

Dear HESS Editor,

21/07/2020

Please find below a point-by-point response (rep) to the second round of reviews (R1b and R2b), including a description of all relevant changes made in the manuscript, and a marked-up manuscript version showing in red modifications performed on the text.

All issues raised by the reviewers were considered. These modifications were grammatical or aimed at improving technical terms.

Thank you for the time spent evaluating our article,

Best regards,
B. Cournoyer

Referee #1: Carles, Louis louis.carles@eawag.ch

Second review - comments :

R1b-1. Please write the units always in the same way (for example, x y-1) throughout the manuscript. L153. “0.4 mg mL-1” should be “0.4 mg mL⁻¹”

Rep-R1b-1: modifications were made accordingly all over the manuscript

R1b-2. L155. Remove the comma after “-80°C”

Rep-R1b-2: the comma was removed

Referee #2: Anonymous

Second review - comments :

R2b-1. Please standardize the spacing between the unit and the number throughout the text (e.g., L92 vs. L113, although different units, suggest maintaining the space throughout).

Rep-R2b-1: fixed accordingly all over the manuscript

R2b-2. L18 – Please remove the hyphen in “micro-organisms” here and throughout the text.

Rep-R2b-2: fixed accordingly all over the manuscript

R2b-3. L49 – Please provide a comma after “i.e.” here and throughout the text.

Rep-R2b-3: fixed accordingly all over the manuscript

R2b-4. L116 – “Urbanistic changes” presents a challenge to me in readability. Perhaps simply “modifications” would suffice.

Rep-R2b-4: fixed as follow, “No significant **modifications** impacting the urban watershed were recorded during the investigation.”

R2b-5. L145 – Please replace “About” with “Approximately”

Rep-R2b-5: modified accordingly

R2b-6. L146 – Please alter “Total DNAs were” to “Total DNA was”. Throughout the text, convert the usage of DNAs plural into DNA singular and alter the referencing verbs accordingly.

Rep-R2b-6: fixed accordingly all over the manuscript

R2b-7. L208 – Here, and throughout, please ensure that a decimal point rather than decimal comma is used in all numbers (0,07 to 0.07).

Rep-R2b-7: fixed accordingly all over the manuscript

R2b-8. L381 – Please add a comma before “but”

Rep-R2b-8: modified accordingly

R2b-9. L547 – Please remove the comma before “and”

Rep-R2b-9: modified accordingly

1 **Coalescence of bacterial groups originating from urban runoffs**
2 **and artificial infiltration systems among aquifer microbiomes**

3 Yannick Colin[‡], Rayan Bouchali¹, Laurence Marjolet¹, Romain Marti¹, Florian Vautrin^{1,2}, Jérémy Voisin^{1,2},
4 Emilie Bourgeois¹, Veronica Rodriguez-Nava¹, Didier Blaha¹, Thierry Winiarski², Florian Mermillod-Blondin²
5 and Benoit Cournoyer¹

6 ¹University of Lyon, UMR Ecologie Microbienne Lyon (LEM), Research Team “Bacterial Opportunistic
7 Pathogens and Environment”, Université Lyon 1, CNRS 5557, INRA 1418, VetAgro Sup, 69680 Marcy L’Etoile,
8 France.

9 ²University of Lyon, UMR Laboratoire d’Ecologie des Hydrosystèmes Naturels et Anthropisés (LEHNA),
10 Université Lyon 1, CNRS 5023, ENTPE, 69622 Villeurbanne, France.

11 [‡]Present address : Normandie Université, UNIROUEN, UNICAEN, UMR CNRS 6143, Morphodynamique
12 Continentale et Côtière, 76000 Rouen, France

13

14 **Running title: Y. Colin et al.: Urban runoff bacteria among recharged aquifer**

15 **Keywords:** Stormwater infiltration; Microbial contamination; Aquifer; Source-tracking; Biofilms

16

17 **Correspondence :** yannick.colin@univ-rouen.fr / benoit.cournoyer@vetagro-sup.fr

18 **Abstract.** The invasion of aquifer microbial communities by aboveground microorganisms, a phenomenon known
19 as community coalescence, is likely to be exacerbated in groundwaters fed by stormwater infiltration systems
20 (SIS). Here, the incidence of this increased connectivity with upslope soils and impermeabilized surfaces was
21 assessed through a meta-analysis of 16S rRNA gene libraries. Specifically, 16S rRNA V5-V6 DNA sequences
22 from free-living and attached aquifer bacteria (i.e., water and biofilm samples) were analysed upstream and
23 downstream a SIS, and compared with those from bacterial communities from watershed runoffs, and surface
24 sediments from the SIS detention and infiltration basins. Significant bacterial transfers were inferred by the
25 SourceTracker Bayesian approach, with 23 to 57% of the aquifer bacterial biofilms being composed of taxa from
26 aboveground sediments and urban runoffs. Sediments from the detention basin were found more significant
27 contributors of taxa involved in the build-up of these biofilms than soils from the infiltration basin. Inferred taxa
28 among the coalesced biofilm community were predicted to be high in hydrocarbon degraders such as *Sphingobium*
29 and *Nocardia*. The 16S rRNA-based bacterial community structure of the downstream SIS aquifer waters showed
30 lower coalescence with aboveground taxa (8 to 38%) than those of biofilms, and higher numbers of taxa predicted
31 to be involved in the N- and S-cycles. A DNA marker named *tpm* enabled the tracking of bacterial species from
32 24 genera including the *Pseudomonas*, *Aeromonas* and *Xanthomonas*, among these communities. Several *tpm*
33 sequence types were found to be shared between the aboveground and aquifer samples. Reads related to
34 *Pseudomonas* were allocated to 50 species, of which 16 were found in the aquifer samples. Several of these aquifer
35 species were found to be involved in denitrification but also hydrocarbon degradation (*P. aeruginosa*, *P. putida*,
36 and *P. fluorescens*). Some *tpm* sequence types allocated to *P. umsongensis* and *P. chengduensis* were found to be
37 enriched among the *tpm*-harboring bacteria, respectively of the aquifer biofilms, and waters. Reads related to
38 *Aeromonas* were allocated to 11 species, but only those from *A. caviae* were recovered aboveground and in the
39 aquifer samples. Some *tpm* sequence types of the *X. axonopodis* phytopathogen were recorded in higher
40 proportions among the *tpm*-harboring bacteria of the aquifer waters than in the aboveground samples. A significant
41 coalescence of microbial communities from an urban watershed with those of an aquifer was thus observed, and
42 recent aquifer biofilms were found to be significantly colonized by runoff opportunistic taxa able to use urban C-
43 sources from aboveground compartments.
44

45 1 Introduction

46 Urbanization exerts multiple pressures on natural habitats and particularly on aquatic environments (Konrad and
47 Booth, 2005; McGrane, 2016; Mejía and Moglen, 2009). The densification of urban areas, combined with the
48 conversion of agricultural and natural lands into urban land-use, led to the replacement of vegetation and open
49 fields by impervious urban structures (*i.e.*, roads, rooftops, side-walks and parking lots) (Barnes et al., 2001). These
50 impervious structures reduce the infiltration capacity of soils. They also exacerbate the speed and volume of
51 stormwater runoff that favor soil erosion, flooding events, and affect adversely natural groundwater recharge
52 processes (Booth, 1991; Shuster et al., 2005). Due to these consequences, stormwater infiltration systems (SIS) or
53 managed aquifer recharged systems (MAR) have been developed over the last decades, and are gaining more
54 interest in developed countries (Pitt et al., 1999). Such practices reduce direct stormwater discharges to surface
55 waters and alleviate water shortages (Barba et al., 2019; Dillon et al., 2008; Marsalek and Chocat, 2002). However,
56 stormwater represents a major source of nonpoint pollution, and its infiltration into the ground may have adverse
57 ecological and sanitary impacts (Chong et al., 2013; Pitt et al., 1999; Vezzaro and Mikkelsen, 2012).

58 The vadose zone of a SIS can act as a natural filter capturing pollutants (hydrocarbons and heavy metals) and
59 microorganisms washed-off by runoffs (e. g., Murphy and Ginn, 2000; Tedoldi et al., 2016). Nevertheless, the
60 effectiveness of SIS in preventing the migration of contaminants towards aquifers is not always optimal (Borchardt
61 *et al.*, 2007; Lapworth et al., 2012; Arnaud et al., 2015; Voisin et al., 2018). The filtering properties of SIS are
62 influenced by various abiotic factors such as the nature of the media (rocks, sand and other soil elements), the
63 physical properties (e. g., granulometry, hydrophobicity index, organization), and the runoff water flow velocity
64 (Lassabatere et al., 2006; Winiarski et al., 2013). These constraints will impact both the water transit time from
65 the top layers to the aquifer, but also the biological properties of these systems including the plant cover, root
66 systems, worm population and composition of microbiota (Barba et al., 2019; Bedell et al., 2013; Crites, 1985;
67 Pignoret et al., 2016). The thickness of the vadose zone was found to be one of the key parameters explaining
68 chemical transfers such as phosphate and organic-carbon sources (Voisin et al., 2018). The situation is much less
69 clear regarding the microbiological communities that flow through these systems (e. g., Barba et al., 2019; Voisin
70 et al., 2018).

71 According to the microbial community coalescence concept conceptualized by Tikhonov, (2016) and adapted
72 to riverine networks by Mansour et al. (2018), urban aquifers fed by SIS should harbor microbiota reflecting the
73 coalescence (community assemblages and selective sorting) of aboveground microbial communities with those of
74 the aquifer. Indeed, during rain events, microbial communities will be re-suspended through runoff-driven surface
75 erosion processes, favoring detachment of microorganisms from plant litter, wastes, soil, and other particles
76 (Mansour et al. 2018). These re-suspended communities will merge and generate novel assemblages. The resulting
77 community will initially match the relative contributions of the various sub-watersheds to the overall
78 microbiological complexity of the assemblages (Mansour et al. 2018). The prevailing ecological constraints among
79 the downward systems will then gradually drive this coalescence towards the most fit community structures. These
80 resulting communities might be highly efficient at degrading urban pollutants trapped among a SIS, but could also
81 disturb the ecological equilibria of the connected and more sensitive systems like those of deep aquifers as
82 suggested by Voisin et al. (2018).

83 This study explores the impact of a SIS with a thick vadose zone (> 10 m) on the coalescence of urban runoff
84 microbial communities in a connected aquifer. Two hypotheses were tested (1) highly specialized taxa (often

85 termed K-strategists e. g., Vadstein et al., 2018) of an aquifer outcompete the intrusive community members of
86 aboveground taxa; and (2) nutrient inputs from runoffs and pollutants drive changes among communities and
87 favour environmental opportunists (often termed r-strategists e. g., Vadstein et al., 2018). The targeted SIS is part
88 of a long-term experimental site (<http://www.graie.org/portail/dispositifsderecherche/othu/>) that records both
89 physico-chemical and biological properties. This SIS is connected to the eastern aquifer of Lyon (France) which
90 is fed by three low hydraulic conductivity corridors (10^{-5} – 10^{-8} m. s⁻¹) separated by moraine hills (Foulquier et al.,
91 2010). The average vadose zone thickness of the SIS is 15 m, and the delay between a rainfall event and the impact
92 on the aquifer waters was estimated at 86±11h (Voisin et al., 2018). A 16S rRNA gene meta-barcoding dataset
93 was assembled for this site, to investigate bacterial community coalescence from the top compartments into the
94 connected aquifer waters but also the biofilm communities developing on inert surfaces. This investigation was
95 also built on the hypothesis that a less significant microbial community coalescence is expected in aquifer waters
96 than biofilms. This is supported by previous reports which suggested the occurrence of transient free-living
97 bacteria among aquifers acting as a traveling seed bank (Griebler et al., 2014). For such monitorings, water grab
98 samples were previously found to give access to snapshots of the diversity found within an aquifer (Voisin et al.,
99 2018) whereas aquifer biofilms developing on artificial surfaces (clay beads) were shown to be more integrative
100 and informative of the groundwater microbiological quality (Mermillod-Blondin et al., 2019). Clay bead biofilms
101 were found to capture the most abundant aquifer taxa, and taxa that could not be detected from grab samples.
102 However, some bacterial taxa were still not detectable by this approach because of a poor ability at colonizing clay
103 beads over short time periods. A field based investigation was thus performed to further explore the relative
104 contributions of a set of sources such as runoffs and urban soils on the observed biofilm assemblages recovered
105 from an aquifer. A Bayesian methodology, named SourceTracker (Knights et al., 2011), was used to investigate
106 community coalescence from 16S rRNA gene – based DNA meta-barcoding datasets. Complementary datasets
107 were then assembled from an additional DNA marker named *tpm* (encoding EC:2.1.1.67 which catalyzes the
108 methylation of thiopurine drugs) (Favre-Bonté et al., 2005). This genetic marker enables finer taxonomic
109 allocations down to the species level to explore the coalescence of a set of waterborne bacterial species and sub-
110 species, including plant and human pathogens, within the aquifer microbial community.

