Bacterial assemblages of urban microbiomes mobilized by runoff waters match land use typologies and harbor core species involved in pollutant degradation and opportunistic human infections

Rayan Bouchali, Claire Mandon, Romain Marti, Jérôme Michalon, Axel Aigle, Laurence Marjolet, Sophie Vareilles, Gislain Lipeme Kouyi, Philippe Polomé, Jean-Yves Toussaint, et al.

Bacterial assemblages of urban microbiomes mobilized by runoff waters match land use typologies and harbor core species involved in pollutant degradation and opportunistic human infections

Rayan Bouchali[1], Claire Mandon[2], Romain Marti[1], Jérôme Michalon[3], Axel Aigle[1], Laurence Marjolet[1], Sophie Vareilles[2], Gislain Lipeme Kouyi[4], Philippe Polomé[5], Jean-Yves Toussaint[2], and Benoit Cournoyer[1]*

[1]Université de Lyon, Université Claude Bernard Lyon 1, VetAgro Sup, UMR Ecologie Microbienne, CNRS 5557, INRA 1418, 69280 Marcy L'Etoile, France;[2]Université de Lyon, INSA Lyon, UMR Environnement, Ville, Société, CNRS 5600, 18 rue Chevreul, 69362 Lyon, France ; ;[3]Université de Lyon, UMR Triangle, CNRS 5206 Université Jean Monnet Saint Etienne, 6 rue Basse des Rives, 42023 Saint-Etienne, France ; [4]Université de Lyon, INSA Lyon, DEEP, EA7429, 11 rue de la physique, 69621 Villeurbanne, France ; [5]Université de Lyon, UMR GATE, CNRS 5824, Université Lumière Lyon 2, 93 chemin des Mouilles, 69131 Ecully, France

***Corresponding author:**
B. Cournoyer, UMR Microbial Ecology, CNRS 5557, CNRS 1418, VetAgro Sup, Main building, aisle 3, 1st floor, 69280 Marcy-L'Etoile, France. Tel. (+33) 478 87 56 47. Fax. (+33) 472 43 12 23. Email: benoit.cournoyer@vetagro-sup.fr

Running title: Pollution-driven urban microbiome

Keywords: Runoff; DNA metabarcoding; Urban microbiology; Petrol engines; Core microbiome

1. Introduction

Humans concentrate in urban settings, with half of the world population now living in cities. This proportion is likely to rise up to 70% in 2050, according to the World Health Organization (2015), representing a population of six billions. Cities can be divided according to their main functional categories (hereafter termed morphotypes) as defined by their commercial, residential and industrial activities (e. g. Revitt et al., 2014). These morphotypes can then be subdivided according to their typologies which are defined according to variables such as: (i) land cover composition including green (e. g. parks) and blue (streams, lake) belts, (ii) land use intensity, (iii) connectivity between impervious surfaces such as car parks and the presence of urban drainage systems, roads and sidewalks, and (iv) other factors including industrial and commercial activities (Göbel et al., 2007; Revitt et al., 2014). These typologies can generate shelters for, among others, several birds, rats, spiders, and favor the development of particular microbiota (e. g. Marti et al., 2017; Aigle et al., 2021).

Ecological trends among microbial communities of urban biomes (conceptualized under the term urban microbiomes) have been investigated for some typologies (see Gilbert and Stephens (2018) for review). To illustrate, the bacterial taxa recovered from surface swabs or air samples from transit systems (e. g. Danko et al., 2021), soils of urban parks (Ramirez et al., 2014), deposits and sediments from streets (Marti et al., 2017; Aigle et al., 2021) and urban waters (McLellan et al., 2015; Voisin et al., 2018) were investigated through 16S rRNA (ribosomal RNA) and *tpm* (thiopurine methyltransferase) meta-barcoding (meaning DNA sequence analysis of PCR products) approaches or metagenomics (meaning an exhaustive sequencing of total bacterial DNA extracts). The diversity observed on city surfaces was found to be significantly different from the ones of agricultural soils and meadows (Ibekwe et al., 2013). Skin-associated bacterial genera were significantly found among air samples of transit systems during peak commuting hours, and relative humidity and air temperature were

inversely linked to the ecological richness of these samples as inferred from 16S rRNA meta-barcoding profilings (Leung et al., 2014). Interestingly, an equilibrium between indoor and outdoor air-borne microbial communities was observed in these systems e. g. Leung et al. (2014); Robertson et al. (2013). Furthermore, the microbiome of urban deposits was found to greatly evolve over time while pollutants were accumulating (Aigle et al., 2021; Marti et al., 2017).

Urban microbiomes are expected to be structured around core and versatile opportunistic microbial groups. These would be representative of the main sources of microbial taxa seeding a particular biotope, and being organized in efficient functional units by the prevailing selective forces. Even though stochastic microbial distribution patterns likely led to some of the observed DNA imprints among these communities (e. g. Hao et al., 2016), one can hypothesize that the principles of the r/K ecological theory (e. g. Song et al., 2017) can apply to these biomes, and gradually lead to a selection of K-like specialists (best fit). In fact, a core urban microbiome for transit systems was recently reported, supporting this hypothesis (Danko et al., 2021). Synurbic (defined as “living in the city”) microbial species thus appear to have been selected over time among urban biomes. However, on the short run, these urban infrastructures are likely to represent transient opportunities for r-ecological strategists (at the other end of this r/K gradient) which can grow rapidly, and efficiently colonize empty or disturbed habitats. These r-strategists are considered better fit for uncrowded microbiome because of their fast growth rate (e. g. Song et al., 2017) and greater metabolic versatility. These r-strategists have often been associated to human- and animal-opportunistic pathogens (Vadstein et al., 2018). However, recent investigations have shown that the numbers of virulence and antibiotic resistance genes can be more abundant among bacterial communities growing under oligotrophic growth conditions than copiotrophic ones (Song et al., 2017). These species could have selected virulence or defense genes to increase their competitivity

when resource supply rates are low. A better knowledge of these r/K gradients among urban biomes could identify risky behaviors (e. g. chemical spills) or practices that can favor undesirable opportunistic bacteria such as the human pathogenic ones.

Here, the incidence of land use specificities and human activities on core (K-strategists) and versatile bacterial taxa (r-strategists) colonizing urban surfaces was investigated. To address these issues, a long term experimental site, the Mi-plaine catchment of the Field Observatory for Urban Water Management (OTHU) located in Lyon (France), was used. This catchment is an impervious area (impervious rate of 75%) drained by a separated stormwater system made of detention and infiltration basins (SIS) located at the outlet, in order to avoid flooding and allow a recharge of the connected aquifer (the one of Lyon) (e. g. Sébastian et al., 2014).

Main objectives of this study were to:

(1) evaluate similarities between sub-catchments based on their socio-urbanistic patterns, relative runoff volumes per rain event, and contents in classical bacterial indicators

(2) investigate relationships between sub-catchments' specificities and the numbers of core bacterial K- and versatile r-strategists recovered over urban surfaces (using runoff waters) through a 16S rRNA gene meta-barcoding approach allowing differentiation at the genus level

(3) investigate relationships in the numbers of core bacterial K- and versatile r-strategists recovered over surfaces (using runoff waters) through a *tpm* gene meta-barcoding approach allowing differentiation at the species level

(4) resolve co-occurrence networks between bacterial taxa inferred from the meta-barcoding profilings, and identify keystone bacterial taxa indicative of sorting processes triggered by socio-urbanistic variables

2. Materials and Methods

2.1. Experimental site and its main socio-urbanistic features

The Mi-plaine urban catchment of the Lyon area (http://www.chassieu.fr/la_zone_industrielle_mi_plaine.html) was investigated in this study. The selected watershed is part of the town of Chassieu, and covers a 210 ha which is impervious at 75% (**Fig. 1 and Fig. 2**). Roads are cleaned every week by a mechanical sweeper. The watershed of Chassieu is divided into sub-catchments (**Fig. S1**) transferring their runoffs into the main rain water network and SIS of the catchment. These sub-catchments were used as references for the socio-urbanistic surveys and definition of typologies, and for inferring relationships with the genetic structures of bacterial communities inferred from 16S rRNA and *tpm* DNA profilings (see section 2.3).

Socio-urbanistic and industrials surveys were performed over three periods (spring, summer, and fall). Technical objects, traces and human activities over the study site that could impact the urban microbiomes were recorded. The first series of surveys (n=3) allowed defining runoff sampling points (n=21) per sub-catchment according to the observed activities and organizations shown in **Table S1**. Urbanistic features and objects separating private properties (gardens and parking lots, buildings) were considered to define the area leading to the observed runoffs at each sampling point (**Table S1, and Fig. 2**). Low concrete walls often separated the private and public areas but sometimes porous delimitations (portal, shrubs, lawn) were observed (**Table S1**). Economical and industrial activities at the sampling points were identified and divided into categories for further analyses (**Table S2a, b, c**). Twelve field surveys of the technical objects, traces and human activities completed these socio-industrial profilings of the watershed (**Table S2**). These surveys were performed over a surface covering a 50 m diameter per sampling zone. These surveys were performed in the morning, afternoon and at night, over 30-minute long sessions per point on working days and

considering peak hours (noon and evening, and n=3 were performed at night). This allowed the identification of 54 variables representing objects and traces indicative of particular behaviors that have impacted the selected sampling sites (**Table S2**).

2.2. Runoff water samplings, and their main physico-chemical and microbiological features

Three rain events (October 16 2014, March 25 2015, and September 17 2015) were considered in this study. Meteorological parameters for these events are shown in **Fig. S1**. These rain events came from different geographical locations. The Canoe urban hydrology computational platform (Chocat, 2013) was used to estimate runoff flow values and volumes per sub-catchment during these events (**Fig. S1**). Runoff waters were collected just before the grate inlet of each sampling site except for the outlet of the separated rain stormwater drainage system (sampling point C23). Waters at the outlet were recovered from the inlet of the detention basin. One liter was collected per sampling point using a single-use manual pump. Electrical conductivity, turbidity, pH, oxygen, and temperature were measured on site using a multi-parametric probe (Horiba, Piscataway, USA) (Table S3). Samples were kept at 4°C until performing the microbiological analyses described below. Numbers of total heterotrophic bacteria were estimated by plating serial dilutions of the runoff waters onto 1/10 diluted TSB amended with 1.5 % agar. Numbers of intestinal *Enterococci* and *Escherichia coli* were estimated using the IDEXX most probable number methods named Colilert and Enterolert (IDEXX, Westbrook, USA).

Microbial DNA extracts were produced from filtered water samples through 0.2 µm pore size polycarbonate filters (Merckmillipore, Burlington, USA) by using the FastDNA SPIN® Kit for Soil (MP Biomedicals, Carlsbad, France). Depending on the amount of suspended matters, between 23 and 100 mL of runoff waters could be filtered. DNA concentrations were measured using a Nanodrop ND1000 (Thermofisher, Waltham, USA). Triplicated qPCR

assays were applied on these DNA extracts using a Bio-Rad CFX96 qPCR device. These assays used the Brilliant II SYBR Green low ROX qPCR mix (Agilent, Vénissieux, France). The CFX Manager 3.0 software (Bio-Rad, Marnes-la-Coquette, France) was used to estimate the numbers of copies of targeted genes per ng DNA (or equivalent runoff volume). The human-specific HF183 *Bacteroides* qPCR assay was performed according to Seurinck et al. (2005), and the assay for the 16S rRNA gene copies of Bacteria was performed according to Park and Crowley (2006) using primers 338F and 518R. Melting T° was 60°C for all assays. Linearized plasmid DNA containing 16S rRNA genes from the targeted DNAs were run as standards using 10-fold dilutions of the plasmids. These plasmids were obtained from Marti et al. (2017). Class 1 integrons (*int1*) were quantified using primers targeting the integrase gene. Presence of inhibitors in the DNA extracts was checked according to Marti et al. (2017). When inhibition was detected, extracts were diluted 10 to 100 times using sterile ultra-pure water until a positive PCR product could be obtained.

2.3. Meta-barcoding analyses of V5-V7 16S rRNA and *tpm* gene PCR products

2.3.1. PCR products and their MiSeq Illumina sequencing

The V5-V7 16S rRNA gene PCR products were generated using DNA primers 799F (barcode + ACCMGGATTAGATACCCKG) and 1193R (CRTCCMCACCTTCCTC) reported by Beckers et al. (2016). PCR amplifications were performed using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) using the 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 for 1 min, with a final elongation step at 72 °C for 5 min. PCR products and blank control samples were verified using a 2% agarose gel and the electrophoretic procedures described in Colin et al. (2020). PCR products obtained from field samples showed sizes around 430 bp but blanks did not show detectable and quantifiable PCR products. Dual-index adapters were ligated to the PCR fragments using the TruSeq® DNA Library Prep Kit which also involved quality controls of

151 the ligation step (Illumina, Paris, France). Illumina V3 Miseq DNA sequencings were
152 performed by MrDNA services (Shallowater, Texas, USA). DNA sequences were paired-end,
153 and set up to obtain around 40K reads per sample.

