medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

Xinyi Wen[1,2,3], Chutong Fang[1,4], Lihan Huang[1,5], Jiazheng Miao[1,6], Yajuan Lin[1,7]*

[1]Division of Natural and Applied Science, Duke Kunshan University, Kunshan, Jiangsu, China

[2]Nicholas School of the Environment, Duke University, Durham, NC, United States

[3]Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC, United States

[4]Blum Center for Developing Economies, University of California - Berkeley, Berkeley, CA, United States

[5]Department of Civil and Environmental Engineering, School of Engineering, Stanford University, Stanford, CA, United States

[6]Department of Biomedical Informatics, Harvard Medical School, Harvard University, Boston, MA, United States

[7]Department of Life Sciences, College of Science, Texas A&M University - Corpus Christi, Corpus Christi, TX, United States

*** Correspondence:**
Yajuan Lin
yajuan.lin@tamucc.edu

Keywords: Citizen Science$_1$, Household drinking water$_2$, Microbial communities$_3$, Waterborne pathogens$_4$, Metabarcoding$_5$.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

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

1 Introduction

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

2 Materials and Methods

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.1 Data Collection

2.1.1 Sites and Participants

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

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

2.1.2 Sample Collection

2.1.2.1 Preparation and Tool Kit

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

To further clarify the procedures and reduce variability in sample collection, an 11-minute video tutorial (720p resolution) was created, providing volunteers a visual guide on the tools and procedures for sample collection, filtration, storage, and shipping (Figure 1B & C and CS 3). Furthermore, in-person demonstrations were provided at DKU Environmental Research Lab for available volunteers following the demonstration model of Willis et al. (Willis et al., 2018). A compact sampling kit was distributed to each volunteer before their trip (Figure 1B). Each kit (in a 78.7×51.2×49.2 inch Styrofoam box) contained one set of all tools mentioned in the protocol and a sampling information form (CS 1 & 2) adapted from Buxton et al.'s (Buxton et al., 2018). This form collected information such as volunteer names, sampling time, geographic coordinate of the sampling site obtained from cell phone, etc. Alongside the form, volunteers were requested to label the Ziploc bag containing the sample with their names for convenient sample tracking in the stage of data collection. Upon recipient in lab, each sample was associated with an anonymous ID number and any data with identifiers was securely eliminated, ensuring no specific identity-related information was present during data processing. The ID number of each sample consists of two parts separated by an underscore – a two-letter abbreviation of the sampling site (city) followed by a four-digit sampling date (Table S1). For example, the sample collected in Beijing on June 20 was named as "BJ_0620" and the full names of other mentioned samples are provided in the "Non-standard abbreviations" section.

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

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

2.1.2.2 Sample Collection and Filtration

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

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

2.1.2.3 Sample Storage Test

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

2.1.3 DNA Extraction, PCR Amplification, and Bacterial 16S Metabarcoding

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(Qiagen) following the manufacturer's instructions. The purified PCR amplicons of each sample was dissolved in 30 μL elution buffer and stored in a -80□ freezer.

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

2.1.4 Auxiliary Data Collection

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

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

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

2.2 Data Analysis

2.2.1 Sequence Processing

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

2.2.2 Assessment of Total Microbial Communities

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

2.2.3 Assessment of Potential Waterborne Bacterial Pathogens

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.2.4 Statistical Analyses

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

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

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

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

3 Results and Discussion

3.1 Sample Validity and PCR Signals

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

3.2 Sample Storage Condition Influence

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

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

3.3 Sequence Reads, ASVs, and Taxonomy Classification

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

3.4 Total Microbial Communities

3.4.1 Compositions of Total Microbial Communities

3.4.1.1 Phylum Level Pattern by Sample

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The phylum-level taxonomic composition for each sample is detailed in Figure 3 and Table S3. In this study, *Proteobacteria* (class α- and γ-*Proteobacteria*) was the most predominant phylum in 20 samples and the second most in others. It accounts for 29.9 – 99.0% of the reads in each sample, with the highest relative abundance detected in Huizhou (HZ_0219). Among all the samples, *Sphingomonas* was the most abundant genus of *Proteobacteria*. This result was consistent with the previous study (Han et al., 2020) that found *Proteobacteria* to be dominant in tap water collected mostly from central and eastern China, and the genus *Sphingomonas* grew during chlorination (Jia et al., 2015) or monochloramine treatment (Chiao et al., 2014).

However, two samples CZ_0220 (Changzhou, Feb. 20) and NC_0210 (Nanchang, Feb.10) were dominated by *Crenarchaeota*, a common archaeal phylum. Specifically, *Crenarchaeota* accounted for 33.3% and 35.0% with the genus *Candidatus Nitrosotenuis* (32.0%) and *Candidatus Nitrosotalea* (34.9%) being most abundant in samples CZ_0220 and NC_0210 respectively. This showed that in addition to a variety of bacteria, archaea can also grow in tap water, which could be supported by the studies which detected the archaeal phylum *Crenarchaeota* in drinking water distribution systems or drinking water-related environments (Roeselers et al., 2015, Inkinen et al., 2021, Dai et al., 2020, Franca et al., 2015, Bautista-de los Santos et al., 2016). In particular, Inkinen et al. (Inkinen et al., 2021) found a high abundance of archaeal reads from the genus *Candidatus Nitrosotenuis* and *Candidatus Nitrosotalea* in drinking water distribution systems supplying non-disinfected waters. This suggests that the disinfection processes of samples CZ_0220 and NC_0210 may be less effective compared to others.

Interestingly, compared to CZ_0220, the pre- and post-typhoon samples collected from the same city household in Changzhou in July were much more similar in composition at the phylum level. However, the post-typhoon sample (CZ_0728) was richer in *Actinobacteria* (increased from 7.1% to 2.2%) and *Cyanobacteria* (increased from 5.2% to 14.9%) which are the phyla containing potential waterborne pathogens and the species that produce cyanotoxins respectively. Elevated levels of the pathogen *Mycobacterium* spp. (more details in Section 3.5 and Supplementary Material: SM1.1) as well as toxin producing *Cyanobacteria* spp. were identified in CZ_0728. Specifically for cyanobacteria, *Microcystis* spp. were higher in RA while *Cylindrospermopsis* sp. and *Dolichospermum* sp. appeared after the typhoon event. Moreover, other toxic species of *Cyanobacteria* including *Aphanizomenon* sp. and *Anabaena* sp. were detected in other tap samples collected from Shanghai, Lanzhou, Xi'an, etc. Many *Cyanobacteria* spp. from those genera can produce a variety of cyanotoxins such as Microcystins and Cylindrospermopsin which can cause damage to the liver and kidney with potential carcinogenicity (EPA, 2022). Similarly, a substantial rise in the RA of *Cyanobacteria* occurred in treated water samples collected from a drinking water treatment plant in Jiangsu was reported after Typhoon Lekima in August 2019 ($P < 0.05$) (Tang et al., 2021). In this study, although remained detectable, the proportion of *Cyanobacteria* decreased to the pre-typhoon level by the third day after the typhoon event.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

3.4.1.2 Class Level Taxonomy

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

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

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

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

3.4.2 Alpha-and Beta-Diversities

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

The extreme rainfall events had a clear impact on the alpha diversity of the microbiome in tap water samples collected from Changzhou and Zhengzhou, the two cities impacted by extreme rainfall events in 2021 summer. Compared to the two samples collected in winter (CZ_0220 & ZZ_0219), higher alpha diversity was observed in the post-weather summer samples in both cities (Figure 5C). Among the three samples collected from the same household in Changzhou, however, the highest diversity was observed in the pre-typhoon summer sample CZ_0712, indicating that outdoor temperatures might be a stronger driver of the water microbiome diversity than the typhoon.

Outdoor temperature plays a significant role in shaping the community composition and structure at the ASV level. The beta diversity of the total microbial communities in sampled tap water is shown in Figure 6. Principal Axis 1, 2, and 3 for PCoA (Bray-Curtis dissimilarity) represent 20.4%, 14.5%, and 8.1% of the variation among the samples respectively. The PCoA ordination illustrates that despite the different sampling regions, the samples in High T category clustered closely except BJ_0620 which was greatly dominated by *Proteobacteria* (87.5%). In contrast, samples in the other two categories, especially Medium T exhibited dispersive distributions. According to PERMANOVA and PERMDISP (Table S7), the influence of outdoor temperatures was statistically significant (R^2 = 0.24, adjusted P = 0.012 between High T and Low T), and the differentiation was not due to differences in group dispersions (P = 0.17). However, when making comparisons among normal samples (Figure S3 and SM 1.2), samples from different regions were greatly mixed and scattered except for the two samples in R7: Southwest (CQ_0715 & LJ_0722), which corresponds to the results of top 20 ASV analysis (Figure S2). A Mantel test (Table S8) indicate a strong correlation between microbial community structures and outdoor temperatures (Spearman's r = 0.898, P = 0.0001), whereas geographic locations had minimal impact (Spearman's r = -0.179, P = 0.9792).