111 2 Material and Methods

112 2.1 Experimental site

113 The Chassieu urban catchment is located in the suburbs of Lyon (France). It has a surface of 185 ha and hosts
114 mainly industrial and commercial activities (*i.e.* wholesaling, recovery and waste management, metal surface
115 treatment, car wash and repair services). The imperviousness coefficient of the catchment area is approximately
116 75%. No significant **modifications** impacting the urban watershed were recorded during the investigation.
117 Stormwater and dry weather flows from industrial activities are drained by a network separated from the sewer.
118 This network transfers waters into the Django-R SIS, which is a part of the OTHU long term experimental
119 observatory dedicated to urban waters (<http://www.graie.org/othu/>). This SIS contains an open and dry detention
120 basin (DB) (32,000 m³) built on a concrete slab, with edges impermeabilized by a thick plastic lining. This DB
121 allows a settling of coarse and medium size particles, resulting in sedimentary deposits which favor plant cover
122 development. The DB water content is delivered within 24h into an infiltration basin (IB) (61,000 m³), which
123 favors the recharge of the connected aquifer (AQ). This infiltration basin had a vadose zone of about 11 m during

124 the investigation, and its geology, hydrology, ecology and pollution levels **had** been previously investigated (e. g.,
125 Barraud et al. 2002; Le Coustumer and Barraud, 2007; El-Mufleh et al. 2014).

126 The Chassieu watershed, the Django-R SIS, and the Lyon aquifer were investigated in this study (Figure 1,
127 Table S1). Watershed runoff waters (hereafter WS) have been collected from sampling points spread over the
128 catchment (21 sub-watersheds over three sampling periods, n=64 samples). Sediments from the detention basin
129 (hereafter DB) were recovered from a 50 cm² area covering the full sediment column down to the concrete slab
130 (n=20 samples). These sediments (or urban soils) often had an herbaceous plant cover and were sampled in four
131 areas defined according to the hydrological forces prevailing in the basin (e. g., Marti et al., 2017; Sébastien et al.,
132 2014). Infiltration basin soil samples (hereafter IB) were collected from 3 zones (the area receiving the inflow
133 waters, the bottom area of the basin, and an upper zone of the basin exposed to inflow waters only during **heavy**
134 rain events) (n=5 samples per zone) at a 0-10 cm depth covering a surface of 50 cm². The aquifer samples were
135 recovered from piezometers located upstream (up, in a zone of the aquifer not influenced by water recharge) and
136 downstream (dw, in a zone of the aquifer influenced by water recharge) of the SIS of the Django-R site at a depth
137 of 2 m below the water table (e. g., Barraud et al., 2002; Voisin et al., 2018) (Figure 1). Groundwater samplings
138 (n=6; named AQ_wat) were performed with an immersed pump, used at a pumping rate of 6–8 L. min⁻¹ (PP36
139 inox, SDEC, Reignac-sur-Indre, France), and previously cleaned with 70 % ethanol. The first 50 L were used to
140 rinse the sampling equipment and were subsequently discarded. The following 6 L were used for the
141 microbiological analyses. The biofilm samples (AQ_bio) from the aquifer were recovered from clay beads
142 incubated in the aquifer over 10 days using the piezometers described above (n=6 samples). Clay beads were used
143 as physical **matrices** to sample groundwater biofilms according to Voisin et al. (2016).

144 **2.2 DNA extractions, 16S rRNA gene qPCR assays, and PCR products DNA sequencings**

145 **Approximately** 600 mg of sediments or soils, or up to 5 L of aquifer or runoff water samples filtered using 0.22
146 µm polycarbonate filters, were used per DNA extraction. Total DNA was extracted from soils/sediments or filters
147 using the FastDNA SPIN® Kit for Soil (MP Biomedicals, Carlsbad, France). For clay bead biofilms, microbial
148 cells were detached by shaking at 2500 rpm for 2 min in 10 mL of 0.8 % NaCl. These suspensions were then
149 filtered and their DNA contents were extracted as indicated above. Blank samples were performed during these
150 extractions for both the soils/sediments or filtered cells. DNA **was** quantified using a nanodrop UV-Vis
151 spectrophotometer. Blank DNA extracts showed values below the detection limit. DNA extracts were visualized
152 after electrophoresis at 6V. cm⁻¹ using a TBE buffer (89 mM Tri-borate, 89 mM boric acid, 2 mM EDTA, (pH
153 8.0)) through a 0.8 % (w/v) agarose gel, and DNA staining with 0.4 mg. mL⁻¹ ethidium bromide. A Gel Doc XR+
154 System (Bio-Rad, France) was used to observe the stained DNA, and confirm their relative quantities (between
155 20-120 ng. µL⁻¹; median value around 40 ng. µL⁻¹) and qualities. DNA **was** kept at -80°C and shipped on ice within
156 24h to the DNA sequencing services when appropriate.

157 Quantitative PCR assays were performed on the DNA extracts to estimate their relative content in 16S rRNA gene
158 copies. These assays were performed on a Bio-Rad CFX96 realtime PCR instrument with Bio-Rad CFX Manager
159 software, version 3.0 (Marnes-la-Coquette, France). The 16S rRNA gene primers 338F and 518R described by
160 Park and Crowley (2006) were used, together with the Brilliant II SYBR green low ROX qPCR master mix for
161 SYBR Green qPCR. Melting T° was 60°C. Linearized plasmid DNA containing a 16S rRNA gene **was** used as a
162 standard, and obtained from Marti et al. (2017). Presence of inhibitors in the DNA extracts was checked by spiking

163 known amount of plasmids harboring *int2* (10^7 copies of plasmid per μL) in the PCR mix. Number of cycles needed
164 to get a PCR signal was compared with wells where only plasmid DNA harboring *int2* was added to the qPCR
165 mix. When a high number of cycles was needed to observe a signal, a 5- or 10-fold dilution of the DNA extract
166 was done, and another round of tests was performed to confirm the absence of PCR inhibitions. Each assay was
167 triplicated on distinct DNA extracts, and technical triplicates were performed. The 16S rRNA gene qPCR datasets
168 are presented in Figure S1. These assays confirmed the high number of bacterial cells per compartment (Figure S1
169 and Table S2): (1) soils from the infiltration basin (IB) had a median content of 1.32×10^{11} 16S rRNA gene copies
170 per g dry weight; (2) sediments from the detention basin (DB) of 1.83×10^{11} 16S rRNA gene copies per g dry
171 weight, (3) the runoff waters (WS) had a median content of 4.75×10^8 16S rRNA gene copies per mL, (4) the
172 aquifer waters (AQ_wat) of 3.10×10^6 16S rRNA gene copies per mL, and (5) the aquifer clay bead biofilms
173 showed 1.35×10^7 16S rRNA gene copies per cm^2 .

174 Sequencing of V5-V6 16S rRNA gene (*rrs*) PCR products were performed by MrDNA DNA sequencing services
175 (Shallowater, Texas, USA) on an Illumina Miseq V3. The PCR products were generated using DNA primers 799F
176 (barcode + ACCMGGATTAGATACCCKG) and 1193R (CRTCCMCACCTTCCTC) reported by Beckers et al.
177 (2016). PCR amplifications were performed using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) using the
178 following temperature cycles: 94 °C for 3 min, followed by 28 cycles of 94 °C for 30 s, 53 °C for 40 s, and 72 °C
179 for 1 min, with a final elongation step at 72 °C for 5 min. PCR products and blank control samples were verified
180 using a 2 % agarose gel and following the electrophoretic procedure described above. PCR products obtained from
181 field samples showed sizes around 430 bp but blanks did not show detectable and quantifiable PCR products.
182 Dual-index adapters were ligated to the PCR fragments using the TruSeq® DNA Library Prep Kit which also
183 involved quality controls of the ligation step (Illumina, Paris, France). Illumina Miseq DNA sequencings of the
184 PCR products were paired-end, and set up to obtain around 40K reads per sample.

185 The *tpm* DNA libraries were also sequenced by the Illumina MiSeq V3 technology but by the Biofidal DNA
186 sequencing services (Vaulx-en-Velin, France). PCR products were generated using the following mix of
187 degenerated PCR primers: ILMN-PTCF2 (5'- P5 adapter tag + universal primer +
188 GTGCCGYTRTGYGGCAAGA-3'), ILMN-PTCF2m (5'- P5 adapter tag + universal primer +
189 GTGCCCYTRTGYGGCAAGT-3'), ILMN-PTCR2 (5'- P7 adapter tag + universal primer +
190 ATCAKYGCGGCGCGGTCRTA-3'), and ILMN-PTCR2m (5'- P7 adapter tag + universal primer +
191 ATGAGBGCTGCCCTGTCRTA-3') targeting conserved regions defined by Favre-Bonté et al. (2005). The
192 universal primer was 5'-AGATGTGTATAAGAGACAG-3'. The P5 adapter tag was : 5'-TCGTCGGCAGCGTC-
193 3'. The P7 adapter tag was : 5'- GTCTCGTGGGCTCGG-3'. PCR reactions were performed using the 5X Hot
194 BIOAmp® master mix (Biofidal, France) containing 12.5 mM MgCl_2 , and 10 % DMSO and 50 ng sample DNA
195 final concentrations. PCR cycles were as follow: (1) a hot start at 94°C for 5 min, (2) 35 cycles consisting of 94°C
196 for 30 s, 58°C for 30 s and 72°C for 30 s, and (3) a final extension of 5 min at 72°C. The mix had two carefully
197 optimized enzymes, the HOT FIREPol® DNA polymerase and a proofreading polymerase. This enzyme blend
198 has both 5'→ 3' exonuclease and 3'→ 5' proofreading activities. This mix exhibits an increased fidelity (up to
199 five fold) compared to a regular *Taq* polymerase. PCR products and blank control samples were verified using a
200 2% agarose gel and following the electrophoretic procedure described above. PCR products obtained from field
201 samples showed sizes around 320 bp but blanks did not show detectable and quantifiable PCR products. Index and
202 Illumina P5 or P7 DNA sequences were added by Biofidal through a PCR procedure using the same Hot

203 BIOAmp® master mix and the above temperatures, but limited to 15 PCR cycles. Indexed P5/P7 tagged PCR
204 products were purified using the SPRIselect procedure (Beckman Coulter, Roissy, France). PCR products and
205 blank control samples were verified using the QIAxcel DNA kit (Qiagen, France), and band sizes around 400 bp
206 were observed but not in the blank samples. Quantification of PCR products by the picogreen approach using the
207 Quantifluor dsDNA kit (Promega, France) and a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, France) was
208 performed, and showed low values among the blanks which were at the limit of detection (around 0.07 ng. μL^{-1}).
209 Still, *tpm* harboring bacteria being in low number among a bacterial community (about 2-3%), these controls were
210 run during the Miseq DNA sequencing of the PCR products. Illumina Miseq DNA sequencings of the *tpm* PCR
211 products were paired-end, and set up to obtain around 40K reads per sample. Blank samples generated low numbers
212 of *tpm* reads (blank 1 = 24 reads; blank 2 = 3 reads, blank 4 = 1028 reads, and blank 5 = 1 read), and these have
213 been listed in Table S3. These reads mainly belonged to unknown species (86%). However, reads from *P.*
214 *fluorescens* (from OTUs not found in the field samples), *P. xanthomarina* (17 reads over all blanks) and *P. fragi*
215 (n=3 reads over all blanks) were recovered but did not have any impact on the coalescence analysis. The 16S rRNA
216 and *tpm* gene sequences reported in this work are available at the European Nucleotide Archive
217 (<https://www.ebi.ac.uk/ena>).

218 2.3 Bioinformatic analyses

219 All paired-end MiSeq reads were processed using Mothur 1.40.4 by following a standard operation protocol (SOP)
220 for MiSeq-based microbial community analysis (Schloss et al., 2009; Kozich *et al.*(2013), so-called MiSeq SOP
221 and available at http://www.mothur.org/wiki/MiSeq_SOP. Due to the large number of sequences to be processed,
222 the cluster.split command was used to assign sequences to OTUs. For the 16S rRNA (*rrs*) gene sequences, reads
223 were filtered for length (>300bp), quality score (mean, ≥ 25), number of ambiguous bases (=0), and length of
224 homopolymer runs (<8) using the trim.seqs script in Mothur, and singletons were discarded. The 16S rRNA gene
225 sequences passing these quality criteria were aligned to the SILVA reference alignment template (release 128).
226 Unaligned sequences were removed. Chimeric sequences were identified using the chimera.uchime command and
227 removed. Variability in the number of cleaned reads per sample was observed but not correlated with variations
228 in the number of 16S rRNA gene sequences (Table S2). These variations were thus considered to be due to the
229 DNA sequencing process. Therefore, a sub-sampled dataset (20,624 reads per sample; with exclusion of samples
230 with total reads below this threshold) was used to mitigate the artifact of sample library sizes. Operational
231 Taxonomic Units (OTUs) were defined using a 97% identity cut-off as recommended by several authors in order
232 to collapse sequences into groups that reduce the incidence of sequence errors on the datasets (e. g., Eren et al.
233 2013; and Johnson et al. 2019). It is to be noted that amplicon sequence variants (ASV) could also be used to build
234 contingency tables (e. g., Callahan et al. 2016; Karstens et al. 2019). However, exact sequence variants can
235 generate uncertainties when using 16S rRNA gene sequences because of variations among species and strains due
236 to the presence of multiple copies per genome (Johnson et al. 2019). Figure S2 shows the OTU rarefaction curves
237 for the full and the sub-sampled datasets. The sub-sampled dataset was used for all downstream analyses except
238 those of the SourceTracker Bayesian approach (see below). OTUs were affiliated to taxonomic groups by
239 comparison with the SILVA reference alignment template if a bootstrap P-value over 80% was obtained.
240 FAPROTAX (Louca et al., 2016) functional inferences were performed on the MACADAM Explore web site
241 (<http://macadam.toulouse.inra.fr/>) according to Le Boulch et al. (2019). For the *tpm* gene sequences, chimeric
242 sequences, primers, barcodes were removed, and the dataset was limited to sequences of a minimum length of 210

243 bp (average length=215 bp). These reads were aligned against the *tpm* database
244 (BD_TPM_Mar18_v1.unique_770seq). Unaligned sequences were removed. The number of sequences per sample
245 was then sub-sampled (4,636 sequences per sample; with exclusion of samples with total reads below this
246 threshold). Operational Taxonomic Units (OTUs) were defined at a 100% identity cut-off. The
247 “BD_TPM_Mar18_v1.unique_770seq” database (<http://www.graie.org/othu/donnees>) was used to classify the
248 sequences using the “Wang” text-based Bayesian classifier (Wang et al., 2007) and a P-bootstrap value above
249 80%. Local BLAST analyses were performed on the “BD_TPM_Mar18_v1.unique_770seq” database using the
250 NCBI BLASTX program to check the quality of the taxonomic affiliations.