154 The *tpm* DNA libraries were also sequenced by the Illumina MiSeq V3 technology but by
155 the Biofidal DNA sequencing services (Vaulx-en-Velin, France). PCR products were
156 generated using the following mix of degenerated PCR primers as reported in Aigle et al.
157 (2021): ILMN-PTCF2 (5'- P5 adapter tag + universal primer +
158 GTGCCGYTRTGYGGCAAGA-'3), ILMN-PTCF2m (5'- P5 adapter tag + universal primer
159 + GTGCCCYTRTGYGGCAAGT-'3), ILMN-PTCR2 (5'- P7 adapter tag + universal primer
160 + ATCAKYGCGGCGCGGTCRTA-'3), and ILMN-PTCR2m (5'- P7 adapter tag + universal
161 primer + ATGAGBGCTGCCCTGTCRTA-'3). The universal primer was 5'-
162 AGATGTGTATAAGAGACAG-'3. The P5 adapter tag was 5'-TCGTCGGCAGCGTC-'3.
163 The P7 adapter tag was: 5'- GTCTCGTGGGCTCGG-'3. PCR reactions were performed and
164 verified as described in Aigle et al. (2021). PCR products obtained from field samples showed
165 sizes around 320 bp but blanks did not show detectable PCR products. Still, *tpm* harboring
166 bacteria being in low number among a bacterial community (about 2-3%), blank samples
167 were run during the Miseq DNA sequencing of the PCR products. Illumina Miseq DNA
168 sequencings of the *tpm* PCR products were paired-end, and set up to obtain around 40K reads
169 per sample. Blank samples generated low numbers of *tpm* reads that were retrieved from the
170 dataset by the decontam procedure. The 16S rRNA and *tpm* gene sequences reported in this
171 work are available at the European Nucleotide Archive (https://www.ebi.ac.uk/ena) under the
172 accession #PRJEB33507.

173 **2.3.2. Processing of meta-barcoding V5-V7 16S rRNA and *tpm* gene DNA sequences, and**
174 **computation of diversity indices and taxonomic allocations**

In order to select the bioinformatic processing packages for the construction of the 16S rRNA gene and *tpm* contingency tables, raw reads were treated by both the Mothur (Schloss et al., 2009) and dada2 packages (Callahan et al., 2016). Regarding dada2 manipulations, linker, barcode and primers were removed from the DNA reads by using the Trimgalore v0.6.5 software (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The Standard Operating Procedure (SOP) for merged reads was used for the processing of the 16S rRNA raw reads (https://benjjneb.github.io/dada2/bigdata_1_2.html) while the SOP for the paired-end reads was used for the *tpm* raw reads (https://benjjneb.github.io/dada2/bigdata_paired.html). Regarding Mothur, sequences were treated according to the pipeline defined by Kozich et al. (2013). In both cases, raw sequences were discarded when they were: (1) shorter than 350 bp and longer than 400 bp for V5-V7 16S rRNA gene, and than 250 bp and 270 bp for *tpm;* (2) had few identities with reference sequences of the database, (3) contained homopolymers longer than 8 bp, (4) contained ambiguous bases. Chimeric DNA sequences were detected and discarded using the chimera.uchime command (Edgar et al., 2011). For Mothur, the number of sequences was normalized between samples by performing a random resampling at n = 10783 for the V5-V7 16S rRNA gene reads and n = 8072 for the *tpm* reads. Operational Taxonomic Units (OTUs) were defined at 97 % identity for the V5-V7 16S rRNA gene reads or 100 % (exact sequence variants) identity for *tpm* reads. Contaminant OTUs which may come from the extraction kit, or the 0.2 µm pore size polycarbonate filters were detected using the Decontam v1.8.0 package (Davis et al., 2018). The Decontam package was used on the OTU tables using the DNA concentration method for the 16S rRNA gene datasets (**Table S4**). The *tpm* OTU contaminants of the blank samples were identified by Decontam (**Table S5**). The R Decontam package detected only 370 16S rRNA contaminant gene reads (representing 0.059 % of the total reads) distributed into 50 OTUs among the dataset (**Table S5**). These low numbers of

contaminant reads did not have any effect on the statistical analyses. The validated OTU contingency tables are given in **Tables S6** and **S7**. Diversity indices were computed from OTU and ASV (amplicon sequence variants), and by grouping the OTU (**Table S8**) and ASV (**Table S9**) per genus. High goodness of fit (R^2) and significant correlations were observed by Spearman correlation tests using both types (OTU against ASV) of contingency tables (**Fig. S2**). Dada2 16S rRNA gene ASV diversity indices were significantly higher than those computed from the Mothur OTU dataset (Wilcoxon-test: p-value < 0.001). For the *tpm* datasets, diversity indices were similar between the ASV/OTU datasets (Wilcoxon test, p-value > 0.05). Spearman correlation tests confirmed the high positive correlations between both datasets (**Fig. S2**). OTUs being found in higher numbers in both the 16S rRNA gene and *tpm* datasets, the following analyses were limited to the OTU contingency Tables. OTUs taxonomic allocations were performed by comparison with the Silva database (16S rRNA genes) or the "BD_TPM_Mar18_v1.unique_770seq" database (Aigle et al., 2021) and using the "Wang" text-based Bayesian classifier (Wang et al., 2007). Only classifications resolved with a bootstrap above 80% were considered. Some functional inferences were made from the taxonomic allocations using the Faprotax software version 1.2.1 (Louca et al., 2016).

2.4. Statistical analyses

Alpha diversity indices (Shannon, Simpson and Evenness) were computed using the Vegan package v2.5-6 for Rv3.5.3 as well as Hellinger transformation. Spearman correlations were computed using the R package Hmisc v4.4.0 and the correlation networks were computed using the Igraph package v1.2.6. The adjusted Wallace coefficient analyses were performed at the UMMI web site of Lisboa University (http://www.comparingpartitions.info/). Chord diagrams showing the relative abundance of 16S rRNA or *tpm* – based taxonomic allocations were made using the "circlize" v0.4.6 package for R. Repartition biases were investigated using the DESeq2 package v1.28.1 for R (Love et al., 2014). Co-occurrence networks and

keystone taxa were resolved using the fast greedy modularity optimization method of the Molecular Ecological Network Analyzes Pipeline (MENAp) and the Cytoscape software v3.8.0 (Deng et al., 2012). Simple linear regressions were computed using the R package ggpubr v0.4.0 and illustrated with ggplot2 v3.3.2. Confrontation between microbial datasets and the explanatory variables were also computed using a multilinear regression model described in **Fig. S8a**.

3. Results

3.1. Similarities in land-use and human behaviors between sub-catchments

Socio-urbanistic indicators were recorded over the Mi-plaine Chassieu catchment, and were used to compare urban typologies and activities between sub-catchments (**Table S2b**; **Table S2c**). Overall, 559 industries were recorded over the Chassieu Mi-plaine catchment (**Fig. 1**). Most significant activities concerned retailing of specialized industrial materials (n = 63 companies), construction (n = 53), and metal transformation (n = 33) (e. g. **Fig. 2**). Hotels offering up to 170 rooms, and restaurants welcoming up to 300 customers during service hours, were found in this area. 139 companies distributed into 44 commercial sectors impacted this catchment, and were involved in wholesale trading (n = 15), specialized construction works (n = 10), financial services (n = 10), metal product manufacturing (n = 8), rental and leasing (n = 7) and real estate (n = 5) (**Table S2**).

The “land use” dataset was converted into a frequency table (**Table S2c**). Spearman correlations were computed from this dataset to analyze co-occurrence patterns (**Fig. 3**). This led to the observation of recurrent significant associations such as those of (i) the chemical and plastic industries, (ii) the entertainment – related ones, and (iii) the building and metal-work industries. Similarities between these partitions per sampling point were tested by computing adjusted Wallace indices. These indices of relatedness between sites are indicated in **Fig. 1**, and led to the identification of two large groups termed I (C01, C02, C05, C06, C09,

C13, C14, C17) and II (C22, C07, C19, C21, C15, C16), but five sub-watersheds showed specific urbanistic organizations as observed in **Table S2c**. Group I sampling points showed more industries involved in specialized construction works; these points were mainly found among quiet streets and dead ends. Group II sampling points were found distributed along a large avenue showing several rental and leasing services.

In order to go deeper in the understanding of variables impacting the urban surface microbiomes, technical objects and social behaviors over the catchment were investigated, and led to a dataset of 26 descriptors (**Table S2b**). Number of employees (total = 3421) and overall frequentations per sub-catchment were compared. Most significant movements over the catchment were associated with gasoline engines with a total number of 3880 events involving these engines per day (n=3 surveys; total = 11640 events) (**Table S2b**). Soft traveling modes such as walking (n = 308 events per day survey period) and cycling (n = 36 events per day survey period) were also frequent. Other technical objects mobilized at daytime and behaviors were observed such as high uses of portals (n=106 mobilizations per day survey period), construction engines (n=63), lorry for horses (n=46), bus shelters (n=27) or cigarettes (n=18). Wastes were also frequently detected, and concerned feces (12 observations per day survey period), food-derivatives (n=9) or hygienic products (n=5) (**Table S2b**). Chemical spills were observed including petrol and oil stains (n=58), clear fluid spills (n=6) or white spills (n=3). Spearman correlations were computed from these traces to highlight relationships between all socio-urbanistic parameters including the economic and industrial activities (**Fig. 3**). Interestingly, sampling points on streets with the highest numbers of gasoline engines ("trucks, "cars", "2 wheels", "commercial vehicles" and "bus") were correlated with restaurants and rental activities of group II (Spearman correlation tests; $p<0.05$) (**Fig. 3**). Areas with the highest numbers of pedestrians and employees were

associated with higher numbers of wastes. Furthermore, quiet streets and dead ends showed more fecal matters, and night activities such as prostitution (Wilcoxon test: p-value <0.05).

To avoid collinearity effects while performing statistical analyses between the socio-urbanistic and microbiological datasets, the following variables were selected (**Fig. 3**): (a1) number of cars; also representing numbers of buses, commercial vehicles, motorized two wheels and trucks, (b1) number of employees; also representing counts of cigarettes, pedestrians and domestic wastes, (c1) number of freight lifts; also representing counts of construction materials, liquid spills and bus shelters (**Fig. 3**). Analysis of these social traces led to the identification of two sectors with high gasoline engines circulation (North-West and South-East) (**Fig. 1 and 2**) which were termed "group III", and were composed of sampling points C02, C03, C05, C06, C07, C08, C15, C16, C17, C19 and C22. These sectors were associated with significantly higher numbers of trucks, cars, two-wheels, commercial vehicles, non-motorized scooters, and buses, but also cigarettes, food and other wastes, employees and pedestrians, and restaurants (Wilcoxon test: p-value <0.05). Dead end (C01, C14, C18 (no industrial or commercial activities)) and quiet streets (C09, C10, C11, C13, C20 or C21) were defined as group IV (**Fig. 1 and 2**) and were characterized by significantly higher numbers of hygienic wastes, prostitution and fecal matters (Wilcoxon test: p-value <0.05).

3.2. Variations in runoff volumes, physico-chemical monitoring and classical bacterial indicators between sub-catchments

Geographical origin of the clouds and rain events which led to the three runoff sampling campaigns investigated in this study are indicated in **Fig. S1**. Runoff waters were subjected to physico-chemical monitorings (**Table S3**). A PCA of these datasets was performed (**Fig. S3**). pH values of the runoff waters of campaign 2 were higher, and their turbidity was lower, in most instances. Campaign 2 had the largest differences between the measured min-max values

for pH (pH 6.3 – 9.1). Lowest value was obtained from runoff waters recovered at site C14, and appeared related to an industrial chemical spill. This pH shift was associated with a drop in turbidity (lowest value recorded in this study) but an increase in conductivity (highest value recorded), explaining the peculiar PCA ordination of this sampling point (**Fig. S3**). A negative correlation between pH and the presence of chemical industries was observed (C14, C2) (Spearman correlation tests; $p<0.05$). High pH values were correlated to rental and leasing activities such as those of group II sampling points (**Fig. 3**) (Spearman correlation tests; $p<0.05$) (**Fig. S4)**. High conductivity values were correlated to food catering activities (Spearman correlation tests; $p<0.05$) (**Fig. S4**), which were positively correlated to the number of food wastes on the streets (**Fig. 3**). Conductivity and turbidity values were positively correlated with the numbers of cars, trucks, commercial vehicles and pedestrians observed on the catchment (**Fig. S5**). Samples from campaign 2 had the lowest temperatures, with an average of 8.5ºC, while the runoffs had temperatures between 16-18ºC for samples of the other campaigns (**Table S3**). Runoff flow values and total volumes were similar between sampling campaigns 1 and 3 but campaign 2 had lower values (**Table S3**).

Concerning the bacterial indicators, runoffs from the watershed showed high level of heterotrophic bacterial plate counts ranging from 4.0×10^5 to 1.8×10^8 CFU (colony forming units) per 100 mL (**Table S3**). These plate counts were positively correlated with sub-group b1 variables (employees, cigarettes, pedestrians, wastes, **Fig. 3**), and merchandise storage (**Fig. S6**; Spearman correlation tests; $p<0.05$). *E. coli* and intestinal enterococci numbers were high all over the catchment (**Fig. 1; Table S3**), and found to be correlated (Spearman correlation test; $p<0.05$). Intestinal *enterococci* cell numbers were negatively correlated with group b1 variables but also merchandise storage and non-motorized scooters (**Fig. S6**, Spearman correlation tests; $p<0.05$). *E. coli* and intestinal enterococci numbers were highest among sites of group IV (dead ends, hygienic wastes, and fecal matters) (Wilcoxon test;

p<0.05), and were associated with a higher occurrence of HF183 DNA marker confirming significant human fecal contaminations at these sites (Spearman correlation test; p<0.05) (**Fig. 1, Table S3**). Twenty-two samples over the three sampling campaigns were found positive to HF183 by qPCR (**Table S3**). Relative ratios of HF183 marker (of human fecal contamination) over numbers of total bacterial 16S rRNA gene targets ranged from 3.1 x 10^{-6} to 3.7 x 10^{-3} (at sampling point C23 which is the outlet of the watershed). The relative numbers of HF183 DNA targets were positively correlated with the mobilization of portals. This technical object was mainly used among quiet places or secondary roads of group IV sites (Spearman correlation tests; p<0.05). Presence of integron 1 (*int1*) was detected among 24 runoff samples (**Fig. 1**). Relative *int1* counts over total 16S rRNA gene copies ranged from 6.2 x 10^{-7} to 3.5 x 10^{-3} (at sampling point C23). Integron 1 was mainly detected among group III sites associated with high counts of gasoline engines (Wilcoxon test; p<0.05).

