data

MDPI

Data Descriptor

Physico-Chemical Quality and Physiological Profiles of Microbial Communities in Freshwater Systems of Mega Manila, Philippines

Marie Christine M. Obusan, Arizaldo E. Castro *, Ren Mark D. Villanueva, Margareth Del E. Isagan, Jamaica Ann A. Caras and Jessica F. Simbahan

Microbial Ecology of Terrestrial and Aquatic Systems Laboratory, Institute of Biology, College of Science, University of the Philippines Diliman, Quezon City 1101, Philippines; mmobusan@up.edu.ph (M.C.M.O.); rdvillanueva7@up.edu.ph (R.M.D.V.)

* Correspondence: aecastro@up.edu.ph

Abstract: Studying the quality of freshwater systems and drinking water in highly urbanized megalopolises around the world remains a challenge. This article reports data on the quality of select freshwater systems in Mega Manila, Philippines. Water samples collected between 2020 and 2021 were analyzed for physico-chemical parameters and microbial community metabolic fingerprints, i.e., carbon substrate utilization patterns (CSUPs). The detection of arsenic, lead, cadmium, mercury, polyaromatic hydrocarbons (PAHs), and organochlorine pesticides (OCPs) was carried out using standard chromatography- and spectroscopy-based protocols. Physiological profiles were determined using the Biolog EcoPlate™ system. Eight samples were free of heavy metals, and none contained PAHs or OCPs. Fourteen samples had high microbial activity, as indicated by average well color development (AWCD) and community metabolic diversity (CMD) values. Community-level physiological profiling (CLPP) revealed that (1) samples clustered as groups according to shared CSUPs, and (2) microbial communities in non-drinking samples actively utilized all six substrate classes compared to drinking samples. The data reported here can provide a baseline or a comparator for prospective quality assessments of drinking water and freshwater sources in the region. Metabolic fingerprinting using CSUPs is a simple and cheap phenotypic analysis of microbial communities and their physiological activity in aquatic environments.

Keywords: drinking water; heavy metals; emerging pollutants; microbial communities; water quality

Citation: Obusan, M.C.M.; Castro, A.E.; Villanueva, R.M.D.; Isagan, M.D.E.; Caras, J.A.A.; Simbahan, J.F. Physico-Chemical Quality and Physiological Profiles of Microbial Communities in Freshwater Systems of Mega Manila, Philippines. *Data* **2023**, *8*, 103. https://doi.org/10.3390/data8060103

Academic Editor: Jamal Jokar Arsanjani

Received: 21 February 2023
Revised: 10 April 2023
Accepted: 29 May 2023
Published: 4 June 2023

1. Summary

Twenty-nine (29) samples from water filtering stations, tributaries of Laguna Lake, water supply reservoirs, water treatment plants, and groundwater sources in Mega Manila, Philippines, were analyzed for their physico-chemical parameters and microbial community physiological profiles. Water samples were collected in triplicate. Samples were stored according to the requirements of analytical tests and immediately processed in the laboratory. Samples A–D were collected during the dry season of 2020 (December), while samples E–G were collected during the dry season of 2021 (May). Samples H-T were collected during the wet season of 2021 (June–July), while samples U–AC were collected in the latter part of 2021, transitioning to the dry season of the following year (November). All water samples were analyzed for temperature, pH, oxidation-reduction potential, absolute conductivity, resistivity, total dissolved solids, salinity, pressure, dissolved oxygen, heavy metal content (arsenic, lead, cadmium, and mercury), and organic content (polyaromatic hydrocarbons and organochlorine pesticides). Data on the physical parameters of water samples were taken in situ using a multiparameter meter. The physiological profiles of microbial communities in water samples were determined using the Biolog EcoPlate™ system. These data are relevant because:

(1) the physico-chemical parameters and physiological profiles of microbial communities in freshwater systems are useful data when assessing water quality, which can inform regulatory and monitoring policies and practices at the national and local levels.
(2) raw and analyzed data can be shared with key agencies in the Philippines, such as the Department of Health (DOH), the Department of Environment and Natural Resources (DENR), and the Laguna Lake Development Authority (LLDA). These government agencies can use our reported data as a reference or as baseline information when evaluating the outcomes of their respective national programs that use water quality as a success metric. For example, LLDA can use our reported quality data to describe whether there has been an improvement in the quality of water from Laguna Lake tributaries, e.g., when comparisons are to be made between specific years.
(3) the raw data reported in this article can be used to compute a single aggregate value, the water quality index (WQI). The WQI is a summary metric that indicates the overall quality of water in a system. The suitability of a body of water as a source of drinking water can be assessed using the weighted arithmetic WQI method, which assigns a relative weight to each parameter considered for the computation of the overall WQI [1].
(4) metabolic fingerprints, which can be determined using CLPP metrics such as standardized absorbance (SA) and substrate average well color development (SAWCD), can serve as in vitro measures of the biochemical activity of aquatic microbes. Phenotype-based metabolic fingerprints of microbial communities in an environmental sample are useful in informing strategies for implementing additional techniques and protocols, i.e., microcosm experiments [2], taxonomic profiling of microbial communities [3–5], amplification of target metabolic genes/functional profiling of microbial communities [6], and isolation of potential bioactive species [7].

2. Data Description

2.1. Study Area

Mega Manila (Figure 1) includes Metro Manila and the surrounding sub-urban provinces of Region III and Region IV-A [8]. Metro Manila is one of the three metropolitan areas in the Philippines and is composed of 16 highly urbanized cities and a municipality. The region covers an area of 239.22 sq mi and is home to a population of 13.5 million as of 2020 [9]. Sampling sites were coded (A–AC) following a non-disclosure agreement with the collaborators involved in the research program.

