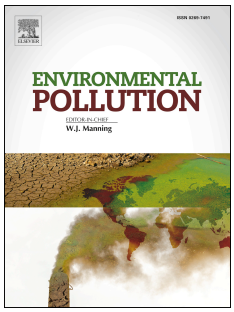
Accepted Manuscript



Could benthic biofilm analyses be used as a reliable proxy for freshwater environmental health?

Yang Pu, Wing Yui Ngan, Yuan Yao, Olivier Habimana

PII: S0269-7491(18)35245-X

DOI: https://doi.org/10.1016/j.envpol.2019.05.111

Reference: ENPO 12656

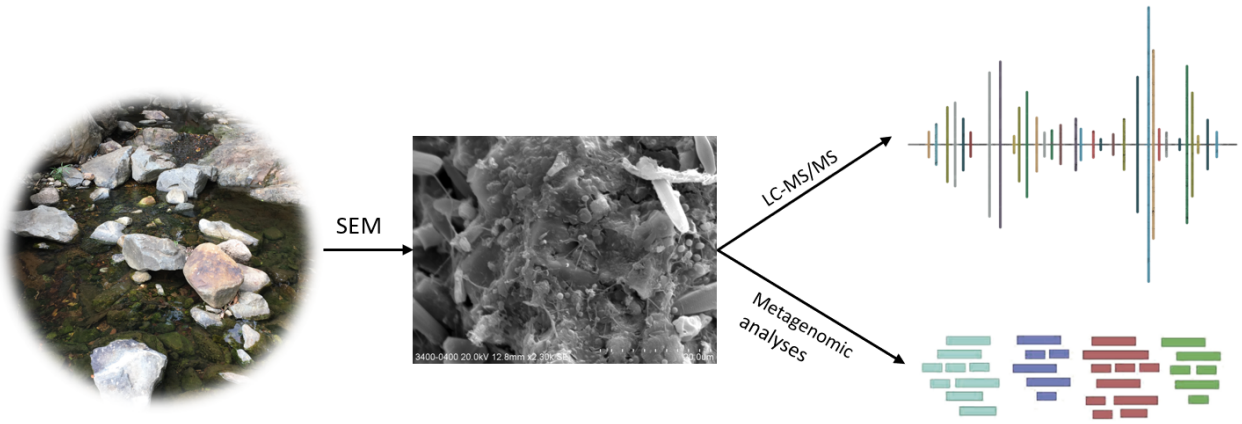
To appear in: *Environmental Pollution*

Received Date: 22 November 2018

Revised Date: 21 May 2019

Accepted Date: 21 May 2019

Please cite this article as: Pu, Y., Yui Ngan, W., Yao, Y., Habimana, O., Could benthic biofilm analyses be used as a reliable proxy for freshwater environmental health?, *Environmental Pollution* (2019), doi: https://doi.org/10.1016/j.envpol.2019.05.111.



Could benthic biofilm analyses be used as a reliable proxy for freshwater environmental health?

Yang PU[1], Wing Yui NGAN[1], Yuan YAO[1], and Olivier HABIMANA[1, 2*]

[1] School of Biological Sciences, Faculty of Science, The University of Hong Kong, Pokfulam Road, Hong Kong *Special Administration Region*, The People's Republic of China.

[2] The University of Hong Kong-Shenzhen Institute of Research and Innovation (HKU-SIRI), Shenzhen, Guangdong Province, The People's Republic of China.

[*]Corresponding Author: Dr. Olivier HABIMANA; email: ohabim@hku.hk

Abstract

The quality of freshwater undoubtedly reflects the health of our surrounding environment, society, and economy, as these are supported by various freshwater ecosystems. Monitoring efforts have therefore been considered a vital means of ensuring the ecological health of freshwater environments. Nevertheless, most aquatic environmental monitoring strategies largely focus on bulk water sampling for analysis of physicochemical and key biological indicators, which for the most part do not consider pollution events that occur at any time between sampling events. Because benthic biofilms are ubiquitous in aquatic environments, pollution released during sporadic events may be absorbed by these biofilms, which can act as repositories of pollutants. The aim of this study was to assess whether benthic biofilm monitoring could provide an efficient way of properly characterizing the extent of pollution in aquatic environments. Here, bulk water and benthic biofilms were sampled from three Hong Kong streams having various pollution profiles, and subsequently compared via high-resolution microscopy, metagenomic analysis, and analytical chemistry. The results indicated that biofilms were, indeed, reservoirs of environmental pollutants, having different profiles compared with that of the corresponding bulk water samples. Moreover, the results also suggested that biofilms sampled in polluted areas were characterized by a higher species richness. While the analytical testing of benthic biofilms still needs further development, the integration of chemical-pollutant profiles and biofilm sequencing data in future studies may provide unique perspectives for understanding and identifying pollution-related biofilm biomarkers.

Keywords: biofilm, LC-MS/MS, 16s rRNA sequencing, SEM

Capsule: Benthic biofilms can act as stable reservoirs of environmental pollutants, hence, future work is needed for identifying pollution-related freshwater biofilm biomarkers.

1. Introduction

The quality of freshwater is irrevocably linked to the health of our surrounding environment, society, and economy, as these are sustained by various freshwater ecosystems (i.e. wetlands, streams, aquifers) (Wang, et al., 2017). Maintaining these freshwater ecosystems, therefore, plays an important role in the preservation of valuable natural resources vital for all living organisms, human health, food production, and economic development (Bouwmeester, et al., 2015; Dudgeon, 2014). The level of biodiversity within an ecosystem is an important benchmark of ecological health, and as such, monitoring the health of these ecosystems requires the ability to construe and characterize ecological changes in an accurate manner using sensitive quantitative indicators (Smetanová, et al. , 2014).

The monitoring of ecological health in aquatic ecosystems is conducted by means of indicators, with physicochemical and biological indicators being the most commonly used (Xie et al., 2016). In the same manner, water quality assessments of freshwater ecosystems are traditionally carried out through a series of physicochemical measurements, which include concentrations of dissolved oxygen, pH, organic carbon, suspended solids, temperature, salinity, and nutrients (nitrogen and phosphorus) (Hounslow, 2018). These measurements may also include monitoring pollutants such as heavy metals, insecticides, or herbicides. Despite the relative ease in acquiring freshwater physicochemical data, large spatial and temporal variability in sampling may lead to improper freshwater quality assessments due to insufficient water sampling numbers (Lear et al., 2012; de Paul Obade & Moore, 2018).

Moreover, the nature of these collected data is highly specific, thereby potentially leading to biased conclusions based on isolated subsets of water quality parameters chosen for the study in question (Lear et al., 2012). For these reasons, physicochemical water measurements are often accompanied using biological indicators, which provide a comprehensive ecosystem health assessment through the monitoring of interferences measured by the presence/absence, condition, and relative abundance and community structure of groups of targeted organisms (Parmar, Rawtani, & Agrawal, 2016).

While macroorganisms, such as fish and macroinvertebrates, are commonly used biological indicators for assessing ecological changes in aquatic ecosystems (Ho & Dudgeon, 2016; Liao et al., 2018), monitoring these may in most cases present constraints with respect to ease of sampling, collection, and analyses, especially in cases where significant expertise in visual recognition of individual organisms is needed for taxon identification (Lear et al., 2012).

The monitoring of microorganisms, such as coliforms, fecal coliforms, and *Escherichia coli*, is widely used for water quality assessment and has traditionally been the preferred standard for estimating the risk of exposure to other types of waterborne pathogens (Pachepsky et al., 2016). Unfortunately, some of the standards used do not necessarily reflect the ecological health of freshwater ecosystems. The occurrence of pathogens in freshwater sources can be sporadic and is contingent on situations characterized by the presence of animals in proximity to the water source, or of environmental factors such as heavy rainfall, where sewage runoff into water sources can dramatically increase the level of pathogens (Draper et al., 2016). While collected water samples are used to determine the level of coliforms and other

indicator microorganisms, they only represent a small fraction of the total microorganisms found embedded within biofilms. Among the different life forms found within freshwater ecosystems, the surfaces of submerged objects such as plant debris and rocks are usually teeming with microorganisms in the form of stream microbial biofilms, also recognized as periphyton (Flemming et al., 2016; Piggott et al., 2015). Such biofilms are made up of complex communities that include bacteria, archaea, fungi, and other microorganisms within a protective gelatinous matrix of extracellular polymeric substances (EPS) and have been shown to play a pivotal role in the functioning of aquatic ecosystems. These biofilms are key components of freshwater ecosystems, in which critical ecological processes such as carbon cycling and nutrient cycling occur (Guasch et al., 2016; Wu, Liu, & Rene, 2018). It is now evident that these biofilm communities are not static but are characterized by changes in community structure related to anthropogenic impacts across a wide range of time scales (Hutchins & Fu, 2017). The presence and abundance of specific organisms within these biofilm communities may, therefore, be used to assess potential cumulative responses to past environmental conditions, and hence should be considered as a compelling indicator to be used in conjunction with physicochemical monitoring techniques, as well as collected data including the presence, abundance, activity, and health of other taxa (Sabater, 2017).

