DISS. ETH NO. 27852

Tracking anthropogenic footprints of antimicrobial resistance in the river system: A Swiss perspective

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

JANGWOO LEE

Master of Science, Pohang University of Science and Technology

Born on 28.11.1987

citizen of Republic of Korea

accepted on the recommendation of

Prof. Dr. Alex Hall, examiner

Dr. Helmut Bürgmann, co-examiner

Prof. Dr. Célia Manaia, co-examiner

2021

Contents

List of Tables and Figures

Summary (in English)

Antimicrobial resistance (AMR) is a serious threat to public health as recognized by many international and governmental entities. Rivers are important routes through which anthropogenic AMR is transmitted to other environments. Various factors could affect the fate of riverine AMR, e.g., hydrogeologic processes, various sources/sinks, and weather. In this study, the interplay among those factors and their impact on riverine resistome were studied in high wastewater-impacted rivers in Switzerland.

In the first research chapter (CH.2), two high wastewater-impacted rivers were studied under dry-flow conditions under the following hypothesis: 1) wastewater treatment plants (WWTPs) are the main sources of AMR contamination in receiving waters, 2) wastewater-origin AMR could be diluted by additional water inputs (i.e., groundwaters and/or tributary inputs), 3) additional sources/sinks could affect the fate of wastewater-borne AMR over the downstream river continuum. Our observation found that sharp decreases of AMR levels over short ranges (2.0 – 2.5 km) in our study sites were due to the dilution by additional water inputs. The impact of additional sources/sinks became more apparent over longer downstream distances. Especially, we found rapid increases of two indicator genes (i.e., *sul1* and *intI1*) at far downstream locations (5 – 6.8 km) in one of the samplings in the River Murg, Münchwilen (MUE). Various lines of evidence revealed that there might not be a significant anthropogenic contamination – the increases might originate from in-system growths.

In the second research chapter (CH.3), the impact of stormwater impact on riverine AMR was studied in one of the study sites from CH.2 (MUE). The main goals were 1) to identify the impact of various stormwater-related sources (i.e., wastewater bypass, effluents, resuspension of river surface sediments, catchment soil runoffs) on riverine antibiotic resistance genes (ARGs) and multi-resistance risk factors (i.e., multi-resistance bacteria and the genomic determinants), and 2) to identify key contributor(s) among the sources. Our results indicated that wastewater bypass was the major source of ARGs and multi-resistance risk factors during stormwater events. Furthermore, the levels of those factors persisted for a while after stormwater events stopped (i.e., up to 22 hours), probably due to the hydraulic transport of upstream inputs. In general, we report that the risk of public exposure to ARGs and multi-resistance risk factors increases profoundly during stormwater events.

In the last research chapter (CH.4), predictive models for AMR contamination under dry-flow conditions were developed taking key observations from the previous chapters and related projects into account: 1) WWTPs are the major sources of AMR contamination in the river, 2) wastewater-borne AMR loadings decrease over downstream distance, 3) intrinsic levels of AMR exist in less-

disturbed waters. The model employing 1) and 3) was finally selected. Using this model, AMR loadings, levels, and potential public exposure to aquatic AMR were predicted at wastewater-receiving waters in the entire river networks. This model provides insight on prioritizing hotspots for AMR contamination where governmental interventions are potentially required.

This thesis provides a comprehensive theoretical and empirical framework that helps to understand 'how the interplay between various factors could affect the riverine AMR'. This thesis also provides important insights which could potentially help to come up with future interventions. In general, WWTPs were found to be key contamination sources of AMR during both dry weather and rainfall. Thus, I propose WWTPs could be important intervention points for halting the dissemination of anthropogenic AMR in the environment. The potential ways for future interventions were further discussed in CH.5.

Summary (in German)

Antibiotikaresistenzen (ABR) sind ernsthafte Bedrohungen für die öffentliche Gesundheit, welche von vielen internationalen und staatlichen Stellen anerkannt sind. Flüsse sind wichtige Wege, über die anthropogene ABR in andere Ökosysteme übertragen werden können. Der Verbleib von ABR in Flüssen kann von verschiedenen Faktoren beeinflusst werden, z.B. hydrogeologische Prozesse, verschiedene Quellen/Senken und das Wetter. In dieser Studie wurde das Zusammenspiel dieser Faktoren und ihr Einfluss auf flussgebundene ABR in stark abwasserbelasteten Flüssen in der Schweiz untersucht.

Im ersten Forschungskapitel (CH.2) wurden zwei stark abwasserbelastete Flüsse unter Trockenbedingungen auf folgende Hypothesen untersucht: 1) Kläranlagen sind die Hauptquellen für ABR-Kontaminationen in den Vorflutern, 2) die aus dem Abwasser stammenden ABR könnten durch zusätzliche Wassereinträge (Grundwasser und/oder Nebenflüsse) verdünnt werden, 3) zusätzliche Quellen/Senken könnten den Verbleib der aus dem Abwasser stammenden ABR flussabwärts beeinflussen. Unsere Beobachtung zeigte, dass starke Abnahmen der ABR-Werte über kurze Distanzen (2,0 - 2,5 km) in den Untersuchungsgebieten auf die Verdünnung durch zusätzliche Wassereinträge zurückzuführen waren. Der Einfluss zusätzlicher Quellen/Senken flussabwärts wurde mit zunehmender Entfernung deutlicher. Insbesondere fanden wir einen raschen Anstieg von zwei Indikatorgenen (sul1 und intI1) an weit flussabwärts gelegenen Standorten (5 - 6,8 km) in einer der Probenahmen in der Murg, Münchwilen (MUE). Basierend auf verschiedenen Anhaltspunkten können wir annehmen, dass es sich nicht um eine signifikante anthropogene Verunreinigung handelte , sondern dass der Anstieg der Indikatorgene auf systeminternem Bewuchs zurückzuführen sein könnte.

Im zweiten Forschungskapitel (CH.3) wurde der Einfluss von Regenwasserereignisse auf die ABR in MUE untersucht, ein Untersuchungsgebiete aus CH.2. Die Hauptziele waren 1) die Identifizierung des Einflusses verschiedener Regenwasserquellen (d.h. Abwasser-Bypass, Abwässer, Resuspension von Flusssedimenten, Bodenabfluss aus dem Einzugsgebiet) auf Antibiotikaresistenzgene (ARG) und Multiresistenz-Risikofaktoren (d.h. multiresistente Bakterien und genomische Determinanten) in Flüssen und 2) die Identifizierung der Hauptquellen von ARG während Regenwasserereignissen. Unsere Ergebnisse zeigten, dass der Abwasser-Bypass die Hauptquelle für ARG und Multi-Resistenz-Risikofaktoren bei Regenereignissen war. Darüber hinaus konnten wir zeigen, dass wahrscheinlich durch die flussaufwärts eingebrachten Abflüsse die erhöhten ARG-Werte noch eine gewisse Zeit (bis zu 22 Stunden) nach den Niederschlägen bestehen blieben . Im Allgemeinen sahen wir, dass das

Risiko einer Exposition der Bevölkerung gegenüber ARGs und multiresistenten Risikofaktoren während Regenwasserereignissen stark ansteigt.

Im letzten Forschungskapitel (CH.4) wurden Modelle zur Vorhersage für ABR-Kontaminationen unter Trockenwetterbedingungen und unter Berücksichtigung der wichtigsten Beobachtungen aus den vorherigen Kapiteln und verwandten Projekten entwickelt: 1) Kläranlagen sind die Hauptquellen der ABR-Kontamination im Fluss, 2) abwasserbedingte ABR-Belastungen nehmen über die flussabwärts gerichtete Entfernung ab, 3) intrinsische ABR-Mengen existieren in Gewässern, die kaum berührt sind. Das Modell, das 1) und 3) verwendet, wurde schließlich ausgewählt. Mit Hilfe dieses Modells wurden die ABR-Belastungen, die Gehalte und die potenzielle Exposition der Bevölkerung gegenüber aquatischen ABR an abwasserführenden Gewässern im gesamten schweizer Flussnetz vorhergesagt. Dieses Modell bietet einen Einblick in die Priorisierung von Hotspots für ABR-Kontaminationen, an denen möglicherweise staatliche Eingriffe erforderlich sind.

Diese Arbeit liefert einen umfassenden theoretischen und empirischen Hintergrund, der hilft “wie das Zusammenspiel verschiedener Faktoren die ABR in Flüssen beeinflussen kann" zu verstehen. Diese Arbeit liefert auch wichtige Erkenntnisse, die möglicherweise helfen können, zukünftige Interventionen zu entwickeln. Im Allgemeinen wurde festgestellt, dass Kläranlagen sowohl bei Trockenwetter als auch bei starken Regenereignissen die Hauptkontaminationsquelle für ABR sind. Daher schlage ich vor, dass Kläranlagen wichtige Interventionspunkte sein könnten, um die Verbreitung von anthropogenen ABR in der Umwelt zu stoppen. Die möglichen Wege für zukünftige Interventionen wurden in Kapitel 5 weiter diskutiert.

CHAPTER 1. Introduction

River sampling under the dry-flow
-The River Suze, Villeret, 2018 Summer

Wastewater effluent discharge from WWTP-Münchwilen

(Left) River sampling during the stormwater event
-The River Murg, Münchwilen, 2019 Summer

River sampling for the model validation (Right)
-The River Glatt, Flawil, 2020 Summer

1.1. Development of Antimicrobial Resistance

Antibiotics – The Core of Infectious Disease Control in the Modern Era

Antibiotics, as defined by naturally occurring or man-made molecules that show antagonistic effects against bacterial or fungal growths (Nicolaou and Rigol, 2018), have been among the most important treatments of human and animal medicine ever since their first discovery. The first antibiotic in human history was mycophenolic acid which was isolated from *Penicillium glaucum*, and reported by an Italian scientist, Bartolomeo Gosio, in 1893 (Gosio, 1893;Nicolaou and Rigol, 2018). It inhibits the growth of *Bacillus anthracis* and suppresses the activities of virus, fungi, tumor, and psoriasis (Kitchin et al., 1997). Despite its early discovery, it was not commercialized until 1995 when US Food and Drug Administration approved it as an immunosuppressant, not an antibiotic(Nicolaou and Rigol, 2018). Another well-known early generation antibiotic is penicillin which was isolated from *P. chrysogenum* and discovered by a Scottish scientist, Alexander Fleming, and his colleagues in 1928. Ever since its growing use in World War II and then commercialization in 1945, it has saved numerous lives (Nicolaou and Rigol, 2018).

As time passes, more antibiotics have been developed and worldwide consumption has increased accordingly. While the degree of increase depends on the region, it is particularly significant in low- and middle-income countries (LMICs). Here the use of antibiotics important to public health increased by 165 % between 2000 – 2015 compared to the 91 % global increase (Sriram et al., 2021).

Switzerland is one of the countries actively trying to reduce misuse and overuse of antibiotics. Owing to this effort, the domestic consumption of antibiotics for livestock decreased by 52 % from 2010 – 2019 (Fig. 1.1B). However, the consumption still increased by 13% for human hospital uses (i.e., inpatient settings) during the same period, and decreased limitedly (by 4 %) for outpatient settings from 2016 – 2019 (Fig. 1.1A).

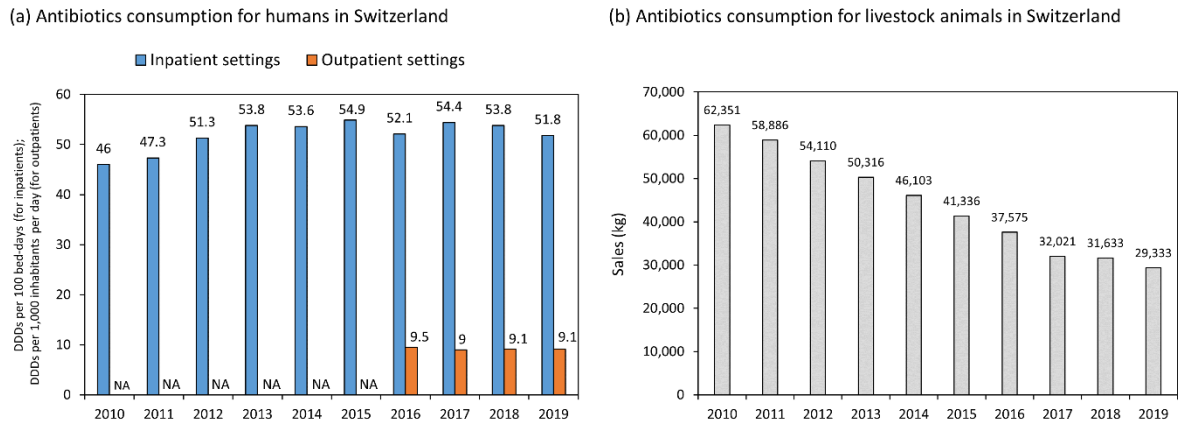


Figure 1.1. Antibiotic consumptions in Switzerland during 2010 – 2019. (a) Antibiotics for inpatient settings (blue) and outpatient settings (orange), (b) Antibiotics licensed for livestock animals (grey). Data source: (FOPH and FSVO, 2020).

Emergence and Evolution of Antimicrobial Resistance

As the number of commercialized antibiotics and their consumption increase, antimicrobial resistance (AMR) has emerged. International and governmental entities recognize it as a serious threat to public health (WHO, 2020). Considering the astounding abundance and diversity of bacteria, some bacteria could intrinsically possess genetic potentials of resistance (Bengtsson-Palme et al., 2018). Adaptive evolution could allow those bacteria showing phenotypic resistance to increase in abundance within a population. For instance, under antibiotic-mediated selection pressure, resistant individuals could have a higher survival and reproduction rate compared to sensitive individuals (Melnyk et al., 2015). However, this evolutionary process takes place slowly and cannot alone explain the astoundingly rapic emergence and development of the many different mechanisms of AMR which has taken place during relatively a short period of antibiotic usage history. It is now understood that the majority of AMR mechanisms observed in pathogens today developed by the acquisition of new mobilized genes (Boerlin and Reid-Smith, 2008).

The development of antibiotic resistance genes (ARGs) could be fostered upon their mobilization in mobile genetic elements, such as integrons, transposons, and plasmids (Bengtsson-Palme et al., 2018). Among them, integrons are regarded as key players in ARG dissemination, especially in gram-negative bacteria (Gillings et al., 2008). The class 1 integron (*intI1*), which is one of the earliest discovered integrons, has been studied well in terms of its ecology and evolutionary history. It encodes a protein catalyzing recombination between the integron recombination site (*attI*) and the *attC* site of gene cassettes (Gillings et al., 2015). As a result, *intI1* often contains many gene cassettes with ARGs in its downstream array (Fig. 1.2).

