundant in Low T.

528 Multiple factors could account for the above differences in the ultimate microbial compositions at the  
529 tap. For example, local geology, the source water properties, water treatment & distribution  
530 processes, (Ji et al., 2015) etc. In addition, historical events that happened in the sampling region,  
531 such as agricultural runoff, could also explain specific patterns of microbial communities (Holinger  
532 et al., 2014).

533

### 534 **3.4.2 Alpha-and Beta-Diversities**

535 Alpha-and beta-diversities have been calculated at the ASV level to further describe the microbial  
536 communities in the samples. Alpha diversity refers to diversity (richness or evenness) within an  
537 ecosystem at the local small scale, while beta diversity measures the amount of differentiation  
538 between different ecosystems or local species communities (Andermann et al., 2022). The alpha  
539 diversity indices among different geographical regions and outdoor temperature categories as well as  
540 the weather samples were calculated (Figure 5 and Table S6). Tap water samples from R3: Northwest

541 & R7: Southwest and High T category (Shannon and Simpson) constituted the most diverse  
542 microbial community while samples from R6: South Coast & R1: Beijing-Tianjin Region and  
543 Medium T category had the lowest alpha diversity. Though all the diversity indices showed some  
544 differences, none of the differences were statistically significant ( $P$  values  $> 0.05$ , adjusted by  
545 multiple methods in Dunn's test). Overall, the alpha diversity of the tap water microbial communities  
546 was not largely region-dependent, and the outdoor temperature was not an influential environmental  
547 driver of the alpha diversity. This may be largely due to the small sample size and the long sampling  
548 period (from late January to late August in 2021, but the sampling events concentrated in February  
549 and July).

550 The extreme rainfall events had a clear impact on the alpha diversity of the microbiome in tap water  
551 samples collected from Changzhou and Zhengzhou, the two cities impacted by extreme rainfall  
552 events in 2021 summer. Compared to the two samples collected in winter (CZ\_0220 & ZZ\_0219),  
553 higher alpha diversity was observed in the post-weather summer samples in both cities (Figure 5C).  
554 Among the three samples collected from the same household in Changzhou, however, the highest  
555 diversity was observed in the pre-typhoon summer sample CZ\_0712, indicating that outdoor  
556 temperatures might be a stronger driver of the water microbiome diversity than the typhoon.

557 Outdoor temperature plays a significant role in shaping the community composition and structure at  
558 the ASV level. The beta diversity of the total microbial communities in sampled tap water is shown  
559 in Figure 6. Principal Axis 1, 2, and 3 for PCoA (Bray-Curtis dissimilarity) represent 20.4%, 14.5%,  
560 and 8.1% of the variation among the samples respectively. The PCoA ordination illustrates that  
561 despite the different sampling regions, the samples in High T category clustered closely except  
562 BJ\_0620 which was greatly dominated by *Proteobacteria* (87.5%). In contrast, samples in the other  
563 two categories, especially Medium T exhibited dispersive distributions. According to PERMANOVA  
564 and PERMDISP (Table S7), the influence of outdoor temperatures was statistically significant ( $R^2 =$   
565 0.24, adjusted  $P = 0.012$  between High T and Low T), and the differentiation was not due to  
566 differences in group dispersions ( $P = 0.17$ ). However, when making comparisons among normal  
567 samples (Figure S3 and SM 1.2), samples from different regions were greatly mixed and scattered  
568 except for the two samples in R7: Southwest (CQ\_0715 & LJ\_0722), which corresponds to the  
569 results of top 20 ASV analysis (Figure S2). A Mantel test (Table S8) indicate a strong correlation  
570 between microbial community structures and outdoor temperatures (Spearman's  $r = 0.898$ ,  $P =$   
571 0.0001), whereas geographic locations had minimal impact (Spearman's  $r = -0.179$ ,  $P = 0.9792$ ).

572 Research spanning China, the Netherlands, and the U.S. has consistently found that temporal  
573 temperature variation, primarily driven by seasonal changes, is a controlling factor for the tap water  
574 microbiome (Zhang et al., 2021b, Zlatanović et al., 2017, Ley et al., 2020). Complementing this,  
575 another study in China revealed that increasing air temperature through indoor heating could lead to  
576 dramatic changes in the composition of the bacterial community in overnight stagnant tap water  
577 (Zhang et al., 2021a). These findings challenge the previous discovery by Han et al. that only a weak  
578 correlation was found between air temperature and tap water bacterial community in China  
579 (Spearman's  $r = 0.088$ ,  $P = 0.092$ ) (Han et al., 2020). The inconsistency might be caused by  
580 differences in sampling year, sites, and geography coverage. Notable, the source water microbiome



581 could exert a substantive influence on tap water microbiome (Han et al., 2020). On the one hand, tap  
582 water in the same city may need different treatment processes depending on the source water. On the  
583 other hand, adequate source water protection and control strategies are necessary for the prevention  
584 of microbial contamination from source water carried over to tap water.

585 The pre-typhoon summer sample CZ\_0712 was distinct from the other two Changzhou samples but  
586 much like other samples in High T category, which further confirmed the impact of high outdoor  
587 temperature. Comparing the samples collected from Changzhou and Zhengzhou, the pre-weather  
588 winter samples differed moderately in microbial community composition and structure, however, the  
589 post-weather samples from the two cities (purple circle and square in Figure 6) were comparable to a  
590 large extent. This indicates a potential contamination mechanism – extreme rainfall events may cause  
591 the same or highly similar shift in tap water microbiomes by favoring certain fast-growing bacteria,  
592 increasing the similarity between dissimilar tap water microbial communities.

593 The above patterns observed in the drinking water microbiome suggests that outdoor temperature and  
594 extreme precipitation are important environmental factors for microbial community compositions and  
595 structures. Given the significant impact of temperature, *in situ* water temperature measurement  
596 should be incorporated in future citizen-based studies (Ley et al., 2020). This could be done by  
597 adding a portable thermometer to each sampling kit. Beyond this citizen science approach, other  
598 drivers should be systematically examined in the future for effective drinking water management. In  
599 addition, the division of Regions 1-7 in this study is subjective to some extent. A potentially more  
600 insightful categorization might focus on drinking water source types: river, lake and reservoir, and  
601 groundwater (Zhang et al., 2022). In a recent study on nationwide biogeography of bacterial  
602 communities in household drinking water in China, they found that source rivers (i.e., major river  
603 systems such as Yangtze River) and precipitation (i.e., local averaged annual rainfall) can also drive  
604 microbial dissimilarity of drinking water (Ma et al., 2022).

605

### 606 **3.5 Occurrence of Potential Pathogenic Bacteria in Drinking Water Microbial Communities**

607 In total, six bacteria genera containing pathogenic species and three pathogenic species were detected  
608 in all the PCR positive samples (n = 22) (Table 3). Five genera and one species that occurred in more  
609 than 30% of the samples (i.e., 7 samples) were categorized as common pathogens while the rest were  
610 grouped into rare pathogens (Table S9A). It is worth mentioning that the overall pattern of pathogens  
611 in the normal samples still holds when the two post-weather samples were included. This finding  
612 suggests that pathogen contamination in tap water could be a widespread phenomenon, regardless of  
613 severe weather events such as extreme rainfall (Ley et al., 2020, Ma et al., 2022).

614 The distribution of these potential pathogens within each water sample are detailed in Figure 7.  
615 Common pathogens' mean relative abundance (RA) in all tap water samples ranged from 0.02% to  
616 2.67% (normal samples: 0.02% - 1.95%) while that of rare pathogens was extremely low (Table  
617 S9B&C). *Mycobacterium* spp. (mean RA 2.67%, including rainfall events), *Acinetobacter* spp.  
618 (1.43%), and *Legionella* spp. (0.46%) occurred in all the samples while *Leptospira* spp. (0.02%) were

619 found in half of the samples. Notably, *Escherichia coli*. (0.06%) that might be *E. coli* O157:H7  
620 according to BLAST – the most common pathogenic strain that causes severe gastrointestinal (GI)  
621 illness in humans (Lim et al., 2010) – were detected in 36.4% of the samples. In addition, two  
622 *Brevundimonas* species, *B. vesicularis* and *B. diminuta*, are particularly considered emerging global  
623 opportunistic pathogens (Ryan and Pembroke, 2018) were found in most of the samples (90.9%).  
624 Compared to normal samples, three more pathogenic bacteria species were detected in the post-  
625 weather samples, including *Salmonella enterica*, *Brevundimonas diminuta*, and *Aeromonas*  
626 *hydrophila*.

627

### 628 **3.5.1 Representative Pathogen Species**

629 Since only the V4 region of the 16S rRNA gene was sequenced in this study, most amplicon  
630 sequences were not able to provide adequate phylogenetic resolution for classification at the species  
631 and strain level. Therefore, some ASVs are not 100% confirmed pathogenic species/strains, but  
632 highly possible (see Table S10). A summary of representative potential pathogens in drinking water  
633 systems and their related diseases is provided in Table S11 (Ramirez-Castillo et al., 2015). A  
634 discussion of the following three as well as two other representative pathogen species that is  
635 primarily for the science communication purpose as provided in SM 1.3.

636 *E. coli*, a common fecal indicator bacteria (WHO, 2017), was detected by this study. The most  
637 polluted tap water samples are CZ\_0728 (RA: 0.61%), BJ\_0620 (0.51%), ZZ\_0802 (0.05%),  
638 CZ\_0712 (0.03%), and JN\_0219 (0.02%). The elevated RA of *E. coli* in the two post-weather  
639 samples suggests that the contamination might be related to the extreme rainfall events. Notably, the  
640 proportion of *E. coli* in BJ\_0620 was dramatically higher than that of other normal samples. It is  
641 worth mentioning that the traditional *E. coli* tests can be a generally reliable indicator of  
642 enteropathogenic serotypes in drinking water, but potential viable while non-culturable *E. coli* cells  
643 could result in underestimations of the actual water contamination (Liu et al., 2008). Therefore, it is  
644 recommended to use PCR or quantitative PCR (qPCR) methods for the monitoring of *E. coli* (Wolf-  
645 Baca and Siedlecka, 2019).