With the latest technological advances in molecular technologies now available, the possibility of linking the microbial diversity and pollutant content of biologically diverse biofilm communities with freshwater ecological health may help develop improved biological indicators for assessing freshwater quality and ecosystem fitness (Scott et al., 2019; Shen et al., 1991). It is now well established that in freshwater ecosystems, biofilm-associated bacterial biomass constitutes a valuable food source

for benthic consumers (Vaughn et al., 2018). In the event of environmental pollution brought about by human activities, significant changes will occur within biofilms' structure, composition, and microbial community (Gerbersdorf et al., 2011). Hence, monitoring benthic biofilm communities' fitness in freshwater environments, as a proxy for freshwater ecological health and pollution level, might help strengthen existing environmental impact assessment frameworks by providing a historical pollution profile of freshwater environments otherwise missed through separate individual water sampling methods. One recent investigation detailing the effects of surface water pollution showed that low pharmaceutical product concentrations could rapidly provoke a variety of functional shifts in stream bacterial communities, thereby affecting important ecosystem processes such as nutrient cycling (Pereda et al., 2019).

It is therefore hypothesized that benthic biofilms might be a suitable candidate for monitoring water pollution and quality, as these biofilms are capable of absorbing and effectively recording chemical and biological pollution events otherwise missed or underestimated by conventional water sampling routines. Within this context, therefore, biofilm monitoring may potentially provide a useful biological indicator for assessing water quality within freshwater environments, by allowing an understanding of *i*) what microbial communities are present within sampled benthic biofilms, and *ii*) the level or concentration of these absorbed pollutants contained within these biofilms.

To test this hypothesis, benthic biofilms and corresponding bulk water samples were sampled from three freshwater stream-sites in Hong Kong and were qualitatively and quantitatively assessed using a series of microscopy, 16S rRNA metagenomics, and LC-MS/MS analyses.

2. Materials and Methods

2.1 Chemicals and reagents

The chemicals used in this study were purchased from Sigma-Aldrich (St. Louis, USA), and all HPLC grade acetonitrile or methanol were supplied by International Laboratory USA and Anaoua Chemicals Supply (Cleveland, USA). The water utilized for analytical purposes was generated by a Millipore water purification system with a resistivity of 18.2 MΩ·cm at ambient temperature. The water was not stored for more than 24 hours to prevent microbial contamination and change in resistivity. A DNA extraction kit was purchased from ZYMO Research (CA, USA).

2.2. Sampling sites

The sampling sites selected for the sampling of rocks attached with benthic biofilms were streams at Sam Dip Tam (ST1), Ho Chung River (ST2), and Lung Fu Shan (ST3) situated in the New Territories, Kowloon, and Hong Kong Island areas, respectively, in Hong Kong. These sites were at the time of sampling heavily influenced by nearby domestic or construction activities. The temperature at the time of sampling was recorded being at a range 20□-25□. The hydrodynamic conditions at sampling stream sites were similar, in which six to nine submerged rocks of identical size were selected. The water sampling routine consisted of collecting approximately 40 mL surface water using 50 mL Falcon® tubes (Thermo Fisher Scientific, USA), as well as corresponding rock samples presenting a visible benthic community, which were

placed in jars and kept submerged in freshwater. All the samples were transported to the laboratory immediately thereafter, for further processing.

The water samples were processed for pollutant analyses, whereas the benthic biofilm samples were aliquoted for scanning electron microscopy, 16S metagenomic profiling, and pollutant analyses. Three separate sample collections were conducted, i.e. one each at ST1, ST2, and ST3, during the month of December 2017.

2.3 Sample pre-treatment

For pollutant analyses, the samples were promptly processed within the first 24 hours of sampling to prevent analyte degradation. Surface water was first acidified with formic acid to pH 3 then filtered with a 0.45 µm syringe filter (PALL Corporation, USA) before short-term storage at -80□.

Rock samples were firstly transferred to glass beakers containing fresh Milli-Q water then sonicated for 7 minutes at 28 kHz frequency (Kan-Pacific, Hong Kong). Following sonication, the rocks were scraped and then discarded, and the remaining resuspended benthic biofilm constituents were then collected following filtration using 0.45 µm membranes. These fouled membranes were used for either pollutant analyses or 16S metagenomics as further described in section 2.5.

2.4 Pollutant extraction and analyses using LC- QTOF-MS/MS

Contrarily to the previously collected and pre-treated acidified freshwater samples described above, the biofilm samples required further treatment for pollutant extraction for qualitative and quantitative analyses. Following biomass collection as

191 described in section 2.3, the biomass was resuspended in 30 mL Milli-Q water by
192 scraping the retentate side of the membrane. To facilitate cell wall damage of the cells
193 embedded within the biofilm, osmotic stress conditions were induced by adding
194 sodium chloride to the resuspended biomass to a final concentration of 1 M. The
195 samples were vortexed at maximum speed for four hours in a cold room (4□),
196 followed by rotation with a speed of 45 r/min, after which the supernatants were
197 transferred to new tubes by centrifugation.

198 Pollutants from the samples originating from bulk freshwater and biofilm treatments
199 were acquired by solid phase extraction (SPE) (HLB, Waters, USA) using methanol
200 as the eluting organic solvent, following the manufacturer's protocols. The samples
201 were then concentrated with a rotavapor (BÜCHI, Switzerland) and reconstituted
202 with 200 µL pure HPLC grade acetonitrile. Before analyses, all samples were filtered
203 with a 0.45 µm syringe filter and stored at -80□ (Núñez et al., 2017).

204 For surface water and biofilm analyses, direct injection of surface water was
205 conducted with a SCIEX X500R QTOF MS/MS (AB SCIEX, USA), which was
206 connected to an ExoionLC^{TM} HPLC (Shimadzu, Japan) coupled with a C18 reversed-
207 phase column (Atlantis® T3 C18, 100 Å, 3 µm, 2.1 mm × 100 mm, Waters). The
208 mobile phases were run with 0.1% formic acid and 100% acetonitrile, respectively,
209 from 5% to 95% gradient in one hour to improve separation performance (Andrés-
210 Costa, Andreu, & Picó, 2017).

211 For qualitative non-targeted screening, several parameters were used to improve the
212 confidence of analyte identification. The detailed instrumental settings used for
213 qualitative non-targeted analyses were as follows: injection volume was 5 µL for each
214 sample, flow rate was 0.30 mL/min, temperature of autosampler was 4□, and column

temperature was 30□. Before analyses, the two pumps and the autosamplers were purged manually. For MS, both modes were surveyed with the information-dependent acquisition (IDA) mode. The detailed conditions were as follows: spray voltage was 5500 V; temperature was 500□; pressure was 45 psi for both gas 1 and 2; curtain gas was 30; collision-activated dissociation (CAD) gas was 7; nitrogen gas was the nebulizer, curtain, and collision gas. For time-of-flight mass spectrometry (TOF MS) scanning, the mass range was set at 100-1000 Da; de-clustering potential was 80 V; collision energy was 10 V; for MS/MS scanning, mass range was 50-1000 Da; de-clustering potential was 80 V; collision energy was 35 V with capillary electrophoreses (CE) spread at 15. In the negative mode, the spray voltage was set at -5000 V.

Pre-calibration was conducted prior to running the samples, thereby ensuring that any errors fell within the default requirements. A second calibration was also required when running the experimental samples to help monitor the stability of the instrument in terms of signal intensity value readings. The acquired data can be viewed and processed by the Explore® and Analytics® software packages, respectively, and can also be accessed in online libraries such as METLIN (Zhu et al., 2013), HMDB (Wishart et al., 2007) and MassBank (Horai et al., 2010).

The quantitative analyses mainly focused on quantifying the level of selected pollutants concentrated in the benthic biofilms and corresponding bulk water samples. Diethyltoluamide (DEET), fenoxycarb, strychnine, and dipyridamole were selected based on their stability and commercial availability compared with other pollutant chemicals. As for most quantitative analyses in LC-MS/MS, method validation was necessary.

When choosing the spiked concentration, we firstly checked the response and selected compound area in the samples and used these to determine the stock solutions as well as the working solutions. The best results were obtained with concentrations of 1 µg/mL, 5 µg/mL, 10 µg/mL, 50 µg/mL, and 100 µg/mL;however, when ng/mL concentrations were prepared and examined by QTOF MS/MS, the responses were relatively low, so only the selected concentrations listed above were used.