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

The pre-typhoon summer sample CZ_0712 was distinct from the other two Changzhou samples but much like other samples in High T category, which further confirmed the impact of high outdoor temperature. Comparing the samples collected from Changzhou and Zhengzhou, the pre-weather winter samples differed moderately in microbial community composition and structure, however, the post-weather samples from the two cities (purple circle and square in Figure 6) were comparable to a large extent. This indicates a potential contamination mechanism – extreme rainfall events may cause the same or highly similar shift in tap water microbiomes by favoring certain fast-growing bacteria, increasing the similarity between dissimilar tap water microbial communities.

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

3.5 Occurrence of Potential Pathogenic Bacteria in Drinking Water Microbial Communities

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

3.5.1 Representative Pathogen Species

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

E. coli, a common fecal indicator bacteria (WHO, 2017), was detected by this study. The most polluted tap water samples are CZ_0728 (RA: 0.61%), BJ_0620 (0.51%), ZZ_0802 (0.05%), CZ_0712 (0.03%), and JN_0219 (0.02%). The elevated RA of *E. coli* in the two post-weather samples suggests that the contamination might be related to the extreme rainfall events. Notably, the proportion of *E. coli* in BJ_0620 was dramatically higher than that of other normal samples. It is worth mentioning that the traditional *E. coli* tests can be a generally reliable indicator of enteropathogenic serotypes in drinking water, but potential viable while non-culturable *E. coli* cells could result in underestimations of the actual water contamination (Liu et al., 2008). Therefore, it is recommended to use PCR or quantitative PCR (qPCR) methods for the monitoring of *E. coli* (Wolf-Baca and Siedlecka, 2019).

Additionally, the RA of total *Legionella* spp. was most abundant in ZZ_0802 (1.32%), CZ_0220 (1.25%), and ZB_0502 (1.09%). Almost all species in the genera *Legionella* are thought to be potential human pathogens, but *L. pneumophila* (on Contaminant Candidate List 5 - CCL 5) is the main cause of Legionnaires' disease (pneumonia) and Pontiac fever (a milder infection) (WHO, 2017, EPA, 2023). Potential *L. pneumophila* ASVs (Blastn best hit results) was detected in 22.7% (5) of the tap water samples (3 from Low T, 1 from Median and High T respectively). Among those samples, the highest *L. pneumophila* RA was observed in ZZ_0219 (RA: 0.237%), followed by MA_0524 (0.035%), NC_0210 (0.019%), XA_0217 (0.011%), and SH_0411 (0.010%). Moreover, some other pathogenic species such as *L. oakridgensis* and *L. maceachernii* occurred in TJ_0715, LZ_0822, XA_0217, LJ_0722, and ZZ_0802.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Salmonella spp., a highly pathogenic bacteria, was detected in one sample despite its known susceptibility to disinfection. Notably, *Salmonella enterica* (ASV 3082, RA: 0.39%) was observed in the post-flood tap water microbiome from Zhengzhou (ZZ_0802). However, the potential health risks remains uncertain, as the severity of the disease depends on the serotype and host factors of *Salmonella* (WHO, 2017). Nonetheless, the presence of *Salmonella enterica* after 2021 Henan Floods indicates potential fecal contamination in the household drinking water after an extreme weather event.

3.5.2 Potential Environmental Drivers: Extreme Precipitation and Hot Weather

The detection of a wide range of pathogens in tap water indicates the great impacts of extreme rainfall events on urban water management. Abnormal redistribution of precipitation patterns or extreme rainfall events could cause more rainfall to be concentrated in a short period, resulting in flooding and a large amount of effluent. These events can lead to huge challenges to urban water safety and pose severe threats on public and environmental health. On the one hand, excessive effluent load on urban wastewater treatment can overwhelm wastewater treatment plants, leading to untreated sewage being released into natural waterways, and resulting in freshwater resource pollution and aquatic ecosystem degradation (Langeveld et al., 2013). On the other hand, destructive rainfall could pollute drinking water sources and cause damage to buildings and infrastructure, such as leaks and breakage of tap water distribution systems. This will eventually lead to drinking water contamination and increased infection risks, which has been supported by the analysis of microbial community and pathogen changes in tap water in this research. As shown in Table S9, the typhoon and flood events generally altered the pathogen profiles in tap water samples by increasing the relative abundance of certain species and the number of species. Specifically, the RA of the total potential pathogens in the two post-weather samples ranked second (CZ_0728, 20.7%) and third (ZZ_0802, 10.7%) among all the samples (Figure 7). In addition, the number of pathogen species in Changzhou samples collected in July increased after Typhoon In-Fa. A similar pattern was observed in Zhengzhou samples and ZZ_0802 was the one with the highest number of pathogen species. More importantly, some devastating pathogen species (*Salmonella* spp. and *Aeromonas hydrophila*) were only detected in the post-weather samples as discussed above. In terms of the elevated RA of specific genera/species, most of the pathogens in the post-typhoon sample CZ_0728 had higher proportions compared to the pre-typhoon summer sample CZ_0712, except for *Legionella* spp. and *Brevundimonas* spp. which saw a decrease in RA, and *Bacillus* sp. being not detected in CZ_0728. Therefore, the rising risk of waterborne pathogens in post-flush tap water samples indicates that extreme rainfall events are a prominent risk factor in public health in terms of water (and food) safety (Phan and Sherchan, 2020). To maintain the standard of the treated discharge water during extreme rainfall events in the future, the government needs to invest more to upgrade the water processing capacity of wastewater treatment plants to be prepared for unpredictable effluent load. Meanwhile, since drinking water resource pollution is expected to happen following flooding, drinking water treatment plants and distribution systems should be designed, monitored, and maintained involving

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

Hot weather (increase of temperature) is another critical climatic factor that will strongly influence the hydrological regime and facilitate the spreading and proliferation of waterborne pathogen in freshwater resources. According to IPCC Technical Paper VI: Climate Change and Water (IPCC, 2008), the increase of surface water temperature has been associated with the atmospheric warming across Asia, North America, and Europe since 1960s. Lipp et al. has discovered that increasing temperature could create conditions that are advantageous for the proliferation of bacteria species less susceptible to temperature variations (Lipp et al., 2002). This could directly trigger the growth and expansion of some pathogenic species, such as *cholera* (*V. cholerae*), *Vibrio parahaemolyticus*, *V. vulnificus* and *V. alginolyticus* (Funari et al., 2012). Our results confirmed the above connections between climate change and increased waterborne infection risk by showing that hot weather is potentially another important environmental driver of pathogen profiles of tap water. The outdoor temperatures higher or equal than 17°C were positively correlated with RAs of total potential pathogens but the correlation is insignificant relative to the alpha level of 0.05 probably due to the small sample size ($n = 12$, $R = 0.46$, $P = 0.13$) (Figure S4). Besides that, higher RAs and a larger number of pathogen species were detected in the summer Changzhou sample CZ_0712 (20.7%, 7 genera, outdoor T: 33.3°C) compared to the winter sample CZ_0220 (1.80%, 4 genera, outdoor T: 17.0°C).

An important aspect to highlight here is that, apart from the separate effects of extreme rainfall and hot extremes on waterborne disease spreading, these extreme weather events are also intercorrelated in a way that creates a detrimental cycle, exacerbating the overall situation. IPCC AR6 report has confirmed that regardless of the emissions scenarios, the Earth's surface temperature will keep increasing until at least the mid-21st century. Unless substantial reductions in CO2 and other greenhouse gas emissions occur in the coming decades, the global warming limits of 1.5°C and 2°C will be exceeded during the 21st century. Most importantly, the report underlined the fact with high confidence that with every additional increment of global warming, the impact on extreme weather events becomes more pronounced. For instance, for each additional 0.5°C rise in global temperatures, there are noticeable and undeniable rises intensity and frequency of extreme heat events, including heatwaves and intense rainfall (Masson-Delmotte et al., 2021). As a result, ensuring tap water safety in the future will become increasingly challenging and require stricter control and maintenance as the effects of climate change continue to intensify.

3.5.3 Potential Monitoring and Control Strategy

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

The incorporation of the two post-weather samples (Figure 8B, n = 24) altered the correlations between different pathogens and revealed new relationships due to the occurrence of other species in the samples. The positive correlation between the RA of *Brevundimonas* spp. and that of *Mycobacterium* spp. became weaker (r = 0.68, P < 0.05) while the RA of *Aeromonas hydrophila* was significantly correlated with *S. enterica* (r = 0.72, P < 0.01). Interestingly, *S. enterica* and *B. diminuta* were only detected in the post-flood sample ZZ_0802, suggesting potential co-occurring relationships between them. Future studies could further investigate the above potential relationships among waterborne pathogens in household drinking water.