251 **2.4 Statistical analyses**

252 All statistical analyses were performed in R (v.3.5.1). For the 16S rRNA gene sequences, alpha-diversity estimates
253 were computed using the function “rarefy” from the ‘Vegan’ package (Oksanen et al., 2015). Richness (S_{obs}) was
254 computed as the number of observed OTUs in each sample. The diversity within each individual sample was
255 estimated using the non-parametric Shannon index. To estimate whether the origin of the samples influenced the
256 alpha-diversity, an ANOVA with Tukey’s post-hoc tests was performed. Shared and unique OTUs were depicted
257 in Venn-diagrams with the “limma” package (Ritchie et al., 2015). Concerning the beta-diversity analyses, a
258 neighbor-joining tree was constructed with a maximum-likelihood approximation method using FastTree (Price et
259 al., 2009). Weighted UniFrac distances were calculated for all pairwise OTU patterns according to Lozupone et
260 al. (2011), and used in a Principal Coordinates Analysis (PCoA) (Anderson and Willis, 2003). Permutation tests
261 of distances (PERMANOVA) (Anderson, 2001) were performed using the “vegan” package (Oksanen et al., 2015),
262 to establish the significance of the observed groupings.

263 **2.5 Bacterial community coalescence analyses**

264 The SourceTracker computer package (Knights et al., 2011) was used to investigate community coalescence.
265 SourceTracker is a Bayesian approach built to estimate the most probable proportion of user-defined “sources”
266 DNA reads in a given “sink” community. In the present analysis, various scenarios of community coalescence
267 were investigated such as the coalescence of bacterial taxa from the watershed runoff waters and sediments from
268 the detention and infiltration basins with those of the downstream SIS aquifer water samples or of recent biofilms
269 developing on clay beads incubated in the aquifer. SourceTracker was run with the default parameters (rarefaction
270 depth = 1000 reads from the original cleaned dataset of 16S rRNA gene reads (Fig. S2a), burn-in: 100, restart: 10)
271 to identify sources explaining the OTU patterns observed among the aquifer samples (waters and clay bead
272 biofilms, n=12). Alpha values were tuned using cross-validation (alpha 1= 0.001 and alpha 2= 1). The relative
273 standard deviation (RSD) based on three runs was used as a gauge to evaluate confidence on the computed values
274 (Henry et al., 2016; McCarthy et al., 2017).

275 **3 Results**

276 **3.1 16S rRNA V5-V6 gene sequences distribution biases and profilings**

277 The analysis of the 16S rRNA V5-V6 gene libraries yielded 2,124,272 high-quality sequences distributed across
278 103 samples as described in Table S2. Subsampling-based normalization was applied (20,624 reads per sample)
279 and sequences were distributed into 10,231 16S rRNA gene OTUs (with > 97 % identity between sequences of an
280 OTU). The rarefaction curves are shown in Figure S2. At all sampling sites, bacterial communities were dominated

281 by Proteobacteria, Bacteroidetes and Actinobacteria (WS=95 % of total reads, DB=84 %; IB=71 %; AQ_bio=99
282 % and AQ_wat=59 %), but 10 other phyla with relative abundances greater than 0.5 % were also detected (Figure
283 2A and Table S4). Alpha-diversity estimates showed that aquifer samples harbored a microbiome with a
284 significantly lower richness (AQ_bio: $S_{obs}=278$ OTUs \pm 106 and AQ_wat: $S_{obs}=490$ OTUs \pm 333) and a less diverse
285 bacterial community (AQ_bio: $H'=2.9 \pm 0.3$ and AQ_wat: $H'=4.3 \pm 0.7$) than the ones of the upper compartments
286 ($S_{obs-WS}=1,288$ OTUs \pm 232; $S_{obs-DB}=1,566$ OTUs \pm 245, $S_{obs-IB}=1,503$ OTUs \pm 177 and $H'_{WS}=5.0 \pm 0.5$; $H'_{DB}=5.4$
287 ± 0.5 , $H'_{IB}=5.7 \pm 0.4$) (ANOVA, $p<0.001$) (Figure 2B and Table S5). Among the surface samples, a greater
288 diversity was observed among the soil samples from the infiltration basin than from samples of watershed runoff
289 waters and sediments recovered from the detention basin (ANOVA, $p<0.05$). In the aquifer, water grab samples
290 were more diverse and showed higher 16S rRNA gene OTU contents than biofilms recovered from the clay beads
291 incubated over a 10-day period (ANOVA, $p<0.05$) (Figure 2B and Table S5).

292 The structure of bacterial communities inferred from V5-V6 16S rRNA gene sequences changed markedly
293 along the watershed. A PCoA ordination of the OTU profiles based on weighted Unifrac distances showed samples
294 to be clustered according to their compartment of origin (*i.e.* WS, DB, IB, AQ_bio and AQ_wat) (Figure 3). These
295 changes in community structures between compartments were supported by PERMANOVA statistical tests
296 ($F=20.7$, $P<0.001$). Bacterial communities per compartment were found to contain core and flexible (defined as
297 not conserved between all sampling periods) bacterial taxa. Within the same compartment, similarities between
298 bacterial community profiles ranged from 65 % (AQ_wat) to 82 % (IB), whereas similarities across compartments
299 ranged from 48 % (DB vs AQ_bio) to 66 % (DB vs IB) (Figure S3). Bacterial community profiles of the aquifer
300 waters were found closer to the ones of the detention basin deposits (57 %) and soils of the infiltration basin (61
301 %) than those of the aquifer biofilms (48 and 49 %, respectively). However, more than 89 % of the 16S rRNA
302 gene OTUs ($n=8,284$) identified above the aquifer (WS, DB and IB) were not detected in groundwater samples
303 (AQ_bio and AQ_wat) (Figure S4). This large group of OTUs was made of minor taxa which accounted for 37 %,
304 44% and 47% of the total reads recovered from the WS, DB and IB samples, respectively.

305 **3.2 Coalescence of surface and aquifer bacterial communities**

306 A SourceTracker analysis was performed to estimate the coalescence of bacterial taxa inferred from V5-V6 16S
307 rRNA gene reads from the watershed and SIS down into the aquifer waters and biofilm bacterial communities.
308 This analysis indicated significant coalescence between the bacterial communities of the runoffs, the
309 soils/sediments of the SIS, and the aquifer samples. The aquifer water microbial community upstream the SIS was
310 found to explain about 40 % of the downstream water microbial community (Table 1), while 16S rRNA gene reads
311 from the runoff waters were found to explain about 5 %, and those of the DB around 8 % of the observed patterns
312 (Table 1). The infiltration basin explained about 7 % of the observed diversity among the SIS impacted aquifer
313 water community. The aquifer biofilm bacterial communities were also found to be assemblages of communities
314 from the surface environments. The origin of more than 94 % of the SIS impacted aquifer biofilms could be
315 explained by the SourceTracker. Main sources of taxa were inferred to be the upstream aquifer waters (59 %), the
316 sediments of the detention basin (22 %), and the runoff waters (8.5 %) (Table 1). Soils from the infiltration basin
317 did not appear to have contributed substantially to the taxa recovered from these aquifer biofilms (<4 %) (Table
318 1). Aquifer biofilms recovered upstream the SIS showed a high proportion of taxa related to those observed among
319 the runoff waters (44 %) and the aquifer waters (49 %). This was not considered surprising because runoff

320 infiltration can occur in several sites upstream of the SIS (although no direct relation with other SIS could be made
321 so far).

322 **3.3. 16S rRNA gene inferred bacterial taxa undergoing coalescence in the aquifer**

323 To identify the bacterial taxa involved in the coalescence process, OTUs of the 16S rRNA gene dataset were
324 allocated to taxonomic groups using the SILVA reference alignment template. These taxonomic allocations
325 indicated that (1) 14 genera were only recorded in the aquifer samples, (2) 421 genera were only recorded in the
326 upper surface compartments of the watershed, and (3) 219 were recorded among aboveground and aquifer
327 compartments (Table S6). The following bacterial genera were exclusively associated to the aquifer bacterial
328 communities: *Turicella*, *Fritschea*, *Metachlamydia*, *Macrococcus*, *Anaerococcus*, *Finegoldia*, *Abiotrophia*,
329 *Dialister*, *Leptospirillum*, *Omnitrophus*, *Campylobacter*, *Sulfurimonas*, *Haemophilus*, and *Nitratireductor*. These
330 bacterial genera were recovered from all water samples, and 5 were also detected in biofilms (Table S6). These
331 genera were associated to 926 16S rRNA gene OTUs that accounted for, respectively, 48 % and 1.8 % of the total
332 reads recovered from aquifer waters and aquifer biofilms developing on clay beads. FAPROTAX functional
333 inferences indicated some of these genera to be host-associated such as *Fritschea*, *Metachlamydia*, *Finegoldia*,
334 *Campylobacter* and *Haemophilus*, with the latter two being well-known to contain potential pathogens.
335 *Campylobacter* and *Sulfurimonas* cells have also been associated with nitrogen and sulfur respiration processes,
336 and *Leptospirillum* with nitrification.

337 Regarding the bacterial taxa of the aboveground communities matching those of the aquifer samples, a total of
338 1,021 16S rRNA gene OTUs was found to be shared between these compartments (Table 2 and Figure S4). These
339 OTUs consisted of abundant taxa as they accounted for 9.7-39.4 % of the total reads for the samples recovered
340 from the surface compartments, and for 33.6-83.4 % and 95.0-99.4 % of the total reads of the water and biofilm
341 aquifer samples, respectively (Table 2). The β - and γ -proteobacteria dominated this group. It is noteworthy that
342 aquifer samples collected upstream of the SIS shared less OTUs with the surface compartments (125 OTUs \pm 41)
343 than samples under the influence of the infiltration system (332 OTUs \pm 85) (Table 2 and Figure S4). The shared
344 OTUs between aquifer samples and the upper compartments represented a higher fraction of bacterial communities
345 in samples recovered downstream of the SIS (81.3 % \pm 22.8 of total reads) compared to those collected upstream
346 (68.9 % \pm 30.9 of total reads) (Table 2). Reads from *Pseudomonas*, *Nitrospira*, *Neisseria*, *Streptococcus*, and
347 *Flavobacterium* were the most abundant (>1 %) of the shared OTUs recovered in the aquifer water samples,
348 whereas those allocated to *Pseudomonas*, *Duganella*, *Massilia*, *Nocardia*, *Flavobacterium*, *Aquabacterium*,
349 *Novosphingobium*, *Sphingobium*, *Perlucidibaca*, and *Meganema* were the most abundant (>1 %) among the
350 aquifer biofilms (Table S6). Most of these aquifer water taxa (except *Streptococcus*) were found to be involved in
351 denitrification or nitrification as inferred from FAPROTAX. The biofilm taxa were most often associated with
352 hydrocarbon degradation (*Novosphingobium*, *Sphingobium*, and *Nocardia*) by FAPROTAX. Several of these
353 biofilm bacterial genera were also found to be containing potential human pathogens (*Duganella*, *Massilia*,
354 *Nocardia*, and *Aquabacterium*) by FAPROTAX (and published clinical records). A set of 14 potentially hazardous
355 bacterial genera was selected from Table S6, and used to illustrate the coalescence of bacterial taxa among the
356 aquifer samples on Figure 4. The 16S rRNA gene reads from *Flavobacterium* prevailed in all upper compartments
357 (WS=6.9 % of total reads, DB=13.4 % and IB=8.3 %) and were in significant numbers among the connected
358 aquifer (AQ_wat = 1.1 % and AQ_bio = 3.1 %) (Figure 4B and Table S6). *Pseudomonas* 16S rRNA gene reads

359 were in relatively lower numbers in the upper compartments (WS = 0.4 % of total reads, DB = 0.4 % and IB <
360 0.05 %) than the aquifer (AQ_wat = 8.4 % and AQ_bio = 35.5 %) (Figure 4B and Table S6). Similar trends were
361 observed for *Nocardia* and *Neisseria* OTUs (Figure 4B). Notably, OTUs exclusively recovered from the upper
362 compartments were mainly allocated to the *Gemmatimonas* (0.2-1.6 % of total reads), *Geodermatophilus* (0.1-1.8
363 %) and *Roseomonas* (0.1-1.0 %) (Table S6).

