Sources and drivers of contamination along an urban tropical river (Ciliwung, Indonesia)

Insights from microbial DNA, isotopes and water chemistry

Duvert, Clément; Priadi, Cindy R.; Rose, Alea M.; Abdillah, Ayik; Marthanty, Dwinanti R.; Gibb, Karen S.; Kaestli, Mirjam

Published in:
Science of the Total Environment

DOI:
10.1016/j.scitotenv.2019.05.189

Published: 10/09/2019

Document Version
Peer reviewed version

Sources and drivers of contamination along an urban tropical river (Ciliwung, Indonesia): insights from microbial DNA, isotopes and water chemistry

Clément Duvert, Cindy R. Priadi, Alea M. Rose, Ayik Abdillah, Dwinanti R. Marthanty, Karen S. Gibb, Mirjam Kaestli

Abstract

Wastewater treatment infrastructure is lacking in many developing countries, often resulting in high loads of contaminants discharged to urban rivers. In these countries, targeted pollution mitigation requires an understanding of where, how and when contaminants enter water bodies. Here we report on contamination of the Ciliwung River, a dynamic, tropical system flowing through the Jakarta metropolitan area (Indonesia). We measured a set of isotopic, chemical and microbial tracers in representative water and contamination sources, as well as longitudinally within the river, to assess the spatial and temporal variations in contaminant levels in and pathways to the river. In the dry season, we observed a tight coupling between locally recharged groundwater sources and the river, whereas in the wet season, one single water source originating from the fractured headwaters predominantly contributed to river flow. Yet, the flushing of upstream waters in the wet season did not always lead to the dilution of contaminants downstream. We delineated several contamination hotspots along the river, particularly active during the wet season due to higher hydrological connectivity between sources and the river. These hotspots may originate from septic tank leakage, as supported by metal ratios and dominant microbial communities, although we could not rule out other potential sources such as urban runoff or sediment resuspension. Bayesian source tracking on the whole microbial community proved useful in outlining processes that conventional tracers did not capture, such as the occurrence of a localised domestic contamination in the upper catchment, and the inflow of agricultural runoff all along the river profile during the wet season. Our study emphasises the role of rivers as biogeochemical reactors that constantly process and transform

contaminants and microbial communities. We also demonstrate the value of using isotopic, chemical and microbial tools together to trace the movement of water and contaminants through urban rivers.

Keywords: 16S rRNA, bacterial community, heavy metals, weathering, sediment, riverine processes

1. Introduction

Although developing countries are experiencing rapid demographic growth and economic development, the level of public investment in wastewater treatment infrastructure rarely keeps pace with increasing demand (McDonald *et al.*, 2011). Consequently, most urban river systems in the developing world are subject to deteriorating water quality (Capps *et al*., 2016). Because of largely unplanned urbanisation, metropolitan areas in these regions can comprise a mixture of land uses that include urban housing, industrial zones, small-scale urban farming as well as informal settlements. Mixed land use, combined with the lack of treatment facilities, leads to a wide range of chemical and biological contaminants entering rivers via both point and diffuse sources (García-Armisen *et al*., 2014; Pongmala *et al*., 2015). Different types of mitigation and remediation strategies are needed for different contaminants (Priadi *et al*., 2017), so it is essential to develop methods that can delineate and distinguish between dominant contamination sources to rivers.

Contaminants are conveyed to rivers via either direct runoff or subsurface flow pathways. Variations in rainfall, at both short-term and seasonal scales, are likely to exert a direct control on the magnitude and timing of contaminant delivery (Mouri *et al.*, 2011; Rochelle-Newall *et al.*, 2016). This is particularly significant in tropical countries, which face the challenge of extreme rainfall intensity and amount that often result in high rates of stormwater runoff on largely impervious areas (Silveira, 2002; Parkinson and Mark, 2005). High rainfall events may trigger the delivery of stormwater-borne contaminants to

46 downstream receiving waters, while at the same time diluting domestic wastewater (e.g. Surbeck *et al*.,
47 2006; Priadi, 2010; Sindern *et al*., 2016). Groundwater systems, in contrast, can temporarily store
48 contaminants before releasing them to rivers upon the activation of subsurface flow pathways (Heeren *et*
49 *al*., 2010). Hence, we need a comprehensive understanding of how different water sources contribute to
50 contaminant transport to rivers, and this requires a detailed assessment of the hydrological response and
51 dominant flow pathways.

52 Identifying water flow pathways and tracking contaminants can be achieved by exploring the spatial and
53 temporal variations of selected tracers in sources and receiving rivers. The stable isotopes of water have
54 long been used to distinguish between different water sources to rivers (e.g. Fontes, 1980; Hooper and
55 Shoemaker, 1986), and, when used in combination with other tracers (carbon isotopes, electrical
56 conductivity, silica), can provide reliable assessments of dominant hydrological processes and flow
57 pathways across catchments (e.g. Schulte *et al.*, 2011; Duvert *et al.*, 2015). Specific chemical elements
58 have been used as contamination source tracers, such as heavy metals for gasoline (e.g. Ayrault *et al*.,
59 2012) and agricultural sources (e.g. Smith and Blake, 2014), organic substances such as stanols for animal
60 manure (Derrien *et al.*, 2011), or polychlorinated biphenyls for pesticides (e.g. Guan *et al.*, 2009). With the
61 advent of next generation sequencing, microbial indicators based on DNA sequencing and complete
62 inventories of bacterial operational taxonomic units have been increasingly used to differentiate
63 contamination sources (e.g. Unno *et al*., 2010; Newton *et al*., 2013; Kabiri *et al*., 2016). These techniques
64 have shed light on the impact of urbanisation (Wang *et al.*, 2016; Hosen *et al.*, 2017), flow conditions
65 (Febria *et al.*, 2015; Wang *et al.*, 2016) and nutrients (Kaestli *et al.,* 2017) on riverine and estuarine
66 microbial communities.

67 Source tracking techniques based on chemical or microbial tracers are relatively recent and still fraught
68 with limitations (e.g. Priadi, 2010; Ayrault *et al*., 2012; Tan *et al.*, 2015; Costa *et al*., 2016). For instance,
69 microbial DNA analyses cannot always distinguish between different contaminant sources (Ibekwe *et al*.,

2016, García-Armisen *et al*., 2014). Combining tracers of contamination (both chemical and microbial) with hydrological tracers may help address the different shortcomings in source tracking, although very few studies have used these tracers simultaneously. Among the few examples we are aware of, Ben Maamar *et al.* (2015) showed that nitrate inputs via infiltrating water had a major effect on microbial communities in recently recharged groundwater. Sugiyama *et al.* (2018) related heavy rainfall to increased spring yields using groundwater-borne microbial composition as a tracer—a process that was not detectable using conventional tracers of water sources. Other studies reported on the linkage between microbial communities, redox conditions and heavy metal behaviour along groundwater flowpaths (Zhang *et al*., 2017; Ma *et al.*, 2019). Generally, there is a need for research that combines the use of contaminant fingerprints and hydrological tracers to provide better understanding of contaminant behaviour, and of the processes driving their spatial and temporal variability.

We assessed the dynamics of several contamination sources (domestic wastewater, agricultural runoff) to the Ciliwung River, a tropical river that flows through the Jakarta metropolitan area (Indonesia). Using a set of isotopic, chemical and microbial tracers, we evaluated the spatial (upstream–downstream continuum) and temporal (dry *vs* wet season) variations in contamination level and pathways to the river. In particular, we tested the following hypotheses: (H1) in this system, there is a spatial trend from predominantly agricultural contamination sources in the upstream reaches to domestic contamination downstream; (H2) in the wet season, due to higher runoff and lower residence time, the river acts as a "passive pipe" (*sensu* Raymond *et al*., 2016) and most contamination sources are affected by dilution.

2. Materials and methods

2.1. Site description

The Ciliwung River is a mesoscale river system in the western part of Java island, Indonesia (Fig. 1). It flows in a south-north direction from its volcanic headwaters to a largely urbanised alluvial floodplain. The catchment covers approximately 400 km^2 and spans a broad elevation gradient, from 3020 m above sea level at its peak in the South (Mount Pangrango) to sea level as the river flows into Jakarta Bay. Average slopes in the upper part of the catchment are > 8%, where the geology is dominated by Quaternary volcanic formations (mostly breccias, lavas and lahars; Irawan *et al.*, 2015). Slopes then rapidly decrease in the middle reaches (< 2%), where the catchment is underlain by a large alluvial fan made up of volcanic deposits (sandy tuff). In the lower reaches (slopes < 0.2%), the geology comprises finer alluvial deposits. The area has a humid tropical climate characterised by high precipitation (1,500–3,500 mm/year), with the bulk of rainfall occurring between November and March (Lubis *et al.*, 2008). The catchment supports a variety of land uses from primary rainforest in the headwaters, a mixture of small-scale farming (paddy fields, cassava, market gardening) and urbanised areas in the middle reaches, to almost entirely urban in the downstream section. In the downstream reaches, the natural flow of the river has been largely diverted via a network of canals.