3.6 Citizen Science Approach

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

3.7 Other Strengths, Limitations, and Future Work

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

4 Conclusions

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

Albus K., Thompson R. & Mitchell F. (2019). Usability of Existing Volunteer Water Monitoring Data: What Can the Literature Tell Us? *Citizen Science: Theory and Practice,* 4.

Allen M. J., Edberg S. C. & Reasoner D. J. (2004). Heterotrophic plate count bacteria—what is their significance in drinking water? *International journal of food microbiology,* 92**,** 265-274.

Altenburger R., Brack W., Burgess R. M., Busch W., Escher B. I., Focks A., Mark Hewitt L., Jacobsen B. N., de Alda M. L. & Ait-Aissa S. (2019). Future water quality monitoring: improving the balance between exposure and toxicity assessments of real-world pollutant mixtures. *Environmental Sciences Europe,* 31**,** 1-17.

Andermann T., Antonelli A., Barrett R. L. & Silvestro D. (2022). Estimating alpha, beta, and gamma diversity through deep learning. *Frontiers in plant science,* 13**,** 839407.

Anderson M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology,* 26**,** 32-46.

Anderson M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics,* 62**,** 245-53. doi: 10.1111/j.1541-0420.2005.00440.x.

Arbizu M. (2020). *pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4* [Online]. github. Available: https://github.com/pmartinezarbizu/pairwiseAdonis [Accessed April 23 2022].

Ashbolt N. J. (2015). Microbial Contamination of Drinking Water and Human Health from Community Water Systems. *Curr Environ Health Rep,* 2**,** 95-106. doi: 10.1007/s40572-014-0037-5.

Baalbaki R., Ahmad S. H., Kays W., Talhouk S. N., Saliba N. A. & Al-Hindi M. (2019). Citizen science in Lebanon-a case study for groundwater quality monitoring. *R Soc Open Sci,* 6**,** 181871. doi: 10.1098/rsos.181871.

Bautista-de los Santos Q. M., Schroeder J. L., Sevillano-Rivera M. C., Sungthong R., Ijaz U. Z., Sloan W. T. & Pinto A. J. (2016). Emerging investigators series: microbial communities in full-scale drinking water distribution systems–a meta-analysis. *Environmental Science: Water Research & Technology,* 2**,** 631-644.

Borchardt M. A., Boehm A. B., Salit M., Spencer S. K., Wigginton K. R. & Noble R. T. (2021). The environmental microbiology minimum information (EMMI) guidelines: qPCR and dPCR quality and reporting for environmental microbiology. *Environmental Science & Technology,* 55**,** 10210-10223.

Brettar I. & Hofle M. G. (2008). Molecular assessment of bacterial pathogens - a contribution to drinking water safety. *Curr Opin Biotechnol,* 19**,** 274-80. doi: 10.1016/j.copbio.2008.04.004.

Brouwer S., Van der Wielen P. W., Schriks M., Claassen M. & Frijns J. (2018). Public participation in science: the future and value of citizen science in the drinking water research. *Water,* 10**,** 284.

Buxton A. S., Groombridge J. J. & Griffiths R. A. (2018). Comparison of two citizen scientist methods for collecting pond water samples for environmental DNA studies. *Ctizen Science: Theory and Practice,* 3.

Cai D. 2023. Climate change worsened Asia's April heatwave by 2C - study. BBC News.

Callahan B. J., McMurdie P. J., Rosen M. J., Han A. W., Johnson A. J. & Holmes S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods,* 13**,** 581-3. doi: 10.1038/nmeth.3869.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K. & Madden T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421. doi: 10.1186/1471-2105-10-421.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Cann K. F., Thomas D. R., Salmon R. L., Wyn-Jones A. P. & Kay D. (2013). Extreme water-related weather events and waterborne disease. *Epidemiol Infect,* 141**,** 671-86. doi: 10.1017/S0950268812001653.

Chen S. (2021). Zhengzhou Rainstorm Once in a thousand years? The meteorological authority responded: A year's rain fell in one day. *Chinanews*.

Chiao T. H., Clancy T. M., Pinto A., Xi C. & Raskin L. (2014). Differential resistance of drinking water bacterial populations to monochloramine disinfection. *Environ Sci Technol,* 48**,** 4038-47. doi: 10.1021/es4055725.

China D. I. T. o. T. S. C. o. t. P. s. R. o. 2022. Investigation report on Zhengzhou 7.20 Rainstrom Disaster. *In:* China, M. o. E. M. o. t. P. s. R. o. (ed.).

Chopyk J., Moore R. M., DiSpirito Z., Stromberg Z. R., Lewis G. L., Renter D. G., Cernicchiaro N., Moxley R. A. & Wommack K. E. (2016). Presence of pathogenic Escherichia coli is correlated with bacterial community diversity and composition on pre-harvest cattle hides. *Microbiome,* 4**,** 1-11.

Cissé G. (2019). Food-borne and water-borne diseases under climate change in low-and middle-income countries: Further efforts needed for reducing environmental health exposure risks. *Acta tropica,* 194**,** 181-188.

Commission E. 2020. COMMISSION STAFF WORKING DOCUMENT: Best Practices in Citizen Science for Environmental Monitoring. European Commission.

Corning. (2023). *Syringe Filters* [Online]. Available: https://ecatalog.corning.com/life-sciences/b2c/US/en/Cell-Culture/Filtration/Syringe-Filters/Syringe-Filters/p/syringeFilters [Accessed May 12 2023].

Dai Z., Sevillano-Rivera M. C., Calus S. T., Bautista-de Los Santos Q. M., Eren A. M., van der Wielen P., Ijaz U. Z. & Pinto A. J. (2020). Disinfection exhibits systematic impacts on the drinking water microbiome. *Microbiome,* 8**,** 42. doi: 10.1186/s40168-020-00813-0.

EPA. (2022). *Learn about Cyanobacteria and Cyanotoxins* [Online]. Available: https://www.epa.gov/cyanohabs/learn-about-cyanobacteria-and-cyanotoxins#what3 [Accessed April 23 2022].

EPA. (2023). *Contaminant Candidate List (CCL) and Regulatory Determination* [Online]. Available: https://www.epa.gov/ccl/contaminant-candidate-list-5-ccl-5 [Accessed May 25 2023].

Franca L., Lopez-Lopez A., Rossello-Mora R. & da Costa M. S. (2015). Microbial diversity and dynamics of a groundwater and a still bottled natural mineral water. *Environ Microbiol,* 17**,** 577-93. doi: 10.1111/1462-2920.12430.

Funari E., Manganelli M. & Sinisi L. (2012). Impact of climate change on waterborne diseases. *Annali dell'Istituto superiore di sanita,* 48**,** 473-487.

Garner E., Davis B. C., Milligan E., Blair M. F., Keenum I., Maile-Moskowitz A., Pan J., Gnegy M., Liguori K. & Gupta S. (2021). Next generation sequencing approaches to evaluate water and wastewater quality. *Water Research,* 194**,** 116907.

Haegeman B., Hamelin J., Moriarty J., Neal P., Dushoff J. & Weitz J. S. (2013). Robust estimation of microbial diversity in theory and in practice. *ISME J,* 7**,** 1092-101. doi: 10.1038/ismej.2013.10.

Han Z., An W., Yang M. & Zhang Y. (2020). Assessing the impact of source water on tap water bacterial communities in 46 drinking water supply systems in China. *Water Res,* 172**,** 115469. doi: 10.1016/j.watres.2020.115469.

Harper L. (2018). *The Importance of Citizen Scientists* [Online]. American Forests. Available: https://www.americanforests.org/blog/the-importance-of-citizen-scientists/ [Accessed April 23 2022].

Ho S. Y., Xu S. J. & Lee F. W. (2020). Citizen science: An alternative way for water monitoring in Hong Kong. *PLoS One,* 15**,** e0238349. doi: 10.1371/journal.pone.0238349.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Holinger E. P., Ross K. A., Robertson C. E., Stevens M. J., Harris J. K. & Pace N. R. (2014). Molecular analysis of point-of-use municipal drinking water microbiology. *Water Res,* 49**,** 225-35. doi: 10.1016/j.watres.2013.11.027.

Hull N. M., Holinger E. P., Ross K. A., Robertson C. E., Harris J. K., Stevens M. J. & Pace N. R. (2017). Longitudinal and source-to-tap New Orleans, LA, USA drinking water microbiology. *Environmental Science & Technology,* 51**,** 4220-4229.