In the determination of matrix effects in environmental samples, especially freshwater samples, it is usually difficult to find matrix-matched matrices. In this study, we utilized water from our laboratory-engineered freshwater system, in which a miniature system is designed to mimic the natural environment of Hong Kong. The water quality was checked by LC-MS/MS, and no emerging pollutants were detected within the detection limit.

The laboratory method validation applied in this study included tests of linearity, recovery, repeatability, intra-laboratory reproducibility, limit of detection (LOD), limit of quantitation (LOQ), and matrix effects, as previously described in other studies with some modifications (Fraselle, Derop et al., 2007; Huerta et al., 2016). Briefly, key parameters having stable values were recorded during MS instrument calibration and before running experimental samples. Experimental samples were spiked with internal standards and run in triplicate on three separate weeks. LOD and LOQ were calculated as the concentrations giving a signal to noise ratio of 3 and 10, respectively. Recoveries were acquired by following the same procedure of SPE, rotary evaporation, and reconstitution described in detail in section 2.4. Matrix effects were evaluated at the spiked concentration of 5 µg/mL.

2.5 Biofilm DNA extraction, sequencing, and analyses

Fouled membranes bearing benthic biofilms were first scraped off to collect the dense biomass into 50 mL Falcon® tubes. The remaining biomass on the filters was recovered through a final sonication step, by submerging the fouled membranes in glass test tubes containing sterile ultrapure water. The pooled biofilm biomass sample was then processed for genomic DNA (gDNA) extraction. Here, a Quick-DNA™ Fecal / Soil Microbe Miniprep Kit (ZYMO, CA, USA) was applied, following the extraction instructions recommended by the kit manufacturers.

A Bio-drop spectrophotometer was used to assess the yield and quality of the gDNA products prior to further processing (Table S1, supplementary materials). All DNA samples were sent to BGI (Shenzhen, China) for amplicon V3-V4 sequencing and analyses. An amplicon library with fusion primers (16S V3-V4) was first constructed, with dual index and adapters for PCR. After sequencing using a MiSeq PE300, the raw data were filtered to eliminate the adapter pollution and low-quality data to obtain clean reads. The paired-end reads with overlap were merged using FAST (Fast Length Adjustment of Short Reads, v1.2.11) and the generated tags were clustered to operational taxonomic units (OTUs) at 97% sequence similarity using USEARCH (v7.0.1090). Taxonomic ranks were assigned to OTU-representative sequences using the Ribosomal Database Project (RDP) Name Bayesian Classifier v.2.2, and rank curves were drawn by R-software (v3.1.1). Finally, features such as alpha diversity and beta diversity were analyzed based on the OTUs and taxonomic ranks. A descriptive bioinformatics analyses pipeline is presented in Figure S1 in the supplementary material section.

2.6 Scanning electron microscopy analyses of benthic biofilms

Rocks were collected and stored at -80□ before analyses. Small pieces were obtained using a hammer. The rocks were washed with saline buffer before fixation in 10% neutral buffered formalin for 24 hours at 4□. The rocks were dehydrated in an increasing series of ethanol concentrations (30%, 50%, 70%, 90%) for 15 min each with 2 changes of solvent and 100% ethanol for 20 min with 3 changes of solvent. The rocks were then dried in a Critical Point Dryer (Baltec CPD 030, Leica, Germany) using liquid carbon dioxide as translational fluid. SEM images were acquired with a Hitachi S-3400N scanning electron microscope (Hitachi, Japan).

2.7 Statistical analyses

For 16s rDNA sequencing analyses, alpha diversity indices including number of species observed, Shannon index, Chao index, abundance-based coverage estimate (ACE), and Simpson index were analyzed by Mothur (v1.31.2) (Schloss et al., 2009). These indices were used to provide different parametric and non-parametric diversity estimates. For species annotation and community structures, out-representative sequences were taxonomically classified using the RDP Classifier trained with the Greengenes database, and the confidence value was 0.8. The Wilcoxon rank-sum test was used for two-group comparison, while the Kruskal-Wallis test was used for multi-group comparison. All statistical analyses were performed with R software (version 3.1.1).

3. Results

3.3 LC-QTOF-MS/MS analyses

To test the validity of biofilms as viable indicators of environmental health, both freshwater and corresponding biofilms sampled in three different areas in Hong Kong were surveyed for non-targeted and targeted pollution analyses using LC-QTOF-MS/MS.

3.3.1 Untargeted analyses of pollutants in both freshwater and corresponding biofilm samples

The non-targeted analyses revealed that most features could not be matched to existing libraries. Note that the parameters in the non-targeted analyses were based on library match confidence. Only when checking the MS/MS spectra of matched annotations could similarities be observed. The final annotation results presented 28 chemicals, most of which were identified as pharmaceutical compounds, as listed in Table 1. Positive mode ionization appeared to have a better performance in library matching, while the parent ion information provided by the negative mode resulted in much fewer matched results in this study.

Differences were found between the non-targeted analytical profiles in freshwater samples and their corresponding biofilms at the collecting point. Specifically, the results unveiled the presence of pollutants in the sampled biofilms that were not identified in the corresponding bulk water samples, which was especially the case for the Lung Fu Shan samples (ST3, ST3B). It is always necessary to cross-validate the matching of the results; thus, the standards were also analyzed by the same method. Taking DEET as an example, Figure S2 (supplementary materials) shows the MS/MS spectra of the DEET standard in the data-independent acquisition mode SWATH

(sequential window acquisition of all theoretical spectra), in which the m/z value of the parent ion is 192.1399, and the adducts of the [M+H]+ ion with distinct daughter ions have m/z values of 119.0495, 91.0545, 72.0451, and 65.0388. The differences indicate neutral loss during the fragmentation process. In IDA mode of the sample analyses (Figure S3, supplementary materials), both MS1 and MS2 spectra were extracted. The MS2 spectra had similar daughter ions to MS1, which indicated the library match was of high confidence.

3.3.2 Concentration determination and method validation

Linearity was obtained by plotting the peak area of five working solutions, and a regression model was used to quantify the data. In the quantitative study using the LC-MS/MS instruments, the priority was set to triple quadrupole model, which enabled the selection of multiple reaction monitoring (MRM). The resolution of QTOF was improved by forfeiting the quantitation capacity. Specifically, the mass detector is easily affected by temperature shifts, resulting in a systematic quantitative error. Therefore, *R* square with a value higher than 0.98 was accepted in this study (Table 2).

In terms of recoveries (Table S3), the overall results were acceptable but with some exceptions, for example DEET had lower recoveries at high concentration while strychnine had poor recovery at low concentration. The minimum LOD and the maximum LOQ were 0.04 µg/mL and 1.96 µg/mL, respectively. Matrix effects were obvious in all four chemicals, especially for dipyridamole at -93%, fenoxycarb at -62%, and strychnine at 69% (positive values indicate ion enhancement, while negative values represent ion suppression). Four samples had relatively acceptable

358 reproducibility and repeatability (Table 3), but dipyridamole and strychnine had
359 higher deviation.

360 After calculating the concentration of pollutant compounds in all samples (Table 4),
361 the biofilms were generally shown to have higher concentrations of pollutants by one
362 order of magnitude compared with the corresponding bulk freshwater samples at that
363 time. In general, the EPS concentrations in the biofilm samples were much higher
364 than in bulk water in general, except for dipyridamole.

365

366 ***3.4 Sequencing analyses***

367 The microbial community structures were analyzed to compare the microbial profiles
368 of benthic biofilms sampled from different freshwater sites through 16S
369 metagenomics.

370 The alpha diversity data revealed that ST3 had the highest species diversity and
371 richness, with 3248 observed species compared with 2129 and 1897 observed species
372 in samples ST2 and ST1, respectively. Among the three sampled benthic biofilms, the
373 highest alpha diversity was observed in sample ST3, with a Shannon index of 6.4
374 compared with 4.5 and 4.03 in samples ST1 and ST2, respectively (Table S2,
375 supplementary materials).

376 To assess the extent of unique or shared OTUs between sampled benthic biofilms, a
377 beta-diversity analysis was performed as depicted in Figure 1. Biofilms from Lung Fu
378 Shan (ST3) had the largest beta-diversity with 1367 unique OTUs, or 42% of its total
379 OTUs, compared with 363 OTUs and 219 OTUs at Sam Dip Tam (ST1) and Ho

Chung Rivers (ST2), respectively (Figure 1). Interestingly, the sampled benthic communities shared 967 common OTUs between ST3, ST2, and ST1.

