Spatio-temporal variations in chemical pollutants found among urban deposits match changes in thiopurine S-methyltransferase-harboring bacteria tracked by the tpm metabarcoding approach

Axel Aigle, Emilie Bourgeois, Laurence Marjolet, Sabine Houot, Dominique Patureau, Benoit Cournoyer, Wessam Galia, Yannick Colin, Rayan Bouchali, Romain Marti, et al.

Version of Record: https://www.sciencedirect.com/science/article/pii/S0048969721004939
Manuscript_470dc6fa60283770264d3efe62129600

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

Axel Aigle[1], Yannick Colin[1#], Rayan Bouchali[1], Emilie Bourgeois[1], Romain Marti[1], Sébastien Ribun[1], Laurence Marjolet[1], Adrien C. M. Pozzi[1], Boris Misery[1Ψ], Céline Colinon[1], Claire Bernardin-Souibgui[1], Laure Wiest[2], Didier Blaha[1], Wessam Galia[1], Benoit Cournoyer[1]

[1]Université de Lyon, Université Claude Bernard Lyon 1, VetAgro Sup, UMR Ecologie Microbienne, CNRS 5557, INRA 1418, Research team «Bacterial Opportunistic Pathogens and Environment», 69280 Marcy L'Etoile, France; and [2]Université de Lyon, Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, CNRS 5280, 5 rue de la Doua, 69100, Villeurbanne, France.

[#Ψ]Present address : Normandie Université, UNIROUEN, UNICAEN, [#]UMR CNRS 6143, Morphodynamique Continentale et Côtière, 76000 Rouen, and [Ψ]ABTE, 14000 Caen, 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: Bacterial exposure to high priority pollutants

Keywords: urban ecology, pollutants, *tpm* metabarcoding, DADA2, Mothur, sediments, ecotoxicology

1. Introduction

Urban road deposits lead to the formation of novel habitats for micro-organisms (Reese et al., 2016). They are made of organic matter, plant debris, solid wastes (sand, gravels, plastics) and mixtures of chemical pollutants (Grimm et al., 2008). Their microbiome can be considered representative of the forces and components which led to their aggregation into structured matrices often termed sediments. These road deposits and sediments represent an opportunity for deciphering the outdoor urban microbiome, and investigate the incidence of combined exposures to high risk priority pollutants on their diversity. During rain events, these deposits are partly washed off, and are often transferred into a separated network of pipes devoted to the management of urban waters. This network will transfer the stormwaters into either rain gardens, ditches, rivers, brooks, or detention and infiltration basins (Sébastian et al., 2014). Detention basins are frequently used in urban contexts to: (1) mitigate runoff peak flows and protect downstream areas from flooding, and (2) remove some pollutants from urban waters by favoring settling of the suspended matters prior transfer of the waters to an infiltration basin (Yan et al., 2014) that will contribute at recharging a connected aquifer (Voisin et al., 2018). While detention basins represent efficient technical devices for handling urban waters, they also offer an opportunity to gain access to urban road deposits and sediments from a well-defined urban watershed. These deposits can then be used as reference matrices to evaluate the local spread of hazardous chemicals (Becouze-Lareure et al., 2018) and biological agents (Bernardin-Souibgui et al., 2018a).

The selective forces and constraints that can impact the microbiome of urban deposits and their related more organized sediments among a detention basin include the (1) chemical interactions that will affect the aggregation / disaggregation of particles (Badin et al., 2008; Crawford et al., 2012), (2) toxicity of the co-mobilized chemical pollutants (Gonzalez-Merchan et al., 2014), (3) the hydrodynamic parameters that will generate variable flows, turbulences, and

shearing forces impacting the structure of the deposits and related sediments (Yan et al., 2014), (4) weather conditions including temperature and duration of dry periods, and (5) biological interactions such as predation e. g. nematodes, amoeba or the development of a plant cover which can lead to a production of antimicrobials or favor the enrichment of certain microbial taxa e. g. phytopathogens (Cournoyer et al., 1995) or rhizosphere bacteria (Balandreau et al., 2001). 16S rRNA gene Operational Taxonomic Unit (OTU) profilings of urban deposits and sediments were recently found to be sufficiently informative to infer the incidence of some of the above constraints on urban microbiomes. Their genetic structures were found indicative of the sources of their physical components such as soil and fecal matters (Marti et al., 2017a). Similarly, 16S rRNA gene OTU patterns of urban soils were found strongly structured by physico-chemical constraints such as pH (Yan et al., 2016), and Mn, Ni, Sb and Tl concentrations (Xu et al., 2014). Their OTU patterns were shown to contain DNA imprints matching the urbanization level of their surrounding area (Xu et al., 2014). These observations supported the hypothesis that urban deposits should contain DNA signatures indicative of their geographical origin but also of human activities and associated pollutants driving the differentiation of their microbiome. In fact, Cd, Cr, Ni, Pb, fluorene, and chrysene concentrations were found to be strongly correlated to 16S rRNA gene OTU patterns of urban deposits when accumulating in a detention basin (Marti et al., 2017a). Harsher environmental constraints building-up over time due to pollutant accumulations in such basins were also found to increase the occurrence of extremophiles such as *Acidibacter* and *Haliangium* (Marti et al., 2017a).

In order to go deeper into the analysis of the incidence of multiple chemical stressors on urban microbiomes, the ecology of the *tpm* gene encoding the bTPMT (bacterial thiopurine methyltransferase) enzyme was explored. bTPMTs are involved in the transformation of toxic S-, Se-, Te-containing substances. It was initially described as a detoxifying process for metalloid-based oxyanions, and to be a key component of resistomes expressed to fight the deleterious

effects of highly oxidative molecules (Cournoyer et al., 1998; Prigent-Combaret et al., 2012). The bTPMT can induce the depletion of these toxic metalloid-based oxyanions by driving their transformation into volatile methylated metalloids (Prigent-Combaret et al., 2012). However, bTPMTs are likely to have a much broader range of substrates. In fact, they were shown to inactivate 6-mercaptopurine (6-MP) drugs through *S*-methylation (Cournoyer et al., 1998). Furthermore, the human (h) and mice TPMT orthologs were found to be involved in the biotransformation of several toxic cellular by-products and pharmaceuticals (e. g. Katara and Kuntal, 2016). The human TPMT is considered essential to circumvent the lethal effects of 6-MPs by favoring chemical conjugation reactions making these drugs more prone to excretion. 6-MP and the related azathioprine are commonly used to treat leukemia, rheumatic diseases, inflammatory bowel diseases, and to favor solid organ transplantations (Coussement et al., 2016; Ford and Berg, 2010). TPMTs were found distributed among about 150 bacterial genera mainly allocated to the γ-proteobacteria such as the *Pseudomonas* and *Aeromonas* (Cournoyer et al., 1998; Colin et al., 2020). The *tpm* gene has highly conserved DNA boxes that were found to delineate highly informative sequences in terms of phylogenetic and taxonomic inferences (Favre-Bonte et al., 2005; Favre-Bonté et al., 2006). This gene was suggested to make possible inferences down to the species and sub-species levels. These levels could not be reached with 16S rRNA gene metabarcoding analyses as shown in Marti et al. (2017a).

Here, we hypothesized that: (1) high risk priority pollutants can explain the organization of *tpm*-harboring bacterial communities among urban sediments, and (2) *tpm* sequence types can be used as indicators of the level of contamination of outdoor systems by multiple chemical pollutants. These hypotheses were tested by using the experimental design reported in Wiest et al. (2018) based on the analysis of urban sediments and deposits from an urban industrial catchment (named Mi-Plaine) connected to a stormwater infiltration system (SIS). This catchment is part of a field observatory of urban waters in Lyon named “OTHU”

(/www.graie.org/othu/). This observatory keeps record of all datasets and samples (DNA, sediments, soils, runoffs) conserved for this catchment, and favors their uses through multi-disciplinary research actions. This work could thus benefit of DNA samples and datasets produced over several years. The OTHU sampling scheme allowed to obtain multiple independent DNA replicates per sampling points of the detention basin (DB) of the selected SIS and its settling pit, but also of a control point located outside the DB. A *tpm* DNA metabarcoding approach was then implemented to build-up large DNA libraries of *tpm* sequences. The *tpm* reads generated by this procedure were analysed using both the DADA2 (Callahan et al. 2016) and Mothur (Schloss et al. 2009) MiSeq standard operating procedures (SOP). Mothur is based on the use of a re-sampled dataset (mothur) and definition of operational taxonomic units (OTUs), and DADA2 makes use of all the sequenced reads through the definition of amplicon sequence variants (ASVs) but implies a deeper analysis of sequence quality scores. Significant correlations between their respective filtered output DNA sequences were observed. Relationships between the monitored physico-chemical variables and distribution patterns of high risk priority pollutants, and the *tpm*-metabarcoding profiles generated by mothur and DADA2 were investigated. This allowed a demonstration of significant changes among bTPMT-harboring bacteria over time matching the accumulation of several priority pollutants (listed at https://ec.europa.eu/environment/water/water-framework/priority_substances.htm) like HAPs and nonylphenols. Assessment of the ecotoxicological consequences of combined exposures to high priority toxic substances was achieved using the *tpm* DNA metabarcoding approach.

2. Material and Methods

2.1. Experimental site and sampling design

Urban deposits were collected from a dry stormwater detention basin named 'Django-Reinhardt' (abbreviated Django-R) located in Chassieu (Auvergne-Rhône-Alpes region, France),

shown in the graphical abstract and in Marti et al. (2017a). This detention basin (DB) collects urban waters from a 185 ha industrial area named Mi-Plaine (45°44'12.3"N; 4°57'28.2"E). This basin handles runoff peak flows, and favors a decantation of polluted particles (sand, gravels and solid wastes) on its surface or in a central settling pit. Deposits are removed every 7 years, and last removals were recorded in February 2006 and in April 2013. Seven sampling campaigns have been considered in this investigation, and covered the 2010 and 2015 periods (**Table S1**). Details about these sampling campaigns can be obtained from Bernardin-Souibgui et al. (2018). Deposits were sampled from four "surface" points (P1, P2, P4 and P7; maximum thickness of 50 cm), the DB settling pit (P12 – at a depth of 150 cm) and a control area outside the DB (P0). The DB surface points were chosen based on the basin hydrodynamics and chemical pollutant contents (Yan et al., 2014). Samples were collected over a 50 cm^2 area by the quartering method (Gy, 2004).

2.2. High priority pollutants monitoring, and qPCR assays

Deposits were characterized by 46 variables including general physico-chemical and microbial qPCR datasets, polycyclic aromatic hydrocarbons (PAH), bisphenol A, metallic traces elements (MTE), pesticides and alkyphenol concentrations (**Table S2**). These datasets were reported in Marti et al. (2017a), Sébastian et al. (2014), and Wiest et al. (2018), and had been obtained from the samples analysed in this study. This is the first study presenting full comparisons between these chemical monitorings, and microbiological DNA patterns.

DNA extracts for the microbiological analyses were obtained from the sediments/soils using the FastDNA SPIN® Kit for Soil (MP Biomedicals, Carlsbad, France). DNAs were quantified using a nanodrop UV-Vis Spectrophotometer. DNA extracts were visualized as indicated in Colin et al. (2020). DNAs were kept at -80°C. The 16S rRNA gene, *int1*, *int2*, *int3* genes of integrons (mobile genetic elements often associated with antibiotic resistances), HF183 (a DNA

126 target for *Bacteroides* species specific of Human fecal matters) and total *Bacteroides* qPCR
127 assays were performed as described in Marti et al. (2017a, 2017b) and Voisin et al. (2019).
128 Assays were all triplicated.

129 **2.3. Phylogenetic analyses, *tpm* PCR screenings, and construction of the *tpm* database**

130 Phylogenetic analyses of *tpm* and encoded bTPMTs (bacterial thiopurine methyltransferases)
131 were performed from sequences extracted from GenBank. Multiple alignments were generated
132 using ClustalX (Higgins and Sharp, 1988). The MEGA6 software was used to compute
133 maximum likelihood or Neighbor-Joining phylogenetic trees from Kimura-2 parameters
134 divergence matrices for DNA sequences or from observed % divergences for the amino acids
135 sequences (Tamura et al., 2013). Pairwise homoplasy index (phi) analyses were performed using
136 Splitstree4 (Bruen et al., 2006; Huson and Bryant, 2006) in order to investigate the impact of
137 recombination on *tpm* sequences. The adjusted Wallace coefficient analysis (computed according
138 to Severiano et al. 2011) using UMMI web site at Lisboa University
139 (http://www.comparingpartitions.info/) was used to compare the resolution of the *tpm* CDS in
140 bacterial classification against other genetic markers.