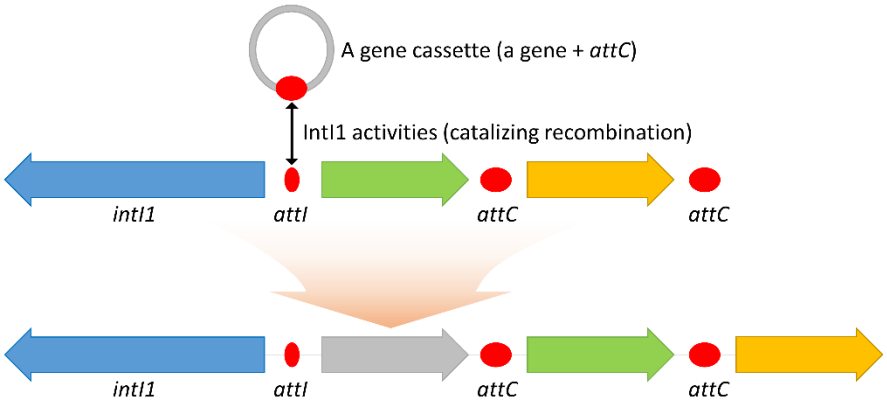


Figure 1.2. A schematic diagram summarizing the recombination activity mediated by the integron integrase class 1 protein (IntI1). Modified from Gillings et al. (2015).

intI1 could be associated with both chromosomes and plasmids. Chromosomal *intI1* has been found in various non-pathogenic and environmental bacteria including *Hydrogenophaga*, *Aquabacterium*, *Acidovorax*, *Imtechium*, *Azoarcus*, and *Thauera* genera (Gillings, 2014). While it is not

mobile on its own, it can be mobilized be if other mobile elements (e.g., transposons and recombinases) were located within its cassette arrays or adjacent to it (Gillings et al., 2008). *intI1* can also be associated with plasmids; in this case, it could be transmitted to other cells more easily via conjugation. For instance, donor cells could transfer the plasmids containing *intI1* and other ARGs in its gene cassette to recipient susceptible cells. In the case where the plasmids contain ARGs, the susceptible cells could become resistant to antibiotics. According to the mobile *intI1* evolutionary theory argued by Gillings et al. (2008), a Tn*402*-like transposon captured the chromosomal *intI1* found in environmental betaproteobacteria via site-specific recombination and then transposed it to a wide variety of plasmids. While this transposition occurred, various genes conferring resistance to various molecules, including antibiotics, were inserted into the cassette array. A quaternary ammonium compound resistance gene, *qacE*, and a sulfonamides resistance gene, *sul1*, are both among the earliest genes that were captured (Gillings et al., 2008). The mobilization of *intI1* is thought to have only occurred recently with the rise in anthropogenic activities (Gillings et al., 2008). Accordingly, *intI1* is often found in pathogens, human-related samples, including human and animal commensals, and in human-impacted environments, such as wastewater and its receiving waters (Gillings, 2014).

There are other mechanisms through which ARGs could be transmitted to susceptible cells, namely transformation and transduction. Transformation is the uptake of exogenous DNA into the cell (Soucy et al., 2015). ARGs could also be transmitted to other cells via transduction. This is mediated by bacteriophages and bacterial viruses which could incorporate ARG-containing fragments of host DNA into their capsid before injecting those fragments into new hosts (Bello-López et al., 2019;Soucy et al., 2015). After that, those fragments could be recombined into the genomes of the new hosts giving them antibiotic resistance (Bello-López et al., 2019). All the mechanisms mentioned above are summarized in Fig. 1.3.

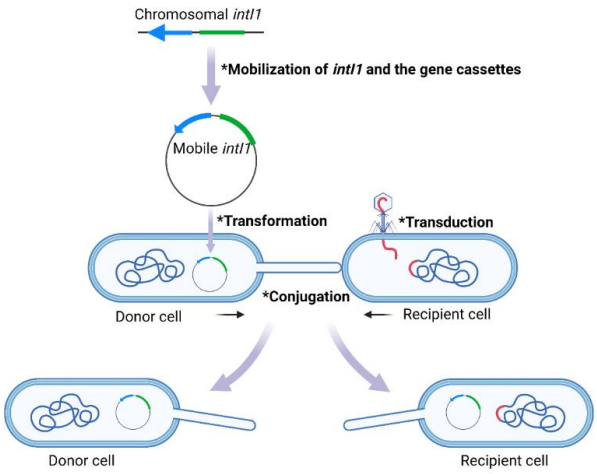


Figure 1.3. Conceptualized diagram summarizing mobilization of *intI1* and horizontal gene transfer processes.

1.2. Dissemination of Antimicrobial Resistance

The One Health and Global Health Concepts – From Humans to Environments

While AMR genetic elements can be mobile by themselves via the abovementioned mechanisms, the host cells (i.e., antibiotic resistant bacteria (ARB)) are also mobile. For instance, ARB that reside in host organisms (i.e., animals or humans) could potentially be transmitted to any organisms or objects with which their host interacts. Considering that human or animal commensal bacteria could be finally discharged to the environment (e.g., via feces), transmission pathways of ARB are interconnected to various parts of the ecosystem. In this sense, the One Health Concept highlights the importance of AMR surveillance in multiple sectors encompassing humans, animals, and ecosystems (Hernando-Amado et al., 2019). Additionally, given that there are no strict environmental borders, animals or environmental components could transmit ARB to other regions. For instance, animals carrying resistance determinants (i.e., ARGs and ARB) could migrate to another country, or environmental resistance determinants could be transported to connected downstream regions via rivers (Hernando-Amado et al., 2019). The modern development of transportation promotes intercontinental travel and trade, meaning humans could also transmit resistance determinants on a global scale. Regarding this, the Global Health Concept indicates that multi-national efforts are important in combating the threat of AMR (Hernando-Amado et al., 2019). The One Health and Global Health Concepts are graphically summarized in Fig. 1.4.

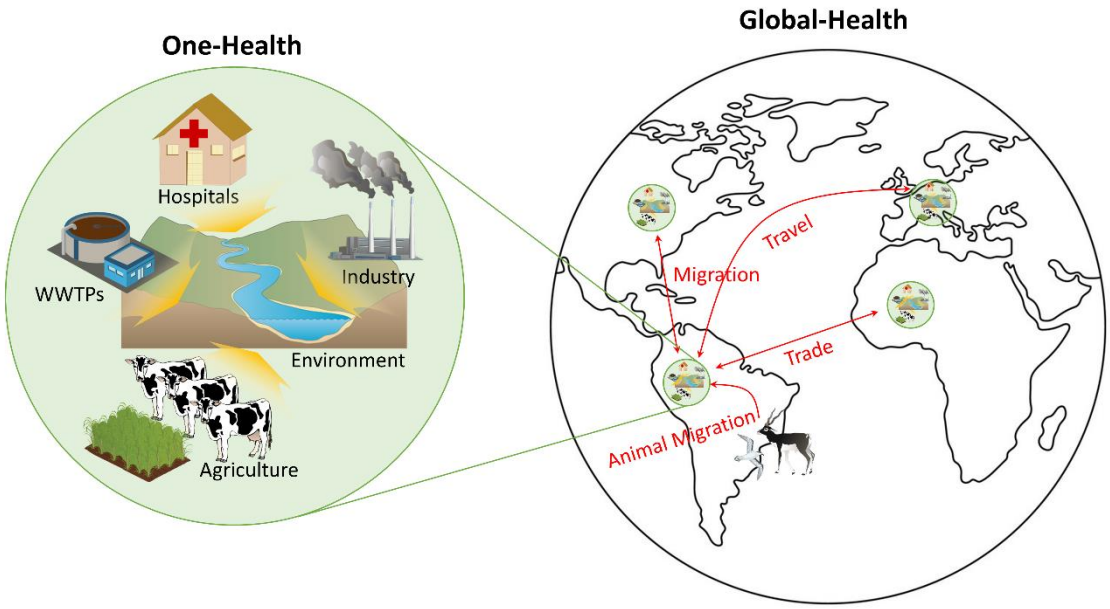


Figure 1.4. The concepts of One-Health and Global-Health. Modified from Hernando-Amado et al. (2019).

In the light of this issue, international entities such as the World Health Organization (WHO) and European Union (EU) have implemented efforts to bring about multi-sectoral and multi-national awareness and action. In 2001, WHO published the Global Strategy for Containment of Antimicrobial Resistance which detailed actions for surveillance and prevention of AMR in addition to disease control (WHO, 2001). In 2012, WHO announced to its member states that multi-disciplinary approaches encompassing human and veterinary medicine as well as farming were now required (WHO, 2012). Later in 2015, WHO announced more concrete plans which highlighted the need for multiple sectors and countries to harmonize tools and standards for combating AMR (WHO, 2020).

In addition, the EU endorsed a resolution on halting AMR (A Strategy Against the Microbial Threat) in 1999 (Gerards, 2011). This has led to the launching of the European Antimicrobial Resistance Surveillance Network (EARS-Net), European Surveillance of Antimicrobial Consumption Network (ESAC-Net), and the European Committee on Antimicrobial Susceptibility Testing (EUCAST). These entities have developed harmonized standards for testing AMR among member countries and have monitored AMR regularly in multiple sectors (Bundesrat, 2015). In Switzerland specifically, the National Research Program 49 (NRP49) launched in 2001 has ensured that the AMR situation has been monitored in humans, animals, and the environment (Bundesrat, 2015). However, the NRP72 will run from 2017 until 2023 and aims to monitor the development and spread of AMR, develop new drugs and new diagnostics, and optimize the use of antibiotics (SNSF, 2019).

All these efforts highlight that scientific investigations and practices required in various sectors (incl. environmental aspects) in light of One Health. Furthermore, AMR is not only a countrywide/region-wide problem. It is also a 'global' problem from the perspective of Global Health. As a result, every country – no matter how much it is seemingly impacted – needs to tackle the issue of the development and dissemination of AMR.

Wastewater Treatment Plants as Major Sources of Anthropogenic Antimicrobial Resistance

Wastewater treatment plants (WWTPs) are important routes through which anthropogenic ARGs and ARB disseminate in the environment. Accordingly, WWTPs are regarded as hotspots of ARGs and ARB (Rizzo et al., 2013). Incoming domestic sewages contain a high abundance of resistance determinants. For example, it has higher abundances than undisturbed freshwaters by more than about 2 orders of magnitude (Czekalski et al., 2012;Rodriguez-Mozaz et al., 2015). Because WWTPs were not originally designed to remove ARGs or ARB, these determinants are not completely removed during treatment (Rizzo et al., 2013). Many studies report that relative abundances of some resistance determinants increase after wastewater treatment processes (Lee et al., 2017;Mao et al., 2015;Rafraf et al., 2016). A study using metatranscriptomics showed that transcriptomic activities of

ARGs were comparatively active in the secondary effluent compared to other treatment steps (Ju et al., 2019). As a result, resistance determinants are present in high abundances even after being treated in WWTPs. The levels in treated effluent are higher than in undisturbed freshwaters by up to approx. 1 order of magnitude (Czekalski et al., 2012;Czekalski et al., 2016b;Rodriguez-Mozaz et al., 2015).

WWTPs can also discharge even higher abundances and loadings of resistance determinants during stormwater events. During heavy rains, combined sewage (i.e., stormwater runoff plus sewage) overflows because WWTPs reach their capacity limits and cannot treat the increased flow (Weyrauch et al., 2010). The portion of untreated combined sewage that exceeds the treatment capacity is discharged directly to the receiving water, which is called 'wastewater bypass' (Toronto, 2021). Considering that untreated sewage contains higher abundances of contaminants as stated above, a profound pollution in the bypass-receiving water is expected during stormwater events. Previous studies revealed that the bypass resulted in a heavy pollution of chemical contaminants (Launay et al., 2016;Weyrauch et al., 2010) and potentially ARGs in receiving waters (Chaudhary et al., 2018;Eramo et al., 2017). On the other hand, other contamination sources such as suspended sediments, surface runoff of catchment soils, and urban runoff could also come into play (Baral et al., 2018;Tsihrintzis and Hamid, 1997). This necessitates a systematic study of various potential AMR sources to figure out the main factor(s) which drive(s) AMR pollution in a receiving river, which has not been comprehensively studied yet.

The abovementioned points make it clear that WWTPs could be major sources of ARGs and ARB under dry weather and stormwater events.

Rivers as Important Routes for the Transmission of Antimicrobial Resistance

Treated wastewater and untreated bypass are largely discharged to rivers. In Switzerland, a total of 631 WWTPs serving 83 % of the population are connected to rivers. Considering that anthropogenic ARGs and ARB could be transported to far downstream locations along the river continuum, the potential roles of rivers in disseminating AMR are significant.

Any pollutants, once in the river, are transported laterally according to hydraulic processes, namely advection, molecular diffusion, and dispersion. Among them, molecular diffusion is typically less important in bulk river water movement because it is slower than the other processes (Nevers and Boehm, 2010). Dispersion, on the other hand, is the mixing of materials due to non-uniform flow and is considered more important than diffusion. Because of dispersion process, molecules are transported differently depending on their concentration gradients which can be mathematically described as follows:

$$J = -D \cdot \frac{dC}{dx}$$

Where, J is the dispersive flux, D indicates the coefficient of dispersion, and dC / dx denotes the concentration gradient (along the direction x) of a pollutant.

Dispersion could occur in both longitudinal (i.e., in the flow direction) and cross-sectional directions, mixing contaminants in all directions. Under the steady-state assumption (i.e., continuous discharges of effluents from WWTPs), longitudinal dispersion does not have to be considered (Fischer et al., 1979). However, cross-sectional mixing could be important in both steady-state and non-steady state conditions. This indicates that a certain downstream distance is required for discharged contaminants to become fully mixed (cross-sectionally). It should thus be considered when sampling surface waters so that representativeness could be guaranteed at each cross-section.

Alternatively, advection is a volumetric transport of materials according to flow velocity. It can be parameterized as:

$$A = v \cdot C$$

Where, A is the advective flux, v indicates the river flow velocity, and C denotes the concentration of a pollutant.

For biologically or chemically reactive contaminants, additional source and sink (or decay) mechanisms are considered. For example, predation, (lower) ambient temperature, sunlight, etc. are considered as sinks and resuspended sediments, intrinsic growths, etc. are regarded as sources for biological contaminants (Nevers and Boehm, 2010).

There have been some studies that investigated the fate of wastewater-borne resistance determinants in rivers. The levels of resistance determinants along the waterway were monitored for instance, in rivers in Beijing, China (Xu et al., 2016), Brussels, Belgium (Proia et al., 2018), and in the Netherlands (Sabri et al., 2018). Those studies reported that the levels of resistance determinants increased after receiving WWTP effluents, but differed on whether they decreased or increased along the downstream waterway. This might be because the various factors mentioned above collectively drive the downstream fate of wastewater-borne resistance determinants. To draw a more comprehensive interpretation, a systematic study involving multiple disciplines is required.

1.3. Methodologies to Analyze Environmental Antibiotic Resistance Genes and Bacteria

Culture Independent Method: Quantitative PCR

One of the ways to detect ARGs is to use quantitative polymerase chain reaction (qPCR). The principle of qPCR is based on the quantitative relationship between the number of PCR cycle when the amplified products start to be detected above the background signal (Cp) and the concentration of the target gene in the sample (Dharmaraj, 2021). As it is, the higher the concentration in the sample, the lower the Cp value. To obtain an accurate detection of the amount of amplified product, fluorescence dyes, such as SYBR Green or TaqMan, are used. SYBR Green dye nonspecifically binds to double-stranded DNA (dsDNA). As the amplified dsDNA increases, the fluorescence emitted by SYBR Green also increases (Dharmaraj, 2021). TaqMan works differently. This is a hydrolysis probe which consists of target specific short sequences with a fluorophore and quencher attached at each end (Dharmaraj, 2021). This probe is specifically annealed to the target sequence of DNA template, and emits fluorescence when Taq polymerase meets its 5'-end while extending primers (Dharmaraj, 2021). Thus fluorescence increases as the number of amplified products increases. Therefore, it is possible to quantify abundances of target genes in a sample by constructing a standard curve of the serially diluted target sequences having known concentrations. qPCR is known as one of the most accurate techniques for quantifying target genes. It was thus relied on to precisely quantify genes of interest in this study.