364 **3.4 Coalescence of *Pseudomonas* and other *tpm*-harboring bacterial species**

365 DNA sequences from *tpm* PCR products generated according to Favre-Bonté et al. (2005) allowed a further
366 exploration of the bacterial species undergoing a coalescence with the aquifer microbiome. A total of 19,129 *tpm*
367 OTUs was recorded among the samples (from datasets re-sampled to reach 4,636 reads per sample). As expected,
368 these *tpm* reads were mainly assigned to the *Proteobacteria* (WS = 92 % of total reads, DB = 87 %; IB = 76 %;
369 AQ_wat = 83 % and AQ_wat = 85 %), but some reads could also be attributed to the *Bacteroidetes*, *Nitrospirae*
370 and *Cyanobacteria* (Table S7). These taxonomic allocations allowed the identification of 24 bacterial genera and
371 91 species whose distributions are summarized in Tables S7 and S8. The *tpm* sequences were mainly allocated to
372 *Pseudomonas* (WS = 36 % of the reads, DB = 27 %; IB = 7 %; AQ_wat = 51 % and AQ_bio = 48 %), *Aeromonas*
373 (WS = 1 % of the reads, DB = 3 %; IB <0.05 %; AQ_wat = 0.07 % and AQ_bio < 0.05 %), *Xanthomonas* (WS =
374 4 % of the reads, DB <0.05 %; IB = 1 %; AQ_wat = 8 % and AQ_bio < 0.05 %), *Herbaspirillum* (WS = 11 % of
375 the reads) and *Nitrosomonas* (DB = 4 % of the reads; IB = 0.2 %) (Table S8). Reads related to *Pseudomonas* were
376 allocated to 50 species, including pollutant-degraders (*P. pseudoalcaligenes*, *P. aeruginosa*, *P. fragi*, *P.*
377 *alcaligenes*, *P. putida* and *P. fluorescens*), phytopathogens (*P. syringae*, *P. viridiflava*, *P. stutzeri*, and *P.*
378 *marginalis*) and human opportunistic pathogens (*P. aeruginosa*, *P. putida*, *P. stutzeri*, *P. mendocina*, *S.*
379 *acidaminiphila*) (Table S9). It is to be noted that blank samples sequenced during the *tpm* meta-barcoding assay
380 revealed 23 *Pseudomonas* OTUs coming from the DNA extraction kit or generated during the PCR product
381 Illumina sequencing process (Table S3). Only OTU00573 was found in high number (867 reads), but this
382 contaminant did not have an impact on the coalescence analysis because of its absence in the below ground datasets
383 (Table S10). Other contaminant OTUs did not represent more than 10 times the ones observed in the field samples
384 for identical OTUs, a criterion used to distinguish significant contaminants (Lukasik et al., 2017). In fact, only
385 seven OTUs found among the blanks matched OTUs recovered from the environmental samples, and only two of
386 these could be related to well-defined species *i. e.*, *P. xanthomarina* (17 reads among all blanks) and *P. fragi* (three
387 reads among all blanks). These reads matched a single OTU over eleven allocated to *P. xanthomarina* in the
388 environmental samples, and one OTU over 52 for *P. fragi* (Table S10). Reads related to the *Aeromonas* were
389 attributed to 11 species, but only those allocated to *A. caviae* could be recovered from the aquifer and aboveground
390 compartments (Table S9). Reads related to the *Xanthomonas* were allocated to 9 species, but only those allocated
391 to the *X. axonopodis/campestris* complex and *X. cannabis* species were recovered from the aquifer and upper
392 compartments (Table S9). Regarding the *Pseudomonas*, *tpm* reads allocated to *P. jessenii*, *P. chlororaphis*, and *P.*
393 *resinovorans* were restricted to the aquifer samples. Reads allocated to *P. aeruginosa*, *P. anguilliseptica*, *P.*
394 *chengduensis*, *P. extremaustralis*, *P. fluorescens*, *P. fragi*, *P. gessardii*, *P. koreensis*, *P. pseudoalcaligenes*, *P.*
395 *putida*, *P. stutzeri*, *P. umsongensis*, and *P. viridiflava*, were recovered from the aquifer and upper compartments
396 (Table S9). FAPROTAX analysis indicated that a significant number of the species detected in the aquifer can be
397 involved in denitrification (*P. aeruginosa*, *P. fluorescens*, *P. putida*, *P. stutzeri*, *S. acidaminiphila*, *X.*
398 *autotrophicus*, *P. chlororaphis*) or nitrification (*Nitrospira defluvii*, *Nitrosomonas oligotropha*) but also in

399 hydrocarbon degradation (*P. aeruginosa*, *P. fluorescens*, *P. putida*). Some of these species were also suggested by
400 FAPROTAX to be human pathogens or invertebrate parasites (e. g., *P. chlororaphis*, *P. aeruginosa*).

401 The *tpm* OTUs (representative of infra-specific complexes) shared between the upper compartments and the
402 aquifer were allocated to 14 species and 5 genera (Table 3 and Table S10). Four of these OTUs led to higher
403 relative numbers of reads in the aquifer samples, in the following decreasing order: *P. umsongensis* (Otu00005) >
404 *P. chengduensis* (Otu00024) > *X. axonopodis/campestris* (Otu00019 & Otu00878) > *P. stutzeri* (Otu00119 &
405 Otu10066). These co-occurrences of OTUs between aboveground and aquifer samples support the hypothesis of
406 significant coalescence between these bacterial communities. The other OTUs showed higher number of reads
407 among the top compartments. The OTU allocated to *X. cannabis* showed the highest relative number of reads of
408 this group among runoff waters. The distribution pattern of this OTU suggested a relative decline when moving
409 down the aquifer. The *P. aeruginosa* Otu00066 was recovered in the runoff waters, and biofilms developing on
410 clay beads incubated in the aquifer.

411 4 Discussion

412 The coalescence of bacterial taxa from runoff and stormwater infiltration systems (SIS) with those of a connected
413 aquifer was investigated. Taxonomic and functional inferences were performed using 16S rRNA gene libraries. In
414 addition, a genetic marker named *tpm* was used to track species and particular sequence types of the *Pseudomonas*,
415 *Aeromonas*, and *Xanthomonas* (and a few other genera) from runoffs down into the SIS impacted aquifer.
416 Estimation of alpha-diversity indices from the 16S rRNA bacterial community profilings indicated that
417 groundwater samples (*i.e.* waters and biofilms) harbored a less diverse microbiome than those of the top
418 compartments (*i.e.* WS, DB, IB). A 2 to 5-fold reduction in bacterial richness was observed from the surface
419 compartments down into the aquifer. This result suggested that a high proportion of bacterial taxa carried by
420 stormwater runoffs or thriving in the detention/infiltration basins were retained and/or eliminated by the vadose
421 zone filtration process. In line with this result, the estimation of the copy number of the bacterial 16S rRNA gene
422 by qPCR revealed that bacterial biomass was much lower in aquifer than in runoff samples. In fact, more than 89
423 % of the 16S rRNA gene OTUs in the top compartments were not detected in the underground samples. This is in
424 agreement with previous works which have shown that immobilization of microorganisms through porous media
425 are high in the top soil layers and triggered by mechanical straining, sedimentation and adsorption (Kristian Stevik
426 et al., 2004; Krone et al., 1958). Moreover, particles that accumulate as water passes through the soil can form a
427 mat that enhances this straining process (Krone et al., 1958). Despite this filtering effect, infiltration induces
428 significant changes in the diversity of groundwater bacterial communities. Both, water and biofilm aquifer samples
429 recovered downstream the SIS had higher bacterial richness than those collected upstream. It is to be noted that
430 soils of the infiltration basin showed higher bacterial diversity than those of the sediments of the detention basin
431 and runoffs. This is most likely related to a development of plant-associated bacteria in this compartment. Indeed,
432 the infiltration basin was covered by several plant species of *Magnoliophyta* like *Rumex* sp. which can disseminate
433 rapidly through rhizomes (Bedell et al., 2013) and generate multiple ecological niches for bacteria.

434 The SourceTracker Bayesian probabilistic approach based on 16S rRNA gene meta-barcoding datasets
435 (Knights et al., 2011) was applied to refine our understanding of the coalescence of microbial communities from
436 aboveground environments down into an aquifer. These inferences revealed variable levels of coalescence in the
437 SIS recharged aquifer depending upon the investigated sink *i.e.* waters or biofilms developing on clay beads

438 incubated in the aquifer. Bacterial community structures of the groundwater samples (upstream and downstream
439 the SIS) were significantly built from aboveground communities (*e. g.*, those from runoff waters). However, the
440 origin of a high proportion of the diversity observed among the aquifer waters downstream the SIS remained
441 undefined. This is likely related to the emergence of novel biomes among the vadose zone of a SIS fed with urban
442 waters and pollutants. These biomes would have emerged from the build-up of novel biotopes during the
443 construction and functioning of the SIS (see Winiarski (2014) for review). The prevailing environmental
444 constraints and pollutants would then have favored minor taxa (not detectable by meta-DNA barcoding
445 approaches) from the aboveground compartments. In fact, chemical pollutants have been shown to be significantly
446 washed-off or transported with particles during rain events (El-Mufleh et al., 2014), and some of these were found
447 to reach aquifers fed by SISs (Pinasseau et al. 2019). Among these pollutants, Bernardin-Souibgui et al. (2018)
448 reported that urban sediments found in the detention basin of the experimental site were heavily polluted by
449 polycyclic aromatic hydrocarbons (PAH). Their contents were estimated at 197 ± 36 ng. g dw⁻¹ (dry weight) for
450 light PAHs, and at 955 ± 192 ng. g dw⁻¹ for heavy PAHs. PCBs were also recorded for the seven congeners of the
451 European norm for a total of 0.2 to 2.1 mg. kg dw⁻¹ (Sebastian et al., 2014). Metallic trace elements (MTE) were
452 recorded in significant amounts, with Cu being found at about 280 mg. kg dw⁻¹, Pb at about 200 mg. kg dw⁻¹, Zn
453 at about 1600 mg, and Cd at about 5 mg. kg dw⁻¹ (Sebastian et al., 2014). MTE, PCBs and PAHs were also recorded
454 in the soils of the infiltration basin at similar concentrations *e. g.*, in average, at 0.26 mg PCBs. kg dw⁻¹, and at
455 more than 940 ng. g dw⁻¹ for PAHs (Winiarski, 2014; Winiarski et al. 2006). These sediments and soils were also
456 found contaminated by dioxins at about 36 ng. g dw⁻¹ (Winiarski, 2014), and by 4-nonylphenols and bisphenol A,
457 at concentrations varying from 6 ng. g dw⁻¹ to 3400 ng. g dw⁻¹ (Wiest et al., 2018). However, MTE and non-polar
458 PAHs found among SISs are unlikely to reach groundwaters. To illustrate, Pb and Cd were not recorded at depths
459 below 1.5 m into the non-saturated zone of SISs (Winiarski et al. 2006). In contrast, polar organic pollutants were
460 found to be transferred into aquifers as shown for some pesticides and pharmaceuticals (Pinasseau et al. 2019).
461 These chemical contaminants represent potential energy- and carbon-sources for microorganisms, and can also be
462 detrimental to the growth of some organisms. They can thus have significant impacts on the biology of the
463 contaminated soils and sediments.

464 Functional inferences from the knowledge on bacterial genera suggested an occurrence of several aquifer taxa
465 involved in the nitrogen and sulfur cycles but also in hydrocarbon degradation. *Campylobacter*, *Flavobacterium*,
466 *Pseudomonas*, *Sulfurimonas* cells have been associated with nitrogen and sulfur respiration processes, and
467 *Nitrospira* and *Leptospirillum* with nitrification. The oligotrophic nature of the aquifer waters (concentrations of
468 biodegradable dissolved organic carbon < 0.5 mg. L⁻¹, Mermillod-Blondin et al., 2015) is thus likely to have
469 induced a significant selective sorting of microbial taxa among the merged community. Most abundant
470 aboveground taxa often require high energy (organic carbon) and nutrient levels to proliferate (Cho and Kim,
471 2000; Griebler and Lueders, 2009). Twice as much dissolved organic carbons were detected among aquifer waters
472 of the experimental site recovered downstream the SIS (1.93 mg. L⁻¹ \pm 0.77) than upstream (0.88 mg. L⁻¹ \pm 0.27)
473 (Mermillod-Blondin et al., 2015), and this effect was confirmed for other SIS (Mermillod-Blondin et al., 2015;
474 Winiarski, 2014). Similarly, a large part of the bacterial taxa identified from aquifer biofilms was attributed to
475 aboveground sources by the SourceTracker approach. Indeed, watershed runoff waters and detention basin
476 deposits were found to have significantly contributed to the build-up of the observed biofilm community structures.
477 These biofilms showed a high content of 16S rRNA gene sequences belonging to the β - and γ -proteobacteria.