141 Full length TPMT multiple alignments were used to confirm the presence of the well-
142 conserved VPLCGK and YDRAAM TPMT motifs reported in Favre-Bonte et al. (2005). The
143 corresponding DNA sequences of these motifs were targeted for the definition of Illumina
144 (ILMN) PCR primers to be used in a high throughput DNA sequencing of *tpm* PCR products
145 (**Table S3**). All PCR primer pairs were tested using reference DNA extracts from type strains
146 and a panel of additional strains per targeted groups of sequences (γ-proteobacteria including the
147 *Acinetobacter*, *Aeromonas*, *Legionella*, *Pseudomonas*, and *Vibrio*). PCR reactions were
148 performed with the 5X HOT BIOAmp ® Blend Master Mix using 12.5 mM $MgCl_2$, 1 µM of each
149 PCR primer, and the HotStart Taq polymerase (Biofidal, Vaulx-en-Velin, France). Only the
150 PTCF2-PTCR2 and PTCF2m-PTCR2m primer pairs yielded positive PCR products from DNA

extracts of urban deposits, and were thus conserved in this study. These primers were adapted for a use in a Nextera Illumina-sequencing pipeline of PCR products, and termed ILMN-*tpm* primers. ILMN-PTCF2 (5'- P5 adapter tag + universal primer + GTGCCGYTRTGYGGCAAGA-'3), ILMN-PTCF2m (5'- P5 adapter tag + universal primer + GTGCCCYTRTGYGGCAAGT-'3), ILMN-PTCR2 (5'- P7 adapter tag + universal primer + ATCAKYGCGGCGCGGTCRTA-'3), and ILMN-PTCR2m (5'- P7 adapter tag + universal primer + ATGAGBGCTGCCCTGTCRTA-'3) were used at equimolar concentrations to generate the PCR products from field DNA extracts (**Table S3**). The universal primer sequence was 5'-AGATGTGTATAAGAGACAG-'3. The P5 adapter tag was : 5'-TCGTCGGCAGCGTC-'3. The P7 adapter tag was : 5'- GTCTCGTGGGCTCGG-'3. PCR reactions were performed using the 5X Hot BIOAmp® master mix (Biofidal, France) containing 12,5 mM MgCl2, and with 10% DMSO and 50 ng DNA final concentrations. PCR cycles were as follow: (1) a hot start at 94°C for 5 min, (2) 35 cycles consisting of 94°C for 30 s, 58°C for 30 s and 72°C for 30 s, and (3) a final extension of 5 min at 72°C. PCR products and blank control samples were visualized as indicated above using a 2% agarose gel. PCR products showed sizes around 320 bp but blanks did not show detectable and quantifiable PCR products. Index and Illumina P5 or P7 DNA sequences were added by Biofidal (Vaulx-en-Velin, France) through a PCR procedure using the same Hot BIOAmp® master mix and the above temperatures except annealing which was performed at 55°C. PCR cycles were limited at 15. Indexed P5/P7 tagged PCR products were purified using the SPRIselect procedure (Beckman Coulter, Roissy, France). PCR products and blank control samples were analysed using the QIAxcel DNA kit (Qiagen, France), and band sizes between 400 bp were observed but not in the blank samples. Quantification of PCR products by the picogreen approach using the Quantifluor dsDNA kit (Promega, France) and a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, France) was performed, and showed low values among the blanks which were at the limit of detection (around 0,07 ng/µl). Still, *tpm*

176 harboring bacteria being in low number among a bacterial community (about 2-3%), these
177 controls were kept during DNA sequencing. MiSeq Illumina V3 paired-end sequencing of the
178 *tpm* PCR products was performed by Biofidal (Vaulx-en-Velin, France –
179 http://www.biofidal.com), and set up to obtain around 40K reads per sample. Blank samples
180 generated low number of sequences, and these have been listed in **Table S4**. Three blanks over
181 six did not yield *tpm* sequences. The other blanks led to less than 34 sequences that were further
182 compared to the *tpm* database, and *tpm* sequences from the field samples. The *tpm* DNA
183 sequences produced in this study were deposited on webin
184 (https://www.ebi.ac.uk/ena/submit/sra/#home) under the accession number PRJEB40142.
185 A database of *tpm* bacterial sequences covering the region between the VPLCGK and
186 YDRAAM motifs was built from representative sequences extracted from GenBank. This
187 database was named "BD_TPM_Mar18_v1.unique_770seq", and is available from
188 https://www.graie.org/othu/donnees. New sequences from undefined bacterial taxa were
189 imported in this database. It is to be noted that redundancy was avoided in the database by
190 keeping a single entry when identical sequences from different taxa were observed.
191
192 **2.4. Manipulation of *tpm* PCR products DNA libraries through the Mothur and DADA2**
193 **bio-informatic packages**
194 The paired-end *tpm* DNA libraries were manipulated using standard operating procedures
195 available in the Mothur suite (v.1.36.1) as described by Schloss et al. (2009), and using the
196 MiSeq standard operating procedure (SOP) developed by Kozich et al. (2013), and available at
197 http://www.mothur.org/wiki/MiSeq_SOP. DADA2 analyses were performed according to the
198 SOP of Callahan et al. (2016). Results from these two approaches were compared because one is
199 based on the use of a re-sampled dataset (Mothur) and definition of operational taxonomic units
200 (OTUs), and the other one makes use of all the sequenced reads through the definition of

amplicon sequence variants (ASVs) but implies a deeper analysis of sequence quality scores and frequencies using a probabilistic noise model for nucleotide transitions. Comparison of these two approaches has rarely been reported. The SOP from Mothur allowed removing chimeric sequences, primers, barcodes, and limited the dataset to sequences of a minimum length of 210 bp (average length=215 bp). Aligned sequences were screened and filtered according to length (going from position 1 to 269). The pre.cluster command was applied on the dataset in order to mask up the remaining sequencing errors. The split.abund cut off was set at 1 to remove singletons. Chimeric sequences were searched by chimera.uchime, and removed using the remove.seq command. A sub-sampling of sequences was performed at 3600 reads on the cleaned dataset (**Table S1)**. OTU clustering was performed at an identity threshold of 100% (leading to exact sequence variants), and these OTUs were allocated to a taxonomic group using the "Wang" text-based Bayesian classifier (Wang et al., 2007) against the "BD_TPM_Mar18_v1.unique_770seq" database, and using a bootstrap cutoff of 80%. The contingency table for the *tpm* OTUs is shown in **Table S5a**, and their matching sequences are shown in **Table S5b**. The original OTU table was converted into genus and species-like groupings (**Table S5c, d, e**).

The DADA2 SOP was applied on the same dataset. Raw reads were trimmed in order to remove barcodes and primers using Trimgalore (--paired, --clip_R1 19, --clip R2 20, -q 15) (https://github.com/FelixKrueger/TrimGalore). DADA2 v3.10 was then used. Briefly, reads were filtered and trimmed (maxN=0, maxEE=c(2,2), truncQ=2, truncLen=c(200,200), minLen = 126) and the learnErrors command was applied, and the reads were merged into a table of ASVs. Chimeras were removed (method="consensus") and taxonomy was assigned using the assignTaxonomy command (minBoot=80) along with the "BD_TPM_Mar18_v1_unique_770seq" database. These taxonomic allocations implied the use of a Wang text-based Bayesian classifier. In order to remove non-*tpm* reads from the ASV table,

the align.seq command from the Mothur package was applied using the "BD_TPM_Mar18_v1_unique_770seq" database as reference and a minimum of 150 nucleotides alignment as a cutoff. The decontam package (Davis et al., 2018) was used along with the phyloseq package (McMurdie and Holmes, 2013) in order to remove contaminant sequences found in the negative controls (**Table S4**). Contaminants were identified with the "frequency" method and removed. The contingency table for the *tpm* ASVs is shown in **Table S6a**, and their matching sequences are shown in **Table S6b**. These ASVs were grouped into genus and species-like groupings for some analyses (**Table S6c, d, e**).

Between 25% (ASV) to 33% (OTU) of the *tpm* sequences from the detention basin could not be attributed to well-defined lineages among the *Bacteria.* The unclassified OTUs (the OTU contingency table showed more of these than the ASV one) were compared by molecular phylogeny, and groups of at least three distinct OTUs (with a bootstrap p-value >90%) were further analysed by BLASTx comparisons (**Table S7**). This allowed the definition of twenty-seven novel sequence types in the database for the α, β, δ, and γ-proteobacteria, and the *Spirochaetes* and *Nitrospirae*. Species contingency tables (**Tables S5c, d, e and S6c, d, e)** were used to define the core *tpm* bacterial groups (i. e. species among all samples) of urban deposits. The DESEQ2 package (Love et al., 2014) was used to highlight species with significant changes in abundance in the deposits/sediments of the DB over time and space, as recommended by Callahan et al. (2016). Kruskal-Wallis (kruskal.test function from the vegan package) and Dunn tests (*post-hoc* Holm analysis, kwAllPairsDunnTest function, PMCMRplus package (Pohlert, 2018); https://CRAN.R-project.org/package=PMCMRplus) were performed to highlight significant differences in distribution patterns of *tpm* reads allocated to *Aeromonas* and *Pseudomonas* species.

2.5. Metabarcoding *tpm* diversity and community analyses

Rarefaction curves were computed according to the ASV and subsampled OTU *tpm* contingency tables using the rarecurve function of the vegan package (Oksanen et al., 2017; https://CRAN.R-project.org/package=vegan) (option raremax). Diversity indices (Shannon, Simpson and Shannon-Evenness) and richness were calculated with the Mothur suite from the *tpm* subsampled OTU, and ASV tables (**Tables S5a** and **S6a**). Kruskal-Wallis and Dunn tests were applied to assess significant differences in diversity indices and richness. Hellinger transformation was applied on both *tpm* subsampled OTU and AVS tables (decostand function, method='total'). Bray-Curtis dissimilarity matrices were then computed (vegdist function from the vegan package, option "bray"), and used to generate Non-metric MultiDimensional Scaling (NMDS) plots (metaNMDS function). The ordispider option was used to visualize groups (drawn from the centroid of the confidence ellipse defined by the standard deviations from the computed ordinations), and adonis tests (999 permutations) were performed to assess the significance of these groupings (pairwise.adonis and vegdist (option "bray") function implemented in the pairwise.adonis package (Arbizu, 2017; https://github.com/pmartinezarbizu/pairwiseAdonis). To validate differences between *tpm* community profiles between urban deposits against a soil control sample recovered outside the DB (P0) (shown in Marti et al. 2017a), discriminatory tests were applied on the diversity indices, richness and pairwise *tpm* OTU dissimilarities.

NMDS (envfit function), Redundancy Analysis (RDA, rda function, according to the Akaike's Information Criterion) and Canonical Correspondance Analysis (CCA, cca function, followed by ANOVA, and ANOVA by term) were performed on each set of variables (nutrients, MTE, PAH, pesticides, bisphenol A, alkyphenols, integrons *int* DNA targets, and MST (microbial source tracking) DNA markers) to highlight those significantly related to the *tpm* community patterns (**Tables S5 and S6**). The *tpm* ASV and subsampled OTU contingency tables were transformed according to the requirements for NMDS, RDA and CCA analyses. Chemical

variables were log10 standardized (decostand function, "log"). Permutations (n=999) were performed to test the significance of the observed correlations.

3. Results

3.1. Phylogenetic reliability of the *tpm*-inferred taxonomic allocations

A total of 721 reference proteins covering the diversity of the bTPMT family involved in the detoxification of metalloid-based toxic substances was extracted from GenBank. Bacterial TPMT coding regions (CDS) were observed among more than 130 genera. To assess the genetic stability of the *tpm* CDS, phi-tests (pairwise homoplasy index) were performed. No significant recombination was found among most of the tested genera (n=9/11) ($p > 0.05$). When signs of recombination were detected, a good congruence between the *tpm* and 16S rRNA gene phylogenetic tree topologies was still observed (**Fig. S1 and S2**). Nevertheless, a few inconsistencies between *tpm* phylogenies and reported taxonomic allocations among the *Xanthomonas* and *Pseudomonas* were observed. These were likely related to misleading classifications (**Fig. S3**). It is to be noted that the taxonomy of *Xanthomonas* has been the subject of considerable debates, and many reclassifications have been done (Schaad et al., 2006, 2005). In a few instances, *tpm* sequences reported from different species of *Pseudomonas* were found to harbor identical *tpm* sequences (**Fig. S4**). To keep record of these discrepancies, complexes of species were defined and indicated in the annotations of the *tpm* database built for the metabarcoding analyses.

Infraspecific genetic variations of the *tpm* CDS were investigated for some bacterial species to further assess the reliability of the *tpm* marker. Specifically, *tpm* CDS from 65 whole genome sequences of *P. aeruginosa* (Pae) were analyzed (**Table S8**). Phi-tests detected no significant level of recombination in this species (p=0.89). Furthermore, the comparison of phylogenetic trees based on the full *tpm* Open Reading Frame (ORF) with the one inferred from the MLST (multi locus sequence typing) scheme gene loci (the one from PubMLST - Public databases for

molecular typing and microbial genome diversity at https://pubmlstg/) confirmed a reliable allocation of *P. aeruginosa tpm* reads into sub-clades termed PAO1, PA14, PA7, and PA39 (**Fig. S5**). Adjusted Wallace coefficient analysis confirmed the full *tpm* ORF (tpm_{FL}) to reliably predict sub-clade classifications among *P. aeruginosa* (tpm_{FL} → *pae* clade, probability (prob) of proper allocation = 82%) (**Fig. S6** and **S7**). The CDS area covered by the ILMN-*tpm* primers being about one third of the ORF, a loss in discriminatory power was observed (tpm_{SL} → clade prob differentiation=59%) but differentiation at subspecies levels remained possible, especially for distinguishing clades PAO1/PA14 from clades PA7 and PA39 (tpm_{SL} → clade prob differentiation=84%) (**Fig. S7 and S8**). Similarly, the reliability of the *tpm* phylogenetic groupings were also demonstrated for *P. syringae* (Psyr) among which several pathovars have been differentiated (**Table S9, Fig. S9**). The distribution of *tpm* sequence types was assessed among *P. syringae* (**Fig. S10 and S11**), and adjusted Wallace coefficients indicated high probabilities to predict a pathovar from both full tpm_{FL} CDS (probability to differentiate pathovars =90%), and the short tpm_{SL} segment used in the metabarcoding scheme (probability to differentiate pathovars =74%) (**Fig. S7**).

These analyses allowed to build the ILMN (Illumina)-*tpm* database used in section 3.4. The taxonomic status of each unique *tpm* sequence found between the ILMN *tpm* primer pairs (**Table S3**) from GenBank was declined into its respective phyla, class, order, family, genus, species and subspecies or pathovar taxonomic level. If several species or sub-species/pathovars were detected for a same tpm_{SL} sequence, all names were kept in the taxonomic record of the sequence. A phylogenetic tree was built from the ILMN-*tpm* alignment to detect redundant sequences. A single sequence was kept per ILMN-*tpm* sequence type. This led to an ILMN-*tpm* database covering 8 bacterial phyla, 15 classes, 35 orders, 66 families, 132 genera, 461 species and 46 subspecies or pathovars. Unique sequences that could not be attributed to well-defined bacterial groups were classified as novel TPMT lineages using letter codes defined according to

the nomenclature of Favre-Bonte et al. (2005). A total of 79 ILMN-*tpm* lineages were defined by this approach.