The Jakarta Greater Region area is home to approximately 30 million people, with an average 3.6% annual population growth rate between 2000 and 2010 (Forbes, 2015). The city's only wastewater treatment plant serves 2% of the population, and it is estimated that 95% of the septic tanks leak (Kerstens *et al.*, 2015). Contamination of the river has been a concern since the 1990s (Palupi *et al.*, 1995), and increasingly high contaminant levels have been reported since, with low dissolved oxygen (Costa *et al.*, 2016), high heavy metals (Yasuda *et al.*, 2011), faecal coliforms (Phanuwan *et al.*, 2006), as well as biochemical and chemical oxygen demand (Tallar and Suen, 2015). Domestic wastewater has been identified as the major source of pollution across the catchment (Umezawa *et al.*, 2009). The wording "domestic wastewater" typically encompasses two distinct sources, i.e. overflow and leakage from septic tanks (often referred to

as “blackwater”), and kitchen and bathroom drainage (“greywater”). In this study, we consider these two domestic wastewater types as two separate contamination sources.

2.2. Sample collection

Water was sampled under dry (September 2017) and wet (February 2018) conditions from six river locations (Fig. 1; C1 to C6) selected based on ease of access and, more importantly, in an effort to capture the urbanisation gradient of the Ciliwung catchment. From upstream to downstream, we sampled the river at Cisarua (C1; considered as km 0 in the following), Gadog Bridge (C2), Otto Iskandar Bridge (Bogor; C3), Cibinong (C4), Kelapa Dua (Depok; C5), and Manggarai (Jakarta; C6; km 105). In addition to these six river locations, we also sampled eight sources representing four source types. Given the dominant land uses in the catchment, we identified three main contamination sources that potentially contribute to river discharge, i.e. greywater (duplicate sources GRY1 and GRY2; drainage from kitchens and laundries), blackwater (duplicate sources BLK1 and BLK2; septic tank) and rural drainage (duplicate sources AGR1 and AGR2; mixture of agricultural runoff and domestic wastewater). While most samples were collected midstream in the catchment, we assumed that each source will be representative of any similar contamination type across the whole catchment. To these three contamination sources, we added a “pristine” groundwater source likely contributing to river flow (duplicate sources GW1 and GW2). GW1 is a spring located in the headwaters of the catchment, while GW2 is within a semiurban land use further downstream.

At each site, samples were collected in acid-washed bottles (2M HNO_3) for chemical analyses (two 500 mL PP bottles), microbial sequencing (between one and six 1 L HDPE bottles), and isotopic analyses (one 50 mL polyethylene tube for water isotopes and one 250 mL HDPE bottle for carbon isotopes). Before sampling, bottles were rinsed thoroughly three times and, when feasible, river samples were collected by immersing the bottle about 30 cm deep below the water surface (Thuong *et al.*, 2013). Greywater samples

were directly taken from kitchen (GRY1) and bathroom (GRY2) discharge pipes, while blackwater samples were taken from inside septic tanks, at a depth of 0.6 to 1 m. The two AGR samples were collected directly from paddy field (AGR1) and cassava (AGR2) irrigation drains. All samples were kept on ice and processed within 24h of collection. Upon arrival to the laboratory, samples for metal analysis were acidified with 98% HNO_3 until pH<2 was reached.

In addition to collecting samples, we also measured discharge during the dry season using a handheld propeller flow meter at C1 to C4. Due to high flow velocities in the channel and potential for large boulder transport, we were not able to enter the river and manually gauge it during the wet sampling round.

2.3. Element and isotopic analyses

Dissolved oxygen, conductivity, pH and temperature were measured in the laboratory immediately after sampling using a multiparameter probe (HACH HQ30D). Major and minor elements were analysed by a certified external laboratory using Inductively Coupled Plasma – Atomic Emission Spectroscopy (ICP-AES), previously digested using 5% nitric acid (APHA 3120 B and 3030 E methods). Only those elements generally returning results above the detection limit are discussed further. Detection limits and analytical precision were 10 µg L^{-1} and 2.44% for zinc (Zn), 20 µg L^{-1} and 0.91% for iron (Fe), 10 µg L^{-1} and 1.49% for aluminium (Al), and 120 µg L^{-1} and 3.84% for silica (Si as SiO_2), respectively.

All water samples were analysed for oxygen ($\delta^{18}O$) and deuterium ($\delta^{2}H$) stable isotopes at Charles Darwin University, Environmental Chemistry and Microbiology Unit (ECMU) using a cavity ring-down spectrometer (Picarro L2120) fitted with a diffusion sampler (Munksgaard *et al*., 2011). The data were calibrated to the Vienna Standard Mean Ocean Water (VSMOW) by analysis of three standard waters of known isotopic composition. Replicate analyses indicate that analytical precision (±σ) was ±0.02 ‰ for $\delta^{18}O$ and ±0.23 ‰ for $\delta^{2}H$. Dissolved inorganic carbon (DIC) and its stable isotope composition ($\delta^{13}C_{DIC}$) were measured with an ISO-CADICA, which consists of an extraction chamber connected to a cavity ring-

down spectrometer (Picarro G2101-i). According to duplicate analyses, analytical error (±σ) was ±0.56 ‰ for $\delta^{13}C_{DIC}$ and ±0.42 mg L^{-1} for DIC. Further details on the use of the ISO-CADICA instrument can be found in Bass *et al*. (2012).

2.4. Microbial data analysis

DNA collection, extraction and processing

Samples were collected in triplicate for the 4x2 sources and in duplicate for the six river locations, totalling 36 samples per sampling round. Collected waters were filtered within 24h of collection through a sterile 0.45µm membrane (Pall Corporation). Each sample was filtered until the membrane was impervious to water, the filtered volumes varying between 15 mL and 2.25 L. Filtered membranes were stored at 4°C before they were shipped to Australia (Biosecurity Permit No. 0001790936) and extracted at Charles Darwin University following the DNeasy PowerWater Kit (QIAGEN) protocol. The quality and purity of extracted DNA was measured using a Nanodrop. Purified DNA was dried and shipped to the Australian Centre for Ecogenomics (ACE) in Brisbane for 16S rRNA gene amplicon sequencing.

Sequence analysis

The 16S rRNA gene was amplified using the 2015 Earth Microbiome Project protocol and universal V4 primers, forward 515FB (GTGYCAGCMGCCGCGGTAA) and reverse 806RB (GGACTACNVGGGTWTCTAAT). The amplicons were sequenced using the Illumina MiSeq platform. Sequences were processed to sequence variants (SVs) starting with the removal of poor quality sequences using the Trimmomatic software (sliding window of 4 bases and an average base quality above 15). All reads with less than 250 bases were excluded and the remainder hard trimmed to 250 bases. Reads were processed to SVs using the default parameters for the QIIME-2 workflow and DADA-2 algorithm (Callahan *et al*., 2016; Caporaso *et al.*, 2010). Taxonomic classifications were assigned to the SVs using BLAST+ and the reference database SILVA.

Processing sequence variant data

The R library Phyloseq was used to process 10,052 SVs. Because of the low biomass of some spring water samples, potential contaminant SVs from lab reagents were excluded from analysis (Salter *et al*., 2014; Karstens *et al*., 2018). The R package "decontam" was used to exclude 14 SVs based on their occurrence in extraction kit negative controls (0.5 threshold prevalence method) (Karstens *et al*., 2018). A further 6,089 SVs (61%) were excluded because they were only found in one sample and 22 SVs were excluded as they did not belong to the Bacteria or Archaea kingdoms. Ten % of samples (7/72) were removed due to low sequence counts of < 6,700 sequences (4 negative controls, one GW, one AGR, and one river sample). The GW1 wet season sample was also excluded due to probable sample contamination. Negative control samples had final sequence counts ranging between 0 and 19 sequences and a hierarchical cluster analysis in Primer-E v7 (Plymouth, UK) showed that no samples clustered with negative controls. The final dataset contained 2,027 different SVs and 64 samples. All remaining samples were rarefied to the lowest common sequence number per sample (6,769 sequences). Rarefaction curves indicated that with this cut-off, most samples reached their SV richness; however, one BLK and two river samples only reached approximately 70-80% of their SV richness.