Figure 1. Map of Mega Manila, Philippines (United States Geological Survey).

2.2. Data on Physico-Chemical Parameters of Water Samples

The physical parameters of water samples are shown in Table 1, wherein the second column contains data on the sample type, e.g., terminal circuit (TC) samples from water filtering stations; Laguna Lake Tributary (LLT) samples from some of the major river tributaries of the lake; before treatment plant (BTP) samples from earth dams, aqueducts, and pre-treated water within treatment plants; after treatment plant (ATP) samples from water pump stations/reservoirs and post-treated water within treatment plants; and deep well (DW) samples from groundwater stored in select deep wells. Columns 3–11 contain the recorded values for the nine physical parameters. The response variables were found to have a multivariate non-normal distribution (Henze-Zirkler p value < 0.05); hence, a non-parametric comparison of samples was performed. The p values ($p = 0$) indicate statistical differences across sample types. Post hoc analysis using Dunn's test of multiple comparisons using rank sums was performed, and sample types with significant differences (adjusted $p < 0.05$) are identified in Table 1.

Table 1. Data on physical parameters of water samples.

Sample Code	Sample Type	Temp. [°C]	pH	ORP [mV]	EC [µS/cm]	R [KOhm-cm]	TDS [ppt]	Sal. [psu]	Press. [psi]	DO [ppm]
A	TC	26.51	7.05	111.00 ^{C}	102.00 ^{A}	0.01	49.00	0.05 ^{A}	14.54 ^{A}	4.73
B	TC	28.94	6.81	122.00 ^{C}	9.00 ^{A}	0.13	4.00	0.00 ^{A}	14.54 ^{A}	4.39
C	TC	26.53	7.28	78.90 ^{C}	133.00 ^{A}	0.01	64.00	0.06 ^{A}	14.54 ^{A}	4.93
D	TC	26.97	7.09	133.20 ^{C}	31.00 ^{A}	0.03	15.00	0.01 ^{A}	14.54 ^{A}	4.68
E	LLT	29.61	6.67	−243.97 ^{A,C}	658.33 ^{A}	334.08	0.30	0.29 ^{A}	14.65 ^{A}	0.14
F	LLT	29.50	7.19	−100.43 ^{A,C}	665.00 ^{A}	0.00	306.67	0.29 ^{A}	14.68 ^{A}	0.85
G	LLT	32.10	7.09	−171.67 ^{A,C}	1181.00 ^{A}	0.00	520.00	0.51 ^{A}	14.63 ^{A}	1.53
H	BTP	26.03	7.61	30.07	147.33	6.91	0.07	0.07	14.59	3.98
I	BTP	26.22	7.48	16.97	147.33	6.93	0.07	0.07	14.59	5.67
J	ATP	26.21	7.46	622.70 ^{A,B}	147.00	6.96	0.07	0.07	14.59	4.25
K	ATP	26.43	7.48	636.83 ^{A,B}	147.00	6.98	0.07	0.07	14.62	7.14
L	BTP	31.08	8.02	5.07	781.33	1.43	0.35	0.34	14.71	0.24
M	BTP	31.29	7.70	2.93	790.00	1.42	0.35	0.34	14.73	0.25
N	ATP	32.02	7.09	524.60 ^{A,B}	810.33	1.40	0.36	0.34	14.72	0.27
O	ATP	26.99	8.05	466.77 ^{A,B}	147.67	7.03	0.07	0.07	14.69	0.23
P	BTP	30.61	7.01	579.80	785.00	1.41	0.35	0.34	14.65	0.38
Q	ATP	27.22	7.36	595.37 ^{A,B}	151.33	0.01	72.67	0.07	14.60	0.25
R	LLT	30.34	7.25	−222.13 ^{A,C}	665.00 ^{A}	1.66	0.30	0.29 ^{A}	14.69 ^{A}	0.41
S	LLT	29.06	6.91	−16.47 ^{A,C}	307.33 ^{A}	111.57	0.14	0.13 ^{A}	14.72 ^{A}	0.00
T	LLT	30.96	7.45	23.10 ^{A,C}	681.67 ^{A}	1.64	0.31	0.29 ^{A}	14.71 ^{A}	0.01
U	DW	29.36	7.04	−8.63 ^{B}	568.33	1.91	0.26	0.25	14.58	0.00
V	DW	29.22	8.65	−70.30 ^{B}	551.00	1.96	0.26	0.24	14.64	0.00
W	DW	29.46	7.04	−77.90 ^{B}	535.00	2.03	0.25	0.23	14.62	0.00
X	DW	28.22	8.62	−87.17 ^{B}	564.33	1.90	0.27	0.25	14.66	0.00
Y	DW	28.04	7.53	−91.90 ^{B}	465.00	2.28	0.22	0.21	14.63	0.00
Z	DW	28.60	7.91	−59.77 ^{B}	402.67	2.66	0.19	0.18	14.69	0.00
AA	BTP	26.12	8.05	−16.13	147.67	6.93	0.07	0.07	14.45	0.00
AB	BTP	23.84	8.57	68.50	139.33	7.03	0.07	0.07	14.59	0.00
AC	BTP	24.72	7.61	24.37	138.33	7.17	0.07	0.06	14.60	0.00

TC—terminal circuit; LLT—Laguna Lake tributary; BTP—before treatment plant; ATP—after treatment plant; DW—deep well; Temp.—temperature; ORP—oxidation reduction potential; EC—absolute EC resolution; R—resistivity; TDS—total dissolved solids; Sal.—salinity; Press.—pressure; DO—dissolved oxygen; For ORP, significant differences (SD) were detected between ATP and LLT samples types, ATP and DW sample types, and LLT and TC sample types (same superscript letters indicate SD per response variable); For EC, Sal., and Press., SD were detected between LLT and TC sample types (same superscript letters indicate SD per response variable).