3.2. Metabarcoding *tpm* diversity and community analyses of urban deposits

The *tpm* metabarcoding dataset was built from 40 samples which led to 1.73 million raw reads among which 540 277 (DADA2 processing) to 568 633 reads (Mothur processing) were confirmed to cover the DNA sequences encoding the VPLCGK and YDRAAM motifs of bTPMTs. ASVs and OTUs were defined at 100% identity (exact sequence variants). This led to a total of 12 217 OTUs and 2 412 ASVs (**Tables S5a and S6a**), ranging from 88 to 634 OTUs, and 95 to 617 ASVs per sample (**Table S1**). Contaminant *tpm* reads among the blanks were either absent or in low numbers (<34 reads) (Table S4). For the ASV-based analyses, sequence variants identified in the blank samples, considered as being contaminants, were deleted from the dataset. Among the OTU dataset, samples with < 3,600 reads were deleted (**Table S1**). The DADA2 approach for filtering reads appeared more stringent than the one used in the Mothur SOP. Rarefaction curves are presented in **Fig. S12**. P0 OTU and ASV rarefaction curves reached the asymptote faster than the ones derived from the urban deposits (P1 to P12). P0 richness was found significantly lower than the one of the P12 samples for the OTU dataset, and of the P1, P2 and P12 samples for the ASV one (**Table S1;** Kruskal-Wallis, $p<0.05$). For the ASV dataset, P0 showed significantly lower values than those of the P1 and P2 samples for the Shannon index, and higher values than P1 for the Simpson index (**Table S1;** Kruskal-Wallis, $p<0.05$). This indicated a significant difference in the *tpm* diversity patterns between sediments recovered in a detention basin (DB), and urban soils. It is noteworthy that indices computed from the OTU and ASV datasets were found to be correlated (**Table S1**). Spearman Rank Correlation tests comparing the richness values based on the OTU and ASV contingency tables gave a rho=0.6, of 0.7 for the Simpson and Shannon indices, of 0.8 for the Shannon-Evenness indices, and $p<0,01$ which are indicative of significant tests.

NMDS ordinations computed from the Bray Curtis matrices of dissimilarities (**Tables S10 & S11**) between pairwise *tpm* ASV or OTU profiles, including all samples but differentiated by sampling zone, confirmed a clear differentiation of the P0 control samples against the DB ones (**Fig. S13, Table S12**, Adonis tests, $p<0.05$). Adonis tests indicated a significant differentiation of the urban sediments *tpm* profiles from P0 using the ASV dataset, and of those from P1, P4 and P12 from those of P0 using the OTU dataset (**Table S12; p<0,05**). The P2 and P7 profiles of the OTU dataset against those of P0 showed a clear differentiation but the adjusted P values for multiple tests were not significant. NMDS ordinations (ASV and OTU) and associated Adonis tests without P0 samples discriminated *tpm* profiles according to the deposits maturation time periods in the detention basin and according to their origin (**Fig. 1, and Fig. S14**). This led to the definition of the following categories for further statistical analyses of the datasets: (1) recent deposits/sediments having settled in the DB between 6 to 15 months, (2) old sediments having settled in the DB for more than 24 months, (3) surface sediments accumulating over the concrete slab, and (4) the settling pit sediments (**Fig. 1**).

3.3. Spatio-temporal variations of high risk toxic pollutants among urban deposits, and their correlations with *tpm* gene metabarcoding profiles

All values of high risk chemical pollutants monitored during these investigations have been compiled in **Table S2**. Principal component analyses (PCA) and other statistical analyses of these variables were previously reported (Marti et al. 2017a; Wiest et al. 2018). Nevertheless, PCA were performed to verify the datasets. These analyses were in agreement with those that were previously published for the PAHs and MTE contents of these deposits/sediments, and confirmed P0 (control zone outside the detention basin (DB)) to be contaminated by chemical pollutants such as PAHs. Furthermore, higher concentrations in alkyphenols (para-ter-octylphenol (A1), nonylphenol-mono-ethoxyle (A4) and nonylphenol-bi-ethoxyle (A5)) were

376 confirmed to be significantly associated with the recent sediment samples from the decantation
377 pit, while higher 4-nonylphenol (A2) concentrations were associated to samples that had settled
378 over a long time period in the decantation pit (**Table S2; Fig. S15**). Bisphenol A (A3)
379 concentrations were confirmed to be significantly higher ($p<0.05$) among recent surface samples
380 (**Fig. S15**). No significant concentrations of dibenzo(a,h)anthracene (H15), diuron, isoproturon,
381 and trifluraline were observed (**Table S2**).

382 To infer the impact of all these substances on the *tpm* community structure defined by OTUs
383 and ASVs, envfit on NMDS ordinations, RDA, and CCA were performed. RDA, CCA and envfit
384 confirmed the significant relationships ($p<0.05$) between accumulation time periods (old against
385 recent) and location in the detention basin (surface against pit), and the *tpm* community profiles
386 (**Table 1**; **Table S13a, b** for RDA and envfit and **Table S13c, d** for CCA). Cadmium, nickel,
387 and chrome contents were found to match significantly both *tpm* datasets through envfit, and
388 partially by CCA analyses (**Table 1**; **Table S13**; $p<0.05$). Regarding PAH concentrations,
389 several forms showed significant variations matching the *tpm* profiles. Most significant
390 relationships were observed for anthracene, chrysene, benzo(b)fluoranthene, fluorene,
391 naphthalene, through envfit, RDA, or CCA (Table 1). Significant relationships were also
392 observed by envfit, RDA, or CCA with the concentrations of chlorpyrifos, 4-nonyphenol and
393 nonylphenol-mono-ethoxylate concentrations (**Table 1; Table S13**). However, relationships
394 between *tpm* ordinations, and contents in total C, N, and P were not significant (**Table S12**).

395 CCA of *tpm* ASV and OTU profiles from urban deposits / sediments constrained by
396 accumulation time periods and / or position of the sampling points, and the monitored
397 concentrations of high risk priority pollutants were performed (**Table S14**, and **Fig. S16**, **S17,**
398 **S18**). All significant CCA (anova) tests are shown in **Table 1**, and most significant ones
399 (observed with the OTU and ASV datasets) are shown in **Fig. 2.** Other significant CCA are
400 presented in the supplementary materials (**Table S14**; **Fig. S19 and S20**). In brief, using both the

OTU and ASV datasets, significant CCA ordination patterns that linked accumulation time periods and sampling location of the urban sediments/deposits, and variations in the concentrations of (i) chrysene, (ii) chlorpyrifos, or (iii) 4-nonylphenols, were observed. These CCA ordination patterns also significantly linked the accumulation time periods and location, and the qPCR datasets indicative of variations in the occurrences of (iv) *int2*, (v) *int3*, or (vi) total *Bacteroides* cells (**Table 1; Fig. S19 and S20**). High chrysene and chlorpyrifos concentrations were found associated to community structures of the recent surface deposits while high nonylphenol contents matched the *tpm* patterns of the old pit deposits (**Fig. 2**). The CCA (anova) of the ASV dataset (but not for the OTU patterns) constrained by anthracene (H6) was also found significant, and H6 was found in higher amounts among the recent surface deposits (**Table 1; Fig. S20**). Naphthalene (H1) concentrations were found related to community structures of the old decantation pit deposits only using the ASV dataset (**Table 1**, **Fig. S20**). Significant relationships with variations in fluorene (H4) were also detected using the ASV ordinations constrained by the accumulation time periods (**Fig. S17**). CCA (anova) of the OTU contingency table constrained by anthracene (H6), or benzo(a)anthracene (H9), and the accumulation period were also significant (**Table 1; Fig. S16**). CCA (anova) of the ASV dataset also showed significant relationships between accumulation periods and sampling position, and bacterial contents in HF183 human fecal contaminations (**Table 1, Fig. S20**). Significant human fecal contaminations (not coming from the sewer system) over the commercial zone appeared to have impacted the *tpm* diversity observed in recent deposits but also of sediments recovered from the DB settling pit.

3.4. Taxonomic allocations of *tpm* OTUs

3.4.1. Spatio-temporal dynamics at the levels of bacterial phyla and genera

425 The full taxonomic allocations of the *tpm* ASVs and OTUs are shown in **Tables S5** and **S6**,
426 and summarized in **Fig. 3**. Most sequences were attributed to the *Proteobacteria* but a group
427 belonged to the *Terrabacteria* (*Crinalium* of the *Cyanobacteria*), and one to the *Nitrospirae*
428 (**Fig. 3**). *Crinalium* are large size cyanobacteria with poor environmental records (Mikhailyuk et
429 al., 2019). This is the first report of *C. epipsammum* among urban samples, and on the French
430 territory. Between 25% (ASV) to 33% (OTU) of the *tpm* sequences from the detention basin
431 (**Tables S5 and S6**) could not be attributed to well-defined lineages among the *Bacteria* using
432 the *tpm* metabarcoding database. Nevertheless, about 20% of these unclassified *tpm* sequences
433 could be grouped into clusters of more than three OTU/ASV (bootstrap p-values > 90%), and
434 were further investigated by BLASTx for similarity searches with well-defined bTPMTs (**Table**
435 **S7**). These translated *tpm* reads showed, respectively, 70% similarities with the TPMT from
436 *Alteromonas*, 92% with the TPMT from *Archangium*, 73% with the one from *Azoarcus*, 82%
437 with the one from *Collimonas*, 88-90% with the one from *Dechloromonas*, 77-79% with the one
438 from *Hyphomicrobium*, 72% with the one from *Haliangium*, 78% with the one from *Leptospira*,
439 71% with the one from *Methylotenera*, 69% with the one from *Methylomonas*, 71 to 83%
440 similarities (CJ lineage) with the one from *Myxococcales*, 85-87% similarities (DE lineage) with
441 the one from *Nitrospira*, 82% similarities with the one from *Nitrotoga*, and 66% with the one
442 from *Sideroxydans*. The status of these novel *tpm* reads was used to improve the *tpm*
443 metabarcoding taxonomic allocations.
444 Among all the resolved taxonomic allocations, reads allocated to the γ class of *Proteobacteria*
445 represented 50% of the lineages, the β class 8%, and the α 3%. Among the γ-*Proteobacteria*, the
446 *tpm* lineages from the *tpm* dataset were attributed to the *Pseudomonadaceae*, *Shewanellaceae*,
447 *Aeromonadaceae*, and *Xanthomonadaceae*. Some of these lineages were found to be related to
448 well-defined genera such as the *Pseudomonas*, *Azotobacter*, and *Aeromonas* (**Fig. 3**). Reads
449 allocated to the *Pseudomonas* were found among all samples, and those allocated to the

450 *Aeromonas* and *Herbaspirillum* were found among all recent sediment samples (**Fig. 4**).
451 DESEQ2 analyses of distribution biases indicated significant variations for about 30 genera
452 among the OTU dataset and 36 among the ASV one. These analyses were in agreement but the
453 DADA2 dataset led to more significant results (**Fig. 4**). DESEQ2 showed a significant ($p<0.05$)
454 decrease in *Pseudomonas* (3.5 log2 Fold Change for the OTU dataset (FC^{OTU})), *Aeromonas* (5.9
455 log2 FC^{OTU}), *Nitrosomonas* (1.9 log2 FC^{OTU}), *Herbaspirillum* (8.6 log2 FC^{OTU}),
456 *Stenotrophomonas* (2.5 log2 FC^{OTU}) and *Aquaspirillum* (3.4 log2 FC^{OTU}) *tpm* reads during the
457 sediment aging process in the DB (**Fig. 4; Tables S15 and S16**). Reads from several unclassified
458 taxa (e. g. CC5-8, CC3, CG1, CZ1, DE, CJ2-3 and DC2) showed a significant increase over time
459 in the DB sediments (**Fig. 4; Tables S15 and S16**). Among these latter taxa, only the DE (1.5
460 log2 FC^{OTU}) and CJ2-3 (6 log2 FC^{OTU}) *tpm* reads could be further related to known taxa i. e.
461 *Nitrospira*, and *Myxococcales*, respectively, through BLASTx analyses (**Table S7**). Most of
462 these unclassified lineages were among the lineages (CG1, CC8, CZ1, DC2, BD, DD2, CJ2,
463 DD1, CB, CL, CR, CJ1, DE; 7 to 2 log2 FC^{OTU}) that were found in higher numbers in the surface
464 deposits rather than the pit ones (**Fig. 4**). A single group named OTU_genus_53 showed a
465 significant increase in the pit sediments by DESEQ2, but could only be allocated to an unknown
466 group of γ-*Proteobacteria* (**Fig. 4**). It is to be noted that analyses performed from the OTU
467 contingency table allowed a detection of slightly more genera than those performed from ASV:
468 (1) *tpm* OTUs were found allocated to eighteen genera (plus twenty-six genus-like lineages) with
469 reads from *Cellvibrio*, *Achromobacter*, *Spirosoma*, candidatus mu-proteobacteria, *Rugosibacter,*
470 and *Methylovulum* only found in this dataset; (2) *tpm* ASV were allocated to sixteen genera (plus
471 thirty genus-like lineages) with reads from *Azoarcus*, *Curvibacter* and *Marinobacter* only found
472 in this latter dataset.
473
474 **3.4.2. Inferred spatio-temporal dynamics at the scale of bacterial species**

All taxonomically well-defined species recorded in this study are indicated in **Fig. 3**. It is to be noted that analyses performed from the OTU contingency table allowed a detection of more species than those performed from ASV. *S. maltophilia/pictorum*, *A. veronii*, *P. fluorescens, P. mendocina, P. synxantha,* and *P. pelagia tpm* reads were recorded among the OTU contingency table but not the ASV one (**Tables S5a and S6a**). Nevertheless, in a one instance, the ASV approach detected *tpm* sequences from a species not recorded in the OTU table i. e. *P. hunanensis*. All OTUs and ASVs of **Tables S5a and S6a** that could be grouped into well-established species or equivalent groupings (species-like) were converted into taxonomic entities named OTU^{sp} (Table S5c) or ASV^{sp} (Table S6c). A total of 119 and 110 species groupings were generated respectively for the OTU and the ASV datasets, with a distribution going from 19 to 56 OTU^{sp} and 21 to 55 ASV^{sp} per sample (Table S5c and S6c). **Fig. 5** indicates the core bacterial species per sample categories. $OTU^{sp}001$/$ASV^{sp}001$ (unknown proteobacteria), $OTU^{sp}003$/$ASV^{sp}011$ (unknown phylum), *Pseudomonas sp.* $OTU^{sp}002$/$ASV^{sp}002$ (PGPPP3-like), $OTU^{sp}005$/$ASV^{sp}003$ (an unknown *gammaproteobacteria*), *Pseudomonas* sp. $OTU^{sp}007$/$ASV^{sp}008$, and the *Proteobacteria* Gp_DE $OTU^{sp}008$/ $ASV^{sp}14$ were found among all DB samples, and were considered to be part of the core microbiome of urban sediments. Considering the recurrent *tpm*-harboring species found among the recent deposits, the following species can also be added to this core microbiome: $OTU^{sp}019$ / $ASV^{sp}022$ of the *Pseudomonadaceae*, $OTU^{sp}012$/$ASV^{sp}009$ of the Gp_CH allocated to the *Proteobacteria*, and $ASV^{sp}016$/$OTU^{sp}013$ allocated to *Herbasprillum aquaticum*. It is to be noted that *Pseudomonas sp.* $OTU^{sp}002$/$ASV^{sp}002$ (PGPPP3-like), *Nitrosomonas oligotropha* $ASV^{sp}007$/$OTU^{sp}010$, *Herbasprillum aquaticum* $ASV^{sp}016$/$OTU^{sp}013$, *P. koreensis* $ASV^{sp}021$/$OTU^{sp}020$, *P. anguilliseptica* $ASV^{sp}016$/$OTU^{sp}021$, and *Stenotrophomonas terrae* $ASV^{sp}025$/$OTU^{sp}027$ and *S. acidaminiphila* $ASV^{sp}031$/$OTU^{sp}033$ showed the highest number of *tpm* reads among the well-defined species (**Tables S5** and **S6**).