Inkinen J., Siponen S., Jayaprakash B., Tiwari A., Hokajarvi A. M., Pursiainen A., Ikonen J., Kauppinen A., Miettinen I. T., Paananen J., Torvinen E., Kolehmainen M. & Pitkanen T. (2021). Diverse and active archaea communities occur in non-disinfected drinking water systems-Less activity revealed in disinfected and hot water systems. *Water Res X,* 12**,** 100101. doi: 10.1016/j.wroa.2021.100101.

IPCC 2008. Climate Change and Water. *In:* Bates, B., Kundzewicz, Z. W., Wu, S. & Palutikof, J. (eds.).

Ivone V.-M., Conceição E., Olga C N. & Célia M M. (2013). Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. *FEMS microbiology ecology,* 83**,** 361-374.

Ji P., Parks J., Edwards M. A. & Pruden A. (2015). Impact of Water Chemistry, Pipe Material and Stagnation on the Building Plumbing Microbiome. *PLoS One,* 10**,** e0141087. doi: 10.1371/journal.pone.0141087.

Jia S., Shi P., Hu Q., Li B., Zhang T. & Zhang X. X. (2015). Bacterial Community Shift Drives Antibiotic Resistance Promotion during Drinking Water Chlorination. *Environ Sci Technol,* 49**,** 12271-9. doi: 10.1021/acs.est.5b03521.

Johnson J. S., Spakowicz D. J., Hong B. Y., Petersen L. M., Demkowicz P., Chen L., Leopold S. R., Hanson B. M., Agresta H. O., Gerstein M., Sodergren E. & Weinstock G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun,* 10**,** 5029. doi: 10.1038/s41467-019-13036-1.

Kozich J. J., Westcott S. L., Baxter N. T., Highlander S. K. & Schloss P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol,* 79**,** 5112-20. doi: 10.1128/AEM.01043-13.

Langeveld J., Schilperoort R. & Weijers S. (2013). Climate change and urban wastewater infrastructure: there is more to explore. *Journal of Hydrology,* 476**,** 112-119.

Laws E. A. (2017). *Aquatic Pollution: An Introductory Text,* Wiley.

Levy K., Smith S. M. & Carlton E. J. (2018). Climate change impacts on waterborne diseases: moving toward designing interventions. *Current environmental health reports,* 5**,** 272-282.

Ley C. J., Proctor C. R., Singh G., Ra K., Noh Y., Odimayomi T., Salehi M., Julien R., Mitchell J. & Nejadhashemi A. P. (2020). Drinking water microbiology in a water-efficient building: stagnation, seasonality, and physicochemical effects on opportunistic pathogen and total bacteria proliferation. *Environmental Science: Water Research & Technology,* 6**,** 2902-2913.

Lim J. Y., Yoon J. W. & Hovde C. J. (2010). A brief overview of Escherichia coli O157: H7 and its plasmid O157. *Journal of microbiology and biotechnology,* 20**,** 5.

Lin Y., Gifford S., Ducklow H., Schofield O. & Cassar N. (2019). Towards Quantitative Microbiome Community Profiling Using Internal Standards. *Appl Environ Microbiol,* 85**,** e02634-18. doi: 10.1128/AEM.02634-18.

Ling F., Whitaker R., LeChevallier M. W. & Liu W. T. (2018). Drinking water microbiome assembly induced by water stagnation. *ISME J,* 12**,** 1520-1531. doi: 10.1038/s41396-018-0101-5.

Lipp E. K., Huq A. & Colwell R. R. (2002). Effects of global climate on infectious disease: the cholera model. *Clinical microbiology reviews,* 15**,** 757-770.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Liu Y., Gilchrist A., Zhang J. & Li X. F. (2008). Detection of viable but nonculturable Escherichia coli O157:H7 bacteria in drinking water and river water. *Appl Environ Microbiol,* 74**,** 1502-7. doi: 10.1128/AEM.02125-07.

Ma L., Jiang X. T., Guan L., Li B. & Zhang T. (2022). Nationwide biogeography and health implications of bacterial communities in household drinking water. *Water Res,* 215**,** 118238. doi: 10.1016/j.watres.2022.118238.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal,* 17**,** 10-12.

Masson-Delmotte V., Zhai P., Pirani A., Connors S. L., Péan C., Berger S., Caud N., Chen Y., Goldfarb L. & Gomis M. (2021). Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change,* 2.

Matsuo Y., Komiya S., Yasumizu Y., Yasuoka Y., Mizushima K., Takagi T., Kryukov K., Fukuda A., Morimoto Y. & Naito Y. (2021). Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC microbiology,* 21**,** 1-13.

McKight P. E. & Najab J. (2010). Kruskal-wallis test. *The corsini encyclopedia of psychology***,** 1-1.

McMurdie P. J. & Holmes S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One,* 8**,** e61217. doi: 10.1371/journal.pone.0061217.

Metcalfe A. N., Kennedy T. A., Mendez G. A. & Muehlbauer J. D. (2022). Applied citizen science in freshwater research. *Wiley Interdisciplinary Reviews: Water,* 9**,** e1578.

MilliporeSigma. (2023). *Sterivex* [Online]. Available: https://www.sigmaaldrich.com/US/en/search/sterivex?focus=products&page=1&perpage=30&sort=relevance&term=sterivex&type=product [Accessed May 12 2023].

Motlagh A. M. & Yang Z. (2019). Detection and occurrence of indicator organisms and pathogens. *Water Environment Research,* 91**,** 1402-1408.

NMC. (2021). National Meteorological Center. Available: http://www.nmc.cn/ [Accessed February 14 2022].

NMC-Typhoon. (2021). *2106 In-fa* [Online]. National Meteorological Centre - Typhoon. Available: http://typhoon.nmc.cn/web.html [Accessed February 24 2022].

Oksanen J., Blanchet F. & Kindt R. 2015. Vegan: commity ecology package. R package version 2.3-0.

Paddison L. 2023. Heat waves in US and Europe would have been 'virtually impossible' without climate change, new report finds. CNN.

Pandey P. K., Kass P. H., Soupir M. L., Biswas S. & Singh V. P. (2014). Contamination of water resources by pathogenic bacteria. *AMB Express,* 4**,** 51. doi: 10.1186/s13568-014-0051-x.

Phan N. K. & Sherchan S. P. (2020). Microbiological Assessment of Tap Water Following the 2016 Louisiana Flooding. *Int J Environ Res Public Health,* 17**,** 1273. doi: 10.3390/ijerph17041273.

Pinto A. J., Schroeder J., Lunn M., Sloan W. & Raskin L. (2014). Spatial-temporal survey and occupancy-abundance modeling to predict bacterial community dynamics in the drinking water microbiome. *mBio,* 5**,** e01135-14. doi: 10.1128/mBio.01135-14.

Pinto A. J., Xi C. & Raskin L. (2012). Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environmental science & technology,* 46**,** 8851-8859.

Pocock M. J., Roy H. E., Fox R., Ellis W. N. & Botham M. (2017). Citizen science and invasive alien species: predicting the detection of the oak processionary moth Thaumetopoea processionea by moth recorders. *Biological conservation,* 208**,** 146-154.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Preheim S. P., Perrotta A. R., Friedman J., Smilie C., Brito I., Smith M. B. & Alm E. (2013). Computational methods for high-throughput comparative analyses of natural microbial communities. *Methods in enzymology.* Elsevier.

Ramirez K. S., Knight C. G., de Hollander M., Brearley F. Q., Constantinides B., Cotton A., Creer S., Crowther T. W., Davison J., Delgado-Baquerizo M., Dorrepaal E., Elliott D. R., Fox G., Griffiths R. I., Hale C., Hartman K., Houlden A., Jones D. L., Krab E. J., Maestre F. T., McGuire K. L., Monteux S., Orr C. H., van der Putten W. H., Roberts I. S., Robinson D. A., Rocca J. D., Rowntree J., Schlaeppi K., Shepherd M., Singh B. K., Straathof A. L., Bhatnagar J. M., Thion C., van der Heijden M. G. A. & de Vries F. T. (2018). Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat Microbiol,* 3**,** 189-196. doi: 10.1038/s41564-017-0062-x.

Ramirez-Castillo F. Y., Loera-Muro A., Jacques M., Garneau P., Avelar-Gonzalez F. J., Harel J. & Guerrero-Barrera A. L. (2015). Waterborne pathogens: detection methods and challenges. *Pathogens,* 4**,** 307-34. doi: 10.3390/pathogens4020307.

Roeselers G., Coolen J., van der Wielen P. W., Jaspers M. C., Atsma A., de Graaf B. & Schuren F. (2015). Microbial biogeography of drinking water: patterns in phylogenetic diversity across space and time. *Environ Microbiol,* 17**,** 2505-14. doi: 10.1111/1462-2920.12739.