Culture Independent Method: Metagenomics

Metagenomics is a study of 'unknown' organisms in the sample using genetic materials directly recovered from them (Ju and Zhang, 2015). Unlike amplicon sequencings in which sequence variants from 'amplified target genes' are analyzed, metagenomics is based on the shotgun sequencing of genetic materials directly originated from environmental samples (Ju and Zhang, 2015). In this way, more 'unbiased' information can be recovered from the samples, enabling us to take a solid 'snapshot' of environmental microorganisms.

Two ways of identifying ARG profiles (i.e., resistome) using metagenomics were employed in this study: *de-novo* assembly-based annotation and short read-based annotation approaches. The former requires the assembly of raw reads using a de Bruijn graph. Several well-established assembly pipelines are available such as Megahit (Li et al., 2015), IDBA-UD (Peng et al., 2012), and metaSPAdes (Nurk et al., 2017). Longer assembled sequences (contigs) are produced after assembly, and gene-like sequences (open reading frame (ORF)) are predicted afterwards (Ju and Zhang, 2015). Finally, ARGs are annotated to each ORF against several published ARG databases, such as SARG (Yin et al., 2018) and CARD (Alcock et al., 2020). *de-novo* assembly approach potentially has the following benefits: 1)

Co-location of ARGs could be identified using contig-based analysis, and 2) It is easier to assign taxonomy information to the assembled contigs because chances are higher for those longer fragments to contain taxonomically conserved-sequences (e.g., housekeeping genes). However, assembly efficiency could be low for some samples, meaning the identified gene profiles could not properly represent the original profiles (Myrold et al., 2014). On the other hand, short-read based annotation does not require assembly process and annotates ARGs directly to short reads. This allows a more representative picture of gene profiles to be obtained. As cons, the association among ARGs cannot be identified and the possibility of taxonomy assignments can be lower because short reads contain less information. Several well-established pipelines are available, including ARGs-OAP (Yin et al., 2018) and DeepARG-SS (Arango-Argoty et al., 2018). The overview of two approaches is summarized below in Fig. 1.5.

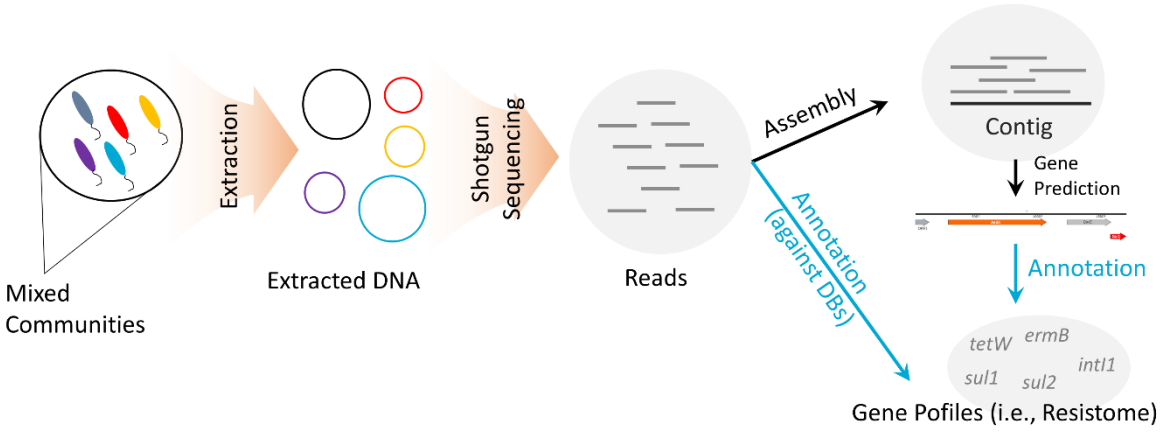


Figure 1.5. The overview of metagenomics using short-read bases annotation and *de-novo* assembly based annotation approaches.

Culture Dependent Method: Heterotrophic Plate Counts

Heterotrophic plate counts were also performed in this study to analyze ARB. Even though this method has disadvantages over culture independent methods (e.g., only about 0.01% of waterborne bacteria are culturable and it can take a long time to cultivate), it is still useful for phenotypically active groups since it allows potentially risky groups to be quantitatively identified (Bartram et al., 2003). This method is also relatively simple to perform and economical (Bartram et al., 2003) and thus a candidate for use in future routine monitoring efforts.

Clarithromycin and tetracycline multi-resistant bacteria as well as sulfamethoxazole, trimethoprim, and tetracycline multi-resistant bacteria were previously reported to be good indicators of wastewaters (Czekalski et al., 2012). Therefore, they were analyzed in this study.

Specifically, water samples were filtered through 0.2 µm pore size filters. Those filters were then cultivated on R2A medium where antibiotics were added. After 72 hours at 25 °C, the colony forming units were enumerated.

1.4. Research Objectives and Hypothesis

In this thesis, the fate of anthropogenic AMR inputs was comprehensively studied in Swiss rivers. Among various anthropogenic sources, WWTPs are regarded as hotspots of AMR. While WWTP-related inputs might be the main sources of AMR, other factors (i.e., dilution by additional water inflows, sources, sinks, and weather) could also play roles in governing the fate of AMR along the river continuum. Therefore, key influential factor(s) of riverine resistome were identified and discussed in the following three research chapters:

Objective 1. (CH. 2) The Downstream Fate of Wastewater-borne Resistance Determinants

In this chapter, riverine resistance determinants (ARGs and ARB) were monitored in two of the most wastewater-impacted rivers in Switzerland under dry-flow conditions.

The main goal in this chapter is:

i) To unravel key driver(s) for governing the downstream fate of wastewater-borne ARGs and ARB.

Under the dry-flow condition, it was hypothesized that dilution by additional water inflows might play a significant role in governing the short-distance fate of wastewater-borne resistance determinants. On the other hand, additional source/sink mechanisms might play more important roles in governing the fate of resistance determinants in longer downstream distance, meaning additional resistance determinants could diffuse into the system from a large catchment area.

Objective 2. (CH. 3) Impact of Stormwater Events on the Riverine Resistome

To assess the effect of weather (i.e., stormwater events), the riverine resistome was monitored during stormwater events in a high wastewater-impacted river in Switzerland.

Two specific purposes of this chapter are:

i) To identify the impact of stormwater-related disturbances (i.e., wastewater bypass, effluent, resuspension of river surface sediments, catchment soil runoff) on the riverine resistome.

ii) To figure out the key source(s) of ARGs in the wastewater-receiving river among stormwater-related disturbances.

It was hypothesized that abundances and/or diversity of ARGs might be different by source. Thus, it was expected that the main driver(s) could be distinguished by comparing different compartments (i.e., stormwater-disturbed waters and different sources) in terms of various parameters, such as micro- and molecular-biological and ecological signatures.

Objective 3. (CH. 4) Model-based Prediction of AMR Contamination in Swiss Rivers

In this chapter, the key observations from the previous chapters were parameterized (if possible), and expanded to the entire Swiss river networks using Graph Theory (network analysis). Two well-known anthropogenic AMR indicators, i.e., *sul1* and *intI1* were used as indicators.

The main purpose is:

i) To predict hotspots of AMR contamination and potentials of public exposure to aquatic AMR in Swiss river networks.

The main underlying hypothesis is 'WWTPs are the main sources of AMR in rivers'. Additional hypothesis mainly on (but not limited to) the downstream fate of WWTP-borne AMR were developed upon finalization of the previous two chapters (CH.2 and CH.3) and will thus be further discussed in CH.4.

Contributions to Related Works

Even though the following works will not be discussed in detail in this thesis, I also contributed to these works in the context of my PhD project:

i) Ju F, Lee J, et al., Phenotypic metagenomics unravels cross-microbiota clinically relevant ESBL and Carbapenemase producers and wastewater-promoted selection in rivers, In Preparation.

ii) Erb S, 2019, Frequency of multidrug resistance in ESBL isolates from water samples, Zurich University of Applied Science (ZHAW), A Semester Project.

iii) Erb S, 2020, Antibiotic resistance in high and low nucleic acid content bacteria, Zurich University of Applied Science (ZHAW), A Bachelor's Thesis.

CHAPTER 2.
Unraveling the riverine antibiotic resistome: the downstream fate of anthropogenic inputs

Jangwoo Lee[1,2], Feng Ju[3,4], Ayella Maile-Moskowitz[5], Karin Beck[1], Andreas Maccagnan[1], Christa S. McArdell[1], Marco Dal Molin[1,6], Fabrizio Fenicia[1], Peter Vikesland[5], Amy Pruden[5], Christian Stamm[1], Helmut Bürgmann[1,*]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, CH-6047 Kastanienbaum or CH-8600 Dübendorf, Switzerland

[2]Department of Environmental Systems Science, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

[3]Key Laboratory of Coastal Environment and Resources of Zhejiang Province, School of Engineering, Westlake University, Hangzhou, China

[4]Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou, China

[5]Department of Civil and Environmental Engineering, Virginia Tech, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, United States of America

[6]The Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, 2000 Neuchâtel, Switzerland

* Corresponding author: Helmut Bürgmann, Eawag, Seestrasse 79, 6047 Kastanienbaum, Switzerland. E-mail: Helmut.Buergmann@eawag.ch

This chapter was published as follows:

Lee et al., 2021. Unraveling the riverine antibiotic resistome: The downstream fate of anthropogenic inputs, Water Research, 197, 117050, https://doi.org/10.1016/j.watres.2021.117050

Abstract

River networks are one of the main routes by which the public could be exposed to environmental sources of antibiotic resistance, that may be introduced e.g. via treated wastewater. In this study, we applied a comprehensive integrated analysis encompassing mass-flow concepts, chemistry, bacterial plate counts, resistance gene quantification and shotgun metagenomics to track the fate of the resistome (collective antibiotic resistance genes (ARGs) in a microbial community) of treated wastewater in two Swiss rivers at the kilometer scale. The levels of certain ARGs and the class 1 integron integrase gene (*intI1*) commonly associated with anthropogenic sources of ARGs decreased quickly over short distances (2-2.5 km) downstream of wastewater discharge points. Mass-flow analysis based on conservative tracers suggested this decrease was attributable mainly to dilution but ARG loadings frequently also decreased (e.g., 55.0-98.5 % for *ermB* and *tetW*) over the longest studied distances (6.8 and 13.7 km downstream). Metagenomic analysis confirmed that ARG of wastewater-origin did not persist in rivers after 5 ~ 6.8 km downstream distance. *sul1* and intI1 levels and loadings were more variable and even increased sharply at 5 ~ 6.8 km downstream distance on one occasion. While input from agriculture and in-situ positive selection pressure for organisms carrying ARGs cannot be excluded, in-system growth of biomass is a more probable explanation. The potential for direct human exposure to the resistome of wastewater-origin thus appeared to typically abate rapidly in the studied rivers. However, the riverine aquatic resistome was also dynamic, as evidenced by the increase of certain gene markers downstream, without obvious sources of anthropogenic contamination. This study provides new insight into drivers of riverine resistomes and pinpoints key monitoring targets indicative of where human sources and exposures are likely to be most acute.

Keywords

Antimicrobial resistance; Wastewater; River system; Metagenomics; Transport; Degradation

2.1. Introduction

Antibiotic resistance is increasingly recognized by international and governmental entities as a growing global public health threat. According to a 2014 report by the Wellcome Trust and the British government, more than 50,000 cases of antibiotic resistant infections occur annually in Europe and the United States and many hundreds of thousands of people die due to infections with resistant bacteria in other regions of the globe (O'Neill, 2014). In the EU and European Economic Area, the annual attributable deaths by infection with antibiotic resistant pathogens have increased significantly between 2007 and 2015, for instance, from 11,000 to 27,000 (Cassini et al., 2019).

Aquatic environments play a potentially important role as routes of dissemination of resistance; environmental niches at the landscape scale are connected to uses including drinking water supply, irrigation, and recreation. Research in this area has greatly intensified over the last decade (Bürgmann et al., 2018;Rizzo et al., 2013;Zhang et al., 2009) and an increasing number of studies have investigated anthropogenic impacts on receiving rivers. Among the earliest investigations were the studies on the Poudre River in Colorado, United States, which proposed quantitative polymerase chain reaction (qPCR)-based quantification of various antibiotic resistance genes (ARGs), along with phylogenetic analysis (e.g., *tetW*), as a framework for tracking anthropogenic inputs. Anthropogenic input of ARGs to the receiving river was well-apparent using this approach (Storteboom et al., 2010). More recently, the advent of shotgun metagenomic sequencing has greatly advanced the resolution in the ability to characterize large-scale impacts of anthropogenic ARG inputs, as was observed in the Han river catchment in Korea (Lee et al., 2020). The authors noted a strong association of fecal contamination as evidence of anthropogenic activities shaping the composition of the downstream antibiotic resistome (collective ARGs in a microbial community). In Switzerland, a study on rivers and lakes identified the occurrence of extended spectrum β lactamase- and carbapenemase-producing Enterobacteriaceae, which presumably originated from anthropogenic activities (Zurfluh et al., 2013). Another recent study revealed that stream microbiota are significantly altered by the input of treated wastewater in natural streams (Mansfeldt et al., 2020).

ARGs and resistant bacteria (ARB) can persist or proliferate in environmental systems by various mechanisms. Horizontal gene transfer may occur, potentially resulting in new combinations of ARGs or the transfer of resistance to environmentally-adapted bacteria that could in turn change the role of the environment as reservoirs of resistance for clinically-relevant bacteria. Furthermore, the possibility of resistance selection under sub-inhibitory concentrations of antibiotics has been reported (Andersson and Hughes, 2012). Recently, first attempts have been made to estimate predicted no effect concentrations (PNECs) for resistance selection in environmental settings

(Bengtsson-Palme and Larsson, 2016). However, current PNECs are an estimate extrapolated from data on isolated bacteria, and could vary substantially under in-situ environmental conditions and with environmental bacteria.

The above examples make clear that treated wastewater discharges have a significant impact on the abundance and types of ARB and ARGs in receiving rivers. Thus, it is crucial that we gain a better understanding of the downstream fate of the anthropogenic antibiotic resistome in receiving rivers. In this sense, few previous studies have attempted to investigate the downstream behavior of various indicators of resistance (e.g., ARBs, ARGs, mobile genetic elements commonly associated with ARGs), and no clear picture of such behavior has as yet emerged. For instance, a study performed in two wastewater treatment plants (WWTPs) and their receiving river in China reported that the levels of wastewater-origin ARGs (*tetC*, *sul1*) and the class 1 integron integrase gene (*intI1*), decreased significantly 1.2 ~ 2.5 km downstream of the wastewater discharge point (Li et al., 2016). On the other hand, a study performed in a Dutch stream showed that the downstream levels of *sul1*, *sul2*, *ermB*, *tetW*, and *intI1* persisted, or even increased for certain genes over a 20 km downstream distance (Sabri et al., 2018). Mass-flow analyses of ARB and ARG are missing. These contradictory results regarding the downstream behavior of resistance determinants could be in principle attributable to various factors – different geo-hydrological conditions, potential inputs from non-point (e.g. agricultural) sources, and the possible existence of biological drivers (i.e., horizontal and/or vertical gene transfer). An improved understanding of the fate of the wastewater-origin antibiotic resistomes and underlying causes would therefore require an integrated approach across multiple disciplines.