The spatio-temporal dynamics of the OTU^{sp} and ASV^{sp} bacterial groups among the DB were investigated (**Fig. 5 and 6**). DESEQ2 analyses indicated significant variations for 28 OTU^{sp} and 38 ASV^{sp}. Overall, eleven over eighteen OTU^{sp} (and 29 over 36 ASV^{sp}) were found to be significantly more abundant in the recent than the old sediments (**Fig. 5**; **Tables S17 and S18**). Using both species contingency tables (ASV^{sp} and OTU^{sp}), *Pseudomonas sp.* ($OTU^{sp}002$ – PGPPP3-like; 5.5 log2 FC^{OTU})), *Pseudomonas anguilliseptica* (2.2 log2 FC^{OTU}), the GpI lineage (belonging to the *Pseudomonadaceae*, $OTU^{sp}016$ / $ASV^{sp}015$; 2.1 log2 FC^{OTU}), *Aeromonas hydrophila* ($OTU^{sp}017$/ $ASV^{sp}030$; 6.6 log2 FC^{OTU}), *Aeromonas* sp. ($OTU^{sp}022$/$ASV^{sp}020$; 4.4 log2 FC^{OTU}), *N. oligotropha* ($OTU^{sp}010$/ $ASV^{sp}007$; 2.8 log2 FC^{OTU}), *H. aquaticum* (8 log2 FC^{OTU}) as well as three additional undefined Proteobacterial species (Gp_CH ($OTU^{sp}012$/ $ASV^{sp}009$), Gp_CF ($OTU^{sp}014$/ $ASV^{sp}012$) and Gp_CO ($OTU^{sp}024$/$ASV^{sp}024$) were found significantly more abundant in the recent sediments (**Fig. 5,** and **Tables S17 and S18**). Furthermore, the ASV^{sp} dataset allowed to detect additional species undergoing a decrease over time in the detention basin such as *A. eucrenophila* (6.4 log2 FC^{ASV}), *A. media* (5.7 log2 FC^{ASV}), and *Nitrospira defluvii* (2.4 log2 FC^{ASV}) (**Fig. 5**; **Table S18**). Regarding the *tpm* bacterial species which had undergone an increase in the older sediments, these were mainly limited to unclassified lineages: CJ2, CZ1-2 DC2, and CC3, CC5, CC7-8. They had DESEQ2 log2 fold change increasing from 4.4 to 7.2, and the CC8 species was among the most abundant *tpm*-harboring species found in the aged sediments (**Tables S17** and **S18**). These taxa were all found among the surface sediments. $OTU^{sp}005$ / $ASV^{sp}003$ was the only species (an unknown *gammaproteobacteria*) found to be significantly more abundant in the pit deposit samples than the surface ones (**Fig. 5**).

Pairwise Kruskal-Wallis tests were performed to complete these analyses because DESEQ2 is highly sensitive to highly variable distribution patterns. These tests were limited to *tpm* reads allocated to species of the *Pseudomonas* and *Aeromonas*. These two genera showed the most

important diversity in terms of number of species detected among the urban sediments. Thirty-three *Pseudomonas* OTU^{sp} / ASV^{sp} lineages were detected and allocated to 27 species (**Fig. 6, Tables S5d and S6d**). Several species were found significantly more abundant in the settling pit than among the surface sediments: *P. koreensis*, *P. rhodesia*, *P. umsongensis*, *P. reinekei*, *P. nitroreducens*, *P. mendocina*, *P. hunanensis*, *P. fluvialis*, and *P. azotoformans*. However, *P. anguilliseptica tpm* reads, *P. anguilliseptica*-like sp. reads, and those allocated to *P.* sp. PGPPP3-like *tpm* sequences (OTU^{sp}002/ASV^{sp}002) and a *P.* unclassified lineage (OTU^{sp}007/ASV^{sp}008) were found significantly more abundant in the recent urban deposits (Kruskal-Wallis tests; $p<0.05$) while *P. syringae tpm* were significantly more abundant in the older sediments (**Fig. 6**; **Tables S5d** and **S6d**). The *P. syringae* OTUs were further allocated to pathovars when analysed by molecular phylogeny using the dataset of **Fig. S11** and **Table S9**. Otu10054 (type= tpm_psyr6:7), otu00171 (tpm_psyr22), and otu04275 (type= tpm_psyr5) (**Table S5a**) were, respectively, attributed to the lineages of (i) *P. s.* pv. *aceris*, (ii) *P. s.* pv. *pisi*, and (iii) *P. s.* pv. *lachrymans/tomato*. Similarly, *tpm* OTUs allocated to *P. aeruginosa*, which were found to have a relatively conserved distribution pattern over time and space, could be allocated to sub-clades using datasets of **Fig. S8** and **Table S8**. Otu05807 matched *tpmG* types pae12:15-17-18-35 (LESB58-like; PAO1 sub-clade); Otu03611 matched *tpmG* pae11 type (PAO1 clade); Otu01667 = *tpmG* pae19:21-25-27:29 (PAO1 clade); Otu00259 = *tpmG* pae8-24 (PA14 clade); Otu01855 = *tpmG* pae32 (PAO1 clade); and Otu03914= *tpmG* pae30 (PA14 clade) (**Table S8**). Concerning the OTU^{sp} / ASV^{sp} allocated to *Aeromonas* species, sixteen species could be differentiated (**Fig. 6**). Most of these species were found significantly more abundant in the recent urban deposits of the DB than the older ones (**Fig. 6; Tables S5e** and **S6e**). The *tpm* reads allocated to *A. eucrenophila*, *Aeromonas* Gp_CI1, *Aeromonas* Gp_CI2, *A. hydrophila*, *A. media*, *A. salmonicida*, and two *Aeromonas* sp. were found significantly more abundant in the recent urban deposits of the detention basin than among the more mature sediments (Kruskal-Wallis tests;

550 p<0.05) (**Fig. 6).** These analyses were in agreement with those performed using DESEQ2 (**Fig.**
551 **5**) but detected more significant trends in the datasets.
552
553 **4. Discussion**
554 The ecological consequences of combined high risk chemical pollutants on urban biomes
555 were investigated through an analysis of road deposits and sediments recovered from a
556 stormwater infiltration system (SIS). The bacterial community involved in the detoxification of
557 metalloid-based toxic substances through the expression of bTPMTs encoded by the *tpm* gene
558 was used to assess the incidence of these pollutants. A *tpm* DNA metabarcoding approach was
559 designed and applied to investigate the spatio-temporal dynamics of these *tpm*-harboring
560 bacterial taxa. This approach was built from exhaustive phylogenetic analyses of the bTPMTs.
561 These bTPMT gene sequences (*tpm*) were found distributed mainly among *Proteobacteria*, and
562 about 60% were allocated to the γ-*Proteobacteria*. They were also recorded among the
563 *Actinobacteria* (*Actinokineospora, Actinoplanes, Saccharothrix, Actinosynnema*), *Bacteroidetes*
564 (*Spirosoma* and *Fibrisoma*), *Chloroflexi* (*Herpetosiphon*), *Cyanobacteria* (*Leptolyngbya,*
565 *Pseudanabaena, Leptolyngbya, Crinalium, Halomicronema,* and *Pseudanabaena*), *Nitrospirae*
566 (*Nitrospira*), *Spirochaetae* (*Leptospira* and *Leptonema*), and *Verrucomicrobia* (*Terrimicrobium*).
567 The *tpm* gene was not found to have undergone horizontal gene transfers (HGT), but to have
568 strictly evolved vertically among the Bacteria which have conserved this gene over time. These
569 analyses confirmed the reliability of the *tpm* gene as a DNA metabarcoding marker enabling
570 taxonomic allocations down to the species level. Infra-specific *tpm* gene sequence
571 diversifications could also be resolved, and be used to differentiate pathovars and sub-clades.
572 The *tpm* metabarcoding scheme reported here allowed an investigation of the diversity among all
573 tested DNA extracts obtained from urban sediments and deposits. The *tpm* community profiles
574 of urban deposits/sediments were found to harbor about 50% of γ-*Proteobacteria*, 8% of β ones,

575 and about 3% of α-*Proteobacteria*, and to contain sequences allocated to three other phyla, the
576 *Cyanobacteria* (*Crinalium*), *Bacteroidetes*, and *Nitrospirae.* Among the γ-*Proteobacteria*, most
577 lineages were attributed to the *Pseudomonadaceae*, *Shewanellaceae*, *Aeromonadaceae*, and
578 *Xanthomonadaceae*.

579 It is to be noted that the *tpm* metabarcoding DNA sequences were analysed using both the
580 Mothur and DADA2 SOP. Comparative analyses of contingency tables generated by these two
581 bio-informatic packages have been rarely reported, so far (Olson et al. 2020; Prodan et al. 2020).
582 DADA2 had the advantage of using the full *tpm* dataset to define amplicon sequence variants
583 (ASV), while Mothur recommended the use of a random sub-sampling to obtain identical
584 number of reads per sample for defining operational taxonomic units (OTU; defined threshold at
585 100% identity). Nevertheless, these approaches led to strongly correlated datasets. The most
586 significant trends in terms of *tpm* diversity indices were found correlated (rho=0.6-0.8, p<0,01),
587 and community changes inferred from NMDS, RDA or CCA were in agreement. However, the
588 Mothur SOP detected more sequence variants (12 217 OTUs against 2 412 ASVs). Taxonomic
589 allocations of these OTUs/ASV led to similar numbers of bacterial genera but each approach
590 detected a few genera not recorded by the other one. These differences concerned taxa with low
591 read numbers. Similarly, the OTU contingency table allowed to detect slightly more species than
592 the ASV one but these also concerned taxa with low read numbers. DADA2 still generated an
593 ASV dataset which revealed more significant changes in species read numbers by the DESEQ2
594 statistical tests. These approaches appeared to be complementary. The ASV and OTUs datasets
595 were always compared in this study to find the most significant ecological changes among the
596 *tpm*-harboring bacteria.

597 Ordination analyses (NMDS, RDA, CCA) of *tpm* ASV and OTU profiles from urban deposits
598 / sediments tested by permutated ANOVA tests were performed. These approaches differentiated
599 significantly (p<0.05) the *tpm* genetic profiles (based on ASV and OTUs) according to the

accumulation time periods (old (over two years) against recent (from 6 to 15 months)) of deposits and sediments in the detention basin of the SIS, and their sampling position (surface or settling pit). Furthermore, these ordinations were found to match variations in chemical pollutants. For PAHs and metallic trace elements (MTE), (i) significantly higher [chrysene] were associated with the *tpm* profiles from recent surface deposits, (ii) significantly higher [fluorene] with those of the settling pit samples, (iii) significantly higher [naphthalene], and [Cd, Cr, Ni] were associated to the *tpm* profiles of the older surface sediments, (iv) significantly higher [anthracene / or benzo(a)anthracene] with the ones of the surface deposits, and (v) significantly higher [benzo(b)fluoranthene] were associated to the *tpm* profiles of the old deposits. These positive relationships between *tpm* metabarcoding profiles, and [MTE] and [PAHs], can be related to multiple resistance or degradation processes that could be expressed from *tpm* collinear genetic loci. Several *tpm*-harboring taxa have been shown to encode key enzymes for the oxidation of hydrocarbons (Rojo and Martínez, 2019) or the transformation of MTEs (Deredjian et al., 2011), and thus likely encoded the metabolic pathways required for a growth among urban deposits. Otherwise, negative relationships between *tpm*-harboring bacteria and chemical pollutants were probably related to the genotoxicity of some of these pollutants and their related derivatives e. g. methylated fluorenes (Rice et al. 1987). In fact, ecotoxicological assays performed by the ostracod (*Heterocypris incongruens*) sub-chronic assay indicated a higher toxicity of the sediments recovered from the settling pit than from the surface of the detention basin (Becouze-Lareure et al., 2018). However, these ostracod assays did not reveal changes in ecotoxicity over time (Becouze-Lareure et al., 2018). The *tpm* metabarcoding approach appeared more sensitive for a detection of significant changes in the ecology of these systems. All these trends were in full agreement with those previously reported from the analysis of 16S rRNA gene meta-barcoding datasets generated for the same samples (Marti et al. 2017b). The 16S rRNA gene profiles of these samples were also found to match with their content in PAHs and

625 MTE. Using the envfit approach on NMDS ordinations, significant relationships between the
626 16S rRNA gene profiles, and Cd, Cr, Ni, fluorene, and chrysene concentrations were observed,
627 and RDA showed high significant relationships with chrysene concentrations. High chrysene
628 levels had previously been observed in the recent surface deposits, fluorene in the decantation
629 pit, and MTE in slightly higher concentrations in the more mature sediments (Marti et al.,
630 2017a). An accumulation over time of these latter pollutants had thus also been observed, and
631 considered to have driven changes in the bacterial community structures. These convergences
632 between the *tpm* and 16S rRNA gene community profilings confirmed the reliability of the *tpm*
633 metabarcoding approach.