Roy H. E., Pocock M. J., Preston C. D., Roy D. B., Savage J., Tweddle J. & Robinson L. (2012). Understanding citizen science and environmental monitoring: final report on behalf of UK Environmental Observation Framework.

Ryan M. P. & Pembroke J. T. (2018). Brevundimonas spp: Emerging global opportunistic pathogens. *Virulence,* 9**,** 480-493. doi: 10.1080/21505594.2017.1419116.

Smouse P. E., Long J. C. & Sokal R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic zoology,* 35**,** 627-632.

Tang W., Liu Y., Li Q., Chen L., Li Q., Li P. & Xia S. (2021). The Impact of Extreme Weather Events on Bacterial Communities and Opportunistic Pathogens in a Drinking Water Treatment Plant. *Water,* 14**,** 54.

Thompson L. R., Sanders J. G., McDonald D., Amir A., Ladau J., Locey K. J., Prill R. J., Tripathi A., Gibbons S. M., Ackermann G., Navas-Molina J. A., Janssen S., Kopylova E., Vazquez-Baeza Y., Gonzalez A., Morton J. T., Mirarab S., Zech Xu Z., Jiang L., Haroon M. F., Kanbar J., Zhu Q., Jin Song S., Kosciolek T., Bokulich N. A., Lefler J., Brislawn C. J., Humphrey G., Owens S. M., Hampton-Marcell J., Berg-Lyons D., McKenzie V., Fierer N., Fuhrman J. A., Clauset A., Stevens R. L., Shade A., Pollard K. S., Goodwin K. D., Jansson J. K., Gilbert J. A., Knight R. & Earth Microbiome Project C. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature,* 551**,** 457-463. doi: 10.1038/nature24621.

Underground W. 2022.

Upton G. J. (1992). Fisher's exact test. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 155**,** 395-402.

van der Wielen P. W. & van der Kooij D. (2010). Effect of water composition, distance and season on the adenosine triphosphate concentration in unchlorinated drinking water in the Netherlands. *Water Res,* 44**,** 4860-7. doi: 10.1016/j.watres.2010.07.016.

Vej G. W. 2007. Bioinformatics explained: BLAST. CLC bio.

Wang T., Sun D., Zhang Q. & Zhang Z. (2021). China's drinking water sanitation from 2007 to 2018: A systematic review. *Sci Total Environ,* 757**,** 143923. doi: 10.1016/j.scitotenv.2020.143923.

Wen X., Chen F., Lin Y., Zhu H., Yuan F., Kuang D., Jia Z. & Yuan Z. (2020). Microbial indicators and their use for monitoring drinking water quality—A review. *Sustainability,* 12**,** 2249.

WHO 2017. Guidelines for Drinking-Water Quality (Fourth Edition).

Wikipedia. (2021). *Typhoon In-fa* [Online]. Wikimedia Foundation. Available: https://en.wikipedia.org/wiki/Typhoon_In-fa [Accessed February 14 2022].

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Willis J. R., Gonzalez-Torres P., Pittis A. A., Bejarano L. A., Cozzuto L., Andreu-Somavilla N., Alloza-Trabado M., Valentin A., Ksiezopolska E., Company C., Onywera H., Montfort M., Hermoso A., Iraola-Guzman S., Saus E., Labeeuw A., Carolis C., Hecht J., Ponomarenko J. & Gabaldon T. (2018). Citizen science charts two major "stomatotypes" in the oral microbiome of adolescents and reveals links with habits and drinking water composition. *Microbiome,* 6**,** 218. doi: 10.1186/s40168-018-0592-3.

Wolf-Baca M. & Siedlecka A. (2019). Detection of pathogenic bacteria in hot tap water using the qPCR method: preliminary research. *SN Applied Sciences,* 1**,** 1-9.

Wu Y., Washbourne C. & Haklay M. (2022). Citizen science in China's water resources monitoring: current status and future prospects. *International Journal of Sustainable Development & World Ecology,* 29**,** 277-290.

Xie J., Jin L., Wu D., Pruden A. & Li X. (2022). Inhalable Antibiotic Resistome from Wastewater Treatment Plants to Urban Areas: Bacterial Hosts, Dissemination Risks, and Source Contributions. *Environ Sci Technol,* 56**,** 7040-7051. doi: 10.1021/acs.est.1c07023.

Yang Y., Xing S., Chen Y., Wu R., Wu Y., Wang Y., Mi J. & Liao X. (2021). Profiles of bacteria/phage-comediated ARGs in pig farm wastewater treatment plants in China: Association with mobile genetic elements, bacterial communities and environmental factors. *J Hazard Mater,* 404**,** 124149. doi: 10.1016/j.jhazmat.2020.124149.

Yu P., Zaleski A., Li Q., He Y., Mapili K., Pruden A., Alvarez P. J. & Stadler L. B. (2018). Elevated levels of pathogenic indicator bacteria and antibiotic resistance genes after Hurricane Harvey's flooding in Houston. *Environmental science & technology letters,* 5**,** 481-486.

Yu X., Lin H. & Zhang S. (2012). Microbial contamination in drinking water: occurrence, measurement and control. *Bulletin of the Chinese Academy of Sciences BCAS,* 26**,** 84-91.

Zachariah M., Philip S., Pinto I., Vahlberg M., Singh R. & Otto F. E. L. 2023. Extreme heat in North America, Europe and China in July 2023 made much more likely by climate change. Grantham Institute for Climate Change.

Zhang H., Xu L., Huang T., Liu X., Miao Y., Liu K. & Qian X. (2021a). Indoor heating triggers bacterial ecological links with tap water stagnation during winter: Novel insights into bacterial abundance, community metabolic activity and interactions. *Environmental Pollution,* 269**,** 116094.

Zhang H., Xu L., Huang T., Yan M., Liu K., Miao Y., He H., Li S. & Sekar R. (2021b). Combined effects of seasonality and stagnation on tap water quality: Changes in chemical parameters, metabolic activity and co-existence in bacterial community. *Journal of Hazardous Materials,* 403**,** 124018.

Zhang J., Chen S., Chen B., Du Y., Huang X., Pan X. & Zhang Q. (2013). Citizen science: integrating scientific research, ecological conservation and public participation. *Biodiversity Science,* 21**,** 738.

Zhang J., Mauzerall D. L., Zhu T., Liang S., Ezzati M. & Remais J. V. (2010). Environmental health in China: progress towards clean air and safe water. *Lancet,* 375**,** 1110-9. doi: 10.1016/S0140-6736(10)60062-1.

Zhang Y., Deng J., Qin B., Zhu G., Zhang Y., Jeppesen E. & Tong Y. (2022). Importance and vulnerability of lakes and reservoirs supporting drinking water in China. *Fundamental Research*.

Zlatanović L., van der Hoek J. P. & Vreeburg J. (2017). An experimental study on the influence of water stagnation and temperature change on water quality in a full-scale domestic drinking water system. *Water research,* 123**,** 761-772.

Zou X.-Y., Peng X.-Y., Zhao X.-X. & Chang C.-P. (2023). The impact of extreme weather events on water quality: International evidence. *Natural Hazards,* 115**,** 1-21.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Non-standard abbreviations

RA: Relative abundance

CS: Citizen science sampling protocol and materials

SM: Supplementary material

NC: Nanchang, Jiangxi

XA: Xi'an, Shanxi

NJ: Nanjing, Jiangsu

HZ: Huizhou, Guangdong

JN: Jinan, Shandong

ZZ: Zhengzhou, Henan

CZ: Changzhou, Jiangsu

SH: Shanghai

ZB: Zibo, Shandong

MA: Macau

BJ: Beijing

TJ: Tianjin

CQ: Chongqing

LJ: Lijiang, Yunnan

LZ: Lanzhou, Gansu

Figure legends

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

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

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

Figure 5 The α diversity (ASV-based).

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

Figure 6 Principal coordinates analysis (PCoA).

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

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

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

Figure 8 Spearman's correlation.

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

Tables

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

Those three represent the coverage, diversity, and the genus spectrum of microbial community in the water samples respectively. [a] The percentage of sequences was the phylum sequences in the total of assigned sequences at the phylum level which was 1,333,322.