Biostatistical data analysis

The rarefied SV data were squared-root transformed and a Bray Curtis distance matrix calculated. PERMANOVA analysis (999 permutations, Type III Sums of Squares) in Primer-E v7 (Plymouth, UK) was used for hypothesis testing to assess whether the bacterial composition differed between the source and river (sink) communities and seasons (wet/dry). Sample sites were included as a random factor nested in Source/Sink and Season status. A distance-based test for homogeneity of multivariate dispersions (PermDISP) was conducted to test for differences in data dispersion between sample groups. The impact of the different filtered volumes on the bacterial community results were tested by including volume as a covariate for the PERMANOVA analysis (Type I Sums of Squares). Non-metric multi-dimensional scaling

(nMDS) in Phyloseq was used to assess the similarity of the bacterial communities. A distance linear model and distance-based redundancy analysis (dbRDA) in Primer-E was used to associate abiotic factors (water physicochemical factors, stable isotopes and metals) with bacterial composition. Co-linear variables δ^2H, Ni, Zn and Al were excluded from the analysis. Model selection was based on the lowest AICc and BEST selection. A result was considered significant if $p<0.05$ unless otherwise stated. Using a R source file (Knights *et al*., 2011), Bayesian source tracking was performed on river samples each for the dry and wet season with the following sources: GW1 (n=3), GW2 (n=3), AGR (n=6), GRY (n=6) and BLK (n=6). Default parameters were used (0.001 for Dirichlet parameters and 1,000 sequence rarefaction). Rare SVs occurring in less than three samples were excluded for this analysis. Source training samples were checked for correct source allocation by leave-one-out validation. Source tracking was repeated three times and the same results were obtained. The relative standard deviation (RSD) ranged between 0 and 27% (mean 7%) except for five GRY source samples with larger RSD of 33-100% due to their small predicted contributions of 1-3% with 1% SD.

3. Results

3.1. Hydrological context

The dry season sampling campaign occurred over one week in September 2017, during which flow rates at C5 (Sugutamu gauging station) averaged 20.5 $m^3 s^{-1}$. This is to be compared with the mean annual discharge of 24.0 $m^3 s^{-1}$ at this station for the period 1992-2000, with mean minimum and maximum daily discharge of 14 and 39 $m^3 s^{-1}$, respectively (UNESCO-IHP, 2004). Our on-the-spot measurements at C1 to C4 ranged from around 2 $m^3 s^{-1}$ at C1 and C2, to 3.5 $m^3 s^{-1}$ at C3 and 6 $m^3 s^{-1}$ at C4. The wet season sampling campaign was conducted after significant rainfall over the catchment (1,800 and 2,720 mm in Jakarta (C6) and Cisarua (C1), respectively, between November 2017 and February 2018), and flow rates at C5

229 averaged 32.6 m^3 s^{-1} during our measurements. While we were not able to gauge the river during the wet
230 sampling round, according to our field observations discharge at these four river locations was at least
231 three–five times higher than during the dry sampling round.

232 **3.2. Water and carbon stable isotopes**

233 The stable isotopic ratios of water varied significantly along the river continuum as well as between dry
234 and wet seasons (Fig. 2). A steady enrichment in ^{18}O and ^{2}H occurred from upstream to downstream
235 during the dry season, with values increasing from –7.5 to –5.4 ‰ ($\delta^{18}O$) and from –45.1 to –32.4 ‰ ($\delta^{2}H$).
236 The bulk of this increase occurred between C1 and C4, i.e. in the first 40 km of the river longitudinal profile.
237 During the wet season, all isotopic values were lower than during the dry season and varied less
238 significantly from upstream to downstream. Values were relatively steady from C1 to C5 (–8.3 to –7.8 ‰
239 for $\delta^{18}O$ and –46.4 to –43.8 ‰ for $\delta^{2}H$) and then increased only at the lowest sampled river location (–7.0
240 ‰ for $\delta^{18}O$ and –37.8 ‰ for $\delta^{2}H$ at C6). Sources also had highly variable isotopic ratios, with again a much
241 larger variability and generally more enriched signatures during the dry *vs* wet season. The most depleted
242 source was the headwater spring GW1 ($\delta^{18}O$ –7.7 and –8.2 ‰ for dry and wet campaigns, respectively),
243 while the most enriched source was the septic water from BLK1 ($\delta^{18}O$ –5.2 and –6.5 ‰ for dry and wet
244 campaigns, respectively).

245 The $\delta^{13}C_{DIC}$ values in river water were highly variable during the dry season, with ratios ranging between
246 –17.7 and –9.5 ‰ (Fig. 2). Values remained within a lower range (–18 to –14 ‰) from C1 to C4 despite a
247 noticeable peak around C2-C3, but then increased dramatically downstream (–12 to –9 ‰; C5-C6).
248 Measured $\delta^{13}C_{DIC}$ were more uniform along the river profile during the wet season, except for the
249 relatively more enriched values between C2 and C4 (range –15.5 to –14.5 ‰), similar to the peak also
250 observed during the dry season at these same sites. Contrasting to the dry season, waters in the
251 downstream reaches were as depleted as upstream waters ($\delta^{13}C_{DIC}$ around –17 ‰ for C1, C5 and C6). The

252 $\delta^{13}C_{DIC}$ signatures of sources spanned a broader range than those of river water. Of consideration were
253 the highly depleted values for the two spring sources (median –21.7 ‰). The most enriched source was
254 the septic water collected at BLK1 (+1.9 ‰), although this value should be taken with caution as it is our
255 only measurement for this source. There were also large differences within single source types (e.g. GRY1
256 –18.8 ‰ *vs* GRY2 –22.2 ‰) and between seasons (e.g. the median values for the AGR source were –16.4
257 and –19.4 ‰ for the dry and wet sampling campaigns, respectively). Lastly, the $\delta^{13}C_{DIC}$ values consistently
258 decreased from dry to wet seasons across all sources.

259 **3.3. Water chemistry**

260 In the following, we discuss only a selection of the full set of parameters analysed (full dataset available
261 in Supplementary Table S1). Electrical conductivity (EC) steadily increased from C1 to C6 during the dry
262 season (from 450 to 890 µS cm^{-1}) but was lower and less variable during the wet season (from 90 to 140
263 µS cm^{-1}; Fig. 3), similar to the pattern observed for the water isotopic ratios (Fig. 2). Dissolved oxygen (DO)
264 in the river was considerably lower during the dry (3.0–4.4 mg L^{-1}) relative to the wet season (7.3–7.9 mg
265 L^{-1}), with the exception of C6 (3.6 mg L^{-1}; Fig. 3). Both EC and DO were highly variable across sources, but
266 values were highest (electrical conductivity; median 1630 µS cm^{-1}) and lowest (oxygen; median 1.0 mg L^{-1}) in the BLK and GRY samples. Measurements from the GW and AGR sources, in contrast, reflected more
268 closely the range of values found in the river.

269 There was no apparent spatial trend in the Ciliwung River for most chemical elements (Al, Fe, Zn), with
270 large fluctuations from site to site, although concentrations in the downstream reaches (C5 and C6)
271 tended to be higher for most elements and for both seasons, while peaks at C3 were observed during the
272 wet season (Supplementary Table S1). Large seasonal differences were observed that affected elements
273 in different ways. Si and Pb concentrations in the river were 4-10 times greater in the dry season, while
274 Fe and Al concentrations were lower in the dry season, up to 2-20 times, and Zn was similar for both

seasons (Fig. 3). Importantly, a number of elements covaried in the river during both sampling campaign, with Fe and Al significantly correlated for both seasons. Additionally, EC and the negative of DO were correlated with Fe and Al in the dry season, and with Fe in the wet season (Table S2). The concentrations of chemical elements measured in sources spanned broad ranges, with the highest concentrations invariably found in the BLK samples for both seasons and all elements. Si and Zn concentrations were also above river concentrations for GRY samples, while concentrations from the GW and AGR sources were always within the range of values measured in the river (Fig. 3; Supplementary Table S1).

3.4. Bacterial community composition

The taxa bar plot (Fig. 4) shows that bacterial families differed between seasons with more *Pseudomonacaea* and order Bacillales in the dry season, while the wet season had more *Flavobacteriacea* and *Comamonadaceae*. The three most abundant families at C1 were *Moraxellaceae*, *Flavobacteriaceae* and *Comamonadaceae* for both dry and wet sampling campaigns, whereas from C2 to C6, the dominant taxa were more variable between the two seasons. Of note is the consistent decreasing trend of *Comamonadaceae* and *Flavobacteriaceae* from C1 to C6 during the wet season (Fig. 4). In the dry season, river water at C1, C3, C4 and C6 contained various Firmicutes, known faecal indicators or mammalian gut colonizers including *Christensenellaceae, Clostridiaceae, Veillonellaceae, Lachnospiraceae, Enterococcaceae* and *Ruminococcaceae*. By contrast, in the wet season, mean counts of these taxa were generally lower, especially for C3 and C6, with the exception of *Veillonellaceae* and *Ruminococcaceae*. Taxa also differed between the sources and river sites, with the exception of the two AGR sources which had a more similar bacterial composition to the river system. There were large differences within source categories, particularly for the GRY and BLK sources and for both seasons. The two AGR sources, in contrast, had similar bacterial communities and abundance. The GRY and BLK sources had unique taxa that did not resemble any other source or river location. GRY samples contained *Syngergistaceae*,

298 *Lactobacillaceae* and *Enterobacteriaceae*, while BLK samples contained *Methanosaetaceae* but also
299 various Firmicutes including the taxa listed above.