634 Analyses of the relations between the occurrences of *tpm*-related bacterial taxa and chemical
635 pollutants were completed by investigating correlations with variations in [chlorpyrifos],
636 [octylphenols] and [bisphenol A]. Chlorpyrifos (an insecticide) was previously shown to
637 negatively impact soil microbial activities including those of the sulfur and nitrogen cycles
638 (Karas et al., 2018). Highly significant CCA linked [chlorpyrifos] in the sediments, their location
639 in the DB and maturation/accumulation periods, and *tpm*-community profiles. Chlorpyrifos was
640 in higher concentrations in the surface deposits. Regarding octylphenols, these appeared to share
641 some structural similarities with 6-mercaptopurines (6-MP) and related drugs which are
642 transformed by human TPMTs (e. g. Cournoyer et al., 1998). Among these, alkylphenol
643 ethoxylates (NPE) are nonionic surfactants commonly used in the formulation of chemical
644 products such as ink, herbicides, and also used as antioxidants to protect polymers like rubber
645 and plastics (Wiest et al., 2018). Nonylphenols (NP), a degradation by-product of NPE, are of
646 very high concern because of their endocrine disruptive properties (ECHA, 2018). This is also
647 the case of bisphenol A which was also monitored in these investigations. Both pollutants were
648 shown to induce the decay of bacteria involved in the sulfur cycle (Van Ginkel et al., 2010).
649 Here, NPs were found in significantly higher concentrations among the oldest sediments of the

650 decantation pit. This was indicative of significant NPE biological degradation in the pit.
651 Interestingly, strong significant relationships by CCA were detected between the [NPs],
652 accumulation time periods and position of the sediments in the DB, and the *tpm* community
653 profiles (with the OTU and ASV datasets). Such significant relationships were not observed with
654 [bisphenol A]. These results suggested that the bacterial community of the pit and its *tpm*-
655 harboring taxa can contribute at the degradation of NPEs. Further analyses will be required to
656 demonstrate a role of the *tpm* or associated genes in this degradation process but a *tpm*-harboring
657 strain of *P. fluorescens* was found as one of the best degrader of NPEs (Ruiz et al., 2013). It is to
658 be noted that the Predicted No Effect Concentration (PNEC) for NPs in sediments (according to
659 the European Union risk assessment report) was estimated at 39 ng/g. The sediments analysed in
660 our study had >1000 ng NP/g dry weight sediment, and were thus considered to be toxic for
661 aquatic organisms. Here, high [NPs] were associated with a significant re-shuffling of *tpm*-
662 harboring bacteria including changes in their relative count numbers. These taxa or *tpm*
663 sequences could thus be used as a toolbox for the development of biological indicators of the
664 hazards associated with multiple contaminations by high risk priority chemical pollutants. To go
665 further with such developments, taxonomic allocations of the *tpm* bacterial taxa selected by the
666 living conditions building-up over time among urban deposits/sediments were carried out.

667 The *tpm* OTUs were allocated to eighteen genera and twenty-six genus-like lineages, and the
668 ASV ones were allocated to sixteen genera and thirty genus-like lineages. These reads could then
669 be respectively allocated to 119 and 110 species. The undefined genus-like lineages are
670 indicative of a significant microbial dark matter among urban biomes as reported for other
671 ecological systems (e. g. Dance (2020) for a review). Reads allocated to well-defined genera
672 such as *Aeromonas*, *Stenotrophomonas*, *Pseudomonas*, *Xanthomonas*, *Aquaspirillum*,
673 *Herbaspirillum* (particularly *H. aquaticum*), and *Nitrosomonas* (particularly *N. oligotropha*)
674 were found in significantly higher relative numbers among the recent deposits (showing higher

675 chlorpyrifos, chrysene (pit), NPEs) of the detention basin than the old ones (showing higher NP,
676 MTE, anthracene, fluorene, naphthalene). The toxic pollutants found in the aged deposits /
677 sediments were related to a significant relative drop in these taxa. In fact, a major decline in the
678 *Aeromonas tpm*-harboring cells was observed in the older sediments. More than twelve species
679 of *Aeromonas* were recorded in the recent deposits, with the most significant ones being *A.*
680 *hydrophila*, *A. media*, *A. caviae*, *A. salmonicida*, and *A. eucrenophila*. These species were all
681 shown to decline over the sediment aging process including those representing human health
682 hazards. Some of these species (*A. media*, *A. rivuli*, *A. salmonicida*, and two undefined
683 *Aeromonas* species) were also found to be enriched by the pit physico-chemical conditions. The
684 pit sediments were found to be enriched in fluorene which can generate toxic degradation by-
685 products such as methylated fluorenes (Rice et al. 1987), and by alkylphenol ethoxylates which
686 can lead to NPs. These bacterial species could thus be tolerant or resistant to these by-products.
687 The harsher living conditions of the old sediments were also associated with the discovery of
688 several undefined *tpm*-harboring species among the *Aeromonas* but also other genera.
689 Regarding the more than 30 species of *Pseudomonas* detected among the *tpm* DB
690 metabarcoding dataset, distribution biases indicated a significant enrichment of *P. syringae tpm*
691 reads among the old surface deposits (showing higher naphthalene, and Cd, Cr, Ni). Further
692 analysis of these *tpm* reads by molecular phylogeny showed that they were part of (i) *P. s.* pv.
693 *aceris*, (ii) *P. s.* pv. *pisi*, and (iii) *P. s.* pv. *lachrymans/tomato*. This higher relative number of *P.*
694 *syringae* in the old deposits could be related to their known ability at using naphthalene as a
695 carbon-source (Wald et al., 2015). The above pathovars could thus likely be used in the
696 development of biological sentinels for detecting multiple chemical contaminations in
697 environmental samples. These bacteria harbor distinct *tpm* DNA signatures that could be tracked
698 by PCR assays (e. g. Favre-Bonte et al., 2005). It is to be noted that the highest numbers of
699 *Pseudomonas tpm* reads were allocated to *P. anguilliseptica / P. anguilliseptica-like species*, *P.*

koreensis, *P. extremaustralis*, and two novel lineages of undefined species with one being affiliated to PGPPP3-like genomes. These latter genomes were found among the bacterial metagenome of the star freshwater diatom named *Asterionella formosa* (GenBank accession number NKIE00000000). PGPPP3-like *tpm* reads were in higher relative numbers in the recent deposits. The association with diatoms could be an important factor explaining the distribution patterns of these bacteria. *P. anguilliseptica* is an emerging opportunistic fish pathogen (Fadel et al., 2018), and was recorded among deep aquifers (Favre-Bonte et al., 2005). A tropism for wet systems thus likely explains its detection in the most recent sediments of the DB, and the gradual decline of its population among older sediments. *P. koreensis* was reported among Korean agricultural soil samples, and not much is known about its ecology (Kwon et al., 2003). However, *P. extremaustralis* was isolated from a temporary pond in Antarctica, and was found resistant to high heat and chemical pollutants (Tribelli et al., 2012); two properties that can explain their survival among urban deposits. *P. extremaustralis* and other known degraders of urban pollutants identified in the *tpm* dataset such as *P. aeruginosa* and *P. putida* did not show significant variations over the sampling periods. This suggests a tolerance towards most chemical contaminants observed in this study. *P. aeruginosa tpm* reads were further classified into sub-clades, and the reads were found to belong to the PAO1 and PA14 lineages. *P. aeruginosa* was previously found not to be well-adapted for growth among natural soils (Nazaret et al., 2014). Here, urban sediments appeared a source of these bacteria. *P. aeruginosa* was the most hazardous pathogenic species detected among these sediments. The detected *tpm* sequence types were attributed to isolates from the following MLST sequence types (https://pubmlst.org/): ST146 which was mainly attributed to isolates from cystic fibrosis (CF) patients in the UK, ST244 grouping isolates from non-CF patients with urinary tract, soft tissues, and blood infections (worldwide distribution), ST267 only reported in Australia so far, ST693 (no epidemiological data available), ST252 from non-CF infections (reported in Spain and Senegal),

725 ST253 (found in CF lung infections and non-CF urinary tract, soft tissues, and blood infections;
726 reported in Europe, Senegal, Peru), ST308 (from CF lung infections and non-CF urinary tract,
727 soft tissues, and blood infections; reported in Australia, Africa, and Europe), and ST319
728 (reported in Spain and Senegal).

729 It is to be noted that *tpm* reads allocated to *P. koreensis*, *P. rhodesia*, *P. umsongensis*, *P.*
730 *reinekei*, *P. nitroreducens*, *P. mendocina*, *P. hunanensis*, *P. fluvialis*, and *P. azotoformans* were
731 found in higher numbers among the pit sediments than the surface ones. As indicated above, the
732 particular growing conditions prevailing in the settling pit including the occurrence of particular
733 pollutants can explain these differences. Still, the old pit deposits were found to be enriched in
734 *tpm* reads from mainly undefined species that appeared better adapted to heavily polluted
735 environments. Their *tpm* reads can be easily differentiated from those of the other species, and
736 could be used for the design of PCR assays for their tracking among outdoor systems impacted
737 by high priority toxic pollutants.

738

739 **5. Conclusions**

740 The *tpm* DNA metabarcoding approach developed in this study was used to evaluate the
741 incidence of combined high risk chemical pollutants on urban microbiomes. The *tpm*-encoding
742 bacteria were found distributed among at least 8 phyla, and to be systematically recovered from
743 urban deposits and sediments. The *tpm* bacterial community of urban deposits was found to
744 change over time and space according to their evolving content in high risk pollutants. Some
745 *tpm*-harboring bacterial taxa were found associated with a degradation of these pollutants such as
746 NPE. A significant finding was the decline in *tpm*-harboring Aeromonads over time among
747 sediments that were heavily contaminated by a mixture of NPs, MTEs and some PAHs like
748 naphthalene. This decline was clearly indicative of harsher living conditions building up in the
749 older urban sediments leading to a decay of all the well-known *tpm*-harboring *Aeromonas*

750 species. However, some species were favored by these living conditions such as *P. syringae*.
751 This makes these bacterial groups and species potential ecological sentinels of environmental
752 health hazards. The harsher living conditions of the old polluted sediments also led to the
753 discovery of a significant microbial dark matter among urban biomes. Old sediments
754 accumulating pollutants for over two years appeared a source of *tpm*-bacterial taxa that can be
755 used as biological indicators for assessing the ecological threat of multiple chemical
756 contaminations.
757

758 **Credit authorship contribution statement**

759 **Axel Aigle:** conceptualization, data curation, formal analysis. **Yannick Colin:**
760 conceptualization, data curation, formal analysis**. Rayan Bouchali:** conceptualization, data
761 curation, formal analysis. **Emilie Bourgeois:** conceptualization, data curation, formal analysis.
762 **Romain Marti**: conceptualization, data curation, formal analysis. **Sébastien Ribun:**
763 conceptualization, data curation, formal analysis. **Laurence Marjolet:** conceptualization, data
764 curation, formal analysis, **Adrien C. M. Pozzi:** data curation, writing - review & editing. **Boris**
765 **Misery:** data curation, formal analysis**, Céline Colinon:** data curation, formal analysis, **Claire**
766 **Bernardin-Souibgui:** conceptualization, data curation. **Laure Wiest:** conceptualization, data
767 curation. **Didier Blaha:** conceptualization, data curation. **Wessam Galia:** data curation, writing -
768 review & editing. **Benoit Cournoyer:** conceptualization, data curation, formal analysis, funding
769 acquisition, project administration and supervision, writing - original draft, writing - review &
770 editing.

771 **Declaration of Competing Interest**

772 The authors declare that they have no known competing financial interests or personal
773 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 Desir and DOmic projects.

Acknowledgements

Authors thank the OTHU network for technical assistance and financial supports, the School of Integrated Watershed Sciences H2O'LYON (ANR 17-EURE-0018), and the Urban School of Lyon (ANR-17-CONV-0004) for their advices and supports in the elaboration of this multi-disciplinary research initiative.

Appendix A. Supplementary materials

The Supplementary Materials for this article can be found online at: https://.....

References

Arbizu, P.M., 2017. pairwiseAdonis: Pairwise Multilevel Comparison using Adonis. R package version 0.0.1.