Further relative abundance heat map analyses (Figure 2) revealed the presence of two taxonomic clusters based on abundance levels in all tested biofilms. *Proteobacteria*, *Cyanobacteria*, *Actinobacteria*, *Acidobacteria*, and *Bacteroidetes* were the major phyla in all biofilm samples. Of the 42 identified phyla, 14 were clustered as the most abundant, and 28 as the least abundant, and included *Chlamydiae*, *Fusobacteria*, *Spirochaetes*, *Fibrobacteres*, and *Caldiserica*. Horizontal clustering of the tested samples presented ST2 and ST1 as one cluster group differentiated from ST3, which displayed a unique abundance profile. More specifically, ST3 was characterized by a much higher abundance of identified taxa compared with ST2 and ST1, including *Actinobacteria*, *Chloroflexi*, *Planctomycetes*, and *Verrucomicrobia*.

Taxonomic composition distribution analyses (Figure 3) showed that *Cyanobacteria* was the most abundant phylum in Ho Chung River (ST2) and Sam Dip Tam (ST1) at 75.01% and 52.01%, respectively, but was found to be the least abundant in the Lung Fu Shan river (ST3) at 10.05%. In contrast, *Proteobacteria* was the most abundant in ST3 at 47.67%, compared with 26.79% and 13.96% in ST1 and ST2. In a like manner, *Actinobacteria* was most abundant in ST3 at 10.12% compared with the ST1 and ST2 biofilms, which had abundance values of 2.56% and 1.84%, respectively. In more detail, a huge decrease in cyanobacterial abundance was observed in the Lung Fu Shan (ST3) samples, but a clear rise in *Proteobacteria* (Figure S4, supplementary materials).

We also used Random Forest to analyze which marker microorganisms played the most critical roles in the response to EPS. The results are shown as the value of mean

decrease in accuracy, which indicates the importance of each phylum-level microorganism (Figure 4). These analyses contrasted with the OTU-based abundance analyses, by showing that *Chlorobi* was the feature that most strongly impacted the freshwater biofilm system, followed by *Nitrospirae*, *Bacteroidetes*, and *Chlamydiae*.

The predicted metabolic function analyses (Figure S5) highlighted membrane transport, amino acid metabolism, and carbohydrate metabolism in all three samples, with slight abundance variations. Compared with the other tested samples, the Lung Fu Shan (ST3) biofilm samples had the lowest and highest abundance in energy metabolism and carbohydrate metabolism, respectively. Moreover, of the total of 22 metabolic features identified in ST3, 18 of these were clustered as least abundant. Based on these analyses, we also analyzed key features based on the Random Forest score (Figure S5, supplementary materials). The results showed that energy metabolism was the first feature in EPS response, followed by amino acids and membrane transport, as visually displayed in Figure S6.

3.5 Scanning electron microscopy analyses

The rock biofilms displayed heterogeneous structures at all three sampling sites (Figure 5). It is clear that diatoms are prevalent in all three aquatic environments and have a variety of shapes. The biofilms were mostly covered by the diatoms. In the ST1 rocks, various layered bacterial shapes were observed, which indicated cross-linked bacterial communities. In the ST2 and ST3 rocks, EPS could be observed, but the bacterial communities were different from those of the ST1 rocks. These images generally demonstrate the presence of a mixture of species on the environmental

biofilms and suggest potentially significant structural differences in response to chemical pollutants at a single time point.

4. Discussion

In our study, three individual sampling sites were selected in different parts of Hong Kong. The Hong Kong government has been monitoring the river quality since the last century, and detailed data can be obtained from the official website: http://wqrc.epd.gov.hk/en/water-quality/river-1.aspx, in which longitudinal analyses of water quality are presented based on traditional physiochemical and coliform indicators. Based on these data, the water quality in Sam Dip Tam (ST1) and Ho Chung River (ST2) was rated as excellent or good, while the Lung Fu Shan (ST3) water quality was unclassified. All sampling sites were not in the vicinity of industrial factories. Based on the identified pollutant profiles in both water and biofilm samples (Table 1), the presence of emerging pollutants may be linked to human activities, as most of the EPS in the non-targeted analyses were prescription drugs and personal care products. Previous studies have placed significant emphasis on the toxicity of these compounds, more specifically with regards to their short- and long-term environmental and health impacts, usually concluding with low risk evaluations based on low collected concentrations (ng/L) (Tsaboula et al., 2016). However, prescription drugs, especially antibiotics, should be of significant concern regardless of concentration levels, as they can potentially accelerate the spread of antibiotic resistance genes in freshwater environments (Marti et al., 2014). In our study, antibiotics such as monensin were identified in both bulk water and corresponding biofilm samples from the ST3 stream (Table 1), highlighting the importance of conducting targeted antibiotic-based analyses for assessing the likelihood and severity

452 of the interaction between these antibiotics and their surrounding environments
453 (Baquero, et al., 2008; Kaeseberg et al., 2018; Devarajan et al., 2015). The main
454 drawback in monitoring emerging pollutants from environmental samples, such as
455 biofilms, is the lack of standardized sampling and processing protocols. The technical
456 aspects of pollutant analyses also have their limitations. In our study, HLB columns
457 were utilized during the SPE extraction step for the qualitative analyses of emerging
458 pollutants; however, these columns may not be suited for alkaline or acidic
459 compounds (Andrade-Eiroa et al., 2016). Similarly, the C18 column used for HPLC is
460 also only suitable for hydrophilic compounds (Cruz et al., 1997). These technical
461 limitations may restrict our understanding of the degree of pollution based on
462 identifiable contaminant profiles. Quantifying the concentrations of these detected
463 pollutants from samples also presents challenges when constructing libraries that are
464 not fully annotated. Using freely accessible libraries such as Metlin and MassBank
465 also has drawbacks, as the collision energy (CE) values may differ from one library to
466 another. Different CE values can result in totally different fragmentations, which can
467 be an obstacle for manual interpretation (Richardson, 2009). In terms of these
468 limitations, several improvements can be proposed based on the above discussion. In
469 the extraction process, other SPE columns such as WCX and MAX could be
470 investigated. Although hydrophilic interaction chromatography (HILIC) HPLC
471 columns are ideal for separating hydrophilic compounds, it is highly recommended
472 that laboratory-specific databases are constructed using LC-MS/MS, despite being a
473 time-intensive task.

474 The association between bacterial communities and water pollution has been
475 extensively studied (Dobor et al., 2012; Logue et al., 2016), to the extent that attempts
476 have been made to use microbial biofilms as markers to trace the origins of pollution

(Gillings et al., 2015; Vierheilig et al., 2015). In our study, 16s rRNA sequencing analyses were utilized to annotate bacterial species in different sampling sites, with varying water pollution profiles. The results revealed widely varying microbial composition profiles among the studied samples, with biofilms from ST3 having the highest diversity and content of identifiable environmental pollutants. It has been argued that stream biofilm structure and composition are mostly impacted by physical conditions, such as pH, temperature, and light (Battin et al., 2016). In our study, the physicochemical properties of stream water were therefore considered, by monitoring parameters such as water temperature and pH prior to sampling. Although no obvious differences were observed between the sampling sites, other water parameters such as shear conditions or water velocity would also need to be compared in future studies. Moreover, other aspects such as biofilm growth stage could not be established from our biofilm samples, which could have potentially contributed to variations in microbial community profiles between samples. Future studies should therefore consider longitudinal investigations, allowing comparison of biofilms from different sites based on their microbial profile.

At the phylum level, *Acidobacteria* and *Armatimonadetes* are the two main classes found in our samples. *Acidobacteria* are commonly found in soil and water, and studies have shown that they have important ecological roles (Eichorst, Breznak, & Schmidt, 2007). Moreover, they actively respond to environmental pollution (Sun, Dafforn, Johnston, & Brown, 2013). *Armatimonadetes* were firstly identified in aquatic environments, and studies have linked them with heavy metal pollution (van Straalen et al., 2014). Two *Actinobacteria,* BHI80-139 and AD3, had significantly higher abundance in ST3. However, the abundance of taxa does not always provide any insight into which bacteria play core roles in the system, whereas the Random

Forest scores indicated that some of the least abundant species could affect the whole biofilm system through their response to EPS. However, little research has been conducted on these low-abundance bacteria, such as *Chlorobi,* as they always show a negative relationship with EPS abundance, while *Cyanobacteria* has the opposite trend (Figure S7). Two hypotheses can be generated, namely, that these bacteria actively respond to the existence of EPS, or that they help to maintain the stability of the biofilm system; more studies are needed to resolve this. In addition to sequencing, the bacterial communities in this study could be observed to a limited extent by SEM. However, the use of SEM may be restricted by the different ages of the biofilms. In any case, it is necessary to conduct a longitudinal study.

Based on our data, there could exist a relationship between the quality of freshwater ecosystems and the corresponding biofilms. Specifically, the biofilms possess more diversity in highly polluted areas, as can be summarized from the 16s rRNA results and the qualitative and quantitative profiling analyses of individual pollutants. Further analyses based on sequencing data revealed key features in the microbial communities, such as the presence of *Chlorobi*. These methods could be combined to uncover biological markers relating to freshwater quality.