646 Additionally, the RA of total *Legionella* spp. was most abundant in ZZ\_0802 (1.32%), CZ\_0220  
647 (1.25%), and ZB\_0502 (1.09%). Almost all species in the genera *Legionella* are thought to be  
648 potential human pathogens, but *L. pneumophila* (on Contaminant Candidate List 5 - CCL 5) is the  
649 main cause of Legionnaires' disease (pneumonia) and Pontiac fever (a milder infection) (WHO, 2017,  
650 EPA, 2023). Potential *L. pneumophila* ASVs (Blastn best hit results) was detected in 22.7% (5) of the  
651 tap water samples (3 from Low T, 1 from Median and High T respectively). Among those samples,  
652 the highest *L. pneumophila* RA was observed in ZZ\_0219 (RA: 0.237%), followed by MA\_0524  
653 (0.035%), NC\_0210 (0.019%), XA\_0217 (0.011%), and SH\_0411 (0.010%). Moreover, some other  
654 pathogenic species such as *L. oakridgensis* and *L. maceachernii* occurred in TJ\_0715, LZ\_0822,  
655 XA\_0217, LJ\_0722, and ZZ\_0802.

656 *Salmonella* spp., a highly pathogenic bacteria, was detected in one sample despite its known  
657 susceptibility to disinfection. Notably, *Salmonella enterica* (ASV 3082, RA: 0.39%) was observed in  
658 the post-flood tap water microbiome from Zhengzhou (ZZ\_0802). However, the potential health risks  
659 remains uncertain, as the severity of the disease depends on the serotype and host factors of  
660 *Salmonella* (WHO, 2017). Nonetheless, the presence of *Salmonella enterica* after 2021 Henan Floods  
661 indicates potential fecal contamination in the household drinking water after an extreme weather  
662 event.

### 663 664 **3.5.2 Potential Environmental Drivers: Extreme Precipitation and Hot Weather**

665 The detection of a wide range of pathogens in tap water indicates the great impacts of extreme  
666 rainfall events on urban water management. Abnormal redistribution of precipitation patterns or  
667 extreme rainfall events could cause more rainfall to be concentrated in a short period, resulting in  
668 flooding and a large amount of effluent. These events can lead to huge challenges to urban water  
669 safety and pose severe threats on public and environmental health. On the one hand, excessive  
670 effluent load on urban wastewater treatment can overwhelm wastewater treatment plants, leading to  
671 untreated sewage being released into natural waterways, and resulting in freshwater resource  
672 pollution and aquatic ecosystem degradation (Langeveld et al., 2013). On the other hand, destructive  
673 rainfall could pollute drinking water sources and cause damage to buildings and infrastructure, such  
674 as leaks and breakage of tap water distribution systems. This will eventually lead to drinking water  
675 contamination and increased infection risks, which has been supported by the analysis of microbial  
676 community and pathogen changes in tap water in this research. As shown in Table S9, the typhoon  
677 and flood events generally altered the pathogen profiles in tap water samples by increasing the  
678 relative abundance of certain species and the number of species. Specifically, the RA of the total  
679 potential pathogens in the two post-weather samples ranked second (CZ\_0728, 20.7%) and third  
680 (ZZ\_0802, 10.7%) among all the samples (Figure 7). In addition, the number of pathogen species in  
681 Changzhou samples collected in July increased after Typhoon In-Fa. A similar pattern was observed  
682 in Zhengzhou samples and ZZ\_0802 was the one with the highest number of pathogen species. More  
683 importantly, some devastating pathogen species (*Salmonella* spp. and *Aeromonas hydrophila*) were  
684 only detected in the post-weather samples as discussed above. In terms of the elevated RA of specific  
685 genera/species, most of the pathogens in the post-typhoon sample CZ\_0728 had higher proportions  
686 compared to the pre-typhoon summer sample CZ\_0712, except for *Legionella* spp. and  
687 *Brevundimonas* spp. which saw a decrease in RA, and *Bacillus* sp. being not detected in CZ\_0728.  
688 Therefore, the rising risk of waterborne pathogens in post-flush tap water samples indicates that  
689 extreme rainfall events are a prominent risk factor in public health in terms of water (and food) safety  
690 (Phan and Sherchan, 2020). To maintain the standard of the treated discharge water during extreme  
691 rainfall events in the future, the government needs to invest more to upgrade the water processing  
692 capacity of wastewater treatment plants to be prepared for unpredictable effluent load. Meanwhile,  
693 since drinking water resource pollution is expected to happen following flooding, drinking water  
694 treatment plants and distribution systems should be designed, monitored, and maintained involving

695 emergency response planning to secure their capacity for resisting the negative impact of extreme  
696 weather events.

697 Hot weather (increase of temperature) is another critical climatic factor that will strongly influence  
698 the hydrological regime and facilitate the spreading and proliferation of waterborne pathogen in  
699 freshwater resources. According to IPCC Technical Paper VI: Climate Change and Water (IPCC,  
700 2008), the increase of surface water temperature has been associated with the atmospheric warming  
701 across Asia, North America, and Europe since 1960s. Lipp et al. has discovered that increasing  
702 temperature could create conditions that are advantageous for the proliferation of bacteria species less  
703 susceptible to temperature variations (Lipp et al., 2002). This could directly trigger the growth and  
704 expansion of some pathogenic species, such as *cholera* (*V. cholerae*), *Vibrio parahaemolyticus*, *V.*  
705 *vulnificus* and *V. alginolyticus* (Funari et al., 2012). Our results confirmed the above connections  
706 between climate change and increased waterborne infection risk by showing that hot weather is  
707 potentially another important environmental driver of pathogen profiles of tap water. The outdoor  
708 temperatures higher or equal than 17°C were positively correlated with RAs of total potential  
709 pathogens but the correlation is insignificant relative to the alpha level of 0.05 probably due to the  
710 small sample size ( $n = 12$ ,  $R = 0.46$ ,  $P = 0.13$ ) (Figure S4). Besides that, higher RAs and a larger  
711 number of pathogen species were detected in the summer Changzhou sample CZ\_0712 (20.7%, 7  
712 genera, outdoor T: 33.3°C) compared to the winter sample CZ\_0220 (1.80%, 4 genera, outdoor T:  
713 17.0°C).

714 An important aspect to highlight here is that, apart from the separate effects of extreme rainfall and  
715 hot extremes on waterborne disease spreading, these extreme weather events are also intercorrelated  
716 in a way that creates a detrimental cycle, exacerbating the overall situation. IPCC AR6 report has  
717 confirmed that regardless of the emissions scenarios, the Earth's surface temperature will keep  
718 increasing until at least the mid-21st century. Unless substantial reductions in CO<sub>2</sub> and other  
719 greenhouse gas emissions occur in the coming decades, the global warming limits of 1.5°C and 2°C  
720 will be exceeded during the 21st century. Most importantly, the report underlined the fact with high  
721 confidence that with every additional increment of global warming, the impact on extreme weather  
722 events becomes more pronounced. For instance, for each additional 0.5°C rise in global temperatures,  
723 there are noticeable and undeniable rises intensity and frequency of extreme heat events, including  
724 heatwaves and intense rainfall (Masson-Delmotte et al., 2021). As a result, ensuring tap water safety  
725 in the future will become increasingly challenging and require stricter control and maintenance as the  
726 effects of climate change continue to intensify.

727

### 728 **3.5.3 Potential Monitoring and Control Strategy**

729 Spearman's correlations between potential pathogens and alpha diversity indexes are shown in Figure  
730 8. When the impacts of extreme rainfall events were not considered (Figure 8A,  $n = 22$ ), the RA of  
731 *Brevundimonas* spp. and that of *Mycobacterium* spp. was positively correlated ( $r = 0.84$ ,  $P < 0.001$ ),  
732 and both were positively correlated with the total RA of potential pathogens ( $r = 0.60$ ,  $0.67$ ,  $P <$   
733  $0.001$ ). Notably, a significant correlation between the low alpha diversity of the microbial community

734 and the samples positive for the presence of *E. coli* (Chao1:  $r = -0.24$ ,  $P < 0.05$ ; Shannon:  $r = -0.21$ ,  $P$   
735  $< 0.05$ ) was observed. Similarly, Chopyk et al. (Chopyk et al., 2016) found the same pattern in their  
736 pre-harvest cattle hide samples. Therefore, it is worth understanding the relationship between  
737 indigenous water microbiomes and pathogenic *E. coli*, which might contribute to the  
738 control/prevention strategies for EHEC in tap water as well as indicators for risk assessment.

739 The incorporation of the two post-weather samples (Figure 8B,  $n = 24$ ) altered the correlations  
740 between different pathogens and revealed new relationships due to the occurrence of other species in  
741 the samples. The positive correlation between the RA of *Brevundimonas* spp. and that of  
742 *Mycobacterium* spp. became weaker ( $r = 0.68$ ,  $P < 0.05$ ) while the RA of *Aeromonas hydrophila* was  
743 significantly correlated with *S. enterica* ( $r = 0.72$ ,  $P < 0.01$ ). Interestingly, *S. enterica* and *B. diminuta*  
744 were only detected in the post-flood sample ZZ\_0802, suggesting potential co-occurring relationships  
745 between them. Future studies could further investigate the above potential relationships among  
746 waterborne pathogens in household drinking water.

747

### 748 **3.6 Citizen Science Approach**

749 This study is among the few studies on tap water microbiome via citizen science which can serve as a  
750 proof of concept for national-scale microbiological monitoring of tap water using citizen science and  
751 demonstrate its competitive advantages compared to non-citizen science sampling method. Firstly,  
752 the adopted citizen science approach was greatly cost-effective, especially in the context of China.  
753 With a relatively small investment in time, money, and personnel, the study had a large spatial and  
754 temporal coverage. Specifically, the sampling sites covered 32 administrative regions spanning 19  
755 provinces/regions in China and different seasons from winter to summer. With more volunteers  
756 participate in the sampling, larger coverage and higher accuracy of monitoring will be obtained. It is  
757 worth mentioning that due to the adoption of Corning® syringe filter (~\$2.5–4/unit) (Corning, 2023),  
758 the expense of per sampling kit was under \$7, which was around 1 – 1.5 times cheaper than using the  
759 traditional MilliporeSigma® Sterivex pressure filter (~\$8–13/unit) (MilliporeSigma, 2023). Our  
760 DNA extraction protocol was new and optimized for this low-cost filter. In addition, a large amount  
761 of travel expense and time was saved. Although the processes of sample collection and filtration  
762 performed by the volunteers were not closely supervised, the high validity of the samples (40 out of  
763 the 50 drinking water samples successfully passed the quality control) and the above consistencies in  
764 results suggest that our citizen science approach was largely applicable and reliable.