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

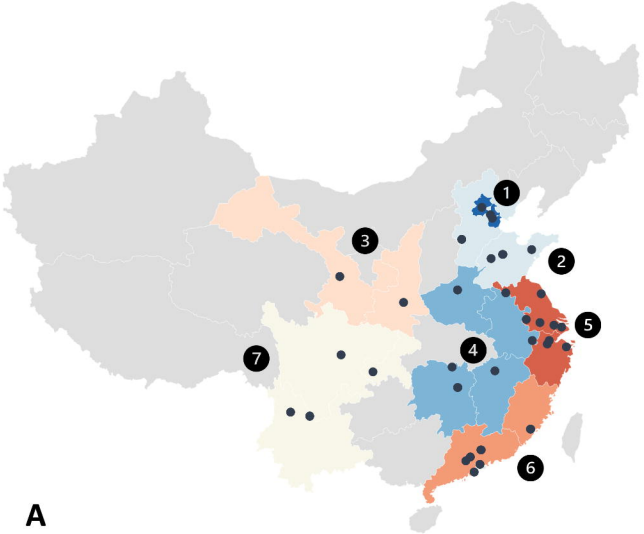
Table 3 Occurrence and mean relative abundance (RA) of major potential pathogens.

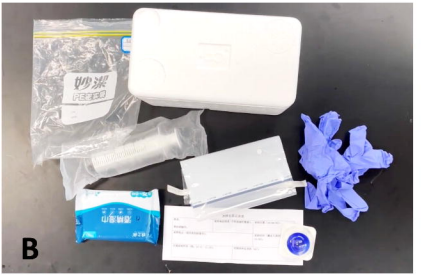
Pathogen Species	Occurrence	Mean RA
Mycobacterium spp.	22 (100.0%)	2.67E-02
Acinetobacter spp.	22 (100.0%)	1.43E-02
Legionella spp.	22 (100%)	4.59E-03
Brevundimonas spp.	20 (90.9%)	6.12E-03
Leptospira spp.	11 (50.0%)	2.04E-04

medRxiv preprint doi: https://doi.org/10.1101/2023.10.16.23297104; this version posted October 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

Figure 1





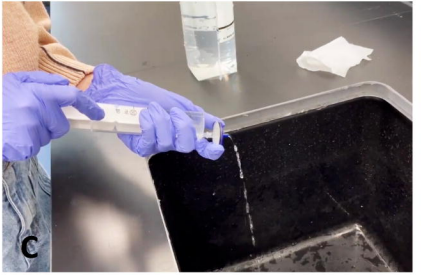


Figure 2

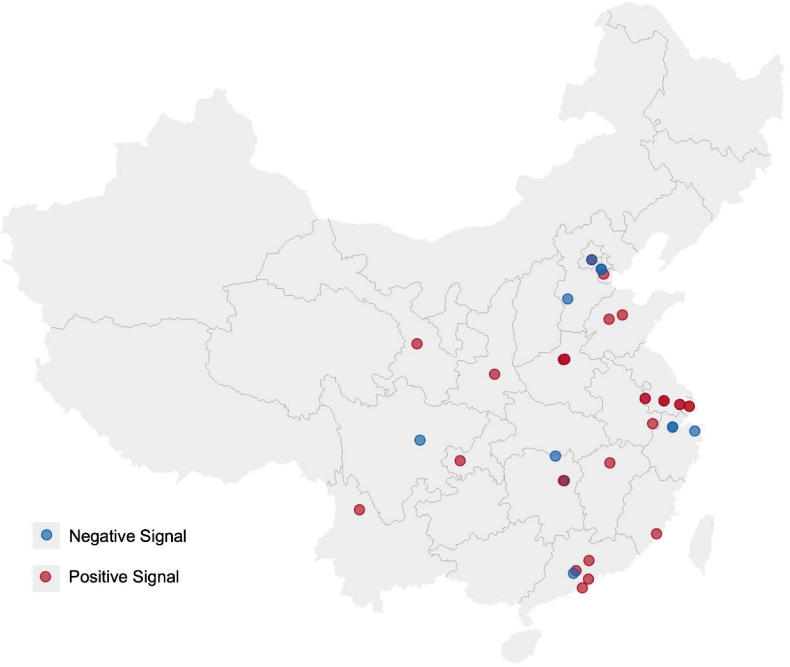


Figure 3

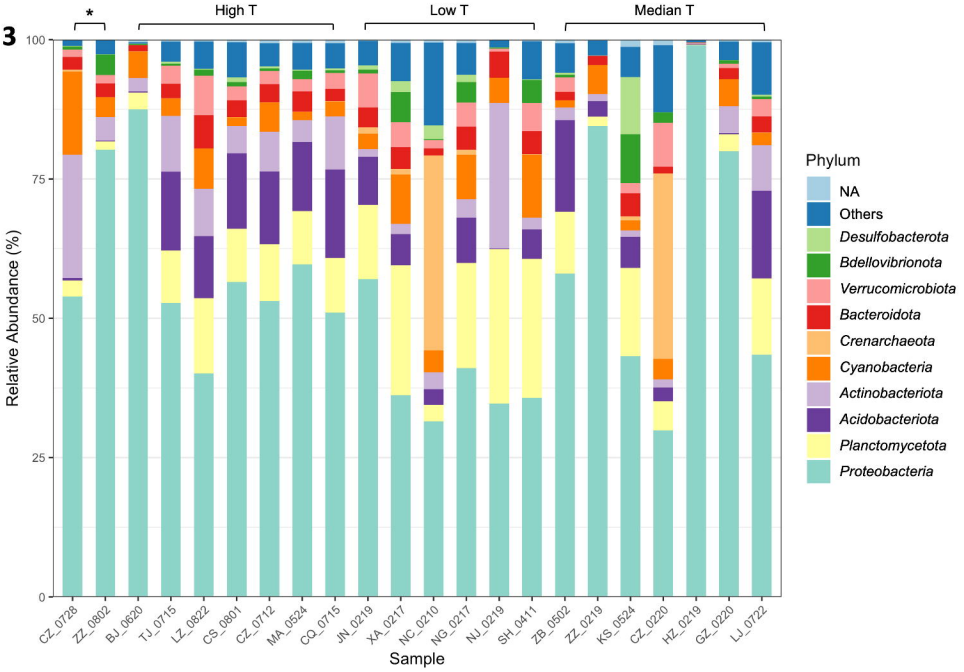

*
High T
Low T
Median T
100
75
50
25
0
Relative Abundance (%)
CZ_0728
ZZ_0802
BJ_0620
TJ_0715
LZ_0822
CS_0801
CZ_0712
MA_0524
CQ_0715
JN_0219
XA_0217
NC_0210
NG_0217
NJ_0219
SH_0411
ZB_0502
ZZ_0219
KS_0524
CZ_0220
HZ_0219
GZ_0220
LJ_0722
Sample
Phylum
NA
Others
Desulfobacterota
Bdellovibrionota
Verrucomicrobiota
Bacteroidota
Crenarchaeota
Cyanobacteria
Actinobacteriota
Acidobacteriota
Planctomycetota
Proteobacteria

Figure 4

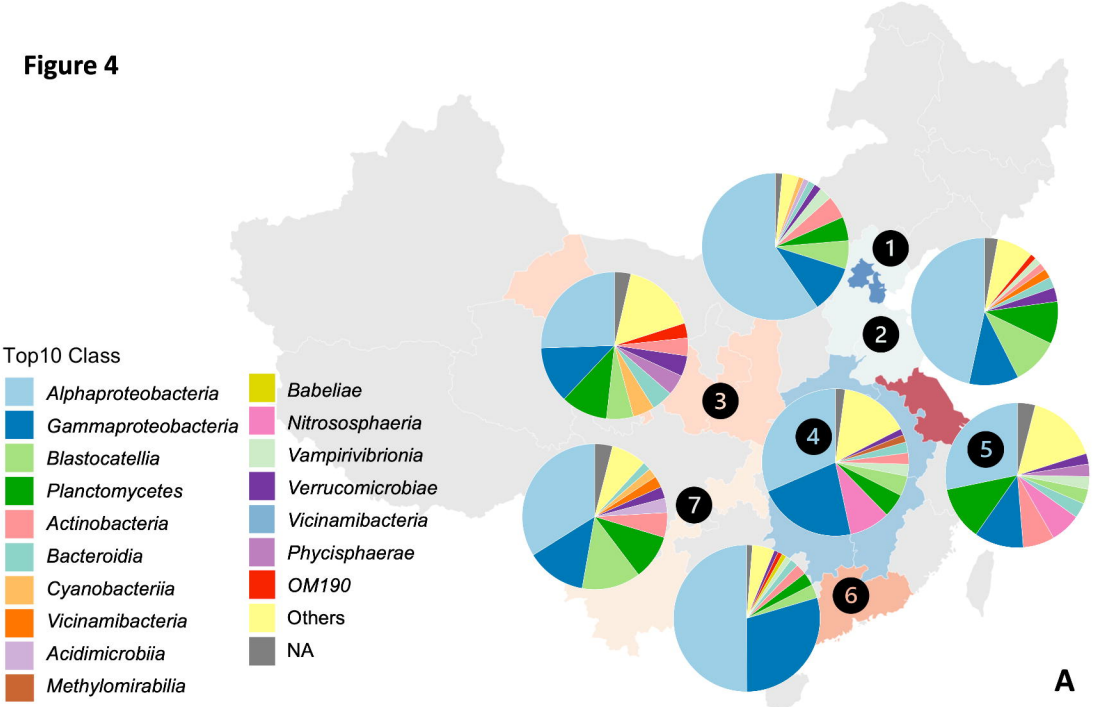
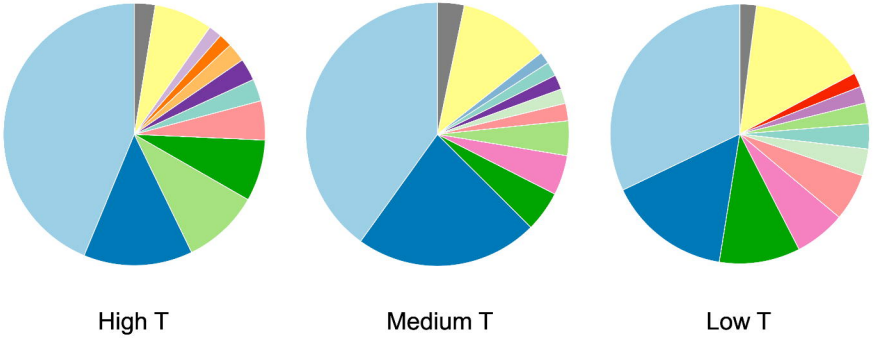