5. Conclusions and Future Perspectives

The pollutant profile in a dynamic and ever-changing bulk water system, such as a stream, could possibly misrepresent and underestimate the true level and diversity of pollutants, through dilution effects and the passage of time since pollution events. Here, we have shown that benthic biofilms may reflect both the current and past pollutant profile of a freshwater environment, by comparing and analyzing both

biofilms and their corresponding bulk water. Although further developments are needed in the analytical testing of benthic biofilms, the integration of chemical-pollutant profiles and sequencing data may enable a monitoring tool based on the use of freshwater microbial communities as an ideal proxy for freshwater environmental health. Further studies in this context, especially with regards to unravelling the fundamental responses of biofilm microbial communities to exposure to key environmental pollutants, should see the development of enhanced targeted biofilm indicators, most specifically towards biofilm biomarkers.

Acknowledgments

This work was funded by Research Grant Council of Hong Kong (NO.: 27200917). The authors would like to thank Prof. David Lee PHILIPS, Dr. Lili DU, and Mr. Wenchao WANG from the Department of Chemistry of The University of Hong Kong, for their technical support on some spectroscopy experiments. The authors also thank HKU's Electron Microscopy Unit for their kind help in sample preparation and scanning electron microscopy.

References:

Andrade-Eiroa, A., Canle, M., Leroy-Cancellieri, V., & Cerdà, V. (2016). Solid-phase extraction of organic compounds: a critical review (Part I). *TrAC Trends in Analytical Chemistry*, 80, 641-654.

Andrés-Costa, M. J., Andreu, V., & Picó, Y. (2017). Liquid chromatography–mass spectrometry as a tool for wastewater-based epidemiology: Assessing new psychoactive substances and other human biomarkers. *TrAC Trends in Analytical Chemistry*, *94*, 21-38.

Baquero, F., Martínez, J.-L., & Cantón, R. (2008). Antibiotics and antibiotic resistance in water environments. *Current opinion in biotechnology*, *19*(3), 260-265.

Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M., & Packmann, A. I. (2016). The ecology and biogeochemistry of stream biofilms. *Nature Reviews Microbiology*, 14(4), 251.

Bouwmeester, H., Hollman, P. C., & Peters, R. J. (2015). Potential health impact of environmentally released micro-and nanoplastics in the human food production chain: experiences from nanotoxicology. *Environmental Science & Technology*, *49*(15), 8932-8947.

Cruz, E., Euerby, M. R., Johnson, C. M., & Hackett, C. A. (1997). Chromatographic classification of commercially available reverse-phase HPLC columns. *Chromatographia*, *44*(3-4), 151-161.

de Paul Obade, V., & Moore, R. (2018). Synthesizing water quality indicators from standardized geospatial information to remedy water security challenges: A review. *Environment international*, 119, 220-231.

Devarajan, N., Laffite, A., Graham, N. D., Meijer, M., Prabakar, K., Mubedi, J. I., ... & Poté, J. (2015). Accumulation of clinically relevant antibiotic-resistance genes, bacterial load, and metals in freshwater lake sediments in Central Europe. *Environmental Science & Technology*, *49*(11), 6528-6537.

Draper, A. D., Doores, S., Gourama, H., & LaBorde, L. F. (2016). Microbial Survey of Pennsylvania Surface Water Used for Irrigating Produce Crops. *Journal of Food Protection*, *79*(6), 902-912. doi:10.4315/0362-028X.JFP-15-479.

Dobor, J., Varga, M., & Záray, G. (2012). Biofilm controlled sorption of selected acidic drugs on river sediments characterized by different organic carbon content. *Chemosphere*, 87(2), 105-110.

Dudgeon, D. (2014). Accept no substitute: biodiversity matters. *Aquatic Conservation-Marine and Freshwater Ecosystems*, *24*(4), 435-440.

Eichorst, S. A., Breznak, J. A., & Schmidt, T. M. (2007). Isolation and characterization of soil bacteria that define Terriglobus gen. nov., in the phylum Acidobacteria. *Applied and Environmental Microbiology*, *73*(8), 2708-2717.

Flemming, H. C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A., & Kjelleberg, S. (2016). Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology*, 14(9), 563.

Fraselle, S., Derop, V., Degroodt, J. M., & Van Loco, J. (2007). Validation of a method for the detection and confirmation of nitroimidazoles and the corresponding hydroxy metabolites in pig plasma by high performance liquid chromatography-tandem mass spectrometry. *Anal Chim Acta*, *586*(1-2), 383-393.

Gerbersdorf, S. U., Hollert, H., Brinkmann, M., Wieprecht, S., Schüttrumpf, H., & Manz, W. (2011). Anthropogenic pollutants affect ecosystem services of freshwater sediments: the need for a "triad plus x" approach. *Journal of Soils and Sediments*, 11(6), 1099-1114.

Gillings, M. R., Gaze, W. H., Pruden, A., Smalla, K., Tiedje, J. M., & Zhu, Y. G. (2015). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *The ISME j*
ournal, *9*(6), 1269-1279.

Guasch, H., Artigas, J., Bonet, B., Bonnineau, C., Canals, O., Corcoll, N., . . . Morin, S. (2016). The use of biofilms to assess the effects of chemicals on freshwater ecosystems. *Aquatic Biofilms*, 125.

Ho, B. S. K., & Dudgeon, D. (2016). Are high densities of fishes and shrimp associated with top-down control of tropical benthic communities? A test in three Hong Kong streams. *Freshwater Biology*, *61*(1), 57-68.

Hounslow, A. (2018). Water quality data: analysis and interpretation: CRC press.

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., . . . Aoshima, K. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of mass spectrometry*, *45*(7), 703-714.

Huerta, B., Rodriguez-Mozaz, S., Nannou, C., Nakis, L., Ruhi, A., Acuna, V., . . . Barcelo, D. (2016). Determination of a broad spectrum of pharmaceuticals and endocrine disruptors in biofilm from a waste water treatment plant-impacted river. *Science of Total Environment*, *540*, 241-249.

Hutchins, D. A., & Fu, F. (2017). Microorganisms and ocean global change. *Nature microbiology*, 2(6), 17058.

Kaeseberg, T., Schubert, S., Oertel, R., Zhang, J., Berendonk, T. U., & Krebs, P. (2018). Hot spots of antibiotic tolerant and resistant bacterial subpopulations in natural freshwater biofilm communities due to inevitable urban drainage system overflows. *Environmental pollution*, *242*, 164-170.

Liao, K., Bai, Y., Huo, Y., Jian, Z., Hu, W., Zhao, C., & Qu, J. (2018). Integrating microbial biomass, composition and function to discern the level of anthropogenic activity in a river ecosystem. *Environment international*, *116*, 147-155.

Logue, J. B., Stedmon, C. A., Kellerman, A. M., Nielsen, N. J., Andersson, A. F., Laudon, H., ... & Kritzberg, E. S. (2016). Experimental insights into the importance of aquatic bacterial

community composition to the degradation of dissolved organic matter. *The ISME journal*, 10(3), 533.

Marti, E., Variatza, E., & Balcazar, J. L. (2014). The role of aquatic ecosystems as reservoirs of antibiotic resistance. *Trends in microbiology*, 22(1), 36-41.

Núñez, M., Borrull, F., Pocurull, E., & Fontanals, N. (2017). Sample treatment for the determination of emerging organic contaminants in aquatic organisms. *TrAC Trends in Analytical Chemistry*, *97*, 136-145.

Pachepsky, Y., Shelton, D., Dorner, S., & Whelan, G. (2016). Can E. coli or thermotolerant coliform concentrations predict pathogen presence or prevalence in irrigation waters?. *Critical reviews in microbiology*, 42(3), 384-393.

Parmar, T. K., Rawtani, D., & Agrawal, Y. (2016). Bioindicators: the natural indicator of environmental pollution. *Frontiers in life science*, 9(2), 110-118.

Pereda, O., Acuña, V., von Schiller, D., Sabater, S., & Elosegi, A. (2019). Immediate and legacy effects of urban pollution on river ecosystem functioning: A mesocosm experiment. *Ecotoxicology and environmental safety*, 169, 960-970.

Piggott, J. J., Salis, R. K., Lear, G., Townsend, C. R., & Matthaei, C. D. (2015). Climate warming and agricultural stressors interact to determine stream periphyton community composition. *Global Change Biology*, 21(1), 206-222.

Richardson, S. D. (2009). Water analysis: emerging contaminants and current issues. *Analytical chemistry*, 81(12), 4645-4677.

Sabater, S. (2017). Microbial Ecotoxicology: Looking to the Future. In Microbial Ecotoxicology (pp. 339-352). Springer, Cham.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Robinson, C. J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537-7541.