765 Secondly, we conducted serendipitous studies on Typhoon In-Fa and 2021 Henan Floods with the  
766 help of four local volunteers in Hangzhou, Changzhou, and Zhengzhou. Given the limited funding  
767 and the emergency of the events, it would be logistically difficult to investigate extreme weather  
768 events without using this citizen science approach. The result of our storage test showed that if  
769 immediate filtration of water samples is not feasible,  $-18^{\circ}\text{C}$  storage for up to 7 days will not lead to  
770 proliferation of bacteria if there are no bacteria in the water samples upon collection. Therefore, our  
771 sampling method will not lead to false positive results if the storage time is controlled within 7 days.

772 Although -18°C storage will lead to bacteria proliferation if there are bacteria in the water samples  
773 upon collection, this proliferation will not influence the identification of the existence of pathogens.

774 Thirdly, the method is not restricted by geographic locations, and therefore can be applied to most  
775 countries upon some modifications in actual implementation. In the U.S., using this simple sampling  
776 kit, community water quality data can be easily collected by citizen scientists and can be potentially  
777 incorporated into EPA database as a supplement to official data. In addition, the study also tried to  
778 build up a bridge between the science community and the public to engage the public in science  
779 projects, enhance public awareness of pathogen pollution in drinking water, and inform them how to  
780 protect their health. As a reward for volunteers helping us collect water samples, we sent the water  
781 quality report to them afterwards.

782

### 783 **3.7 Other Strengths, Limitations, and Future Work**

784 Combining citizen science sampling and culture independent metabarcoding, this study provided  
785 comparatively comprehensive profiles of total microbial communities and waterborne pathogens  
786 from 18 provinces of China – many of which are undetectable by HPC. Therefore, it demonstrated  
787 that the actual tap water contamination is generally underestimated by conventional monitoring  
788 methods.

789 However, the approach also has some limitations and can be improved in various ways. First, despite  
790 the result consistencies with previous studies that were not citizen-based, to further minimize the  
791 sample collection variability, closer supervision of volunteers (e.g., video recording), duplicate  
792 sample collection, and more reliable shipping methods are needed. Due to limited funding, we only  
793 collected tap water samples from one household in each representative city as a proof-of-concept.  
794 However, one city may have several different water supply systems which serve different  
795 communities and districts. Therefore, the samples in our study cannot reflect the actual full picture of  
796 the water quality of the city. In future studies, it is better to choose several sampling points covering  
797 the major water systems within each city. Similarly, we were not able to examine the potential  
798 temporal pattern of the microbial communities within a household due to the low sequence reads of  
799 two summer samples from the same household in Shanghai and Nanjing, respectively. To better  
800 understand the temporal variations, further studies should conduct more systematic time-series  
801 collection of tap water samples from the same location. In addition, since our study took samples  
802 from household water taps, we were not able to identify the source of the pollution if harmful  
803 pathogens were detected. For future studies that aim to find out the source of bacteria, collecting  
804 water samples from the effluent of the drinking water treatment and distribution system will be  
805 helpful to understand if the pollution is caused by Substandard water treatment or pipeline pollution.  
806 Besides that, tools for measurements of more environmental parameters such as tap water  
807 temperature and pH can be added to the sampling kit such that more interactions between abiotic  
808 factors and tap water microbiome can be analyzed. Third, for better taxonomy resolution and thus  
809 more accurate risk assessments of harmful bacteria in tap water, more variable regions of 16S rRNA  
810 or even the full length could be sequenced (Matsuo et al., 2021). In addition, methods to accurately

811 quantify tap water microorganisms (especially waterborne pathogens) such as qPCR and digital PCR  
812 (dPCR) with proper controls could be used in further studies (Borchardt et al., 2021). Our study did  
813 not measure the absolute abundance of each species. To evaluate if the level of detected pathogens in  
814 our water samples could be health-threatening, we need to measure the absolute abundance to further  
815 provide reliable evidence to the public health sectors for more targeted actions.

816 To improve our citizen science sampling method, sample shipping temperature should be strictly  
817 controlled. This is not hard to achieve in countries with cold-chain shipping services, but as more  
818 people use the service, the overall expenditure on transportation will increase dramatically. As a  
819 result, setting up cooperation with some shipping companies to reduce the cost of sample shipping  
820 will be the key focus in the next step. Due to limited time and support, all the volunteers we recruited  
821 in this study were students and faculty members in our university and their family members. To  
822 increase our spatial and temporal coverage of sampling, volunteers from diverse institutions and  
823 organizations should be included. This will require us to publicize our citizen science research  
824 outside the campus and promote social volunteer network building. Meanwhile, as we expand our  
825 volunteer network to collect more data, we plan to set up an online platform to present our research  
826 discoveries to a wider audience to better facilitate science education and communication, which  
827 could also help encourage more citizens to participate in our research.

828

## 829 **4 Conclusions**

830 Ensuring household drinking water safety is vital for public health due to the risks associated with  
831 microbiological contamination. Utilizing a combination of citizen science and metabarcoding, we  
832 profiled the total microbial communities and potential waterborne pathogens in household drinking  
833 water across China, examining the impacts of temperature and extreme rainfall. Our results,  
834 consistent with previous studies, validate the efficacy of our citizen science sampling approach. This  
835 method, which extends beyond basic water collection and observation (e.g., water transparency),  
836 suggest that well-structured collaborations between professional agencies and citizen science can  
837 effectively monitor water quality on a broad scale.

838 In this study, a total of 7635 prokaryotic ASVs were detected from 40 household drinking water  
839 samples covering 28 cities of 18 provinces/regions in China. Outdoor temperature was found to  
840 significantly influence tap water microbiome community structure, with high temperatures associated  
841 with increases in potential pathogens. Extreme events, such as typhoons and floods, can exacerbate  
842 the presence of pathogens (e.g., *Escherichia coli* and *Salmonella* spp.) as well as toxin producing  
843 *Cyanobacteria* spp. in tap water. These observations underscore that conventional disinfection  
844 practices in the surveyed regions might be inadequate. Although not all bacteria species of the  
845 identified genera are pathogenic or toxin-producing, the potential health risks to humans are  
846 nonnegligible.

847 Considering these findings, enhanced water treatment measures and management during and after  
848 extreme rainfall events and periods of hot weather are warranted. This is particularly crucial in the

849 current context of climate change, which may heighten the frequency and intensity of such  
850 conditions.

851



## 852 References

- 853 Albus K., Thompson R. & Mitchell F. (2019). Usability of Existing Volunteer Water Monitoring  
854 Data: What Can the Literature Tell Us? *Citizen Science: Theory and Practice*, 4.
- 855 Allen M. J., Edberg S. C. & Reasoner D. J. (2004). Heterotrophic plate count bacteria—what is their  
856 significance in drinking water? *International journal of food microbiology*, 92, 265-274.
- 857 Altenburger R., Brack W., Burgess R. M., Busch W., Escher B. I., Focks A., Mark Hewitt L.,  
858 Jacobsen B. N., de Alda M. L. & Ait-Aissa S. (2019). Future water quality monitoring:  
859 improving the balance between exposure and toxicity assessments of real-world pollutant  
860 mixtures. *Environmental Sciences Europe*, 31, 1-17.
- 861 Andermann T., Antonelli A., Barrett R. L. & Silvestro D. (2022). Estimating alpha, beta, and gamma  
862 diversity through deep learning. *Frontiers in plant science*, 13, 839407.
- 863 Anderson M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral*  
864 *ecology*, 26, 32-46.
- 865 Anderson M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*,  
866 62, 245-53. doi: 10.1111/j.1541-0420.2005.00440.x.
- 867 Arbizu M. (2020). *pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version*  
868 *0.4* [Online]. github. Available: <https://github.com/pmartinezarbizu/pairwiseAdonis>  
869 [Accessed April 23 2022].
- 870 Ashbolt N. J. (2015). Microbial Contamination of Drinking Water and Human Health from  
871 Community Water Systems. *Curr Environ Health Rep*, 2, 95-106. doi: 10.1007/s40572-014-  
872 0037-5.
- 873 Baalbaki R., Ahmad S. H., Kays W., Talhouk S. N., Saliba N. A. & Al-Hindi M. (2019). Citizen  
874 science in Lebanon-a case study for groundwater quality monitoring. *R Soc Open Sci*, 6,  
875 181871. doi: 10.1098/rsos.181871.
- 876 Bautista-de los Santos Q. M., Schroeder J. L., Sevillano-Rivera M. C., Sungthong R., Ijaz U. Z.,  
877 Sloan W. T. & Pinto A. J. (2016). Emerging investigators series: microbial communities in  
878 full-scale drinking water distribution systems—a meta-analysis. *Environmental Science: Water*  
879 *Research & Technology*, 2, 631-644.
- 880 Borchardt M. A., Boehm A. B., Salit M., Spencer S. K., Wigginton K. R. & Noble R. T. (2021). The  
881 environmental microbiology minimum information (EMMI) guidelines: qPCR and dPCR  
882 quality and reporting for environmental microbiology. *Environmental Science & Technology*,  
883 55, 10210-10223.
- 884 Brettar I. & Hofle M. G. (2008). Molecular assessment of bacterial pathogens - a contribution to  
885 drinking water safety. *Curr Opin Biotechnol*, 19, 274-80. doi: 10.1016/j.copbio.2008.04.004.
- 886 Brouwer S., Van der Wielen P. W., Schriks M., Claassen M. & Frijns J. (2018). Public participation  
887 in science: the future and value of citizen science in the drinking water research. *Water*, 10,  
888 284.
- 889 Buxton A. S., Groombridge J. J. & Griffiths R. A. (2018). Comparison of two citizen scientist  
890 methods for collecting pond water samples for environmental DNA studies. *Citizen Science:*  
891 *Theory and Practice*, 3.
- 892 Cai D. 2023. Climate change worsened Asia's April heatwave by 2C - study. BBC News.
- 893 Callahan B. J., McMurdie P. J., Rosen M. J., Han A. W., Johnson A. J. & Holmes S. P. (2016).  
894 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*, 13,  
895 581-3. doi: 10.1038/nmeth.3869.
- 896 Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K. & Madden T. L. (2009).  
897 BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421. doi: 10.1186/1471-  
898 2105-10-421.