In terms of the oxidation reduction potential, significant differences were found between three groups of sample types: ATP vs. LLT, ATP vs. DW, and LLT vs. TC. In terms of the absolute EC resolution, salinity, and pressure, significant differences were found between the LLT and TC sample types.

The results of the chemical analyses are shown in Table 2, with the concentrations of arsenic, lead, cadmium, and mercury shown in columns 2–5. Arsenic was detected in one sample, while cadmium was detected in three samples. Lead was detected in 12 samples, while mercury was detected in 19 samples. Eight water samples were found to be free of heavy metals. None of the samples were found to contain PAHs or OCPs.

Table 2. Data on chemical parameters of water samples.

Sample Code	Sample Type	Arsenic [mg/L]	Lead [mg/L]	Cadmium [mg/L]	Mercury [mg/L]	PAHs [ug/L]	OCPs [ug/L]
A	TC	<MDL	<MDL	<MDL	<MDL	ND	ND
B	TC	<MDL	<MDL	<MDL	<MDL	ND	ND
C	TC	<MDL	<MDL	<MDL	<MDL	ND	ND
D	TC	<MDL	<MDL	<MDL	<MDL	ND	ND
E	LLT	<MDL	0.734	0.014	0.00018	<MDL	<MDL
F	LLT	0.008	0.22	0.012	0.0017	<MDL	<MDL
G	LLT	<MDL	1.686	0.006	0.002	<MDL	<MDL
H	BTP	<MDL	0.074	<MDL	0.0012	<MDL	<MDL
I	BTP	<MDL	0.068	<MDL	0.0013	<MDL	<MDL
J	ATP	<MDL	0.062	<MDL	0.0011	<MDL	<MDL
K	ATP	<MDL	0.054	<MDL	0.0012	<MDL	<MDL
L	BTP	<MDL	0.064	<MDL	0.0012	<MDL	<MDL
M	BTP	<MDL	0.026	<MDL	0.0012	<MDL	<MDL
N	ATP	<MDL	0.024	<MDL	0.0014	<MDL	<MDL
O	ATP	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
P	BTP	<MDL	0.001	<MDL	<MDL	<MDL	<MDL
Q	ATP	<MDL	0.001	<MDL	<MDL	<MDL	<MDL
R	LLT	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
S	LLT	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
T	LLT	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
U	DW	<MDL	<MDL	<MDL	0.0002	<MDL	<MDL
V	DW	<MDL	<MDL	<MDL	0.0003	<MDL	<MDL
W	DW	<MDL	<MDL	<MDL	0.0002	<MDL	<MDL
X	DW	<MDL	<MDL	<MDL	0.0004	<MDL	<MDL
Y	DW	<MDL	<MDL	<MDL	0.0007	<MDL	<MDL
Z	DW	<MDL	<MDL	<MDL	0.003	<MDL	<MDL
AA	BTP	<MDL	<MDL	<MDL	0.0005	<MDL	<MDL
AB	BTP	<MDL	<MDL	<MDL	0.0002	<MDL	<MDL
AC	BTP	<MDL	<MDL	<MDL	0.0004	<MDL	<MDL

Notes on detection limits: 1. Samples A–D (terminal circuit water): 0.01 mg/L for arsenic, lead, and mercury; 0.003 mg/L for cadmium; 0.001 ppm or 0.001 mg/L for PAHs and OCPs. 2. Samples E–AC (non-terminal circuit water): 0.004 mg/L for arsenic; 0.001 mg/L for lead and cadmium; 0.0001 mg/L for mercury; 4 µg/L or 0.004 mg/L or 0.004 ppm for benzo(b)fluoranthene and benzo(k)fluoranthene; 2 µg/L or 0.002 mg/L or 0.002 ppm for other analytes; 0.01 ug/L or 0.00001 mg/L or 0.00001 ppm for OCPs. 3. ND—not detected; MDL—minimum detection limit; Cells highlighted in green indicate heavy metal concentrations that are below the MDL

2.3. Microbial Community Physiological Profiles

Data on microbial community physiological profiles included overall average well color development (AWCD) and community metabolic diversity (CMD), which are shown in Table 3. At the end of the three-day incubation, all LLT sample types (n = 6), five BTP sample types, and three DW sample types had AWCD values that were greater than or equal to 0.5. A total of 13 samples had CMD values that were greater than or equal to 15 utilized substrates (six LLT sample types, five BTP sample types, and two DW sample types). Since AWCD and CMD are multivariate non-normal (Henze-Zirkler p value < 0.05), a non-parametric comparison of multivariate samples was performed. The p values ($p < 0.05$) indicate statistical differences across sample types. A post hoc analysis using Dunn's test of multiple comparisons using rank sums was performed, and sample types with significant differences (adjusted $p < 0.05$) are identified in Table 3.

Table 3. AWCD and CMD values of microbial communities in water samples at 72 h post incubation.