3.3. Core K- and opportunistic r-strategists observed over urban surfaces (using runoff waters) through 16S rRNA gene meta-barcoding analyses

- general features

Removal of chimeric sequences, short reads and a random resampling based on the lowest number of V5-V7 16S rRNA gene reads (n=10783) recovered for a runoff water sample were performed. This led to a dataset of 4 048 699 16S rRNA gene reads. These reads were divided into 9155 OTUs. Rarefaction curves showed a plateau phase in the relation between the number of reads analyzed and the number of OTUs observed per sample (**Fig. S7a**). Most samples had between 700 to 1000 OTUs (**Table S8**). Shannon, Simpson and Evenness indices showed the samples from the 1st campaign to have a significantly higher diversity than those of the 2nd and 3rd campaigns (Wilcoxon test; $p < 0.05$), and their content in 16S rRNA gene copies per 100 mL were slightly lower (**Table S3**). These indices showed that samples from group I (construction industries) (**Fig. 1**) had significantly more diversity than those of group

II (large avenue) (Wilcoxon test; $p < 0.05$) but not of groups III (high motor engines) and IV (quiet streets). NMDS analysis confirmed the significant relationships between the socio-urbanistic groupings (defined in section 3.1) and the 16S rRNA gene meta-barcoding profilings. All four socio-urbanistic groups defined above were associated with significantly different profiles as computed from Adonis analyses using the Bray-Curtis dissimilarities between samples **(Fig. 4**).

- standard partition analyses

Runoff OTUs were affiliated to bacterial taxa using the Silva database in order to identify the core K- and versatile r-like bacterial taxa thriving in the investigated urban system. Thirty-five phyla were inferred from the dataset (for more than 99% of the reads; **Table S10**). Sample C14_2 (impacted by a chemical spill) harbored the smallest number of phyla (n=11). Sample C16_1 (group III, **Figs. 1 & 4**) showed the greatest number of phyla (n= 24). Nine core phyla were found over the industrial catchment: *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Gemmatimonadetes*, *Planctomycetes*, *Acidobacteria*, *Firmicutes*, *Saccharibacteria* and *Armatimonadetes*. Reads allocated to the *Proteobacteria*, *Actinobacteria* and *Bacteroidetes* represented in average 93.6 % of the dataset and varied respectively from 39 % to 81 %, 0.70 % to 54 % and 2% to 49 % per campaign (**Table S10**). *Proteobacteria* was the dominant phylum (showing an equivalent of 2×10^9 to 2×10^5 16S rRNA gene copies per 100 mL; median value = $2{,}4 \times 10^8$) in all samples except C07_3 and C08_3 which were dominated by *Actinobacteria*. It is to be noted that C07 and C08 are part of the group III sites (**Figs. 1 & 4**) associated with high traffic of gasoline engines. In average, *Actinobacteria* were the second most dominant group over the catchment with 19.75 % of relative abundance followed by the *Bacteroidetes* (15.48 %). However, *Bacteroidetes* were more abundant than the *Actinobacteria* in 29 samples over 60 (**Table S10**). This dataset was characterized by a very low number of *Firmicutes* with scores ranging from 0.02 to 3.5% per

374 sample representing an equivalent of about 4 x 10^3 to 2.5 x 10^7 16S rRNA gene copies per
375 100 mL (median=1 x 10^6). Sample C14_2 (showing a toxic chemical spill) was again
376 characterized by unusual values with a very low relative number of reads allocated to the
377 *Actinobacteria* (0.7 %; highly sensitive to the chemical spill) but very high (more than 81%;
378 representing an equivalent of about 10^8 16S rRNA gene copies per 100 mL) for reads
379 allocated to *Proteobacteria* (likely more resistant to the chemical spill). Similar trends were
380 observed for sample C14_3 but this sample also showed the highest score for
381 *Saccharibacteria* (5.40 %) of the TM7 phylum which are related to obligate epibiont of
382 *Actinomyces* spp. (He et al., 2015).

383 Deeper analysis of these phyla showed a differentiation into 717 genera (**Table S11**).
384 Twenty-eight of these genera including the *Pseudomonas* (representing an equivalent of about
385 3 x 10^3 to 3.8 x 10^7 16S rRNA gene copies per 100 mL; median=7 x 10^5) were conserved
386 among all samples, representing almost 90 % of the total 16S rRNA gene reads, and were thus
387 considered to be related to K-strategists. **Fig. 5** shows the genera with a relative abundance
388 over 0.5 % over the dataset, and highlights those which were classified as core bacterial taxa
389 (in all samples) likely grouping K-strategists. The C14 sampling site was confirmed to be
390 peculiar with high numbers of reads allocated to *Novosphingobium* (α-proteobacteria) (**Table**
391 **S6; Table S11**). C14 also showed high numbers of DNA reads allocated to the
392 *Sulfurospirillum* (ε-proteobacteria) (5.61 %) and *Paludibacter* (*Bacteroidetes*) (11.6 %).
393 These latter enriched genera were considered r-like opportunistic strategists.

394 Distribution biases of the above genera against the behavioral (groups III or IV; **Fig. 2**) and
395 industrial/commercial activities (groups I to II; **Fig. 1)** were investigated by DESeq2 (**Table**
396 **S12).** Reads from *Novosphingobium* were found in significantly higher proportions in the
397 runoff water samples from group I (construction industries) and IV (quiet areas) (DESeq2;
398 $p<0.05$). Reads from *Actinobacteria* such as *Geodermatophilus, Marmoricola,* and *Dietzia,*

399 and from *Methylobacterium* (α-proteobacteria) were found in higher proportions in samples
400 from group II (large avenues) (DESeq2; $p<0.05$). Reads affiliated to *Bdellovibrio* were found
401 in significantly higher proportions in the samples from group I (DESeq2; $p<0.05$), and those
402 from *Roseomonas* and *Arthrobacter* in group III (DESeq2; $p<0.05$). Reads allocated to
403 *Pseudomonas* and *Gemmatimonas* were in higher proportion among group IV samples
404 (DESeq2; $p<0.05$). These strong relationships with particular socio-urbanistic patterns are
405 indicative of the occurrences of K-like strategists over the catchment.

406 The sources of the above taxa were explored by comparing their relatedness with taxa in
407 higher numbers (i) among old urban sediments (*Acidibacter*, *Haliangium*) that had
408 accumulated metallic trace elements and naphthalene, or (ii) among recent urban sediments of
409 a detention basin which had high chrysene contents (*Aquabacterium*, *Paludibacter*) according
410 to Marti et al. (2017). 16S rRNA gene reads allocated to taxa known to be enriched in old
411 urban sediments were in very low relative abundances among the sampled runoff waters
412 (**Table S11**). However, reads affiliated to taxa in higher numbers in recent sediments/deposits
413 found in a detention basin (*Peredibacter*, *Aquabacterium*, *Paludibacter* and *Cloacibacterium*)
414 were recorded among the sampled runoffs, with relative abundances going from 0.20 to 3.4 %
415 (**Table S11**). These taxa did not show distribution biases matching the socio-urbanistic
416 parameters of the sub-catchments, and were likely to contain K-strategists.

417 ***- correlation network analyses***

418 Co-occurrence patterns between numbers of 16S rRNA gene reads and their allocation to
419 bacterial genera were further explored through multiple Spearman correlation tests (**Fig. 6a**),
420 and were considered indicative of the occurrence of K-strategists. Statistics of the most
421 significant networks are presented in **Table S14**, and the modules are shown in **Fig. 6a**. The
422 largest module of correlated taxa was made of 17 genera mainly grouping *Actinobacteria*
423 involved in hydrocarbon degradation (based on Faprotax and literature searches). Only two

genera from another phylum were recorded in this module and belonged to the *Roseomonas* and *Methylobacterium* (*α-proteobacteria*). This module was in line with the above DESeq2 analyses which had previously shown positive relationships between some of these genera and the group II urban typologies (**Fig. 1 and Fig. 3;** large avenue with high numbers of gasoline engines). The second largest module (Mod2) showed correlations between three *Bacteroidetes* (*Cytophaga*, *Flavobacterium* and *Pedobacter*) and a connection between *Flavobacterium* read numbers and those allocated to *Pseudospirillum* (hydrocarbon degrader) of the *γ-proteobacteria* which were further connected to *Perlucidibaca*, and *Arenimonas* 16S rRNA gene reads distribution patterns (**Fig. 6a**). These last three genera are known to harbor bacterial species involved in oil / hydrocarbon degradation processes (Morais et al., 2016). Module-EigenGene analyses highlighted several significant (p-value < 0.001) Spearman correlations between the selected explanatory variables and the above modules (**Fig. 6a**). For example, module 1 (Mod1) showed positive correlations with sub-group a1 (gasoline engines) variables represented by car numbers (ratio=0.45), and which contributed to the definition of group III socio-urbanistic patterns (**Fig. 2**). Genera from Mod1 showed negative relationships with hygienic wastes (-0.47) associated with group IV sampling sites located in more quiet streets (**Fig. 2** and **Fig. 6a**). Mod2 genera were positively correlated to higher uses of construction engines (0.69) and freight_lift – like variables (0.49) of sub-group c1, and, as Mod1, were negatively associated with hygienic wastes (-0.46).

Further computations using a multilinear regression model (described in **Fig. S8a**) were performed to confirm these relationships between the meta-barcoding, behavioral and organizational explanatory variables (**Fig. S8a**). Interestingly, this analysis showed that core genera likely harboring K-strategists had lower numbers of significant R^2 correlation values than those of the flexible r-like bacterial taxa (Wilcoxon-test: p-value < 0.001). Reads from seven r-like genera including *Thauera*, *Sulfurospirillum*, *Noviherbaspirillum*, hgcl_clade,

Novosphingobium, *Microcella*, and *Phenylobacterium* showed significant correlations with the physico-chemical variables ($R^2 > 0.1$). Number of reads of *Sulfurospirillum* were negatively correlated with the pH values ($R^2 = -0.14$), and positively correlated with the conductivity ones ($R^2 = 0.08$) (**Fig. S8a**), confirming an opportunistic development matching with the occurrence of a chemical spill at C14. Read numbers affiliated to *Novosphingobium* also showed a negative relationship with the pH values. *Thaurea* read numbers were negatively correlated to temperature and positively with turbidity. Several other significant relationships were detected but with low R^2 (**Fig. S8c**).

3.4. Core K- and opportunistic r-strategists observed over urban surfaces (using runoff waters) through *tpm* gene meta-barcoding analyses

- general features

A high quality dataset of 910 048 *tpm* sequences was generated, re-sampled (n=8087), and used to infer the effects of urban land use typologies and human behaviors on *tpm*-harboring bacterial species. These *tpm* sequences were divided into 9 641 OTUs (exact sequence variants) (**Tables S7 and S8**). Rarefaction curves showed a plateau phase (**Fig. S7b**). Most samples had between 150 and 300 *tpm* OTUs. Diversity indices have been computed and are shown in **Table S8**. These diversity indices were compared with those computed from the 16S rRNA gene meta-barcoding dataset through linear regression analyses. No correlations were observed for the Shannon, Simpson and Evenness indices. In the same way, no significant differentiation was observed between diversity indices computed from the three sampling campaigns and the four socio-urbanistic groups of sampling points defined in section 3.1.

NMDS and RDA analyses of the Bray-Curtis dissimilarities between the *tpm* gene read patterns were performed. Adonis analyses did not differentiate the *tpm* profilings according to the groups of samples defined from the analyzed urban typologies and activities (groups I to IV). However, RDA indicated a significant relationship between the ordinations and the

activities of group III sampling points such as high traffic of gasoline engines. Anova.cca and ordistep tests confirmed the significance of the computed RDA model and relationships with traffic ($p<0.01$). To illustrate, the position of *tpm* Otu00135 affiliated to *Pseudomonas extremaustralis* in the RDA plot showed higher read numbers among the runoff waters of group III samples. This OTU showed positive relationships with cars (and correlated variables shown in **Fig. 3**) and the quantity of food wastes observed over the watershed. Similar positive relationships on the RDA were recorded for *tpm* OTU00154 (a non-classified *Pseudomonas* species). The goodness of fit test confirmed the significance of these relationships.

- standard partition analyses

Taxonomic allocations at the genus level allowed a classification of 62 % of the *tpm* reads into 20 well-defined genera (**Table S15**). The genera inferred by the *tpm* meta-barcoding approach represented on average an equivalence of 6.28 % (0.35 to 21.53 % per sample) of the 16S rRNA gene dataset (computed from the shared genera between both datasets). Only *Pseudomonas tpm* reads were recorded in all the runoff samples. They represented in average 35.79 % of the total *tpm* reads (from 0.21 to 75.93 %) (**Table S15**). Other *tpm*-harboring genera were found among at least half of the runoff samples such as *Herbaspirillum*, *Xanthomonas* and *Aeromonas* representing respectively 11, 5 and 0.8 % of the total *tpm* reads. These four genera were considered core *tpm*-harboring taxa circulating over the investigated urban catchment likely harboring K-strategists (**Table S15**).