- 899 Cann K. F., Thomas D. R., Salmon R. L., Wyn-Jones A. P. & Kay D. (2013). Extreme water-related  
900 weather events and waterborne disease. *Epidemiol Infect*, 141, 671-86. doi:  
901 10.1017/S0950268812001653.
- 902 Chen S. (2021). Zhengzhou Rainstorm Once in a thousand years? The meteorological authority  
903 responded: A year's rain fell in one day. *Chinanews*.
- 904 Chiao T. H., Clancy T. M., Pinto A., Xi C. & Raskin L. (2014). Differential resistance of drinking  
905 water bacterial populations to monochloramine disinfection. *Environ Sci Technol*, 48, 4038-  
906 47. doi: 10.1021/es4055725.
- 907 China D. I. T. o. T. S. C. o. t. P. s. R. o. 2022. Investigation report on Zhengzhou 7.20 Rainstorm  
908 Disaster. In: China, M. o. E. M. o. t. P. s. R. o. (ed.).
- 909 Chopyk J., Moore R. M., DiSpirito Z., Stromberg Z. R., Lewis G. L., Renter D. G., Cernicchiaro N.,  
910 Moxley R. A. & Wommack K. E. (2016). Presence of pathogenic *Escherichia coli* is  
911 correlated with bacterial community diversity and composition on pre-harvest cattle hides.  
912 *Microbiome*, 4, 1-11.
- 913 Cissé G. (2019). Food-borne and water-borne diseases under climate change in low-and middle-  
914 income countries: Further efforts needed for reducing environmental health exposure risks.  
915 *Acta tropica*, 194, 181-188.
- 916 Commission E. 2020. COMMISSION STAFF WORKING DOCUMENT: Best Practices in Citizen  
917 Science for Environmental Monitoring. European Commission.
- 918 Corning. (2023). *Syringe Filters* [Online]. Available: [https://ecatalog.corning.com/life-](https://ecatalog.corning.com/life-sciences/b2c/US/en/Cell-Culture/Filtration/Syringe-Filters/Syringe-Filters/p/syringeFilters)  
919 [sciences/b2c/US/en/Cell-Culture/Filtration/Syringe-Filters/Syringe-Filters/p/syringeFilters](https://ecatalog.corning.com/life-sciences/b2c/US/en/Cell-Culture/Filtration/Syringe-Filters/Syringe-Filters/p/syringeFilters)  
920 [Accessed May 12 2023].
- 921 Dai Z., Sevillano-Rivera M. C., Calus S. T., Bautista-de Los Santos Q. M., Eren A. M., van der  
922 Wielen P., Ijaz U. Z. & Pinto A. J. (2020). Disinfection exhibits systematic impacts on the  
923 drinking water microbiome. *Microbiome*, 8, 42. doi: 10.1186/s40168-020-00813-0.
- 924 EPA. (2022). *Learn about Cyanobacteria and Cyanotoxins* [Online]. Available:  
925 <https://www.epa.gov/cyanohabs/learn-about-cyanobacteria-and-cyanotoxins#what3>  
926 [Accessed April 23 2022].
- 927 EPA. (2023). *Contaminant Candidate List (CCL) and Regulatory Determination* [Online]. Available:  
928 <https://www.epa.gov/ccl/contaminant-candidate-list-5-ccl-5> [Accessed May 25 2023].
- 929 Franca L., Lopez-Lopez A., Rossello-Mora R. & da Costa M. S. (2015). Microbial diversity and  
930 dynamics of a groundwater and a still bottled natural mineral water. *Environ Microbiol*, 17,  
931 577-93. doi: 10.1111/1462-2920.12430.
- 932 Funari E., Manganelli M. & Sinisi L. (2012). Impact of climate change on waterborne diseases.  
933 *Annali dell'Istituto superiore di sanita*, 48, 473-487.
- 934 Garner E., Davis B. C., Milligan E., Blair M. F., Keenum I., Maile-Moskowitz A., Pan J., Gnegy M.,  
935 Liguori K. & Gupta S. (2021). Next generation sequencing approaches to evaluate water and  
936 wastewater quality. *Water Research*, 194, 116907.
- 937 Haegeman B., Hamelin J., Moriarty J., Neal P., Dushoff J. & Weitz J. S. (2013). Robust estimation of  
938 microbial diversity in theory and in practice. *ISME J*, 7, 1092-101. doi:  
939 10.1038/ismej.2013.10.
- 940 Han Z., An W., Yang M. & Zhang Y. (2020). Assessing the impact of source water on tap water  
941 bacterial communities in 46 drinking water supply systems in China. *Water Res*, 172, 115469.  
942 doi: 10.1016/j.watres.2020.115469.
- 943 Harper L. (2018). *The Importance of Citizen Scientists* [Online]. American Forests. Available:  
944 <https://www.americanforests.org/blog/the-importance-of-citizen-scientists/> [Accessed April  
945 23 2022].
- 946 Ho S. Y., Xu S. J. & Lee F. W. (2020). Citizen science: An alternative way for water monitoring in  
947 Hong Kong. *PLoS One*, 15, e0238349. doi: 10.1371/journal.pone.0238349.

- 948 Holinger E. P., Ross K. A., Robertson C. E., Stevens M. J., Harris J. K. & Pace N. R. (2014).  
949 Molecular analysis of point-of-use municipal drinking water microbiology. *Water Res*, 49,  
950 225-35. doi: 10.1016/j.watres.2013.11.027.
- 951 Hull N. M., Holinger E. P., Ross K. A., Robertson C. E., Harris J. K., Stevens M. J. & Pace N. R.  
952 (2017). Longitudinal and source-to-tap New Orleans, LA, USA drinking water microbiology.  
953 *Environmental Science & Technology*, 51, 4220-4229.
- 954 Inkinen J., Siponen S., Jayaprakash B., Tiwari A., Hokajarvi A. M., Pursiainen A., Ikonen J.,  
955 Kauppinen A., Miettinen I. T., Paananen J., Torvinen E., Kolehmainen M. & Pitkanen T.  
956 (2021). Diverse and active archaea communities occur in non-disinfected drinking water  
957 systems-Less activity revealed in disinfected and hot water systems. *Water Res X*, 12, 100101.  
958 doi: 10.1016/j.wroa.2021.100101.
- 959 IPCC 2008. Climate Change and Water. In: Bates, B., Kundzewicz, Z. W., Wu, S. & Palutikof, J.  
960 (eds.).
- 961 Ivone V.-M., Conceição E., Olga C N. & Célia M M. (2013). Bacterial diversity from the source to  
962 the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods.  
963 *FEMS microbiology ecology*, 83, 361-374.
- 964 Ji P., Parks J., Edwards M. A. & Pruden A. (2015). Impact of Water Chemistry, Pipe Material and  
965 Stagnation on the Building Plumbing Microbiome. *PLoS One*, 10, e0141087. doi:  
966 10.1371/journal.pone.0141087.
- 967 Jia S., Shi P., Hu Q., Li B., Zhang T. & Zhang X. X. (2015). Bacterial Community Shift Drives  
968 Antibiotic Resistance Promotion during Drinking Water Chlorination. *Environ Sci Technol*,  
969 49, 12271-9. doi: 10.1021/acs.est.5b03521.
- 970 Johnson J. S., Spakowicz D. J., Hong B. Y., Petersen L. M., Demkowicz P., Chen L., Leopold S. R.,  
971 Hanson B. M., Agresta H. O., Gerstein M., Sodergren E. & Weinstock G. M. (2019).  
972 Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis.  
973 *Nat Commun*, 10, 5029. doi: 10.1038/s41467-019-13036-1.
- 974 Kozich J. J., Westcott S. L., Baxter N. T., Highlander S. K. & Schloss P. D. (2013). Development of  
975 a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data  
976 on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*, 79, 5112-20. doi:  
977 10.1128/AEM.01043-13.
- 978 Langeveld J., Schilperoort R. & Weijers S. (2013). Climate change and urban wastewater  
979 infrastructure: there is more to explore. *Journal of Hydrology*, 476, 112-119.
- 980 Laws E. A. (2017). *Aquatic Pollution: An Introductory Text*, Wiley.
- 981 Levy K., Smith S. M. & Carlton E. J. (2018). Climate change impacts on waterborne diseases:  
982 moving toward designing interventions. *Current environmental health reports*, 5, 272-282.
- 983 Ley C. J., Proctor C. R., Singh G., Ra K., Noh Y., Odimeyomi T., Salehi M., Julien R., Mitchell J. &  
984 Nejadhashemi A. P. (2020). Drinking water microbiology in a water-efficient building:  
985 stagnation, seasonality, and physicochemical effects on opportunistic pathogen and total  
986 bacteria proliferation. *Environmental Science: Water Research & Technology*, 6, 2902-2913.
- 987 Lim J. Y., Yoon J. W. & Hovde C. J. (2010). A brief overview of Escherichia coli O157: H7 and its  
988 plasmid O157. *Journal of microbiology and biotechnology*, 20, 5.
- 989 Lin Y., Gifford S., Ducklow H., Schofield O. & Cassar N. (2019). Towards Quantitative Microbiome  
990 Community Profiling Using Internal Standards. *Appl Environ Microbiol*, 85, e02634-18. doi:  
991 10.1128/AEM.02634-18.
- 992 Ling F., Whitaker R., LeChevallier M. W. & Liu W. T. (2018). Drinking water microbiome assembly  
993 induced by water stagnation. *ISME J*, 12, 1520-1531. doi: 10.1038/s41396-018-0101-5.
- 994 Lipp E. K., Huq A. & Colwell R. R. (2002). Effects of global climate on infectious disease: the  
995 cholera model. *Clinical microbiology reviews*, 15, 757-770.