Sample Code	Sample Type	AWCD	CMD
A	TC	0.01483870968 (0.01296822339) ^{B}	0 ^{B}
B	TC	0.1164086022 (0.05841774577) ^{B}	5 ^{B}
C	TC	0.0116344086 (0.005071941683) ^{B}	0 ^{B}
D	TC	0.0008924731183 (0.001545808785) ^{B}	0 ^{B}
E	LLT	1.82394623700 (0.0859611085) ^{A,B}	31 ^{A,B}
F	LLT	1.734043011 (0.06011458304) ^{A,B}	30 ^{A,B}
G	LLT	2.037892473 (0.08508079244) ^{A,B}	31 ^{A,B}
H	BTP	1.119387097 (0.5999576216)	25
I	BTP	1.143666667 (0.2180662634)	24
J	ATP	0.008462365591 (0.003574148907) ^{A}	0 ^{A}
K	ATP	0.01991397849 (0.01713581463) ^{A}	0 ^{A}
L	BTP	1.797129032 (0.1438013863)	30
M	BTP	1.67955914 (0.2894580503)	29
N	ATP	0.008516129032 (0.006719673076) ^{A}	0 ^{A}
O	ATP	0.005677419355 (0.005017451335) ^{A}	0 ^{A}
P	BTP	0.02705376344 (0.02855444319)	0
Q	ATP	0.01431182796 (0.009223794803) ^{A}	0 ^{A}
R	LLT	1.879408602 (0.01568074879) ^{A,B}	31 ^{A,B}
S	LLT	1.719462366 (0.1569285392) ^{A,B}	30 ^{A,B}
T	LLT	1.553505376 (0.344758706) ^{A,B}	30 ^{A,B}
U	DW	0.6966236559 (0.2111123366)	13
V	DW	0.7065591398 (0.6286307092)	14
W	DW	0.3399784946 (0.4529326825)	9
X	DW	0.4320322581 (0.5781944527)	10
Y	DW	1.102913978 (0.4458875064)	19
Z	DW	0.4518602151 (0.06537119126)	16
AA	BTP	0.2282043011 (0.2012599862)	9
AB	BTP	0.8999032258 (0.1105057384)	18
AC	BTP	0.2719892473 (0.1583313756)	10

AWCD—average well color development, reported values are expressed as mean (SD); CMD—community metabolic diversity; Reported AWCD and CMD are mean values of three replicates per sample; For AWCD and CMD, significant differences (SD) were detected between ATP and LLT sample types as well as LLT and TC sample types (same superscript letters indicate SD per response variable).

For both AWCD and CMD, significant differences were found between two groups of sample types, i.e., ATP vs. LLT and LLT vs. TC.

In the heatmap (Figure 2), four clusters are evident. The first cluster (the leftmost portion of the heatmap) comprises one terminal circuit (TC) sample (c), four after treatment plant (ATP) samples (j, k, n, and q), and two deep well (DW) samples (w and x). Laguna Lake Tributary (LLT) samples (e, f, g, r, s, and t), together with four before treatment plant (BTP) samples (l, m, ab, and ac), and two deep well (DW) samples (u and v), clustered as one group. The third cluster is composed of one terminal circuit (TC) sample (b), two before treatment plant (BTP) samples (h and i), and two deep well (DW) samples (y and z). The cluster at the rightmost portion of the heatmap is made up of two before treatment plant (BTP) samples (p and aa), one after treatment plant (ATP) sample (o), and two terminal circuit (TC) samples (a and d).

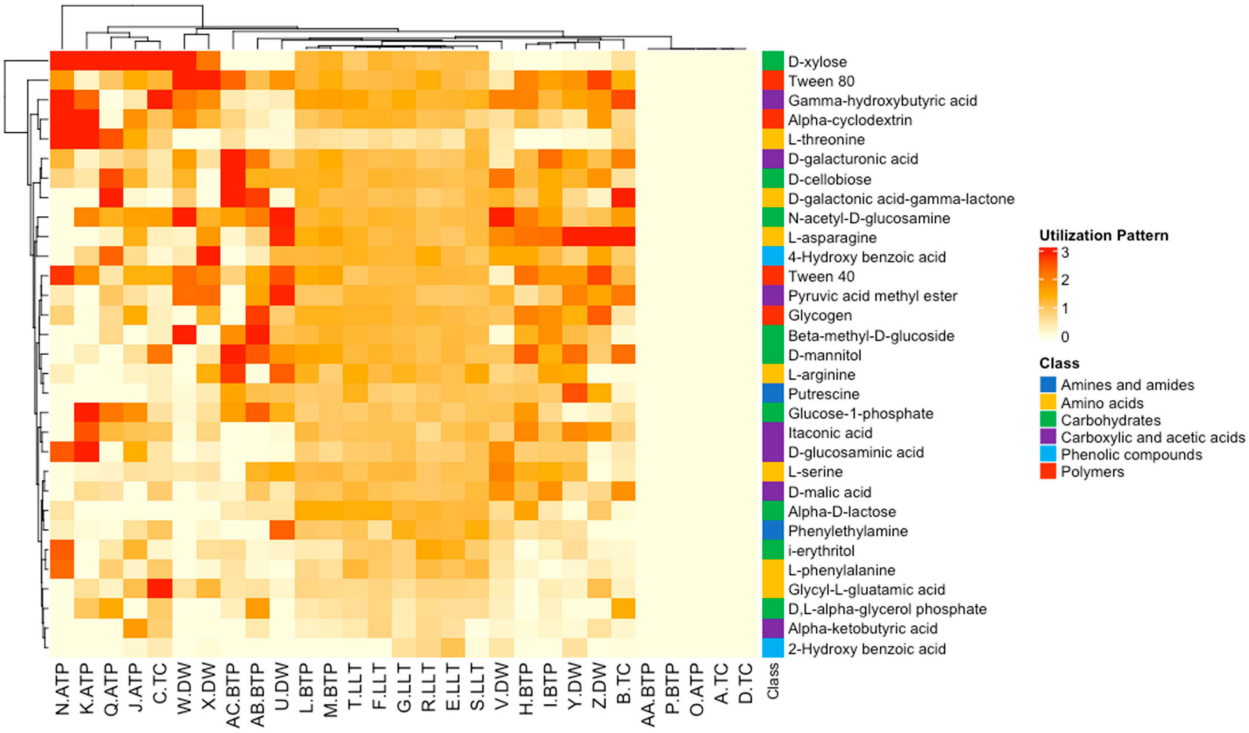

Figure 2. Metabolic fingerprints of microbial communities in all water samples for 31 carbon-containing substrates.

