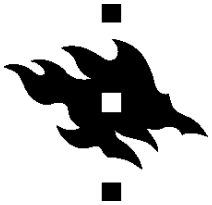


HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

The Use of epicPCR to study Antimicrobial Resistance Genes and Their Bacterial Hosts in Human Impacted Environments in West Africa

Anniina Sarekoski
Master of Science Thesis
University of Helsinki
May 2020

Tiedekunta – Fakultet – Faculty Faculty of Agriculture and Forestry		Koulutusohjelma – Utbildningsprogram – Degree Programme Microbiology and Microbial Biotechnology
Tekijä – Författare – Author Anniina Sarekoski		
Työn nimi – Arbetets titel – Title The Use of epicPCR to study Antimicrobial Resistance Genes and Their Bacterial Hosts in Human Impacted Environments in West Africa		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Microbiology		
Työn laji – Arbetets art – Level Master of Science Thesis	Aika – Datum – Month and year May 2020	Sivumäärä – Sidoantal – Number of pages 42

Tiivistelmä – Referat – Abstract

The emerging crisis of antimicrobial resistance is especially worrisome in low-income countries that lack controlled antibiotic policy and have poor infrastructure. Inadequate hygiene practices combined with ability of microbes to quickly evolve and adapt to changes rise the concern of resistance of infectious pathogens to many first-line antimicrobial drugs. Moreover, wastewaters that are widely used as irrigation water in urban gardening in sub-Saharan Africa, can function as vehicle for the dissemination of bacteria that carry antimicrobial resistance genes into the surrounding environment.

In this study, eight anthropogenically impacted water samples were collected from Burkina Faso and Mali and differences in their microbial communities were evaluated by 16S rRNA gene sequencing. Also, the presence of antimicrobial resistance genes was examined with SmartChip qPCR. The bacterial host range of bla_{NDM}, bla_{CTX-M}, bla_{OXA} and *qacEΔ1* was profiled using a novel culture-independent technique, Emulsion, Paired Isolation and Concatenation PCR (epicPCR).

The presence of 202 genes associated with antimicrobial resistance were detected with SmartChip qPCR array analysis, including carbapenemase genes that can transfer horizontally. Worryingly, sixteen taxonomical units, including possible human pathogens *Acinetobacter*, *Klebsiella*, *Escherichia* and *Pseudomonas,* were found to carry all the four genes investigated with epicPCR. The most abundant genus *Arcobacter* along with *Dechloromonas, Methylotenera, MM1* and *Methylophilus* were new discoveries as bla_{NDM} hosts. Furthermore, a considerable number of bla_{OXA} and clinical class 1 integron marker *qacEΔ1* gene hosts were discovered in every sample. Lastly, putative events of horizontal gene transfer in two WWTP samples were observed.

Broad host range of bla_{OXA} and *qacEΔ1* genes suggests a heavy antimicrobial resistance genes burden in West Africa and the results support the theory that environmental bacteria can function as resistance gene reservoirs. These results show occurrence of horizontally transferrable bla_{NDM} and bla_{CTX-M} genes in pathogens especially in hospital wastewater, and a threat of their spread into the environment and to the community. However, to decipher their role in the infectious disease burden in Africa, more research is needed.

Avainsanat – Nyckelord – Keywords
Antimicrobial resistance gene, wastewater, epicPCR, culture-independent, West Africa

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Kaisa Haukka, Antti Karkman, Marko Virta

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information
Funded by the Academy of Finland Develop programme

Tiedekunta – Fakultet – Faculty Maatalous-metsätieteellinen tiedekunta	Koulutusohjelma – Utbildningsprogram – Degree Programme Mikrobiologia ja mikrobibiotekniikka	
Tekijä – Författare – Author Anniina Sarekoski		
Työn nimi – Arbetets titel – Title Mikrobilääkeresistenssigeenien ja niiden isäntäbakteereiden tutkiminen ihmisen vaikutuksen alaisissa ympäristöissä Länsi-Afrikassa epicPCR-menetelmää käyttäen		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Mikrobiologia		
Työn laji – Arbetets art – Level Maisterintutkielma	Aika – Datum – Month and year Toukokuu 2020	Sivumäärä – Sidoantal – Number of pages 42

Tiivistelmä – Referat – Abstract

Mikrobilääkeresistenssi on lisääntynyt erityisen huolestuttavasti matalan tulotason maissa, joista puuttuu kontrolloitu mikrobilääkepolitiikka ja joissa on heikko infrastruktuuri. Riittämättömät hygieniakäytännöt yhdistettynä mikrobien nopeaan evoluutioon herättävät huolen taudinaiheuttajien resistenssistä yhä useammalle mikrobilääkkeelle. Mikrobilääkeresistenssigeenejä kantavien bakteereiden leviämistä ympäristöön edesauttavat erilaiset jätevedet, joita käytetään laajalti kasteluvesinä kaupunkiviljelyssä Saharan eteläpuolisessa Afrikassa.

Tässä tutkimuksessa kerättiin kahdeksan antropogeenistä vesinäytettä Burkina Fasosta ja Malista ja niiden mikrobiyhteisöjä kuvailtiin ja eroja arvioitiin 16S rRNA-geenin sekvenssien perusteella. Mikrobilääkeresistenssigeenien esiintymistä näytteissä tutkittiin SmartChip qPCR:n avulla. Lisäksi neljän resistenssigeenin, bla_{NDM}, bla_{CTX-M}, bla_{OXA} ja *qacEΔ1*:n isäntäkirjoa profiloitiin hiljattain kehitetyllä molekulaarisiin menetelmiin perustuvalla epicPCR-menetelmällä (eng. Emulsion, Paired Isolation and Concatenation PCR).

Yhteensä 202 erilaista mikrobilääkeresistenssiin linkittyvää geeniä, mukaan lukien horisontaalisesti siirtyviä karbapenemaasigeenejä, havaittiin SmartChip qPCR -analyysillä. Kuudentoista bakteeri-isännän, myös mahdollisesti ihmisille tautia aiheuttavien *Acinetobacter*, *Klebsiella*, *Escherichia* ja *Pseudomonas* -sukujen havaittiin kantavan kaikkia neljää epicPCR:llä tutkittua geeniä. Näytteissä kaikkein yleisimpänä esiintynyt bakteerisuku *Arcobacter* sekä *Dechloromonas*, *Methylotenera*, *MM1* ja *Methylophilus* olivat uusia bla_{NDM} -geenin kantajia. Lisäksi jokaisesta näytteestä löydettiin huomattavan suuri määrä bla_{OXA} ja *qacEΔ1* -geenien kantajia. Kahden näytteen perusteella jätevedenpuhdistamon vesissä havaittiin horisontaalista geenien siirtoa.

Laaja bla_{OXA} ja *qacEΔ1* -geenien isäntäkirjo viittaa Länsi-Afrikan vakavaan mikrobilääkeresistenssigeenien taakkaan ja tulokset tukevat teoriaa, jonka mukaan ympäristöbakteerit voivat toimia resistenssigeenien varastoina. Nämä tulokset osoittavat horisontaalisesti siirtyvien kliinisesti merkittävien bla_{NDM} ja bla_{CTX-M} -geenien esiintymisen patogeeneissä erityisesti sairaalan jätevesissä, ja uhan niiden leviämisestä ympäristöön ja yhteisöön. Niiden rooli Afrikan tartuntatautien taakassa on kuitenkin epäselvä ja sen selvittämiseksi tarvitaan enemmän sekä kansallista että maailmanlaajuista yhteisyötä ja tutkimusta.

Avainsanat – Nyckelord – Keywords
Mikrobilääkeresistenssigeeni, jätevesi, epicPCR, molekulaariset menetelmät, Länsi-Afrikka

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Kaisa Haukka, Antti Karkman, Marko Virta

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information
Suomen Akatemian Kehitystutkimus-ohjelman rahoittama

Introduction

Treating bacterial infections is getting increasingly complicated due to the ability of bacteria to evolve multitudinous defense mechanisms against antimicrobial agents (Viens and Littmann, 2015). Antimicrobial resistance genes (ARG) have been existing and capable of conferring resistance to naturally occurring antibiotics long before our civilization existed (Barlow and Hall, 2002). Regardless of the emerging problem of antimicrobial resistance (AMR) that has been contributed by ancient evolutionary events, human actions have considerably facilitated the rise of post-antibiotic era (Barlow and Hall, 2002).

Antimicrobial resistance is especially increasing in low-income countries in sub-Saharan Africa that lack controlled antibiotic policy and have poor infrastructure (Bernabé et al., 2017). Inadequate hygiene practices combined with ability of microbes to quickly evolve and adapt to changes rise the concern of resistance of infectious pathogens to many first-line medicines. Furthermore, infectious diseases are still the leading cause of death in the African continent (World Health Organization, 2017). Top three critical priority pathogens listed by World Health Organization (WHO) include carbapenem-resistant *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* (World Health Organization, 2017). However, their prevalence in West Africa is still largely unknown, as in addition to many other deficiencies and vulnerabilities related to public health care, the area has major gaps in AMR data (Bernabé et al., 2017).

One of the main sources of anthropogenic pollution is effluent from urban wastewater treatment plants (Rizzo et al.,2013). In fact, urban wastewater treatment plants (WWTP) are considered as hotspots for ARG dissemination and transfer due to the presence of waters rich in nutrients, antibiotics, mobile genetic elements and bacteria (Rizzo et al., 2013; Guo et al., 2017). Rapidly growing population and gradual industrialization especially in West Africa's urban areas has led to situation where the capacity to treat wastewaters is exceeded and only part of water are treated (Mafuta et al., 2018). Also, wastewaters are widely used for irrigation in urban gardening (Drechsel et al., 2008). That is of concern, since wastewaters function as vehicle for dissemination of bacteria that carry ARGs into the surrounding environment. Nonpathogenic bacteria living in their natural water environments can become reservoirs of resistance genes, when in contact with drug-resistant bacteria originating from anthropogenic

waters, such as hospital wastewater (Dolejska et al., 2011). However, the use of wastewater for irrigation enables food production for growing population (Drechsel et al., 2008; Bougnom et al., 2019). Consequently, people suffering from endemic poverty are directly exposed to untreated wastewater due to lack of clean water and sanitation (Bougnom et al., 2019).

Horizontal dissemination of ARGs is of concern especially if clinically relevant pathogens acquire resistance genes, which highlights a need to study the bacteria carrying the ARGs. Recently developed culture-independent Emulsion, Paired Isolation, and Concatenation PCR (epicPCR) enables the study of the host range of ARGs in environmental samples (Spencer et al., 2016). Taxonomic identity is revealed by linking phylogenetic markers and target genes in single-cell level. With a polymerization process, cells are first captured inside polyacrylamide beads. Next, the target gene and 16S rRNA gene in a cell are fused with a linker primer and lastly, the fusion PCR product is amplified in a nested PCR.

In this study, four resistance genes that are able to spread horizontally via plasmids were chosen for examination. bla_{CTX-M} is the most common type of extended spectrum β-lactamase (ESBL) gene, and bacteria that carry the gene are able to hydrolyze third generation cephalosporins (Eckert, Gautier and Arlet, 2006). New Delhi metallo-β-lactamase, bla_{NDM} is a highly mobile carbapenemase gene that is able to hydrolyze almost all β-lactam antibiotics, and very limited treatment options against infections caused by NDM-positive strains exist (Wu et al., 2019). bla_{OXA-10} is a class D β-lactamase gene and a so called non-ESBL, a narrow spectrum β-lactamase. However, bacteria carrying the gene are resistant to many other groups of antibiotics (Naas and Nordmann, 1999). The *qacEΔ1* gene is a biocide resistance gene, conferring resistance to a quaternary ammonium disinfectant and located in Class I integron (Romão et al., 2011). Class I integron is an anthropogenic impact marker and genetic element that has promoted dissemination of antimicrobial resistance and, moreover, is extensively responsible of AMR in gram-negative bacteria (Hegstad et al., 2010; Romão et al., 2011; Pärnänen et al., 2019).

This study is a part of a consortium project called Antimicrobial Resistance in West Africa (AMRIWA). The consortium consists of expertise from different fields, so together microbiologists, medical doctors and sociologists explore the flow of antimicrobial resistance genes between water, soil, animals, food and humans from the perspective of One Health. To

understand the mechanisms behind the evolution of AMR in general, different kinds of environmental and clinical samples from various sampling sites in Burkina Faso, Benin and Mali have been collected.

The aims of this research were to describe and compare microbial communities in eight samples obtained from anthropogenically impacted water environments in Burkina Faso and Mali. Further, the purpose was to investigate the bacterial host range of four resistance genes, namely bla_{NDM-1}, $bla_{CTX-M-15}$, bla_{OXA-10} and *qacEΔ1*. Based on the bacterial host range data obtained, speculations and observations of putative horizontal gene transfer in two WWTP samples were included in this study. Finally, this research aimed to conceptualize and create information on antimicrobial resistance in West Africa, especially of the emergence of infectious pathogens carrying antimicrobial resistance genes in hospital wastewater.

Materials and methods

Wastewater sample collection

All together eight sampling sites were chosen, five of them in Burkina Faso and three in Mali (Table 1). Samples were collected during dry season in early 2019 together with local students and co-workers from the AMRIWA consortium. Three biological replicates were taken of each sample. The town of Nanoro (Table 1) is located in rural area in central Burkina Faso. The hospital wastewater sample (sample No 1.) was collected from a septic tank of the St. Camille Hospital of Nanoro. Ouagadougou is the capital of Burkina Faso. Currently the city is estimated to have close to 2.8 million inhabitants with an annual growth rate of 4.8 % (https://worldpopulationreview.com/world-cities/ouagadougou-population/). Yalgado Ouedraogo University Hospital in Ouagadougou is the largest hospital in the country. The Yalgado hospital wastewater sample (sample No 2.) was collected from a sewer pipe connected to the sewer system. The WWTP influent (sample No 3.) was collected from an intersection point of the sewer system, where several wastewater sources come together, to be further directed to the Ouagadougou WWTP. The effluent sample (sample No 4.) was collected from red-colored shallow runoff originating from the WWTP. The sample No. 5 was wastewater from an abattoir located in an industrial area next to the WWTP and the sampled water had undergone a biological pretreatment process.

Bamako is the capital of Mali, and its population is estimated to be 2.6 million. In addition, the annual population growth rate of Bamako is 3.4 % (https://worldpopulationreview.com/world-cities/bamako-population/) The hospital wastewater (sample No 6.) was originating from the university hospital of Bamako, Gabriel Touré, and the sample was collected from a discard pipe next to the Niger river. Lastly, the sample No. 7 represent the urban area canals and No. 8 the gutters that are intended for heavy rainwater flow but are often used as open sewage systems.

Samples were collected to clean plastic or glass bottles size of approximately one liter. Bottles were kept in a cooler with ice blocks and processed in a laboratory within 24 h. Samples collected from Burkina Faso were processed in a laboratory of the Clinical Research Unit of Nanoro (CRUN) and samples from Mali were processed in Charles Mérieux Centre of Infectiology in Bamako (CICM). Treated sample volume varied depending on the sample, location and laboratory (Table 1).

Table 1. Human impacted waters collected from Burkina Faso and Mali, from three different areas, namely Nanoro, Ouagadougou and Bamako.

Sample No.	Sample description	Country	Area	Location	Centrifuged volume for epicPCR (ml)
1	hospital wastewater	Burkina Faso	Nanoro	12°41'20.4"N 2°11'26.6"W	36
2	hospital wastewater	Burkina Faso	Ouagadougou	12°23'01.9"N 1°30'16.9"W	24
3	WWTP influent water	Burkina Faso	Ouagadougou	12°23'36.2"N 1°29'57.9"W	24
4	WWTP effluent water	Burkina Faso	Ouagadougou	12°25'27.4"N 1°27'53.1"W	24
5	slaughter site wastewater	Burkina Faso	Ouagadougou	12°24'59.0"N 1°28'32.1"W	24
6	hospital wastewater	Mali	Bamako	12°38'01.9"N 7°59'38.2"W	80
7	canal water	Mali	Bamako	12°39'39.7"N 7°57'29.0"W	40
8	street gutter water	Mali	Bamako	12°38'20.6"N 8°00'30.7"W	40

Sample pretreatment for epicPCR

Burkinabe samples were centrifuged 7700 × *g* for one hour at approximately 4 °C (ROTANTA 460 R, Andreas Hettich GmbH & Co. KG, Tuttlingen, Germany). Supernatant was discarded and pellet resuspended into 20 % v/v glycerol. Samples were kept at – 70 °C and shipped to Finland on dry ice in 15 ml tubes (Corning, Fisher Scientific, UK). The same procedure was used to Malian samples, however, the relative centrifugal force in centrifugation was 4420 × *g* (Sigma 3-18K, Sigma).