The purpose of this study was to track wastewater-origin antibiotic resistomes and identify the key mechanisms governing their fate in two of the most substantially wastewater-impacted rivers in Switzerland. It was hypothesized, that short-distance (up to 1~2 km from wastewater discharge point) behavior of wastewater-origin resistance determinant concentrations would be governed mostly by hydrological effects such as mixing and dilution. Thus, we used conservative chemical tracers to determine dilution effects and further investigated the contribution of dilution on downstream dynamics of resistance determinants. On the other hand, it was expected that over longer distances (more than 1~2 km; up to 13.7 km to the next downstream WWTP) fate of ARGs and ARB would depend also on additional source/sink mechanisms, such as from biological processes (e.g., death or growth of wastewater-origin ARB, in-situ resistance (co)selection by antibiotics or metals, and horizontal and/or vertical gene transfer), and/or non-point sources (e.g., agricultural runoff) which are expected to diffuse into the system continuously from a large catchment area. Therefore, the potential effects of biological drivers over long downstream distances were

investigated after accounting for hydrological effects. To provide a comprehensive assessment of indicators of antibiotic resistance in the environment, we combined various approaches: cultivation of heterotrophic bacteria on media containing antibiotics, quantification of key indicators of anthropogenic sources of antibiotic resistance by qPCR, and broad profiling of the resistome in selected samples using shotgun metagenomic sequencing. Our study contributes to a systematic, interdisciplinary understanding of the mechanisms driving the fate of the wastewater antibiotic resistome in anthropogenically-impacted rivers.

2.2. Materials and Methods

2.2.1. Site description and field work

A list of WWTP effluent-receiving Swiss rivers without known upstream point-source inputs (e.g., other WWTPs) was obtained from a database provided by Eawag, the Swiss Federal Institute of Aquatic Science and Technology (retrieved 2018) (Eawag, 2014). Two sites were selected according to the following criteria: 1) Greatest proportion of effluent discharge to river discharge, 2) Least number of side streams (for minimum dilution effect from side streams), and 3) Longest distance until the receiving river reaches another downstream WWTP. The selected sites were the river Suze in Villeret (VIL) in canton Bern, and the river Murg in Münchwilen (MUE) in canton Thurgau (maps with all sampling points, see supplementary Fig. S2.1 and S2.2). At the sampled sections, both are shallow (generally <30 cm depth under low flow conditions, maximum depth 1 m) rivers of Strahler order number 3 and 6, and a mean annual runoff of 2.03 and 1.61 m^3/s, respectively. The river beds are mostly gravel. To avoid elevated flow conditions we sampled only under dry weather conditions at the time of sampling and during at least the previous 36 hours.

At VIL, we studied a 23.7 km stretch of the Suze that we sampled from 10 km upstream (US5) of the effluent (EF) discharge point of WWTP Villeret and at 8 downstream sites located from 0.5 km (D1) to 13.7 km flow distance downstream (D8) before the Suze reaches the discharge point of another WWTP. Four sampling campaigns were performed in 2018 on July 09 (VI L1), July 19 (VIL2), July 30 (VIL3), and November 05 (VIL4). Different combinations of locations were sampled in each sampling campaign as described in detail in the SI (pp. 45-46). Daily discharge measures from two gauging stations, one near US, and the other near D8 were obtained and are given in Dataset S1.

At MUE, we studied a 7.0 km stretch of the Murg that we sampled from 0.2 km upstream (US) of WWTP Münchwilen and at 8 downstream sites located from 0.5 (D1) to 6.8 km flow distance downstream (D8 before the Murg reaches the discharge point of another WWTP. Three sampling campaigns were performed in 2018 on July 26 (MUE1), August 03 (MUE2), and August 06 (MUE3). Discharge data was obtained from a gauging station near D4 (Dataset S1).

Samplings were performed according to other projects performed in Swiss rivers and WWTPs (Ju et al., 2019;Mansfeldt et al., 2020). At each river sampling location grab samples (5L in sterilized water containers) were obtained by combining water from just below the surface at three points along a river transect: in the middle and roughly equidistant from the banks to each side of the middle point. EF samples were obtained from the final effluent of the WWTPs prior to discharge. Temperature (°C), conductivity (µS/cm), pH, and dissolved oxygen (DO, mg/L), were measured on site in an aliquot of the sample using a portable multi-parameter probe (Multi 3630 IDS, WTW, Germany) at the time of sampling. To make sure EF was fully mixed with receiving water at D1, the conductivity values across the cross-section were measured, and no significant deviation was observed (< 0.5%). All samples were cooled at 4°C in the dark while transported to our laboratory on the same day. Samples were processed on the same and next day within 36 hours. For organic micropollutant analysis, water samples were obtained separately, and stored in pre-combusted glass bottles on site, cooled at 4°C during transportation, and frozen at -20°C in the dark upon arrival at the laboratory until analyzed. Sediment samples were obtained from 5 select locations (US, D1, D2, D5, and D8) for select campaigns (VIL1~3 and MUE1~3), and frozen at -20°C upon arrival at our laboratory.

To better constrain flow velocities in the rivers, salt tracer experiment using NaCl and flow-velocity measurement were performed in separate sampling campaigns in August 2019 (August 23, 2019 for VIL, and August 27, 2019 for MUE) under comparable flow conditions. The results were summarized in Dataset S4 (e.g., flow-velocity and hydraulic residence time).

Further details on sites, field sampling procedures and hydrological experiments are given in the Method section of the SI (pp. 46-47). The exact sampling locations (GPS-coordinates) are given in Table S2.1.

2.2.2. Heterotrophic plate count of antibiotic resistant bacteria (ARB)

Levels (colony forming units (CFUs) per mL) of ARB cultivable on R2A agar plates were determined in the presence of two combinations of antibiotics: 1) clarithromycin (4.0 mg/L) and tetracycline (16.0 mg/L) (CLR/TET), and 2) sulfamethoxazole (76.0 mg/L), trimethoprim (4.0 mg/L), and tetracycline (16.0 mg/L) (SMX/TMP/TET) referring to the resistance breakpoints for Enterobacteriaceae suggested by Clinical and Laboratory Standards Institute (Cockerill et al., 2013) and also one of our previous publications (Czekalski et al., 2012). The detailed protocol is available in the SI (pp. 47).

2.2.3. DNA extraction and quantitative PCR

Two aliquots of each water sample were filtered through two 0.2 µm pore size membrane filters, using 0.5 L for EF and 1.0 L for river water samples. Replicate filters were then processed separately.

DNA was extracted from the filters using DNeasy PowerWater Kit (Qiagen, Germany) following the manufacturer's instructions. For sediment samples, DNA was extracted from about 20 g of wet sediment using the DNeasy PowerMax Soil Kit (Qiagen, Germany). Extraction blanks confirmed absence of DNA contamination (see SI, p.8). The concentration and qualities of extracted DNA were measured using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, USA) (Dataset S2).

Presence and abundance of key indicator genes for anthropogenic ARG inputs (*sul1*, *tetW*, *ermB*, bla_{CTX} and integron integrase class 1 gene *intI1*) (Berendonk et al., 2015;Gillings et al., 2015;Ju et al., 2019), were determined by qPCR as described previously (Czekalski et al., 2012;Czekalski et al., 2014). The detailed qPCR protocols are reported in the SI. Absence of contamination from filtration and extraction procedures was confirmed using an experimental control by qPCR analysis as shown in the SI.

2.2.4. Metagenome and 16S rRNA gene amplicon sequencing analysis

Shotgun metagenomics and 16S rRNA gene amplicon sequencing analysis were performed using Illumina platforms for samples from three selected sampling campaigns. Samples were selected for sequencing according to the following rationale: For VIL the samples were selected only from campaign VIL1 (i.e., 6 samples: US, EF, D1, D2, D5, D8) as the samples from other campaigns (VIL2-3) showed similar patterns of resistance determinant profiles downstream. For MUE, 6 samples from MUE2 and MUE3 campaigns (i.e., US, EF, D1, D3, D5, D8 from each sampling) were selected as the far downstream behaviors of certain ARG (e.g., *sul1*) and *intI1* were significantly different from each other in those campaigns. DNA extracts from replicated filters were pooled. All library construction and sequencing was performed by Novogene (Hong Kong).

A detailed description of the bioinformatics workflow is given in the SI. Briefly, metagenomic data were analyzed as follows: 1) After quality controls of metagenome reads, de-novo assembly was performed using MEGAHIT v1.1.3 (Li et al., 2015), 2). Open reading frames (ORFs) were predicted from the assembled contigs using Prodigal v2.6.3 (Hyatt et al., 2010), and annotated to ARGs using the Structured Antibiotic Resistance Genes (SARG) v2.0 database (Yin et al., 2018), 3). After read mapping to contigs and ORFs using Bowtie2 (Langmead and Salzberg, 2012) and Samtools (Li et al., 2009), the coverage information for contigs and ORFs was calculated according to (Albertsen et al., 2013). 4) Using the coverage information, abundance metrics were calculated as described in Table S2. 5) Further downstream analyses were performed, such as contig-based taxonomy assignment using Kaiju v1.7.2 (Menzel et al., 2016), Kraken2 (Wood et al., 2019) and BLASTN, and detailed annotation and visualization of ARG-containing contigs.

To analyze 16S rRNA gene amplicon sequencing data, we used the DADA2 pipeline (Callahan et al., 2016), and followed the work-flow suggested by the developers. The detailed protocol is described in the SI.

2.2.5. Chemical analysis

Metals, ions (i.e., dissolved cations and anions), nutrients, and dissolved organic carbon were measured as described in Ju et al. (2019) using high-resolution inductively coupled plasma mass spectrometry, ion chromatography, flow-injection analysis, and total organic carbon analyzer, respectively, as described in the SI. Dissolved micropollutants (i.e., pharmaceuticals, antibiotics) were measured as described in Ju et al. (2019) using liquid chromatography triple quad mass spectrometry with electrospray ionization in the SI. Total dried solid (TS) were measured in sediment samples according to standard methods (APHA-AWWA-WPCF., 1981).

2.2.6. Estimating the dilution effect on downstream levels of resistance determinants

Under continuous discharge and after complete horizontal and vertical mixing, the discharged load of a conservative tracer (e.g., sodium) entering the river through EF is expected to be conserved along the river continuum. Under this assumption, any change in the concentration of the conservative tracer would be due to dilution effects by additional water inflows (i.e., groundwater and/or tributary inputs) and additional inputs of the tracer with these inflows. We used sodium and two micropollutants as conservative tracers (i.e., 4/5-methylbenzotriazole, carbamazepine) because these substances had high concentrations in EF compared to the US river and are known to not substantially degrade or adsorb in the river system. The rationale for selecting the conservative tracers is described in more detail in the SI.

Starting with these mass conservation assumptions, under steady state conditions, the dilution parameter (*DP*, the ratio between external water inflow and streamflow at the downstream section between any two points A and B along a river stretch) can be estimated from a ratio of tracer concentrations according to Eq.1:

$$DP_{A\rightarrow B} = (C_B - C_A)/(\overline{C} - C_A) \qquad \textbf{(Eq.1),}$$

where A indicates an upstream location; B denotes a downstream location; C indicates the concentration of a tracer; $\overline{C}$ denotes the average concentration of the tracer in the external inflow between A and B.

The derivation of Eq.1 is schematized in Fig. 2.1, and also described in detail in the SI.

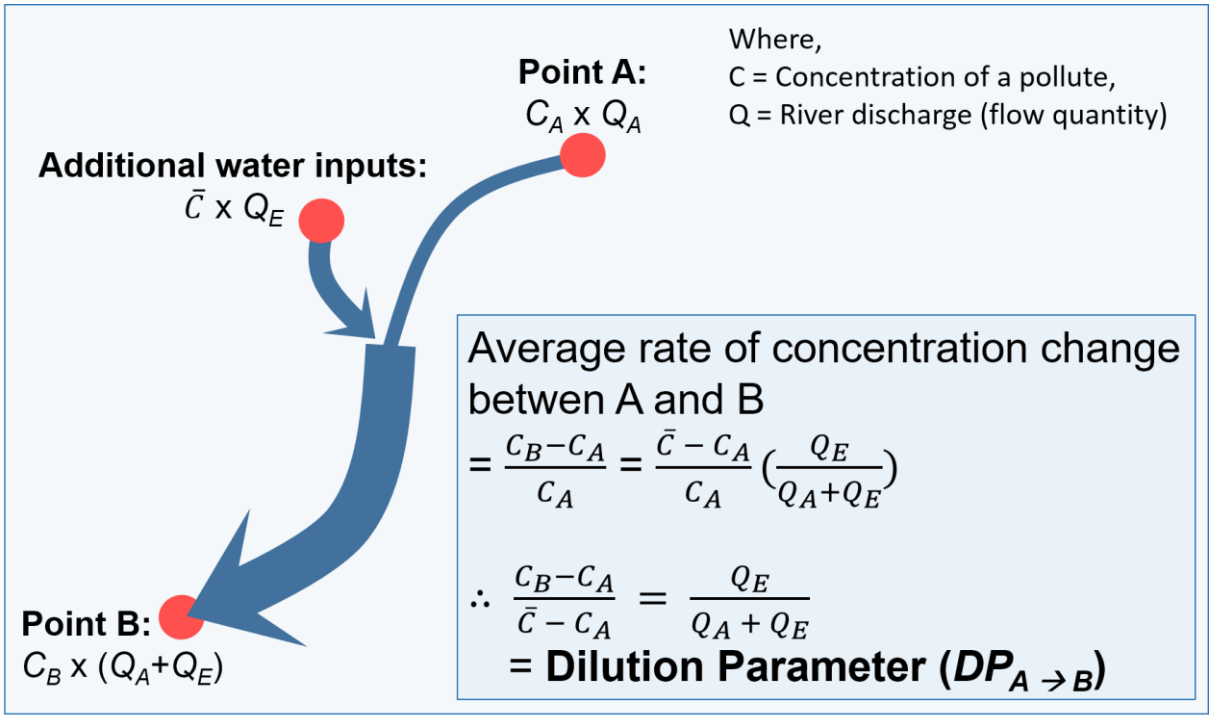


Figure 2.1. Derivation of dilution parameter (DP) from an upstream point A to the downstream point B using the concentration of a pollute as a marker under the mass-flow assumption.

Concentrations C_A or C_B were measured directly for all compounds, $\bar{C}$ was estimated according to the following equation (Eq.2) for sodium (the values shown in Dataset S9.3) and assumed to be 0 for 4/5-methylbenzotriazole and carbamazepine. In short, the difference in sodium loadings (mass per time) between the point of EF discharged and the downstream point where gauging stations were located (D8 for VIL; D4 for MUE) was divided by the quantity of additional water inflows:

$$\overline{Na_{in}} \approx \frac{Na_{D8\ or\ D4} \times Q_{D8\ or\ D4} - (Na_{US} \times Q_{US} + Na_{EF} \times Q_{EF})}{Q_{D8\ or\ D4} - (Q_{US} + Q_{EF})} \quad \textbf{(Eq. 2)}$$

Where, $Na_{D8\ or\ D4}$ denotes the sodium concentration measured at D8 or D4; $Q_{US,EF,D8\ or\ D4}$ indicates the river flow quantity or wastewater effluent discharge (volume per time) at US, EF, D8 or D4.

The value *DP* should be the same for all conservative tracers. To test our hypothesis that "the short distance dynamics of resistance determinants is largely governed by dilution effects", we calculated *DP* over short distance (i.e., $DP_{D1 \rightarrow D2}$ for VIL, and $DP_{D1 \rightarrow D3}$ for MUE), and compared *DP* values over the same distance for resistance determinants (*sul1*, *intI1*, *ermB*, *tetW*, and CLR/TET resistant bacteria) with values for the conservative tracers. Higher DP values for resistance determinants would indicate a lower than expected concentration in the downstream and thus removal.