In general, samples in the first cluster demonstrated utilization patterns ranging from very low to very high for most of the carbon-containing substrates. On the other hand, samples in the second cluster showed moderate utilization of the substrates, particularly the LLT samples. In the third cluster, moderately high to high utilization can be seen for 27 of the 31 substrates (D-xylose to L-phenylalanine). In the last cluster, five samples showed no utilization for all substrates. Considering the sample type, significant differences were detected in 7 out of the 31 tested substrates. For pyruvic acid methyl ester, DW samples differed when compared with ATP, BTP, and TC samples. For alpha-D-lactose, LLT samples differed when compared to ATP and TC samples. For D-xylose, BTP and ATP samples differed significantly. For i-erythritol, LLT samples differed when compared to BTP and TC samples. For L-asparagine, ATP and DW samples differed significantly. For L–phenylalanine, LLT and BTP samples differed significantly. Lastly, for putrescine, ATP samples differed when compared with DW and LLT samples.

Overall, there was active utilization of all substrate classes by the microbial communities in the samples that were grouped in the first cluster (Figure 3). This cluster includes all LLT samples (e, f, g, r, s, and t) and two BTP samples (l and m). It is worth noting that the moderately high to high utilization rates obtained by the LLT samples may be attributable to the diversity of microbial communities in the samples, as these are untreated freshwater from some of the major river tributaries of the Laguna Lake. The second cluster, composed of all DW samples (u, v, w, x, y, and z) and three BTP samples (h, i, and ab), had low to moderate utilization of all substrate classes but mostly made use of polymers, carbohydrates, carboxylic and acetic acids, as well as amino acids. The third cluster of samples showed utilization rates ranging from none to low for the same substrate classes compared to the samples in the first and second clusters. Within the third cluster (one BTP sample (p), three TC samples (a, c, and d), and five ATP samples (j, k, n, o, and q)), nine samples showed no utilization of any of the substrate classes, while the remaining three samples (one TC sample (b) and two BTP samples (ac and aa)) showed very low to low utilization of all substrate classes except phenolic compounds. Taking into account the sample type, significant differences were detected in all substrate classes. For polymers,

LLT samples significantly differed compared to ATP samples ($p = 0.0010$) and TC samples ($p = 0.0012$). For carbohydrates, LLT samples significantly differed from ATP samples ($p = 0.0002$) and TC samples ($p = 0.0014$). Another statistically different group pairing for the carbohydrate substrate class was the BTP and ATP samples ($p = 0.0147$). For carboxylic and acetic acids, LLT samples significantly differed from ATP samples ($p = 0.0002$) and TC samples ($p = 0.0023$). For amino acids, LLT samples significantly differed from ATP samples ($p = 0.0001$) and TC samples ($p = 0.0016$). For amines and amides, LLT samples significantly differed from ATP samples ($p = 0.0002$) and TC samples ($p = 0.0031$). Lastly, for phenolic compounds, LLT samples significantly differed from ATP samples ($p = 0.0010$) and TC ($p = 0.0063$) samples.