DNA extraction for 16S rRNA gene sequencing and SmartChip qPCR analysis

The volume of filtered water varied from 100 ml to 300 ml, depending on turbidity: 200 ml for sample number 1, 100 ml for samples 2-6, and 300 ml for sample 7. DNA from sample number 8 (Bamako gutter water) was extracted with DNeasy PowerLyzer PowerSoil Kit (Qiagen, Germany) due to the muddy and turbid consistency of the sample. Samples 1 to 7 were filtered through a 0.2 µm polycarbonate filter (Whatman™, GE Healthcare Life Sciences) using a portable vacuum pump (Millivac-Mini Vacuum Pump XF54, Millipore, Merck, Germany). Subsequently, the filters were placed in PowerWater Bead tubes with sterile tweezers and kept at -20 °C until the start of DNA extraction with DNeasy PowerWater Kit (Qiagen, Germany). DNA was extracted according to manufacturer's instructions together with local students in Burkina Faso and Mali.

16S rRNA gene sequencing

In Finland, for 16S rRNA gene sequencing, the V3-V4 region of the 16S rRNA gene was amplified with following a protocol provided by the enzyme's manufacturer (Thermo High–Fidelity DNA Polymerase Product Information Sheet). Extracted DNA of each sample was used as a template. The amount of template DNA was adjusted with PCR-grade water to 75-140 ng per reaction. Also, three negative controls (PCR-grade water as template) were included. PCR reactions were run in a volume of 50 µl and cycling conditions were starting with initial denaturation at 98 °C (30 s), followed by 14 cycles of denaturation (98 °C, 10 s), annealing (59 °C, 30 s) and extension (72 °C, 30 s), and ending to a final extension (72 °C, 5 min). An equimolar mix of both 16S rRNA forward and reverse primers Illum_519F_1-4, and Illum_785R_1-4, respectively, were used (Table 2). After purification of PCR products (Monarch PCR & DNA Cleanup Kit New England Biolabs, Inc., Ipswich, MA, USA), DNA

quality and concentrations were determined by Qubit Broad-Range Assay Kit (Thermo Fisher Scientific, Waltham, USA). Lastly, 16S rRNA gene PCR products were sequenced with Illumina MiSeq sequencing platform at the Institute of Biotechnology of the University of Helsinki.

Table 2. Primers used in the study. F1 (fusion PCR) and F1_TS (nested PCR, Illumina TruSeq adapter sequences underlined) are target gene forward primers, R1-F2' is a linker primer (16S forward primer part bolded) used in fusion PCR and pH' in turn is a universal 16S rRNA gene primer used in fusion PCR.

Primer	Sequence (5'-3')	Source
blaOXA10 F1	CGCAATTATCGGCCTAGAAACT	Stedtfeld et al., 2018
blaOXA10 R1-F2'	**GWATTACCGCGGCKGCTGCTG**TTGGCTTTCCGTCCCATTT	This study; Stedtfeld et al., 2018
blaOXA10 F1_TS	ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCAATTATCGGCCTAGAAACT	This study; Stedtfeld et al., 2018
blaCTX-M F1	GCCGCGGTGCTGAAGA	Stedtfeld et al., 2018
blaCTX-M R1-F2'	**GWATTACCGCGGCKGCTGCTG**ATCGGATTATAGTTAACCAGGTCAGATTT	This study; Stedtfeld et al., 2018
blaCTX-M F1_TS	ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCGCGGTGCTGAAGA	This study; Stedtfeld et al., 2018
blaNDM F1	GGCCACACCAGTGACAATATCA	Stedtfeld et al., 2018
blaNDM R1-F2'	**GWATTACCGCGGCKGCTGCTG**CAGGCAGCCACCAAAAGC	This study; Stedtfeld et al., 2018
blaNDM F1_TS	ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCACACCAGTGACAATATCA	This study; Stedtfeld et al., 2018
qacEΔ1 F1	TCGCAACATCCGCATTAAAA	Eckert, Gautier and Arlet, 2006

qacEΔ1 R1-F2'	**GWATTACCGCGGCKGCTG**ATGGATTTCAGAACCAGAGAAAGAAA	Karkman et al., 2016
qacEΔ1 F1-TS	ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCAACATCCGCATTAAAA	Karkman et al., 2016
Universal 16S rRNA gene primer pH’	AAGGAGGTGATCCAGCCGCA	Edwards et al., 1989
Blocking primer U519F block_F	TTTTTTTCAGCMGCCGCGGTAATWC/3SpC3/	Spencer et al., 2016
Blocking primer U519R block_R	TTTTTTTGWATTACCGCGGCKGCTG/3SpC3/	Spencer et al., 2016
16S forward primer Illum_519F_1	CAGCMGCCGCGGTAATWC	Baker, Smith and Cowan, 2003
16S forward primer Illum_519F_2	gtCAGCMGCCGCGGTAATWC	Baker, Smith and Cowan, 2003
16S forward primer Illum_519F_3	agagCAGCMGCCGCGGTAATWC	Baker, Smith and Cowan, 2003
16S forward primer Illum_519F_4	tagtgtCAGCMGCCGCGGTAATWC	Baker, Smith and Cowan, 2003
16S reverse primer Illum_785R_1	GACTACHVGGGTATCTAATCC	Herlemann et al., 2011
16S reverse primer Illum_785R_2	aGACTACHVGGGTATCTAATCC	Herlemann et al., 2011

16S reverse primer Illum_785R_3	tctGACTACHVGGGTATCTAATCC	Herlemann et al., 2011
16S reverse primer Illum_785R_4	ctgagtgGACTACHVGGGTATCTAATCC	Herlemann et al., 2011

16S rRNA gene analysis

16S rRNA gene analysis started with using UNIX tools and bioinformatics software. Primers and adapters were removed from the raw reads by using cutadapt version 2.7 (Martin, 2011) and FastQ version 0.11.8 was used for quality control (Andrews, 2011). Next, in R, DADA2 package (Table 3) was used to resolve amplicon sequence variants (ASV). Before merging the forward and reverse reads with *mergePairs* command, *filterAndTrim* command was used to truncate forward and reverse reads to a length of 240 and 160 nucleotides, respectively, while maximum number of expected errors allowed was two. ASVs were combined to a table with *makeSequenceTable* command, chimeras were removed with *removeBimeraDenovo*, and taxonomy was assigned with *assignTaxonomy* and *addSpecies* commands. Silva version 132 database was used as the reference for taxonomic classification (Quast et al., 2013).

Phyloseq object was constructed with using the phyloseq package (Table 3). Sequence data was contained with command *Biostrings::DNAStringSet* that is included in the biostrings package (Table 3). Next, the commands *taxa_names* and *merge_phyloseq* were used to obtain the taxa names and for merging all the arguments into one phyloseq object. Alpha diversity was estimated with phyloseq's *plot_richness* command and ggplot package was used for visualization (Table 3). Next, data were transformed to proportions for Bray-Curtis dissimilarity index with *transform_sample_counts* command and ordinated with *ordinate* command and NMDS method. Ordination was visualized with *plot_ordination* command. With *tax_glom* command, ASVs that had the same taxonomic rank at genus level were merged. Genus abundance was deciphered with commands *transform_sample_counts* and *prune_taxa*. Abundance list was arranged with packages tibble and dplyr (Table 3) with commands *group_by*, *summarize*, *arrange* and *mutate*. Finally, twenty most relatively abundant genera among all samples were visualized with ggplot2 (Table 3).

Table 3. R and its packages.

Software/package and version	Used in	Source
R 3.6.2	epicPCR data / 16S rRNA gene analysis	RStudio Team, 2019
DADA2 1.14.1	16S rRNA gene analysis	Callahan et al., 2016
Phyloseq 1.30.0	16S rRNA gene analysis	McMurdie and Holmes, 2013
Dplyr 0.8.5	epicPCR data / 16S rRNA gene analysis	Wickham et al., 2020
Ggplot2 3.3.0	epicPCR data / 16S rRNA gene analysis	Wickham, 2016
Tibble 2.1.3	16S rRNA gene analysis	Müller and Wickham, 2019
Data.table 1.12.8	epicPCR data analysis	Dowle and Srinivasan, 2019
Biostrings 2.54.0	16S rRNA gene analysis	Pagès et al., 2019

SmartChip qPCR assay for selection of genes to epicPCR

DNA concentrations were measured with NanoDrop™ One (ThermoFisher Scientific, Waltham, USA) and adjusted to approximately 10 ng μl^{-1} with PCR-grade water. Biological replicates were pooled. The presence of 384 antimicrobial resistance genes was assessed with SmartChip qPCR array analysis (Takara Bio Inc., USA), service provided by Resistomap oy, Helsinki, Finland. Raw data was analyzed using R version 3.6.2 (RStudio Team, 2019), tidyverse version 1.3.0 (Wickham et al., 2019) and pheatmap version 1.30.0 (Kolde, 2019).

epicPCR

EpicPCR protocol was modified from the original version by Spencer et al. (2016). A volume of 100 µL of each concentrated sample was centrifuged (Eppendorf® Centrifuge 5424, Merck, Germany) for 1 min at 12 000 × *g* to remove glycerol. Next, pellets were resuspended to 100 µL PCR-grade water. The samples were further diluted to PCR-grade H_2O prior polymerization process, with dilution factor ranging from zero to 1:50. Polymerization process was completed with the modifications to Spencer et al. (2016) introduced by Hultman et al. (2018). The validation of the most optimal dilutions was made based on fluorescence microscopy (Nikon ECLIPSE E600, Nikon Instruments Inc., Melville, USA) and staining of the cells with Sybr Green II DNA Gel Stain (Thermo Fisher Scientific, Waltham, USA). The most optimal dilution

was chosen by observing the prevalence of empty beads and beads containing a singular cell. Cell density of more than one cell in a bead was to be avoided.

EpicPCR primers blaOXA10, blaCTX-M and blaNDM, listed in Table 2, were designed based on the qPCR array 2.0 previously published primers (Stedtfeld et al., 2018). The primers were used unchanged as forward primers in fusion PCR and by adding Illumina TrueSeq adapter overhangs to them they were used as forward primers in nested PCR.

Briefly, fusion PCR protocol (Spencer et al., 2016) was modified as presented in Hultman et al. (2018) with following reagents: 1 × GC buffer (New England Biolabs, Inc., Ipswich, MA, USA), 1 mM MgCl2 (New England Biolabs, Inc., Ipswich, MA, USA), 0.25 mM dNTPs (BioNordika, Finland), 1 μM forward primer F1 (see table 2 for primers), 1 μM reverse primer pH', 0.01 μM linker primer R1-F2' and 0.16 U μl^{-1} Phusion Hot Start Flex DNA polymerase (New England Biolabs, Inc., Ipswich, MA, USA). Unlike in Hultman et al. (2018), in diethyl ether extraction for ABIL EM 90 emulsions, 50 μL of water was added to the sample after first addition of water-saturated diethyl ether and centrifugation. All PCR steps were run using a C1000 Touch Thermal Cycler (Bio-Rad Laboratories, Hercules, USA) and the PCR products were purified with Monarch PCR & DNA Cleanup Kit (New England Biolabs, Inc., Ipswich, MA, USA).

Modified from Spencer et al. (2016), an additional PCR with using only blocking primers was implemented to prevent the amplification of unfused products. Blocking PCR was completed in 50 μL reaction volume duplicates and reagents used were as follows: 1 × GC buffer (Thermo Fisher Scientific, Waltham, USA), 0.2 mM dNTPs (BioNordika, Finland), 0.02 U μl^{-1} Phusion polymerase (Thermo Fisher Scientific, Waltham, USA), 3.2 μM block_F and 3.2 μM block_R. The thermal cycling started with an initial denaturation at 98 °C for 30 s, followed by 30 cycles of denaturation (98 °C, 10 s), annealing (55 °C, 30 s) and extension (72 °C, 30 s) and ending in final extension of 5 mins at 72 °C.

In nested PCR, 2 μL of purified blocking PCR product was used as a template. Nested PCR was run in quadruplicate in 25 μL reaction volume, and deviating from Spencer et al. (2016), annealing temperature was raised to 60 °C. Also, no additional amplification with Illumina adapters was conducted, as the adapter overhangs were included in nested PCR. Reagents were

analogous to the blocking PCR, however, concentration of the blocking primers was lowered to 0.32 µM, and 0.3 µM forward nested primer F1_TS and 0.3 µM reverse 16S primer were added. For each run of nested PCR, a negative control (PCR-grade water as template) was included. PCR replicates were pooled, and product fragment sizes visualized with E-Gel EX Agarose gel, 2% (Invitrogen, Carlsbad, USA). Finally, the PCR products were purified and submitted for sequencing with Illumina MiSeq sequencing platform at the Institute of Biotechnology of the University of Helsinki.

epicPCR data analysis

epicPCR sequence processing was performed with UNIX tools and bioinformatics software (Table 4). For the most part, instructions for epicPCR analysis created by Katariina Pärnänen (https://github.com/KatariinaParnanen/epicPCR_analysis) were followed. Adapters and primers were removed from reads with cutadapt, followed by a quality check with FastQC (Table 4). Further, reads were assembled with PEAR and converted from fastq to fasta format (Table 4). Again, cutadapt was used to separate target gene fragment and 16S rRNA gene region by using the linker primer sequence. Chimeras were removed with reference-based chimera detection from the 16S rRNA reads with VSEARCH tool and the Silva "Gold" reference database (Quast et al., 2013, database obtained from https://github.com/torognes/vsearch/wiki/VSEARCH-pipeline). Using BLASTn (Table 4), β-lactamase gene part sequences were aligned to the resfinder database (Zankari et al., 2012) to verify their accuracy. The accuracy of *qacEΔ1* sequences was confirmed with aligning them to a published *qacEΔ1* gene sequence (Yoon, Park and Kim, 2015). Out of 1 153 bla_{NDM} gene reads, five reads were aligning to other than the target gene and they were removed manually. Next, based on the BLASTn analysis, the headers of the correct ARG reads were used with seqtk tool to find the corresponding 16S rRNA gene reads from a FASTA file. This had to be done, as unexpectedly, even after the size selection, there were more 16S rRNA reads than there were target gene reads. Finally, mothur (Table 4) was used for taxonomic classification with Silva version 132 database as the taxonomy reference. After taxonomic assignment, unassigned sequences were removed using mothur. Analysis of the 16S rRNA sequence reads was continued in R (Table 3).

The 16S rRNA sequence reads of each target gene were handled separately. A genus level approach to study the host range was adopted. Further, only the presence of the genus in each

sample's replicate was confirmed, thus the number of reads designated to each genus was irrelevant. Occurrences of false positives was scaled down with the replicates, as if a genus was only detected in one of the three replicates, it was ignored. This reduced the number of genera considerably, however, it was considered necessary to produce reliable results. Absence/presence information of all genes was constructed using dplyr and data.table packages (Table 3). For bla_{NDM} and bla_{CTX-M} , the host range was visualized with ggplot2 (Table 3). A phylogenetic tree of bla_{OXA} and *qacEΔ1* genes was built with FastTree by first aligning the sequences with muscle (Table 4). Finally, bacterial host range of aforesaid genes was visualized using Anvi'o interactive interface (Table 4).

Table 4. Bioinformatics tools for epicPCR analysis.