- 996 Liu Y., Gilchrist A., Zhang J. & Li X. F. (2008). Detection of viable but nonculturable *Escherichia*  
997 *coli* O157:H7 bacteria in drinking water and river water. *Appl Environ Microbiol*, 74, 1502-7.  
998 doi: 10.1128/AEM.02125-07.
- 999 Ma L., Jiang X. T., Guan L., Li B. & Zhang T. (2022). Nationwide biogeography and health  
1000 implications of bacterial communities in household drinking water. *Water Res*, 215, 118238.  
1001 doi: 10.1016/j.watres.2022.118238.
- 1002 Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.  
1003 *EMBnet. journal*, 17, 10-12.
- 1004 Masson-Delmotte V., Zhai P., Pirani A., Connors S. L., Péan C., Berger S., Caud N., Chen Y.,  
1005 Goldfarb L. & Gomis M. (2021). Climate change 2021: the physical science basis.  
1006 *Contribution of working group I to the sixth assessment report of the intergovernmental panel*  
1007 *on climate change*, 2.
- 1008 Matsuo Y., Komiya S., Yasumizu Y., Yasuoka Y., Mizushima K., Takagi T., Kryukov K., Fukuda A.,  
1009 Morimoto Y. & Naito Y. (2021). Full-length 16S rRNA gene amplicon analysis of human gut  
1010 microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC*  
1011 *microbiology*, 21, 1-13.
- 1012 McKight P. E. & Najab J. (2010). Kruskal-wallis test. *The corsini encyclopedia of psychology*, 1-1.
- 1013 McMurdie P. J. & Holmes S. (2013). phyloseq: an R package for reproducible interactive analysis  
1014 and graphics of microbiome census data. *PLoS One*, 8, e61217. doi:  
1015 10.1371/journal.pone.0061217.
- 1016 Metcalfe A. N., Kennedy T. A., Mendez G. A. & Muehlbauer J. D. (2022). Applied citizen science in  
1017 freshwater research. *Wiley Interdisciplinary Reviews: Water*, 9, e1578.
- 1018 MilliporeSigma. (2023). *Sterivex* [Online]. Available:  
1019 [https://www.sigmaaldrich.com/US/en/search/sterivex?focus=products&page=1&perpage=30](https://www.sigmaaldrich.com/US/en/search/sterivex?focus=products&page=1&perpage=30&sort=relevance&term=sterivex&type=product)  
1020 [&sort=relevance&term=sterivex&type=product](https://www.sigmaaldrich.com/US/en/search/sterivex?focus=products&page=1&perpage=30&sort=relevance&term=sterivex&type=product) [Accessed May 12 2023].
- 1021 Motlagh A. M. & Yang Z. (2019). Detection and occurrence of indicator organisms and pathogens.  
1022 *Water Environment Research*, 91, 1402-1408.
- 1023 NMC. (2021). National Meteorological Center. Available: <http://www.nmc.cn/> [Accessed February  
1024 14 2022].
- 1025 NMC-Typhoon. (2021). *2106 In-fa* [Online]. National Meteorological Centre - Typhoon. Available:  
1026 <http://typhoon.nmc.cn/web.html> [Accessed February 24 2022].
- 1027 Oksanen J., Blanchet F. & Kindt R. 2015. Vegan: commity ecology package. R package version 2.3-  
1028 0.
- 1029 Paddison L. 2023. Heat waves in US and Europe would have been ‘virtually impossible’ without  
1030 climate change, new report finds. CNN.
- 1031 Pandey P. K., Kass P. H., Soupir M. L., Biswas S. & Singh V. P. (2014). Contamination of water  
1032 resources by pathogenic bacteria. *AMB Express*, 4, 51. doi: 10.1186/s13568-014-0051-x.
- 1033 Phan N. K. & Sherchan S. P. (2020). Microbiological Assessment of Tap Water Following the 2016  
1034 Louisiana Flooding. *Int J Environ Res Public Health*, 17, 1273. doi: 10.3390/ijerph17041273.
- 1035 Pinto A. J., Schroeder J., Lunn M., Sloan W. & Raskin L. (2014). Spatial-temporal survey and  
1036 occupancy-abundance modeling to predict bacterial community dynamics in the drinking  
1037 water microbiome. *mBio*, 5, e01135-14. doi: 10.1128/mBio.01135-14.
- 1038 Pinto A. J., Xi C. & Raskin L. (2012). Bacterial community structure in the drinking water  
1039 microbiome is governed by filtration processes. *Environmental science & technology*, 46,  
1040 8851-8859.
- 1041 Pocock M. J., Roy H. E., Fox R., Ellis W. N. & Botham M. (2017). Citizen science and invasive  
1042 alien species: predicting the detection of the oak processionary moth *Thaumetopoea*  
1043 *processionea* by moth recorders. *Biological conservation*, 208, 146-154.

- 1044 Preheim S. P., Perrotta A. R., Friedman J., Smilie C., Brito I., Smith M. B. & Alm E. (2013).  
1045 Computational methods for high-throughput comparative analyses of natural microbial  
1046 communities. *Methods in enzymology*. Elsevier.
- 1047 Ramirez K. S., Knight C. G., de Hollander M., Brearley F. Q., Constantinides B., Cotton A., Creer S.,  
1048 Crowther T. W., Davison J., Delgado-Baquerizo M., Dorrepaal E., Elliott D. R., Fox G.,  
1049 Griffiths R. I., Hale C., Hartman K., Houlden A., Jones D. L., Krab E. J., Maestre F. T.,  
1050 McGuire K. L., Monteux S., Orr C. H., van der Putten W. H., Roberts I. S., Robinson D. A.,  
1051 Rocca J. D., Rowntree J., Schlaeppli K., Shepherd M., Singh B. K., Straathof A. L., Bhatnagar  
1052 J. M., Thion C., van der Heijden M. G. A. & de Vries F. T. (2018). Detecting  
1053 macroecological patterns in bacterial communities across independent studies of global soils.  
1054 *Nat Microbiol*, 3, 189-196. doi: 10.1038/s41564-017-0062-x.
- 1055 Ramirez-Castillo F. Y., Loera-Muro A., Jacques M., Garneau P., Avelar-Gonzalez F. J., Harel J. &  
1056 Guerrero-Barrera A. L. (2015). Waterborne pathogens: detection methods and challenges.  
1057 *Pathogens*, 4, 307-34. doi: 10.3390/pathogens4020307.
- 1058 Roeselers G., Coolen J., van der Wielen P. W., Jaspers M. C., Atsma A., de Graaf B. & Schuren F.  
1059 (2015). Microbial biogeography of drinking water: patterns in phylogenetic diversity across  
1060 space and time. *Environ Microbiol*, 17, 2505-14. doi: 10.1111/1462-2920.12739.
- 1061 Roy H. E., Pocock M. J., Preston C. D., Roy D. B., Savage J., Tweddle J. & Robinson L. (2012).  
1062 Understanding citizen science and environmental monitoring: final report on behalf of UK  
1063 Environmental Observation Framework.
- 1064 Ryan M. P. & Pembroke J. T. (2018). Brevundimonas spp: Emerging global opportunistic pathogens.  
1065 *Virulence*, 9, 480-493. doi: 10.1080/21505594.2017.1419116.
- 1066 Smouse P. E., Long J. C. & Sokal R. R. (1986). Multiple regression and correlation extensions of the  
1067 Mantel test of matrix correspondence. *Systematic zoology*, 35, 627-632.
- 1068 Tang W., Liu Y., Li Q., Chen L., Li Q., Li P. & Xia S. (2021). The Impact of Extreme Weather  
1069 Events on Bacterial Communities and Opportunistic Pathogens in a Drinking Water  
1070 Treatment Plant. *Water*, 14, 54.
- 1071 Thompson L. R., Sanders J. G., McDonald D., Amir A., Ladau J., Locey K. J., Prill R. J., Tripathi A.,  
1072 Gibbons S. M., Ackermann G., Navas-Molina J. A., Janssen S., Kopylova E., Vazquez-Baeza  
1073 Y., Gonzalez A., Morton J. T., Mirarab S., Zech Xu Z., Jiang L., Haroon M. F., Kanbar J.,  
1074 Zhu Q., Jin Song S., Kosciolk T., Bokulich N. A., Lefler J., Brislawn C. J., Humphrey G.,  
1075 Owens S. M., Hampton-Marcell J., Berg-Lyons D., McKenzie V., Fierer N., Fuhrman J. A.,  
1076 Clauset A., Stevens R. L., Shade A., Pollard K. S., Goodwin K. D., Jansson J. K., Gilbert J.  
1077 A., Knight R. & Earth Microbiome Project C. (2017). A communal catalogue reveals Earth's  
1078 multiscale microbial diversity. *Nature*, 551, 457-463. doi: 10.1038/nature24621.
- 1079 Underground W. 2022.
- 1080 Upton G. J. (1992). Fisher's exact test. *Journal of the Royal Statistical Society: Series A (Statistics in*  
1081 *Society)*, 155, 395-402.
- 1082 van der Wielen P. W. & van der Kooij D. (2010). Effect of water composition, distance and season  
1083 on the adenosine triphosphate concentration in unchlorinated drinking water in the  
1084 Netherlands. *Water Res*, 44, 4860-7. doi: 10.1016/j.watres.2010.07.016.
- 1085 Vej G. W. 2007. Bioinformatics explained: BLAST. CLC bio.
- 1086 Wang T., Sun D., Zhang Q. & Zhang Z. (2021). China's drinking water sanitation from 2007 to 2018:  
1087 A systematic review. *Sci Total Environ*, 757, 143923. doi: 10.1016/j.scitotenv.2020.143923.
- 1088 Wen X., Chen F., Lin Y., Zhu H., Yuan F., Kuang D., Jia Z. & Yuan Z. (2020). Microbial indicators  
1089 and their use for monitoring drinking water quality—A review. *Sustainability*, 12, 2249.
- 1090 WHO 2017. Guidelines for Drinking-Water Quality (Fourth Edition).
- 1091 Wikipedia. (2021). *Typhoon In-fa* [Online]. Wikimedia Foundation. Available:  
1092 [https://en.wikipedia.org/wiki/Typhoon\\_In-fa](https://en.wikipedia.org/wiki/Typhoon_In-fa) [Accessed February 14 2022].