The expected downstream concentrations of resistance determinants considering dilution as a main driver can be calculated using the $DP_{A \to B}$ of conservative tracers according to the following relationship:

$$\boldsymbol{C_{resist-B} = C_{resist-A} - C_{resist-A} \times DP_{A \to B\,(for\,X)}} \textbf{ (Eq.3),}$$

where $C_{resist-A}$ indicates the concentration of a resistance determinant at an upstream location A; $C_{resist-B}$ denotes the concentration of a resistance determinant at a downstream location B; $DP_{A \to B(for\,X)}$ indicates the *DP* of a conservative tracer (X) between A and B

Eq.3 assumes that resistance determinants behave conservatively over the studied distances and that there are no significant inputs of resistance determinants from the diluting water inflows (i.e., $\bar{C}$ for resistance determinants ≈ 0 in Eq.1). Therefore, deviations from measured to predicted values can indicate violation of these assumptions. We calculated the predicted concentration by dilution effects for each resistance determinant under these assumptions for all downstream sections of the rivers.

2.2.7. Estimating the river discharge over downstream distance

The river discharge (*Q*) at was estimated for several downstream locations where there were not gauging stations. The estimated *Q* values were used when calculating loadings of chemical and resistance indicators over downstream distance. The Q_{EF} values were obtained from each WWTP, and Q_{US} values were either obtained from gauging station (for VIL), or calculated as shown in Eq.8 in the SI (for MUE).

$$\boldsymbol{Q_{D1} = Q_{US} + Q_{EF},}$$

$$\textbf{If n > 2, } \boldsymbol{Q_{D(n)} = Q_{D(n-1)} + Q_{D(n-1)} \times \frac{DP_{D(n-1)\to n}}{1 - DP_{D(n-1)\to n}}} \textbf{ (Eq.4)}$$

Where, $Q_{D(n)}$ indicates the river discharge (*Q*) at the downstream location D(n) (2 ≤ n ≤ 8); $DP_{D(n-1)\to n}$ denotes the dilution parameter between D(n-1) and D(n).

2.3. Results and Discussion

2.3.1. Upstream water quality and WWTP effluent

In agreement with the criteria for site selection, the levels of intI1 and target ARGs upstream of the WWTP were generally low, except for the MUE2 campaign where we observed elevated upstream levels of *ermB*, and *intl1* (Fig. 2.2; Fig. S2.4 and S2.5 in the SI). Chemical water quality likewise did not suggest significant pollution inputs from either tributaries or upstream locations for either VIL or MUE as most micropollutants were below the limit of quantification (Dataset S10~11).

Certain micropollutants (e.g., 4/5-methylbenzotriazole, benzotriazole, and diclofenac) were sporadically detected in very low quantities. For cultivable multi-resistant bacteria (Fig. 2.3), especially CLR/TET resistance, relatively high upstream values were observed in VIL2 and MUE2 US samples. These findings indicate that while there is no indication of significant upstream pollution, some pollution, probably from periodical urban or agricultural activities, may affect the river. There was a settlement upstream of the WWTP and livestock farming activities (i.e., pastures and meadows for livestock) in the catchments, including the upstream sections, in both sites (BAFU, 2013). While we assume surface runoff from the agricultural sites to be minimal as our samplings were performed under dry-weather conditions, it cannot be ruled out that some inputs from agricultural activity occasionally affected the river. Further investigations into the nature of these transient microbial contaminations were not carried out in this project, but future work could employ microbial source tracking or microbial fingerprinting approaches to determine their sources.

The effluent from both WWTPs contained considerable levels of pollutants. For instance, effluent concentrations were higher than the upstream levels by approximately 1 order of magnitude for sodium, 1~2 order of magnitude for ARGs and *intI1*, and more than 2 orders of magnitude for micropollutants (Datasets S8~11). These results are in line with previous results from a large-scale investigation of micropollutants in Swiss streams (Stamm et al., 2016).

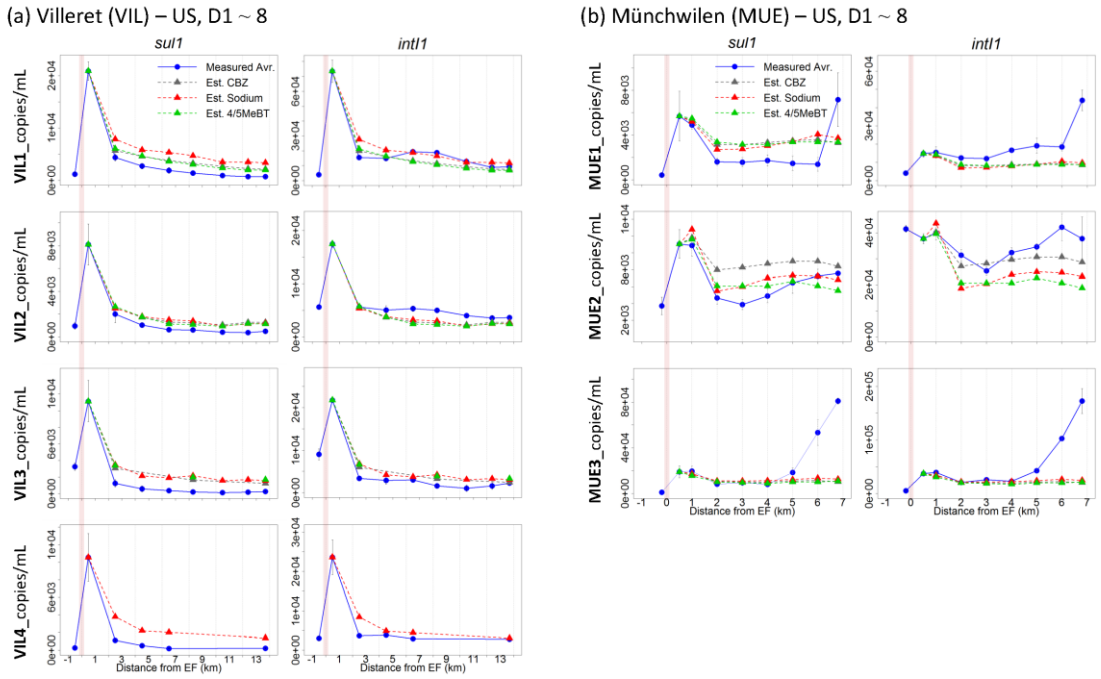


Figure 2.2. Levels (gene copies/mL) of *sul1* and *intI1* in the upstream near effluent discharge point, and downstream river water quantified by qPCR. Average *sul1* and *intI1* concentrations in the (a) river Suze near Villeret (VIL) and (b) river Murg near Münchwilen (MUE). The dotted lines are the estimated levels considering only dilution as a major driver according to the Eq. 3 using sodium, carbamazepine (CBZ), and 4/5-methylbenzotriazole (4/5MeBT) as conservative tracers. The point of EF discharged was indicated by a light red

vertical line. Symbols indicate the average and tips of error bars are the lower and upper values of biological duplicates. The limit of detection for both sul1 and intI1 is 12.5 copies/mL for all the samples shown here.

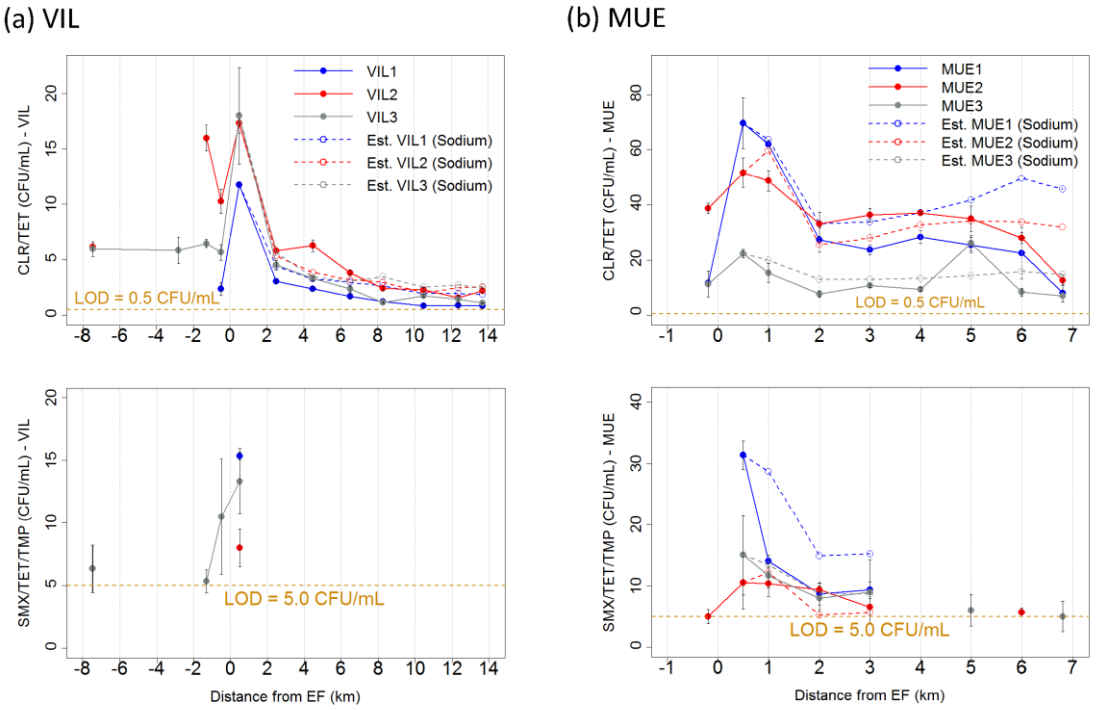


Figure 2.3. Heterotrophic plate counts (CFU/mL) for clarithromycin and tetracycline multi-resistant (CLR/TET) and sulfamethoxazole, trimethoprim, and TET multi-resistant bacte- ria (SMX/TMP/TET) from the upstream and downstream river water in (a) Villeret (VIL), and (b) in Münchwilen (MUE). The predicted values were calculated using a selected conservative tracer (i.e., sodium) according to Eq. 3 , and are shown as dotted lines in red, blue, or black. The limit of detection (LOD) was 0.5 CFU/mL for CLR/TET multi- resistant bacteria and 5.0 CFU/mL for SMX/TMP/TET multi-resistant bacteria. The LOD for SMX/TMP/TET is shown as a yellow dotted horizontal line. The error bars indicate standard errors among technical triplicates.

2.3.2. Short range fate of antibiotic resistance determinants in the downstream river

Focusing on the immediate impact of the WWTP effluents (US versus D1 to D3 sites), there were significant impacts of WWTP effluents on the receiving rivers in both VIL and MUE. The estimated proportions of EF in the downstream receiving waters (D1) estimated by conductivity were 10.5 ~ 35.9 % for VIL1~4, and 33.0 ~ 38.0 % for MUE1~3 (Dataset S9.2). Accordingly, significant increases of *sul1*, *ermB*, *tetW* and *intI1* as quantified by qPCR were observed at D1 compared to US ($p<0.01$ paired t-test; Fig. 2.2 & Fig. S2.5). However, the measured levels of these antibiotic resistance indicator genes rapidly decreased nearly to upstream levels over 2.5 and 2 km downstream distance (D2 or D3 locations) in VIL and MUE, respectively.

The same dynamic was also observed for multi-resistant bacteria (Fig. 2.3), especially CLR/TET resistance. SMX/TMP/TET resistance was often below the limit of detection (5.0 CFU/mL), but clearly exceeded it in the D1 samples and was thus also higher there than further downstream (from D2 on).

Several processes may contribute to the observed decrease of resistance determinants, including dilution by additional water inflows via groundwater and/or tributary inputs, biological deterioration (e.g. cell death or dormancy due to exposure to sunlight, lower ambient temperature, predation, etc.), and cell sedimentation.

2.3.3. Dilution effects strongly affect short distance dynamics of effluent resistance determinants

To determine the importance of dilution effects, we compared DP calculated over a short distance (D1 to D2 for VIL; D1 to D3 for MUE) downstream of the WWTP discharge point ($DP_{D1\rightarrow2}$ for VIL; $DP_{D1\rightarrow3}$ for MUE) from conservative chemical tracer concentrations (e.g., sodium, 4/5-methylbenzotriazole, carbamazepine) as well as ARG and *intI1* levels (Fig. 2.4). The average $DP_{D1\rightarrow2\ or\ 3}$ of the target antibiotic resistance indicator genes levels were always higher than for conservative tracers, indicating possible removal mechanisms at play. However, according to the paired t-test under the null-hypothesis of "No significant differences of dilution parameters between different pairs of bio- and conservative indicators", only the differences between *sul1* and *tetW* versus sodium were significant at p<0.05 (p-adjusted using Benjamini-Hochberg method) (Fig. 2.4), confirming non-conservative behavior and additional removal mechanisms. As $DP_{D1\rightarrow2\ or\ 3}$ for sodium took up a large portion of the values for *sul1* and *tetW* (i. e. $DP(Na^{+})_{D1\rightarrow2\ or\ 3} = 0.72 \sim 0.92 \times DP(sul1)_{D1\rightarrow2\ or\ 3}$ and $0.59 \sim 0.96 \times DP(tetW)_{D1\rightarrow2\ or\ 3}$), the dilution effects quantified by sodium nonetheless explain the majority of the concentration decrease for these parameters. This result implies that the observed rapid decrease in the downstream levels of wastewater-origin resistance determinants immediately downstream of the WWTPs was mainly governed by dilution in the studied systems. Dilution effects thus need to be carefully considered in studies of the environmental fate of resistance determinants, and loadings instead of concentrations need to be determined to accurately assess environmental fate.