Software/ command line tool and version	Commands and parameters given	Used for	Reference
Cutadapt 2.7 (with Python 3.7.3)	cutadapt -a -A -o -p	removal of adapters	Martin, 2011
	cutadapt --trimmed only -max 5 -q 20 -m 300 -M 500	removal of primers	
	cutadapt --trimmed-only -0 15	selection of 16S region and target gene region from reads	
FastQC v0.11.8	fastqc -o	quality checks	Andrews, 2011
PEAR v0.9.6	pear -y 150M -j 8	assembly of reads	Zhang et al., 2014
FASTX Toolkit 0.0.14	fastq_to_fasta -i	FASTQ file transformation to FASTA file	Gordon (http://hannonlab.cshl.edu/fastx_toolkit/)
VSEARCH v2.14.1	vsearch --uchime_ref –db –sizeout –nochimeras	identification of chimeric sequences	Rognes et al., 2016
mothur v.1.40.4	classify.seqs fasta= reference=	classification of sequences	Schloss et al., 2009

	taxonomy= cutoff= 60 processors=1 probs= F		
	remove.lineage taxonomy= taxon=unknown	removal of unassigned sequences	
BLAST 2.10.0+	blastn -subject -query -outfmt 6 -max_target_seqs 1 -task blastn-short*	alignment of sequences to known databse	Altschul et al., 1990
seqtk 1.3-r106	seqtk -subseq	processing sequences in FASTA formats	(https://github.com/lh3/seqtk)
MUSCLE v3.8.31	muscle -in -out	alignment of sequences	Edgar, 2004
FastTree 2.1.10	fasttree -nt -gtr	creation of phylogenetic trees	Price et al., 2010
Anvi'o 6.2	anvi-interactive -t -p -manual -d	visualizations	Eren et al., 2015

* Used only for bla_{NDM} gene reads.

Results

Microbial community determined by 16S rRNA gene sequencing

Based on the 16S rRNA gene sequences, both α- and β-diversity of the microbial communities within and between the samples were estimated. The Shannon diversity index was used to assess the bacterial diversity in microbial communities in each sample (Figure 1). Both bacterial abundance and evenness in the sample (α-diversity) was weighted based on the 16S rRNA sequences. Species diversity is highly skewed in Ouagadougou WWTP effluent sample as two out of three replicates have substantially little diversity in the microbial community. Samples 2, 6 and 7 have the highest diversity of the samples and they also had very little variability between the replicates.

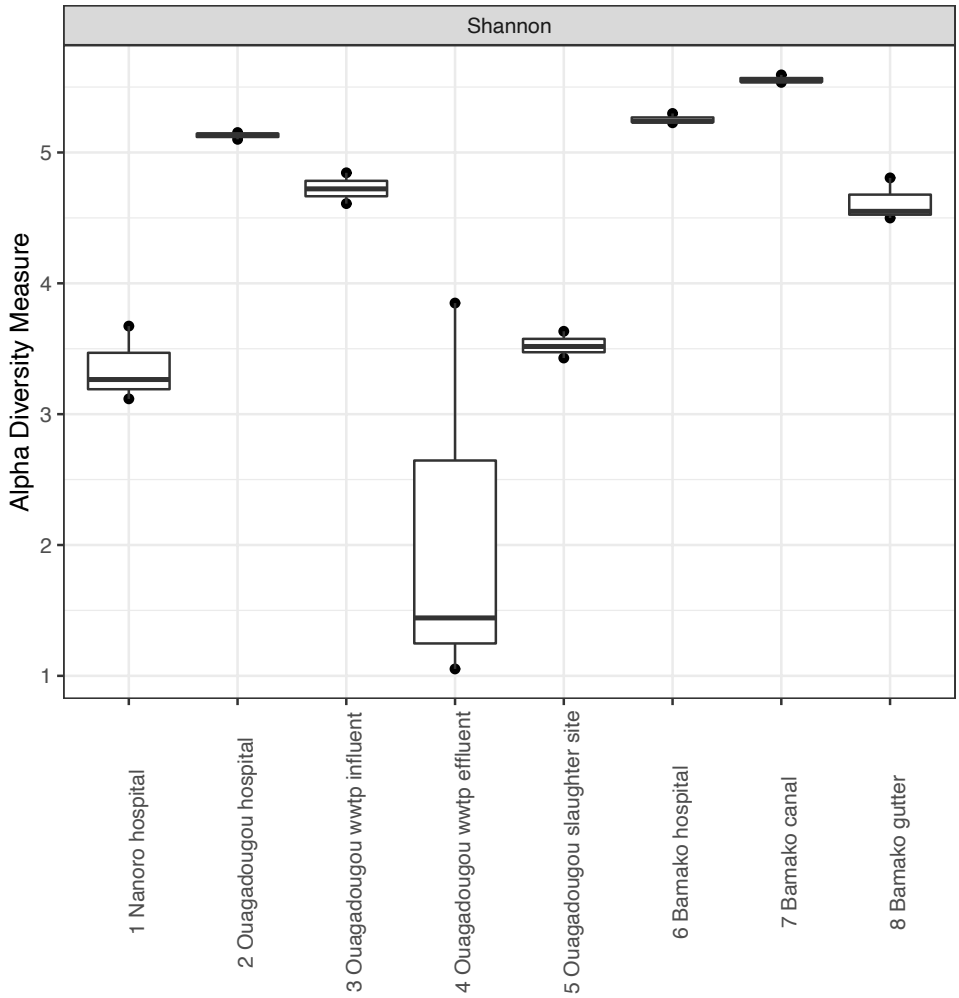


Figure 1. Alpha diversity boxplot of the microbial communities of each sampling site, calculated with Shannon diversity index. The more variability there is in the 16S rRNA gene sequences, the higher the interquartile range value is on the y-axis. Dots indicate two replicates and variability outside the upper and lower quartiles. The third replicate is shown as a median by the thickened horizontal line drawn over the box. Unbalanced box indicates skewness in the α-diversity. For example, sample number 4 has a lot of variety in its replicates as one of the replicates has a lot more species diversity than the other two. This creates skewness.

Bray-Curtis dissimilarity that takes into account species presence/absence and abundance, was used to quantify β-diversity, and it was visualized with a NMDS scaling plot (Figure 2). Stress value of the plot was 0.13, which indicates a good reliability. The biological replicates of the samples are situated close to each other, which suggest that they share a rather homogeneous microbial community. Only the Ouagadougou WWTP effluent is having notably scattered replicates. The effluent sample was collected to a bottle that had been tied to a stick. Thus, the sampling method of these effluent replicates combined with the heavy current of the shallow water might explain the discrepancy. Moreover, based on the ordination, the hospital wastewater samples (samples 1, 2 and 6) are dissimilar. Also, the samples from Bamako

hospital wastewater and Ouagadougou WWTP influent are overlapping, suggesting similarities in their microbial communities. Overall, no clear pattern or clustering is observed.

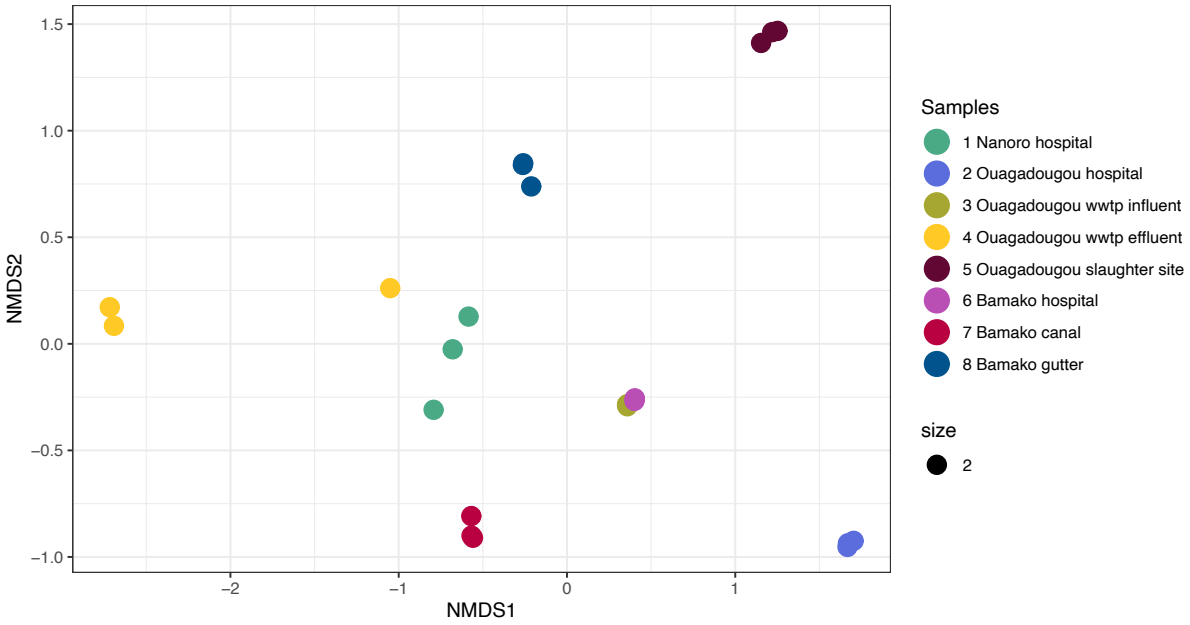


Figure 2. Bray-Curtis based non-metric multidimensional scaling plot (NMDS) illustrating the similarities of the microbial communities within samples. Distance between points in NMDS plot indicates dissimilarities in species composition. Each sample has three replicates. (NMDS plot stress value 0.13).

The most abundant genus of the microbial communities was *Arcobacter* and it was detected in all the samples (Figure 3). In some genera, also species was successfully classified. *Arcobacter cryarophilus* was the most abundant species of the genus in question. The second most abundant genus *Thiolamprovum* was found abundant only in the Ouagadougou WWTP effluent sample. *Thiolamprovum* constituted 91 % of the microbial community in Ouagadougou WWTP effluent replicate A, 81% in replicate B and 19% in replicate C. A strictly anaerobic species *Anaerocella delicata*, belonging to the third most abundant genus *Anaerocella*, was prominent especially in the Ouagadougou slaughter site sample. Other abundant genera were *Azonexus, Prevotella_9, Pseudomonas, Macellibacteroides, Dechloromonas, Acinetobacter, Paludibacter, Bacteroides, Rhodocyclus, Streptococcus, C39, Tolumonas, Comamonas, Proteiniclasticum, Blautia, Acidovorax* and *Feacalibacterium.*

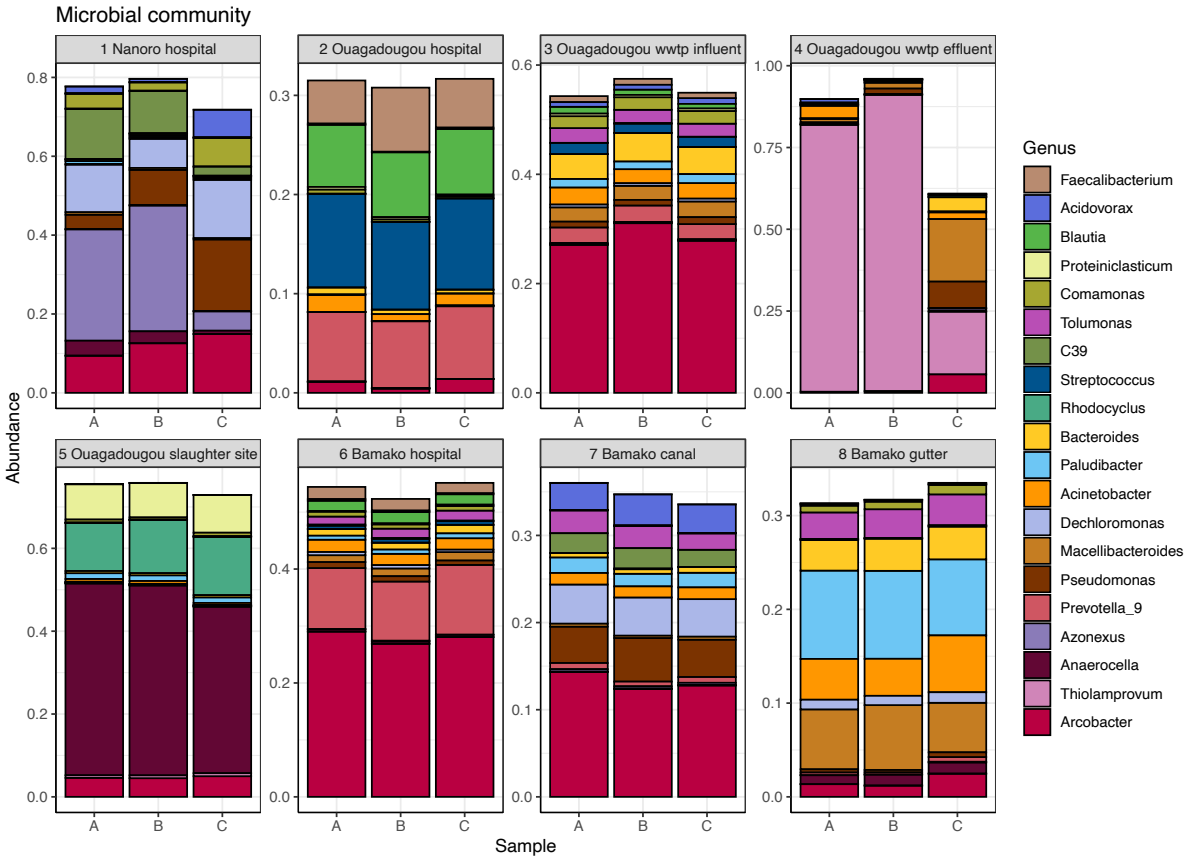


Figure 3. Twenty most abundant genera among all the samples, with the bottom one (*Arcobacter*) being the most abundant. Y axis shows relative abundances (%) of the genera that were found from the sample in question. A, B and C are the biological replicates.

The dominant genera are different in different samples (Table 5). Their relative abundances differ too, and in samples 2 (Ouagadougou hospital wastewater), 7 (Bamako canal water) and 8 (Bamako gutter water) roughly a third of the microbial community corresponds to the twenty most abundant genera (Figure 3). For example, the most abundant genus in Ouagadougou hospital water was *Streptococcus* (Table 5), which constitutes 9 % of the microbial community in the hospital water. Further, the most abundant genera for Bamako canal and gutter water samples were *Arcobacter* and *Paludibacter,* and their quota in the microbial community abundances were 14 % and 9 %, respectively.

Lastly, three negative controls were included into the 16S rRNA gene PCR and sequencing. However, they turned out as negative since no reads were left in the control samples after trimming of all the reads.

Table 5. Three most abundant (%) genera in each sample.

Sampling site	1.	2.	3.
1 Nanoro hospital	*Azonexus*	*Pseudomonas*	*Arcobacter*

2 Ouagadougou hospital	*Streptococcus*	*Prevotella_9*	*Enterococcus*
3 Ouagadougou WWTP influent	*Arcobacter*	*Bacteroides*	*Cloacibacterium*
4 Ouagadougou WWTP effluent	*Thiolamprovum*	*Macellibacteroides*	*Pseudomonas*
5 Ouagadougou slaughter site	*Anaerocella*	*Rhodocyclus*	*Proteiniclasticum*
6 Bamako hospital	*Arcobacter*	*Prevotella_9*	*Thiothrix*
7 Bamako canal	*Arcobacter*	*Aquabacterium*	*Flavobacterium*
8 Bamako street gutter	*Paludibacter*	*Macellibacteroides*	*Acinetobacter*

Antimicrobial resistance gene occurrence determined by qPCR

SmartChip qPCR array analysis revealed presence of high number of different antimicrobial resistance genes and resistance related genes in both countries and especially in hospital wastewaters. A total of 202 genes out of the tested 384 genes were found in at least one sample (Supplementary Figure 1). Genes associated to class 1 integrons (*intI1*, *qacE*Δ) and genes conferring resistance to tetracyclines (*tetM*), aminoglycosides (*aadA*) and sulfonamides (*sul*) were ubiquitous. Also, various horizontally transferrable carbapenemase genes (bla_{NDM}, bla_{IMP}, bla_{VIM}) were detected. The genes bla_{NDM} and bla_{VIM} were detected in two Burkinabe samples, in Nanoro hospital wastewater and in Ouagadougou hospital wastewater. The latter sample contained also bla_{IMP}. In addition, bla_{VIM} was detected in the Bamako gutter sample. The gene bla_{OXA-10} was detected in every sample. Five different primers binding to several gene variants of bla_{CTX-M} revealed that ESBL gene was present in all, but in Ouagadougou slaughter site and Bamako gutter samples. The qPCR array CTX-M primer that was later utilized in epicPCR revealed the presence of bla_{CTX-M} in three Burkinabe samples, namely in Nanoro and Ouagadougou hospital wastewater samples and in the WWTP influent.