Scott, W. C., Breed, C. S., Haddad, S. P., Burket, S. R., Saari, G. N., Pearce, P. J., ... & Brooks, B. W. (2019). Spatial and temporal influence of onsite wastewater treatment systems, centralized effluent discharge, and tides on aquatic hazards of nutrients, indicator bacteria, and pharmaceuticals in a coastal bayou. *Science of the Total Environment*, 650, 354-364.

Smetanová, S., Bláha, L., Liess, M., Schäfer, R., & Beketov, M. (2014). Do predictions from Species Sensitivity Distributions match with field data? *Environmental pollution*, *189*, 126-133.

Sun, M. Y., Dafforn, K. A., Johnston, E. L., & Brown, M. V. (2013). Core sediment bacteria drive community response to anthropogenic contamination over multiple environmental gradients. *Environmental microbiology*, *15*(9), 2517-2531.

Tsaboula, A., Papadakis, E. N., Vryzas, Z., Kotopoulou, A., Kintzikoglou, K., & Papadopoulou-Mourkidou, E. (2016). Environmental and human risk hierarchy of pesticides: A prioritization method, based on monitoring, hazard assessment and environmental fate. *Environment international*, 91, 78-93.

van Straalen, N. M., van Gestel, C. A., Zhou, J., He, Z., Wen, C., & Roling, W. F. (2014). Microbial community composition and functions are resilient to metal pollution along two forest soil gradients. *FEMS microbiology ecology*, *91*(2015).

Vaughn, C. C., & Hoellein, T. J. (2018). Bivalve Impacts in Freshwater and Marine Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, 49, 183-208.

Vierheilig, J., Savio, D., Ley, R. E., Mach, R. L., Farnleitner, A. H., & Reischer, G. H. (2015). Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. *Water Science and Technology*, 72(11), 1962-1972.

Wang, R., Zimmerman, J. B., Wang, C., Font Vivanco, D., & Hertwich, E. G. (2017). Freshwater Vulnerability beyond Local Water Stress: Heterogeneous Effects of Water-Electricity Nexus Across the Continental United States. *Environmental Science & Technology*, *51*(17), 9899-9910.

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., . . . Sawhney, S. (2007). HMDB: the human metabolome database. *Nucleic acids research*, *35*(suppl_1), D521-D526.

Wu, Y., Liu, J., & Rene, E. R. (2018). Periphytic biofilms: a promising nutrient utilization regulator in wetlands. *Bioresource technology*, 248, 44-48.

Xie, Y., Wang, J., Wu, Y., Ren, C., Song, C., Yang, J., . . . Zhang, X. (2016). Using in situ bacterial communities to monitor contaminants in river sediments. *Environmental pollution*, *212*, 348-357.

Yunfen, S., Manru, G., & Weisong, F. (1991). Assessment of water system in Changde City based on microbial communities. *Chinese Journal of Applied Ecology*, 2.

673 Zhu, Z.-J., Schultz, A. W., Wang, J., Johnson, C. H., Yannone, S. M., Patti, G. J., & Siuzdak, G. (2013).
674 Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of
675 metabolites guided by the METLIN database. *Nature protocols*, *8*(3), 451.

676

677

Compound	Formula	Found at Mass	CAS#	Mode	ST1	ST2	ST3	ST1B	ST2B	ST3B
Pharmaceuticals										
Acecarbromal	C9H15BrN2O3	279.1593	77-66-7	+	Y	Y		Y	Y	Y
Procyclidine	C19H29NO	288.2907	77-37-2	+	Y	Y			Y	Y
Ziprasidone	C21H21CIN4OS	413.2666	146939-27-7	+	Y	Y		Y	Y	Y
Dipyridamole	C24H40N8O4	505.2641	58-32-2	+	Y		Y	Y	Y	Y
Dilazep	C31H44N2O10	604.2996		+			Y			Y
Helvolic acid	C33H44O8	569.2953	29400-42-8	+	Y		Y	Y	Y	Y
Gamithromycin	C40H76N2O12	777.0376	145435-72-9	+			Y			Y
Monensin	C36H62O11	670.4292	17090-79-8	+			Y			Y
Oxymetazoline	C16H24N2O	260.1889	1491-59-4	+						Y
Phenyltoloxamine	C17H21NO	255.1623	92-12-6	+						Y
Tolycaine	C15H22N2O3	279.1720	3686-58-6	+				Y		
Thymopentin	C30H49N9O9	680.5152	69558-55-0	+		Y				
Fumigaclavine A	C18H22N2O2	297.1534		-	Y	Y		Y	Y	Y
Vardenafil	C23H32N6O4S	488.6030	224785-90-4	+						Y
Pesticides										
DEET	C12H17NO	192.1394	134-62-3	+	Y	Y		Y	Y	
Etofenprox	C25H28O3	376.2038	80844-07-1	+						Y
Hydramethylnon	C25H24F6N4	494.1905	67485-29-4	+			Y			Y
Fenoxycarb	C17H19NO4	302.2704	72490-01-8	+				Y	Y	
Strychine	C21H22N2O2	335.1482	57-24-9	+	Y	Y			Y	Y
Dodine	C15H33N3O2	228.2694	2439-10-3	+	Y			Y		
Industrial ingredients										
Dibutyl phthalate	C16H22O4	278.1518	84-74-2							Y
PEG-7/8/9/10/11/12mer				+	Y	Y	Y	Y	Y	Y
Toxin										
Sambucinol	C15H22O4	265.1479	90044-33-0	-	Y	Y		Y	Y	Y

Table 1 MS/MS spectra library match chemical profiles: B at the end of ST means biofilms. The shade columns indicate further quantitative study chemicals. Y means chemicals detected in the sample

685

686 Table 2 Linearity of standards

	Linear regression curve	*R* square
DEET	y = 3.46684e6 x + 4.30092e6	0.98023
Dipyridamole	*y* = 5.08285e6 *x* + 4.29770e7	0.99147
Fenoxycarb	*y* = 4.09234e4 *x* + 2.05930e5	0.98837
Strychnine	y = 1.02018e6 x + 2.79777e7	0.99809

687

688

689

690

691 Table 3 Reproductivity and repeatability of standards in matrix-match matrices, data was shown by
692 mean ± RSD (relative standard deviation) of five different spiked concentrations

	Reproducibility					Repeatability				
	1μg/mL	5μg/mL	10μg/mL	50μg/mL	100μg/mL	1μg/mL	5μg/mL	10μg/mL	50μg/mL	100μg/mL
DEET	4.73e6±1.57	3.15e7±1.38	5.30e7±1.47	1.47e8±4.59	3.38e8±0.33	6.18e6±7.52	2.22e7±7.81	3.48e7±8.52	1.10e8±7.38	1.56e8±0.66
Dipyridamole	3.24e7±6.93	1.08e8±7.22	1.38e8±7.14	1.78e8±6.73	2.34e8±6.41	2.05e7±13.83	4.62e7±14.35	5.08e7±13.95	9.81e7±1.37	1.16e8±12.52
Fenoxycarb	1.94e5±0.74	6.04e5±0.52	8.79e5±0.23	2.29e6±1.00	3.30e6±0.30	5.74e4±10.10	1.82e5±11.10	1.94e5±10.70	3.38e5±7.03	1.07e6±7.32
Strychnine	5.60e7±0.33	1.69e8±3.24	1.96e8±1.48	3.12e8±1.47	4.48e8±1.30	5.37e7±10.80	1.24e8±9.13	1.77e8±8.36	2.66e8±8.13	5.84e8±12.50

693

694

695

696

697

698 Table 4 Concentration calculation in samples, NA-not available in detection

	Pollutant Concentration (μg/L) in freshwater samples			Pollutant concentration (μg/L*) in benthic biofilm samples		
	ST1	ST2	ST3	STB1	STB2	STB3
DEET	97.7	266	NA	258	246	NA
Dipyridamole	133	NA	1020	409	452	1080
Fenoxycarb	NA	NA	NA	1530	1820	NA
Strychnine	187	349	NA	1310	780	NA

699 *here the biofilms were disrupted, and pollutants were dissolved in the aqueous

700

701 **Captions to Figures and Tables**

702 Figure 1: A Venn diagram displaying the degree of overlap of bacterial OTUs among
703 biofilm samples collected from Sam Dip Tam (ST1), Ho Chung (ST2), and Lung Fu
704 Shan (ST3). The numbers of shared and unique OTUs between biofilms samples are
705 shown. A significantly greater ($P < 0.001$) number of unique OTUs was found in the
706 biofilm sample from Lung Fu Shan (ST3) river, previously characterized as the most
707 polluted of all three tested sites.