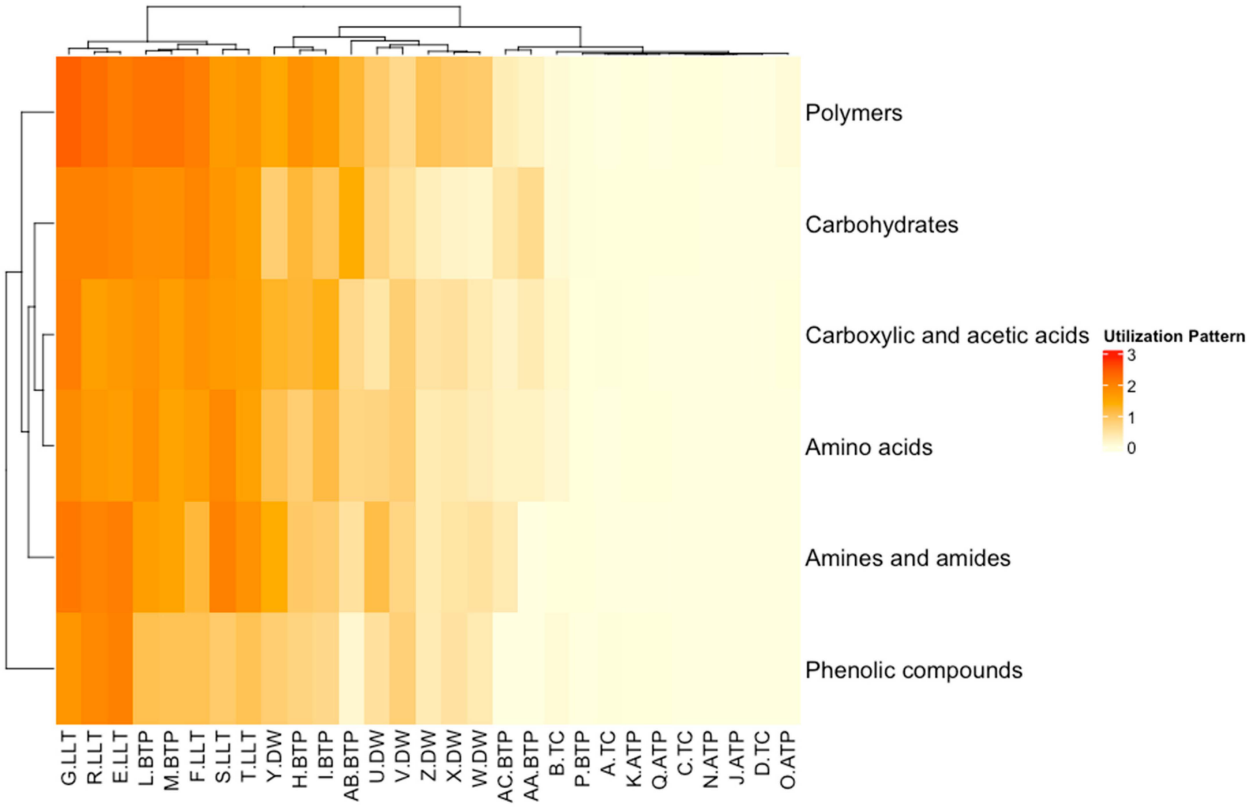

Figure 3. Metabolic fingerprints of microbial communities in all water samples for 6 substrate classes.

The clustering of the samples based on shared CSUPs and the detected significant differences in the standardized absorbances for specific substrates suggest that metabolic fingerprint data are useful for describing the functional potential of microbial communities in water samples. The functional activity of aquatic microbial communities has been reported to be associated with water resource type and pollution level [10,11]. Therefore, metabolic fingerprints, expressed as substrate utilization (phenotype-based outputs reflecting microbial community activity), are informative data that must be considered as potential biological indicators of water quality. Microbial community activity in aquatic environments is influenced by water quality, as microbes play an important role in the biogeochemical cycling of nutrients and organic matter in these environments. Different physical, chemical, and biological factors affect how microbial communities interact with one another and with water as their matrix. The usual quality parameters that impact microbial community activity include pH, temperature, dissolved oxygen, and concentrations of nutrients such as nitrogen and phosphorus. Thus, investigating the metabolic activity of microbial communities can aid in assessing the quality of water in freshwater resources. Our findings support the utility of metabolic fingerprinting as a relatively easy and cheap method for analyzing the physiological responses of microbial communities

in water. Microbial community level physiological profiles may be water resource-type-specific and, when combined with data on community functional richness (i.e., CMD) and overall metabolic rates (i.e., AWCD), are demonstrated to be concordant. Additionally, metabolic fingerprint data can serve as a phenotype-based supplement to support the results of DNA-based taxonomic and functional profiling.

3. Methods

Water samples from different source types were collected in triplicate. Five general source types were sampled, i.e., (1) TC for drinking water refilling stations; (2) LLT for Laguna Lake river tributaries; (3) BTP for earth dams, aqueducts, and pre-treatment water; (4) ATP for post-treatment water, water pump stations, and post-treatment water reservoirs; and (5) DW for groundwater sources (deep wells). All necessary permits from governing agencies were secured prior to sampling. Sampling was conducted from December 2020 to November 2021. All samples were analyzed for physical parameters, chemical composition, and physiological profiles of microbial communities. A multiparameter meter (HI-98194 multiparameter meter, Hanna Instruments, Woonsocket, RI, USA,) was used according to the manufacturer's instruction [12] to carry out an in situ measurement of physical parameters reported in Table 1. The distribution of the physical parameter data was assessed using Henze-Zirkler's test for multivariate normality. The test for multivariate normality was performed in RStudio (version 2022.07.2) using the MVN package. A non-parametric comparison of multivariate samples was conducted to test for statistical differences across all sample types using the physical parameters reported in Table 1. The non-parametric comparison was performed in RStudio (version 2022.07.2) using the nonpartest function of the npmv package with the permreps argument set to 1000. In order to identify which sample types significantly differ, a post hoc analysis was conducted using Dunn's test. Post hoc analysis was performed in RStudio (version 2022.07.2) using the dunn.test package with p values adjusted following the Benjamini-Hochberg correction in order to minimize false discovery rates.

Arsenic was tested in the samples via the 3114 B: Manual Hydride Generation/AAS method [13], while mercury was tested using inductively coupled plasma-optical emission spectroscopy (ICP-OES) [14]. Lead and cadmium were tested in the samples via 3111-B: Direct Air-Acetylene Flame method [15]. Maximum allowable limits set for drinking water as per the Philippine National Standards for Drinking Water of 2017 are 0.01 mg/L for arsenic and lead, 0.001 mg/L for mercury, and 0.003 mg/L for cadmium. Gas chromatography–mass spectrometry and gas chromatography–electron-capture detector methods [15,16] were used to analyze the samples for the presence of polyaromatic hydrocarbons namely acenaphthene, acenaphthylene, anthracene, benz(a)anthracene, benzo(a)pyrene, benzo(k)fluoranthene, benzo(g,h,i)perylene, benzo(b)fluoranthene, chrysene, dibenz(a,h)anthracene, fluoranthene, fluorene, indeno(1,2,3-cd), pyrene, napthalene, phenanthrene, pyrene, 1-methylnaphthalene, and 2-methylnaphthalene. The same methods were used to detect organochlorine pesticides, namely, a-BHC, g-BHC, b-BHC, heptachlor, d-BHC, aldrin, heptachlor epoxide, g-chlordane, a-chlordane, endosulfan I, 4,4′-DDE, dieldrin, endrin, 4,4′-DDD, endosulfan II, 4,4′-DDT, endrin, aldehyde, endosulfan sulfate, methoxychlor, endrin ketone, 2,6-Di-tert-butyl-p-cresol, aldrin and dieldrin (combined), chlordane, lindane, total organochlorine pesticides, and toxaphene.