Badin, A.-L., Faure, P., Bedell, J.-P., Delolme, C., 2008. Distribution of organic pollutants and natural organic matter in urban storm water sediments as a function of grain size. Sci. Total Environ. 403, 178–187. https://doi.org/10.1016/j.scitotenv.2008.05.022

Balandreau, J., Viallard, V., Cournoyer, B., Coenye, T., Laevens, S., Vandamme, P., 2001. B*urkholderia cepacia* genomovar III is a common plant-associated bacterium. Appl. Environ. Microbiol. 67, 982–985. https://doi.org/10.1128/AEM.67.2.982-985.2001

Becouze-Lareure, C., Lipeme Kouyi, G., Gonzalez-Merchan, C., Bazin, C., Sebastian, C., Barraud, S., Perrodin, Y., 2018. Spatial and temporal dynamics of sediment ecotoxicity in urban stormwater retention basins: Methodological approach and application to a pilot site close to Lyon in France. J. Environ. Sci. Health Part A 53, 1123–1130. https://doi.org/10.1080/10934529.2018.1529894

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

Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172, 2665–2681. https://doi.org/10.1534/genetics.105.048975

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

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

Cournoyer, B., Orvis, J. D., Astuto, A., Gibbon, M. J., Taylor, J., Vivian, A., 1995. Molecular Characterization of the *Pseudomonas syringae* pv. *pisi* plasmid-borne avirulence gene *avrPpiB* which matches the *R3* resistance locus in pea. Mol. Plant. Microbe Interact. 8, 700. https://doi.org/10.1094/MPMI-8-0700

822 Cournoyer, B., Watanabe, S., Vivian, A., 1998. A tellurite-resistance genetic determinant from
823 phytopathogenic pseudomonads encodes a thiopurine methyltransferase: evidence of a
824 widely-conserved family of methyltransferases. Biochim. Biophys. Acta BBA - Gene Struct.
825 Expr. 1397, 161–168. https://doi.org/10.1016/S0167-4781(98)00020-7
826 Coussement, J., et al. for the European Study Group for Nocardia in Solid Organ
827 Transplantation, 2016. *Nocardia* infection in solid organ transplant recipients: a multicenter
828 european case-control study. Clin. Infect. Dis. 63, 338–345.
829 https://doi.org/10.1093/cid/ciw241
830 Crawford, J.W., Deacon, L., Grinev, D., Harris, J.A., Ritz, K., Singh, B.K., Young, I., 2012.
831 Microbial diversity affects self-organization of the soil–microbe system with consequences
832 for function. J. R. Soc. Interface 9, 1302–1310. https://doi.org/10.1098/rsif.2011.0679
833 Dance, A., 2020. The search for microbial dark matter. Nature 582, 301–303.
834 https://doi.org/10.1038/d41586-020-01684-z
835 Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., Callahan, B.J., 2018. Simple statistical
836 identification and removal of contaminant sequences in marker-gene and metagenomics data.
837 Microbiome 6, 226. https://doi.org/10.1186/s40168-018-0605-2
838 Deredjian, A., Colinon, C., Brothier, E., Favre-Bonté, S., Cournoyer, B., Nazaret, S., 2011.
839 Antibiotic and metal resistance among hospital and outdoor strains of *Pseudomonas*
840 *aeruginosa*. Res. Microbiol. 162, 689–700. https://doi.org/10.1016/j.resmic.2011.06.007
841 Fadel, A., Mabrok, M., Aly, S., 2018. Epizootics of *Pseudomonas anguilliseptica* among
842 cultured seabream (*Sparus aurata*) populations: Control and treatment strategies. Microb.
843 Pathog. 121, 1–8. https://doi.org/10.1016/j.micpath.2018.04.021
844 Favre-Bonté, S., Ranjard, L., Champier, L., Cournoyer, B., Nazaret, S., 2006. Distribution and
845 genetic diversity of bacterial thiopurine methyltransferases in soils emitting dimethyl selenide.
846 Biochimie 88, 1573–1581. https://doi.org/10.1016/j.biochi.2006.09.005

847 Favre-Bonte, S., Ranjard, L., Colinon, C., Prigent-Combaret, C., Nazaret, S., Cournoyer, B.,
848 2005. Freshwater selenium-methylating bacterial thiopurine methyltransferases: diversity and
849 molecular phylogeny. Environ. Microbiol. 7, 153–164. https://doi.org/10.1111/j.1462-
850 2920.2004.00670.x
851 Ford, L.T., Berg, J.D., 2010. Thiopurine S-methyltransferase (TPMT) assessment prior to
852 starting thiopurine drug treatment; a pharmacogenomic test whose time has come. J. Clin.
853 Pathol. 63, 288–295. https://doi.org/10.1136/jcp.2009.069252
854 Gonzalez-Merchan, C., Barraud, S., Bedell, J.-P., 2014. Influence of spontaneous vegetation in
855 stormwater infiltration system clogging. Environ. Sci. Pollut. Res. 21, 5419–5426.
856 https://doi.org/10.1007/s11356-013-2398-y
857 Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X., Briggs, J.M., 2008.
858 Global Change and the Ecology of Cities. Science 319, 756–760.
859 https://doi.org/10.1126/science.1150195
860 Gy, P., 2004. Sampling of discrete materials—a new introduction to the theory of sampling.
861 Chemom. Intell. Lab. Syst. 74, 7–24. https://doi.org/10.1016/j.chemolab.2004.05.012
862 Higgins, D.G., Sharp, P.M., 1988. CLUSTAL: a package for performing multiple sequence
863 alignment on a microcomputer. Gene 73, 237–244. https://doi.org/10.1016/0378-
864 1119(88)90330-7
865 Huson, D.H., Bryant, D., 2006. Application of Phylogenetic Networks in Evolutionary Studies.
866 Mol. Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030
867 Karas, P.A., Baguelin, C., Pertile, G., Papadopoulou, E.S., Nikolaki, S., Storck, V., Ferrari, F.,
868 Trevisan, M., Ferrarini, A., Fornasier, F., Vasileiadis, S., Tsiamis, G., Martin-Laurent, F.,
869 Karpouzas, D.G., 2018. Assessment of the impact of three pesticides on microbial dynamics
870 and functions in a lab-to-field experimental approach. Sci. Total Environ. 637–638, 636–646.
871 https://doi.org/10.1016/j.scitotenv.2018.05.073

872 Katara, P., Kuntal, H., 2016. TPMT polymorphism: when shield becomes weakness. Interdiscip.
873 Sci. Comput. Life Sci. 8, 150–155. https://doi.org/10.1007/s12539-015-0111-1
874 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of
875 a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data
876 on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120.
877 https://doi.org/10.1128/AEM.01043-13
878 Kwon, S.W., Kim, J.S., Park, I.C., Yoon, S.H., Park, D.H., Lim, C.K., Go, S.J., 2003.
879 *Pseudomonas koreensis* sp. nov., *Pseudomonas umsongensis* sp. nov. and *Pseudomonas*
880 *jinjuensis* sp. nov., novel species from farm soils in Korea. Int. J. Syst. Evol. Microbiol. 53,
881 21–27. https://doi.org/10.1099/ijs.0.02326-0
882 Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for
883 RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-
884 0550-8
885 Marti, R., Bécouze-Lareure, C., Ribun, S., Marjolet, L., Bernardin Souibgui, C., Aubin, J.-B.,
886 Lipeme Kouyi, G., Wiest, L., Blaha, D., Cournoyer, B., 2017a. Bacteriome genetic structures
887 of urban deposits are indicative of their origin and impacted by chemical pollutants. Sci. Rep.
888 7.
889 Marti, R., Ribun, S., Aubin, J.-B., Colinon, C., Petit, S., Marjolet, L., Gourmelon, M., Schmitt,
890 L., Breil, P., Cottet, M., Cournoyer, B., 2017b. Human-driven microbiological contamination
891 of benthic and hyporheic sediments of an intermittent peri-urban river assessed from MST and
892 16S rRNA genetic structure analyses. Front. Microbiol. 8.
893 https://doi.org/10.3389/fmicb.2017.00019
894 McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis
895 and graphics of microbiome census data. PLoS ONE 8, e61217.
896 https://doi.org/10.1371/journal.pone.0061217

Mikhailyuk, T., Vinogradova, O., Holzinger, A., Glaser, K., Samolov, E., Karsten, U., 2019. New record of the rare genus *Crinalium* Crow (Oscillatoriales, Cyanobacteria) from sand dunes of the Baltic Sea, Germany: epitypification and emendation of *Crinalium magnum* Fritsch et John based on an integrative approach. Phytotaxa 400, 165. https://doi.org/10.11646/phytotaxa.400.3.4

Nazaret, S., Deredjian, A., Colinon-Dupuech, C., Ranjard, L., Dequiedt, S., Hartmann, A., Brothier, E., Youenou, B., Cournoyer, B., Hien, E., Saby, N.P.A., Houot, S., Jolivet, C., 2014. Low occurrence of *Pseudomonas aeruginosa* in agricultural soils with and without organic amendment. Frontiers Cellular Infect. Microbiol. 4:53. https://doi.org/10.3389/fcimb.2014.00053

Olson, N. D., Kumar, M. S., Li, S., Braccia, D. J., Hao, S., Timp, W., Salit, M. L., Stine, O. C., Bravo, H. C., 2020. A framework for assessing 16S rRNA marker-gene survey data analysis methods using mixtures. Microbiome 8, 35. https://doi.org/10/ghtr9w

Prigent-Combaret, C., Sanguin, H., Champier, L., Bertrand, C., Monnez, C., Colinon, C., Blaha, D., Ghigo, J.-M., Cournoyer, B., 2012. The bacterial thiopurine methyltransferase tellurite resistance process is highly dependent upon aggregation properties and oxidative stress response. Environ. Microbiol. 14, 2645–2660. https://doi.org/10.1111/j.1462-2920.2012.02802.x

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., Levin, E., 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One, 15, e0227434. https://doi.org/10/ggtjbr

Reese, A.T., Savage, A., Youngsteadt, E., McGuire, K.L., Koling, A., Watkins, O., Frank, S.D., Dunn, R.R., 2016. Urban stress is associated with variation in microbial species composition—but not richness—in Manhattan. ISME J. 10, 751–760. https://doi.org/10.1038/ismej.2015.152

922 Rice, J.E., Rivenson, A., Braley, J., LaVoie, E.J., 1987. Methylated derivatives of pyrene and
923 fluorene: Evaluation of genotoxicity in the hepatocyte/DNA repair test and tumorigenic
924 activity in newborn mice. J. Toxicol. Environ. Health 21, 525–532.
925 https://doi.org/10.1080/15287398709531040
926 Rojo, F., Martínez, J.L., 2019. Hydrocarbon degraders as pathogens, in: Goldfine, H. (Ed.),
927 Health consequences of microbial interactions with hydrocarbons, oils, and lipids. Springer
928 International Publishing, Cham, pp. 1–15. https://doi.org/10.1007/978-3-319-72473-7_22-1
929 Ruiz, Y., Medina, L., Borusiak, M., Ramos, N., Pinto, G., Valbuena, O., 2013. Biodegradation of
930 polyethoxylated nonylphenols. ISRN Microbiol. 2013, 1–9.
931 https://doi.org/10.1155/2013/284950
932 Schaad, N.W., Postnikova, E., Lacy, G., Sechler, A., Agarkova, I., Stromberg, P.E., Stromberg,
933 V.K., Vidaver, A.K., 2006. Emended classification of xanthomonad pathogens on citrus. Syst.
934 Appl. Microbiol. 29, 690–695. https://doi.org/10.1016/j.syapm.2006.08.001
935 Schaad, N.W., Postnikova, E., Lacy, G.H., Sechler, A., Agarkova, I., Stromberg, P.E.,
936 Stromberg, V.K., Vidaver, A.K., 2005. Reclassification of *Xanthomonas campestris* pv. *citri*
937 (ex Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (ex Hasse) sp.
938 nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifolii* (ex Gabriel 1989) sp. nov. nom. rev.
939 comb. nov., and *X. alfalfae* subsp. *citrumelo* (ex Riker and Jones) Gabriel et al., 1989 sp. nov.
940 nom. rev. comb. nov.; *X. campestris* pv *malvacearum* (ex Smith 1901) Dye 1978 as *X. smithii*
941 subsp. *smithii* nov. comb. nov. nom. nov.; *X. campestris* pv. *alfalfae* (ex Riker and Jones,
942 1935) Dye 1978 as *X. alfalfae* subsp. *alfalfae* (ex Riker et al., 1935) sp. nov. nom. rev.; and
943 "var. *fuscans*" of *X. campestris* pv. *phaseoli* (ex Smith, 1987) Dye 1978 as *X. fuscans* subsp.
944 *fuscans* sp. nov. Syst. Appl. Microbiol. 28, 494–518.
945 https://doi.org/10.1016/j.syapm.2005.03.017

946 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski,
947 R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van
948 Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-Source, platform-independent,
949 community-supported software for describing and comparing microbial communities. Appl.
950 Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09
951 Sébastian, C., Barraud, S., Ribun, S., Zoropogui, A., Blaha, D., Becouze-Lareure, C., Kouyi,
952 G.L., Cournoyer, B., 2014. Accumulated sediments in a detention basin: chemical and
953 microbial hazard assessment linked to hydrological processes. Environ. Sci. Pollut. Res. 21,
954 5367–5378. https://doi.org/10.1007/s11356-013-2397-z
955 Severiano, A., Pinto, F.R., Ramirez, M., Carriço, J.A., 2011. Adjusted Wallace coefficient as a
956 measure of congruence between typing methods. J. Clin. Microbiol. 49, 3997–4000.
957 https://doi.org/10.1128/JCM.00624-11
958 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular
959 evolutionary genetics analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729.
960 https://doi.org/10.1093/molbev/mst197
961 Tribelli, P.M., Raiger Iustman, L.J., Catone, M.V., Di Martino, C., Revale, S., Mendez, B.S.,
962 Lopez, N.I., 2012. Genome sequence of the polyhydroxybutyrate producer *Pseudomonas*
963 *extremaustralis*, a highly stress-resistant antarctic bacterium. J. Bacteriol. 194, 2381–2382.
964 https://doi.org/10.1128/JB.00172-12
965 Van Ginkel, S.W., Hassan, S.H.A., Oh, S.-E., 2010. Detecting endocrine disrupting compounds
966 in water using sulfur-oxidizing bacteria. Chemosphere 81, 294–297.
967 https://doi.org/10.1016/j.chemosphere.2010.05.056
968 Voisin, J., Cournoyer, B., Marjolet, L., Vienney, A., Mermillod-Blondin, F., 2019. Ecological
969 assessment of groundwater ecosystems disturbed by recharge systems using organic matter