708

709 Figure 2: Log-scaled percentage heat map of taxonomic assignments at phylum level
710 for biofilm samples collected from Sam Dip Tam (ST1), Ho Chung (ST2), and Lung
711 Fu Shan (ST3) rivers.

712

713 Figure 3: Taxonomic assignments at phylum level for biofilm samples collected from
714 Sam Dip Tam (ST1), Ho Chung (ST2), and Lung Fu Shan (ST3) rivers.

715

716 Figure 4: Marker organism shift based on Random Forest score in three samples.
717 Mean decrease in accuracy value indicates the importance of each feature.

718

719 Figure 5: SEM observation of rocks from the three sampling sites, from left to right:
720 Sam Dip Tam (ST1), Ho Chung River (ST2), and Lung Fu Shan (ST3). Scale bar: 10
721 µm.

722

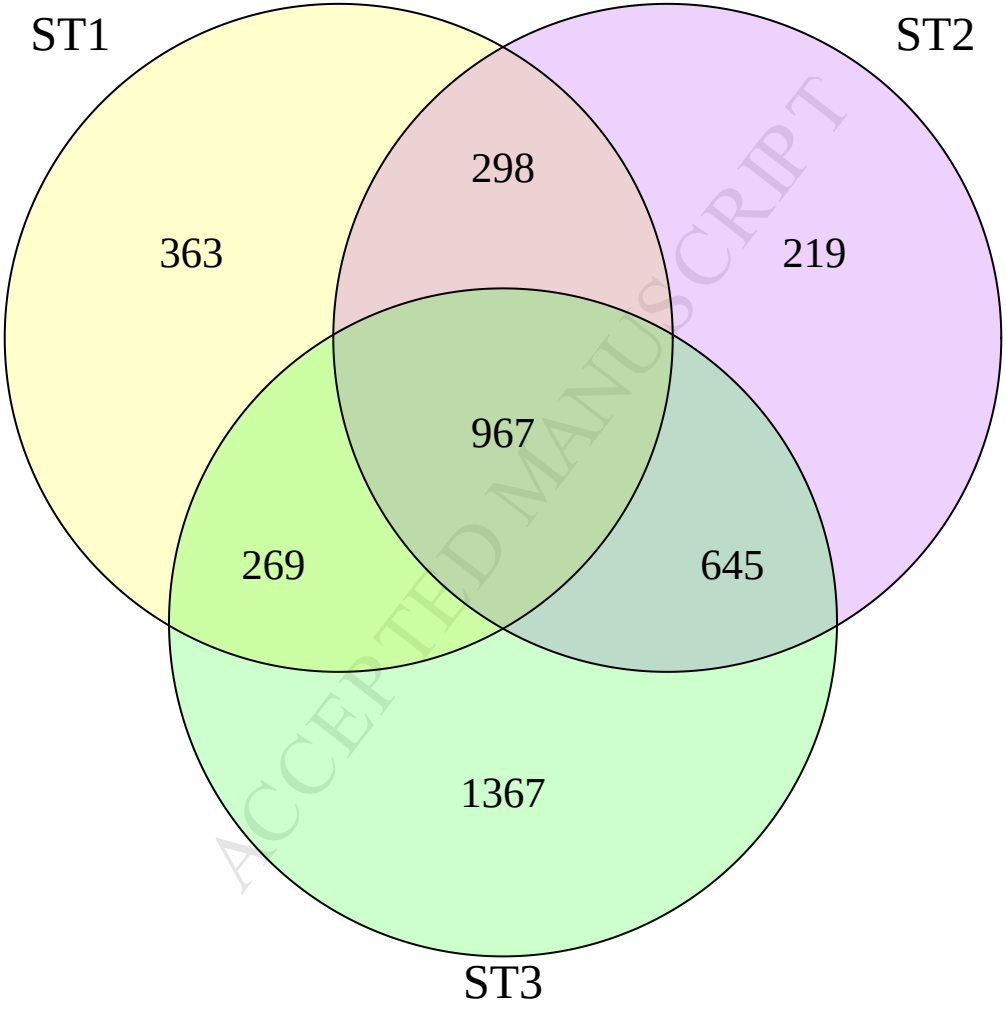
723 Table 1: MS/MS spectra library match chemical profiles: B at the end of ST means
724 biofilms. The shade

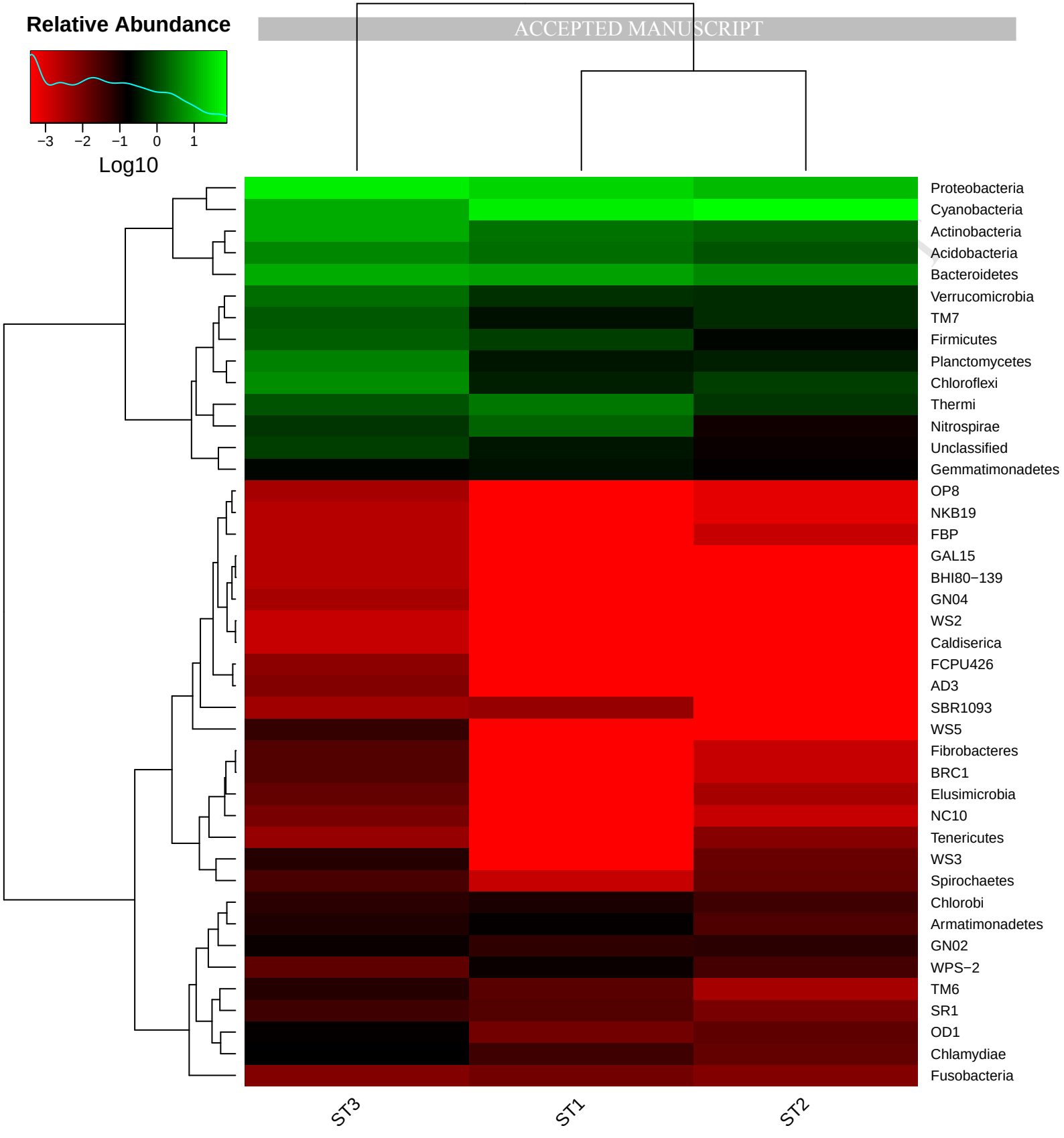
columns indicate further quantitative study chemicals. Y means chemicals detected in the sample