300 The above results were also confirmed by the PERMANOVA hypothesis testing which showed that the
301 bacterial communities significantly differed between the wet and dry seasons, with only an average 10.7%
302 similarity between seasons (Pseudo-F (1,37) = 2.7, P = 0.002). Furthermore, the bacterial communities
303 significantly differed between the sources and river locations (Pseudo-F (1,37) = 3.3, P = 0.001) and the
304 impact of location upon the microbiota also changed between seasons (Pseudo-F (24,37) = 10.0, P =
305 0.001). Volume was also associated with changes in the bacterial community (Pseudo-F (1,36) = 2.2, P =
306 0.001) but did not alter the significance of the above results. The average similarity in the microbiota
307 between the sources and river locations was lower in the dry than the wet season (11.9 % and 18.1 %
308 similarity, respectively). Riverine bacterial communities were more spatially variable along the upstream–
309 downstream continuum in the dry than the wet season, with no clustering of riverine communities
310 according to location in the wet season (Fig. 5). For both seasons, riverine communities were similar to
311 the two AGR sources and to one spring water source (GW2), but different from the communities
312 encountered in GRY and BLK sources, with the exception of C6 in the dry season. There was also limited
313 similarity between the headwater spring bacterial composition (GW1) and the riverine communities, even
314 those present in the upstream section of the river (C1 and C2; Fig. 5).

315 A distance based linear model showed that for the dry season EC, phosphate and DO explained most of
316 the bacterial composition (marginal results 14-15 % each, P=0.001) while for the wet season, these were
317 Si, phosphate and DO (18-22 % each, P=0.001). The first two axes of the corresponding dbRDA ordination
318 explained 28.8 % and 41.8 % of the total variation in the bacterial community in the dry and wet season
319 respectively (Supplementary Fig. S1). As indicated by the dbRDA axes, in the wet season, the chemistry
320 explained a larger proportion of the variability in bacterial communities, with multivariable analysis
321 indicating that Si was positively correlated (22 %) with BLK source communities. Regardless of the season,

communities from BLK sources were correlated with low oxygen and high electrical conductivity, phosphate, $\delta^{13}C$ and $\delta^{18}O$. Conversely, bacterial communities from the river, AGR and GW2 sources were associated with high oxygen and lower isotopic ratios, nutrient and electrical conductivity. Although trends were less apparent for the GRY source communities, they were relatively well correlated with lower pH, oxygen and Pb concentrations (Supplementary Fig. S1).

3.5. Bayesian microbial source tracking

We used a Bayesian source tracking approach (Knights *et al.*, 2011) to identify the relative contributions of each source for each river location (Fig. 6). We identified a significant domestic contamination at the upstream site (C1), for both seasons, with the GRY source accounting for about 8% and 26% of the riverine communities found at C1 for the dry and wet seasons, respectively. By contrast, the contribution from GRY sources was not detected in the downstream river sites, except for C3 and C5 in the dry (≈6 and 4%), C4 in the wet (≈16%), and C6 in both seasons (≈3-5%). Very low contributions from BLK sources were detected in the riverine samples, with slightly higher proportions at C6 and C3 during the dry season (both < 1%; RSD < 0.7), and C1 and C4 during the wet season (both < 1%; RSD < 0.6). We also note a large contribution of agricultural drainage to all river sites, but particularly to C2 and C3, and in both seasons. In the wet season, AGR sources were dominant in the river from upstream to downstream (up to 70% of river flow at C2 and C3). Lastly, the Bayesian modelling suggests a low contribution from the upstream spring GW1, even along the upstream reaches of the river, whereas GW2 contribution was prevalent during the wet season, with increasing contributions from upstream (starting at C2; 8%) to downstream (C6; 41%).

4. Discussion

4.1 Water sources and dominant flow pathways

Dry season

Distinct isotopic and elemental signatures were measured for the dry and wet seasons, which can be linked to specific hydrological processes. In the dry season, $\delta^{2}H$ and $\delta^{18}O$ in the river became gradually enriched from upstream to downstream (Fig. 2), likely the effect of progressive depletion in heavy isotopes upon orographic uplift or air masses (> 900 m elevation gradient between C6 and C1). This altitude effect is typical in high-gradient tropical terrains (e.g. Gonfiantini *et al.*, 2001), and its occurrence in the Ciliwung River suggests a tight coupling between locally recharged groundwater sources and the river during the dry season. It is interesting to note that the water isotopic signatures that we measured in sources closely reflected the elevation where each source was sampled. For instance, GW1 (spring draining volcanic headwaters) was highly depleted in the heavier isotopes (similar to C1), consistent with its high elevation within the catchment. Both AGR samples and GW2 had a water isotopic signature similar to that of C2 and C3, which also matches the elevation range where these samples were collected (Fig. 2).

The high degree of resemblance between water isotopic ratios measured at GW1 and C1 further indicates that during the dry season, the bulk of upstream river flow originated from springs located in the headwaters. This is supported by a largely ^{13}C-depleted signature at C1 (Fig. 2), indicative of dissolved carbon originating from a mostly C3 (tree) photosynthetic pathway (Farquhar *et al.*, 1989). Waters infiltrating through the headwater forested areas would have accumulated dissolved carbon of similar signature. Another line of evidence for this close relationship between local sources of subsurface waters and the upstream river were the Si concentrations (Fig. 3). Very high Si concentrations at GW1 and C1 likely resulted from silicate weathering of the breccia and lava deposits, as often reported in high-gradient, high rainfall settings (e.g. Calmels *et al.*, 2011). But the influence of these headwater springs on river flow was restricted to the upstream reaches only. Si concentrations then decreased downstream due to the dilution with other groundwater sources, which illustrates the negligible role of upstream inputs in delivering downstream river flow. This gradual addition of locally-sourced waters is also demonstrated by

368 the large discharge differential from upstream to downstream (ten-fold increase in flow rate between C1
369 and C5 in our dry season sampling campaign).

370 *Wet season*

371 During the wet season, a different range of processes controlled water movement within the catchment.
372 The $\delta^{2}H$ and $\delta^{18}O$ ratios in the Ciliwung River were less variable and more depleted than in the dry season,
373 apart from values at C6 (Fig. 2), which suggests the overriding contribution of a single water source. We
374 interpret this as the rapid replenishment of highly productive breccia and lava deposit aquifers in the
375 upper catchment (Lubis *et al*., 2008; Irawan *et al*., 2015), which then contributed substantial river flow to
376 downstream areas via both surface and subsurface pathways. It is likely that these upstream water stores
377 also recharged downstream alluvial deposits, which in turn contributed water to river flow all along the
378 river profile, with little additional inflow from locally-sourced groundwaters. The mechanism we describe
379 here is akin to the concept of mountain block recharge, common in high-gradient areas underlain by
380 fractured bedrock (e.g. Wilson and Guan, 2004; Ajami et al., 2011). Additional lowland water sources may
381 have contributed to river flow at C6, though, as suggested by the different isotopic composition for this
382 downstream site (Fig. 2).

383 Our interpretation of the hydrological response of the Ciliwung catchment to high rainfall is corroborated
384 by the spatial variations in other tracers. Both Si and electrical conductivity showed little variations along
385 the river continuum as compared to the dry season (Fig. 3), indicating a mostly upstream water source to
386 river flow. Si in particular was transported downstream with minimal dilution from other non-igneous
387 sources. The $\delta^{13}C_{DIC}$ values also preserved relatively depleted signatures all along the longitudinal profile
388 (–17.5 to –14.5 ‰), confirming again the rapid transfer of upstream, ^{13}C-depleted water sources to
389 downstream areas. Rather than the addition of other water sources, we think that the discrepancy in
390 $\delta^{13}C_{DIC}$ between GW1 (–25.1 ‰) and our river samples may be explained by turbulent flow, which is known
391 to cause preferential evasion of the lighter ^{12}C (e.g. Doctor *et al.*, 2008).