2.3.4. Additional source/sink effects become apparent over longer downstream distances

We hypothesized that additional source/sink mechanisms affect the downstream behaviors of antibiotic resistance indicator genes over longer distances. To analyze this in more detail the daily loading (copies/day) for the target ARGs and *intI1* at the point of discharge (as the sum of upstream and EF loadings), and for short (D2 for VIL; D3 for MUE) and far downstream distances (D8) were calculated by multiplying resistance levels (copies/m^3) with the discharge (m^3/day) at each location and then compared. The discharge was either obtained directly from nearby gauging stations, or

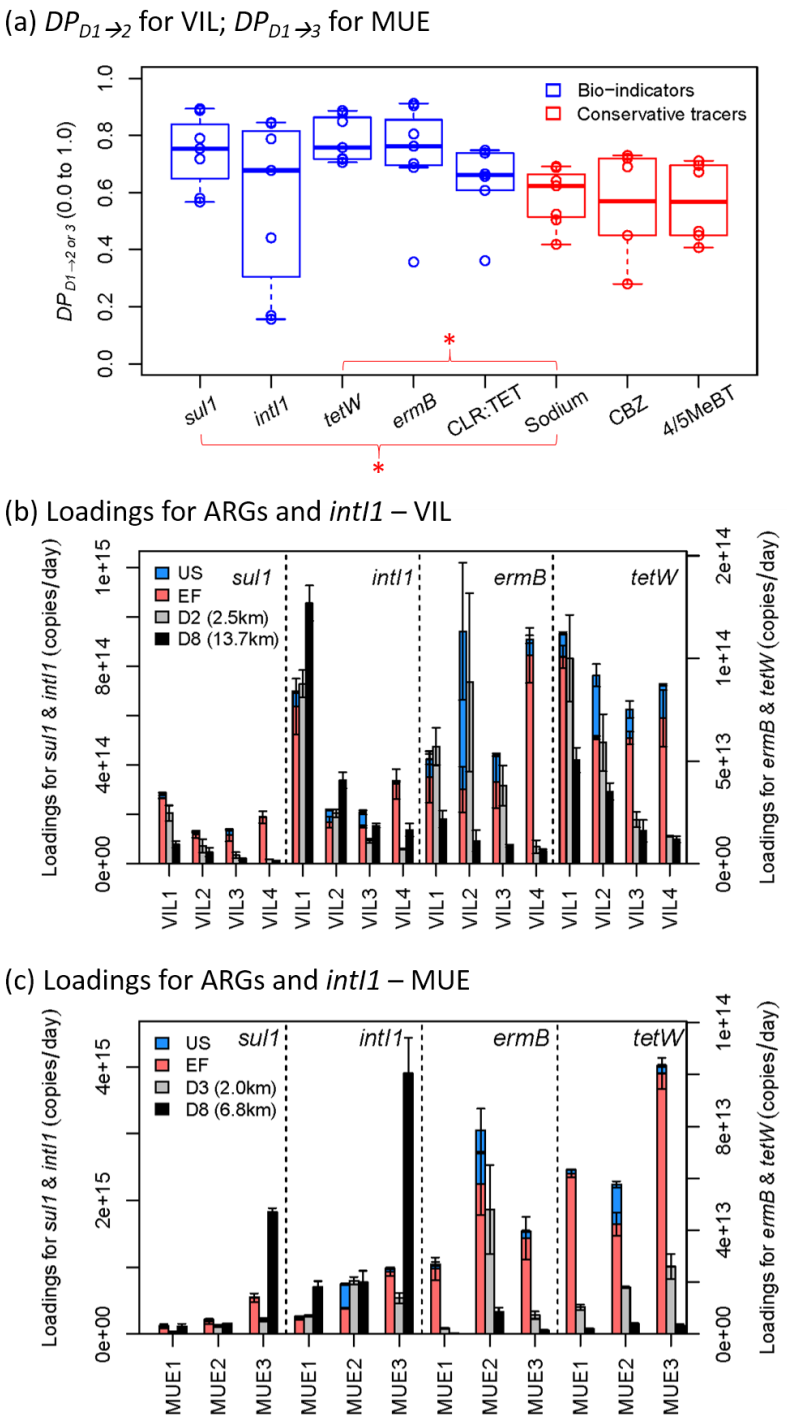


Figure 2.4. (a) Dilution parameter (*DP*) values over short downstream distance (i.e., D1 to D2 for VIL, D1 to D3 for MUE) among different biological and conservative indicators; (b) ARGs and *intI1* loadings at upstream near EF (US), treated wastewater effluents (EF), short downstream (D2 or 3); and long downstream distance (D8) in Villeret (VIL), and (c) in Münchwilen (MUE). The treatment pairs with significant difference in between were asterisked (∗) in (a). The error bars represent upper and lower values of biological duplicates for each gene in (b ~c).

estimated under consideration of the EF discharge (m^3/day) using sodium as an indicator, and according to the equation derived under the mass-conservation assumption (Eq. 4).

The downstream behaviors of the target antibiotic resistance indicator genes varied by indicator and also by sampling campaign. For instance, the load decrease from wastewater discharge (US + EF) to the furthest downstream point (~ D8) was pronounced and consistent for *ermB* and *tetW* in all the samplings (Fig. 2.4b & c). The average load reduction was 81±17 % for *ermB*, and 70±15 % for *tetW* over 13.7 km distance in VIL1~4; 95±5 % for *ermB*, and 96±2 % for *tetW* over the 6.8 km distance in MUE1~3 (Dataset S13). In contrast, the downstream behavior of the *sul1* loadings was inconsistent between sampling campaigns. A pronounced decrease over distance (64-94 % at D8) was seen in VIL1 ~ 4, but little reduction over distance in MUE1 ~ 2 (7 and 29% at D8), and a strong increase in MUE3. The downstream fate of *intI1* was also variable, for instance, as *intI1* loads did not decrease and in some instances even increased .

To further analyze if there is a break point where *sul1* and *intI1* start to deviate from conservative behavior, we calculated the predicted levels of resistance determinants over the whole study distance considering dilution as a major driver using Eq. 3 (Dataset S8). The predictions based on three different conservative tracers are visualized for *sul1* and *intI1* in Fig. 2.2. In VIL, measured levels are always below predictions, except for *intI1* in VIL2. For MUE, in contrast, we see measured values exceeding predicted values in several instances, for *intI1* even for most downstream locations. In MUE3 where the pronounced increase of *sul1* loading was observed between D5 and D8, the level of *sul1* started to exceed predicted values between 5 ~ 6.8 km distance. The concentration of *intI1* increased also very rapidly between 5 ~ 6.8 km downstream distance in MUE3, which indicates either a pronounced proliferation or a non-point source of *sul1* and *intI1* in this stretch of the river. We will discuss potential mechanisms (e.g., biological drivers, on-site selection, additional anthropogenic source input, and surface sediment inputs) in section 2.3.10.

A number of mechanisms may contribute to the generally observed removal: Sedimentation (especially of cell aggregates or flocs) and cell death by predation, UV light, or various other environmental conditions unfavorable to wastewater bacteria. With the available data we are not able to determine the contribution of various mechanisms. Future studies could investigate the persistence of resistant bacteria or molecular resistance markers in micro- or mesocosm experiments or in a turbulent flow system mimicking natural streams to answer such questions. Modeling transport and sedimentation of wastewater-origin particles using the information on particle size, mass, and flow characteristics could provide information on the importance of sedimentation.

2.3.5. WWTP effluent affects the downstream riverine resistome

To obtain a broader view of the river antibiotic resistome we retrieved the ARG content of metagenomes obtained for selected sampling campaigns (VIL1, MUE2, and MUE3). Overall, 65 ARG

subtypes were identified, 49 of them occurred in both upstream and downstream river samples (Dataset S7). The antibiotic resistome in the receiving water closest to the discharge point (D1) was clearly influenced by the input from EF. For instance, a total 28 out of 36 ARG subtypes found in D1 were also detected in EF (B, C, F, G in Fig. 2.5a) while 16 of these were not observed US (B, C in Fig. 2.5a). The 16 EF-derived resistance genes confer resistance to the following antibiotic classes: 1 x aminoglycoside, 4 x beta-lactam antibiotics, 1 x chloramphenicol, 2 x macrolide, 1 x quinolone, 4 x tetracycline; 3 subtypes were multidrug resistance genes. Of these 16 genes, 14 were no longer detected in the far downstream (6.8 ~ 13.7 km downstream distance in MUE and VIL, designated as D_Far in Fig. 2.5a). This is in agreement with the results for qPCR enumeration of ARGs and *intI1* reported above and implies that the majority of ARGs that occurred exclusively in EF do not persist at detectable levels for a long distance in rivers where significant amounts of additional water inflows and additional removal mechanisms are expected.

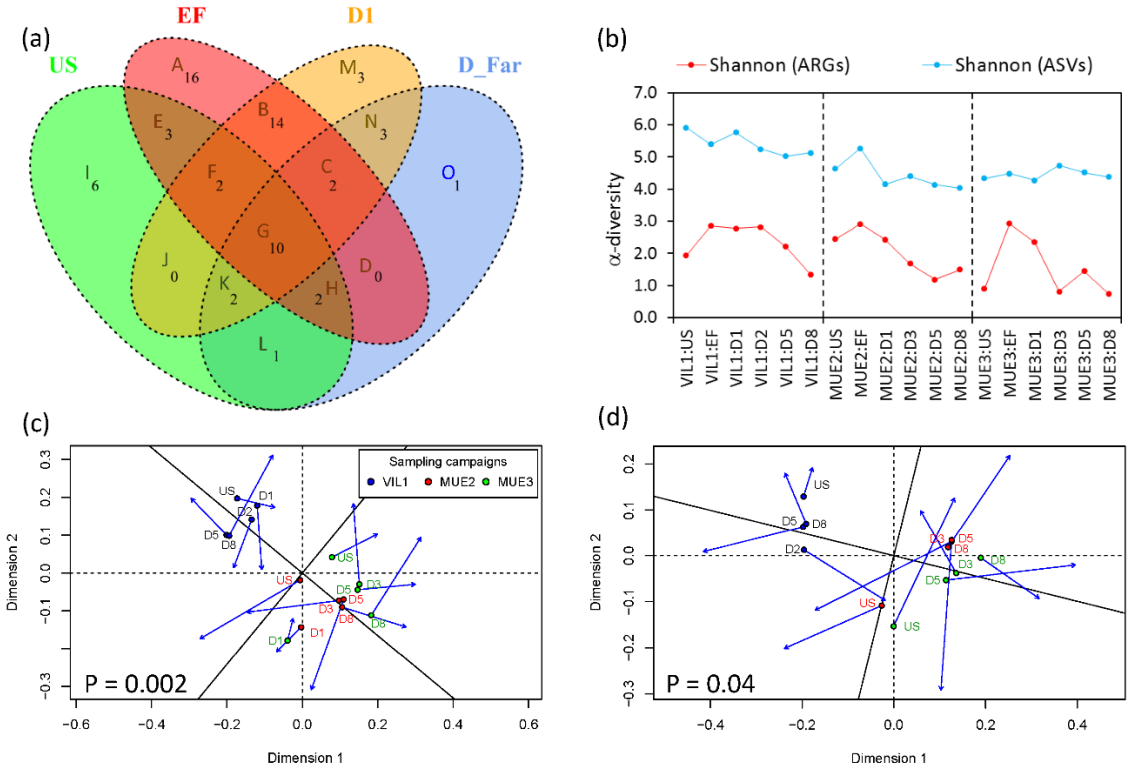


Figure 2.5. Metagenomic analysis of effluent and river antibiotic resistomes at Villeret (VIL) and Muenchwilen (MUE) sites for the selected sampling campaigns (VIL1, MUE2, and MUE3). (a) Venn diagram showing occurrence of antibiotic resistance gene subtypes in the treated wastewater (EF) and in river water upstream (US) and downstream (D1, 0.5 km distance) of the effluent discharge point and in the far downstream (D_Far, 6.8 –13.7 km distance). The presence of ARGs were counted from all three (VIL1, MUE2, MUE3) consolidated-campaigns for each treatment using the presence-absence table shown in Dataset S6. (b) Shannon α-diversity of ASVs (blue) and metagenome- assembled ARG subtypes (red). Procrustes analysis between ASVs (round dot symbols) and resistome (blue arrow tips) where EF was included (c) and EF & D1 were excluded (d). The length of blue arrows indicates the size of Procrustes errors. The error bars represent upper and lower values of biological duplicates for each gene.

2.3.6. Diversity of the river resistome and microbiome along the river continuum

We analyzed the alpha-, and beta-diversity of river resistomes and microbiomes as another way to observe potential effects of the WWTP discharge and to see if the dynamics in the resistome are strongly correlated with the microbial community, as noted e.g. for changes observed during wastewater treatment (Ju et al., 2019). As expected, Shannon alpha-diversity of ARGs was higher in EF compared to US samples by 20.2 ~ 225.4 % (Fig. 2.5b). Accordingly, the impact of the EF resistome was observed, especially for VIL1 and MUE3, as an increase in ARG diversity in river water at the D1 sites. The ARG alpha-diversity decreased downstream in all sampling campaigns. However, for the microbial community as represented by amplicon sequence variants (ASVs), alpha-diversity was not consistently higher in EF versus US, and consequently also did not change significantly from US to D1 and did not consistently decrease downstream (Fig. 2.5b). This indicates that the downstream dynamics of the overall microbial community and the antibiotic resistome were decoupled.

Similar conclusions were obtained from beta-diversity analysis. Procrustes analysis between microbial communities and antibiotic resistome (Fig. 2.5c) revealed a strong structural correlation between microbial communities and antibiotic resistome only when the most strongly effluent-affected sites (D1) were considered ($p = 0.002$; Fig. 2.5c). When the D1 samples were excluded, the correlation was barely significant ($p = 0.04$; Fig. 2.5d), indicating that the structural correlation between microbial communities and resistome largely resulted from the impacts of WWTPs on D1. The weak structural correlation in less impacted waters suggests a lack of strong drivers, such as selective pressures or the influx of external ARB.

2.3.7. Resistome analysis confirms effluent effect and abatement

To quantitatively investigate the dynamics of the resistome along the river continuum in more detail, the seven most prevalent ARG subtypes that appeared in more than 9 out of 18 samples were chosen for detailed analysis. We calculated the proportion of each gene in this set based on relative abundance (GPM, gene per million) (Fig. 2.6a). The proportions (%) of each of six genes (*aph(3'')-I*, *aph(6)-I*, *mexT*, *tetQ*, *aadA*, and *sul1*) to the whole seven genes were lower in US than in EF and D1 (Fig. 2.6a). The *bacA* gene, in contrast, comprised a larger proportion in US (i.e., up to 83 % in VIL1:US), D5 and D8 (i.e., up to 93 % in VIL1:D8) than in EF and D1. It was therefore excluded from the plots of relative and cumulative abundance of the assembled ARG in the metagenome in Fig. 2.6b, and their individual and cumulative environmental level (Gene per liter) in Fig. 2.6c. Both relative and absolute abundances showed a similar pattern – the abundances of the selected ARG were higher in EF and D1 compared to US (Fig. 2.6b). The abundances of those six genes decreased along the downstream locations except for *sul1* in the far downstream location (D8) in the MUE3

campaign (Fig. 2.6b). This analysis confirmed a quantitative effect of the effluent on the abundance of prevalent resistance genes in the river resistome, and suggests additional candidates for tracking anthropogenic sources of resistance in future studies (*aph(3'')-I*, *aph(6)-I*, *mexT*, *tetQ*, *aadA*, and *sul1*) that may be useful for tracking resistance determinants from wastewater. Several of these genes have been used as resistance indicators for environmental samples mainly in combination with culture-dependent approaches(Rizzo et al., 2013;Zhang et al., 2009), but much less frequently with culture-independent approaches (Rizzo et al., 2013;Sharma et al., 2016).