970 quality, biofilm characteristics, and bacterial diversity. Environ. Sci. Pollut. Res.
971 https://doi.org/10.1007/s11356-019-06971-5
972 Voisin, J., Cournoyer, B., Vienney, A., Mermillod-Blondin, F., 2018. Aquifer recharge with
973 stormwater runoff in urban areas: Influence of vadose zone thickness on nutrient and bacterial
974 transfers from the surface of infiltration basins to groundwater. Sci. Total Environ. 637–638,
975 1496–1507. https://doi.org/10.1016/j.scitotenv.2018.05.094
976 Wald, J., Hroudova, M., Jansa, J., Vrchotova, B., Macek, T., Uhlik, O., 2015. Pseudomonads
977 rule degradation of polyaromatic hydrocarbons in aerated sediment. Front. Microbiol. 6.
978 https://doi.org/10.3389/fmicb.2015.01268
979 Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid
980 assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol.
981 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07
982 Wiest, L., Baudot, R., Lafay, F., Bonjour, E., Becouze-Lareure, C., Aubin, J.-B., Jame, P.,
983 Barraud, S., Kouyi, G.L., Sébastian, C., Vulliet, E., 2018. Priority substances in accumulated
984 sediments in a stormwater detention basin from an industrial area. Environ. Pollut. 243, 1669–
985 1678. https://doi.org/10.1016/j.envpol.2018.09.138
986 Xu, H.-J., Li, S., Su, J.-Q., Nie, S., Gibson, V., Li, H., Zhu, Y.-G., 2014. Does urbanization
987 shape bacterial community composition in urban park soils? A case study in 16 representative
988 Chinese cities based on the pyrosequencing method. FEMS Microbiol. Ecol. 87, 182–192.
989 https://doi.org/10.1111/1574-6941.12215
990 Yan, B., Li, J., Xiao, N., Qi, Y., Fu, G., Liu, G., Qiao, M., 2016. Urban-development-induced
991 Changes in the Diversity and Composition of the Soil Bacterial Community in Beijing. Sci.
992 Rep. 6, 38811. https://doi.org/10.1038/srep38811
993 Yan, H., Lipeme Kouyi, G., Gonzalez-Merchan, C., Becouze-Lareure, C., Sebastian, C.,
994 Barraud, S., Bertrand-Krajewski, J.-L., 2014. Computational fluid dynamics modelling of

995 flow and particulate contaminants sedimentation in an urban stormwater detention and settling
996 basin. Environ. Sci. Pollut. Res. 21, 5347–5356. https://doi.org/10.1007/s11356-013-2455-6
997
998

Figure captions

Fig. 1. NMDS of Bray-Curtis dissimilarities computed from *tpm* ASV profiles of all deposits / sediment samples (but without P0) investigated in this study according to (a) their accumulation and maturation time periods (old - green or recent - red), and (b) position in the detention basin (surface of the detention basin - blue, and settling pit - orange). See **Table S1** for a description of the samples. Adonis tests confirmed the significance of the differentiations between the colored groups ($p<0.001$) but also showed a lack of significant differentiation within each group. See **Fig. S13** for the NMDS computed for the *tpm* OTU profiles.

Fig. 2. CCA plots computed from *tpm* (a) OTU and (b) ASV profiles constrained by the accumulation and maturation time periods and the position of the urban sediments as well as by a set of high priority chemical pollutants (see **Table 1** and suppl. files for other significant relationships). The *tpm* OTU and ASV profiles were significantly related to [4-nonylphenol], [chlorpyrifos], and [chrysene]. Eigenvalues are indicated in **Table S14**.

Fig. 3. Illustration of the diversity of *tpm* sequences and of their respective taxonomic allocations obtained by the metabarcoding approach presented in this work and applied on DNA extracts from urban deposits and sediments. OTUs and ASVs were used to build this figure. Phylogenetic lineages and groupings match those inferred from a maximum-likelihood (ML) phylogenetic analysis. Only the most significant bootstraps are shown. Tree branches were collapsed for divergences within genera. A total of 194 DNA sites were considered to build this cladogram, and 694 bootstrap replicates were performed. Only plain lines match the bar scale. DNA sequences were retrieved from **Tables S5** and **S6**.

Fig. 4. Relative abundance of the bacterial genera inferred from the *tpm* OTU and ASV contingency tables. These genera showed significant changes in abundance according to the maturation/accumulation time periods (recent vs old) or the position of the urban

deposits/sediments in the detention basin (surface vs pit). These changes were estimated by DESEQ2 using **Tables S5c** and **S6c**. Names of higher taxonomic levels than genera are highlighted by the use of slashes. Overlaps between bubbles have no biological significance. NA: not attributed to defined taxa.

Fig. 5. Relative abundance of the bacterial species inferred from the *tpm* OTU and ASV contingency tables. These species showed significant changes in abundance according to the maturation/accumulation time periods (recent vs old) or the position of the urban deposits/sediments in the detention basin (surface vs pit). These changes were estimated by DESEQ2 using **Tables S5c** and **S6c**. Names of higher taxonomic levels than genera are highlighted by the use of slashes. *Pseudomonas* sp. corresponds to the PGPPP3-like *tpm* sequences matching OTU ^{sp}002 and ASV^{sp}002. Overlaps between bubbles have no biological significance. NA: not attributed to defined taxa.

Fig. 6. Bar plots showing changes in the relative abundance of the species inferred from the *tpm* OTUs and ASVs belonging to the *Aeromonas* or *Pseudomonas* genus according to the maturation state (old (o) or recent (r)) or position (surface (s) or pit (p)) of the urban deposits in the detention basin. Kruskal-Wallis tests were performed to detect significant changes; a letter code on a bar indicates a significant difference but also the group with the highest values. See **Tables S5d, e** and **S6d, e** for more details. Two *tpm* species-like entities are not shown, *P. sp.* (PGPPP3-like) and *P. unclassified,* for simplification of the figure but their distribution patterns are illustrated in **Fig. 5**.

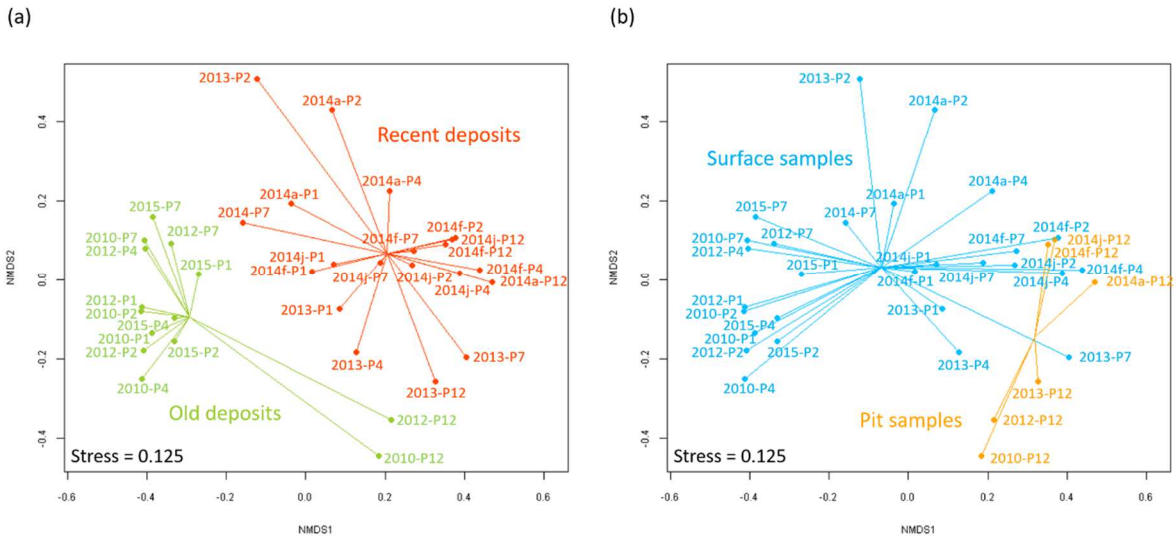

Figure 1. Aigle et al.

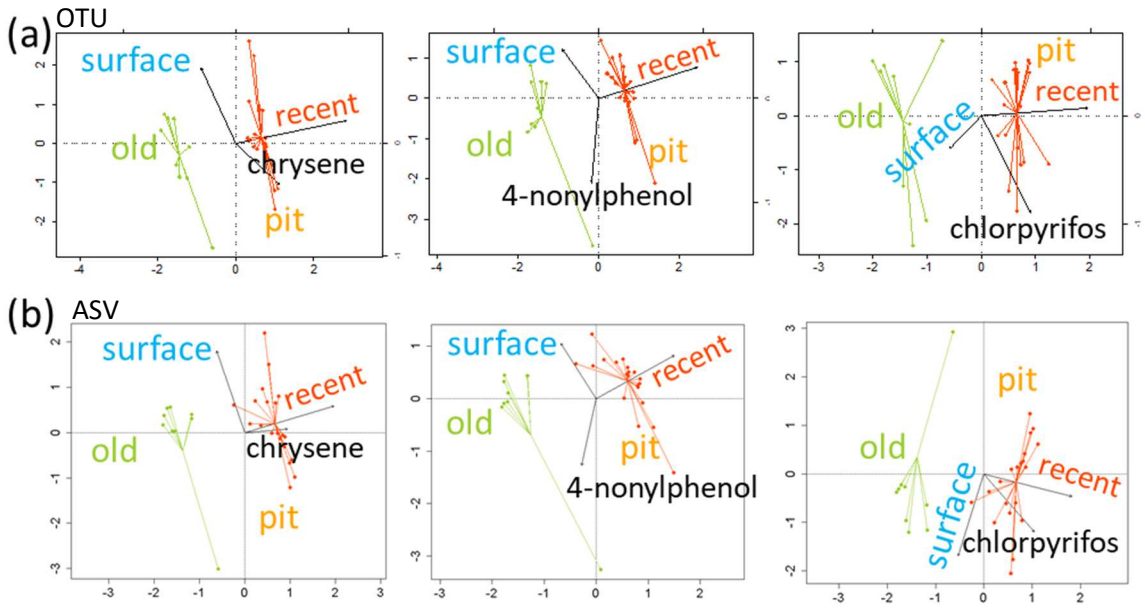

Figure 2 Aigle et al.

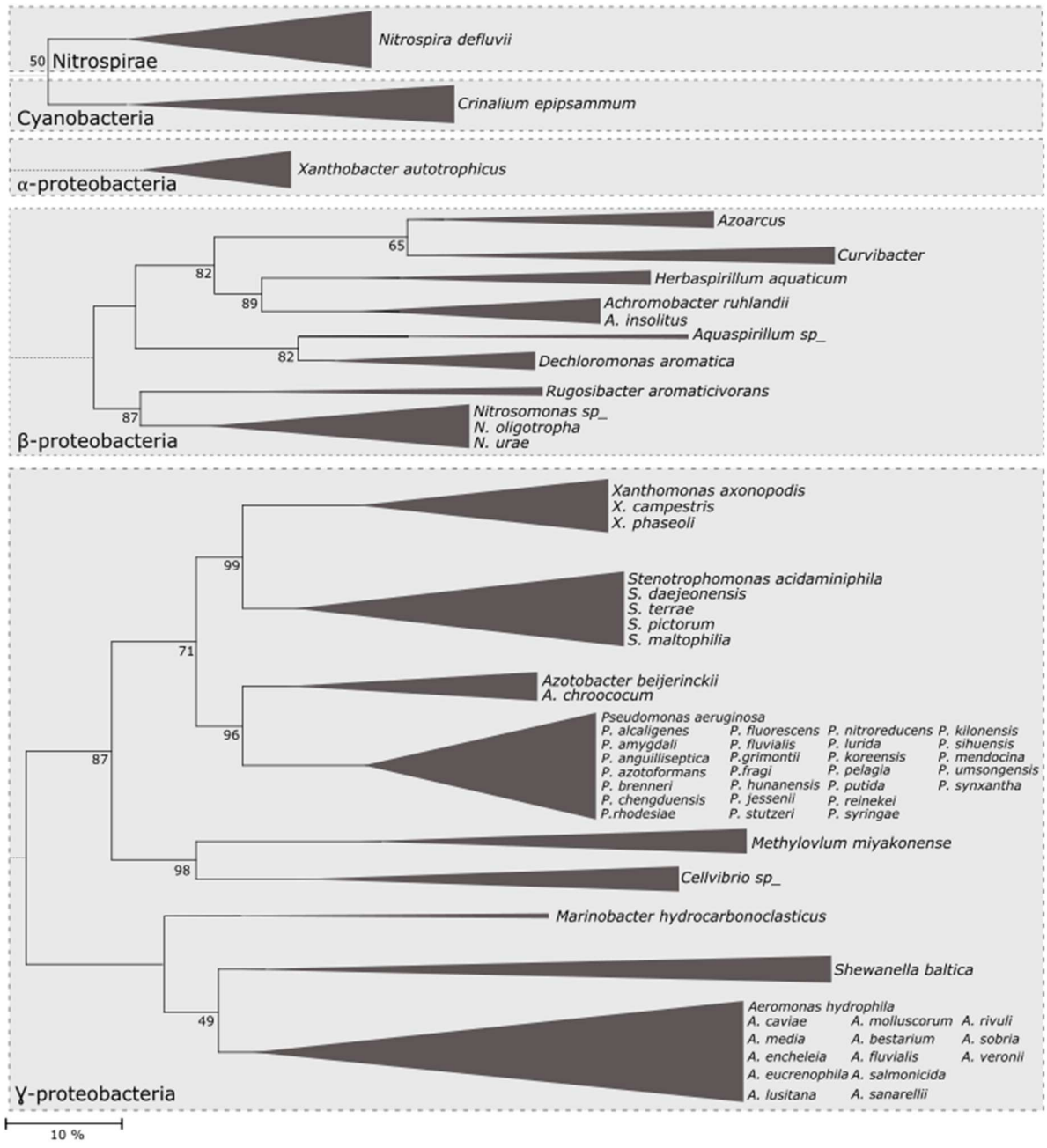

Figure 3 Aigle et al.