392 **4.2 Delineating hotspots of contamination**

393 As predicted, a major contamination hotspot in the downstream, highly urbanised reaches of the Ciliwung
394 River was evident during both seasons, with particularly high levels of heavy metals at C5 and C6 (Fig. 3;
395 Supplementary Table S1). During the dry season, when flow rates were lower and the dilution by upstream
396 runoff sources was minimal, the Al/Si ratio at these two downstream sites (≈ $8x10^{-2}$) was up to 82 times
397 that measured at C1, and higher than all but the two BLK sources (4–24). This may indicate an increased
398 contribution of septic waters at C5 and C6. Another line of evidence is provided by the measured increase
399 in $\delta^{13}C_{DIC}$ values at C5 and C6 during the dry season (Fig. 2), which can be related to the highly enriched
400 $\delta^{13}C_{DIC}$ signature in the BLK source (+1.9 ‰). We consider that this ^{13}C enrichment is a result of
401 methanogenic activity in anoxic environments, typical of septic tanks (Hobson *et al.*, 1974). While the
402 isotopic enrichment in the river could also be due to the accumulation of geogenic carbon from carbonate
403 weathering, this is unlikely because the increase in $\delta^{13}C_{DIC}$ did not coincide with an increase in DIC
404 (Supplementary Table S1). The occurrence of a widespread contamination from septic waters during the
405 dry season in the downstream reaches of the river is further corroborated by the microbial data, where
406 the riverine communities present at C6 were more similar to those in the two BLK sources as compared
407 to C1–C5 (Fig. 5). Samples taken from C3, C4 and C6 contained various Firmicutes including
408 *Christensenellaceae*, *Veillonellaceae*, *Lachnospiraceae*, *Enterococcaceae* or *Ruminococcaceae*, which are
409 all well-described faecal indicators or mammalian gut colonizers (e.g. Biddle *et al.,* 2013). Earlier studies
410 in the Ciliwung River also outlined the potential for high domestic wastewater loads to enter the surface
411 network downstream, where the river banks and surrounding areas are most densely populated (e.g.
412 Phanuwan *et al.*, 2006).

413 Another hotspot of contamination in the river was further upstream, at C3—a site located within the city
414 of Bogor. During the wet season, and despite the significant addition of water from upstream sources, the
415 Fe and Al concentrations measured at C3 were two to three times higher than “background” values at C1,

C2 and C4. They were also higher than in agricultural runoff, greywater and groundwater sources, and similar to the more contaminated, downstream sites C5 and C6 (Fig. 3). Phanuwan *et al*. (2006) and Zhang et al. (2003) also reported exceptionally high metal levels in floodwaters of the Ciliwung River, and so our values at C3 were not entirely unexpected. The source of this contamination was unclear, however, as according to microbial data, the source was unlikely to be domestic wastewater (Fig. 5) as mean faecal indicator counts were generally lower, especially for C3 and C6. Interestingly, this increase in Fe and Al at C3 (but also at C5 and C6) was concomitant with an increase in the heavy metals Pb, Cu, Ni, and Zn (Supplementary Table S1). Simultaneous increases in these elements have been related to metals that are adsorbed to Fe (Tessier *et al*., 1985; Priadi *et al*., 2011) and Al oxyhydroxides (Priadi *et al*., 2012). A recent study of the Jakarta Bay sediments also mentioned the likely role of Fe and Al oxyhydroxides in transporting heavy metals through the catchment (Sindern *et al*., 2016). We argue that during the dry season, relatively low DO in the water column may have triggered the precipitation of Fe and Al, making the sediment deposited at the bottom of the river a temporary sink for these contaminants. As discharge increased during the wet season, higher turbulence led to the doubling of DO concentrations (Fig. 3), which may in turn have caused the oxidation and release of sediment-bound contaminants in the water column. While we did not directly measure the heavy metals contained in suspended sediment, it is reasonable to assume that this mechanism was responsible for the high levels of metals at C3, as reported in other settings (Radakovitch *et al*., 2008; Du Laing *et al*., 2009; Le Pape, 2012). The fact that this increase was not observed further downstream (C4) suggests that the source of metals was highly localised and within short distance upstream of this river reach, consistent with the high density of small manufacturing industries in the Bogor urban area.

4.3 The added value of Bayesian source tracking

To obtain a more quantitative understanding of contaminant source and transport in the river, we used a Bayesian approach that relies on the entire microbial community data to model contamination in the river as a mixture of the identified sources (Knights *et al.*, 2011). This approach enabled us to determine the relative microbial contributions of each source for each river location. Our results (Fig. 6) provide further understanding of contamination pathways, especially where chemical tracers did not enable the distinction between different sources or the detection of other contamination sources (greywater and agricultural runoff for instance).

An important finding of the Bayesian source tracking was the identification of a significant domestic contamination at C1 for both seasons. This contamination was not detected by other tracers, and can be explained by the high density of informal settlements along the river, even in this less urbanised, upper part of the catchment. Furthermore, based upon the chemical and nMDS analyses, we were expecting large inputs of domestic (blackwater and greywater) wastewater, particularly in the lower part of the catchment. But only small proportions of BLK sources were detected in our riverine samples according to the Bayesian source tracking, with contributions < 1%. This counterintuitive result may be attributed to inadequate sampling for septic tank sources and faecal contamination decay (e.g. Staley *et al.*, 2018), which we discuss further in the following section.

Another significant insight of this analysis was the large contribution of agricultural drainage to most river sites, but particularly to C2 and C3, and in both seasons. We were able to identify this major and consistent contribution of agricultural runoff along those reaches because our two AGR source samples were taken in the vicinity of C2 and C3, where land use mostly comprises small-scale farming. We interpret the dominance of AGR sources in the wet season as the flushing of all upstream tributaries and small drainage channels across the catchment during the wet season. More surprising was the low contribution from the upstream spring GW1 during that same (wet) sampling campaign, which we would not have predicted based on our water source tracing. One explanation may be that while transiting through agricultural

462 fields before reaching the river, upstream groundwaters may have lost some of the microbial taxa typical
463 of largely anaerobic, underground environments, and rapidly acquired instead a microbial signature
464 characteristic of agricultural runoff. The low contribution from GW1 likely also reflects the lower
465 abundance of microorganisms in upstream groundwater compared to the other sources.

466 We used the microbial communities measured in the urban spring (GW2) as a proxy for contributions of
467 urban groundwater to river flow. Interestingly, there was little contribution of this source to the river in
468 the dry season, except for C3 and C4 – which might be due to the vicinity of GW2 to these river sites. It is
469 conceivable that due to local biogeochemical processes, GW2 was not representative enough of all urban
470 groundwater. Conversely, the increasing contributions of GW2 from upstream to downstream during the
471 wet season suggest a large contribution of the urban alluvial groundwater to river flow during the wet
472 season, which is in agreement with prior studies on groundwater–river interactions in the lower Ciliwung
473 (Irawan *et al.*, 2015; Costa *et al.*, 2016) and might be related to the slightly evolving isotopic signature at
474 C5 and C6 in the wet season (Fig. 2) relative to C1–C4.

475 Overall, we believe that there is potential for such source tracking methods to not only delineate hotspots
476 of contamination in rivers, but also trace different water sources carrying different bacterial taxa (e.g.
477 Sugiyama *et al.*, 2018).

478 **4.4 Limitations associated with source selection**

479 Unravelling the complexity of contaminant transport through river systems requires multiple approaches
480 and the application of tracers at multiple scales. We found some value in applying isotopic, chemical and
481 microbial tools together to trace the movement of water and contaminants. However, also evident in Fig.
482 6 is the large proportion of unknown sources making up the riverine microbiota, meaning that the sources
483 we sampled did not explain the whole range of microbial sequence variants found in the river. This can
484 stem from (i) missing source(s) and/or (ii) a lack of representativeness of sampled source(s) and/or (iii)

485 the non-conservative properties of microbial communities within each source. In the following we discuss
486 these three potential shortcomings.

487 *Did we miss a source?*

488 A substantial increase in Fe and Al occurred at C3, C5 and C6 from the dry to the wet season, as also
489 observed by Phanuwan *et al.* (2006) and Zhang *et al.* (2003). While these authors did not provide any
490 specific explanation for such high metal concentrations at high flow, we argue that Fe and Al might
491 originate from a source that we did not sample. First, urban stormwater runoff may be a preferred
492 pathway for contaminants, as the runoff originating from urban impervious areas has been shown to
493 effectively concentrate metals (Priadi, 2010). Alternatively, a likely source for Al and Fe would be via
494 increased exchange with the riverbed sediment. The occurrence of such mechanism would mean that
495 significant biogeochemical processes can occur within the riverine environment, the river acting as a
496 "reactor" that actively processes contaminants even during the wet season. We also note that industrial
497 sources of contamination may be an important source that was not sampled, especially for heavy metal
498 loads.

499 *How representative were our sources?*

500 Our study design consisted of two sources per source type (e.g. GRY1 and GRY2 for greywater). Both
501 chemical (Fig. 3) and microbiological (Fig. 4) tracers showed varying degrees of variability between single
502 source types, for instance, the bacterial composition of the two greywater sources GRY1 and GRY2
503 differed considerably (Fig. 4). This suggests that using two sources per source type may not have been
504 sufficient to encapsulate all the chemical and microbiological variability within each source type. In
505 contrast to greywater, chemical elements and microbiota were very similar between AGR1 and AGR2,
506 giving at first glance more confidence in our sampling design. However, these sources were likely to also
507 comprise upstream water contributions and domestic inputs, therefore encompassing more sources than

pure runoff from agricultural fields. A source overlap of this kind could also potentially explain why the Bayesian source tracker did not model a significant septic water contribution into the river, i.e. microbes of septic origin also occurred in other sources sampled such as runoff.