Relative abundances inferred from the *tpm* gene dataset showed the *Pseudomonas* to be fairly homogenous across the sampling points except for C16_S2 (0.21 %; group III variables associated with traffic of gasoline engines), C10_S3 (5.41 %; group IV variables associated with quiet streets) and C09_S1 (75.93 %; group I (construction industries) and IV). *Herbaspirillum* was not detected among the 3rd sampling campaign except for C05_S3 and

C06_S3 but with a very low relative abundance (0.01 and 0.02 %), and a high one for C14_S3 (31.48 %) (**Table S15**). Interestingly, the pollution observed at C14 was associated with an enrichment in *Herbaspirillum*. C16_S2 (96.35 %) also showed an environmental burst in *Herbaspirillum tpm* reads (**Table S15**). *Xanthomonas tpm* reads also had a heterogeneous distribution pattern with relative abundances going from 0 to 50% (**Table S15**). The 3rd sampling campaign led to high relative abundances of *Xanthomonas* reads at C01_S3 (48.86 %), C03_S3 (33.81 %) and C07_S3 (20.72 %). *Xanthomonas* reads were not detected during the 2nd sampling campaign except at C18_S2 but in low abundance (0.08 %). This approach also led to the recovery of *Stenotrophomonas* (0.65 %), *Shewanella* (0.30 %) and *Curvibacter* (0.2 %) *tpm* DNA reads (**Table S15**).

At the species level, more than 50 % of the *tpm* reads could be allocated to 78 well-defined species (**Table S16; Fig. 7**). DNA reads from several species were recovered in more than 50 % of the samples such as *Herbaspirillum aquaticum*, *Pseudomonas koreensis*, *Pseudomonas rhodesiae*, *Pseudomonas stutzeri*, *Pseudomonas anguilliseptica* and *Pseudomonas aeruginosa* (**Fig. 7**). These species were considered K-strategists. The most abundant species was *Herbaspirillum aquaticum* (**Table S16**). *H. aquaticum tpm* reads represented 31.32 % of the total number of reads obtained from the C14_S3 sample. *Pseudomonas koreensis* was the second most abundant species with an average relative abundance of 4.7 %, followed by *Xanthomonas cannabis* (3.85 %), *Pseudomonas anguilliseptica-like* (2 %), and *Pseudomonas aeruginosa* (2 %) (**Table S16, Fig. 7**). Other *tpm*-harboring species showing more variable distribution patterns like *Aeromonas jadensis*, *P. putida*, *P. mendocina*, *P. pseudoalcaligenes* and *P. fluorescens* appeared r-like strategists taking benefits for their development of some specialized C-sources available over the catchment. These latter bacteria could, in fact, be associated to hydrocarbon degradation as inferred from Faprotax analyses. It is to be noted that some of the above r/K-like taxa detected by the *tpm* meta-barcoding approach were found

to be well-known human, animal and plant pathogens e. g. *A. hydrophila*, *P. mendocina*, *A. caviae*, *P. aeruginosa*, *P. syringae*, *Stenotrophomonas maltophilia*, *Xanthomonas axonopodis*, and *X. cannabis*. The latter species was the most common *tpm*-harboring phytopathogen recorded in the samples, and was previously described to infect more than 350 plant species including *Cannabis sativa* (Netsu et al., 2014).

The *tpm* reads allocated to *P. aeruginosa* human pathogens were recurrent in the runoff *tpm* dataset. They were recorded in high numbers at C21_S2 (19.25 %; group IV), C07_S1 (10.36 %; group III), C22_S1 (10.06 %; group III) and C19_S3 (9.90 %; group III) (**Table S16**). Conversion of these latter relative abundances into *Pseudomonas* 16S rRNA gene copies according to **Table S3** indicated an equivalence of about 10^4 *P. aeruginosa* 16S rRNA gene numbers per 100 mL in these samples. These *P. aeruginosa tpm* reads were further allocated to sub-lineages, and mapped over the urban watershed (**Fig. S9**). These reads (n = 5594) were divided into 72 OTUs, and 26 were allocated to the sub-clade PAO1 (n = 2273 reads), and 20 to the sub-clade PA14 (n = 214 reads). The *P. aeruginosa tpm* sequence types were allocated to strain characterized by Multi Locus Sequence Typing (https://pubmlst.org/) (**Table S17a**) using the correspondence table reported by Colin et al. (2020). The *tpm* OTU00592 (n = 181 reads) was found to match ST252 isolates from non-CF (cystic fibrosis) infections; the one of OTU00987 (n = 85 reads) matched ST2123/ST226 strains respectively isolated from CF patients and clinical environments; and OTU00072 (n = 1802 reads) matched ST175/ST2042 found in urinary tract and blood infections. This significant occurrence of *P. aeruginosa tpm* reads over the catchment led us to isolate the matching bacterial strains during the third sampling campaign. More than 200 hundred isolates were collected, and their *tpm* sequences were sequenced after PCR amplification (**Table S17b**). Most of the *P. aeruginosa* urban isolates harbored the *tpmG* type 12_15_17_18_35 which was found to be the most abundant among the *tpm* metabarcoding dataset. However, strains from

549 some other dominant meta-barcoding *tpm* sequence types were found more difficult to isolate.
550 Only 7 strains harboring the *tpmG* type 08_24 (matching Otu00056 which showed 2256
551 sequences) were obtained, and 2 showing the *tpmG* type 1 sequence (matching OTU021
552 which showed 1878 sequences). It is to be noted that OTU numbers were attributed by using a
553 larger *tpm* meta-barcoding dataset than the one reported in this study, and this explains the
554 lack of match between OTU code numbers and number of reads per OTU among the dataset
555 presented here. Interestingly, some strains harbored *tpmG* types not detected by
556 metabarcoding such as *tpmG*3, *tpmG*4, *tpmG*16-26, *tpmG*30 and *tpmG*34 (**Table S17**). The
557 *tpm* reads allocated to *P. syringae* could also be further allocated at the level of pathovars
558 such as *P. s.* pathovar *aceris* (3875 reads distributed into 84 OTUs), *P. s. panici* (4 OTUs, 203
559 reads), *P. s. japonica* (3 OTUs, 167 reads), *P. s. aptata* (1 OTUs, 46 reads), *P. s. maculicola*
560 (1 OTUs, 26 reads) and *P. s. pisi* (1 OTUs, 7 reads). DESeq2 computations were performed to
561 evaluate changes in the *tpm* OTU patterns according to the urban typologies and activities.
562 Otu00056 affiliated to *P. aeruginosa* (subclade PA14-like) showed significantly more reads
563 among group I (large avenue) than group II, and among group III (high gasoline engines) than
564 IV. Otu00072 (subclade PAO1-like) reads were more abundant among samples from group
565 IV (those from dead ends) **(Table S18)**.

566 The sources of *tpm* bacterial taxa among runoffs were explored by DNA sequence
567 comparisons with previously reported sequence types recovered from urban deposits and
568 sediments accumulating at C23 (which is the detention basin of the experimental site
569 collecting the runoff waters) (Aigle et al., 2021). 17 OTUs from the runoff water datasets
570 matched at 100 % identity with *tpm* sequences from sediments. Runoff *tpm* of Otu00013
571 (unclassified *Proteobacteria*, n=5891 reads), Otu00027 (Gp_DE, n=3791 reads) matched
572 respectively with ASV_17, ASV_32 found in old deposits containing high concentrations of
573 poorly degradable pollutants including metallic trace elements (Cd, Cr and Ni) and

naphthalene (**Table S19**). 14 additional *tpm* runoff OTUs matched with sequences reported among recently mobilized urban deposits and containing high chrysene concentrations. To illustrate, Otu00002 (*Pseudomonadaceae*, n=11433 reads), Otu00016 (unclassified *Gammaproteobacteria*, n=5291), Otu00026 (unclassified *Gammaproteobacteria*, n=4410), and Otu00028 (Proteobacteria, n=4388) sequences from runoff water matched Otu013, ASV_38, ASV_51 and ASV_35 of Aigle et al. (2021) (**Table S19**). Otu00036 affiliated to *P. syringae* pathovar *aceris* showed higher number of reads among old sediments containing high concentrations of naphthalene, cadmium, chrome and nickel (Aigle et al., 2021).

- correlation network analyses

Co-occurrence networks computed by the MENA approach led to the identification of K-like bacterial modules (termed Mod-*tpm*) supported by Spearman Rho factors > 0.7 (**Table S21; Fig. 6b**). Mod1-*tpm* showed significant (p-value < 0.001) positive correlations with the relative quantities of faeces (0.42) and hygienic wastes (0.42) differentiating social behaviors associated with group IV (**Fig. 1**). Mod2-*tpm* species were negatively correlated with faeces (-0.50), and Mod3-*tpm* species showed positive relationships with merchandise storage activities (0.51) (**Fig. 6b**). The freight lift activities representing the c1 variables (**Fig. 3**) were negatively related to Mod4-*tpm* (-0.90). Multiple linear regression analyses (performed as described in **Fig. S8a**) were run to find r-like distribution patterns in the *tpm* dataset (**Fig. S10**). Numbers of *tpm* reads allocated to several species were found correlated to physico-chemical parameters: (i) those of *Xanthomonas hortorum* and *Pseudomonas amygdali* were negatively correlated to conductivity values (R^2 = -0.023 and -0.05), and (ii) those of *Pseudomonas sp. Gp_BS* were positively correlated to turbidity (R^2 = 0.05).

4. Discussion

Urban surface microbial communities are expected to be made of (i) opportunistic colonizers (also termed r-strategists) coming from exogenous sources, and transiently taking

benefit of the nutrients available in these systems, and (ii) of core taxa (termed K-strategists) adapted to the living conditions prevailing among the colonized catchment including high variations in temperature, water content, and chemical pollutants. The ecology of these core and opportunistic taxa among urban biomes remain to be defined in order to avoid growth of undesirable species and hazards such as human exposure to pathogens. Furthermore, the spread and impact of synurbic bacterial taxa over the earth need to be evaluated. The Anthropocene era might be gradually leading to a major invasion of the earth's ecological systems by these taxa. To evaluate these impacts, the core components of urban microbiomes need to be defined, and be tracked over other systems.

Field surveys were carried out to identify key socio-urbanistic variables impacting the investigated catchment and related microbiomes. This led to the identification of sub-catchments (sampling points) with common economical and industrial activities, leading to similar social behaviors, that could be used as references to infer relationships with particular bacterial taxa. More than 500 industries, and more than 10 000 observations were considered to assess the main socio-urbanistic forces impacting the investigated urban catchment. This catchment was found to be divided into two groups (termed I and II) of industrial / commercial activities, and two groups (termed III and IV) of social behaviors. Interestingly, streets with high numbers of gasoline engines (group III sites) were significantly associated with restaurants while areas with high numbers of pedestrians were associated with higher numbers of wastes. Furthermore, quiet streets and dead ends showed more traces of fecal matters, and night activities (group IV). Previous investigations have shown that traffic was a major contributor of surface sediments and pollutants (Revitt et al., 2014). The above groups and panels of activities were used to investigate relationships with FIB, MST (microbial source tracking) DNA markers, and DNA meta-barcoding datasets. All these datasets

confirmed the informative character of runoff waters to estimate bacterial diversity over sub-catchments, and identify K-bacterial strategists living in an urban catchment.

Relationships between socio-urbanistic features and runoff contents in classical bacterial indicators

Significant contents in FIB were observed all over the investigated catchment among the runoff waters. These taxa (*E. coli* and intestinal *Enterococci*) were clearly core K-like bacterial strategists because of recurrent fecal contaminations coming from multiple sources including man, making available the appropriate C-sources for their survival and growth. Their high values were in line with those of Paule-Mercado et al. (2016) which showed an increase in FIB among urban settings, and related these increases to surface impermeabilization. Group IV sub-catchments (quiet streets) showed the highest FIB numbers in their runoffs, and were associated with the highest occurrences of HF183 MST DNA signatures indicative of human fecal matters. Group a1 variables grouping all sorts of gasoline engines, and variables associated with merchandise storage areas were negatively correlated to FIB contents of the runoffs. These results are in agreement with a previous study which showed road pollutants such as PAHs to be negatively correlated to FIB contents (Bernardin-Souibgui et al., 2018). Such pollutants were suggested to reduce their numbers (Ukpaka et a al., 2014). In fact, only variables indicative of the occurrences of wastes and fecal matters were found positively associated with high FIB contents in the runoffs. These observed high contents of FIB in city runoff samples were not surprising, and similar trends had previously been reported for other urban contexts (Ellis, 2004; Lee et al. 2020; Ahmed et al., 2010; Sauer et al., 2011; Sidhu et al., 2012; Chong et al., 2013). However, the relationships between FIB, urbanistic issues, and social behaviors were inferred for the first time.

Integrons of class I, previously shown to have an abundance positively correlated with the ones of urban pollutants (Wright et al., 2008), were also observed in the collected runoff

waters. These integrons can harbor resistance genes toward antibiotic, metallic elements and disinfectants. Class I integron DNA targets were detected in about 40% of the runoff water samples, and showed a higher occurrence among group III sites associated with high numbers of motor-engines generating petrol-related pollutants. As expected, the distribution of integron 1 genetic elements did not match the one of fecal pollutions, and confirmed that runoff waters DNA imprints can differentiate ecological trends between sub-catchments.