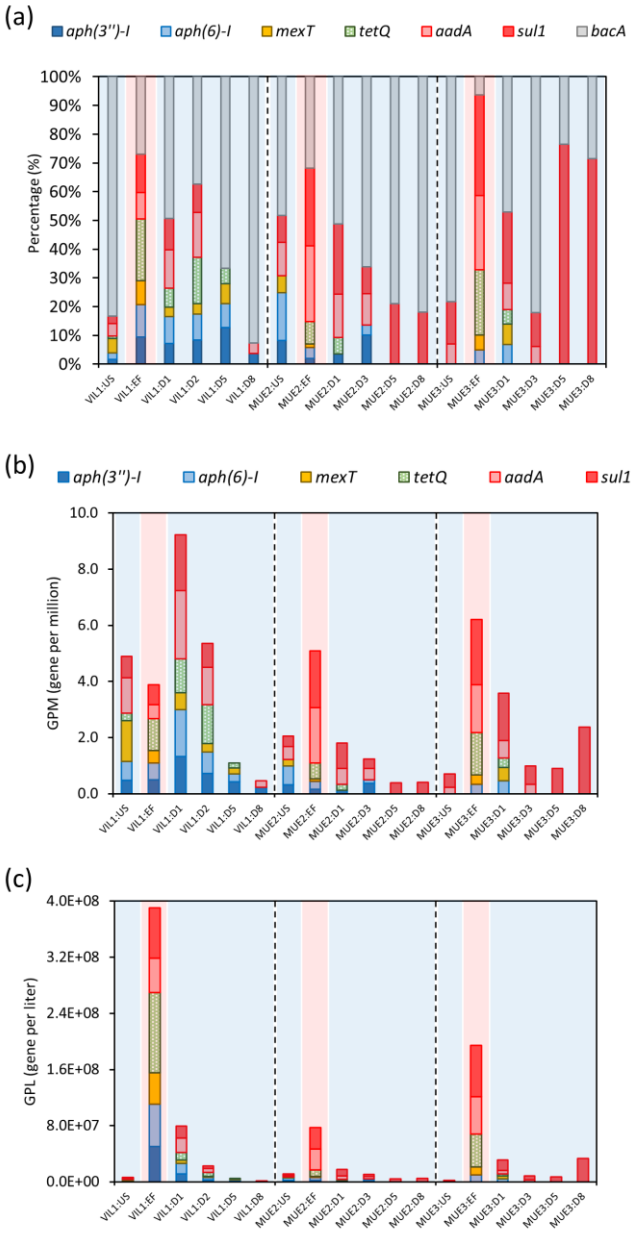


Figure 2.6. Dynamics of prevalent and widespread metagenome-assembled ARGs along the river continuum. (a) The proportion of each gene among the 7 most frequently occurring and widespread ARGs (*aph(3'')-I*, *aph(6)-I*, *mexT*, *tetQ*, *aadA*, *sul1*, and *bacA*). (b) and (c) Stacked bar charts of the abundance of the 6 ARGs that were effluent-associated (omitting *bacA*); (b) relative abundance (GPM, gene per million) and (c) absolute abundance (GPL, gene per liter). Sample EF is shaded in red and the other river water samples are shaded in blue.

2.3.8. Contig analysis indicates ARG co-location

We hypothesized that the similar dynamics of some ARGs could derive from their co-location in the same host or presence on the same genetic elements, so we analyzed their loci within the contigs. *aph(3'')-I* and *aph(6)-I* were indeed found to be located on the same contig in many samples from VIL1 (EF, D1, D2, and D5) (Fig. 2.7a), which may explain the strongly similar dynamics between *aph(3'')-I* and *aph(6)-I* GPM in VIL1 with $R^2 = 0.98$ ($p < 0.001$) (Fig. 2.6b). Another case of co-location between ARGs was observed between *sul1* and *aadA*. The contigs containing those two genes were found in D1 in VIL1, and EF in MUE3 (Fig. S2.13c; Fig. 2.7b). Unlike *aph(3'')-I* and *aph(6)-I* in VIL1 however, the dynamics of *sul1* was not similar to that of *aadA* especially in D5 and D8 in MUE3 where high abundance of *sul1* was observed while no *aadA* was assembled (Fig. 2.6b). In agreement with this observation, the sul1-containing contigs retrieved from D5 and D8 in MUE3 did not contain *aadA*, and this was the only type that was identified in those samples (Fig. 2.7b). This indicates a significant shift of the bacterial populations that yielded *sul1*-containing contigs between D1 and D3 in MUE3, with wastewater-derived contigs containing *sul1* – *aadA* pairs not persisting.

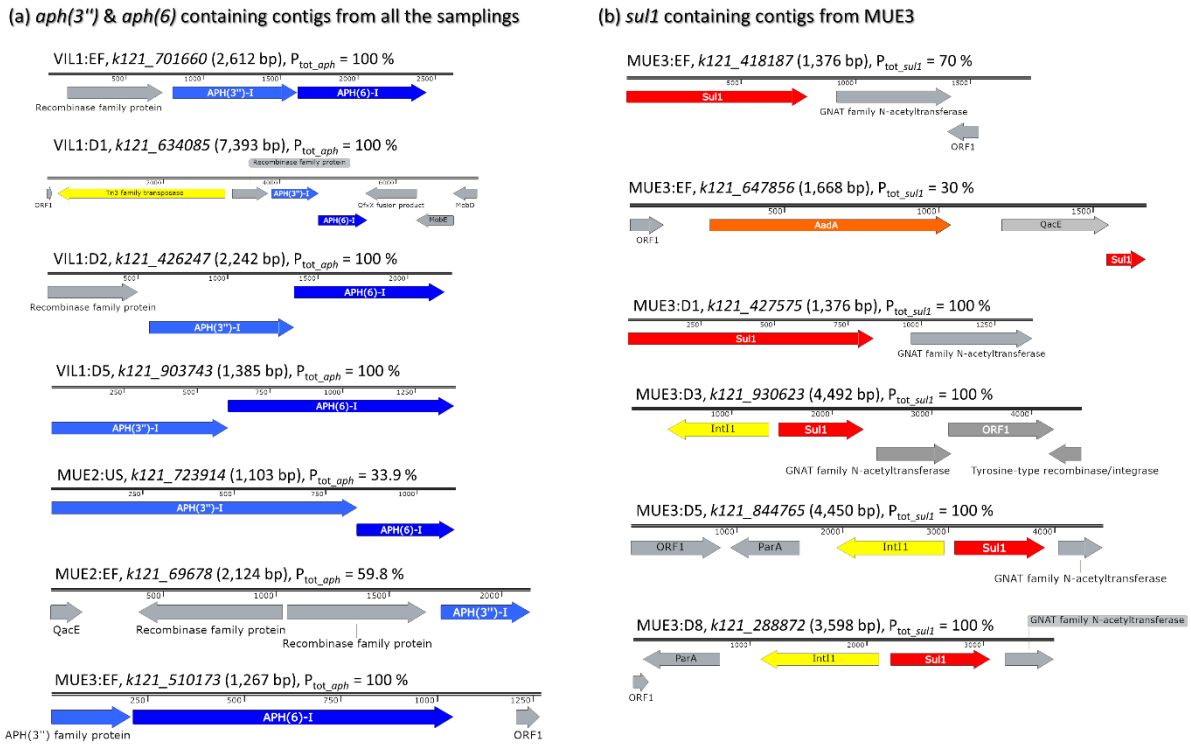


Figure 2.7. Gene arrangement on contigs containing *aph* and *sul1* genes. (a) Contigs containing *aph(3'')* and *aph(6)*, retrieved from all samplings (VIL1, MUE2 & 3). (b) Contigs containing *sul1* retrieved from MUE3. All annotated genes showed > 90.0 % percent identity (P_{Ident}) at the protein level to reference proteins, using DIAMOND protein search against NCBI nr protein database. The contig IDs are italicized (e.g. k121_XXXXXX). P_{tot_aph} indicates the proportion of average coverage for the *aph*-containing contig to the sum of average coverages for all the *aph*-containing contigs identified in the sample. P_{tot_sul1} denotes the proportion of average coverage for the *sul1*-containing contig to the sum of average coverages for all the *sul1*-containing contigs identified in the sample. Only contigs with lengths > 1,000 bp are shown.

2.3.9. *bacA*, an ARG with high natural prevalence in environmental bacteria

Unlike the other prevalent genes, the proportion of *bacA* (also known as UppP, undecaprenyl-diphosphate or -pyrophosphate phosphatase) was greatest in US samples, and was also abundant in many downstream locations (except D5 and D8 in MUE3 where *sul1* occupied the largest proportion) (Fig. 2.6a). In order to further assess whether *bacA* was intrinsic in our river water samples, we identified potential hosts by assigning taxonomy to the metagenome-assembled contigs using a combination of methods. The contigs for which all three methods agreed at the genus level are shown in Table 2.1. The four genera identified as potential hosts of *bacA* contigs derived from less-disturbed freshwater samples (US, D2, D5, D8) were *Pseudomonas*, *Acidovorax*, *Limnohabitans*, and *Aeromonas*. Among them, *Pseudomonas*, *Acidovorax*, and *Limnohabitans* are typical inhabitants of freshwater and soil environments (Peix et al., 2009;Willems, 2014). However, considering that the proportions of *bacA* in the contigs to the total *bacA* in the sample in terms of reads per kilobase (RPK) were low for river water samples (except for D8 in MUE3), we assume that homologues of *bacA* could be present in many other environmental bacteria. Thus, our data suggests that *bacA* is probably unsuited for tracking anthropogenic sources of antibiotic resistance.

2.3.10. Exploring the potential reasons for rapid increase of *sul1* in far downstream locations in MUE3

Both qPCR-based, and metegenomic analysis confirmed that *sul1* and *intI1* increased in the downstream of MUE, and especially strongly in one of the sampling campaigns (i.e., MUE3) between 5.0 – 6.8 km downstream distance (Fig. 2.2b & 6).

To figure out if there was a biological driver for this unexpected increase of *sul1* and *intI1*, we first characterized the *sul1*-containing contigs. The *sul1* gene is known to be highly mobilized and is often associated with *intI1* (Gillings et al., 2008;Gillings et al., 2015). Indeed, all contigs containing *sul1* associated with *intI1* retrieved from the river (D3 ~ 8) were homologs of a single dominant type that appeared to be also plasmid-associated as it contained the plasmid-associated gene *parA* (Davis et al., 1992). We could unsurprisingly not obtain a meaningful taxonomic assignment for these sequences. It could thus not be demonstrated whether the downstream increase in MUE3 was due to an increase in an EF-derived or an environmental organism or from a local contamination source. Further information could be obtained in future studies by isolation or construction of metagenome-assembled genomes. We therefore turned to chemical indicators to further study the potential for local or non-point sources of contamination as an explanation.

Table 2.1. Metagenome-assembled *bacA*-containing contigs to which taxonomy was successfully assigned at genus level. Taxonomy assignment was performed using Kaiju, Kraken2, and the basic local alignment tool for nucleotides (Blastn), and only contigs with consensus from all three approaches at the genus level are shown. For Blastn, the quality criteria were P_{Ident} > 90.0 %, and Q_{cov} > 90 %. P_{tot_bacA} indicates the proportion of *bacA* in the contig to the total *bacA* in the sample in terms of reads per kilobase. P_{Ident} indicates the percentage of identical match. Q_{Cov} indicates the query coverage.

Campaign	P_{tot_bacA} (%)	Contig Information			Kaiju	Kraken2	Blastn		
		Contig ID	Length (bp)	Coverage			Classification	P_{Ident} (%)	Q_{Cov} (%)
VIL1:US	2.4	k121_11403	455	4.0	*Pseudomonas* sp. Bc-h	*Pseudomonas azotoformans*	*Pseudomonas azotoformans* strain P45A	92.1	100
VIL1:US	2.2	k121_184413	333	3.6	*Pseudomonas cichorii*	*Pseudomonas cichorii* JBC1	*Pseudomonas cichorii* JBC1	97.6	100
VIL1:US	1.7	k121_761665	320	2.8	*Pseudomonas cichorii*	*Pseudomonas* spp.	*Pseudomonas cichorii* JBC1	90.2	99.1
VIL1:US	1.7	k121_867352	378	2.8	*Pseudomonas* spp.	*Pseudomonas fluorescens* SBW25	*Pseudomonas* sp. NS1(2017)	94.4	99.5
VIL1:US	3.4	k121_892755	1039	5.3	*Acidovorax temperans*	*Acidovorax* sp. 1608163	*Acidovorax* sp. 1608163	98.3	100
VIL1:US	1.5	k121_1008895	517	2.3	*Aeromonas* spp.	*Aeromonas* sp. CA23	*Aeromonas* sp. CA23	97.7	100
VIL1:EF	7.5	k121_372232	409	1.8	*Aeromonas* spp.	*Aeromonas media* WS	*Aeromonas media* strain MC64	99.5	100
VIL1:EF	20.0	k121_402216	594	4.7	*Aeromonas* spp.	*Aeromonas media* WS	*Aeromonas media* WS	99.7	100
VIL1:D1	4.9	k121_187307	4400	8.8	*Aeromonas media*	*Aeromonas media* WS	*Aeromonas media* WS	98.8	100
VIL1:D1	7.4	k121_695922	693	6.1	*Aeromonas media*	*Aeromonas media* WS	*Aeromonas media* WS	99.7	96.0
VIL1:D2	9.4	k121_274506	354	2.5	*Acidovorax* spp.	*Acidovorax* sp. 1608163	*Acidovorax* sp. 1608163	95.7	99.2
VIL1:D5	11.0	k121_307936	680	4.0	*Acidovorax* spp.	*Acidovorax* sp. 1608163	*Acidovorax* sp. 1608163	97.1	100
VIL1:D8	10.6	k121_916544	401	2.8	*Acidovorax* spp.	*Acidovorax* sp. 1608163	*Acidovorax* sp. 1608163	97.8	100
MUE2:D8	55.7	k121_6061	84229	12.0	*Aeromonas veronii*	*Aeromonas veronii* B565	*Aeromonas veronii* strain 17ISAe	93.3	95.8
MUE3:EF	69.6	k121_139490	591	3.3	*Aeromonas media*	*Aeromonas media* WS	*Aeromonas media* WS	99.5	100
MUE3:D1	11.3	k121_69935	869	3.6	*Aeromonas media*	*Aeromonas media* WS	*Aeromonas media* strain MC64	96.7	100
MUE3:D8	80.9	k121_528493	314	3.0	*Limnohabitans* sp. 63ED37-2	*Limnohabitans* sp. 63ED37-2	*Limnohabitans* sp. 63ED37-2	94.3	94.3

To evaluate non-point-source inputs of pollutants, we chose to evaluate a few micropollutants that may serve as indicators of contamination. Sulfamethazine (also known as sulfadimidine) is used in pig husbandry (Stoob et al., 2007), and mecoprop is a weed control agent used primarily in urban settings in Switzerland (Wittmer et al., 2010). It was assumed that the levels of these pollutants in downstream locations would increase or be persistently high if a pronounced agricultural or urban surface runoff existed, which could accompany resistance genes and bacteria potentially existing in agricultural or urban areas. In VIL, sulfamethazine was below detection (LOD ~ 0.8 ng/L) in all samples except one US sample, while in MUE there appeared to be a source in WWTP effluent especially during the MUE3 campaign, but the compound was not observed to increase in downstream locations. The concentrations and downstream dynamics of mecoprop varied between campaigns (Fig. S2.8). For VIL1~3 and MUE1 and MUE3 mecoprop concentrations were low (< 60 ng/L), while there seemed to be a strong, effluent-associated input for MUE2 and concentrations remained high further downstream (> 200 ng/L). A slight increase in the downstream range > 5km observed in MUE2 and between 1.0 – 2.0 km in MUE3 may be due to fluctuating input of mecoprop from the WWTP effluent. Concentrations did not further increase in the far downstream locations (D8) where the sudden increase of *sul1* and *intI1* was observed (Fig. S2.8). Based on these, but also the other analyzed micropopllutants, we found no evidence for significant downstream contamination sources. However, these chemical indicators are not conclusive, as the analyzed compounds were not a comprehensive selection to trace non-point sources in the downstream river section (e.g., from manure or pesticide applications, although these are not very likely under dry-weather conditions). So while we found no evidence for such contamination we can also not conclusively rule them out as an explanation for the marked *sul1* and *intI1* increase observed for MUE3.