Our microbial analyses were based on triplicate samples per source and season. Triplicate sampling may have failed to capture the variability in microbial taxa of each source. In particular, spatial or temporal variability in the microbial community in the upstream groundwater source might have contributed to the low observed microbial contributions of GW1 to the upstream river site. It is indeed well known that underground microbial communities can be highly spatially heterogeneous (Griebler and Lueders, 2009) and a function of hosting rock formations and redox conditions (Zhang *et al.*, 2017; Ma *et al.*, 2019), and so we can assume that our triplicate sampling at GW1 did not capture all this heterogeneity.

More generally, there are limitations related to the use of Bayesian source tracking. Staley *et al*. (2018) raise a number of points such as the need for sufficient individual samples per source, but also the need for locally sampled sources when it comes to faecal contamination tracking, without which the technique will not yield satisfactory results. In future research, and as noted by Jardé *et al.* (2018), it will be necessary to ensure adequate replication of each source in order to capture all – and only – the variability inherent to each source.

How conservative were our sources?

Lim *et al*. (2017) make the point that fingerprinting microbial contaminants poses additional challenges as compared to chemical contaminants, because microbial tracers are often less stable than chemicals once in the water column. The microbial composition of domestic wastewater, and blackwater in particular, might change rapidly between the source (i.e. our sampling point) and the points of discharge into the river. For instance, methanogenic bacteria developing in anoxic environments may rapidly disappear before contaminants have reached the river. Staley *et al*. (2018) argue that the likely decay of

faecal contamination once in the river system may affect the predictive capacity of source tracking techniques. The same mechanism may occur at spring outflow points, with sudden changes in physicochemical conditions potentially leading to sudden changes in water-borne microbial communities. Another point is the natural assimilation (self-purification) processes that can take place within the river itself, as illustrated by the large site-to-site variations in river communities (Figs 4 and 6), particularly during the dry season when flow was lower and water residence times were higher. We know that rivers are reactors that actively process nutrients and contaminants (e.g. Karrasch *et al.*, 2006; Raymond *et al*., 2016), and since many bacterial indicators are inherently less stable, using them at large, catchment scales remains a challenge.

5. Conclusions

In this study, we confirmed the added value of microbial DNA barcoding to document contaminant behaviour in urban rivers of the tropics. Despite some identified limitations, the technique proved useful to delineate domestic contamination in the upper reaches of the Ciliwung River, where the contamination risk is usually considered low. Microbial DNA analyses also showed potential for the detection of other water sources (agricultural runoff, groundwater) across the catchment.

We did not observe a clear spatial trend in contamination sources, from agricultural upstream to domestic downstream, as first expected (H1). While agricultural runoff was prevalent in the middle reaches of the river during the wet season, we also detected signs of domestic contamination in upstream reaches. Rather than a large-scale trend across the catchment, we found the spatial variability in contaminants to be primarily controlled by local variations in land use and, to a lower extent, flow regime.

Flow regime was found to be an important driver of contaminant delivery to the river (H2). Under low flow conditions, the relatively low hydrological connectivity meant that localised sources of contamination

prevailed, with a particularly large domestic wastewater fingerprint in the downstream, highly urbanised reaches. Under high flow conditions, however, the transfer of large quantities of upstream water did not always cause the dilution of contaminants, since high turbulence was assumed to locally trigger the release of heavy metals. Delineating the source of these contaminants with more confidence will require additional work, such as focusing on a single river reach and higher-resolution spatial and temporal coverage of sources.

Acknowledgements

This research was funded by grants from the Australian National University – Indonesia Project, Charles Darwin University (EHSE Faculty, SPG17/#06) and Universitas Indonesia (PITTA grant 2408/UN2.R3.1/HKP.05.00/2018). We are thankful to landowners for access to sampling sites. Nadine Hafiza and UI students are thanked for fieldwork and laboratory assistance; Pak Bagyo for flow gauging measurements; Dionisia Lambrinidis (ECMU) for isotopic analyses; and Faris Zulkarnain for compiling Figure 1. Discharge and rainfall data were obtained from BPBD DKI Jakarta. We also would like to thank Jonathan Wynn for discussing our carbon isotope results, as well as two anonymous reviewers for their constructive feedback.

Author contributions

CD, CRP, MK, DRM and KSG designed the research. AA, CD and DRM conducted fieldwork. AA, AMR and CD conducted lab work. MK, AMR, CRP, CD and DRM analysed the data. The manuscript was written by CD and CRP, and edited by MK, AMR and KSG.

572 **References**

573 Ajami, H., Troch, P. A., Maddock, T., Meixner, T., & Eastoe, C. (2011). Quantifying mountain block recharge
574 by means of catchment-scale storage–discharge relationships. Water Resources Research, 47,
575 W04504.

576 Ayrault, S., Roy-Barman, M., Le Cloarec, M.-F., Priadi, C. R., Bonté, P., & Göpel, C. (2012). Lead
577 contamination of the Seine River, France: Geochemical implications of a historical perspective.
578 Chemosphere, 87, 902–910.

579 Bass, A. M., Bird, M. I., Munksgaard, N. C., & Wurster, C. M. (2012). ISO-CADICA: Isotopic – continuous,
580 automated dissolved inorganic carbon analyser. Rapid Communications in Mass Spectrometry, 26,
581 639–644.

582 Ben Maamar, S., Aquilina, L., Quaiser, A., Pauwels, H., Michon-Coudouel, S., Vergnaud-Ayraud, V., …
583 Dufresne, A. (2015). Groundwater isolation governs chemistry and microbial community structure
584 along hydrologic flowpaths. Frontiers in Microbiology, 6, 1457.

585 Biddle, A., Stewart, L., Blanchard, J., & Leschine, S. (2013). Untangling the genetic basis of fibrolytic
586 specialization by Lachnospiraceae and Ruminococcaceae in diverse gut communities. Diversity, 5,
587 627–640.

588 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2:
589 high-resolution sample inference from Illumina amplicon data. Nature Methods, 13, 581–583.

590 Calmels, D., Galy, A., Hovius, N., Bickle, M., West, A. J., Chen, M. C., & Chapman, H. (2011). Contribution
591 of deep groundwater to the weathering budget in a rapidly eroding mountain belt, Taiwan. Earth and
592 Planetary Science Letters, 303, 48–58.

593 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R.
594 (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7,
595 335–336.

596 Capps, K. A., Bentsen, C. N., & Ramírez, A. (2016). Poverty, urbanization, and environmental degradation:
597 urban streams in the developing world. Freshwater Science, 35, 429–435.

598 Costa, D., Burlando, P., & Priadi, C. (2016). The importance of integrated solutions to flooding and water
599 quality problems in the tropical megacity of Jakarta. Sustainable Cities and Society, 20, 199–209.

600 Derrien, M., Jardé, E., Gruau, G., & Pierson-Wickmann, A.-C. (2011). Extreme variability of steroid profiles
601 in cow feces and pig slurries at the regional scale: implications for the use of steroids to specify fecal
602 pollution sources in waters. Journal of Agricultural and Food Chemistry, 59, 7294–7302.

603 Doctor, D. H., Kendall, C., Sebestyen, S. D., Shanley, J. B., Ohte, N., & Boyer, E. W. (2008). Carbon isotope
604 fractionation of dissolved inorganic carbon (DIC) due to outgassing of carbon dioxide from a
605 headwater stream. Hydrological Processes, 22, 2410–2423.

606 Du Laing, G., Meers, E., Dewispelaere, M., Vandecasteele, B., Rinklebe, J., Tack, F. M. G., & Verloo, M. G.
607 (2009). Heavy metal mobility in intertidal sediments of the Scheldt estuary: field monitoring. Science
608 of the Total Environment, 407, 2919–2930.

609 Duvert, C., Raiber, M., Owen, D. D. R., Cendón, D. I., Batiot-Guilhe, C., & Cox, M. E. (2015). Hydrochemical
610 processes in a shallow coal seam gas aquifer and its overlying stream–alluvial system: implications
611 for recharge and inter-aquifer connectivity. Applied Geochemistry, 61, 146–159.

612 Farquhar, G. D., Ehleringer, J. R., & Hubick, K. T. (1989). Carbon isotope discrimination and photosynthesis.
613 Annual Review in Plant Physiology and Plant Molecular Biology, 40, 503–537.

614 Febria, C. M., Hosen, J. D., Crump, B. C., Palmer, M. A., & Williams, D. D. (2015). Microbial responses to
615 changes in flow status in temporary headwater streams: A cross-system comparison. Frontiers in
616 Microbiology, 6, 522.