Top10 Class
Alphaproteobacteria
Gammaproteobacteria
Blastocatellia
Planctomycetes
Actinobacteria
Bacteroidia
Cyanobacteriia
Vicinamibacteria
Acidimicrobiia
Methylomirabilia
Babeliae
Nitrososphaeria
Vampirivibrionia
Verrucomicrobiae
Vicinamibacteria
Phycisphaerae
OM190
Others
NA
1
2
3
4
5
6
7
A

Figure 4



B

Figure 5

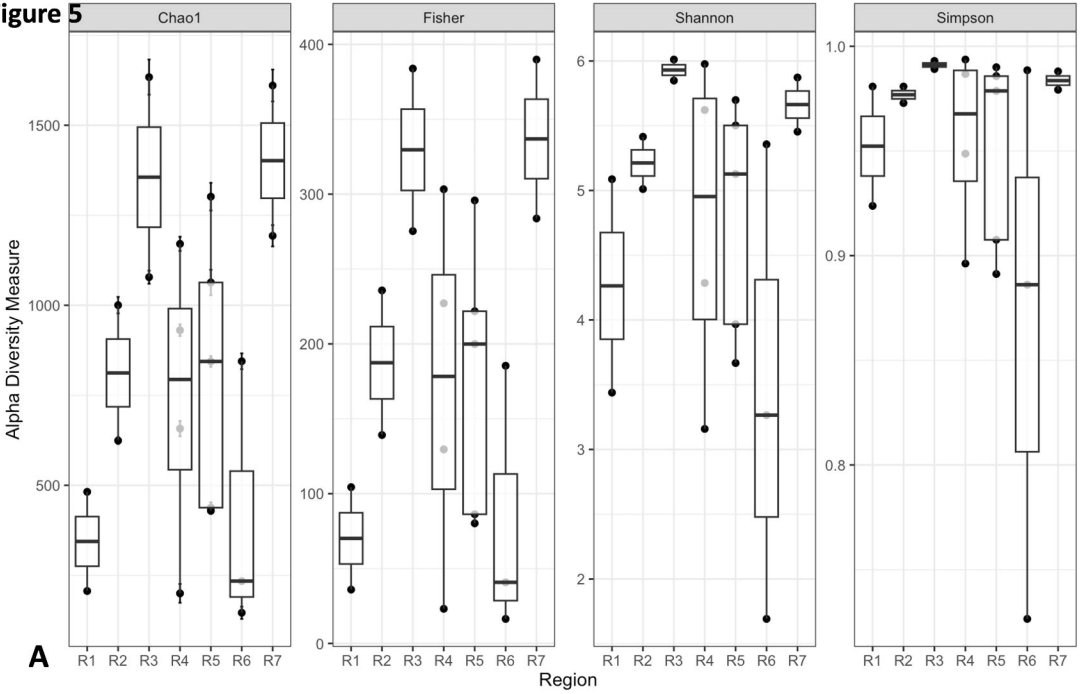
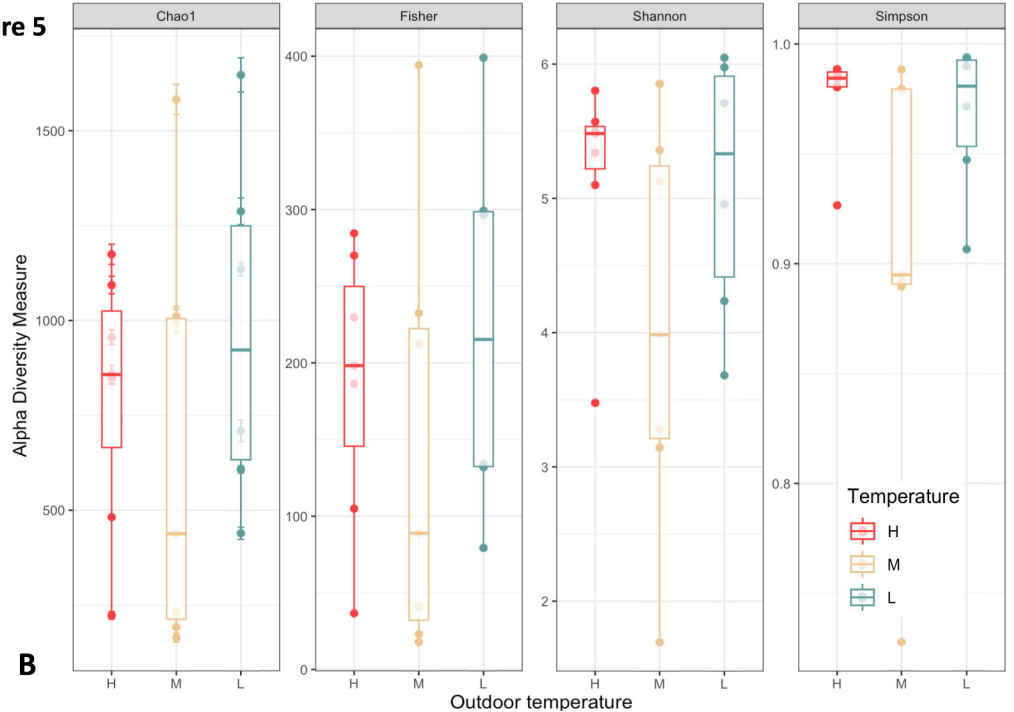
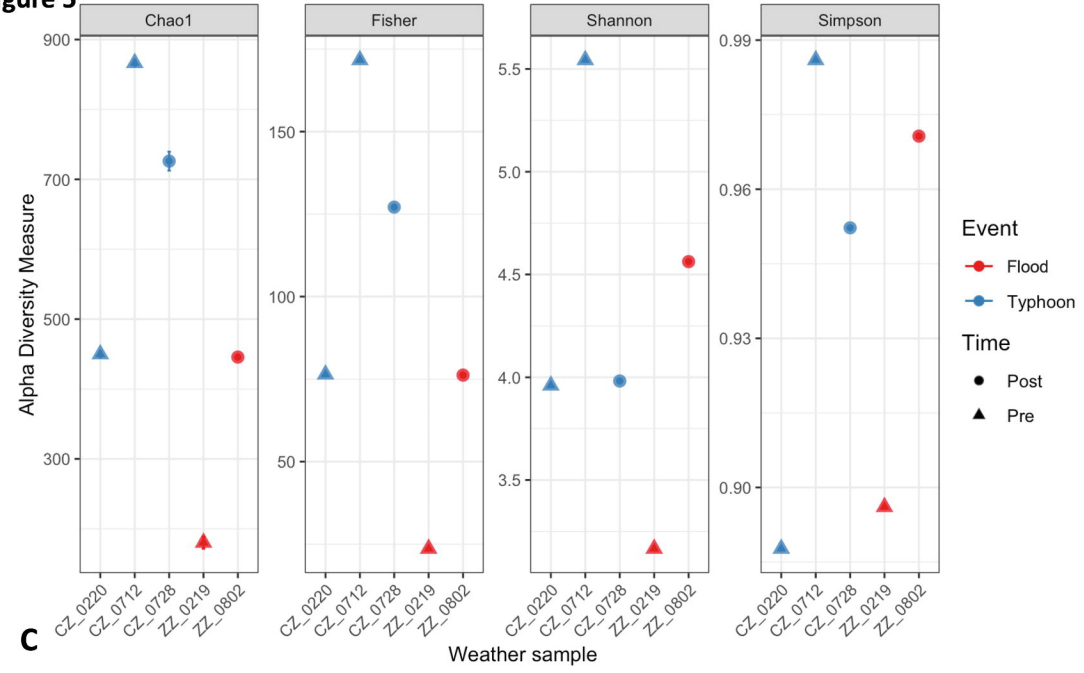