Resistance-associated genes

Sequences of three ARGs, bla_{NDM}, bla_{CTX-M} and bla_{OXA} and a biocide resistance gene *qacE*Δ*1* were obtained by sequencing the fused epicPCR product, i.e. the fusion of fragments of the ARG gene and 16S rRNA gene inside one cell. Expected product sizes for the resistance gene fragments for bla_{NDM}, bla_{CTX-M}, bla_{OXA} and *qacE*Δ*1* were 66 bp, 92 bp, 71 bp and 115 bp, respectively. A blast search of the functional gene part showed that the primers amplified bla_{NDM-24}, $bla_{CTX-M-15}$, and $bla_{OXA-246}$ gene variants. In addition, the number of reads

varied greatly between the studied genes, as 1 084, 1 148, 829 013 and 1 202 200 reads were obtained from bla_{CTX-M}, bla_{NDM}, $qacE\Delta 1$ and bla_{OXA} genes, respectively.

The host range of bla_{NDM}

A total of 25 carbapenemase gene bla_{NDM} host bacteria were detected in two of the eight samples, when the presence of the gene in at least two out of three replicates was a requirement for further analysis (Figure 4). Nearly all of the host bacteria (23) were discovered from the Ouagadougou hospital wastewater. However, hosts classified to six taxa (*Dechloromonas*, *Sulfurospirillum, Methylotenera,* Proteobacteria (unclassified), Gammaproteobacteria (unclassified) and *Moraxellaceae* (unclassified)) were found in the gutter sample. Two of them, namely *Sulfurospirillum* and *Dechloromonas* were detected only in the Bamako gutter sample. All the detected bla_{NDM} gene carrying bacteria were classified to two phyla, with a vast majority to the phylum of Proteobacteria and remaining two to Epsilonbacteraeota.

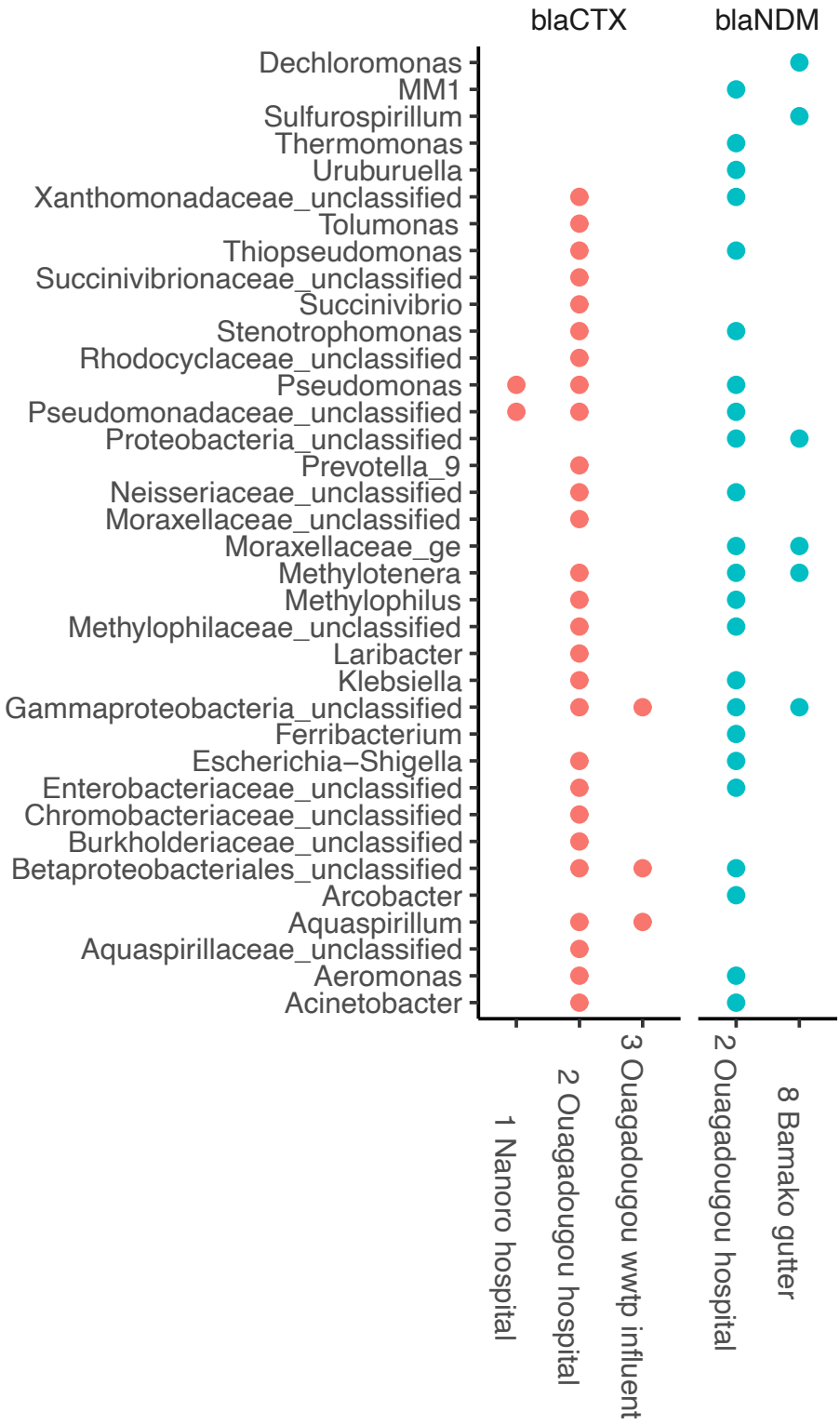


Figure 4. Bacterial host range of bla_{CTX-M} and bla_{NDM} genes. These genes were not detected by epicPCR in the other samples.

The host range of bla_{CTX-M}

The ESBL gene bla_{CTX-M} was discovered from 27 bacteria (Figure 4). The host bacteria belonged to two phyla, to Proteobacteria and Bacteroidetes, and all of them were found from the Ouagadougou hospital wastewater sample. In addition, two bla_{CTX-M} gene hosts (*Pseudomonas* and *Pseudomonadaceae* (unclassified)) were detected from the Nanoro village hospital and three hosts (Gammaproteobacteria (unclassified), Betaproteobacteriales (unclassified) and *Aquaspirillum*) from the Ouagadougou WWTP influent sample. Similar to bla_{NDM}, the gene was considered present in the sample when it was detected in at least two out of three replicates. The genes bla_{CTX-M} and bla_{NDM} shared 16 host taxa; *Acinetobacter*, *Aeromonas*, Betaproteobacteriales (unclassified), *Enterobacteriaceae* (unclassified), *Escherichia-Shigella*, Gammaproteobacteria (unclassified), *Klebsiella*, *Methylophilaceae* (unclassified), *Methylophilus*, *Methylotenera*, *Neisseriaceae* (unclassified), *Pseudomonadaceae* (unclassified), *Pseudomonas*, *Stenotrophomonas*, *Thiopseudomonas* and *Xanthomonadaceae* (unclassified).

The host range of bla_{OXA}

In comparison to the other studied β-lactamase genes bla_{CTX-M} and bla_{NDM} that were associated to less than thirty taxonomical units, bla_{OXA} presented an opposite trend, as in total, 411 taxonomical units in 20 different phyla were identified as bla_{OXA} gene hosts (Figure 5). Of all the identified bla_{OXA} hosts, 243, 86 and 38 belonged to the dominant phyla of Proteobacteria, Firmicutes and Bacteroidetes, respectively. In addition, more than third of the identified hosts were from the class Gammaproteobacteria. As a plethora of taxonomical units were revealed to be hosts of the gene in question, hosts that were only detected in all of the replicates were included in the analysis. Ouagadougou WWTP effluent sample harbored 303 taxonomical units associated with bla_{OXA}, whereas the WWTP influent harbored 185 bla_{OXA} hosts. However, in the Ouagadougou slaughter site and Bamako gutter water samples only 32 and 11 hosts were detected.

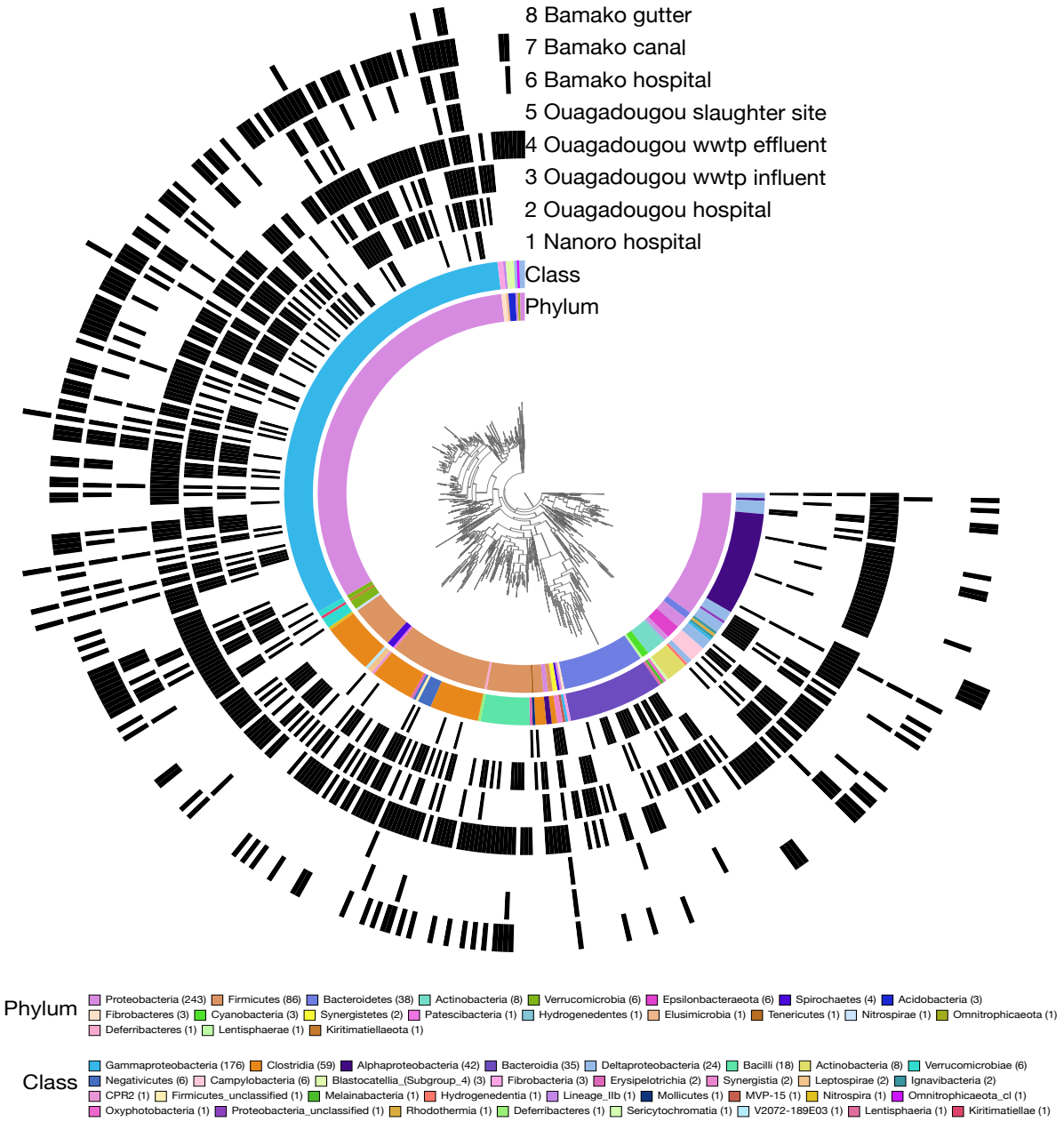


Figure 5. Bacterial host range of the bla_{OXA} gene was comprised of altogether 411 taxonomical units. In the numerical order of the samples, 95, 172, 185, 303, 32, 64, 159 and 11 taxonomical units were assigned to each sample, respectively. An occurrence of a taxonomical unit (host bacterium) in a sample is indicated with a black line. The hosts belonged to 20 phyla and 33 classes.

The host range of *qacEΔ1*

A total of 338 biocide resistance gene *qacEΔ1* hosts were detected. As with bla_{OXA}, hosts present in all three replicates were included. The genus *Rivicola* and unclassified Gammaproteobacteria were the only discovered *qacEΔ1* hosts in the Bamako gutter sample. Similarly, unclassified Gammaproteobacteria and unclassified *Rhodocyclaceae* were

associated with the *qacEΔ1* gene in the Ouagadougou slaughter site. Among the other samples, the gene was present in numerous taxa (Figure 6). The gene *qacEΔ1* was detected in 338 taxonomical units in 15 different phyla and in 24 different classes. Proteobacteria (214), Firmicutes (72) and Bacteroidetes (24) were the dominant phyla. Again, as with bla_{OXA}, Ouagadougou WWTP effluent sample contained by far the largest variety of host bacteria (294). Ouagadougou WWTP influent revealed 119 hosts, which is 60 % less than in the effluent.

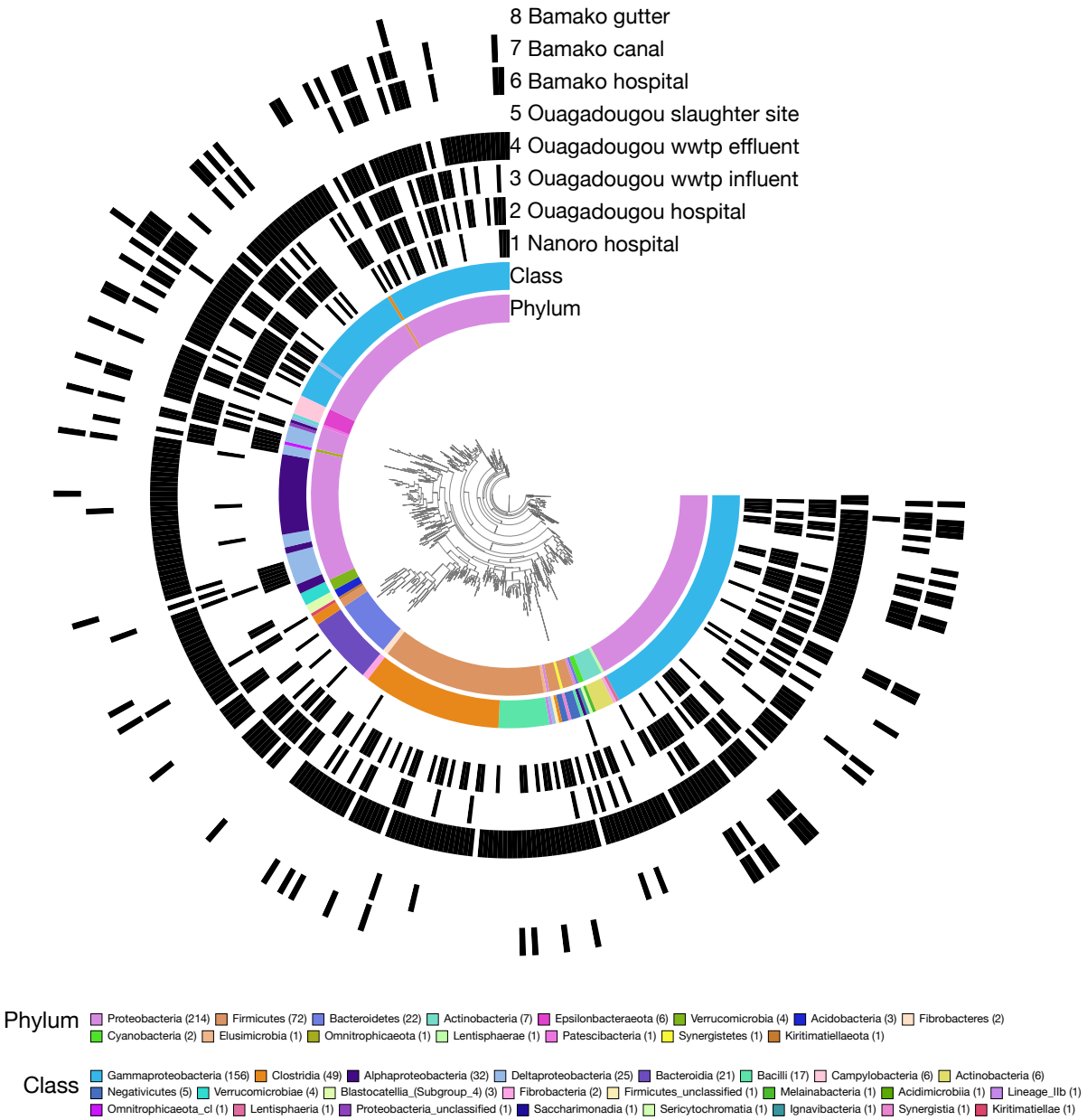


Figure 6. Bacterial host range of *qacEΔ1* gene. In total, 338 taxonomical units were identified. In the numerical order of the samples, 71, 150, 119, 294, 2, 59, 65 and 2 taxonomical units were assigned to each sample, respectively. An occurrence of a taxonomical unit (host

bacterium) in a sample is indicated with a black line. The hosts were assigned to 15 phyla and 24 classes.