617 Fontes, J. C. (1980). Environmental isotopes in groundwater hydrology. In: Handbook of environmental
618 isotope geochemistry. Vol. 1.

619 Forbes, 2015. The World's Fastest-growing Megacities. URL:
620 http://www.forbes.com/pictures/edgl45fdfe/no-10-Jakarta-Indonesia/.

621 García-Armisen, T., İnceoğlu, Ö., Ouattara, N. K., Anzil, A., Verbanck, M. A., Brion, N., & Servais, P. (2014).
622 Seasonal Variations and Resilience of Bacterial Communities in a Sewage Polluted Urban River. PLoS
623 ONE, 9, e92579.

624 Gonfiantini, R., Roche, M. A., Olivry, J. C., Fontes, J. C., & Zuppi, G. M. (2001). The altitude effect on the
625 isotopic composition of tropical rains. Chemical Geology, 181, 147–167.

626 Griebler, C., & Lueders, T. (2009). Microbial biodiversity in groundwater ecosystems. Freshwater Biology,
627 54, 649–677.

628 Guan, Y. F., Wang, J. Z., Ni, H. G., & Zeng, E. Y. (2009). Organochlorine pesticides and polychlorinated
629 biphenyls in riverine runoff of the Pearl River Delta, China: Assessment of mass loading, input source
630 and environmental fate. Environmental Pollution, 157, 618–624.

631 Heeren, D. M., Miller, R. B., Fox, G. A., Storm, D. E., Halihan, T., & Penn, C. J. (2010). Preferential flow
632 effects on subsurface contaminant transport in alluvial floodplains. Transactions of the ASABE, 53,
633 127–136.

634 Hobson, P. N., Bousfield, S., Summers, R., & Kirsch, E. J. (1974). Anaerobic digestion of organic matter. C
635 R C Critical Reviews in Environmental Control, 4, 131–191.

636 Hooper, R. P., & Shoemaker, C. A. (1986). A Comparison of Chemical and Isotopic Hydrograph Separation.
637 Water Resources Research, 22, 1444–1454.

638 Hosen, J. D., Febria, C. M., Crump, B. C., & Palmer, M. A. (2017). Watershed urbanization linked to
639 differences in stream bacterial community composition. Frontiers in Microbiology, 8, 1452.

640 Ibekwe, A. M., Ma, J., & Murinda, S. E. (2016). Bacterial community composition and structure in an Urban
641 River impacted by different pollutant sources. Science of The Total Environment, 566, 1176–1185.

642 Irawan, D. E., Silaen, H., Sumintadireja, P., Lubis, R. F., Brahmantyo, B., & Puradimaja, D. J. (2015).
643 Groundwater--surface water interactions of Ciliwung River streams, segment Bogor--Jakarta,
644 Indonesia. Environmental Earth Sciences, 73, 1295–1302.

645 Jardé, E., Jeanneau, L., Harrault, L., Quenot, E., Solecki, O., Petitjean, P., … Gourmelon, M. (2018).
646 Application of a microbial source tracking based on bacterial and chemical markers in headwater and
647 coastal catchments. Science of the Total Environment, 610, 55–63.

Kabiri, L., Alum, A., Rock, C., McLain, J. E., & Abbaszadegan, M. (2016). A tool box strategy using Bacteroides genetic markers to differentiate human from non-human sources of fecal contamination in natural water. Science of The Total Environment, 572, 897–905.

Kaestli, M., Skillington, A., Kennedy, K., Majid, M., Williams, D., McGuinness, K., … Gibb, K. (2017). Spatial and temporal microbial patterns in a tropical macrotidal estuary subject to urbanization. Frontiers in Microbiology, 8, 1313.

Karrasch, B., Parra, O., Cid, H., Mehrens, M., Pacheco, P., Urrutia, R., … Zaror, C. (2006). Effects of pulp and paper mill effluents on the microplankton and microbial self-purification capabilities of the Biobío River, Chile. Science of the Total Environment, 359, 194–208.

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W. T., Wolfe, A. J., … McWeeney, S. (2018). Controlling for contaminants in low biomass 16S rRNA gene sequencing experiments. BioRxiv, pre-print.

Kerstens, S. M., Leusbrock, I., & Zeeman, G. (2015). Feasibility analysis of wastewater and solid waste systems for application in Indonesia. Science of The Total Environment, 530, 53–65.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., … Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. Nature Methods, 8, 761–763.

Le Pape, P. (2012). Etude de la dynamique, des sources et de la spéciation des éléments traces dans le bassin versant de l'Orge (Essonne, France). PhD thesis, Université Paris Sud – Paris XI.

Lim, F. Y., Ong, S. L., & Hu, J. (2017). Recent Advances in the Use of Chemical Markers for Tracing Wastewater Contamination in Aquatic Environment: A Review. Water, 9, 143.

Lubis, R. F., Sakura, Y., & Delinom, R. (2008). Groundwater recharge and discharge processes in the Jakarta groundwater basin, Indonesia. Hydrogeology Journal, 16, 927–938.

Ma, J., Liu, H., Tong, L., Wang, Y., Chen, R., Liu, S., Zhao, L., Li, Z., & Cai, L. (2019). Relationships between microbial communities and groundwater chemistry in two pristine confined groundwater aquifers in central China. Hydrological Processes, in press.

Malakoff, D. (2002). Microbiologists on the Trail of Polluting Bacteria. Science, 295, 2352–2353.

McDonald, R. I., Green, P., Balk, D., Fekete, B. M., Revenga, C., Todd, M., & Montgomery, M. (2011). Urban growth, climate change, and freshwater availability. Proceedings of the National Academy of Sciences, 108, 6312–6317.

Mouri, G., Takizawa, S., & Oki, T. (2011). Spatial and temporal variation in nutrient parameters in stream water in a rural-urban catchment, Shikoku, Japan: Effects of land cover and human impact. Journal of Environmental Management, 92, 1837–1848.

Munksgaard, N. C., Wurster, C. M., & Bird, M. I. (2011). Continuous analysis of δ18O and δD values of water by diffusion sampling cavity ring-down spectrometry: A novel sampling device for unattended field monitoring of precipitation, ground and surface waters. Rapid Communications in Mass Spectrometry, 25, 3706–3712.

Newton, R. J., Bootsma, M. J., Morrison, H. G., Sogin, M. L., & McLellan, S. L. (2013). A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microbial Ecology, 65, 1011–1023.

687 Palupi, K., Sumengen, S., Inswiasri, S., Agustina, L., Nunik, S. A., Sunarya, W., & Quraisyn, A. (1995). River
688 water quality study in the vicinity of Jakarta. Water Science and Technology, 31, 17–25.

689 Parkinson, J., & Mark, O. (2005). Urban stormwater management in developing countries. IWA Publishing,
690 London (England).

691 Phanuwan, C., Takizawa, S., Oguma, K., Katayama, H., Yunika, A., & Ohgaki, S. (2006). Monitoring of human
692 enteric viruses and coliform bacteria in waters after urban flood in Jakarta, Indonesia. Water Science
693 and Technology, 54, 203–210.

694 Pongmala, K., Autixier, L., Madoux-Humery, A.-S., Fuamba, M., Galarneau, M., Sauvé, S., … Dorner, S.
695 (2015). Modelling total suspended solids, E. coli and carbamazepine, a tracer of wastewater
696 contamination from combined sewer overflows. Journal of Hydrology, 531, 830–839.

697 Priadi, C. R. (2010). Caractérisation des phases porteuses: Métaux particulaires en Seine. PhD thesis,
698 Université Paris Sud – Paris XI.

699 Priadi, C. R., Suleeman, E., Darmajanti, L., Novriaty, S., Suwartha, N., Resnawati, R., … Tjahjono, T. (2017).
700 Water Recycling Opportunity in the Business Sectors of Greater Jakarta, Indonesia. International
701 Journal of Technology, 8, 1031–1039.

702 Priadi, C., Ayrault, S., Pacini, P., & Bonté, P. (2011). Urbanization impact on metals mobility in riverine
703 suspended sediment: Role of metal oxides. International Journal of Environmental Science and
704 Technology, 8, 1–18.

705 Priadi, C., Le Pape, P., Morin, G., Ayrault, S., Maillot, F., Juillot, F., Hochreutener, R., Llorens, I., Testemale,
706 D., Proux, O., & Brown, G. E. (2012). X-ray absorption fine structure evidence for amorphous zinc
707 sulfide as a major zinc species in suspended matter from the Seine River downstream of Paris, Ile-de-
708 France, France. Environmental Science and Technology, 46, 3712–3720.