478 According to the ecological concept of r/K selection, these *Proteobacteria* are often considered as r-strategists,
479 able to respond quickly to environmental fluctuations, and colonize more efficiently newly exposed surfaces than
480 other groups of bacteria (Araya et al., 2003; Fierer et al., 2007; Lladó and Baldrian, 2017; Manz et al., 1999;
481 Pohlen et al., 2010). Moreover, because they tend to concentrate nutrients (Flemming et al., 2016), biofilms are
482 likely to favor the survival of **such** opportunistic bacterial cells capable of exploiting spatially and temporally
483 variable carbon and nutrient sources. Here, taxa recovered from aquifer biofilms were previously recorded to have
484 the ability to use hydrocarbons as carbon- and energy sources e. g., *Nocardia*, *Pseudomonas*, *Sphingobium*, and
485 *Novosphingobium*. As indicated above, SIS and urban runoffs are well known to be highly polluted by such
486 molecules (e. g., Winiarski, 2014; Marti et al., 2017; Wiest et al., 2018). The r/K selection ecological concept thus
487 seems to apply to the biofilm community assemblages observed in this work.

488 Taxonomic allocations of the 16S rRNA OTUs suggested that the aquifer waters and biofilms likely harbored
489 opportunistic human, plant and animal pathogens of the genus *Fingoldia*, *Campylobacter*, *Haemophilus*,
490 *Duganella*, *Massilia*, *Nocardia*, *Aquabacterium*, *Flavobacterium*, *Pseudomonas*, *Streptococcus*, and *Aeromonas*.
491 A striking observation was the enrichment of 16S rRNA gene reads allocated to the *Nocardia* (about 4% of the
492 total reads) and *Pseudomonas* (about 35% of the total reads) in the biofilms recovered from clay beads incubated
493 downstream of the SIS. *Nocardia* and *Pseudomonas* 16S rRNA gene sequences were in much lower relative
494 proportions in the aboveground compartments. The genus *Pseudomonas* was previously found to be abundant
495 under low flow conditions and was often associated with biofilm formation (Douterelo et al., 2013). Moreover,
496 Pseudomonads are well-known for their ability at using hydrocarbons as energy and C-sources. Regarding the
497 *Nocardia* cells, there is a poor knowledge of their ecology, but a few reports indicated a tropism for hydrocarbon
498 polluted urban soils and sediments (e. g., Bernardin-Souibgui et al., 2018; Sébastien et al., 2014). There was no
499 additional approach to further investigate the molecular ecology of *Nocardia* cells found among the investigated
500 urban watershed. However, a *tpm* meta-barcoding analytical scheme **could be** applied on DNA extracts to further
501 explore the taxonomic allocations of the Pseudomonads and some other *tpm*-harboring genera. The applied *tpm*
502 meta-barcoding approach allowed an investigation of the coalescence of about 90 species among the investigated
503 watershed including 50 species of *Pseudomonas*, 11 species allocated to the *Aeromonas*, and some additional
504 species allocated to the *Nitrospira*, *Nitrosomonas*, *Stenotrophomonas*, *Xanthobacter*, and *Xanthomonas*. A single
505 *Aeromonas* species, *A. caviae*, was recorded among the above- and under-ground environments. More than 10
506 *Pseudomonas* species thriving in the recharged aquifer were detected among the aboveground compartments. *P.*
507 *umsongensis* and *P. chengduensis* *tpm* OTUs were detected aboveground, and represented a significant fraction of
508 the *tpm*-harboring bacteria retrieved from the aquifer samples. These two species were initially isolated from farm
509 soil and landfill leachates (Kwon et al., 2003; Tao et al., 2014), further supporting the hypothesis that such soil-
510 associated bacteria can be transferred from runoffs and urban sediments down to natural hydrosystems, and can
511 merge with aquifer communities. Regarding the *Pseudomonas* species that may pose health threats to humans, a
512 *tpm* OTU affiliated to *P. aeruginosa* was found to be shared between the surface compartments and the biofilm
513 *tpm* community developing on clay beads incubated downstream the SIS. *P. aeruginosa* thus had the properties
514 allowing an opportunistic development among the aquifer. This species is known for its metabolic versatility and
515 ability to thrive on hydrocarbons. This is an example of bacterial r-strategist being able to get established
516 opportunistically in aquifer biofilm communities impacted by urban pollutants. Apart from *P. aeruginosa*, the
517 species *P. putida* and *P. stutzeri*, frequently detected in soils and wastewater treatment plants (e.g. Igbinosa et al.,

518 2012; Luczkiewicz et al., 2015; Miyahara et al., 2010), were also recovered along the watershed and the aquifer.
519 Although these two species were identified in human infections (Fernández et al., 2015; Noble and Overman,
520 1994), information about their virulence remains scarce. These species are therefore considered to be of less
521 concern than *P. aeruginosa* and *A. caviae*, another opportunistic infectious agent (Antonelli et al., 2016) found in
522 the aquifer. *P. putida* isolates have been shown to be involved in hydrocarbon degradation, and *P. stutzeri* can play
523 part in the N-cycle either through denitrification or nitrogen-fixation.

524 5 Conclusions

525 The knowledge gained from the present study demonstrated that coalescence of microbial communities from an
526 urban watershed with those of an aquifer can occur and yield novel assemblages. Specialized bacterial
527 communities of aquifer waters were slightly re-shuffled by the aboveground communities. However, the
528 assemblages observed among recent aquifer biofilms were found to be largely colonized by opportunistic r-
529 strategists coming from aboveground compartments, and often associated with the ability to degrade hydrocarbons
530 e. g., the Pseudomonads, *Nocardia* and *Novosphingobium* cells. The aquifer of the investigated site was found, for
531 the first time, to be specifically colonized by species like *P. jessenii*, *P. chlororaphis*, and *P. resinovorans* but also
532 undesirable human opportunistic pathogens such as *P. aeruginosa* and *A. caviae*. Artificial clay beads incubated
533 in the aquifer through piezometers appeared highly efficient trapping systems (termed “germcatchers”) to evaluate
534 the ability of a SIS at preventing transfer of undesirable r-strategists to an aquifer. Nevertheless, the long term
535 incidence of allochthonous bacteria on the integrity of aquifer microbiota remains to be investigated.

536 Free-living aquifer bacterial communities are not likely to be much impaired by exogenous cells. However,
537 microbial communities developing as biofilms on inert surfaces might be significantly re-shuffled through
538 selective sorting likely induced, in part, by aboveground chemical pollutants. Microbial biofilms are key structures
539 in the transformation processes of several chemicals and nutrients. They often display much higher cell densities
540 than free-living populations (Crump and Baross, 1996; Crump et al., 1998; van Loosdrecht et al., 1990). Here, we
541 have demonstrated that runoff and SIS bacterial taxa can colonize solid matrices of a deep aquifer. These modified
542 communities could (i) alter geochemical processes which can indirectly impact other groundwater inhabitants e.
543 g., the amphipod *Niphargus rhenorhodanensis* and other taxa presented in Foulquier et al. (2011), or (ii) directly
544 impact these inhabitants by inducing a modification of their microbial contents, and potentially of their behavior.
545 The stygofauna feed on bacteria, and is well known to be significantly colonized by bacteria (e. g., Smith et al.,
546 2016). The next step in these studies will be to investigate whether native aquifer biofilm communities can resist
547 to repeated invasions by opportunistic r-strategists and if these allochthonous bacteria will impact the ecological
548 health of the stygofauna.

549

550 *Data availability.* The 16S rRNA gene sequences are available at the European Nucleotide Archive
551 (www.ebi.ac.uk/ena) using the following accession numbers: PRJEB33510 (IB), PRJEB21348 (DB),
552 PRJEB29925 (AQ), and PRJEB33507 (WS), and the *tpm* gene sequences using the PRJEB33622 accession
553 number.

554

555 *Supplement.* The supplementary materials related to this article is available online at:

556

557 *Author contribution.* BC coordinated and designed the experiments. YC, VRN, TW, FMB, RB, LM, RM, FV, EB,
558 DB, JV, and BC performed the experiments and contributed at the analysis of the datasets. YC and BC prepared
559 the manuscript with contributions from all co-authors.

560

561 *Competing interests.* The authors declare that they have no conflict of interest.

562 *Acknowledgments.* This work was partly funded by the French national research program for environmental and
563 occupational health of ANSES under the terms of project “Iouqmer” EST 2016/1/120, l’Agence Nationale de la
564 Recherche through the ANR-16-CE32-0006 (FROG) and ANR-17-CE04-0010 (Infiltron) projects, by Labex IMU
565 (Intelligence des Mondes Urbains), the Greater-Lyon Urban Community, the School of Integrated Watershed
566 Sciences H2O'LYON (ANR 17-EURE-0018), the MITI CNRS project named Urbamic, and the Urban School of
567 Lyon (ANR-17-CONV-0004). Authors thank the OTHU network for technical assistance and financial supports,
568 and the reviewers of this paper for their constructive comments.

569

570 Edited by:
571 Reviewed by:
572

573 **References**

- 574 Anderson, M.J.: A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.*, 26, 32–46,
575 <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>, 2001.
- 576 Anderson, M.J., Willis, T.J.: Canonical analysis of principal coordinates: a useful method of constrained ordination
577 for ecology. *Ecology*, 84, 511–525, <https://www.jstor.org/stable/3107905>, 2003.
- 578 Antonelli, A., D’Andrea, M.M., Montagnani, C., Bartolesi, A.M., Di Pilato, V., Fiorini, P., Torricelli, F., Galli, L.,
579 Rossolini, G.M.: Newborn bacteraemia caused by an *Aeromonas caviae* producing the VIM-1 and SHV-12 β -
580 lactamases, encoded by a transferable plasmid. *J. Antimicrob. Chemother.*, 71, 272–274, [https://doi:](https://doi.org/10.1093/jac/dkv304)
581 [10.1093/jac/dkv304](https://doi.org/10.1093/jac/dkv304), 2016.
- 582 Araya, R., Tani, K., Takagi, T., Yamaguchi, N., Nasu, M.: Bacterial activity and community composition in stream
583 water and biofilm from an urban river determined by fluorescent in situ hybridization and DGGE analysis.
584 *FEMS Microbiol. Ecol.*, 43, 111–119, [https://doi: 10.1111/j.1574-6941.2003.tb01050.x](https://doi.org/10.1111/j.1574-6941.2003.tb01050.x), 2003.
- 585 Arnaud, E., Best, A., Parker, B.L., Aravena, R., Dunfield, K.: Transport of *Escherichia coli* through a thick vadose
586 zone. *J. Environ. Qual.*, 44, 1424, [https://doi:10.2134/jeq2015.02.0067](https://doi.org/10.2134/jeq2015.02.0067), 2015.
- 587 Barba, C., Folch, A., Gaju, N., Sanchez-Vila, X., Carrasquilla, M., Grau-Martínez, A., Martínez-Alonso, M.:
588 Microbial community changes induced by managed aquifer recharge activities: linking hydrogeological and
589 biological processes. *Hydrol. Earth Syst. Sci.*, 23, 139–154. <https://doi.org/10.5194/hess-23-139-2019>, 2019.
- 590 Barnes, K.B., Iii, J.M.M., Roberge, M.C.: Impervious surfaces and the quality of natural and built environments,
591 Department of Geography and Environmental Planning. Towson University, 2001.

592 Barraud, S., Gibert, J., Winiarski, T., Bertrand Krajewski, J.-L.: Implementation of a monitoring system to measure
593 impact of stormwater runoff infiltration. *Water Sci. Technol.*, 45, 203-210, [https://iwaponline.com/wst/article-](https://iwaponline.com/wst/article-pdf/45/3/203/425178/203.pdf)
594 [pdf/45/3/203/425178/203.pdf](https://iwaponline.com/wst/article-pdf/45/3/203/425178/203.pdf), 2002.

595 Beckers, B., Beeck, M., Thijs, S., Truyens, S., Weyens, N., Boerjan, W., Vangronsveld, J.: Performance of 16S
596 rDNA primer pairs in the study of rhizosphere and endosphere bacterial microbiomes in metabarcoding studies.
597 *Front Microbiol* 7, 1–15, <https://doi.org/10.3389/fmicb.2016.00650>, 2016.

598 Bedell, J.-P., Mourier, B., Provot, J., Winiarski, T.: Influences on the establishment and dominance of vegetation
599 in stormwater infiltration basins. *Water Sci. Technol.*, 68, 2576–2583, <https://doi.org/10.2166/wst.2013.526>,
600 2013.

601 Bernardin-Souibgui, C., Barraud, S., Bourgeois, E., Aubin, J.-B., Becouze-Lareure, C., Wiest, L., Marjolet, L.,
602 Colinon, C., Lipeme Kouyi, G., Cournoyer, B., Blaha, D.: Incidence of hydrological, chemical, and physical
603 constraints on bacterial pathogens, *Nocardia* cells, and fecal indicator bacteria trapped in an urban stormwater
604 detention basin in Chassieu, France. *Environ. Sci. Poll. Res.*, 25, 24860–24881,
605 <https://doi.org/10.1007/s11356-018-1994-2>, 2018.

606 Booth, D.B.: Urbanization and the natural drainage system-impacts, solutions, and prognoses. *The Northwest*
607 *Environ. J.*, 7, 93–118, 1991.