K-like bacterial strategists among urban surface microbiomes

- **Inferences from 16S rRNA meta-barcoding analyses**

As indicated above, *E. coli* and intestinal enterococci were clearly demonstrated, in this study, to be part of the K-bacterial taxa colonizing urban surfaces. DNA meta-barcoding analytical schemes further confirmed the success of such fecal bacterial taxa over the city. In fact, fecal bacterial DNA reads reported by McLellan and Roguet (2019) and Noyer et al. (2020) among the 16S rRNA meta-barcoding dataset could be tracked, and their distribution patterns were in line with the ones of the FIB and HF183 MST marker for human fecal contaminations. The 16S rRNA gene reads from *Bacteriodes*, *Cloacibacterium,* and *Macellibacteroides* showed a high prevalence over the catchment, and were found to be part of the core urban taxa likely harboring K-strategists. However, DNA sequence reads allocated to fecal *Arcobacter* and the *Ruminococcus* showed more variations in their distribution patterns, and these genera were considered to be part of the r-opportunistic taxa.

Overall, twenty-eight core genera likely harboring K-strategists were found conserved among the sampling points of the catchment. Most of these genera had previously been reported among other urban settings such as *Blastococcus*, *Massilia*, *Flavobacterium*, and *Sphingomonas* (McLellan et al., 2015; Marti et al., 2017*;* Leung et al., 2014; Robertson et al., 2013). Nevertheless, a few discrepancies in the bacterial taxa distribution patterns between those recorded in this work and other reports were observed. In fact, *Spirosoma, Alkanindiges, Rubellimicrobium*, and *Novosphingonium* 16S rRNA gene reads were recorded all over the investigated catchment but not always in the other urban settings. Conversely, some taxa had been described as recurrent in other reports among the urban microbiome but here showed more versatile r-like patterns as observed for *Acidibacter*, *Psychrobacter* and *Propionibacterium*. This is in line with the hypothesis of a selection of specialized taxa by certain urban morphotypes as suggested by Saxena et al. (2015). Here, the investigated

morphotype showed high industrial and economical activities with little green areas and no residential zones. This morphotype was previously shown to be associated with a high occurrence of chemical pollutants and related microbial degraders through the analysis of sediments and deposits transferred into the detention basin located at the extremity of the catchment (Aigle et al., 2021; Marti et al., 2017). Correlation network analyses confirmed this conclusion. The largest modules of correlated distribution patterns between genera (mainly made of *Actinobacteria*) was associated to hydrocarbon-degraders. The most significant K-like actinobacterial taxa among these modules were *Modestobacter, Williamsia, Marmoricola,* and *Blastococcus,* but also involved three *Bacteroidetes* (*Cytophaga*, *Flavobacterium* and *Pedobacter*) and a *γ-proteobacteria* (*Pseudospirillum*). This module was positively correlated with variables of group II (large avenue) and III (high number of gasoline engines), but was inversely correlated to group IV variables related to hygienic wastes. It is to be noted that bacterial taxa previously recorded in high numbers among road sediments/deposits (*Peredibacter*, *Aquabacterium*, *Paludibacter* and *Cloacibacterium*) by Marti et al. (2017) were found among the sampled runoffs. The broad distribution pattern of these taxa confirmed their tropism for city surface habitats.

- **Inferences from *tpm* meta-barcoding analyses**

The *tpm* meta-barcoding marker allowed a deeper characterization of some of the K core taxa down to the species level. Nine species were highly conserved over the catchment: *Pseudomonas koreensis*, *Herbaspirillum aquaticum*, *P. rhodesiae*, *P. stutzeri*, *P. syringae*, *P. anguilliseptica* and *P. aeruginosa*. Interestingly, three of these species, *P. stutzeri*, *P. aeruginosa*, and *P. koreensis*, were also part of the most common taxa observed in urban transit systems (Danko et al., 2021), confirming their synurbic status. Some of these K core species were previously shown to use hydrocarbons as C-sources such as *P. aeruginosa* and *P. stutzeri*. Moreover, *P. koreensis* was reported to degrade, among others, hexadecane,

engine oil, pyrene and phenanthrene (Bučková et al., 2013). *H. aquaticum* and *P. stutzeri* are also known producers of polyhydroxyalkanoates and polyhydroxybutyrate storage molecules allowing survival under extreme growth conditions (Yan et al., 2008; Langenbach et al., 2006; Timm and Steinbuchel, 1992) such as those occurring over road surfaces. *H. aquaticum tpm* reads were highest among the most polluted sites (those showing a drop in pH values). These bacteria are known to be plant growth-promoting rhizobacteria (Marques et al., 2015) but had not been associated with urban pollutants, so far. Correlation network analyses were performed with the *tpm* metabarcoding dataset to further evaluate the occurrence of K-like strategists, and revealed a large module of correlated species associated with hygienic wastes.

r-like bacterial strategists among urban surface microbiomes

Multilinear regression models and DESeq2 analyses highlighted r-like genera. Point source pollutions revealed by changes in pH values were related to environmental bursts of *Novosphingobium* and *Sulfurospirillum* in the runoff waters of a few sub-catchments. These bursts were associated with an unusual drop in *Actinobacteria* (0.7 %). The BASOL database of the French ministry of the ecological and solidarity transition (https:// basol.developpement-durable.gouv.fr/) indicated a recurrent pollution by trichloroethylene and halogenated solvents in this area. *Sulfurospirillum* was previously found to be enriched by organohalides, arsenate and selenite, and chlorinated ethenes (Luijten, 2003; Nijenhuis et al., 2005; Buttet et al., 2018). Similarly, *Novosphingobium* was previously found to be enriched by phenol, aniline, nitrobenzene and phenanthrene (Liu, 2005; Sohn, 2004; Jiao et al., 2017).

RDA of the *tpm* meta-barcoding dataset also revealed r-like taxa such as *Pseudomonas extremaustralis*. In fact, one OTU of this species showed positive relationships with car numbers and the quantity of food wastes over the catchment. This species was previously shown to degrade phenols and alkanes (Tribelli et al., 2012). Other r-like *tpm*-harboring species likely taking benefits from the occurrence of hydrocarbon pollutants over the

730 catchment were further detected by Faprotax analyses such as *Aeromonas jadensis*, *P. putida*,
731 *P. mendocina*, *P. pseudoalcaligenes* and *P. fluorescens*. Furthermore, three OTUs showed a
732 distribution pattern indicative of an opportunistic development among old deposits containing
733 high concentrations of poorly degradable pollutants including metallic trace elements (Cd, Cr
734 and Ni) and naphthalene, and fourteen among recent deposits/sediments previously shown
735 enriched in chrysene content (Aigle et al., 2021).

736 **Microbiological hazards associated with urban surface microbiomes**

737 Previous reports suggested an increase in virulence gene loads among K-strategists from
738 highly selective systems (e. g. Song et al., 2017). Bacterial pathogens were thus searched
739 among the *tpm* meta-barcoding dataset, and their distribution biases were analyzed. The
740 following bacterial pathogens were recorded over the catchment: *Aeromonas hydrophila*, *A.*
741 *caviae*, *Stenotrophomonas maltophilia*, *P. aeruginosa, P. anguilliseptica, P. mendocina*, *P.*
742 *syringae*, *P. amygdali*, *Xanthomonas axonopodis*, *X. hortorum* and *X. cannabis*. The latter
743 species, *X. cannabis,* was the most common *tpm*-harboring phytopathogen recorded in the
744 samples, and was previously described to infect more than 350 plant species including
745 *Cannabis sativa* (Netsu et al., 2014). However, most of these *tpm*-harboring bacterial
746 pathogens appeared to have r-like distribution patterns except *P. aeruginosa* and *P. syringae*.
747 Viable *P. aeruginosa* cells had previously been recorded among urban sediments of the
748 detention basin of this catchment at about 10^5 CFU g dw^{-1} (Bernardin-Souibgui et al., 2018).
749 Furthermore, *P. aeruginosa* was found to have a conserved distribution pattern in these
750 sediments, whatever their pollution content, through the *tpm* meta-barcoding approach (Aigle
751 et al., 2021). This latter species was thus clearly part of the K strategists being able to survive
752 and grow over the city surfaces. Interestingly, variations in *P. aeruginosa* cell counts or of
753 their matching *tpm* DNA read numbers were not found correlated with the FIB dataset
754 suggesting distinct habitats over the catchment.

Urban socio-urbanistic patterns associated with the distribution of the *tpm* sequence types of *P. aeruginosa* and *P. syringae* were investigated at the sub-species level. This led to the first observation of *P. s.* pathovar *aceris, P. s. panici*, *P. s. japonica*, *P. s. aptata*, *P. s. maculicola*, and *P. s. pisi* among urban runoffs. Otu00036 affiliated to *P. syringae* pathovar *aceris* previously showed higher number of reads among old sediments containing high concentrations in naphthalene, cadmium, chrome and nickel (Aigle et al., 2021). However, this latter OTU did not show any particular distribution biases over the catchment. Regarding *P. aeruginosa*, using the correspondence table reported by Colin et al. (2020), the recorded *tpm* sequences could be attributed to PAO1- and PA14-like sub-clades and MLST groups such as ST252 involved in non-CF (cystic fibrosis) infections, and ST2123/ST226 and ST175/ST2042 found in lung, urinary tract and blood infections. The PA14 OTU00056 showed a tropism for sampling sites with high numbers of gasoline engines, suggesting a use of petrol-derivatives as C-sources. The above classifications were validated through the analysis of *P. aeruginosa* isolates obtained from runoffs. These analyses confirmed the occurrence of bacterial pathogens among urban surface K-strategists. However, contribution of their virulence properties in their establishment among urban surface microbiomes remains to be demonstrated.

5. Conclusions

Understanding the complexity of urban surface microbiomes requires (i) the identification of the sources of microbial taxa seeding these systems, and (ii) of the key selective forces triggering changes in their assemblages, and favoring the emergence of efficient functional units. Here, socio-urbanistic surveys were undertaken to better define these forces and sources. The surveyed variables led to significant groupings of urban typologies matching segregations observed among the microbiological datasets (FIB, MST, DNA meta-barcoding) generated from runoff samples. This was clearly indicative of well-established K-like

780 synurbic bacterial taxa over the catchment. The most significant selective forces explaining
781 the success of these K taxa were related to urban pollutants. A large module of co-occurring
782 actinobacterial taxa involved in hydrocarbon degradation was resolved from the 16S rRNA
783 gene dataset. Among these K-like hydrocarbon degraders, bacterial pathogens were observed
784 such as *P. aeruginosa* and *P. syringae*. These urban surface bacterial assemblages were also
785 found to be made of opportunistic r-like taxa. In fact, an industrial point source pollution
786 likely involving trichloroethylene and halogenated solvents was associated with a significant
787 re-shuffle of the microbiota, and a burst in *Novosphingobium* and *Sulfurospirillum*. The high
788 impact of urban pollutants on these urban microbiomes was also further supported by the
789 observation of r-like *tpm*-harboring species known to play part in hydrocarbon degradation
790 such as *P. extremaustralis*.

791 The generated meta-barcoding datasets presented here, and the investigated long – term
792 experimental site, will be highly useful for studies to come on the evaluation of the ecological
793 benefits associated with a reduction of gasoline engines over a city. The evolving state of the
794 genetic structures of a surface urban microbiome is likely to be a very good indicator of the
795 ecological benefits associated with such changes.

Credit authorship contribution statement

Rayan Bouchali: conceptualization, data curation, formal analysis, writing - original draft. **Claire Mandon:** conceptualization, data curation, formal analysis. **Romain Marti**: conceptualization, data curation, formal analysis. **Jérôme Michalon:** conceptualization, formal analysis. **Axel Aigle:** formal analysis. **Laurence Marjolet:** data curation, formal analysis. **Sophie Vareilles:** conceptualization, data curation. **Gislain Lipeme-Kouyi:** conceptualization, data curation and analyses, writing - review & editing. **Philippe Polomé:** data curation and analyses, writing - review & editing. **Jean-Yves Toussaint:** data curation, writing - review & editing. **Benoit Cournoyer:** conceptualization, data curation, formal analysis, funding acquisition, project administration and supervision, writing - original draft, writing - review & editing.

Declaration of Competing Interest

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

Funding

This work was partly funded by the French national research program for environmental and occupational health of ANSES under the terms of project “Iouqmer” EST 2016/1/120, l’Agence Nationale de la Recherche through the ANR-17-CE04-0010 (Infiltron) project, by Labex IMU (Intelligence des Mondes Urbains), the MITI CNRS project named Urbamic, the Greater Lyon Urban Community, and the French water agency for the Rhône, Mediterranean and Corsica areas through the DOmic project.

Acknowledgements

Authors thank the OTHU network for technical assistance and financial supports, E. Bourgeois and Y. Colin (UMR Ecologie Microbienne) for the initial quality validation of the

tpm meta-barcoding dataset presented here and advices on the bio-informatic manipulations, and the Urban School of Lyon (ANR-17-CONV-0004) for their advices and supports in the elaboration of this multi-disciplinary research initiative. This work was also performed within the framework of the EUR H2O'Lyon (ANR-17-EURE-0018) of Université de Lyon (UdL), within the program "Investissements d'Avenir" (PIA) operated by the French National Research Agency (ANR).