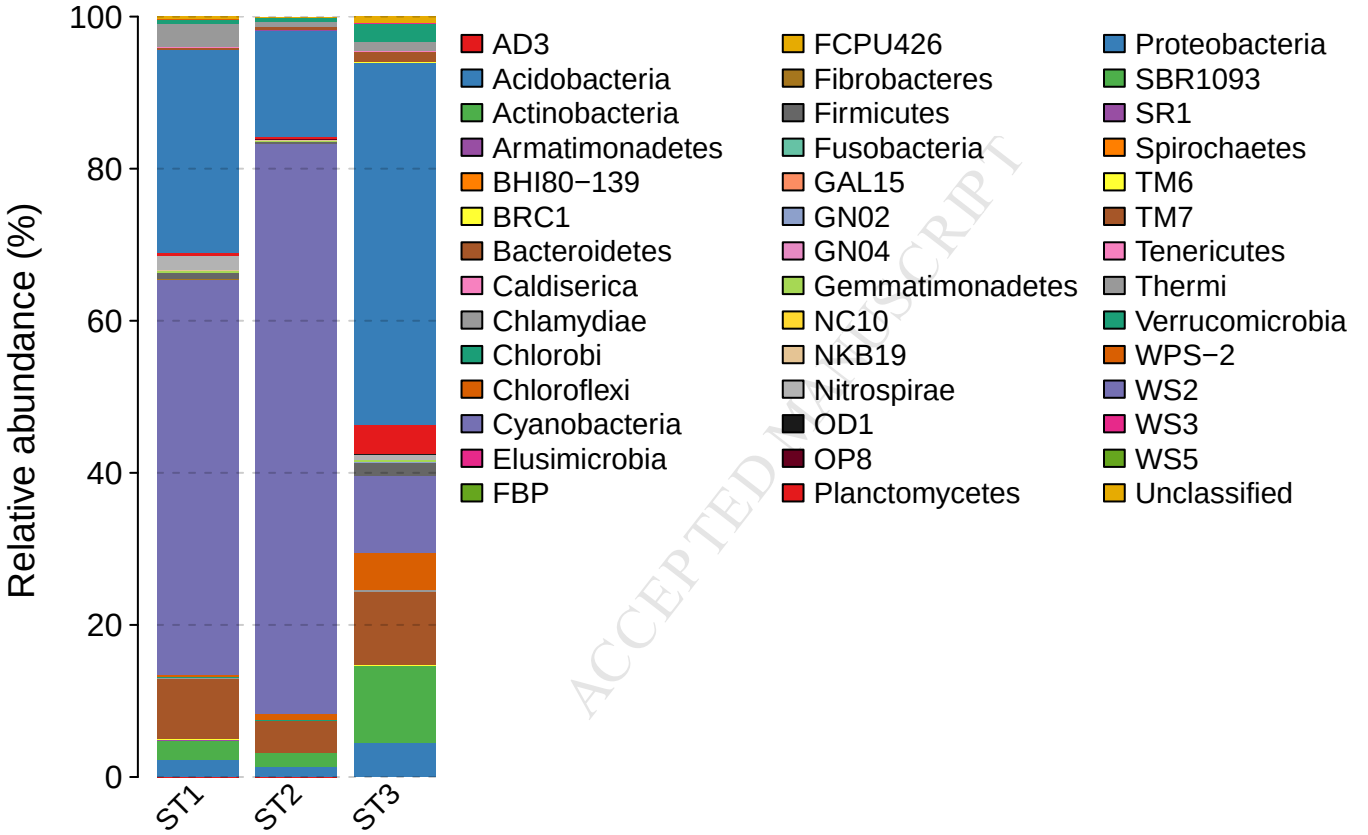
Table 2: Linearity of standards

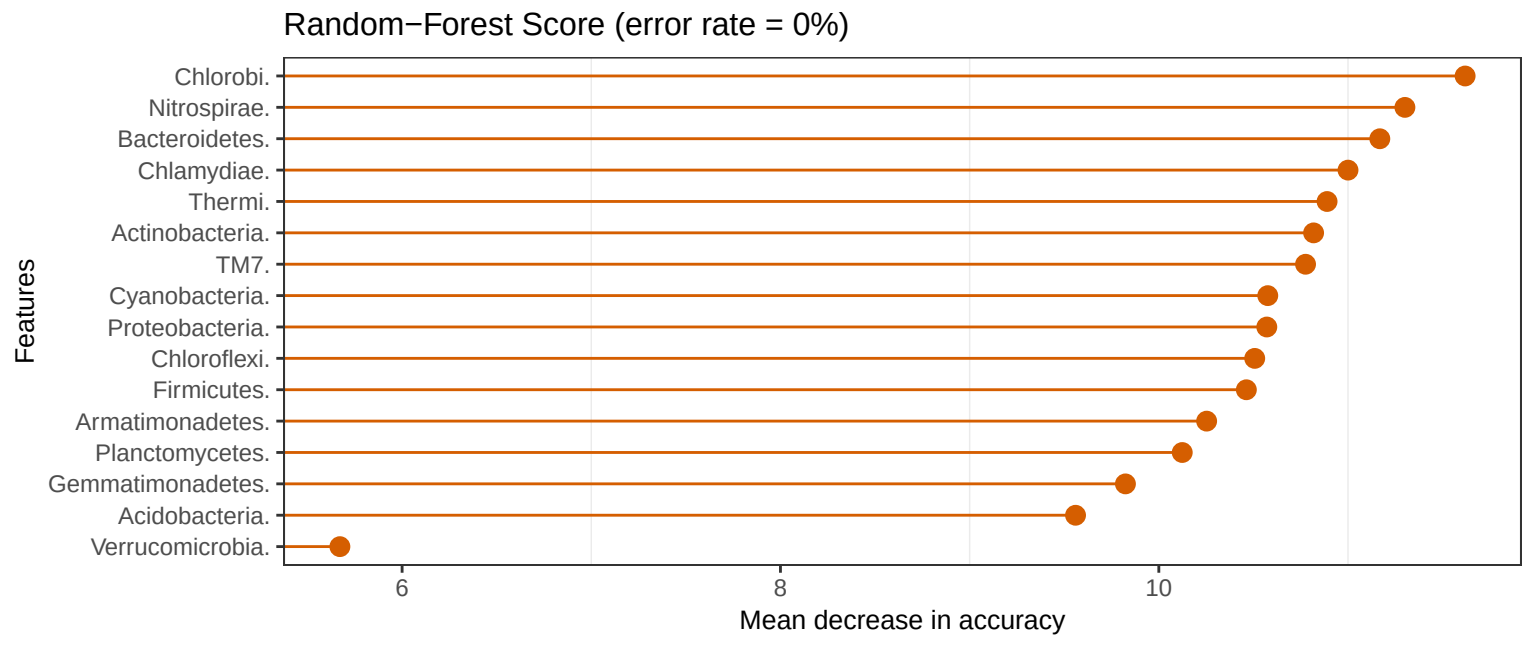
Table 3: Reproductivity and repeatability of standards in matrix-match matrices, data was shown by mean ± RSD (relative standard deviation) of five different spiked concentrations

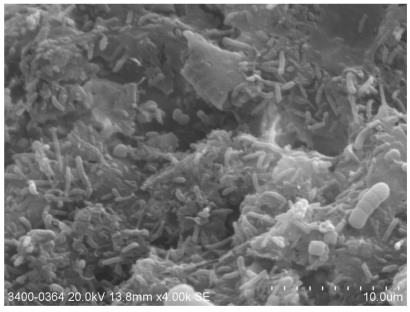
Table 4: Concentration calculation in samples, NA-not available in detection

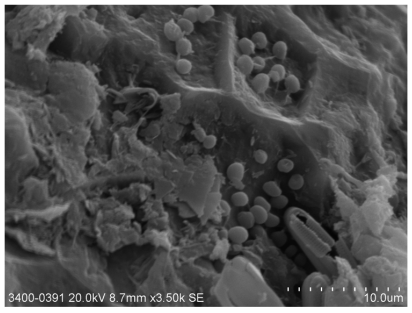


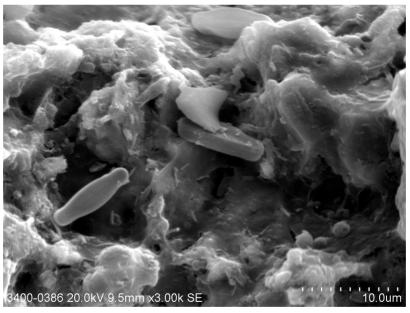





Random−Forest Score (error rate = 0%)
Features
Chlorobi.
Nitrospirae.
Bacteroidetes.
Chlamydiae.
Thermi.
Actinobacteria.
TM7.
Cyanobacteria.
Proteobacteria.
Chloroflexi.
Firmicutes.
Armatimonadetes.
Planctomycetes.
Gemmatimonadetes.
Acidobacteria.
Verrucomicrobia.
6
8
10
Mean decrease in accuracy







Highlights

1. Direct environmental sampling was investigated from different areas in this study.
2. Chemical profiling was obtained by surveying LC-QTOF-MS/MS with library matching.
3. Metagenomic analysis was performed to elucidate detailed taxonomy information.
4. Biofilm analysis as potential ideal proxy for monitoring freshwater environments health.
5. Future work needed for identifying pollution-related freshwater biofilm biomarkers.

.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.