608 Borchardt, M.A., Bradbury, K.R., Gotkowitz, M.B., Cherry, J.A., Parker, B.L.: Human enteric viruses in
609 groundwater from a confined bedrock aquifer. *Environ. Sci. Technol.*, 41, 6606–6612,
610 <https://doi.org/10.1021/es071110+>, 2007.

611 Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P.: Bioconductor workflow for
612 microbiome data analysis: from raw reads to community analyses. *F1000Research*, 5, 1492,
613 <https://doi.org/10.12688/f1000research.8986.2>, 2016.

614 Cho, J.-C., Kim, S.-J.: Increase in bacterial community diversity in subsurface aquifers receiving livestock
615 wastewater input. *Appl. Environ. Microbiol.*, 66, 956–965, [https://doi: 10.1128/aem.66.3.956-965.2000](https://doi:10.1128/aem.66.3.956-965.2000), 2000.

616 Chong, M.N., Sidhu, J., Aryal, R., Tang, J., Gernjak, W., Escher, B., Toze, S.: Urban stormwater harvesting and
617 reuse: a probe into the chemical, toxicology and microbiological contaminants in water quality. *Environ.*
618 *Monit. Assess.*, 185, 6645–6652. <https://doi.org/10.1007/s10661-012-3053-7>, 2013.

619 Crites, R.W.: Micropollutant removal in rapid infiltration, in: *Artificial Recharge of Groundwater*. Butterworth
620 Publishers, Boston, Massachusetts, pp. 579–608, 1985.

621 Crump, B., Baross, J.: Particle-attached bacteria and heterotrophic plankton associated with the Columbia River
622 estuarine turbidity maxima. *Mar. Ecol. Prog. Ser.*, 138, 265–273, 1996.

623 Crump, B.C., Simenstad, C.A., Baross, J.A.: Particle-attached bacteria dominate the Columbia River estuary.
624 *Aquat. Microb. Ecol.*, 14, 7–18, 1998.

625 Dillon, P., Page, D., Vanderzalm, J., Pavelic, P., Toze, S., Bekele, E., Sidhu, J., Prommer, H., Higginson, S., Regel,
626 R., Rinck-Pfeiffer, S., Purdie, M., Pitman, C., Wintgens, T.: A critical evaluation of combined engineered and

627 aquifer treatment systems in water recycling. *Water Sci. Technol.*, 57, 753–762.
628 <https://doi.org/10.2166/wst.2008.168>, 2008.

629 Douterelo, I., Sharpe, R.L., Boxall, J.B.: Influence of hydraulic regimes on bacterial community structure and
630 composition in an experimental drinking water distribution system. *Water Res.*, 47, 503–516. <https://doi.org/10.1016/j.watres.2012.09.053>, 2013.

632 El-Mufleh, A., Béchet, B., Ruban, V., Legret, M., Clozel, B., Barraud, S., Gonzalez-Merchan, C., Bedell, J.-P.,
633 Delolme C.: Review on physical and chemical characterizations of contaminated sediments from urban
634 stormwater infiltration basins within the framework of the French observatory for urban hydrology (SOERE
635 URBIS). *Environ. Sci. Poll. Res.*, 21, 5329–5346, <https://doi.org/10.1007/s11356-013-2490-3>, 2014.

636 Eren, A.M., Vineis, J.H., Morrison, H.G., Sogin, M.L.: A filtering method to generate high quality short reads
637 using Illumina paired-end technology. *PLoS ONE*, 8, e66643, <https://doi.org/10.1371/journal.pone.0066643>,
638 2013.

639 Favre-Bonté, S., Ranjard, L., Colinon, C., Prigent-Combaret, C., Nazaret, S., Cournoyer, B.: Freshwater selenium-
640 methylating bacterial thiopurine methyltransferases: diversity and molecular phylogeny. *Environ. Microbiol.*
641 7, 153–164, <https://doi.org/10.1111/j.1462-2920.2004.00670.x>, 2005.

642 Fernández, M., Porcel, M., de la Torre, J., Molina-Henares, M.A., Daddaoua, A., Llamas, M.A., Roca, A., Carriel,
643 V., Garzón, I., Ramos, J.L., Alaminos, M., Duque, E.: Analysis of the pathogenic potential of nosocomial
644 *Pseudomonas putida* strains. *Front Microbiol.* 6:871, <https://doi.org/10.3389/fmicb.2015.00871>, 2015.

645 Fierer, N., Bradford, M.A., Jackson, R.B.: Toward an ecological classification of soil bacteria. *Ecology*, 88, 1354–
646 1364. <https://doi.org/10.1890/05-1839>, 2007.

647 Flemming, H.-C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., Kjelleberg, S.: Biofilms: an emergent
648 form of bacterial life. *Nat. Rev. Microbiol.*, 14, 563–575, <https://doi.org/10.1038/nrmicro.2016.94>, 2016.

649 Foulquier, A., Malard, F., Mermillod-Blondin, F., Montuelle, B., Dolédec, S., Volat, B., & Gibert, J.: Surface
650 water linkages regulate trophic interactions in a groundwater food web. *Ecosystems*, 14, 1339–1353,
651 <https://doi.org/10.1007/s10021-011-9484-0>, 2011.

652 Foulquier, A., Simon, L., Gilbert, F., Fourel, F., Malard, F., Mermillod-Blondin, F.: Relative influences of DOC
653 flux and subterranean fauna on microbial abundance and activity in aquifer sediments: new insights from ¹³C-
654 tracer experiments. *Freshwater Biol.*, 55, 1560–1576. <https://doi.org/10.1111/j.1365-2427.2010.02385.x>, 2010.

655 Griebler, C., Lueders, T.: Microbial biodiversity in groundwater ecosystems. *Freshwater Biol.*, 54, 649–677,
656 <https://doi.org/10.1111/j.1365-2427.2008.02013.x>, 2009.

657 Griebler, C., Malard, F., Lefébure, T.: Current developments in groundwater ecology—From biodiversity to
658 ecosystem function and services. *Current Opinion Biotechnol.*, 27, 159–167, <https://doi.org/10.1016/j.copbio.2014.01.018>, 2014.

660 Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., Grant, T., Cottam, D., O’Brien, P., Deletic,
661 A.: Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water*
662 *Res.*, 93, 242–253, <https://doi.org/10.1016/j.watres.2016.02.029>, 2016.

- 663 Igbinoso, I.H., Nwodo, U.U., Sosa, A., Tom, M., Okoh, A.I.: Commensal *Pseudomonas* species isolated from
664 wastewater and freshwater milieus in the Eastern Cape Province, South Africa, as reservoir of antibiotic
665 resistant determinants. *Int. J. Environ. Res. Public Health*, 9, 2537–2549,
666 <https://doi.org/10.3390/ijerph9072537>, 2012.
- 667 Johnson, JS, Spakowicz, DJ, Hong, BY, Petersen, LM, Demkowicz, P, Chen, L, Leopold, SR, Hanson, BM,
668 Agresta, HO, Gerstein, M, Sodergren, E, Weinstock, GM.: Evaluation of 16S rRNA gene sequencing for
669 species and strain-level microbiome analysis. *Nat. Commun.*, 10, 5029, <https://doi.org/10.1038/s41467-019-13036-1>, 2019.
- 671 Karstens, L., Asquith, M., Davin, S., Fair, D., Thomas Gregory, W., Wolfe, A.J., Braun, J., McWeeney, S.:
672 Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems* 4, e00290-
673 19, <https://doi.org/10.1128/mSystems.00290-19>, 2019.
- 674 Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R.,
675 Kelley, S.T.: Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods*, 8, 761,
676 <https://doi.org/10.1038/nmeth.1650>, 2011.
- 677 Konrad, C.P., Booth, D.B.: Hydrologic changes in urban streams and their ecological significance, in: *American*
678 *Fisheries Society Symposium*, 47, 157-177, 2005.
- 679 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D.: Development of a dual-index
680 sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina
681 sequencing platform. *Appl. Environ. Microbiol.*, 79, 5112–5120, <https://doi.org/10.1128/AEM.01043-13>,
682 2013.
- 683 Kristian Stevik, T., Kari Aa, Ausland, G., Fredrik Hanssen, J.: Retention and removal of pathogenic bacteria in
684 wastewater percolating through porous media: a review. *Water Res.*, 38, 1355–1367,
685 <https://doi.org/10.1016/j.watres.2003.12.024>, 2004.
- 686 Krone, R.B., Orlob, G.T., Hodgkinson, C.: Movement of coliform bacteria through porous media. *Sewage Ind.*
687 *Wastes*, 30, 1–13, 1958.
- 688 Kwon, S.W., Kim, J.S., Park, I.C., Yoon, S.H., Park, D.H., Lim, C.K., Go, S.J., *Pseudomonas koreensis* sp. nov.,
689 *Pseudomonas umsongensis* sp. nov. and *Pseudomonas jinjuensis* sp. nov., novel species from farm soils in
690 Korea. *Int. J. Syst. Evol. Microbiol.*, 53, 21–27, <https://doi.org/10.1099/ijs.0.02326-0>, 2003.
- 691 Lapworth, D.J., Baran, N., Stuart, M.E., Ward, R.S.: Emerging organic contaminants in groundwater: A review of
692 sources, fate and occurrence. *Environ. Pollut.* 163, 287–303, <https://doi.org/10.1016/j.envpol.2011.12.034>,
693 2012.
- 694 Lassabatere, L., Angulo-Jaramillo, R., Soria Ugalde, J.M., Cuenca, R., Braud, I., Haverkamp, R.: Beerkan
695 estimation of soil transfer parameters through infiltration experiments—BEST. *Soil Sci. Soc. Am. J.*, 70, 521–
696 532, <https://doi.org/10.2136/sssaj2005.0026>, 2006.
- 697 Le Boulch, M., Déhais, P., Combes, S., Pascal, G.: The MACADAM database: a MetAboliC pAthways DATabase
698 for Microbial taxonomic groups for mining potential metabolic capacities of archaeal and bacterial taxonomic
699 groups. *Database* 2019, <https://doi.org/10.1093/database/baz049>, 2019.

700 Le Coustumer, S., Barraud, S.: Long-term hydraulic and pollution retention performance of infiltration systems.
701 Water Sci. Technol., 55, 235–243, <https://doi.org/10.2166/wst.2007.114>, 2007.

702 Lladó, S., Baldrian, P.: Community-level physiological profiling analyses show potential to identify the
703 copiotrophic bacteria present in soil environments. PLoS One, 12: e0171638,
704 <https://doi.org/10.1371/journal.pone.0171638>, 2017.

705 Louca, S., Parfrey, L.W., Doebeli, M.: Decoupling function and taxonomy in the global ocean microbiome.
706 Science, 353, 1272–1277, <https://doi.org/10.1126/science.aaf4507>, 2016.

707 Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R.: UniFrac: an effective distance metric for
708 microbial community comparison. ISME J., 5, 169–172, <https://doi.org/10.1038/ismej.2010.133>, 2011.

709 Luczkiewicz, A., Kotlarska, E., Artichowicz, W., Tarasewicz, K., Fudala-Ksiazek, S.: Antimicrobial resistance of
710 *Pseudomonas* spp. isolated from wastewater and wastewater-impacted marine coastal zone. Environ. Sci.
711 Pollut. Res., 22, 19823–19834, <https://doi.org/10.1007/s11356-015-5098-y>, 2015.

712 Lukasik, P., Newton, J.A., Sanders, J.G., Hu, Y., Moreau, C.S., Kronauer, D.J.C., O'Donnell, S., Koga, R., Russell
713 J.A.: The structured diversity of specialized gut symbionts of the New World army ants. Mol. Ecol., 26, 3808–
714 3825, <https://doi.org/10.1111/mec.14140>, 2017.

715 Mansour, I., Heppell, C.M., Ryo, M., Rillig, M.C.: Application of the microbial community coalescence concept
716 to riverine networks: Riverine microbial community coalescence. Biol. Rev., 93, 1832–1845,
717 <https://doi.org/10.1111/brv.12422>, 2018.

718 Manz, W., Wendt-Potthoff, K., Neu, T.R., Szewzyk, U., Lawrence, J.R.: Phylogenetic composition, spatial
719 structure, and dynamics of lotic bacterial biofilms investigated by fluorescent in situ hybridization and confocal
720 laser scanning microscopy. Microb. Ecol., 37, 225–237, <https://doi.org/10.1007/s002489900148>, 1999.

721 Marsalek, J., Chocat, B.: International report: Stormwater management. Water Sci. Technol. 46, 1–17, 2002.

722 Marti, R., Bécouze-Lareure, C., Ribun, S., Marjolet, L., Bernardin Souibgui, C., Aubin, J.-B., Lipeme Kouyi, G.,
723 Wiest, L., Blaha, D., Cournoyer, B.: Bacteriome genetic structures of urban deposits are indicative of their
724 origin and impacted by chemical pollutants. Sci. Rep., 7: 13219, <https://doi.org/10.1038/s41598-017-13594-8>,
725 2017.

726 McCarthy, D.T., Jovanovic, D., Lintern, A., Teakle, I., Barnes, M., Deletic, A., Coleman, R., Rooney, G., Prosser,
727 T., Coutts, S.: Source tracking using microbial community fingerprints: method comparison with
728 hydrodynamic modelling. Water Res., 109, 253–265, <https://doi.org/10.1016/j.watres.2016.11.043>, 2017.