Lastly, a total of 25 reads from the negative controls in epicPCR passed the same size selection filter than the actual reads and five of them contained a resistance gene fragment (2 033 445 actual epicPCR reads). Negative controls from sample Nos. 1, 2, 5 and 8 were combined before sequencing to reduce the number of negative control samples. Four $bla_{OXA-246}$ gene hosts were found in the negative control in question, namely *Prevotella_9, Prevotellaceae_Ga6A1_group, Pseudomonas* and *Rhodocyclaceae* (unclassified). Also, one read from the negative control of sample No. 4 was associated with *qacEΔ1* and it was classified as *Arcobacter.* The genes bla_{NDM} and bla_{CTX-M} were not found in the negative controls. Low number of contaminants in the negative controls indicates that their presence is possibly a result of cross-contamination coming from the samples. Thus, no bacteria detected in the negative controls were removed from the analysis.

Discussion

The purpose of this study was to describe microbial communities in eight anthropogenically impacted water samples collected from Burkina Faso and Mali. Thus, the V3-V4 region of the 16S rRNA gene was amplified and sequenced. Further, using SmartChip qPCR, the presence of ARGs and other resistance related genes among these samples was investigated and the results were utilized in choosing four genes of interest for closer examination. With epicPCR, the bacterial host range of bla_{NDM}, bla_{CTX-M}, bla_{OXA} and *qacEΔ1*genes was detected.

Microbial community composition of the West African wastewater samples

Microbial communities were analyzed with 16S rRNA gene sequencing. *Arcobacter* was shown to be the most abundant genus among the samples. In addition, the most abundant species of the dominant genus was *Arcobacter cryarophilus*. This dominance was an expected finding, as the genus is in general highly abundant in wastewaters correlating with fecal pollution levels (Fisher et al., 2014, Collado et al., 2008). Moreover, *Arcobacter* has been described as a potential zoonotic agent and an emerging food-borne pathogen (Collado et al., 2011; Patyal et al., 2011).

The species *Anaerocella delicata* dominated in the Ouagadougou slaughter site sample, which was a logical finding since originally it has been isolated from a methanogenic reactor treating cattle farm waste in Japan (Abe et al., 2012). Also other amplicon sequence variants (ASVs) were classified as anaerocellas, however, their sequence resolution was not sufficient for species classification. This is interesting, as the species is to date the only classified species of the novel genus *Anaerocella* (Abe et al., 2012).

The bacterial abundances varied greatly among the samples. In Ouagadougou WWTP effluent, *Thiolamprovum* comprised the vast majority of the ASVs. However, one of the effluent replicates had substantially less *Thiolamprovum* and more *Macellibacteroides* than the other two replicates, and this difference in microbial community composition was also observed with the Shannon diversity index describing the α-diversity of the sample. The replicate samples were taken a few meters a part and it is possible that the sampling bottle has for example touched the sediment layer underneath the shallow water, however, it is difficult to determine the reason for the difference without more samples. The genus *Thiolamprovum* is a purple sulphur bacterium and its high abundance probably explains the red color of the Ouagadougou WWTP effluent. Overload of organic material in WWTP systems leads to process failure, which the 'red-water' phenomenon is a sign of (Belila et al., 2013). A periodical change in the effluent color is a result of exponentially growing purple sulphur bacteria that thrive in anoxic conditions caused by malfunction of the plant (Belila et al., 2013). *Macellibacteroides* is an obligately anaerobic bacterium, that has been isolated from abattoir wastewater in Tunisia (Jabari et al., 2012), hence its abundance in the effluent sample is rational.
In the Ouagadougou hospital wastewater and in the Bamako canal and gutter water samples, the distribution of microbial community compositions of the replicates was notably less skewed than in the other samples. A relatively even distribution in bacterial abundances suggests species richness, structural complexity and co-occurrence of species (Verberk, 2011). The β-diversity of the samples indicated a great variability between the samples in terms of species abundances and prevalence.

Resistance genes and their host bacteria

SmartChip qPCR array analysis revealed high prevalence of resistance genes in the samples (Supplementary Figure 1). The presence of bla_{VIM}, bla_{IMP} and bla_{NDM}, carbapenemase genes that belong to the clinically most relevant subclass B1 of β-lactamases (Mojica, Bonomo and

Fast, 2016), was detected in the Ouagadougou hospital wastewater. Moreover, the qPCR results were mainly in concordance with the data obtained by epicPCR, as bla_{CTX-M}, *qacEΔ1* and bla_{OXA} genes were detected in the same samples. However, the occurrence of bla_{NDM} gene in two samples was different than in the epicPCR results, as with qPCR, the gene was detected in the sample No. 1 (Nanoro hospital wastewater) but not in the sample No. 8 (Bamako gutter water). Also, with qPCR, fewer resistance genes were discovered in Ouagadougou WWTP effluent than in the influent.

Using blastn and the resfinder database (Zankari et al., 2012), the ARG sequences obtained with epicPCR were discovered to be $bla_{OXA-246}$, bla_{NDM-24} and $bla_{CTX-M-15}$. Thus, different bla_{OXA} and bla_{NDM} gene variants were identified than were expected with the used primers. The variant $bla_{OXA-246}$ has a mutation compared to its progenitor bla_{OXA-10}, however, this mutation does not affect the kinetic properties of the enzyme that is encoded by the gene $bla_{OXA-246}$ (Qing et al., 2014). Also, two bla_{OXA-10} variants that are rarely mentioned in the literature were detected, namely bla_{OXA-56} and $bla_{OXA-454}$. Further, the gene variant bla_{NDM-24} is one of the over twenty-six bla_{NDM-1} variants currently identified (Liu et al., 2019). The mutation to shape the enzyme that is encoded by the gene bla_{NDM-24} induces the enzyme activity but decreases the protein stability to some extent (Liu et al., 2019). Nevertheless, bacteria expressing this carbapenemase gene variant are able to hydrolyze the same β-lactams than with the first discovered bla_{NDM-1} (Wu et al., 2019). It is noteworthy that even if the identity of these genes is known, their functionality and expression in these samples remains unknown, since only a fragment of the resistance genes was amplified.

The host bacteria of bla_{NDM}, bla_{CTX-M}, bla_{OXA} and *qacEΔ1* were identified with epicPCR. Known New Delhi metallo-β-lactamase producers, namely *Escherichia* and *Klebsiella*, and *Acinetobacter* and *Pseudomonas* that are frequently associated with the carbapenemase gene (Wu et al., 2019), were discovered as bla_{NDM} hosts also in the Ouagadougou hospital wastewater. Interestingly, none of the mentioned genera were associated with the bla_{NDM} gene in the Bamako gutter sample, even though, according to the 16S rRNA gene sequencing, *Acinetobacter* was the third most abundant genus in that sample. In addition, all these genera are known to cause clinically relevant infections (Alvarez-Uria et al., 2018; Pendleton, Gorman and Gilmore, 2013). One could speculate that bacteria carry the ARG in the hospital environment but not in the street gutter water due to existing antibiotic residues in the hospital and its wastewater. Nonetheless, these potential pathogenic hosts of bla_{NDM} do pose a greater

risk to people and to the community than nonpathogenic environmental hosts of bla_{NDM}. Interestingly, potentially pathogenic *Arcobacter* along with *Methylotenera*, *MM1*, *Methylophilus* and *Dechloromonas* were new discoveries as bla_{NDM} hosts. The same genera were recently found as bla_{NDM} gene hosts in another study, where the host range of bla_{NDM} among other resistance genes was investigated in a Finnish WWTP (Pezzutto, 2019).

The carbapenemase gene bla_{NDM} and the ESBL gene bla_{CTX-M} had a rather similar host range. The genera *Acinetobacter*, *Pseudomonas, Prevotella_9* and *Tolumonas* that were detected to be abundant with 16S rRNA gene sequencing, were also detected to be bla_{CTX-M} gene hosts with epicPCR. The genus *Aquaspirillum* was revealed to be an ESBL gene host in the Ouagadougou hospital wastewater and WWTP influent samples. *Aquaspirillum* spp consists of aerobic freshwater bacteria that are also common in WWTPs. However, to the extent of my knowledge, it has not been described as a bla_{CTX-M} carrier before. A strict anaerobe *Prevotella* is a potential fecal pollution indicator (Okabe et al., 2007). The genus also comprises of numerous pathogenic species and it has been reported as an ESBL-producer that can potentially protect *Pseudomonas aeruginosa* from ceftazidime treatment in cystic fibrosis (Sherrard et al., 2016). This symbiotic relationship of two antimicrobial resistant pathogens is a worrying reminder how complex interactions microbes can have in the battle against antimicrobials.

The heavy burden of ARGs in West Africa

WHO has listed carbapenem-resistant *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* as critical priority pathogens (World Health Organization, 2017). Worryingly, *Acinetobacter, Pseudomonas* and members of the family *Enterobacteriaceae*, namely *Escherichia*/*Shigella* and *Klebsiella,* were all detected as bla_{NDM} gene hosts in hospital wastewater in Ouagadougou. The host range of the other studied genes also included these potential pathogens. Other hosts carrying all the studied genes were *Aeromonas, Betaproteobacteriales* (unclassified), *Enterobacteriaceae* (unclassified), Gammaproteobacteria (unclassified), *Methylophilaceae* (unclassified), *Methylophilus, Methylotenera, Neisseriaceae* (unclassified), *Pseudomonadaceae* (unclassified), *Stenotrophomonas, Thiopseudomonas* and *Xanthomonadaceae* (unclassified). Interestingly, they were always found at least in the Ouagadougou hospital wastewater sample. Water samples collected from the other two hospitals, in Nanoro and in Bamako, differed greatly from the Ouagadougou hospital water. With epicPCR, no carbapenemase producers were detected

in the Nanoro and Bamako hospital samples. In addition, only two taxonomical units were associated with the bla_{CTX-M} gene in the Nanoro hospital water and no bla_{CTX-M} carriers were detected in the Bamako hospital wastewater. The differences in the hospital wastewater samples could be explained by the capacity and locations of the hospitals, and possible differences in antibiotic use. Also, the way to treat and store the wastewaters, especially in the Nanoro hospital where the wastewater is collected in the septic tanks, is likely to have an effect to the results. However, as only one sample with replicates per hospital was taken, the significance of these differences cannot be confirmed. Thus, more sampling and research is needed.

The twenty most abundant genera detected with 16S rRNA gene sequencing were discovered with epicPCR to be hosts of either *qacEΔ1,* bla_{OXA} or as in most of the cases, both of the genes. The two genes were consistently found in every sample, with qPCR and epicPCR. The broad host range of $bla_{OXA-246}$ and *qacEΔ1* genes suggests heavy ARG burden in West Africa. Actually, from a geographical point of view, tropical climate can play a major role in contributing to AMR persistence (Pärnänen et al., 2019). Indeed, a growing body of evidence suggests that globally Africa appears to have the highest prevalence of AMR (Hendriksen et al., 2019). The high number of bacteria harboring ARGs in wastewaters that are further utilized as irrigation waters in urban gardening, can significantly contribute to ARG dissemination into the environment and eventually into the community. Also, a recently published study found that household cockroaches in Ghana (neighboring country of Burkina Faso) carried CTX-M-15-, OXA-48- and NDM-1-producing bacteria (Obeng-Nkrumah et al., 2019). The finding highlights the multitudinous possibilities of occurrence and flow of ARGs.

The bacterial host ranges of bla_{OXA} and *qacEΔ1* genes that were detected in the two WWTP samples (influent and effluent), were compared to evaluate possible horizontal gene transfer occurrences in the WWTP. The microbial community information obtained with 16S rRNA gene sequencing and the host bacteria results obtained with epicPCR were combined. A few genera that carried both bla_{OXA} and *qacEΔ1* genes were found in the Ouagadougou WWTP effluent but not in the influent (Supplementary Figure 2.). However, these genera, namely *Acetobacterium, Akkermansia, Bifidobacterium, Smithella* and *Terrisporobacter*, were detected with 16S rRNA gene sequencing in the microbial communities of both samples. On the other hand, genera that were present in both samples but were carrying bla_{OXA} only in the

influent were *Alistipes, Azonexus, Erysipelothrix, Macellibacteroides, Petrimonas, Prevotella_2* and *Saccharofermentans.* From those, *Macellibacteroides* was the second most abundant genus in the effluent. In addition, genera carrying bla_{OXA} only in the effluent but were present in both samples were *Flavobacterium, Lactobacillus, Moraxella, Parasutterella, Pleomorphomonas, Sphingobacterium* and *Thiothrix.* Thus, based on these two samples, possible horizontal gene transfer was observed in the Ouagadougou WWTP. However, the microbial community composition was remarkably different between the samples, and many of the numerous genera detected as resistance gene carriers with epicPCR were not detected in the microbial communities with 16S rRNA gene sequencing. The notable difference in host number between the two samples is suspicious and it raises the question whether the WWTP effluent runoff has other sources of water than the water originating from the WWTP. After all, the effluent was collected from a runoff next to the WWTP rather than from a specific discard pipe. Hence, in the light of these notes, evidence for horizontal gene transfer remains scarce.

Limitations in the study plan and its implementation

Sample size was fairly limited and without further time points, it is not possible to estimate whether these findings present sporadic or constant situation in the studied environments. To overcome this limitation, a study plan with several time points and sampling also during rainy season should be designed. Also, more harmonized approach regarding sampling and sample processing would have been beneficial. For example, now sample processing was carried out in three laboratories, in Burkina Faso, Mali and Finland. Moreover, although SmartChip qPCR and epicPCR results are mainly concordant, two differences were observed. Regarding the bla_{NDM} gene, epicPCR did not give positive results on one sample (Nanoro hospital wastewater) that qPCR did. As a novel method, epicPCR has some room for optimization, and detection limit is still somewhat unknown, which can partly explain the differences. Furthermore, the reproducibility of epicPCR results remains a question. Due to the encapsulation of random cells before fusion PCR, it is not guaranteed that exactly the same hosts are found every time, and thus deviating results do not necessarily invalidate the previously obtained results. Also, due to the sequencing technology used in epicPCR, the length of the amplified 16S rRNA gene fragment is too short to reveal the species of host bacteria. Knowing the species that carry the resistance gene would increase the significance of these findings.