Chao1
Fisher
Shannon
Simpson
Alpha Diversity Measure
Region
R1
R2
R3
R4
R5
R6
R7
A

Figure 5


Chao1
Fisher
Shannon
Simpson
Alpha Diversity Measure
1500
1000
500
400
300
200
100
0
6
5
4
3
2
1.0
0.9
0.8
H
M
L
Outdoor temperature
Temperature
H
M
L

B

Figure 5


Chao1
Fisher
Shannon
Simpson
Alpha Diversity Measure
900
700
500
300
150
100
50
5.5
5.0
4.5
4.0
3.5
0.99
0.96
0.93
0.90
CZ_0220
CZ_0712
CZ_0728
ZZ_0219
ZZ_0802
Weather sample
Event
Flood
Typhoon
Time
Post
Pre

C

Figure 6

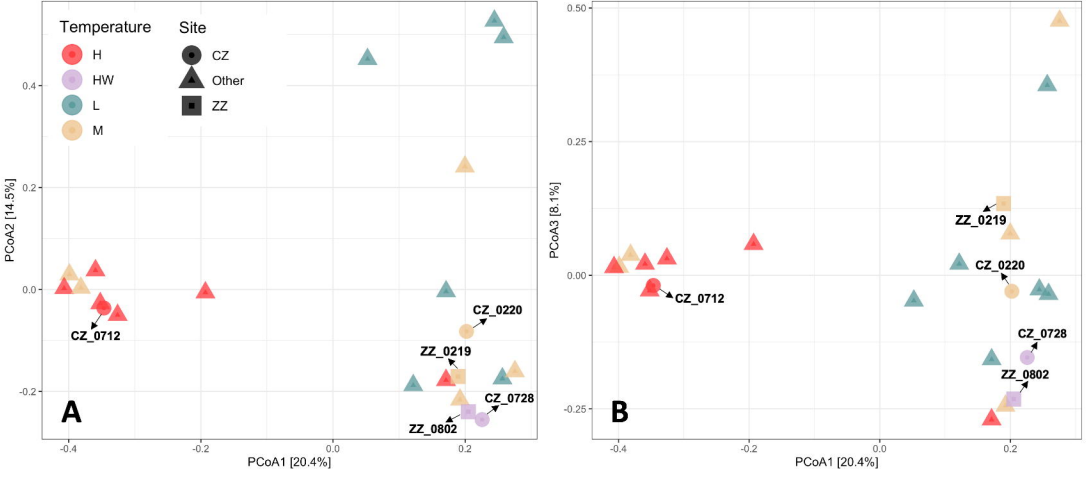

Temperature
H
HW
L
M
Site
CZ
Other
ZZ
A
PCoA2 [14.5%]
PCoA1 [20.4%]
CZ_0712
CZ_0220
ZZ_0219
CZ_0728
ZZ_0802
B
PCoA3 [8.1%]
PCoA1 [20.4%]
ZZ_0219
CZ_0220
CZ_0712
CZ_0728
ZZ_0802

Figure 7

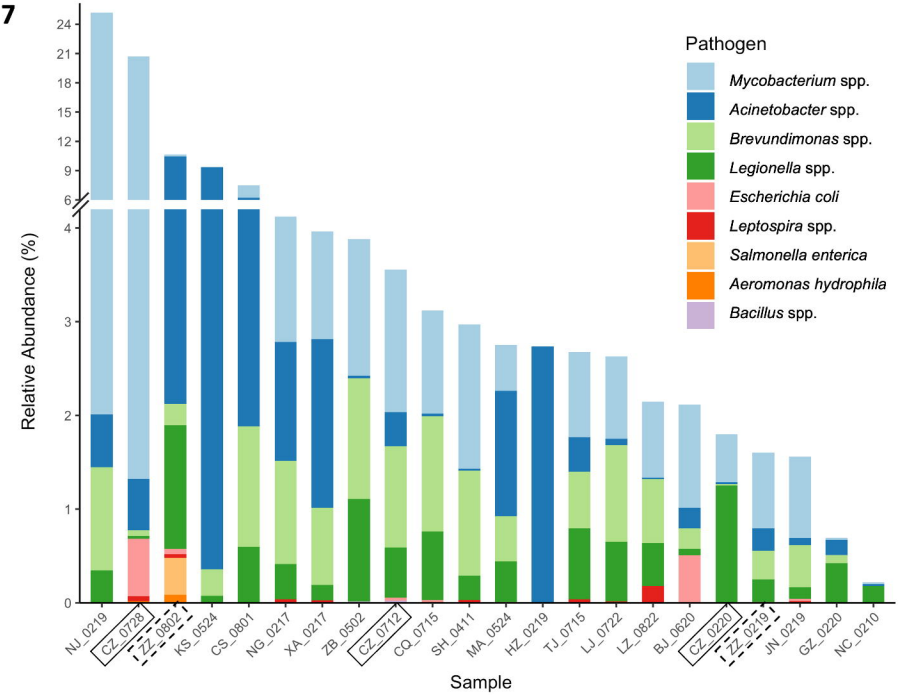
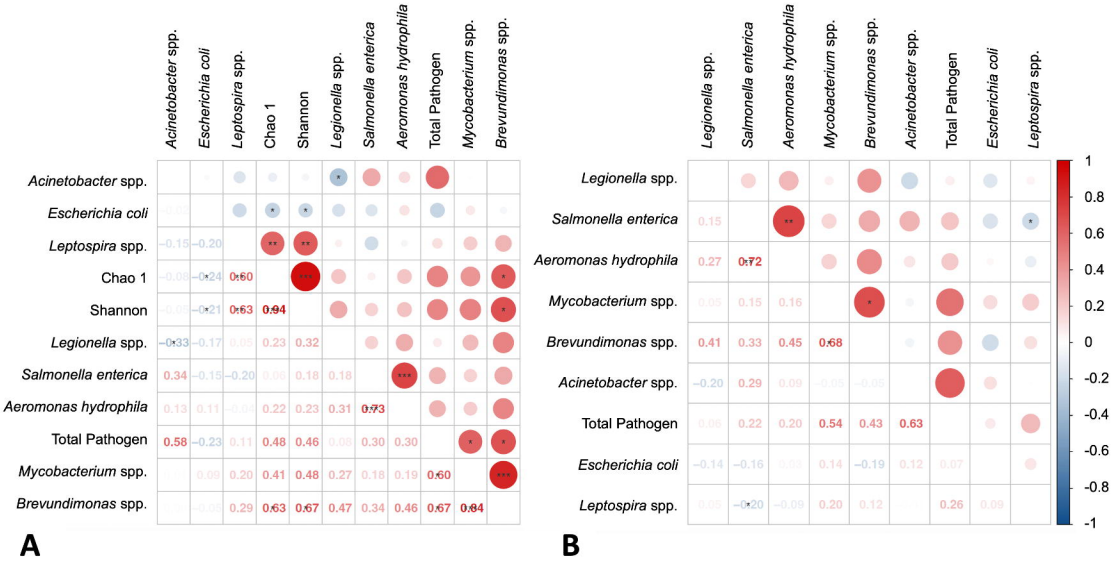

Pathogen
Mycobacterium spp.
Acinetobacter spp.
Brevundimonas spp.
Legionella spp.
Escherichia coli
Leptospira spp.
Salmonella enterica
Aeromonas hydrophila
Bacillus spp.
Relative Abundance (%)
24
21
18
15
12
9
6
4
3
2
1
0
NJ_0219
CZ_0728
ZZ_0802
KS_0524
CS_0801
NG_0217
XA_0217
ZB_0502
CZ_0712
CQ_0715
SH_0411
MA_0524
HZ_0219
TJ_0715
LJ_0722
LZ_0822
BJ_0620
CZ_0220
ZZ_0219
JN_0219
GZ_0220
NC_0210
Sample

Figure 8


Acinetobacter spp.
Escherichia coli
Leptospira spp.
Chao 1
Shannon
Legionella spp.
Salmonella enterica
Aeromonas hydrophila
Total Pathogen
Mycobacterium spp.
Brevundimonas spp.
A
Legionella spp.
Salmonella enterica
Aeromonas hydrophila
Mycobacterium spp.
Brevundimonas spp.
Acinetobacter spp.
Total Pathogen
Escherichia coli
Leptospira spp.
B
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1