- 1093 Willis J. R., Gonzalez-Torres P., Pittis A. A., Bejarano L. A., Cozzuto L., Andreu-Somavilla N.,  
1094 Alloza-Trabado M., Valentin A., Ksiezopolska E., Company C., Onywera H., Montfort M.,  
1095 Hermoso A., Iraola-Guzman S., Saus E., Labeeuw A., Carolis C., Hecht J., Ponomarenko J. &  
1096 Gabaldon T. (2018). Citizen science charts two major "stomatotypes" in the oral microbiome  
1097 of adolescents and reveals links with habits and drinking water composition. *Microbiome*, 6,  
1098 218. doi: 10.1186/s40168-018-0592-3.
- 1099 Wolf-Baca M. & Siedlecka A. (2019). Detection of pathogenic bacteria in hot tap water using the  
1100 qPCR method: preliminary research. *SN Applied Sciences*, 1, 1-9.
- 1101 Wu Y., Washbourne C. & Haklay M. (2022). Citizen science in China's water resources monitoring:  
1102 current status and future prospects. *International Journal of Sustainable Development &*  
1103 *World Ecology*, 29, 277-290.
- 1104 Xie J., Jin L., Wu D., Pruden A. & Li X. (2022). Inhalable Antibiotic Resistome from Wastewater  
1105 Treatment Plants to Urban Areas: Bacterial Hosts, Dissemination Risks, and Source  
1106 Contributions. *Environ Sci Technol*, 56, 7040-7051. doi: 10.1021/acs.est.1c07023.
- 1107 Yang Y., Xing S., Chen Y., Wu R., Wu Y., Wang Y., Mi J. & Liao X. (2021). Profiles of  
1108 bacteria/phage-comediated ARGs in pig farm wastewater treatment plants in China:  
1109 Association with mobile genetic elements, bacterial communities and environmental factors. *J*  
1110 *Hazard Mater*, 404, 124149. doi: 10.1016/j.jhazmat.2020.124149.
- 1111 Yu P., Zaleski A., Li Q., He Y., Mapili K., Pruden A., Alvarez P. J. & Stadler L. B. (2018). Elevated  
1112 levels of pathogenic indicator bacteria and antibiotic resistance genes after Hurricane  
1113 Harvey's flooding in Houston. *Environmental science & technology letters*, 5, 481-486.
- 1114 Yu X., Lin H. & Zhang S. (2012). Microbial contamination in drinking water: occurrence,  
1115 measurement and control. *Bulletin of the Chinese Academy of Sciences BCAS*, 26, 84-91.
- 1116 Zachariah M., Philip S., Pinto I., Vahlberg M., Singh R. & Otto F. E. L. 2023. Extreme heat in North  
1117 America, Europe and China in July 2023 made much more likely by climate change.  
1118 Grantham Institute for Climate Change.
- 1119 Zhang H., Xu L., Huang T., Liu X., Miao Y., Liu K. & Qian X. (2021a). Indoor heating triggers  
1120 bacterial ecological links with tap water stagnation during winter: Novel insights into  
1121 bacterial abundance, community metabolic activity and interactions. *Environmental Pollution*,  
1122 269, 116094.
- 1123 Zhang H., Xu L., Huang T., Yan M., Liu K., Miao Y., He H., Li S. & Sekar R. (2021b). Combined  
1124 effects of seasonality and stagnation on tap water quality: Changes in chemical parameters,  
1125 metabolic activity and co-existence in bacterial community. *Journal of Hazardous Materials*,  
1126 403, 124018.
- 1127 Zhang J., Chen S., Chen B., Du Y., Huang X., Pan X. & Zhang Q. (2013). Citizen science:  
1128 integrating scientific research, ecological conservation and public participation. *Biodiversity*  
1129 *Science*, 21, 738.
- 1130 Zhang J., Mauzerall D. L., Zhu T., Liang S., Ezzati M. & Remais J. V. (2010). Environmental health  
1131 in China: progress towards clean air and safe water. *Lancet*, 375, 1110-9. doi: 10.1016/S0140-  
1132 6736(10)60062-1.
- 1133 Zhang Y., Deng J., Qin B., Zhu G., Zhang Y., Jeppesen E. & Tong Y. (2022). Importance and  
1134 vulnerability of lakes and reservoirs supporting drinking water in China. *Fundamental*  
1135 *Research*.
- 1136 Zlatanović L., van der Hoek J. P. & Vreeburg J. (2017). An experimental study on the influence of  
1137 water stagnation and temperature change on water quality in a full-scale domestic drinking  
1138 water system. *Water research*, 123, 761-772.
- 1139 Zou X.-Y., Peng X.-Y., Zhao X.-X. & Chang C.-P. (2023). The impact of extreme weather events on  
1140 water quality: International evidence. *Natural Hazards*, 115, 1-21.
- 1141

1142 **Non-standard abbreviations**

1143 RA: Relative abundance

1144 CS: Citizen science sampling protocol and materials

1145 SM: Supplementary material

1146 NC: Nanchang, Jiangxi

1147 XA: Xi'an, Shanxi

1148 NJ: Nanjing, Jiangsu

1149 HZ: Huizhou, Guangdong

1150 JN: Jinan, Shandong

1151 ZZ: Zhengzhou, Henan

1152 CZ: Changzhou, Jiangsu

1153 SH: Shanghai

1154 ZB: Zibo, Shandong

1155 MA: Macau

1156 BJ: Beijing

1157 TJ: Tianjin

1158 CQ: Chongqing

1159 LJ: Lijiang, Yunnan

1160 LZ: Lanzhou, Gansu

1161

1162 **Figure legends**

1163 **Figure 1** Geographical locations of sampling sites, sampling regions, sampling kit, and procedure.

1164 (A) Regions 1 to 7 represent **Beijing-Tianjin region, North** (Shandong and Hebei), **Northwest**  
1165 (Gansu and Shanxi), **Central** (Henan, Anhui, Jiangxi, and Hunan), **Central Coast** (Jiangsu,  
1166 Shanghai, and Zhejiang), **South Coast** (Fujian, Guangdong and Macau), and **Southwest** (Sichuan,  
1167 Chongqing, and Yunnan) respectively (Software: Datawrapper) (left). (B) A screenshot from the  
1168 demonstration video showing the sampling kit (upper right). (C) A screenshot from the  
1169 demonstration video showing the filtering process (lower right).

1170 **Figure 2** The distribution of tap water samples showing positive (red dots) PCR signal(s) and  
1171 negative (blue dots) PCR signal.

1172 Darker color denotes more samples. For cities with more than one sample: 1) samples collected from  
1173 Zhengzhou, Nanjing, Kunshan, and Shanghai all showed positive PCR signal(s); 2) samples collected  
1174 from Hangzhou all showed negative PCR signal; 3) samples collected from Beijing, Tianjin, and  
1175 Changsha showed inconsistent PCR signals in each city.

1176 **Figure 3** Taxonomic composition and relative abundance of microbiota in household drinking water  
1177 in China at the phylum level.

1178 “\*” denotes the samples collected after extreme rainfall events. Normal samples are grouped based  
1179 on the outdoor temperature. Within each temperature category, samples are in the order of Regions 1-  
1180 7. Only the top 10 phyla (1-10: bottom to top in the legend) and not assigned (NA) ones are shown.

1181 **Figure 4** Taxonomic composition and relative abundance of microbiota in household drinking water  
1182 in China at the class level.

1183 The pie charts show the mean composition of top 10 bacterial and archaeal classes in **(A)** seven  
1184 geographic regions and **(B)** three temperature categories.

1185 **Figure 5** The  $\alpha$  diversity (ASV-based).

1186 **(A)** geographic regions, **(B)** outdoor temperature categories, and **(C)** samples related to extreme  
1187 rainfall events.

1188 **Figure 6** Principal coordinates analysis (PCoA).

1189 Based on the distance matrix of Bray-Curtis dissimilarity of microbial profiles (ASV-based) among  
1190 all samples ( $n = 22$ ). HW: the weather samples collected in summer; CZ: samples collected in the  
1191 same household in Changzhou; ZZ: samples collected in two nearby households in Zhengzhou.

1192 **Figure 7** Relative abundance of potential bacterial pathogens in all the samples.

1193 Pathogens in the figure legend are in the descending order of mean RA of each pathogen (from top to  
1194 bottom). Samples are in the descending order of total RA of all pathogens in each sample (from left  
1195 to right).

1196 **Figure 8** Spearman’s correlation.

1197 **(A)** alpha diversity indexes and potential pathogens detected in normal samples; **(B)** potential  
1198 pathogens detected in all the samples. 1-3 asterisk(s) denote the significance level of 0.05, 0.01, and  
1199 0.001 respectively.

1200

1201 **Tables**



1202 **Table 1** Sampling seasons: outdoor temperature categories.

Category	Range (°C)	Number of samples
High	T > 25	7
Median	15 < T ≤ 25	7
Low	T ≤ 15	6

1203

1204 **Table 2** Number of sequences, ASVs, and genera for the top 10 phyla in the water samples.

1205 Those three represent the coverage, diversity, and the genus spectrum of microbial community in the  
 1206 water samples respectively. <sup>a</sup> The percentage of sequences was the phylum sequences in the total of  
 1207 assigned sequences at the phylum level which was 1,333,322.

Phylum	Sequence <sup>a</sup>	ASV	Genus
<i>Proteobacteria</i>	697642 (52.3%)	2693	281
<i>Planctomycetota</i>	164487 (12.3%)	998	23
<i>Actinobacteriota</i>	82262 (6.2%)	279	56
<i>Acidobacteriota</i>	77415 (5.8%)	296	17
<i>Cyanobacteria</i>	72219 (5.4%)	402	25
<i>Crenarchaeota</i>	43624 (3.3%)	31	5
<i>Bacteroidota</i>	38869 (2.9%)	603	61
<i>Verrucomicrobiota</i>	33769 (2.5%)	411	27
<i>Bdellovibrionota</i>	31056 (2.3%)	400	8
<i>Desulfobacterota</i>	21608 (1.6%)	59	11
Others	70371 (5.3%)	1209	93
N.A.	6133 (0.5%)	254	/

1208

1209 **Table 3** Occurrence and mean relative abundance (RA) of major potential pathogens.