Microbial community analysis was carried out using Biolog EcoPlates according to the manufacturer's instructions [17]. A total of 100 µL inoculum from the membrane-filtered water samples was transferred to each well of an EcoPlate. The plates were covered with foil and placed inside sterile plastic containers with moist paper towels. Incubation of the plates took place in a dark room at ambient temperature. Optical density (OD) at 590 nm was measured after three days. Two metrics of community metabolic activity were measured for all samples, average well color development (AWCD) and community metabolic diversity (CMD). AWCD refers to the average respiration of carbon substrates by the microbial communities, calculated as the average of the mean difference between

the optical density (OD) of the carbon substrate-containing wells and the control well. CMD refers to the number of substrates utilized by the microbial community (analogous to community functional richness), calculated by adding the frequencies of metabolized carbon substrates (wells with corrected absorbances greater than 0.25). From the absorbance data, the AWCD was calculated using the following equation:

$$\text{AWCD} = \frac{\sum_{i=1}^{N} \text{OD}_i}{\text{N}} \quad (1)$$

wherein the average of blank-corrected absorbance values (OD_i) was taken over the 31 substrates (N). Negative AWCD values that indicate very little response were coded as zeros prior to heatmap analysis [18]. The distribution of the CLPP metrics data was assessed using Henze-Zirkler's test for multivariate normality. The test for multivariate normality was performed in RStudio (version 2022.07.2) using the MVN package. A non-parametric comparison of multivariate samples was conducted to test for statistical differences across all sample types using the CLPP metrics parameters reported in Table 3. The non-parametric comparison was performed in RStudio (version 2022.07.2) using the nonpartest function of the npmv package with the permreps argument set to 1000. In order to identify which sample types significantly differ, a post hoc analysis was conducted using Dunn's test. Post hoc analysis was performed in RStudio (version 2022.07.2) using the dunn.test package, with *p* values adjusted following the Benjamini-Hochberg correction in order to minimize false discovery rates.

The substrates can be further subdivided into classes, as outlined by Sala et al. (2010) [19]. Standardized absorbances (SA) for each well with a specific carbon-containing substrate were computed using the following formula:

$$\text{SA} = \frac{\text{A}_k - \text{A}_o}{\text{AWCD}} \quad (2)$$

wherein A_k represents the absorbance reading for a specific well and A_o is the absorbance reading of the blank well. The difference between the values was then divided by the AWCD for the 72-h time point. Negative SA values indicate very little response and were coded as zeros prior to heatmap analysis. For each substrate class, the substrate average well color development (SAWCD) can be calculated using the following formula [2,20]:

$$\text{SAWCD} = \frac{\sum_{i=1}^{N} \text{OD}_i}{\text{N}} \quad (3)$$

wherein the average of blank-corrected absorbance values (OD_i) was taken over the total number of substrates (N) for each class. Negative SAWCD values indicate very little response and were coded as zeros prior to heatmap analysis. SA and SAWCD were used for heatmap analysis to determine and compare substrate utilization patterns. Substrate classification was used to visualize the primary carbon source metabolism. Heatmap analysis was conducted in RStudio (version 2022.07.2) using a hierarchical clustering method implemented through the ComplexHeatmap package. SA and SAWCD were also analyzed for significant differences, using sample type as the independent variable. The distribution of the SA and SAWCD datasets was assessed using the Henze-Zirkler's test for multivariate normality. The test for multivariate normality was performed in RStudio (version 2022.07.2) using the MVN package. A non-parametric comparison of multivariate samples was conducted to test for statistical differences across all sample types using the SA and SAWCD data reported in the Supplementary Materials. The non-parametric comparison was performed in RStudio (version 2022.07.2) using the nonpartest function of the npmv package with the permreps argument set to 1000. To identify which sample types significantly differ, a post hoc analysis was conducted using Dunn's test. Post hoc analysis was performed in RStudio (version 2022.07.2) using the dunn.test package, with *p* values adjusted following the Benjamini-Hochberg correction to minimize false discovery rates.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/data8060103/s1.