Lastly, it is noteworthy to say that the occurrence of the gene bla_{CTX-M} (qPCR) and number of bla_{CTX-M} hosts (epicPCR) was unexpectedly low as according to literature, CTX-M ESBLs are widespread and it is estimated that 110 million Africans are ESBL-carriers (Cantón and Coque, 2006; Semret and Haraoui, 2019). In addition, bla_{CTX-M} has been reported as the dominant ESBL gene globally (Bevan, Jones and Hawkey, 2017). Moreover, four different primers of CTX-M were tested for epicPCR, and one was selected. Lack of an optimal and universal "ESBL primer" or even universal CTX-M primer can partly explain the limited host range of ESBL in this study. In addition, more studies on ESBL prevalence in the environment is needed.

The significance of this study to the battle against AMR in West Africa

WHO has listed 5 strategic objectives in a global action plan to fight against AMR (World Health Organization, 2015). One of the objectives is to strengthen knowledge of AMR through surveillance and research. Another is to reduce the incidence of infection which could be reached with improving sanitation and health in developing countries. This study offers a glimpse of the current AMR situation in Burkina Faso and Mali. Surveillance of hospital wastewater can offer an overview of the level of resistance in clinical settings. In WWTPs, effluent disinfection processes can be introduced to reduce the occurrence of waterborne pathogenic microorganisms in effluent. Decreasing the number of pathogens in urban wastewaters would be of great importance knowing that people in African cities are often exposed to WWTP effluents and other wastewaters. Emphasis on the need to provide information for stakeholders so that actions against AMR can be implemented along with effective routine surveillance is required. These actions could eventually limit the global burden of AMR.

Culture-independent approaches, such as epicPCR and qPCR that were used in this study, are providing new information regarding AMR, complementing the more conventional microbiological methods, such as culturing. This combination of the methods (16S rRNA gene sequencing, SmartChip qPCR and epicPCR) provided both comprehensive and detailed findings, that could not have been reached with culturing. Culture-independent approaches enable the detection of both abundant and rare taxa. Hence, more frequent introduction of the PCR-based methods for research purposes in low income countries would be useful. Moreover, this study emphasizes the need for One Health research. The emergence of new zoonotic

diseases and antimicrobial drug-resistant microbes is enhanced by human actions, such as urbanization, international travel and intensive industrial livestock production (Mackenzie and Jeggo, 2019). In this study, many multiresistant and potentially pathogenic bacteria (e.g. *Acinetobacter, Pseudomonas, Enterobacteriaceae*) in human impacted environments were identified. This research provides a basis for deciphering the role of these pathogens in the infectious disease burden in Africa. Both national and global research collaboration is needed to design concrete actions to diminish this burden.

Acknowledgements

I wish to acknowledge the CSC—IT Center for Science, Finland, for providing the computational resources and the Academy of Finland for financial support. I thank my supervisors Kaisa Haukka, Antti Karkman and PI Marko Virta for the guidance and valuable feedback I have received during all the stages of my Master's thesis. I also thank the sampling teams in Burkina Faso and Mali. I had many amazing and memorable moments in Africa with you. Finally, I am grateful to my lab mates for the mental support and for the great team spirit we have had and especially thanks to Sirja Viitamäki for her kind advice in data analysis in general.

References

Abe K, Ueki A, Ohtaki Y, Kaku N, Watanabe K, Ueki K. (2012). *Anaerocella delicata* gen. nov., sp. nov., a strictly anaerobic bacterium in the phylum Bacteroidetes isolated from a methanogenic reactor of cattle farms. *The Journal of General and Applied Microbiology*, doi: 10.2323/jgam.58.405.

Altschul SF, Gish W, Miller W, Meyers EW, Lipman, DJ. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215:403-410, doi: 10.1016/S0022-2836(05)80360-2

Alvarez-Uria G, Gandra S, Mandal S, Laxminarayan R. (2018). Global forecast of antimicrobial resistance in invasive isolates of *Escherichia coli* and *Klebsiella pneumoniae*. *International Journal of Infectious Diseases*, 68:50–53, doi: 10.1016/j.ijid.2018.01.011.

Andrews S. (2011). FastQC: A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Last accessed on 20th of April 2020].

Baker GC, Smith JJ, Cowan DA. (2003). Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods*, 55:541-555, doi: 10.1016/j.mimet.2003.08.009.

Barlow M, Hall BG. (2002). Phylogenetic analysis shows that the OXA b-lactamase genes have been on plasmids for millions of years. *Journal of Molecular Evolution*, 55:314-321, doi: 10.1007/s00239-002-2328-y.

Belila A, Abbas B, Fazaa I, Saidi N, Snoussi M, Hassen A, Muyzer G. (2013). Sulfur bacteria in wastewater stabilization ponds periodically affected by the 'red-water' phenomenon. *Applied Microbiology and Biotechnology*, 97:379–394, doi: 10.1007/s00253-012-3931-5.

Bernabé KJ, Langendorf C, Ford N, Ronat J-B, Murphy RA. (2017). Antimicrobial resistance in West Africa: a systematic review and meta-analysis. *International Journal of Antimicrobial Agents*, 50:629-639, doi: 10.1016/j.ijantimicag.2017.07.002.

Bevan ER, Jones AM, Hawkey PM. (2017). Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. *Journal of Antimicrobial Chemotherapy*, 72:2145–2155, doi: 10.1093/jac/dkx146.

Bougnom BP, Zongo C, McNally A, Ricci V, Etoa FX, Thiele-Bruhn S, and Piddock LJV. (2019). Wastewater used for urban agriculture in West Africa as a reservoir for antibacterial resistance dissemination. *Environmental Research*, 168:14-24, doi: 10.1016/j.envres.2018.09.022.

Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13:581- 583, doi: 10.1038/nmeth.3869

Cantón R, Coque TM. (2006). The CTX-M β-lactamase pandemic. *Current Opinion in Microbiology*, 9:466-475, doi: 10.1016/j.mib.2006.08.011.

Collado L, Inza I, Guarro J, Figueras MJ. (2008). Presence of *Arcobacter* spp. in environmental waters correlates with high levels of fecal pollution. *Environmental Microbiology*, 10:1635-1640, doi: 10.1111/j.1462-2920.2007.01555.x.

Collado L, Figueras MJ. (2011). Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clinical Microbiology Reviews*, 24:174-192, doi: 10.1128/CMR.00034-10.

Dolejska M, Frolkova P, Florek M, Jamborova I, Purgertova M, Kutilova I, *et al*. (2011). CTX-M-15-producing *Escherichia coli* clone B2- O25b-ST131 and *Klebsiella* spp. isolates in municipal wastewater treatment plant effluents. *Journal of Antimicrobial Chemotherapy*, 66:2784–2790, doi: 10.1093/jac/dkr363.

Dowle M, Srinivasan A. (2019). data.table: Extension of `data.frame`. R package version 1.12.8. Available online at: https://CRAN.R-project.org/package=data.table [Last accessed on 20th of April 2020].

Drechsel P, Keraita B, Amoah P, Abaidoo RC, Raschid-Sally L, Bahri A. (2008). Reducing health risks from wastewater use in urban and peri-urban sub-Saharan Africa: Applying the

2006 WHO guidelines. *Water Science and Technology*, 57:1461-1466, doi: 10.2166/wst.2008.245.

Eckert C, Gautier V, Arlet G. (2006). DNA sequence analysis of the genetic environment of various *blaCTX-M* genes. *Journal of Antimicrobial Chemotherapy*, 57:14–23, doi: 10.1093/jac/dki398.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32:1792-1797, doi: 10.1093/nar/gkh340.

Edwards U, Rogall T, Blocker H, Emde M, Böttger EC. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research*, 17:7843–53, doi: 10.1093/nar/17.19.7843.

Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ,* 3:e1319, doi: 10.7717/peerj.1319.

Fisher JC, Levican A, Figueras MJ, McLellan SL. (2014). Population dynamics and ecology of *Arcobacter* in sewage. *Frontiers in Microbiology*, 5:525, doi: 10.3389/fmicb.2014.00525.

Guo J, Li J, Chen H, Bond PL, Yuan Z. (2017). Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Research,* 123:468-478, doi: 10.1016/j.watres.2017.07.002.

Hegstad K, Langsrud S, Lunestad BT, Scheie AA, Sunde M, Yazdankhah SP. (2010). Does the wide use of quaternary ammonium compounds enhance the selection and spread of antimicrobial resistance and thus threaten our health? *Microbial Drug Resistance*, 16:91-104, doi: 10.1089/mdr.2009.0120.

Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, Lukjancenko O, *et al.* (2019). Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature Communications*, 10:1124, doi: https://doi.org/10.1038/s41467-019-08853-3.

Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME Journal*, 5:1571–1579, doi: 10.1038/ismej.2011.41.

Hultman J, Tamminen M, Pärnänen K, Cairns J, Karkman A, Virta M. (2018). Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent, *FEMS Microbiology Ecology*, 94:1-10, doi: 10.1093/femsec/fiy038.

Jabari L, Gannoun H, Cayol J-L, Hedi A, Sakamoto M, Falsen E, *et al.* (2012). *Macellibacteroides fermentans* gen. nov., sp. nov., a member of the family *Porphyromonadaceae* isolated from an upflow anaerobic filter treating abattoir wastewaters. *International Journal of Systematic and Evolutionary Microbiology,* doi: https://doi.org/10.1099/ijs.0.032508-0.

Karkman A, Johnson TA, Lyra C, Stedtfeld RD, Tamminen M, Tiedje JM, Virta M. (2016). High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS Microbiology Ecology*, doi: 10.1093/femsec/fiw014.

Kolde R. (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12. Available online at: https://CRAN.R-project.org/package=pheatmap [Last accessed on 20th of April 2020].

Liu Z, Piccirilli A, Liu D, Wang Y, Shen J. (2019). Deciphering the role of V88L substitution in NDM-24 metallo-β-lactamase. *Catalysts*, doi: 10.3390/catal9090744.

Mackenzie JS, Jeggo M. (2019). The One Health approach-Why is it so important?. *Tropical Medicine and Infectious Disease*, 4:88, doi: https://doi.org/10.3390/tropicalmed4020088.

Mafuta C, Barnes R, Plummer L, Westerveld L. (2018). Sanitation and wastewater in Africa. GRID-Arendal. Available online at: https://www.arcgis.com/apps/Cascade/index.html?appid=caf411c40c3442b782406de631bddb2f [Last accessed on 20th of April 2020].

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, doi: https://doi.org/10.14806/ej.17.1.200.

McMurdie PJ, Holmes S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *Public Library of Science* (*PLoS) One*, 8*:*e61217, doi: https://doi.org/10.1371/journal.pone.0061217.

Mojica MF, Bonomo RA, Fast W. (2016). B1-metallo-β-lactamases: Where do we stand? *Current Drug Targets*, 17:1029–1050, doi: 10.2174/1389450116666151001105622.

Müller K, Wickham H. (2019). tibble: Simple Data Frames. R package version 2.1.3. Available online at: https://CRAN.R-project.org/package=tibble [Last accessed on 20th of April 2020].

Naas T, Nordmann P. (1999). OXA-type β -lactamases. In: Current pharmaceutical design, 5:865-79. Bentham Science Publishers. ISSN:1381-6128.

Obeng-Nkrumah N, Labi A, Blankson H, Awuah-Mensah G, Oduro-Mensah D, Anum J, *et al.* (2019). Household cockroaches carry CTX-M-15-, OXA-48- and NDM-1-producing enterobacteria, and share beta-lactam resistance determinants with humans. *BMC Microbiology,* 19:272, doi: https://doi.org/10.1186/s12866-019-1629-x.

Okabe S, Okayama N, Savichtcheva O, Ito T. (2007). Quantification of host-specific *Bacteroides–Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Applied Microbiology and Biotechnology,* 74:890–901, doi: 10.1007/s00253-006-0714-x.

Pagès H, Aboyoun P, Gentleman R, DebRoy S. (2019). Biostrings: Efficient manipulation of biological strings. R package version 2.54.0, doi: 10.18129/B9.bioc.Biostrings.

Patyal A, Rathore RS, Mohan HV, Dhama K, Kumar A. (2011). Prevalence of *Arcobacter* spp. in humans, animals and foods of animal origin including sea food from India. *Transboundary and Emerging Diseases*, 58:402-410, doi: 10.1111/j.1865-1682.2011.01221.x.

Pendleton J, Gorman SP, Gilmore BF. (2013). Clinical relevance of the ESKAPE pathogens. *Expert Review of Anti-infective Therapy*, 11:297-308, doi: 10.1586/eri.13.12.

Pezzutto D. (2019). Host range of antibiotic resistance genes in influent, effluent and dried sludge of a wastewater treatment plant. Master's thesis, University of Helsinki, Helsinki. Available online at: https://helda.helsinki.fi/bitstream/handle/10138/312361/Denise_Pezzutto_Pro_gradu_2019.pdf?sequence=2&isAllowed=y [Last accessed on 1st of May 2020].

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments (FastTree 2). *Public Library of Science* (*PLoS) One,* 5:e9490, doi: https://doi.org/10.1371/journal.pone.0009490.

Pärnänen K, Narciso-Da-Rocha C, Kneis D, Berendonk T, Cacace D, Do T, *et al.* (2019). Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Science Advances, 5:*eaau9124, doi: 10.1126/sciadv.aau9124.

Qing Y, Cao KY, Fang Z L, Huang YM, Zhang XF, Tian GB, *et al*. (2014). Outbreak of PER-1 and diversity of β-lactamases among ceftazidime-resistant *Pseudomonas aeruginosa* clinical isolates. *Journal of Medical Microbiology*, 63:386–392, doi: 10.1099/jmm.0.069427-0.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al*. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research,* 41:D590, doi: 10.1093/nar/gks1219.

Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, *et al*. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. *Science of the Total Envinronment*, 447:345-60, doi: 10.1016/j.scitotenv.2013.01.032.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. (2016). VSEARCH: a versatile open source tool for metagenomics. *Peer-reviewed journal* (*PeerJ)*, 4e2584, doi: 10.7717/peerj.2584.

Romão C, Miranda C, Silva J, Clementino M, de Filippis I, Asensi M. (2011). Presence of *qacE* delta 1 gene and susceptibility to a hospital biocide in clinical isolates of *Pseudomonas aeruginosa* resistant to antibiotics. *Current Microbiology*, 63:16-21, doi : 10.1007/s00284-011-9934-0.

RStudio Team. (2019). RStudio: Integrated development for R. RStudio, Inc., Boston, MA Available online at: http://www.rstudio.com/ [Last accessed on 1st of May 2020].

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et a*l. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75:7537-7541, doi: 10.1128/AEM.01541-09.

Semret M, Haraoui L-P. (2019). Antimicrobial resistance in the tropics. *Infectious Disease Clinics of North America*, 33:231-245, doi: 10.1016/j.idc.2018.10.009.

Sherrard LJ, McGrath SJ, McIlreavey L, Hatch J, Wolfgang MC, Muhlebach MS, *et al*. (2016). Production of extended-spectrum β-lactamases and the potential indirect pathogenic role of *Prevotella* isolates from the cystic fibrosis respiratory microbiota. *International Journal of Antimicrobial Agents*, 47:140-145, doi: 10.1016/j.ijantimicag.2015.12.004.

Spencer SJ, Tamminen MV, Preheim SP, Guo MT, Briggs AW, Brit IL, *et al*. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME Journal*, 10:427–436, doi: 10.1038/ismej.2015.124.

Stedtfeld RD, Guo X, Stedtfeld TM, Sheng H, Williams MR, Hauschild K, *et al*. (2018). Primer set 2.0 for highly parallel qPCR array targeting antibiotic resistance genes and mobile genetic elements. *FEMS Microbiology Ecology*, 94:fiy130, doi: 10.1093/femsec/fiy130.

Verberk W. (2011) Explaining general patterns in species abundance and distributions. *Nature Education Knowledge,* 3:38.

Viens AM, Littmann J. (2015). Is antimicrobial resistance a slowly emerging disaster? *Public Health Ethics*, 8: 255–265, doi: 10.1093/phe/phv015.