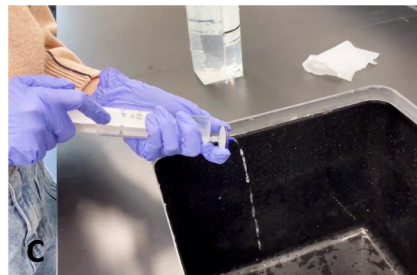
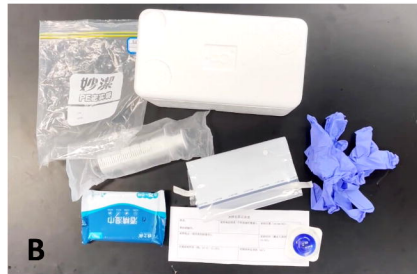
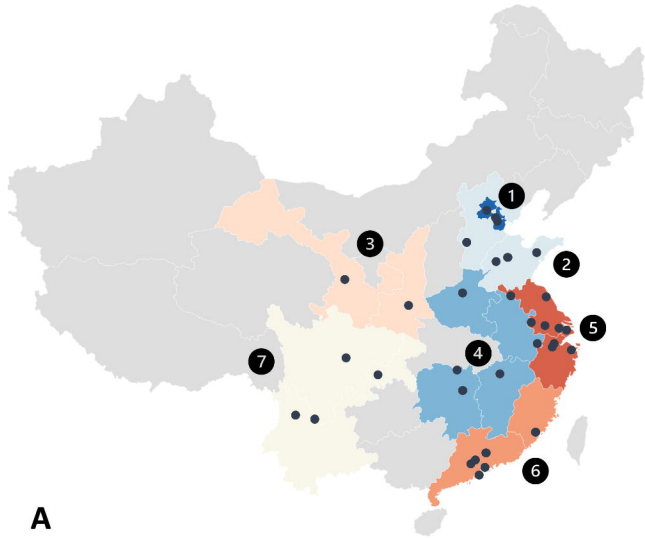
Pathogen Species	Occurrence	Mean RA
<i>Mycobacterium</i> spp.	22 (100.0%)	2.67E-02
<i>Acinetobacter</i> spp.	22 (100.0%)	1.43E-02
<i>Legionella</i> spp.	22 (100%)	4.59E-03
<i>Brevundimonas</i> spp.	20 (90.9%)	6.12E-03
<i>Leptospira</i> spp.	11 (50.0%)	2.04E-04

<i>Escherichia coli</i>	8 (36.4%)	5.80E-04
<i>Bacillus</i> spp.	3 (13.6%)	2.01E-05
<i>Aeromonas hydrophila</i>	2 (9.1%)	4.76E-05
<i>Salmonella enterica</i>	1 (4.5%)	1.79E-04

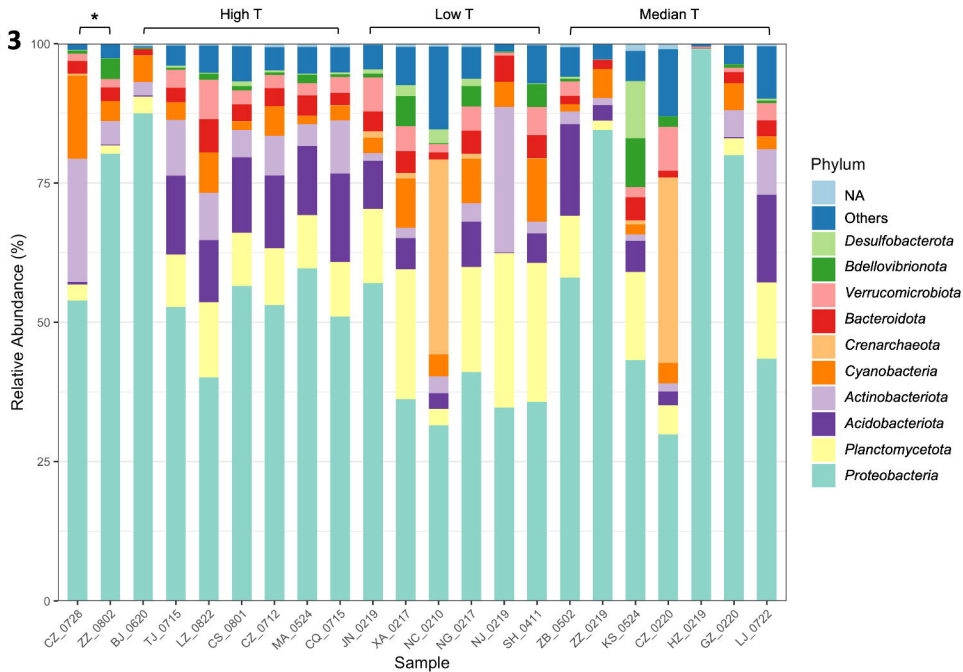
---

1210

**Figure 1**

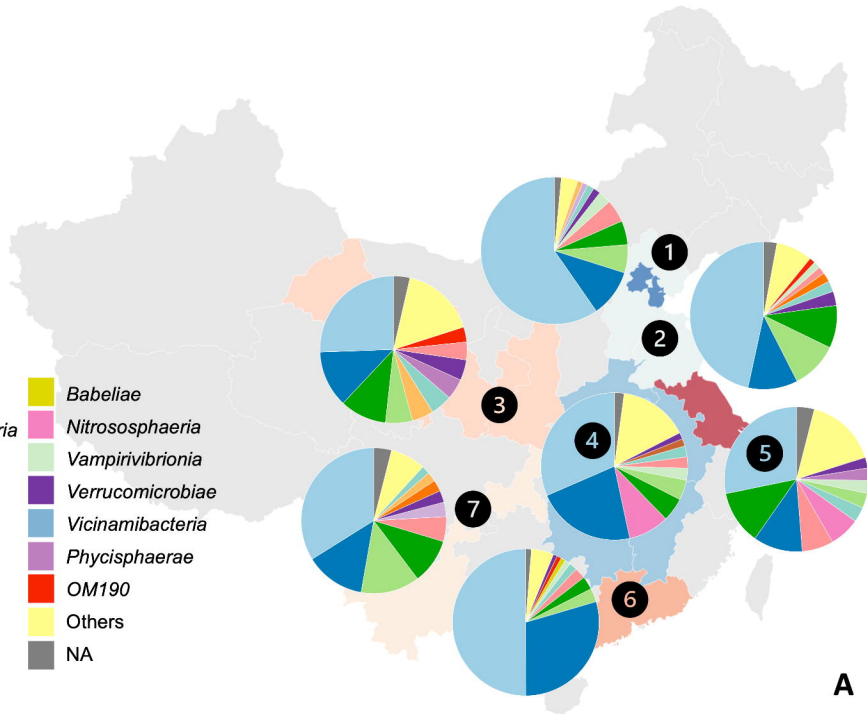
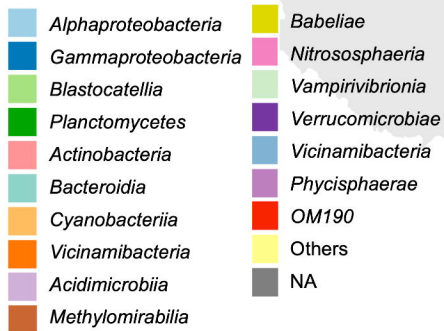