Author Contributions: Conceptualization of the study, funding acquisition, research design, data curation, data validation, writing—review and editing: M.C.M.O. and J.F.S.; Conceptualization of the study, methodology, data gathering, data curation, data validation, statistical analyses and data visualization, writing—review and editing: A.E.C.; data gathering, data curation, data validation, and writing: R.M.D.V., M.D.E.I. and J.A.A.C. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the Department of Science and Technology- Philippine Council for Health Research and Development through Project 1: Microbial communities as sentinels of environmental pollutants in Mega Manila's resources—Water assessment of teratogens and endocrine disruptors on fetal-maternal health (Water FeMaH) research program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available as the content of reported tables and as attached Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Akter, T.; Jhohura, F.T.; Akter, F.; Chowdhury, T.R.; Mistry, S.K.; Dey, D.; Barua, M.K.; Islam, M.A.; Rahman, M. Water quality index for measuring drinking water quality in rural Bangladesh: A cross-sectional study. *J. Health Popul. Nutr.* **2016**, *35*, 4. [CrossRef] [PubMed]
2. Németh, I.; Molnár, S.; Vaszita, E.; Molnár, M. The biolog ecoplate™ technique for assessing the effect of metal oxide nanoparticles on freshwater microbial communities. *Nanomaterials* **2021**, *11*, 1777. [CrossRef] [PubMed]
3. Ge, Z.; Du, H.; Gao, Y.; Qiu, W. Analysis on metabolic functions of stored rice microbial communities by Biolog Eco microplates. *Front. Microbiol.* **2018**, *9*, 1375. [CrossRef] [PubMed]
4. Tobias-Hünefeldt, S.P.; Wing, S.R.; Baltar, F.; Morales, S.E. Changes in microbial community phylogeny and metabolic activity along the water column uncouple at near sediment aphotic layers in fjords. *Sci. Rep.* **2021**, *11*, 19303. [CrossRef] [PubMed]
5. Dumontet, S.; Cavoski, I.; Ricciuti, P.; Mondelli, D.; Jarrar, M.; Pasquale, V.; Crecchio, C. Metabolic and genetic patterns of soil microbial communities in response to different amendments under Organic Farming System. *Geoderma* **2017**, *296*, 79–85. [CrossRef]
6. Koner, S.; Chen, J.-S.; Hsu, B.-M.; Tan, C.-W.; Fan, C.-W.; Chen, T.-H.; Hussain, B.; Nagarajan, V. Assessment of carbon substrate catabolism pattern and functional metabolic pathway for Microbiota of Limestone Caves. *Microorganisms* **2021**, *9*, 1789. [CrossRef] [PubMed]
7. Gałązka, A.; Marzec-Grządziel, A.; Varsadiya, M.; Niedźwiecki, J.; Gawryjołek, K.; Furtak, K.; Przybyś, M.; Grządziel, J. Biodiversity and metabolic potential of bacteria in bulk soil from the peri-root zone of black alder (Alnus glutinosa), silver birch (betula pendula) and Scots pine (pinus sylvestris). *Int. J. Mol. Sci.* **2022**, *23*, 2633. [CrossRef] [PubMed]
8. National Economic and Development Authority—Roadmap for Transport Infrastructure Development for Metro Manila and its Surrounding Areas (Region III and Region IV-A). Available online: https://neda.gov.ph/roadmap-transport-infrastructure-development-metro-manila-surrounding-areas-region-iii-region-iv/ (accessed on 8 April 2023).
9. Philippine Statistics Authority—2020 Census of Population and Housing (2020 CPH) Population Counts Declared Official by the President. Available online: https://psa.gov.ph/content/2020-census-population-and-housing-2020-cph-population-counts-declared-official-president (accessed on 23 January 2023).
10. Jałowiecki, Ł.; Chojniak, J.M.; Dorgeloh, E.; Hegedusova, B.; Ejhed, H.; Magnér, J.; Płaza, G.A. Microbial community profiles in wastewaters from onsite Wastewater Treatment Systems Technology. *PLoS ONE* **2016**, *11*, e0147725. [CrossRef] [PubMed]
11. Melita, M.; Amalfitano, S.; Preziosi, E.; Ghergo, S.; Frollini, E.; Parrone, D.; Zoppini, A. Physiological profiling and functional diversity of groundwater microbial communities in a municipal solid waste landfill area. *Water* **2019**, *11*, 2624. [CrossRef]
12. HANNA Instruments—HI98194, HI98195, HI98196 Multiparameter Meters Instruction Manual. Available online: https://hanna-worldwide.com/mwdownloads/download/link/id/27 (accessed on 23 January 2023).
13. APHA. Standard Methods for Estimation of Water and Wastewater. Available online: https://beta-static.fishersci.com/content/dam/fishersci/en_US/documents/programs/scientific/technical-documents/white-papers/apha-water-testing-standard-methods-introduction-white-paper.pdf (accessed on 23 January 2023).
14. United States Environmental Protection Agency (USEPA). SW-846 Test Method 6010D: Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES). Available online: https://www.epa.gov/hw-sw846/sw-846-test-method-6010d-inductively-coupled-plasma-optical-emission-spectrometry-icp-oes (accessed on 23 January 2023).

15. United States Environmental Protection Agency (USEPA). Method 8270D Semivolatile Organic Compounds by Gas Chromatography/Mass Spectrometry (GC/MS). Available online: https://archive.epa.gov/epa/sites/production/files/2015-12/documents/8270d.pdf (accessed on 23 January 2023).
16. United States Environmental Protection Agency (USEPA). Method 8081B Organochlorine Pesticides by Gas Chromatography Methods. Available online: https://www.epa.gov/sites/default/files/2015-12/documents/8081b.pdf (accessed on 23 January 2023).
17. Biolog—Microbial Community Analysis with EcoPlates. Available online: https://www.biolog.com/products-portfolio-overview/microbial-community-analysis-with-ecoplates/ (accessed on 23 January 2023).
18. Weber, K.P.; Legge, R.L. Community-level physiological profiling. In *Methods in Molecular Biology*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 263–281.
19. Sala, M.M.; Arrieta, J.M.; Boras, J.A.; Duarte, C.M.; Vaqué, D. The impact of ice melting on Bacterioplankton in the Arctic Ocean. *Polar Biol.* **2010**, *33*, 1683–1694. [CrossRef]
20. Chou, Y.-M.; Shen, F.-T.; Chiang, S.-C.; Chang, C.-M. Functional diversity and dominant populations of bacteria in banana plantation soils as influenced by long-term organic and conventional farming. *Appl. Soil Ecol.* **2017**, *110*, 21–33. [CrossRef]