729 McGrane, S.J.: Impacts of urbanisation on hydrological and water quality dynamics, and urban water management:
730 a review. Hydrol. Sci. J., 61, 2295–2311, <https://doi.org/10.1080/02626667.2015.1128084>, 2016.

731 Mejía, A.I., Moglen, G.E.: Spatial patterns of urban development from optimization of flood peaks and
732 imperviousness-based measures. J. Hydrol. Eng., 14, 416–424, [https://doi.org/10.1061/\(ASCE\)1084-
733 0699\(2009\)14:4\(416\)](https://doi.org/10.1061/(ASCE)1084-0699(2009)14:4(416)), 2009.

- 734 Mermillod-Blondin, F., Simon, L., Maazouzi, C., Foulquier, A., Delolme, C., Marmonier, P.: Dynamics of
735 dissolved organic carbon (DOC) through stormwater basins designed for groundwater recharge in urban area:
736 assessment of retention efficiency. *Water Res.*, 81, 27–37, <https://doi.org/10.1016/j.watres.2015.05.031>, 2015.
- 737 Mermillod-Blondin, F., Voisin, J., Marjolet, L., Marmonier, P., Cournoyer, B.: Clay beads as artificial trapping
738 matrices for monitoring bacterial distribution among urban stormwater infiltration systems and their connected
739 aquifers. *Environ. Monit. Assess.*, 191, 58, <https://doi.org/10.1007/s10661-019-7190-0>, 2019.
- 740 Miyahara, M., Kim, S.-W., Fushinobu, S., Takaki, K., Yamada, T., Watanabe, A., Miyauchi, K., Endo, G., Wakagi,
741 T., Shoun, H.: Potential of aerobic denitrification by *Pseudomonas stutzeri* TR2 to reduce nitrous oxide
742 emissions from wastewater treatment plants. *Appl. Environ. Microbiol.*, 76, 4619–4625,
743 <https://doi.org/10.1128/AEM.01983-09>, 2010.
- 744 Murphy, E.M., Ginn, T.R.: Modeling microbial processes in porous media. *Hydrogeol. J.*, 8, 142–158,
745 <https://doi.org/10.1007/s100409900043>, 2000.
- 746 Noble, R.C., Overman, S.B.: *Pseudomonas stutzeri* infection. A review of hospital isolates and a review of the
747 literature. *Diagn. Microbiol. Infect. Dis.*, 19, 51–56, [https://doi.org/10.1016/0732-8893\(94\)90051-5](https://doi.org/10.1016/0732-8893(94)90051-5), 1994.
- 748 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., Simpson, G.L., Solymos, P.,
749 Stevens, M.H.H., Wagner, H.: vegan: Community Ecology Package. R package version 2.3–0. 2015, 2015.
- 750 Park, J.-W., Crowley, D.E.: Dynamic changes in *nahAc* gene copy numbers during degradation of naphthalene in
751 PAH-contaminated soils. *Appl. Microbiol. Biotechnol.* 72, 1322–1329, [https://doi.org/10.1007/s00253-006-](https://doi.org/10.1007/s00253-006-0423-5)
752 [0423-5](https://doi.org/10.1007/s00253-006-0423-5), 2006.
- 753 Pigneret, M., Mermillod-Blondin, F., Volatier, L., Romestaing, C., Maire, E., Adrien, J., Guillard, L., Roussel, D.,
754 Hervant, F.: Urban pollution of sediments: Impact on the physiology and burrowing activity of tubificid worms
755 and consequences on biogeochemical processes. *Sci. Total Environ.*, 568, 196–207,
756 <https://doi.org/10.1016/j.scitotenv.2016.05.174>, 2016.
- 757 Pinasseau, L., Wiest, L., Fildier, A., Volatier, L., Fones, G.R., Mills, G.A., Mermillod-Blondin, F., Vulliet E.: Use
758 of passive sampling and high resolution mass spectrometry using a suspect screening approach to characterise
759 emerging pollutants in contaminated groundwater and runoff. *Sci. Total Environ.*, 672, 253–263.
760 <https://doi.org/10.1016/j.scitotenv.2019.03.489>, 2019.
- 761 Pitt, R., Clark, S., Field, R.: Groundwater contamination potential from stormwater infiltration practices. *Urban*
762 *Water*, 1, 217–236, [https://doi.org/10.1016/S1462-0758\(99\)00014-X](https://doi.org/10.1016/S1462-0758(99)00014-X), 1999.
- 763 Pohlson, E., Marxsen, J., Küsel, K.: Pioneering bacterial and algal communities and potential extracellular enzyme
764 activities of stream biofilms. *FEMS Microbiol. Ecol.*, 71, 364–373, [https://doi.org/10.1111/j.1574-](https://doi.org/10.1111/j.1574-6941.2009.00817.x)
765 [6941.2009.00817.x](https://doi.org/10.1111/j.1574-6941.2009.00817.x), 2010.
- 766 Price, M.N., Dehal, P.S., Arkin, A.P.: FastTree: computing large minimum evolution trees with profiles instead of
767 a distance matrix. *Mol. Biol. Evol.*, 26, 1641–1650, <https://doi.org/10.1093/molbev/msp077>, 2009.

768 Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K.: limma powers differential
769 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, 43, e47–e47,
770 <https://doi.org/10.1093/nar/gkv007>, 2015.

771 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley,
772 B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Horn, D.J.V., Weber, C.F.:
773 Introducing mothur: open-source, platform-independent, community-supported software for describing and
774 comparing microbial communities. *Appl. Environ. Microbiol.*, 75, 7537–7541,
775 <https://doi.org/10.1128/AEM.01541-09>, 2009.

776 Sébastien, C., Barraud, S., Ribun, S., Zoropogui, A., Blaha, D., Becouze-Lareure, C., Kouyi, G.L., Cournoyer, B.:
777 Accumulated sediments in a detention basin: chemical and microbial hazard assessment linked to hydrological
778 processes. *Environ. Sci. Pollut. Res.*, 21, 5367–5378, <https://doi.org/10.1007/s11356-013-2397-z>, 2014.

779 Shuster, W.D., Bonta, J., Thurston, H., Warnemuende, E., Smith, D.R.: Impacts of impervious surface on
780 watershed hydrology: A review. *Urban Water J.*, 2, 263–275, <https://doi.org/10.1080/15730620500386529>,
781 2005.

782 Smith, R.J., Paterson, J.S., Launer, E., Tobe, S.S., Morello, E., Leijs, R., Marri, S., Mitchell, J.G.: Stygofauna
783 enhance prokaryotic transport in groundwater ecosystems *Sci. Rep.*, 6, 32738, <https://doi.org/10.1038/srep32738>,
784 2016.

785 Tao, Y., Zhou, Y., He, X., Hu, X., Li, D.: *Pseudomonas chengduensis* sp. nov., isolated from landfill leachate. *Int.*
786 *J. Syst. Evol. Microbiol.*, 64, 95–100, <https://doi.org/10.1099/ijs.0.050294-0>, 2014.

787 Tedoldi, D., Chebbo, G., Pierlot, D., Kovacs, Y., Gromaire, M.-C.: Impact of runoff infiltration on contaminant
788 accumulation and transport in the soil/filter media of sustainable urban drainage systems: A literature review.
789 *Sci. Total Environ.*, 569–570, 904–926, <https://doi.org/10.1016/j.scitotenv.2016.04.215>, 2016.

790 Tikhonov, M.: Community-level cohesion without cooperation. *eLife*, 5. <https://doi.org/10.7554/eLife.15747>,
791 2016.

792 Vadstein, O., Attramadal, K.J.K., Bakke, I., Olsen, Y.: K-Selection as Microbial Community Management
793 Strategy: A Method for Improved Viability of Larvae in Aquaculture. *Front. Microbiol.*, 9:2730,
794 <https://doi.org/10.3389/fmicb.2018.02730>, 2018.

795 van Loosdrecht, M.C., Lyklema, J., Norde, W., Zehnder, A.J.: Influence of interfaces on microbial activity.
796 *Microbiol. Rev.*, 54, 75–87, 1990.

797 Vezzaro, L., Mikkelsen, P.S.: Application of global sensitivity analysis and uncertainty quantification in dynamic
798 modelling of micropollutants in stormwater runoff. *Environ. Modell. Softw.*, 27, 40–51,
799 <https://doi.org/10.1016/j.envsoft.2011.09.012>, 2012.

800 Voisin, J., Cournoyer, B., Mermillod-Blondin, F.: Assessment of artificial substrates for evaluating groundwater
801 microbial quality. *Ecol. Indicators*, 71, 577–586, <https://doi.org/10.1016/j.ecolind.2016.07.035>, 2016.

802 Voisin, J., Cournoyer, B., Vienney, A., Mermillod-Blondin, F.: Aquifer recharge with stormwater runoff in urban
803 areas: Influence of vadose zone thickness on nutrient and bacterial transfers from the surface of infiltration

804 basins to groundwater. *Sci. Total Environ.*, 637–638, 1496–1507,
805 <https://doi.org/10.1016/j.scitotenv.2018.05.094>, 2018.

806 Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R.: Naive Bayesian classifier for rapid assignment of rRNA
807 sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*, 73, 5261–5267,
808 <https://doi.org/10.1128/AEM.00062-07>, 2007.

809 Wiest, L., Baudot, R., Lafay, F., Bonjour, E., Becouze-Lareure, C., Aubin, J.B., Jame, P., Barraud, S., Kouyi
810 Lipeme, G., Sébastien, C., Vulliet, E.: Priority substances in accumulated sediments in a stormwater detention
811 basin from an industrial area. *Env. Poll.*, 243, 1669–1678, <https://doi.org/10.1016/j.envpol.2018.09.138>, 2018.

812 Winiarski, T.: Fonction filtration d'un ouvrage urbain - conséquence sur la formation d'un anthroposol, French
813 Ministry of Ecological Transition and Solidarity, TEMIS archives, [http://temis.documentation.developpement-](http://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0081/Temis-0081850/21970_A.pdf)
814 [durable.gouv.fr/docs/Temis/0081/Temis-0081850/21970_A.pdf](http://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0081/Temis-0081850/21970_A.pdf), 2014.

815 Winiarski, T., Bedell, J.-P., Delolme, C., Perrodin, Y.: The impact of stormwater on a soil profile in an infiltration
816 basin. *Hydrogeol. J.* 14, 1244–1251, <https://doi.org/10.1007/s10040-006-0073-9>, 2006.

817 Winiarski, T., Lassabatere, L., Angulo-Jaramillo, R., Goutaland, D.: Characterization of the heterogeneous flow
818 and pollutant transfer in the unsaturated zone in the fluvio-glacial deposit. *Procedia Environ. Sci.*, 19, 955–
819 964, <https://doi.org/10.1016/j.proenv.2013.06.105>, 2013.

820

Table 1. Coalescence of surface and aquifer bacterial communities inferred by the SourceTracker Bayesian approach and the 16S rRNA gene meta-barcoding dataset*

samples	WS			DB			IB			AQ_wat_up			unknown		
	mean	rsd		mean	rsd		mean	rsd		mean	rsd		mean	rsd	
1 - waters	AQ_wat_dw1	0.3%	33.3	0.3%	43.3		7.5%	42.6		19.7%	30.6		72.3%	4.8	
	AQ_wat_dw2	10.2%	50.6	17.6%	10.3		9.7%	18.8		25.7%	15.4		36.9%	6.7	
	AQ_wat_dw3	5.0%	9.0	5.0%	29.1		3.8%	32.0		70.7%	1.9		15.5%	2.3	
1 - biofilms	AQ_bio_dw1	8.6%	23.5	25.0%	19.1		3.9%	74.1		56.7%	6.9		5.8%	7.9	
	AQ_bio_dw2	13.6%	28.0	28.4%	14.1		2.9%	46.0		48.2%	6.52		6.8%	11.8	
	AQ_bio_dw3	3.4%	17.1	13.9%	18.4		5.5%	39.3		72.1%	1.85		5.2%	29.8	
2 - biofilms	AQ_bio_up1	32.2%	14.5							61.3%	9.5		6.8%	23.7	
	AQ_bio_up2	56.6%	12.6							36.4%	18.3		7.0%	15.7	
	AQ_bio_up3	44.0%	6.6							48.1%	8.1		7.8%	10.8	

* Two analyses are shown from sub-sampled datasets set at 1000 reads: (1) reads from WS, DB, IB, and aquifer waters from upstream the SIS were considered as the sources of taxa for the aquifer samples downstream the SIS; (2) reads from WS and the aquifer waters upstream the SIS were considered as the sources of taxa for the aquifer biofilms recovered upstream the SIS. SourceTracker was run 3 times using the 16S rRNA gene OTU contingency table and the default parameters. Relative contributions of the sources were averaged. Relative standard deviations (%RSD) are indicated, and used as confidence values. RSD > 100% indicates low confidence on the estimated value. WS: Watershed runoff waters; DB: Detention basin sediments; IB: Infiltration basin sediments. Sequences that could not be attributed to one of the tested sources were grouped under the term unknown.

Table 2. Aquifer 16S rRNA gene (*rrs*) OTUs detected in the upper compartments of the investigated watershed and SIS*.