References

Ahmed, W., Yusuf, R., Hasan, I., Goonetilleke, A., Gardner, T., 2010. Quantitative PCR assay of sewage-associated *Bacteroides* markers to assess sewage pollution in an urban lake in Dhaka, Bangladesh. Can. J. Microbiol. 56, 838–845. https://doi.org/10.1139/W10-070

Aigle, A., Colin, Y., Bouchali, R., Bourgeois, E., Marti, R., Ribun, S., Marjolet, L., Pozzi, A.C.M., Misery, B., Colinon, C., Bernardin-Souibgui, C., Wiest, L., Blaha, D., Galia, W., Cournoyer, B., 2021. Spatio-temporal variations in chemical pollutants found among urban deposits match changes in thiopurine S-, Se-methyltransferase-harboring bacteria tracked by the tpm metabarcoding approach. Sci. Total Environ. 145425. https://doi.org/10.1016/j.scitotenv.2021.145425

Almakki, A., Jumas-Bilak, E., Marchandin, H., Licznar-Fajardo, P., 2019. Antibiotic resistance in urban runoff. Sci. Total Environ. 667, 64–76. https://doi.org/10.1016/j.scitotenv.2019.02.183

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

Bučková, M., Puškarová, A., Chovanová, K., Kraková, L., Ferianc, P., Pangallo, D., 2013. A simple strategy for investigating the diversity and hydrocarbon degradation abilities of cultivable bacteria from contaminated soil. World J. Microbiol. Biotechnol. 29, 1085–1098. https://doi.org/10.1007/s11274-013-1277-5

Buttet, G.F., Murray, A.M., Goris, T., Burion, M., Jin, B., Rolle, M., Holliger, C., Maillard, J., 2018. Coexistence of two distinct Sulfurospirillum populations respiring tetrachloroethene—genomic and kinetic considerations. FEMS Microbiol. Ecol. 94. https://doi.org/10.1093/femsec/fiy018

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869

Chocat, B., 2013. CANOE: An Urban Hydrology Software Package, in: Tanguy, J.-M. (Ed.), Modeling Software. John Wiley & Sons, Inc., Hoboken, NJ USA, pp. 209–218. https://doi.org/10.1002/9781118557891.ch17

Colin, Y., Bouchali, R., Marjolet, L., Marti, R., Vautrin, F., Voisin, J., Bourgeois, E., Rodriguez-Nava, V., Blaha, D., Winiarski, T., Mermillod-Blondin, F., Cournoyer, B., 2020. Coalescence of bacterial groups originating from urban runoffs and artificial infiltration systems among aquifer microbiomes. Hydrol. Earth Syst. Sci. 24, 4257–4273. https://doi.org/10.5194/hess-24-4257-2020

Danko, D., Bezdan, D., Afshin, E.E., et al., 2021. A global metagenomic map of urban microbiomes and antimicrobial resistance. Cell S0092867421005857. https://doi.org/10.1016/j.cell.2021.05.002

Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., Callahan, B.J., 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6. https://doi.org/10.1186/s40168-018-0605-2

Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular ecological network analyses. BMC Bioinformatics 13, 113. https://doi.org/10.1186/1471-2105-13-113

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Ellis, J.B., 2004. Bacterial sources, pathways and management strategies for urban runoff. J. Environ. Plan. Manag. 47, 943–958. https://doi.org/10.1080/0964056042000284910

Gilbert, J.A., Stephens, B., 2018. Microbiology of the built environment. Nat. Rev. Microbiol. 16, 661–670. https://doi.org/10.1038/s41579-018-0065-5

Göbel, P., Dierkes, C., Coldewey, W.G., 2007. Storm water runoff concentration matrix for urban areas. J. Contam. Hydrol. 91, 26–42. https://doi.org/10.1016/j.jconhyd.2006.08.008

Hao, Y.-Q., Zhao, X.-F., Zhang, D.-Y., 2016. Field experimental evidence that stochastic processes predominate in the initial assembly of bacterial communities: Stochastic assembly of bacterial communities. Environ. Microbiol. 18, 1730–1739. https://doi.org/10.1111/1462-2920.12858

He, X., McLean, J.S., Edlund, A., Yooseph, S., Hall, A.P., Liu, S.-Y., Dorrestein, P.C., Esquenazi, E., Hunter, R.C., Cheng, G., Nelson, K.E., Lux, R., Shi, W., 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc. Natl. Acad. Sci. 112, 244–249. https://doi.org/10.1073/pnas.1419038112

Ibekwe, A.M., Leddy, M., Murinda, S.E., 2013. Potential human pathogenic bacteria in a mixed urban watershed as revealed by pyrosequencing. PLoS ONE 8, e79490. https://doi.org/10.1371/journal.pone.0079490

Jiao, S., Zhang, Z., Yang, F., Lin, Y., Chen, W., Wei, G., 2017. Temporal dynamics of microbial communities in microcosms in response to pollutants. Mol. Ecol. 26, 923–936. https://doi.org/10.1111/mec.13978

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl. Environ. Microbiol. 79, 5112–5120. https://doi.org/10.1128/AEM.01043-13

Langenbach, S., Rehm, B.H.A., Steinbüchel, A., 2006. Functional expression of the PHA synthase gene phaC1 from *Pseudomonas aeruginosa* in *Escherichia coli* results in poly(3-hydroxyalkanoate) synthesis. FEMS Microbiol. Lett. 150, 303–309. https://doi.org/10.1111/j.1574-6968.1997.tb10385.x

Lavenir, R., Petit, S.M.-C., Alliot, N., Ribun, S., Loiseau, L., Marjolet, L., Briolay, J., Nazaret, S., Cournoyer, B., 2014. Structure and fate of a *Pseudomonas aeruginosa* population originating from a combined sewer and colonizing a wastewater treatment lagoon. Environ. Sci. Pollut. Res. 21, 5402–5418. https://doi.org/10.1007/s11356-013-2454-7

Lee, S., Suits, M., Wituszynski, D., Winston, R., Martin, J., Lee, J., 2020. Residential urban stormwater runoff: A comprehensive profile of microbiome and antibiotic resistance. Sci. Total Environ. 723, 138033. https://doi.org/10.1016/j.scitotenv.2020.138033

Leung, M.H.Y., Wilkins, D., Li, E.K.T., Kong, F.K.F., Lee, P.K.H., 2014. Indoor-air microbiome in an urban subway network: diversity and dynamics. Appl. Environ. Microbiol. 80, 6760–6770. https://doi.org/10.1128/AEM.02244-14

Liu, Z.-P., 2005. *Novosphingobium taihuense* sp. nov., a novel aromatic-compound-degrading bacterium isolated from Taihu Lake, China. Int. J. Syst. Evol. Microbiol. 55, 1229–1232. https://doi.org/10.1099/ijs.0.63468-0

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

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Luijten, M.L.G.C., 2003. Description of *Sulfurospirillum halorespirans* sp. nov., an anaerobic, tetrachloroethene-respiring bacterium, and transfer of *Dehalospirillum multivorans* to the

genus *Sulfurospirillum* as *Sulfurospirillum multivorans* comb. nov. Int. J. Syst. Evol. Microbiol. 53, 787–793. https://doi.org/10.1099/ijs.0.02417-0

Marques, A.C.Q., Paludo, K.S., Dallagassa, C.B., Surek, M., Pedrosa, F.O., Souza, E.M., Cruz, L.M., LiPuma, J.J., Zanata, S.M., Rego, F.G.M., Fadel-Picheth, C.M.T., 2015. Biochemical characteristics, adhesion, and cytotoxicity of environmental and clinical isolates of *Herbaspirillum* spp. J. Clin. Microbiol. 53, 302–308. https://doi.org/10.1128/JCM.02192-14

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

McLellan, S.L., C. Fisher, J., J. Newton, R., 2015. The microbiome of urban waters. Int. Microbiol. 141–149. https://doi.org/10.2436/20.1501.01.244

McLellan, S.L., Roguet, A., 2019. The unexpected habitat in sewer pipes for the propagation of microbial communities and their imprint on urban waters. Curr. Opin. Biotechnol. 57, 34–41. https://doi.org/10.1016/j.copbio.2018.12.010

Morais, D., Pylro, V., Clark, I.M., Hirsch, P.R., Tótola, M.R., 2016. Responses of microbial community from tropical pristine coastal soil to crude oil contamination. PeerJ 4, e1733. https://doi.org/10.7717/peerj.1733

Netsu, O., Kijima, T., Takikawa, Y., 2014. Bacterial leaf spot of hemp caused by Xanthomonas campestris pv. cannabis in Japan. J. Gen. Plant Pathol. 80, 164–168. https://doi.org/10.1007/s10327-013-0497-8

Nijenhuis, I., Andert, J., Beck, K., Kastner, M., Diekert, G., Richnow, H.-H., 2005. Stable isotope fractionation of tetrachloroethene during reductive dechlorination by *Sulfurospirillum multivorans* and *Desulfitobacterium* sp. strain PCE-S and abiotic reactions with cyanocobalamin. Appl. Environ. Microbiol. 71, 3413–3419. https://doi.org/10.1128/AEM.71.7.3413-3419.2005

Noyer, M., Reoyo-Prats, B., Aubert, D., Bernard, M., Verneau, O., Palacios, C., 2020. Particle-attached riverine bacteriome shifts in a pollutant-resistant and pathogenic community during a Mediterranean extreme storm event. Sci. Total Environ. 732, 139047. https://doi.org/10.1016/j.scitotenv.2020.139047

Park, J.-W., Crowley, D.E., 2006. Dynamic changes in *nahAc* gene copy numbers during degradation of naphthalene in PAH-contaminated soils. Appl. Microbiol. Biotechnol. 72, 1322–1329. https://doi.org/10.1007/s00253-006-0423-5

Ramirez, K.S., Leff, J.W., Barberán, A., Bates, S.T., Betley, J., Crowther, T.W., Kelly, E.F., Oldfield, E.E., Shaw, E.A., Steenbock, C., Bradford, M.A., Wall, D.H., Fierer, N., 2014. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. Proc. R. Soc. B Biol. Sci. 281, 20141988. https://doi.org/10.1098/rspb.2014.1988

Revitt, D.M., Lundy, L., Coulon, F., Fairley, M., 2014. The sources, impact and management of car park runoff pollution: A review. J. Environ. Manage. 146, 552–567. https://doi.org/10.1016/j.jenvman.2014.05.041

Robertson, C.E., Baumgartner, L.K., Harris, J.K., Peterson, K.L., Stevens, M.J., Frank, D.N., Pace, N.R., 2013. Culture-independent analysis of aerosol microbiology in a metropolitan subway system. Appl. Environ. Microbiol. 79, 3485–3493. https://doi.org/10.1128/AEM.00331-13

Sauer, E.P., VandeWalle, J.L., Bootsma, M.J., McLellan, S.L., 2011. Detection of the human specific *Bacteroides* genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment. Water Res. 45, 4081–4091. https://doi.org/10.1016/j.watres.2011.04.049

Saxena, G., Marzinelli, E.M., Naing, N.N., He, Z., Liang, Y., Tom, L., Mitra, S., Ping, H., Joshi, U.M., Reuben, S., Mynampati, K.C., Mishra, S., Umashankar, S., Zhou, J., Andersen, G.L., Kjelleberg, S., Swarup, S., 2015. Ecogenomics reveals metals and land-use pressures on microbial communities in the waterways of a megacity. Environ. Sci. Technol. 49, 1462–1471. https://doi.org/10.1021/es504531s

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn,

975 D.J., Weber, C.F., 2009. Introducing mothur: Open-source, platform-independent,
976 community-supported software for describing and comparing microbial communities. Appl.
977 Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09
978 Sébastian, C., Barraud, S., Ribun, S., Zoropogui, A., Blaha, D., Becouze-Lareure, C., Kouyi, G.L.,
979 Cournoyer, B., 2014. Accumulated sediments in a detention basin: chemical and microbial
980 hazard assessment linked to hydrological processes. Environ. Sci. Pollut. Res. 21, 5367–5378.
981 https://doi.org/10.1007/s11356-013-2397-z
982 Seurinck, S., Defoirdt, T., Verstraete, W., Siciliano, S.D., 2005. Detection and quantification of the
983 human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for
984 assessment of human faecal pollution in freshwater. Environ. Microbiol. 7, 249–259.
985 https://doi.org/10.1111/j.1462-2920.2004.00702.x
986 Sidhu, J.P.S., Hodgers, L., Ahmed, W., Chong, M.N., Toze, S., 2012. Prevalence of human pathogens
987 and indicators in stormwater runoff in Brisbane, Australia. Water Res. 46, 6652–6660.
988 https://doi.org/10.1016/j.watres.2012.03.012
989 Sohn, J.H., 2004. N*ovosphingobium pentaromativorans* sp. nov., a high-molecular-mass polycyclic
990 aromatic hydrocarbon-degrading bacterium isolated from estuarine sediment. Int. J. Syst.
991 Evol. Microbiol. 54, 1483–1487. https://doi.org/10.1099/ijs.0.02945-0
992 Song, H.-K., Song, W., Kim, M., Tripathi, B.M., Kim, H., Jablonski, P., Adams, J.M., 2017. Bacterial
993 strategies along nutrient and time gradients, revealed by metagenomic analysis of laboratory
994 microcosms. FEMS Microbiol. Ecol. 93. https://doi.org/10.1093/femsec/fix114
995 Timm, A., Steinbuchel, A., 1992. Cloning and molecular analysis of the poly(3-hydroxyalkanoic acid)
996 gene locus of *Pseudomonas aeruginosa* PAO1. Eur. J. Biochem. 209, 15–30.
997 https://doi.org/10.1111/j.1432-1033.1992.tb17256.x
998 Tribelli, P.M., Raiger Iustman, L.J., Catone, M.V., Di Martino, C., Revale, S., Mendez, B.S., Lopez,
999 N.I., 2012. Genome sequence of the polyhydroxybutyrate producer *Pseudomonas*
1000 *extremaustralis*, a highly stress-resistant antarctic bacterium. J. Bacteriol. 194, 2381–2382.
1001 https://doi.org/10.1128/JB.00172-12
1002 Ukpaka, C.P., 2014. The influence of chemical and biochemical oxygen demands on the kinetics of
1003 crude oil degradation in salt water pond. Sky J. Biochem. Res. 3(1):001-013.
1004 Vadstein, O., Attramadal, K.J.K., Bakke, I., Olsen, Y., 2018. K-selection as microbial community
1005 management strategy: a method for improved viability of larvae in aquaculture. Front.
1006 Microbiol. 9. https://doi.org/10.3389/fmicb.2018.02730
1007 Voisin, J., Cournoyer, B., Vienney, A., Mermillod-Blondin, F., 2018. Aquifer recharge with
1008 stormwater runoff in urban areas: Influence of vadose zone thickness on nutrient and bacterial
1009 transfers from the surface of infiltration basins to groundwater. Sci. Total Environ. 637–638,
1010 1496–1507. https://doi.org/10.1016/j.scitotenv.2018.05.094
1011 Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve bayesian classifier for rapid assignment
1012 of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–
1013 5267. https://doi.org/10.1128/AEM.00062-07
1014 Wright, M.S., Baker-Austin, C., Lindell, A.H., Stepanauskas, R., Stokes, H.W., McArthur, J.V., 2008.
1015 Influence of industrial contamination on mobile genetic elements: class 1 integron abundance
1016 and gene cassette structure in aquatic bacterial communities. ISME J. 2, 417–428.
1017 https://doi.org/10.1038/ismej.2008.8
1018