Wickham H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York, ISBN 978-3-319-24277-4, doi: 10.1007/978-3-319-24277-4.

Wickham H, Averick M, Bryan J, Chang W, D'Agostino McGowan L, François R, *et al*. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4:1686, doi: https://doi.org/10.21105/joss.01686.

Wickham H, François R, Henry L, Müller K. (2020). dplyr: A grammar of data manipulation. R package version 0.8.5. Available online at: https://CRAN.R-project.org/package=dplyr [Last accessed on 20th of April 2020].

World Health Organization. (2015). Global action plan on antimicrobial resistance. World Health Organization. Available online at: https://apps.who.int/iris/handle/10665/193736 [Last accessed on 20th of April 2020].

World Health Organization. (2017). Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug resistant bacterial infections, including tuberculosis. World Health Organization. Available online https://apps.who.int/iris/handle/10665/311820 [Last accessed on 20th of April 2020].

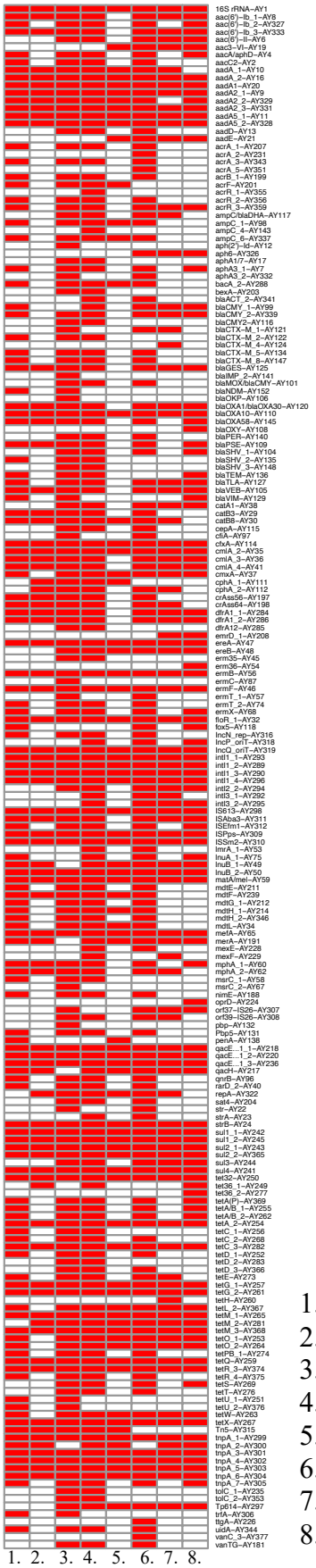
Wu W, Feng Y, Tang G, Qiao F, McNally A, Zong Z. (2019). NDM metallo-β-lactamases and their bacterial producers in health care settings. *Clinical Microbiology Reviews,* 32:e00115-18, doi: 10.1128/CMR.00115-18.

Yoon SH, Park YK, Kim JF. (2015). PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. *Nucleic Acids Research*, 43:D624-D630. *qacEΔ1* gene sequence obtained from: http://paidb.re.kr/view_gene.php?gna=53&gma=NC_017659&m=g. [Last accessed on 25th of April 2020].

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, *et al*. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, 67:2640-2644, doi: 10.1093/jac/dks261.

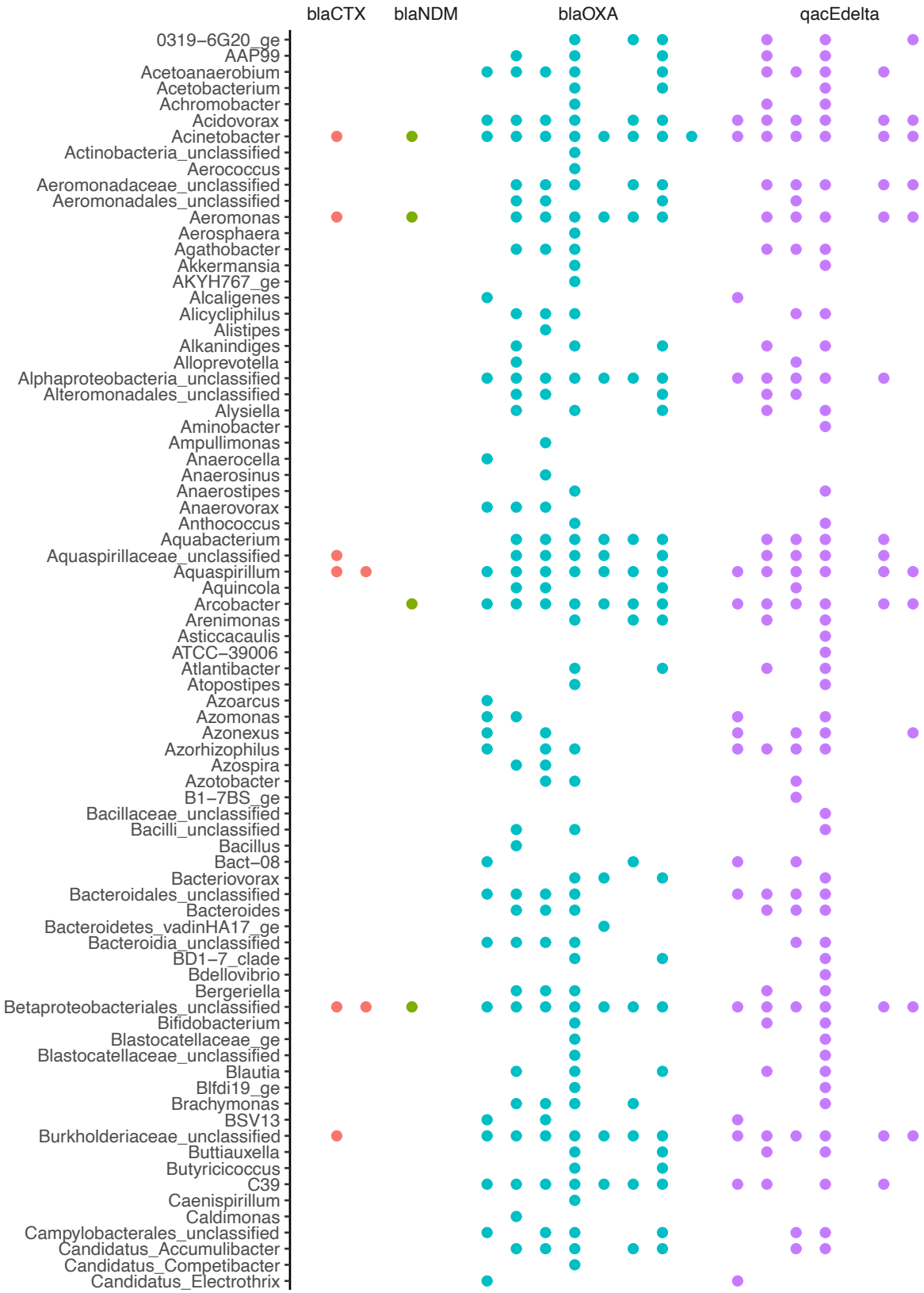
Zhang J, Kobert K, Flouri T, Stamatakis A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30:614-620, doi: 10.1093/bioinformatics/btt593.