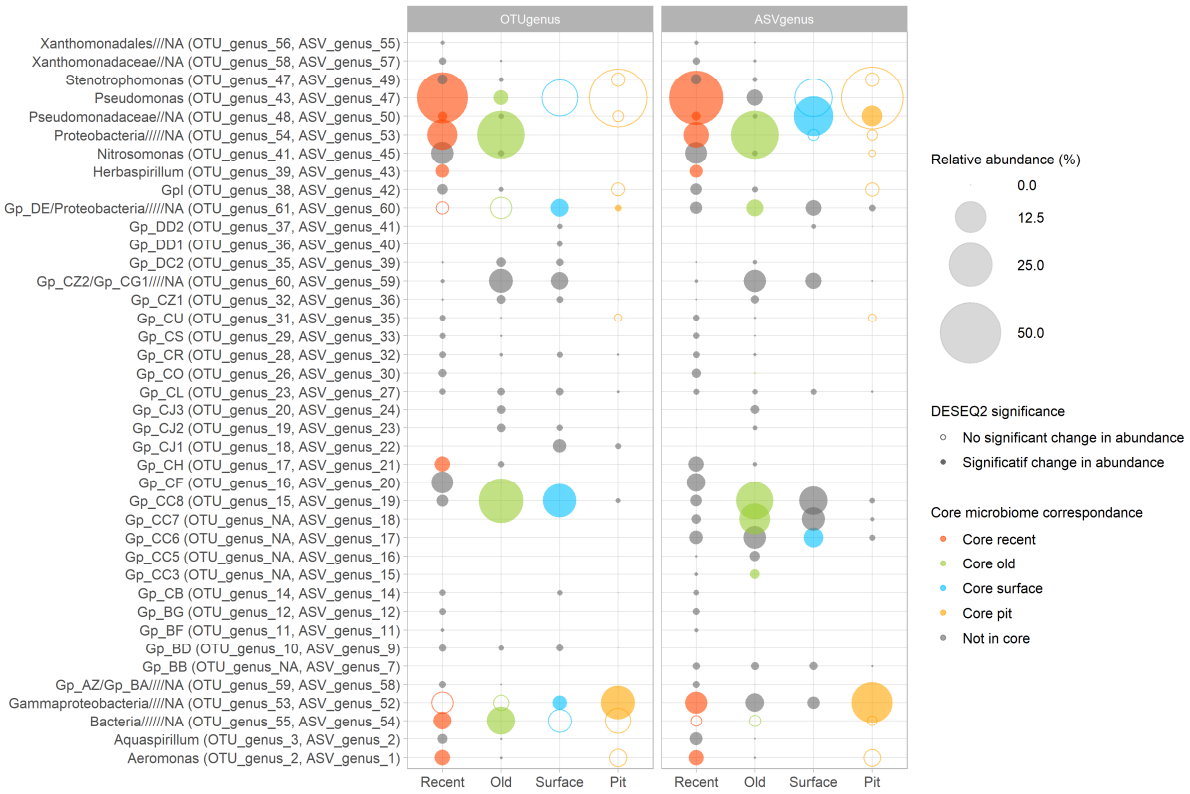

Figure 4 Aigle et al.

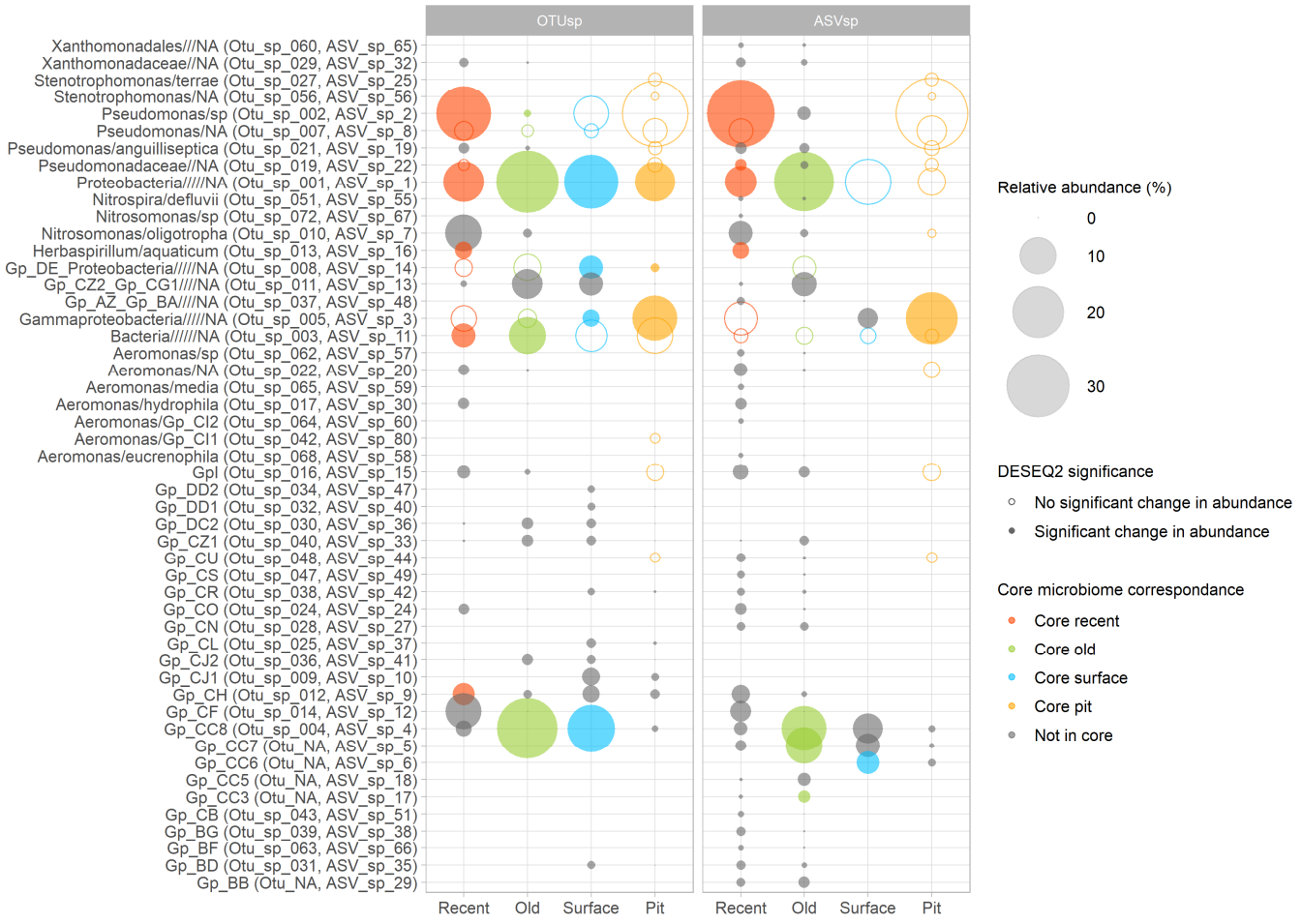

Figure 5 Aigle et al.

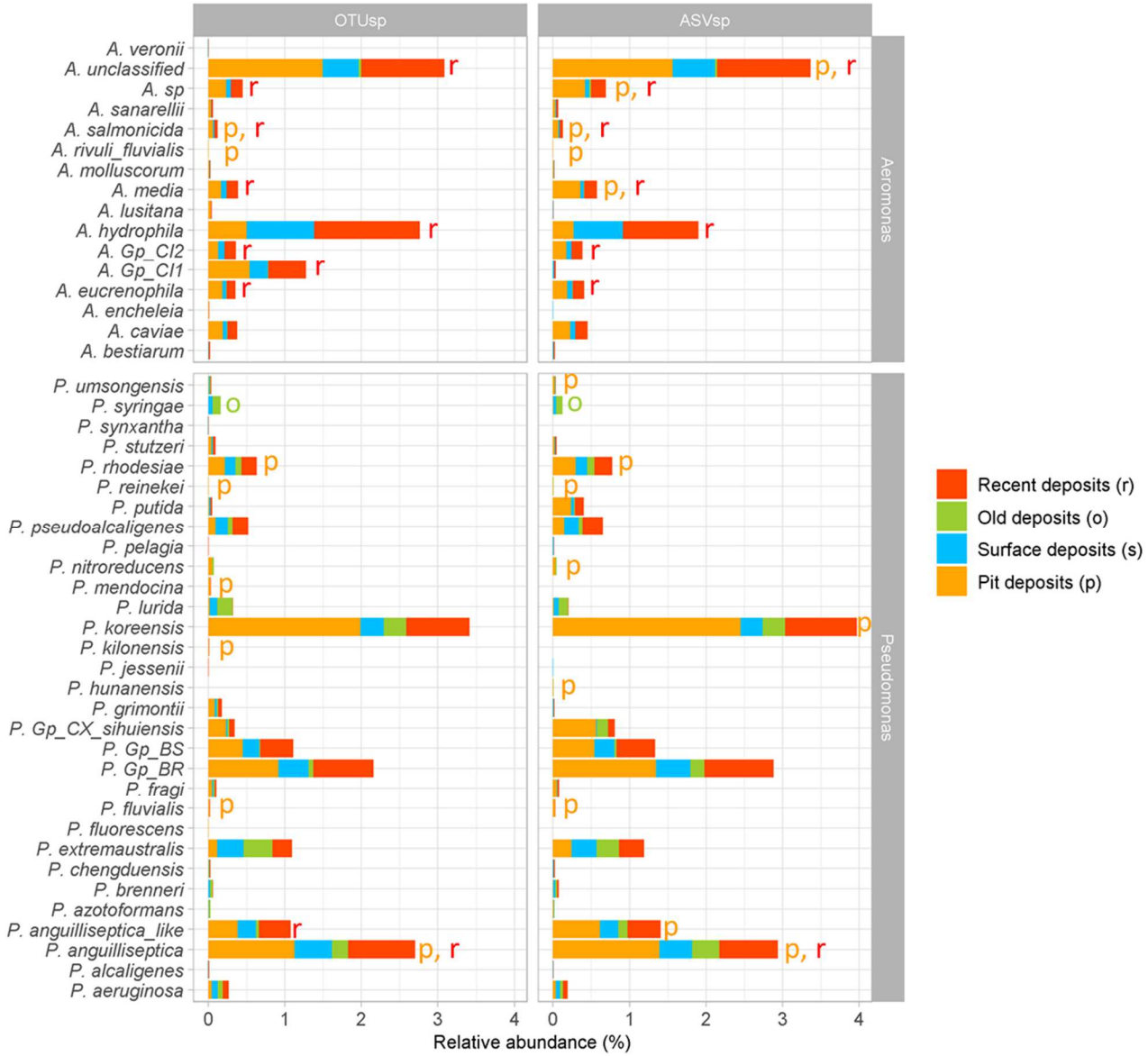

Figure 6 Aigle et al.

Table 1. Envfit (NMDS), RDA and CCA (anova) investigating the relationships between contents of urban sediments in high priority chemicals, and their *tpm* OTU or ASV profiles, while considering their origin (position) and accumulation / maturation time periods among the detention basin. See Table S2 for the full dataset.

Most significant variables (V= total number)	NMDS envfit	RDA	CCA (anova)	CCA - anova (constrained by a variable and aging)	CCA - anova (constrained by a variable and position)	CCA - anova (constrained by a variable, aging, and position)
Global field parameters						
Sediment accumulation / maturation periods	**0.001/0.001**	0.005/0.005	**0.001/0.001**	-	-	-
Position in the detention basin	0.004/0.002	0.065/0.010	0.002/**0.001**			
Metallic pollutants (V=6; no 2010/2015 samples)						
Cadmium (Cd)	**0.001/0.001**	NA[1]/NA[1]	0.002/**0.001**	ns/**0.001**	ns/**ns**	NA[2]/NA
Chrome (Cr)	0.022/0.019	NA[1]/NA[1]	0.672/0.011	**0.002**/**ns**	0.038/ns	NA/NA
Nickel (Ni)	0.010/0.011	NA[1]/NA[1]	0.008/0.011	**0.001/0.001**	0.011/ns	NA/NA
Polycyclic aromatic hydrocarbons (V=16, no 2010 samples)						
Naphthalene (H1)	0.098/0.010	0.200/0.600	0.002/**0.001**	**0.001/0.001**	0.004/**0.001**	ns/**0.001**
Fluorene (H4)	0.7/0.190	0.101/0.520	0.101/0.029	ns/**0.001**	ns/ns	NA/NA
Anthracene (H6)	0.606/0.268	0.065/0.210	0.010/0.003	**0.001/0.001**	ns/0.003	NA/**0.001**
Benzo(a)anthracene (H9)	0.315/0.185	0.765/0.795	0.004/0.021	**0.001**/ns	0.064/ns	NA/NA
Chrysene (H10)	0.274/0.250	0.005/0.005	0.090/0.040	**0.001/0.001**	0.013/**0.001**	**0.001/0.001**
benzo(b)fluoranthene (H11)	0.107/0.089	0.005/0.005	**0.001**/0.002	ns/**ns**	ns/ns	NA/NA
Pesticides (no 2010 samples)						
Chlorpyrifos	0.417/0.068	0.020/0.010	**0.001/0.001**	**0.001/0.001**	0.002/**0.001**	**0.001/0.001**
Alkyphenols (V=5, no 2010 samples)						
4-nonylphenol (A2)	0.591/0.043	0.040/0.010	0.012/0.002	**0.001/0.001**	0.007/0.003	**0.001/0.001**
Nonylphenol-mono-ethoxyle (A4)	0.010/0.013	0.120/0.165	0.005/0.016	**0.001/0.001**	ns/ns	NA/NA
Nonylphenol-bi-ethoxyle (A5)	0.151/0.024	0.545/0.565	0.86/0.621	ns/**0.001**	ns/ns	NA/NA
Integrons (V=4)						
int2	**0.001/0.001**	0.435/0.385	**0.001/0.001**	**0.001/0.001**	**0.001/0.001**	**0.001/0.001**
int3	**0.001/0.001**	0.120/0.015	0.075/0.118	**0.001/0.001**	**0.001/0.001**	**0.001/0.001**
Microbial source tracking (MST)						
HF183 (human fecal contaminations) DNA marker	0.004/0.044	0.495/0.710	0.009/0.026	**0.001/0.001**	0.017/0.006	NA/0.001
All Bacteroides spp. DNA	**0.001/0.001**	0.005/0.025	**0.001**/0.003	**0.001/0.001**	**0.001/0.001**	**0.001/0.001**

[1]Not applicable because of missing values; [2]Not significant. Significant p-values are in black and the most significant are in bold characters (≤0.001). Non-significant relations with the tested variables (p>0.05) are shown in grey. CCA plots of underlined statistical tests are shown in Figure 2.

GRAPHICAL ABSTRACT