	Upstream SIS					
	AQ_bio_up1	AQ_bio_up2	AQ_bio_up3	AQ_wat_up1	AQ_wat_up2	AQ_wat_up3
(A) Number of aquifer <i>rrs</i> OTUs shared with the upper compartments	185/220	110/160	118/173	93/143	80/164	165/464
(B) Relative abundance of the shared <i>rrs</i> OTUs in the aquifer (in %)	99.4	95.0	96.4	43.8	45.4	33.6
(C) Relative abundance of the shared <i>rrs</i> OTUs in the upper compartments (in %)	24.9	15.5	15.8	9.7	9.8	11.3
	Downstream SIS					
	AQ_bio_dw1	AQ_bio_dw2	AQ_bio_dw3	AQ_wat_dw1	AQ_wat_dw2	AQ_wat_dw3
(A) Number of aquifer <i>rrs</i> OTUs shared with the upper compartments	340/403	308/353	321/362	203/523	357/594	468/1052
(B) Relative abundance of the shared <i>rrs</i> OTUs in the aquifer (in %)	99.4	99.4	99.6	52.2	83.4	53.7
(C) Relative abundance of the shared <i>rrs</i> OTUs in the upper compartments (in %)	29.7	30.7	39.4	12.5	32.0	24.2

*the number or relative number of shared aquifer *rrs* OTUs found in the upper compartments (WS, DB, IB) was computed per aquifer sample recovered upstream (up) or downstream (dw) the SIS (see Fig. 1 for the sampling design), after a re-sampling of the reads set at 20,624 per sample. AQ_wat: Aquifer waters; AQ_bio: Aquifer clay beads biofilms; up: upstream the SIS, dw: downstream the SIS.

Table 3. Relative distribution of *tpm* reads per OTU (mean \pm sd) shared between the upper compartments and the aquifer, and that were allocated to well-defined species.¹

Genus	Species	OTU code ²	WS	DB	IB	AQ_Wat_up	AQ_Bio_up	AQ_Wat_dw	AQ_Bio_dw
<i>Nitrosomonas</i>	<i>oligotrophia</i>	Otu00035	nd	1.5 \pm 3.40	0.15 \pm 0.30	nd	+	+	nd
<i>Pseudomonas</i>	<i>aeruginosa</i>	Otu00066	0.42 \pm 1.13	nd	+	nd	nd	nd	0.17 \pm 0.30
<i>Pseudomonas</i>	<i>chengduensis</i>	Otu00024	nd	+	+	20.43 \pm 35.39	nd	+	nd
<i>Pseudomonas</i>	<i>extremaustralis</i>	Otu04178	nd	+	nd	nd	nd	+	nd
<i>Pseudomonas</i>	<i>fragi</i>	Otu00197	0.61 \pm 4.05	nd	nd	nd	nd	+	nd
<i>Pseudomonas</i>	<i>pseudoalcaligenes</i>	Otu00197	0.07 \pm 0.38	+	nd	+	nd	nd	nd
<i>Pseudomonas</i>	<i>putida</i>	Otu00800	+	+	nd	nd	nd	+	nd
<i>Pseudomonas</i>	<i>stutzeri</i>	Otu00119 & Otu10066	0.06 \pm 0.33	nd	+	3.06 \pm 5.29	nd	nd	+
<i>Pseudomonas</i>	<i>umsongensis</i>	Otu00005	+	+	nd	0.41 \pm 0.71	17.79 \pm 20.11	5.34 \pm 8.58	11.71 \pm 13.17
<i>Pseudomonas</i>	<i>viridiflava</i>	Otu00204	0.06 \pm 0.31	nd	0.3 \pm 1.09	nd	nd	0.07 \pm 0.12	nd
<i>Stenotrophomonas</i>	<i>acidaminiphila</i>	Otu00072 & Otu01119	0.09 \pm 0.42	0.29 \pm 0.91	0.06 \pm 0.22	nd	nd	+	nd
<i>Xanthobacter</i>	<i>autotrophicus</i>	Otu00501	+	+	nd	nd	nd	0.06 \pm 0.11	+
<i>Xanthomonas</i>	<i>axonopodis/campestris</i>	Otu00019 & Otu00878	0.25 \pm 0.75	nd	1.24 \pm 2.07	16.04 \pm 27.78	nd	nd	+
<i>Xanthomonas</i>	<i>cannabis</i>	Otu00004	3.74 \pm 9.47	nd	nd	nd	+	+	+

¹ All reads from *tpm* OTUs shared between the upper compartments and the aquifer were used to compute the relative abundances.

² *tpm* sequences of the OTUs are shown in Table S8. WS: Watershed runoff waters; DB: Detention basin deposits; IB: soil of the infiltration basin; AQ_water: Aquifer waters; AQ_bio: Aquifer biofilms. + : OTUs with a relative abundance $<$ 0.05%. nd : not detected.

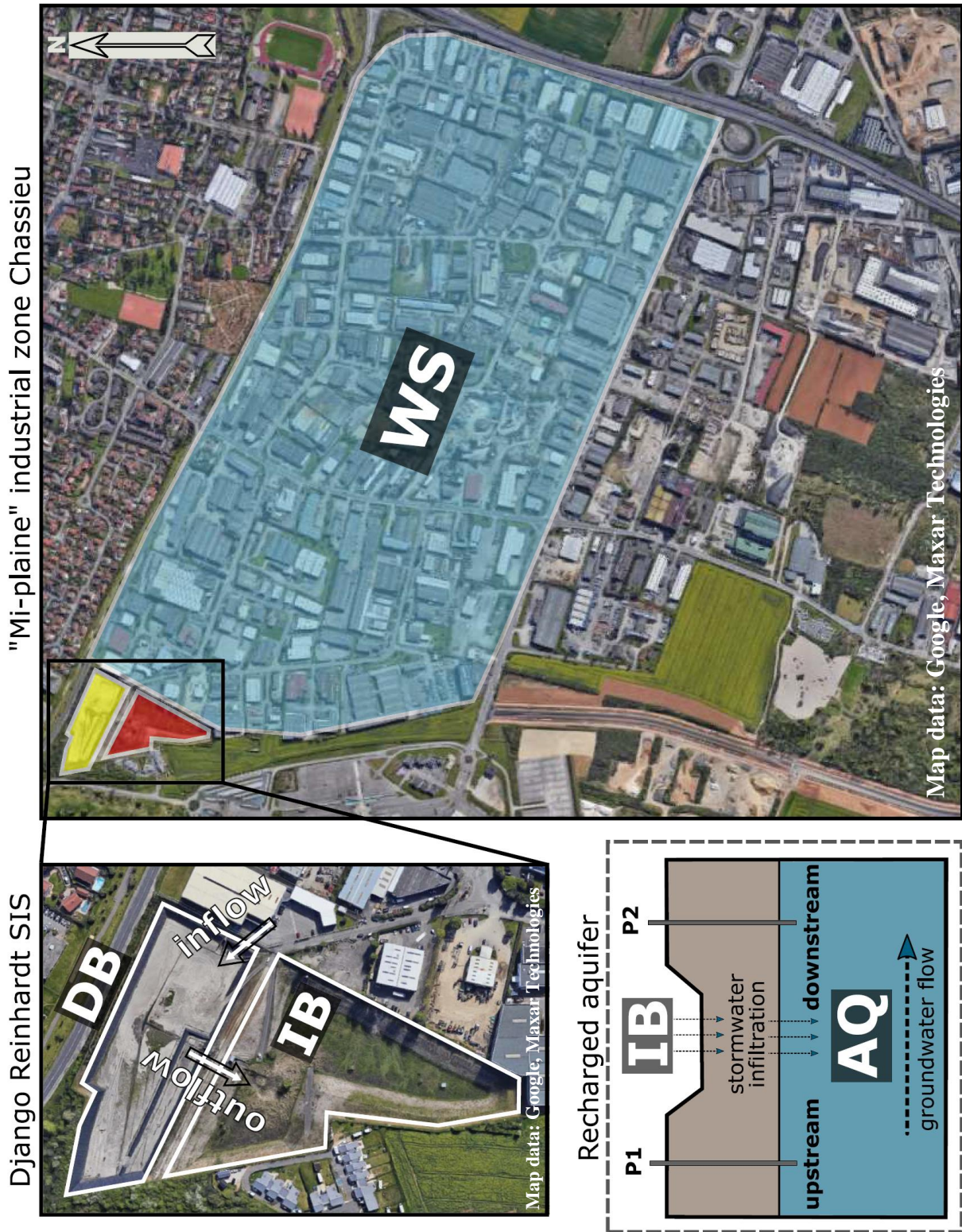
824 **Figure captions**

825 **Figure 1.** Scheme illustrating the trajectory of urban runoffs from the industrial watershed (WS) towards the
826 stormwater infiltration system (SIS) investigated in this study. The urban watershed is located in Chassieu
827 (France). The SIS is made of a detention basin (DB) and an infiltration basin (IB), and is connected to the Lyon
828 200 km² east aquifer (AQ).

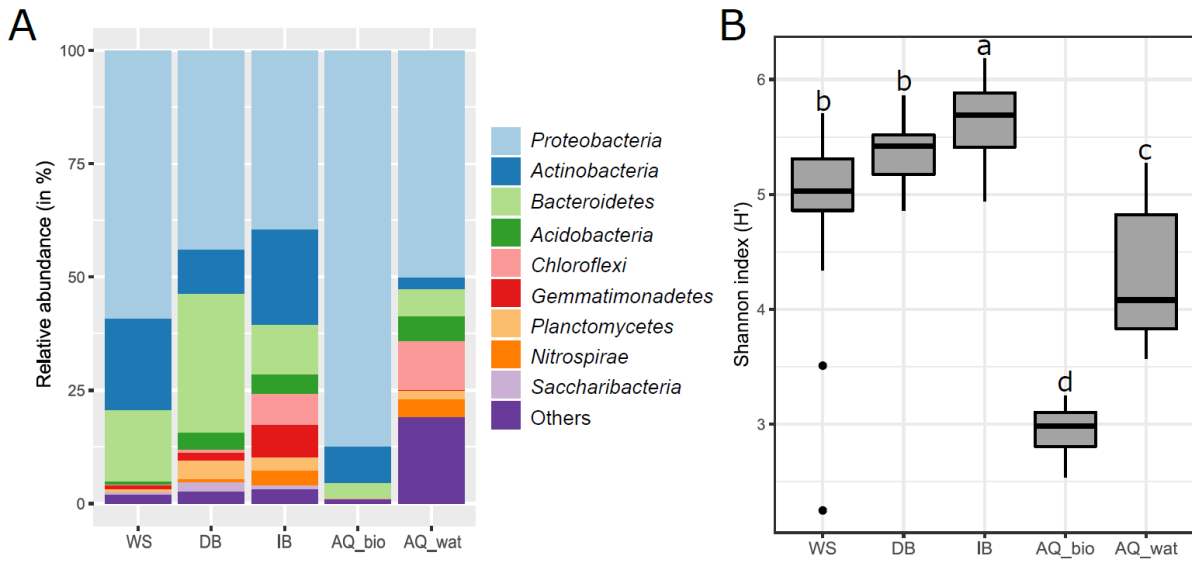
829 **Figure 2.** General features of the V5-V6 16S rRNA gene meta-barcoding DNA sequences obtained from runoff,
830 SIS, and aquifer samples. See Figure 1 for a description of the experimental design. Panel (A) illustrates the relative
831 abundance of the main bacterial phyla observed per compartment, and (B) shows boxplots illustrating the
832 variations observed per compartment between the Shannon diversity indices computed from the V5-V6 16S rRNA
833 gene OTU contingency table (sub-sampled dataset). One-way ANOVA with multiple Tukey post hoc tests were
834 performed to investigate the differences between compartments. Different letter codes indicate significant
835 differences ($p < 0.05$). WS, runoff waters from the watershed; DB: sediments from the detention basin; IB: soils
836 from the infiltration basin; AQ_water: Aquifer waters; AQ_bio: Aquifer clay beads biofilms.

837 **Figure 3.** PCoA analysis of weighted UniFrac dissimilarities between the V5-V6 16S rRNA gene OTU profiles
838 of the watershed runoff waters (WS), urban sediments and soils from the connected detention (DB) and infiltration
839 (IB) basins, and waters (AQ_water) and biofilms (AQ_bio) from the connected aquifer. See Figure 1 for a
840 description of the experimental design. Ellipses are representative of the variance observed (standard error)
841 between the ordinations of a group of samples. PERMANOVA tests confirmed the significance ($p < 0.001$) of the
842 groupings. The proportion of the eigenvalue per axis over the sum of all eigenvalues was 25.0% for PCoA1 and
843 17.6% for PCoA2. The sub-sampled V5-V6 16S rRNA gene dataset was used for the computations.

844 **Figure 4.** Relative numbers of potentially pathogenic bacterial genera along the watershed. The abundance (rel.
845 abund.) of bacterial genera exclusively detected in upper compartments (A) or both in the upper compartments
846 and aquifer (B) are presented. Size of bubbles is proportional to the relative abundance (in %) of each bacterial
847 genus per sampled compartment. WS: runoff waters from the watershed; DB: sediments from the detention basin;
848 IB: sediments from the infiltration basin; AQ_water: Aquifer waters; AQ_bio: Aquifer clay beads biofilms.



852



853