709 Radakovitch, O., Roussiez, V., Ollivier, P., Ludwig, W., Grenz, C., & Probst, J.-L. (2008). Input of particulate
710 heavy metals from rivers and associated sedimentary deposits on the Gulf of Lion continental shelf.
711 Estuarine, Coastal and Shelf Science, 77, 285–295.

712 Raymond, P. A., Saiers, J. E., & Sobczak, W. V. (2016). Hydrological and biogeochemical controls on
713 watershed dissolved organic matter transport: pulse-shunt concept. Ecology, 97, 5–16.

714 Rochelle-Newall, E. J., Ribolzi, O., Viguier, M., Thammahacksa, C., Silvera, N., Latsachack, K., … Soulileuth,
715 B. (2016). Effect of land use and hydrological processes on Escherichia coli concentrations in streams
716 of tropical, humid headwater catchments. Scientific Reports, 6, 32974.

717 Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014).
718 Reagent and laboratory contamination can critically impact sequence-based microbiome analyses.
719 BMC Biology, 12, 87.

720 Schulte, P., van Geldern, R., Freitag, H., Karim, A., Négrel, P., Petelet-Giraud, E., … Barth, J. A. C. (2011).
721 Applications of stable water and carbon isotopes in watershed research: Weathering, carbon cycling,
722 and water balances. Earth-Science Reviews, 109, 20–31.

723 Silveira, A. L. L. da. (2002). Problems of modern urban drainage in developing countries. Water Science
724 and Technology, 45, 31–40.

Sindern, S., Tremöhlen, M., Gronen, L., Siregar, T. H., Ariyani, F., & Irianto, H. E. (2016). Heavy metals in river and coast sediments of the Jakarta Bay region (Indonesia) — Geogenic versus anthropogenic sources. Marine Pollution Bulletin, 110, 624–633.

Smith, H. G., & Blake, W. H. (2014). Sediment fingerprinting in agricultural catchments: A critical re-examination of source discrimination and data corrections. Geomorphology, 204, 177–191.

Staley, C., Kaiser, T., Lobos, A., Ahmed, W., Harwood, V. J., Brown, C. M., & Sadowsky, M. J. (2018). Application of SourceTracker for accurate identification of fecal pollution in recreational freshwater: a double-blinded study. Environmental Science and Technology, 52, 4207–4217.

Sugiyama, A., Masuda, S., Nagaosa, K., Tsujimura, M., & Kato, K. (2018). Tracking the direct impact of rainfall on groundwater at Mt. Fuji by multiple analyses including microbial DNA. Biogeosciences, 15, 721–732.

Surbeck, C. Q., Jiang, S. C., Ahn, J. H., & Grant, S. B. (2006). Flow fingerprinting fecal pollution and suspended solids in stormwater runoff from an urban coastal watershed. Environmental Science & Technology, 40, 4435–4441.

Tallar, R. Y., & Suen, J.-P. (2015). Identification of waterbody status in Indonesia by using predictive index assessment tool. International Soil and Water Conservation Research, 3, 224–238.

Tan, B. F., Ng, C., Nshimyimana, J. P., Loh, L. L., Gin, K. Y. H., & Thompson, J. R. (2015). Next-generation sequencing (NGS) for assessment of microbial water quality: Current progress, challenges, and future opportunities. Frontiers in Microbiology, 6, 1027.

Tessier, A., Rapin, F., & Carignan, R. (1985). Trace metals in oxic lake sediments: possible adsorption onto iron oxyhydroxides. Geochimica et Cosmochimica Acta, 49, 183–194.

Thuong, N. T., Yoneda, M., Ikegami, M., & Takakura, M. (2013). Source discrimination of heavy metals in sediment and water of To Lich River in Hanoi City using multivariate statistical approaches. Environmental Monitoring and Assessment, 185, 8065–8075.

Umezawa, Y., Hosono, T., Onodera, S., Siringan, F., Buapeng, S., Delinom, R., … Taniguchi, M. (2008). Sources of nitrate and ammonium contamination in groundwater under developing Asian megacities. Science of The Total Environment, 404, 361–376.

UNESCO-IHP (2004). Catalogue of rivers for southeast Asia and the Pacific. Volume V, May 2004. In Y. Tachikawa, R. James, K. Abdullah and Mohd. Nor bin Mohd. Desa (Editors), Chapter 3.11, Kali Ciliwung. 12 pp.

Unno, T., Jang, J., Han, D., Kim, J. H., Sadowsky, M. J., Kim, O. S., Chun, J., & Hur, H. G. (2010). Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. Environmental Science & Technology, 44, 7777–7782.

Wang, K., Ye, X., Zhang, H., Chen, H., Zhang, D., & Liu, L. (2016). Regional variations in the diversity and predicted metabolic potential of benthic prokaryotes in coastal northern Zhejiang, East China Sea. Scientific Reports, 6, 38709.

Wilson, J. L., & Guan, H. (2004). Mountain-block hydrology and mountain-front recharge. In J.F. Hogan, F.M. Phillips, B.R. Scanlon (Editors), Groundwater Recharge in a Desert Environment: The Southwestern United States, 113–137. American Geophysical Union, Washington D.C. (USA).

764 Yasuda, M., Yustiawati, Syawal, M. S., Sikder, M. T., Hosokawa, T., Saito, T., … Kurasaki, M. (2011). Metal
765 concentrations of river water and sediments in West Java, Indonesia. Bulletin of Environmental
766 Contamination and Toxicology, 87, 669–673.

767 Zhang, J., Ren, J. L., Liu, S. M., Zhang, Z. F., Wu, Y., Xiong, H., & Chen, H. T. (2003). Dissolved aluminum and
768 silica in the Changjiang (Yangtze River): Impact of weathering in subcontinental scale. Global
769 Biogeochemical Cycles, 17, 1077.

770 Zhang, J., Ma, T., Feng, L., Yan, Y., Abass, O. K., Wang, Z., & Cai, H. (2017). Arsenic behavior in different
771 biogeochemical zonations approximately along the groundwater flow path in Datong Basin, northern
772 China. Science of the Total Environment, 584, 458–468.

773

774 **Figure Captions**

775 Figure 1. Location of the Ciliwung River and sampling sites along the river. Note that the contour interval
776 is 100m. Map compiled from data provided by Department of Geography (UI), Geospatial Information
777 Agency (BIG), and Bureau of Meteorology, Climatology and Geo-physics (BMKG).
778
779 Figure 2. $\delta^{2}H$, $\delta^{18}O$ and $\delta^{13}C_{DIC}$ measurements during both dry and wet conditions (i) along the Ciliwung
780 River from C1 (origin of the x-axes) to C6, and (ii) within the four identified sources, the range for each
781 source corresponding to two replicate sites (groundwater (GW1 and GW2), agricultural runoff (AGR1 and
782 AGR2), greywater (GRY1 and GRY2) and blackwater (BLK1 and BLK2)). The error bars on river samples
783 correspond to analytical precision (±σ). Triangles denote measurements that go beyond the range
784 displayed on the graphs (see Supplementary Table S1 for exact values).
785
786 Figure 3. Electrical conductivity (EC), dissolved oxygen (DO), silica (Si), aluminium (Al), iron (Fe), and zinc
787 (Zn) concentrations during both dry and wet conditions (i) along the Ciliwung River from C1 (origin of the
788 x-axes) to C6, and (ii) within the four identified sources, the range for each source corresponding to two
789 replicate sites (groundwater (GW1 and GW2), agricultural runoff (AGR1 and AGR2), greywater (GRY1 and
790 GRY2) and blackwater (BLK1 and BLK2)). Note that the y-axes of the EC panel are log-scaled. Triangles
791 denote measurements beyond the range displayed on the graphs (see Supplementary Table S1 for exact
792 values and other elements).
793
794 Figure 4. Bacterial taxa at the family level (where known) for the wet and dry seasons. Labels indicate the
795 sample ID. C1-C6 = Ciliwung River samples, GW1: headwater spring; GW2: semi-urban spring; AGR1 and
796 AGR2: agricultural drainage; GRY1 and GRY2: household greywater; BLK1 and BLK2: septic water. Only
797 families with >5% abundance are shown.
798
799 Figure 5. Non-metric multidimensional scaling (nMDS) of the bacterial communities in sources and in the
800 Ciliwung River per season. 2D Stress value 0.14. Labels indicate the sample ID. C1-C6 = Ciliwung River
801 samples, GW1: headwater spring; GW2: semi-urban spring; AGR1 and AGR2: agricultural drainage; GRY1
802 and GRY2: household greywater; BLK1 and BLK2: septic water.
803
804 Figure 6. Bayesian microbial source tracking of the relative contributions of each source from C1 to C6 and
805 for both dry and wet sampling campaigns. Labels indicate the sample ID. C1-C6 = Ciliwung River samples,
806 GW1: headwater spring; GW2: semi-urban spring; AGR: agricultural drainage; GRY: household greywater;
807 BLK: septic water.

808