Figure captions

Fig. 1. Graphical representations of (A) the Mi-plaine catchment of Chassieu (Suburbs of Lyon, France) and (B) of the adjusted Wallace inferences indicating organizational proximities between the observed industrial and commercial activities of this catchment. (A) Relative importance of shops, restaurants and industries in terms of number of employees was indicated by a grey scale (dark values = high numbers). Runoff water sampling sites (C code) are numbered and positioned on the map. The outlet of the watershed is located at C23, and corresponds to a stormwater infiltration system (SIS). Industrial, commercial, and social activities and their traces are listed in **Tables S1** and **S2**. These activities were analyzed over an area of 50 m of diameter around each sampling point. In (B), Adjusted Wallace inferences allowed to identify two large groups, I (brownish dotted lines) and II (dark green dotted lines) of industrial/commercial activities. In (A), groups III (red background – high intensity area) and IV (blue background – low intensity area) were defined according to significant relationships inferred from the behavioral variables. Representative sites of groups III and IV are shown in **Fig. 2**. **Fig.** 3 indicates the main socio-urbanistic variables impacting these sites. Yellow circles in (A) indicate human fecal contaminations, yellow and green circles indicate a human fecal contamination associated with the presence of integron 1, and green circles of integron 1. *E. coli* and intestinal enterococci were recorded among all sub-catchments (**Table S3**).

Fig. 2. Representative urban typologies for the socio-urbanistic groups III and IV observed over the Mi-plaine catchment. Group III was illustrated by sampling point C19, and group IV by C18. See **Table S1** for a description of the other sampling points. Aerial views show the runoff water flow directions which led to a delimitation of the sub-catchment. Economical activities, traffic intensity, human behaviors and wastes recorded during the socio-urbanistic surveys are also indicated by symbols defined on the figure. Photographs (the yellow arrow

1044 indicates the direction of the selected view) illustrate the main typologies observed over the
1045 selected sub-catchments, and were taken from the sections indicated by dotted yellow
1046 rectangles on the aerial views. Position of the sampling code numbers (C18 or C19) indicated
1047 the point of collect of runoff waters. Human fecal contamination is indicated by a small
1048 yellow circle while the yellow/green circle indicates a human fecal contamination associated
1049 with integron 1.

1050 **Fig. 3.** Spearman correlation tests exploring relationships between the repartition of industries
1051 and commercial activities, and of general descriptors of human activities, over the Mi-plaine
1052 urban catchment. Only Rho factors > 0.6 that linked significantly ($p<0.05$) these variables
1053 were considered to draw these networks. Green circles represent the surveyed industries;
1054 orange ones represent general descriptors of human activities. Variables in agreement with the
1055 definition of Group II (shown in Fig. 1) are circled with a plain line. Socio-urbanistic
1056 variables with high correlations are shown with dotted circles, and the selected representative
1057 variables of these groups were circled with a plain line. The datasets are shown in Tables S2b,
1058 c.

1059 **Fig. 4.** NMDS representation of the Bray-Curtis dissimilarity matrix (**Table S13**) computed
1060 from the 16S rRNA gene OTU profiles (**Table S6**) of runoff water samples (except C14)
1061 according to the groups of sampling points defined by the socio-urbanistic surveys. (a) group I
1062 and II and (b) group III and IV. See **Figs. 1 and 2** and text for a description of these groups.
1063 Adonis statistical tests showed significant differentiation between 16S rRNA gene OTU
1064 profiles of (a) groups I and II, and (b) groups III and IV, ($p<0.01$).

1065 **Fig. 5.** Chord diagram showing the distribution of the 19 dominant 16S rRNA-inferred genera
1066 (relative abundance > 0.5 %) among the runoff water samples from the investigated urban
1067 catchment. Full dataset of taxonomic allocations is shown in **Table S11**. These genera

grouped about half of the reads and almost 80% of those that could be allocated to well-defined taxa. The average relative abundance (among the three sampling campaigns) is indicated for each genus. Genera in bold were affiliated to the core-microbiome (present in all the samples).

Fig. 6. Correlation networks computed from the fast greedy modularity optimization method of the MENA pipeline for (a) the 16S rRNA gene inferred genera and (b) the *tpm*-harboring bacterial species. Only bacterial taxa detected in at least 25 (for the 16S rRNA gene dataset) or 10 (for the *tpm* one) runoff water samples were used to compute the correlation networks. Spearman correlations above 0.7 are presented. Global properties of the two networks are indicated in Tables S14 and S21. Module-EigenGene analyses were computed to highlight relationships between 16S rRNA-based or *tpm* modules, and the physico-chemical or socio-urbanistic variables. The significant correlation ratios between the modules and the explanatory variables are indicated in boxes. Only correlations over 0.5 (for the 16S rRNA gene dataset) or 0.4 (for the *tpm* one) are shown. Asterisks indicate Spearman correlations p-values < 0.001. Genera and specie showed in bold indicate taxa involved in hydrocarbon degradation.

Fig. 7. Chord diagram showing the distribution of the 44 dominant *tpm*-harboring bacterial species (relative abundance > 0.1 %) among the runoff water samples from the Mi-Plaine catchment of Chassieu (**Table S16**). These species grouped 45 % of the *tpm* reads. Average relative abundances (among the three sampling campaign) of each *tpm*-harboring bacterial species are indicated. Species in bold were affiliated to the core-microbiome of the runoff water and had a prevalence >50%.

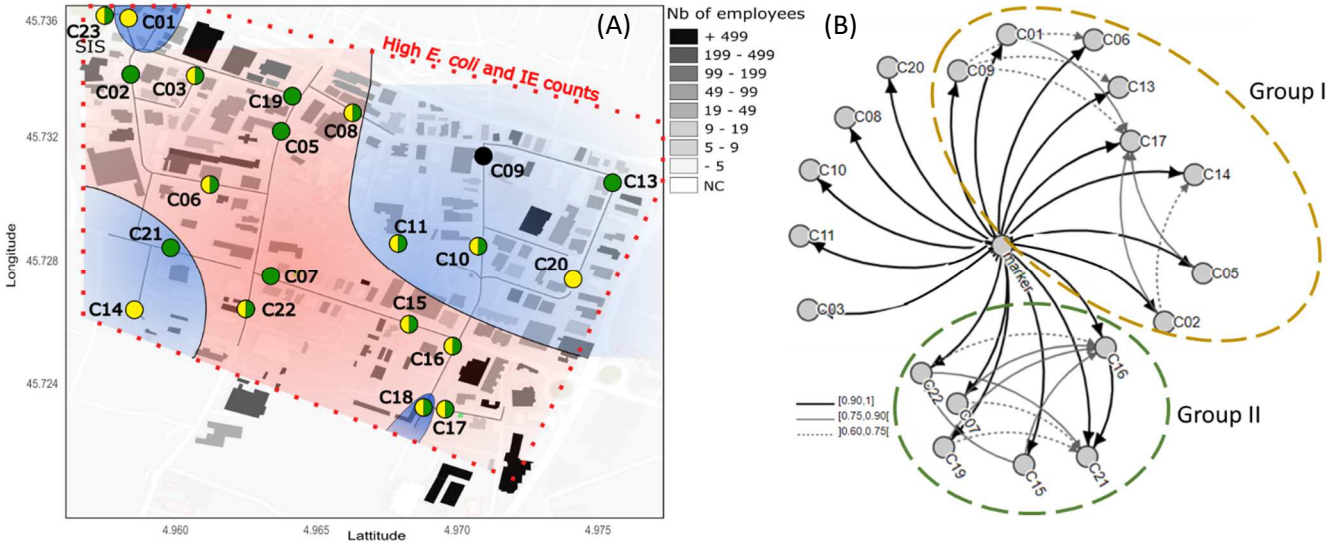
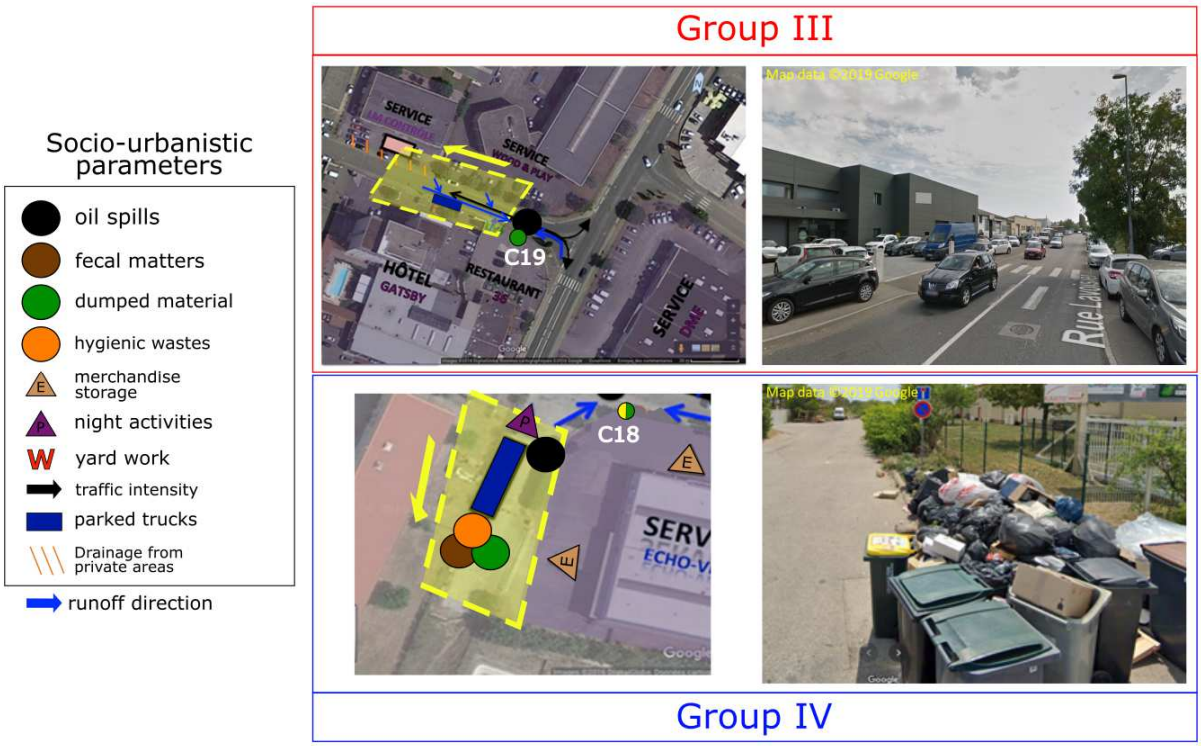
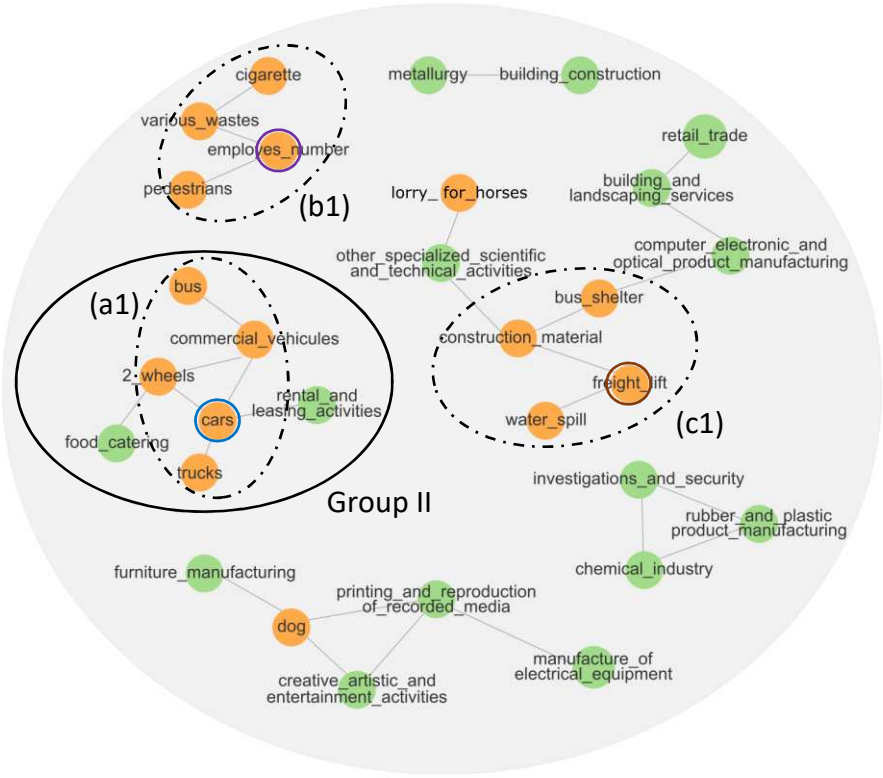