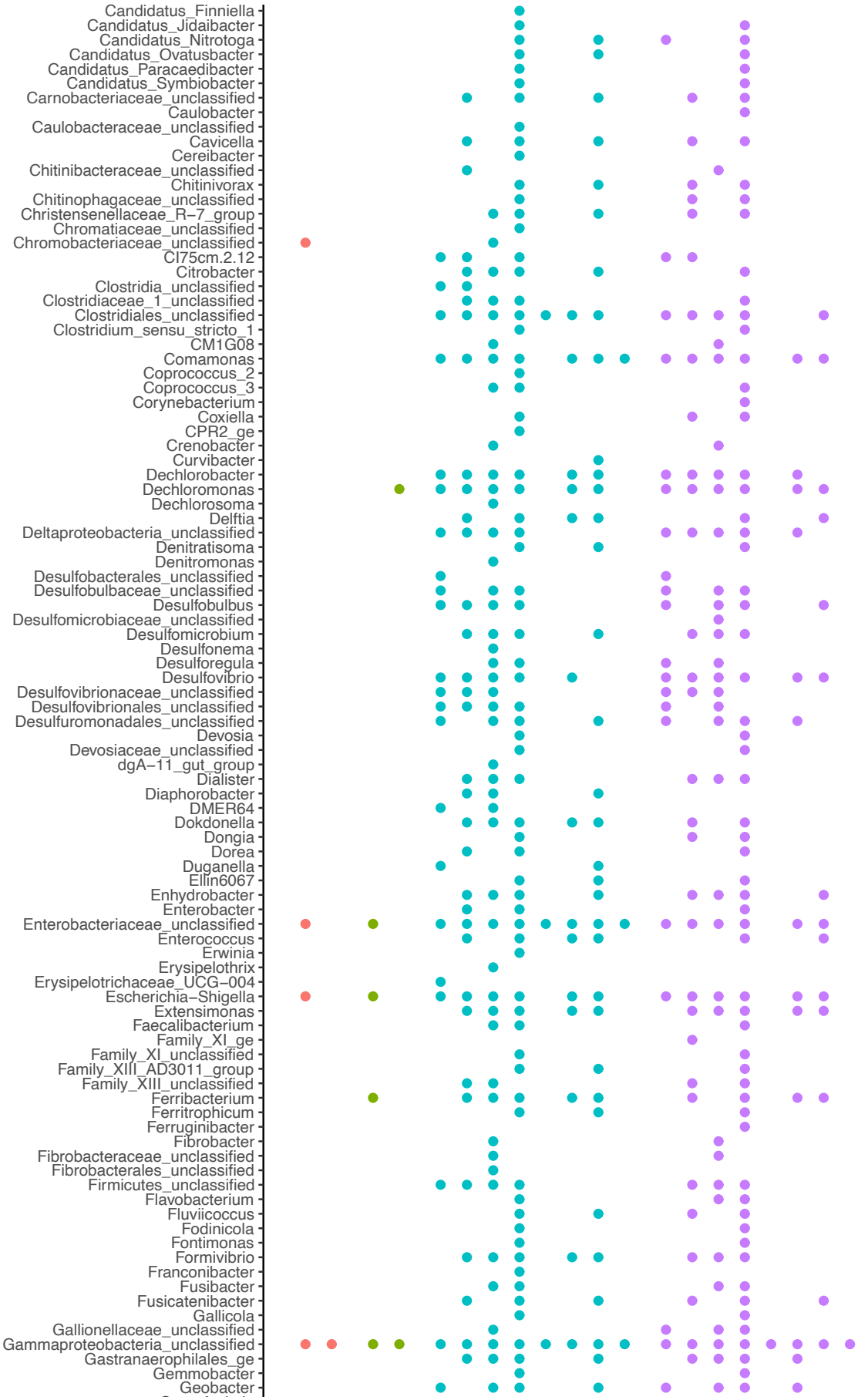
Supplementary data

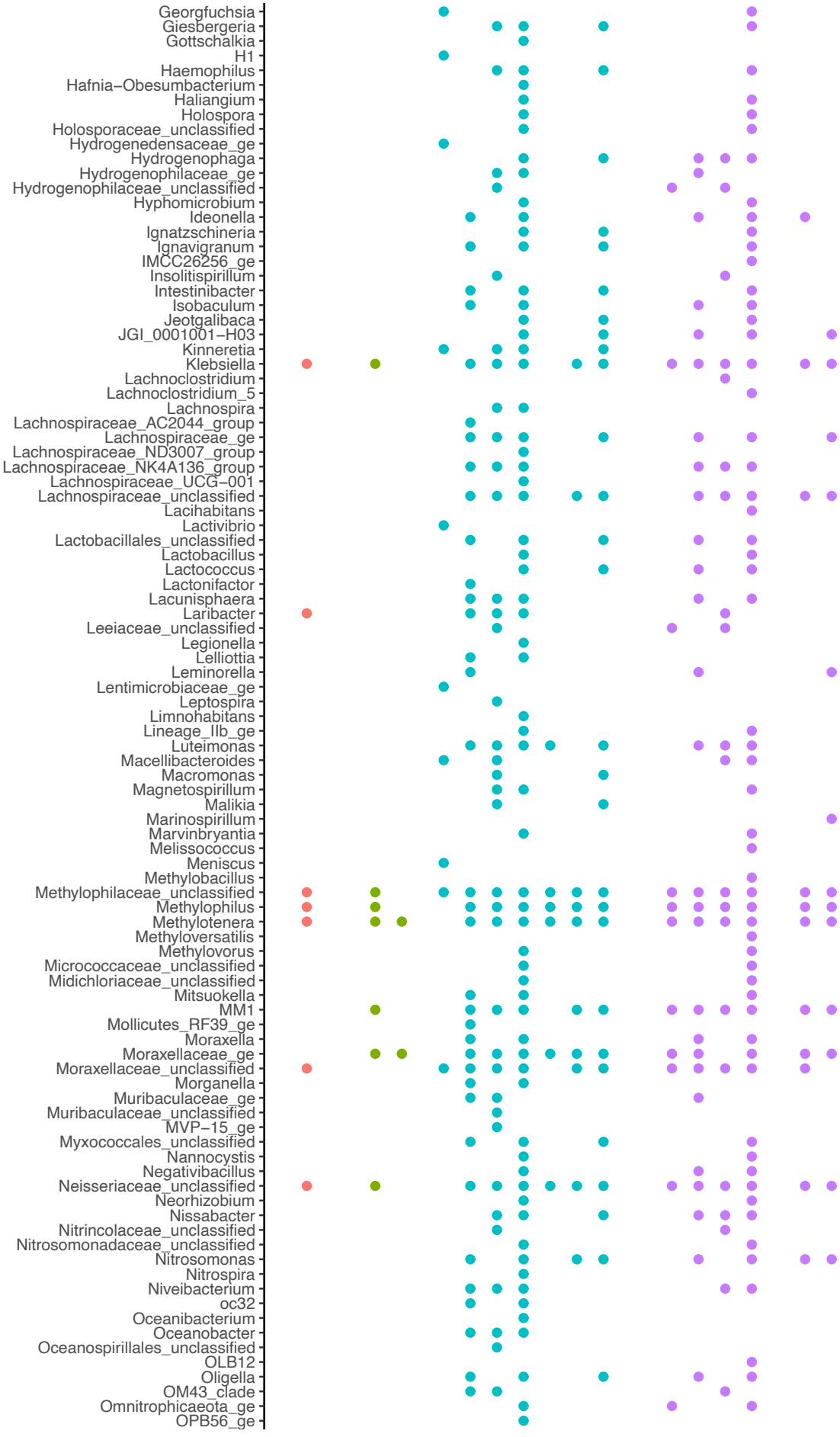


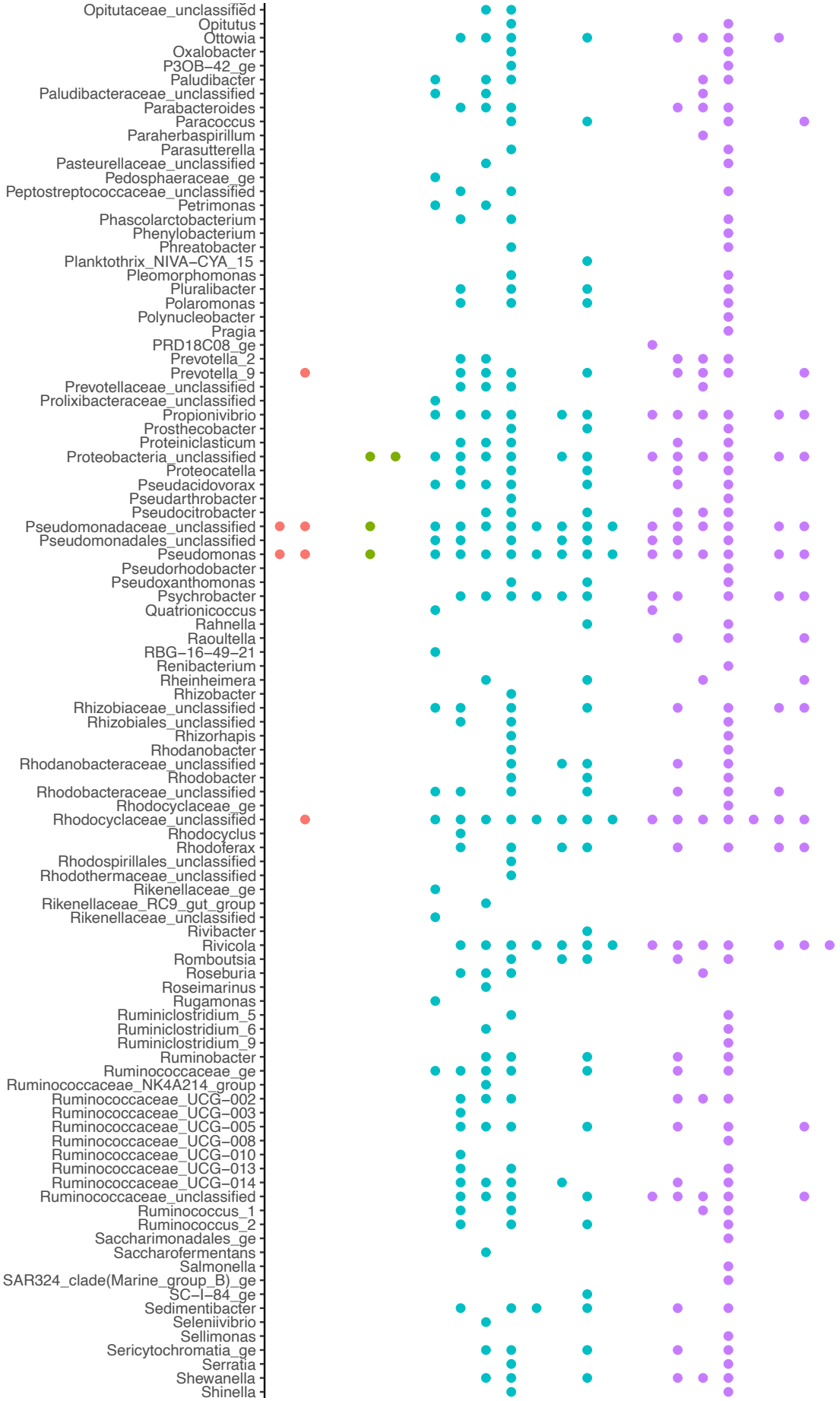
1. Nanoro hospital water
2. Ouagadougou WWTP effluent
3. Ouagadougou hospital water
4. Ouagadougou WWTP influent
5. Ouagadougou slaughter site
6. Bamako hospital water
7. Bamako canal water
8. Bamako gutter water

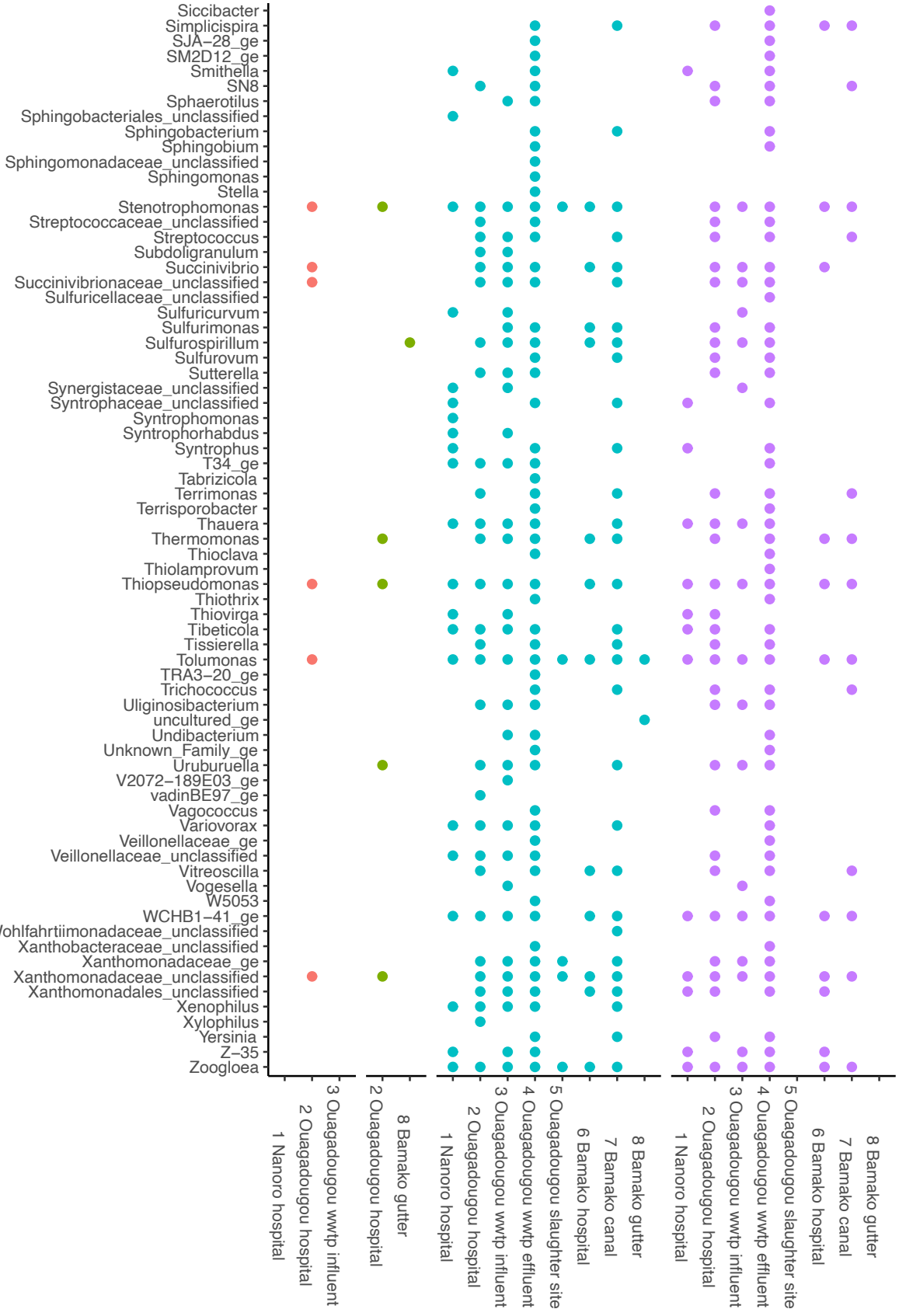
Supplementary Figure 1. Heatmap of SmartChip qPCR results. A total of 202 genes were detected. Red color indicates presence of the gene.


blaCTX
blaNDM
blaOXA
qacEdelta
0319−6G20_ge
AAP99
Acetoanaerobium
Acetobacterium
Achromobacter
Acidovorax
Acinetobacter
Actinobacteria_unclassified
Aerococcus
Aeromonadaceae_unclassified
Aeromonadales_unclassified
Aeromonas
Aerosphaera
Agathobacter
Akkermansia
AKYH767_ge
Alcaligenes
Alicycliphilus
Alistipes
Alkanindiges
Alloprevotella
Alphaproteobacteria_unclassified
Alteromonadales_unclassified
Alysiella
Aminobacter
Ampullimonas
Anaerocella
Anaerosinus
Anaerostipes
Anaerovorax
Anthococcus
Aquabacterium
Aquaspirillaceae_unclassified
Aquaspirillum
Aquincola
Arcobacter
Arenimonas
Asticcacaulis
ATCC−39006
Atlantibacter
Atopostipes
Azoarcus
Azomonas
Azonexus
Azorhizophilus
Azospira
Azotobacter
B1−7BS_ge
Bacillaceae_unclassified
Bacilli_unclassified
Bacillus
Bact−08
Bacteriovorax
Bacteroidales_unclassified
Bacteroides
Bacteroidetes_vadinHA17_ge
Bacteroidia_unclassified
BD1−7_clade
Bdellovibrio
Bergeriella
Betaproteobacteriales_unclassified
Bifidobacterium
Blastocatellaceae_ge
Blastocatellaceae_unclassified
Blautia
Blfdi19_ge
Brachymonas
BSV13
Burkholderiaceae_unclassified
Buttiauxella
Butyricicoccus
C39
Caenispirillum
Caldimonas
Campylobacterales_unclassified
Candidatus_Accumulibacter
Candidatus_Competibacter
Candidatus_Electrothrix


Candidatus_Finniella
Candidatus_Jidaibacter
Candidatus_Nitrotoga
Candidatus_Ovatusbacter
Candidatus_Paracaedibacter
Candidatus_Symbiobacter
Carnobacteriaceae_unclassified
Caulobacter
Caulobacteraceae_unclassified
Cavicella
Cereibacter
Chitinibacteraceae_unclassified
Chitinivorax
Chitinophagaceae_unclassified
Christensenellaceae_R−7_group
Chromatiaceae_unclassified
Chromobacteriaceae_unclassified
Cl75cm.2.12
Citrobacter
Clostridia_unclassified
Clostridiaceae_1_unclassified
Clostridiales_unclassified
Clostridium_sensu_stricto_1
CM1G08
Comamonas
Coprococcus_2
Coprococcus_3
Corynebacterium
Coxiella
CPR2_ge
Crenobacter
Curvibacter
Dechlorobacter
Dechloromonas
Dechlorosoma
Delftia
Deltaproteobacteria_unclassified
Denitratisoma
Denitromonas
Desulfobacterales_unclassified
Desulfobulbaceae_unclassified
Desulfobulbus
Desulfomicrobiaceae_unclassified
Desulfomicrobium
Desulfonema
Desulforegula
Desulfovibrio
Desulfovibrionaceae_unclassified
Desulfovibrionales_unclassified
Desulfuromonadales_unclassified
Devosia
Devosiaceae_unclassified
dgA−11_gut_group
Dialister
Diaphorobacter
DMER64
Dokdonella
Dongia
Dorea
Duganella
Ellin6067
Enhydrobacter
Enterobacter
Enterobacteriaceae_unclassified
Enterococcus
Erwinia
Erysipelothrix
Erysipelotrichaceae_UCG−004
Escherichia−Shigella
Extensimonas
Faecalibacterium
Family_XI_ge
Family_XI_unclassified
Family_XIII_AD3011_group
Family_XIII_unclassified
Ferribacterium
Ferritrophicum
Ferruginibacter
Fibrobacter
Fibrobacteraceae_unclassified
Fibrobacterales_unclassified
Firmicutes_unclassified
Flavobacterium
Fluviicoccus
Fodinicola
Fontimonas
Formivibrio
Franconibacter
Fusibacter
Fusicatenibacter
Gallicola
Gallionellaceae_unclassified
Gammaproteobacteria_unclassified
Gastranaerophilales_ge
Gemmobacter
Geobacter


Georgfuchsia
Giesbergeria
Gottschalkia
H1
Haemophilus
Hafnia−Obesumbacterium
Haliangium
Holospora
Holosporaceae_unclassified
Hydrogenedensaceae_ge
Hydrogenophaga
Hydrogenophilaceae_ge
Hydrogenophilaceae_unclassified
Hyphomicrobium
Ideonella
Ignatzschineria
Ignavigranum
IMCC26256_ge
Insolitispirillum
Intestinibacter
Isobaculum
Jeotgalibaca
JGI_0001001−H03
Kinneretia
Klebsiella
Lachnoclostridium
Lachnoclostridium_5
Lachnospira
Lachnospiraceae_AC2044_group
Lachnospiraceae_ge
Lachnospiraceae_ND3007_group
Lachnospiraceae_NK4A136_group
Lachnospiraceae_UCG−001
Lachnospiraceae_unclassified
Lacihabitans
Lactivibrio
Lactobacillales_unclassified
Lactobacillus
Lactococcus
Lactonifactor
Lacunisphaera
Laribacter
Leeiaceae_unclassified
Legionella
Lelliottia
Leminorella
Lentimicrobiaceae_ge
Leptospira
Limnohabitans
Lineage_IIb_ge
Luteimonas
Macellibacteroides
Macromonas
Magnetospirillum
Malikia
Marinospirillum
Marvinbryantia
Melissococcus
Meniscus
Methylobacillus
Methylophilaceae_unclassified
Methylophilus
Methylotenera
Methyloversatilis
Methylovorus
Micrococcaceae_unclassified
Midichloriaceae_unclassified
Mitsuokella
MM1
Mollicutes_RF39_ge
Moraxella
Moraxellaceae_ge
Moraxellaceae_unclassified
Morganella
Muribaculaceae_ge
Muribaculaceae_unclassified
MVP−15_ge
Myxococcales_unclassified
Nannocystis
Negativibacillus
Neisseriaceae_unclassified
Neorhizobium
Nissabacter
Nitrincolaceae_unclassified
Nitrosomonadaceae_unclassified
Nitrosomonas
Nitrospira
Niveibacterium
oc32
Oceanibacterium
Oceanobacter
Oceanospirillales_unclassified
OLB12
Oligella
OM43_clade
Omnitrophicaeota_ge
OPB56_ge


Opitutaceae_unclassified
Opitutus
Ottowia
Oxalobacter
P3OB−42_ge
Paludibacter
Paludibacteraceae_unclassified
Parabacteroides
Paracoccus
Paraherbaspirillum
Parasutterella
Pasteurellaceae_unclassified
Pedosphaeraceae_ge
Peptostreptococcaceae_unclassified
Petrimonas
Phascolarctobacterium
Phenylobacterium
Phreatobacter
Planktothrix_NIVA−CYA_15
Pleomorphomonas
Pluralibacter
Polaromonas
Polynucleobacter
Pragia
PRD18C08_ge
Prevotella_2
Prevotella_9
Prevotellaceae_unclassified
Prolixibacteraceae_unclassified
Propionivibrio
Prosthecobacter
Proteiniclasticum
Proteobacteria_unclassified
Proteocatella
Pseudacidovorax
Pseudarthrobacter
Pseudocitrobacter
Pseudomonadaceae_unclassified
Pseudomonadales_unclassified
Pseudomonas
Pseudorhodobacter
Pseudoxanthomonas
Psychrobacter
Quatrionicoccus
Rahnella
Raoultella
RBG−16−49−21
Renibacterium
Rheinheimera
Rhizobacter
Rhizobiaceae_unclassified
Rhizobiales_unclassified
Rhizorhapis
Rhodanobacter
Rhodanobacteraceae_unclassified
Rhodobacter
Rhodobacteraceae_unclassified
Rhodocyclaceae_ge
Rhodocyclaceae_unclassified
Rhodocyclus
Rhodoferax
Rhodospirillales_unclassified
Rhodothermaceae_unclassified
Rikenellaceae_ge
Rikenellaceae_RC9_gut_group
Rikenellaceae_unclassified
Rivibacter
Rivicola
Romboutsia
Roseburia
Roseimarinus
Rugamonas
Ruminiclostridium_5
Ruminiclostridium_6
Ruminiclostridium_9
Ruminobacter
Ruminococcaceae_ge
Ruminococcaceae_NK4A214_group
Ruminococcaceae_UCG−002
Ruminococcaceae_UCG−003
Ruminococcaceae_UCG−005
Ruminococcaceae_UCG−008
Ruminococcaceae_UCG−010
Ruminococcaceae_UCG−013
Ruminococcaceae_UCG−014
Ruminococcaceae_unclassified
Ruminococcus_1
Ruminococcus_2
Saccharimonadales_ge
Saccharofermentans
Salmonella
SAR324_clade(Marine_group_B)_ge
SC−I−84_ge
Sedimentibacter
Seleniivibrio
Sellimonas
Sericytochromatia_ge
Serratia
Shewanella
Shinella



Supplementary Figure 2. All detected resistance gene carriers assigned to each gene in alphabetical order. Red color stands for bla_{CTX-M}, green for bla_{NDM}, turquoise for bla_{OXA} and purple for *qacEΔ1*.