**Figure 3**

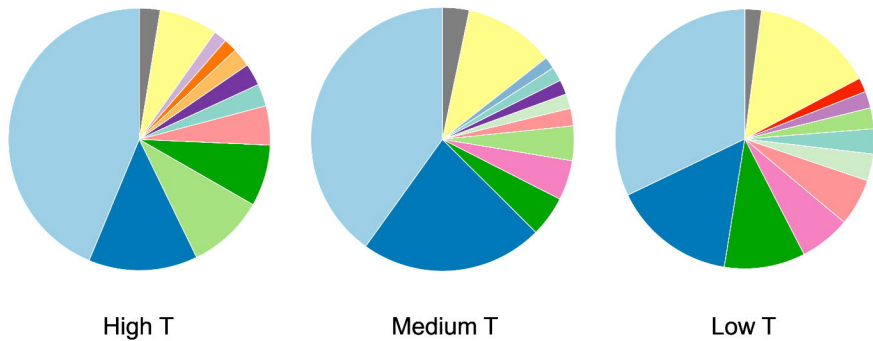
**Figure 4**

Top10 Class

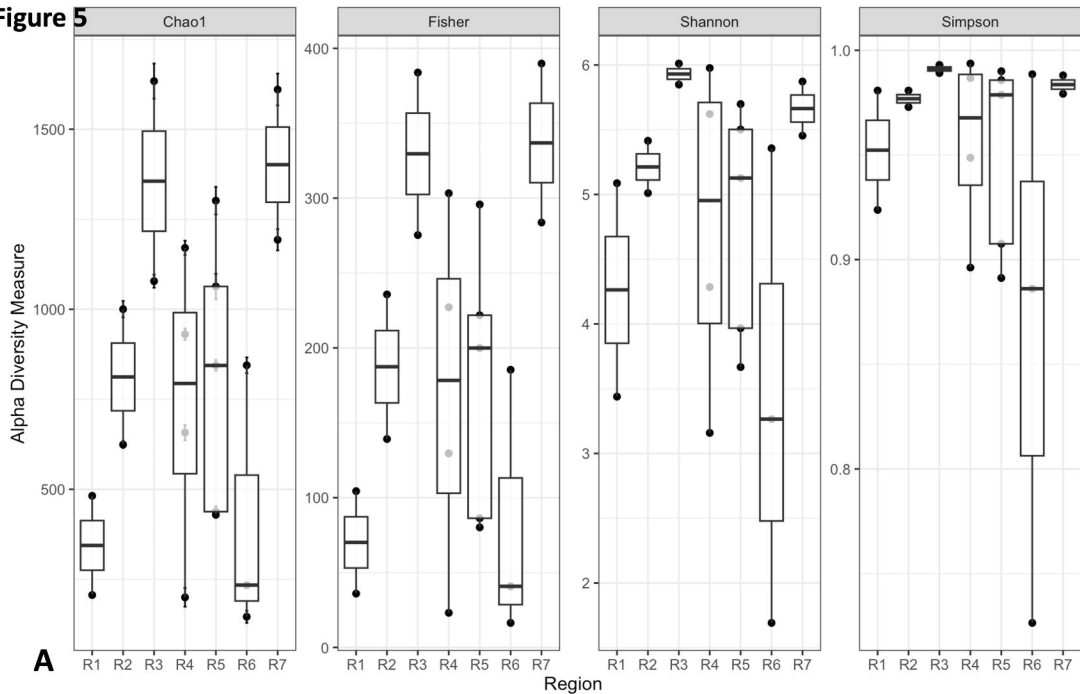


**A**

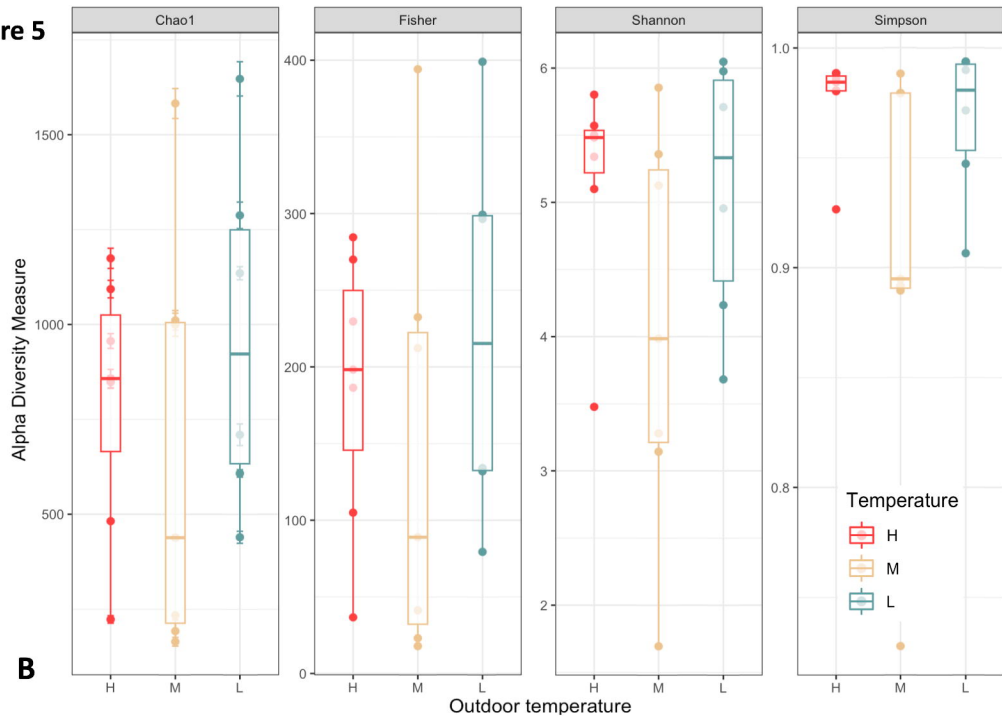
**Figure 4**

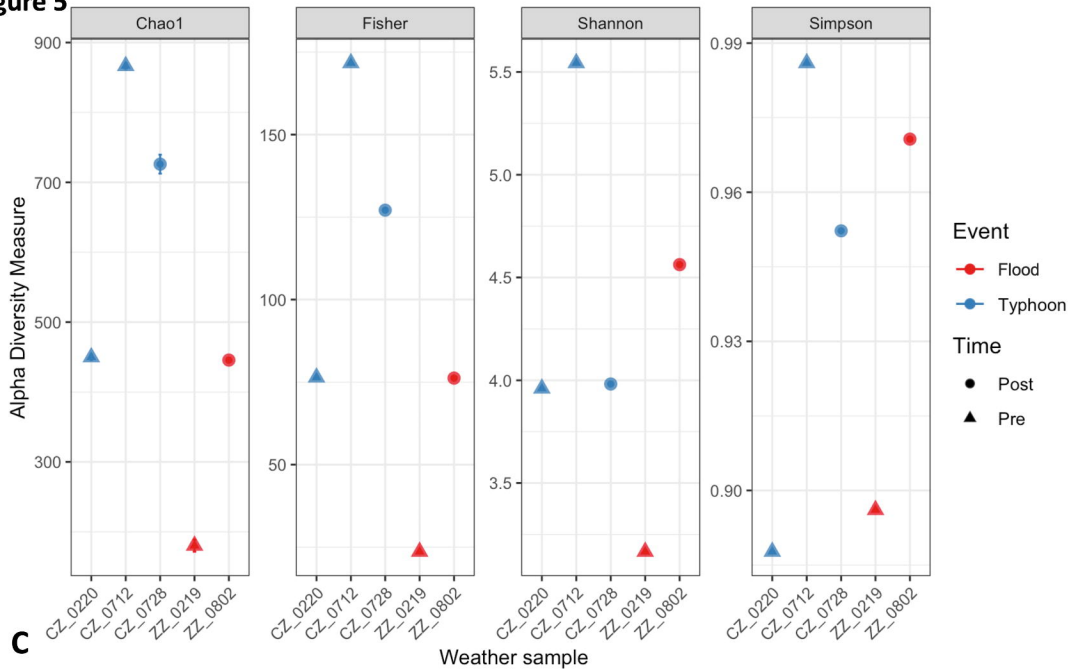


**B**

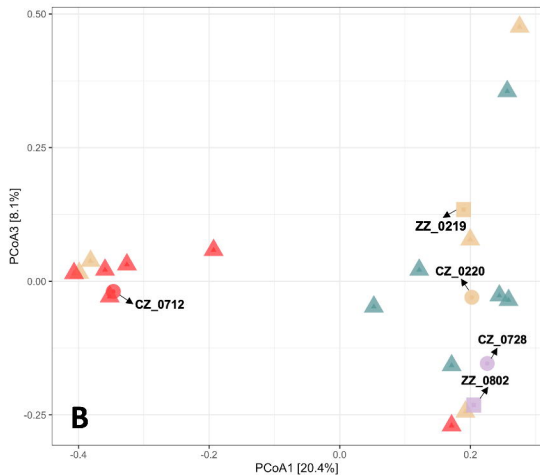
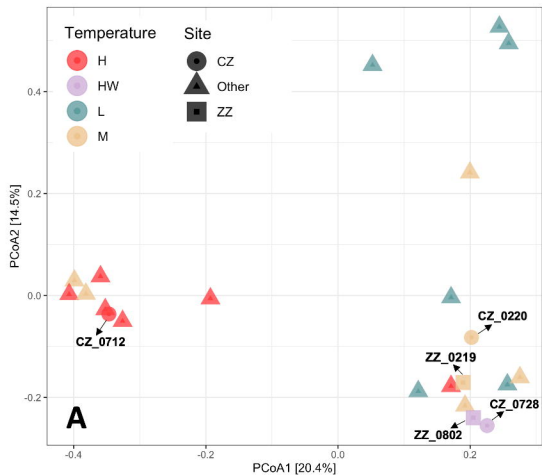
**Figure 5****A**

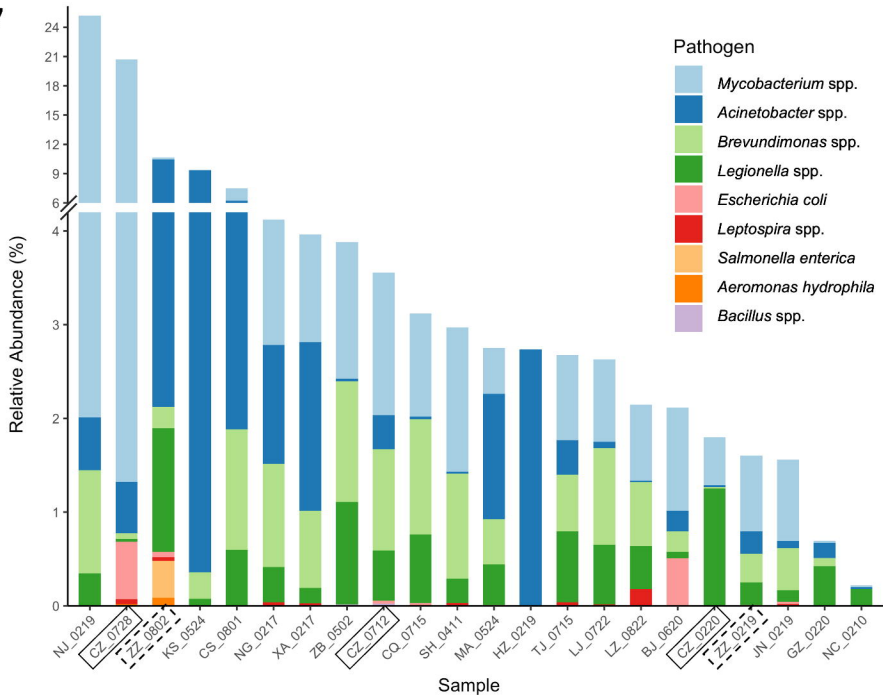


**Figure 5**

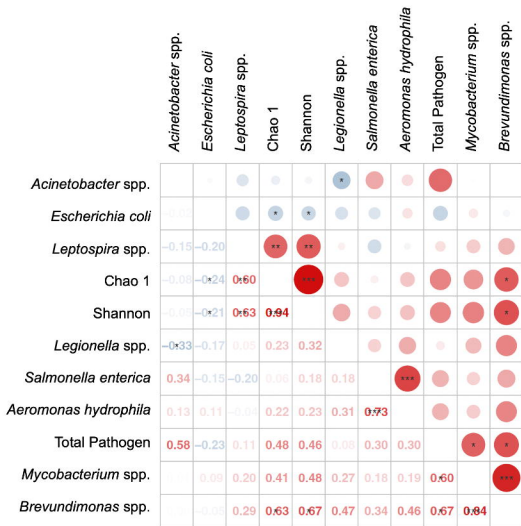
**Figure 5**

**Figure 6**

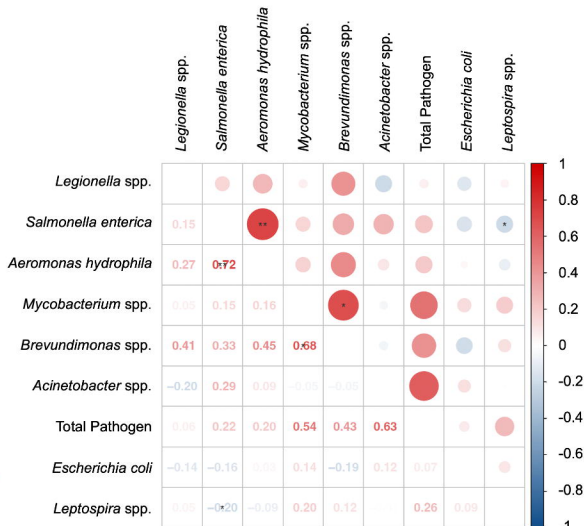


**Figure 7**

# Figure 8



**A**



**B**