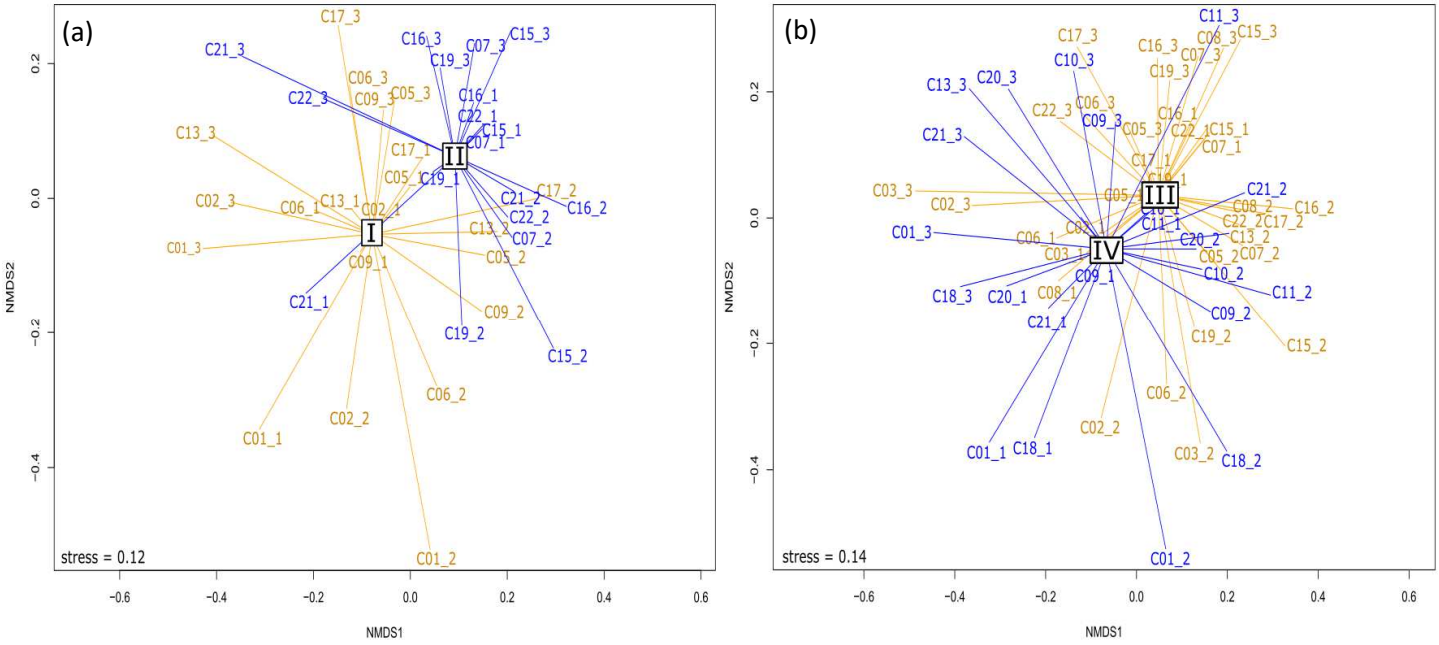
Figure 1. Bouchali et al.



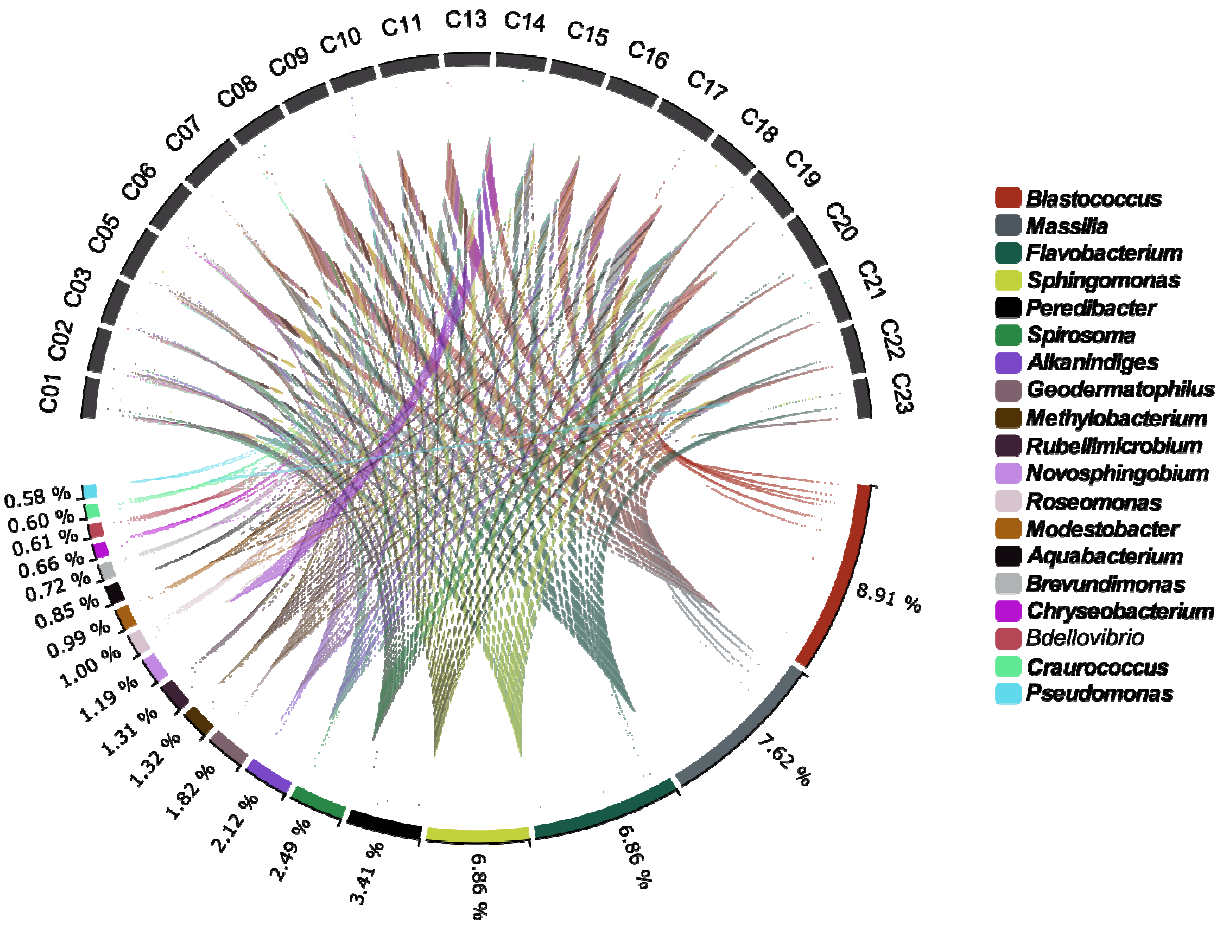
Bouchali et al. Fig. 2



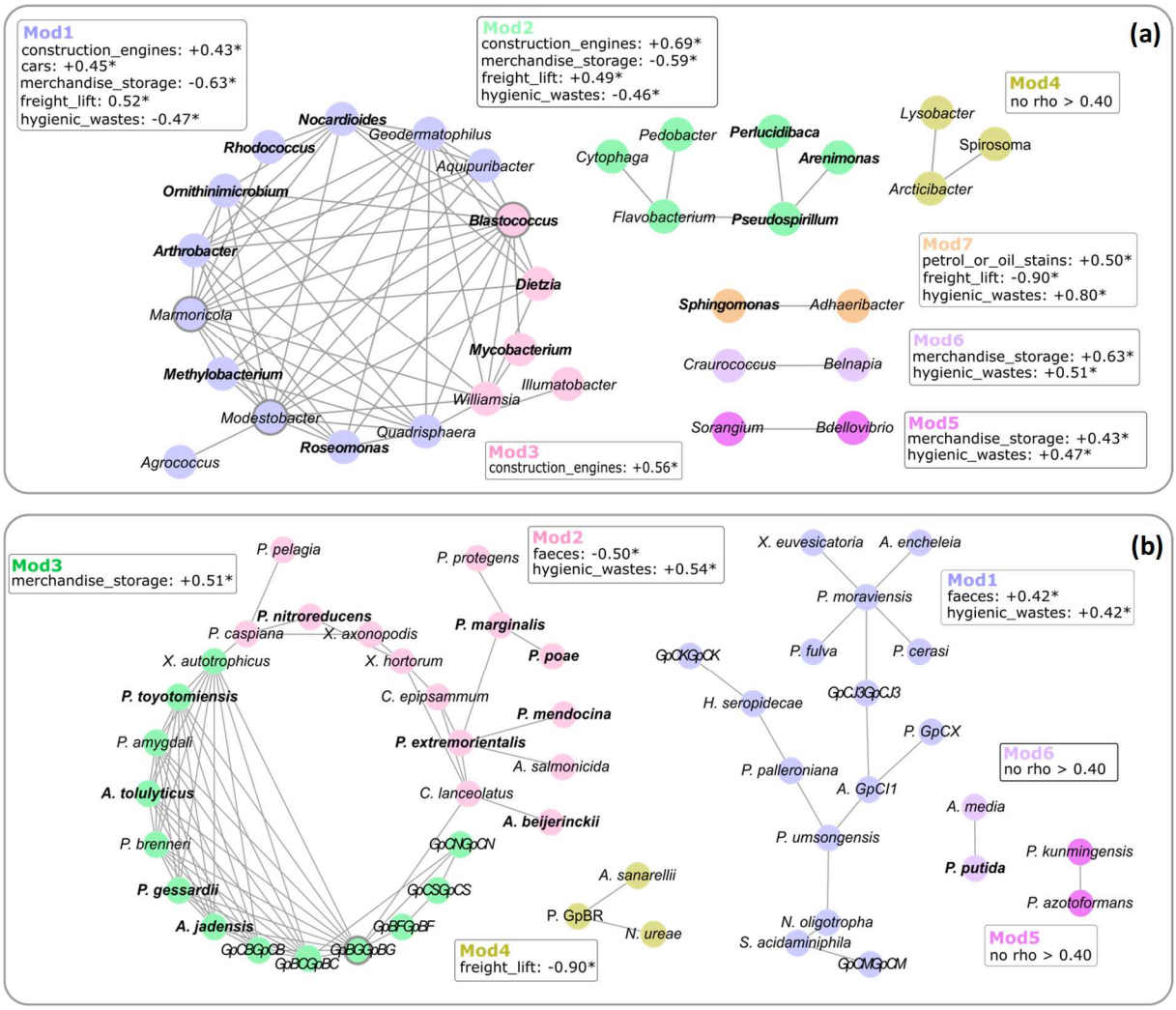
Bouchali et al. Fig. 3



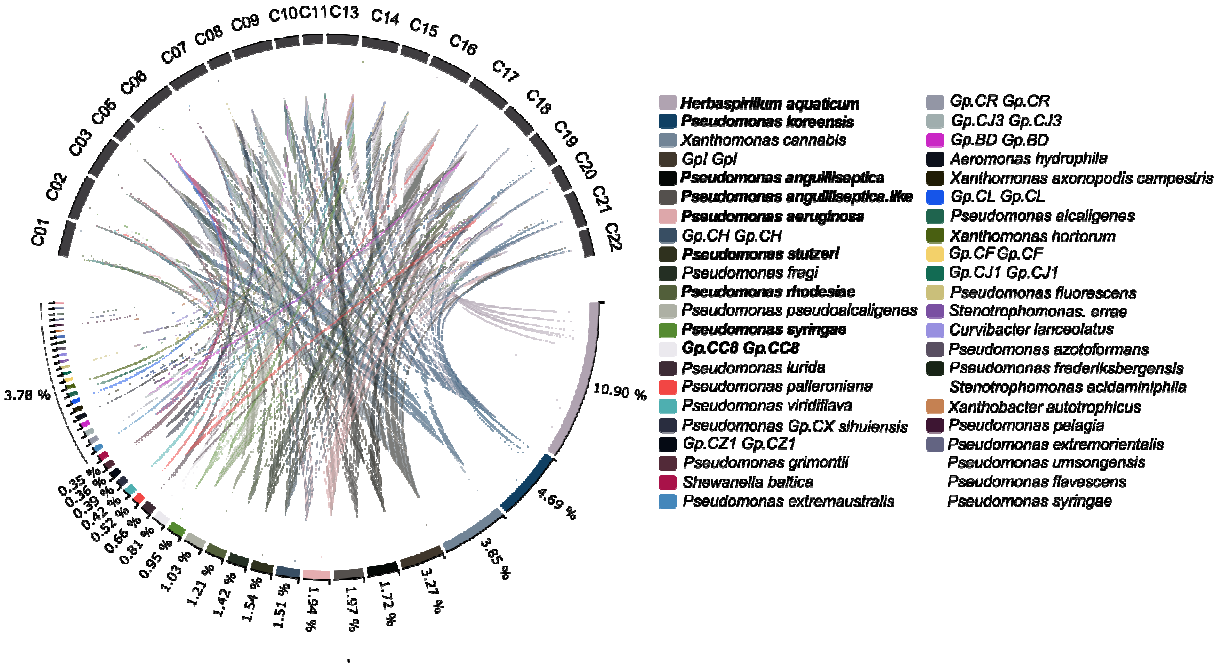
Bouchali et al. Fig.4



Bouchali et al. Fig. 5



Bouchali et al. Fig. 6



Bouchali et al. Fig. 7

GRAPHICAL ABSTRACT