Finally, the potential for in-situ resistance selection in the water was assessed using the concentration of antibiotics and metals in downstream locations in MUE3. Sulfamethoxazole and its derivative (N4-acetylsulfamethoxazole) were the antibiotics with the highest concentration among all the antibiotics analyzed, but downstream concentration (sulfamethoxazole in the range of 33 to 95 ng/L in MUE3) remained far below the published PNEC for resistance selection (e.g., 16,000 ng/L) (Bengtsson-Palme and Larsson, 2016). The concentration of trimethoprim, which is usually prescribed together with sulfamethoxazole, was also much lower than its PNEC for resistance selection (e.g. 500 ng/L) (Bengtsson-Palme and Larsson, 2016). Even though the vast majority of metals analyzed in this study remained below the limit of quantification or below their estimated minimum co-selective concentrations for dissolved metals in water ($MCC_{waterDC}$), the concentrations of two metals (i.e., copper and nickel) were higher than their $MCC_{waterDC}$ (1.5 µg/L for Cu, and 0.29 µg/L for Ni) (Seiler and Berendonk, 2012). However, $MCC_{waterDC}$ is a predictive value and actual

selective levels could be higher, also their levels in far downstream locations (D8) in MUE3 were not specifically higher than at other locations within the same sampling campaign, nor at the same locations than in other samplings where the increase of *sul1* or *intI1* was less pronounced (Dataset S12). Furthermore, we did not observe co-localization between sul1 and any other genes potentially conferring Cu, Ni or any other metal resistance based on contig-based co-localization search in MUE (Fig. 2.7b). Overall no convincing evidence for in-situ resistance co-selection by Cu and Ni as an explanation for the downstream increase of *sul1* and *intI1* was found. We further note that the estimated river retention time per km was relatively short (i.e., 51.4 and 49.5 mins/km for VIL and MUE, respectively, Dataset S4), which makes the likelihood of in-situ resistance selection in the water even less plausible.

As a final possible explanation we considered the possibility of cell migration from other river compartments to the water. According to qPCR enumeration of ARGs and *intI1* in surface sediments, we did not observe the increase of *sul1* and *intI1* levels in D5 and D8 in MUE3 in terms of either absolute and relative abundance (Fig. S2.6). Furthermore, the relative abundance of *sul1* and *intI1* in sediment was generally similar to, or lower than the values for water in MUE1~3 (Fig. S2.7). If sediment resuspension was a major source for aquatic *sul1* and *intI1* elevation in MUE3 D8, relative abundances of *sul1* and *intI1* in water samples would be expected to remain unchanged or to drop. While we could not completely exclude the possibility of contribution of sediment resuspension, we assume that there could be other sources (e.g., stream biofilms) where *sul1* and *intI1* were selectively enriched in terms of both absolute and relative abundance. Considering the downstream levels of both resistance determinants and nutrients remained relatively high in MUE due to high EF inputs and low downstream dilution effects, especially in the third campaign (Fig. S2.5 & Dataset S9), in-system growth is also a plausible hypothesis.

The reason for and the nature of the striking increase of *sul1* and *intI1* (but not of *tetW*, *ermB*, bla_{CTX}), during the MUE3 campaign thus remains open and would require further study to resolve. What the observation shows unambiguously, is that unexpected and perhaps not directly anthropogenic contamination-driven increases of ARGs are possible. As in particular *sul1* and *intI1* are commonly applied for tracking anthropogenic sources of ARG in the environment (Berendonk et al., 2015;Gillings et al., 2015), we caution that monitoring strategies should employ a multi-target strategy to be robust.

Conclusions

- Downstream levels of antibiotic resistance determinants decreased rapidly over 2.0 – 2.5 km distance due to dilution effects and decay over longer distance due to other removal mechanisms.

This would suggest that public exposure to wastewater-origin antibiotic resistance might be most acute only over short distances (few kilometers) from points of discharge, especially if a pronounced input of additional water inflows exists.

- We also observed at least one instance where *sul1* and *intI1* dynamically increased in the river, without being able to establish any link to a local anthropogenic contamination. Other river compartments where in-system growth of biomass could take place (i.e., stream biofilms) could be included as a monitoring target in future studies.
- Metagenomics-based resistome analysis yielded consistent conclusions with qPCR analysis of select targets (e.g., *sul1*) and also identified promising targets for future monitoring of anthropogenic sources of antibiotic resistance (e.g., *aph(3'')-I*, *aph(6)-I*, *mexT*, *tetQ*, and *aadA*). In general metagenomics, qPCR and cultivation-based assays yielded consistent trends. Public health advice could be based on quantifying indicator genes or technically simpler cultivation-based indicators.
- A weak structural correlation between resistome and microbiome, and low levels of (co)selective agents revealed a lack of driving forces in less-disturbed river waters (downstream over 3 km distance, plus upstream locations).
- We showed that contig-based taxonomic assignment and analysis of the genetic neighborhood of assembled ARG can reveal important, if limited, additional information about shifts in ARG host identities, mobilization, and co-localisation of ARG that would otherwise remain hidden.

Data Availability

The raw sequencing data both for metagenome and 16S rRNA amplicon sequencing are available at the European Nucleotide Archive under the project ID – PRJEB39697 for primary, and ERP123247 for secondary accession. All the other research datasets (including additional minor datasets that were not shown in this manuscript, and R codes) are available at the Eawag Open Research Data repository (https://opendata.eawag.ch/https://doi.org/10.25678/0003N6).

Author Contributions

J.L and H.B designed this study, and participated in all stages of the work as main authors. All authors provided feedback and inputs throughout field/laboratory works or manuscript writing stages. F.J, A.M.M, and K.B participated in field/laboratory works and biological data analysis. C.S.M, A.M, participated in designing the study, and performed chemical analysis of micropollutants. M.D.M, and F.F and helped design hydrological measurements and experiments and together with C.S. helped with analysis and interpretation of hydrological data. P.J.V, A.P, and C.S provided important input and a critical review of the manuscript.

Supplementary Information

Site description and field sampling

To avoid impact of high flow conditions all sampling was performed under dry conditions and at least 24 hours after the last precipitation in the catchment. The actual time between last precipitation and sampling was 37 hours for MUE2; 38 hours for VIL2; 41 hours for VIL3 and > 48 hours for the rest of sampling campaigns. The residence time of all the Swiss rivers is less than one day (24 hours) until they either reach a reservoir (lake) or leave the country (Ort et al., 2009), so direct precipitation effects were avoided by this strategy.

In the first sampling in Villeret (VIL1), water grab samples (5L in sterilized water containers) were obtained from 0.5 km upstream of the discharge point (US), wastewater treatment plant final effluent (EF), and downstream locations located at 0.5 km (D1), 2.5 km (D2), 4.5 km (D3), 6.5 km (D4), 8.3 km (D5), 10.5 km (D6), 12.4 km (D7), 13.7 km (D8) flow distance from the point of discharge. A small (watershed area < 3.0 km^2) tributary enters the Suze between sites D4 and D5 (see location S1 in Fig. S2.1). We obtained samples also from the location S1. The water depth was < 30 cm at D1, and no more than 1.0 m at the farthest downstream point (D8).

For campaign VIL2, all the sampling points from VIL1 and additional upstream locations and their small tributaries were sampled, for instance, the upstream locations at 1.3 km (US2), 7.5 km (US4) distance from EF discharge point, and a small upstream tributary located between US and US2 (see location S0 in Fig. S2.1).

For campaign VIL3, water sampling was performed at all the sampling points from VIL2 and one additional upstream point (US3, 2.8km) (Fig. S2.1). In VIL4, one additional upstream point located at 10 km from EF discharge point (US5), and additional small tributary near US4 (see location S3, Fig. S2.1) were sampled together with US, US1 ~ 4, EF, D1 ~ 4, and D8.

At MUE, we studied a 7.0 km stretch of the Murg downstream of WWTP Münchwilen without other WWTP inputs and one location upstream of the discharge point. One major tributary located between D5 and D6 (see location S5 in Fig. S2.2; watershed area > 3.0km^2; discharge approximately 3.8% of the receiving Murg in terms of mean runoff (m^3/d) estimated by FOEN, https://map.geo.admin.ch/), and four minor tributaries (watershed areas < 3.0) enter the Murg in the studied section.

In all sampling campaigns (MUE 1~3), water grab samples were obtained from 0.5 km upstream of the discharge point (US), wastewater treatment plant final effluent (EF), and downstream locations located at 0.5 km (D1), 1.0 km (D2), 2.0 km (D3), 3.0 km (D4), 4.0 km (D5), 5.0 km (D6), 6.0 km (D7),

6.8 km (D8) flow distance from the point of discharge. Water sample was obtained also from one of the side streams located between D5 and D6 (see location S5 in Fig. 2.2). The water depth was < 30 cm at D1, and no more than 1.0 m at the farthest downstream points (D8).

Contrary to VIL where EF was directly discharged to the main water body, the EF in MUE is discharged first to a small rivulet (S1), and the combined discharge (EF + S1) then enters the Murg after ca. 65 meters (Fig. S2.3). The proportion of EF in the side channel was estimated from the conductivities of S1, EF, and EF + S1 according to Ort and Siegrist (2009) to be in excess of 80%, therefore the contribution of S1 was not further considered. In addition, one sample was taken from a major downstream tributary (S5), of which mean runoff (m^3/s) comprises approximately 4%, estimated by a runoff-model run by BAFU (BAFU, 2000). GPS coordinates for all sampling locations are given in Table S1.

On site physicochemical measurements

In order to make sure that EF and river water were completely mixed at and after D1, the horizontal conductivity profiles of the river cross-section were measured at D1 in both systems using a portable multi-parameter probe (Multi 3630 IDS, WTW, Germany). No significant differences in conductivity were noted across the cross-section in any campaign (standard deviation < 0.5% of the average in conductivity profiles across cross-section). Four on-site parameters – Temperature (°C), conductivity (µS/cm), pH, and dissolved oxygen (DO, mg/L), were measured for each water sample as soon as it was taken using the multi-parameter probe (Dataset S9). The conductivity values were corrected to the temperature-compensated values (at 25 °C) using the default algorithms embedded in the instrument.

Sediment sampling

Surface (< 5cm of subsurface) sediment samples were obtained from 5 selected locations (US, D1, D2, D5, and D8) for campaigns VIL1 ~ 3 and from 5 locations (US, D1, D2, D5, and D8) for campaigns MUE1 ~ 3. For each sampling location, sediments were collected from at least 5 random locations regardless of their types and properties, and pooled in order to obtain a representative sample. To reduce eukaryotic DNA contamination We tried to avoid taking sediments from spots where there were a lot of algal growths, which we typically observed in MUE D5 and D8.

On-site hydrological measurements

To improve hydrological understanding of the discharge sites hydrological measurements were performed in August 2019 (August 23, 2019 for VIL, and August 27, 2019 for MUE) under dry conditions, comparable to conditions during sampling. River flow velocity (m/sec) was determined

using a flowmeter (Handgerät HFA up-Flowtherm, Höntzsch, Germany) at 150 and 500 m downstream locations in VIL, and at 150, 226, and 500 m downstream locations in MUE. At each location, three measurements were made over 3 different cross-sections (Dataset S4). Additionally, a tracer spike test was performed to confirm the measured flow rates according to the protocol modified from (Velísková et al., 2014). A pre-prepared NaCl (3 % w/v for VIL, and 4 % w/v for MUE) solution was poured into the river at the EF discharge point (20 L in VIL, and 25 L in MUE) and the time was noted. Conductivity was measured in a time series at 150 m downstream distance in VIL, and 226 m downstream distance in MUE (Fig. S2.10). Flow rate measurements and spike test results were generally in agreement, for instance, the expected arrival time was 437 ~ 489 sec in VIL, and 530 ~ 1,443 sec in MUE using the flow rates in Dataset S4 (0.31 ~ 0.34 m/s at 150 m distance in VIL; 0.16 ~ 0.43 m/s at 150 ~ 226 m distance in MUE).

Heterotrophic plate count of antibiotic resistant bacteria (ARB)

Other than the combinations of antibiotic described in the main text, agar further contained pimaricin (50.0 mg/L) to suppress fungal growth. Water samples were concentrated onto the 0.2 µm pore size membrane filters (Cellulose Nitrate Membrane Filter, Sartorious Germany). The sample filtration volume was pre-optimized in such a way that we obtained between 5 ~ 300 colonies per plate. Dilution was performed if needed, using 0.85% NaCl solution. The limit of detection varied by plate type, 0.5 CFU/mL for CLR/TET plates, and 5.0 CFU/mL for SMX/TMP/TET plates, because significant fungal growth was observed on SMX/TMP/TET plates when we filtered more than 10 mL volume. The total number of colonies on the plates were counted after 72 hours of incubation at 25 °C. Each sample was plated in triplicate and a sterile control plate receiving a sterile filter through which 10 mL of 0.85% NaCl was filtered was performed for each plate type for each sampling campaign.

Quantitative PCR analysis

Bacterial 16S rRNA genes, resistance genes *sul1* and *tetW*, and intergron-integrase gene *intI1* were analyzed by qPCR (Roche LightCycler 480 II, Roche, Switzerland) using LightCycler 480 Probes Master kit. Two other resistance genes (*ermB*, and bla_{CTX}) were quantified using LightCycler 480 SYBR Green I Master (Roche, Switzerland). Bacterial 16S rRNA genes (V3 – 4) were quantified as a proxy for the abundance of the bacterial community. Primer sequences and references are provided in Table S2. DNA extracts were diluted 1/10 to minimize PCR inhibition. Absence of PCR inhibition was further confirmed by comparing results for 1/10 with 1/100 diluted samples for selected samples (EF samples from VIL1 ~ 3, and MUE1~3) for a selected marker gene (*sul1*) (data not shown). The detailed PCR programs were described previously (Ju et al., 2019). At the end of program, a melt

curve analysis (65 – 95 °C) was performed for *ermB* and bla_{CTX} in order to confirm target specificity. The qPCR efficiencies for standard curves calculated using the equation, $E = 10^{(-1/slope)}$ using the algorithm embedded in the LightCycler® 480 Software 1.5. (Roche, Switzerland). The qPCR standard curves were prepared ranging from 50 to 50,000,000 copies, and measured repeatedly in quintuplicate in every batch runs. Determination of limits of detection and quantitation for qPCR assays was modified from (Czekalski et al., 2012; Czekalski et al., 2014), the most diluted standard (i.e., 50 copies) was always set as LOD except for bla_{CTX} where the S.D. of Cq among quintuplicates at 50 copies was too large (i.e., S.D. = 1.6) and the second-most diluted standard (i.e., 500 copies) was chosen as LOD. In case the S.D. of Cq values at LOD was > 0.5, the value range between LOD and the next most diluted standard of which S.D. was ≤ 0.5 was regarded as less reliable, and highlighted in red in Dataset S8. The selected key information for qPCR validation recommended by MIQE guideline is given in Table S5 (Bustin et al., 2009). The DNA extracts from both filter replicates of each sample were analyzed in triplicate. Any measurements of which S.D. of Cq among technical triplicates was > 0.5, or the mean was below LOD were indicated as "Detected, but not-quantifiable". Furthermore, if ≥ 2 out of 3 technical triplicates were negative, or the average Cq values were > the values for negative controls, the target was considered as "Not-detected (N.D.)". Negative controls (i.e., molecular-grade distilled water) were checked in triplicates in every batch runs, and were always below LOD (Table S5). The differences between biological duplicate values were visually displayed in Figures as error bars. An experiment control was prepared by filtrating 500 mL sterilized nanopure distilled water through a membrane filter followed by DNA extraction. The DNA concentration was below the detection limit (< 0.2 ng/µL) of Qubit (Qubit™ dsDNA High Sensitivity Assay Kit, Thermo Fisher Scientific, USA) in our settings, and no contamination neither during filtration nor extraction was confirmed by measuring qPCR using selected markers, i.e., N.D. for *ermB*; D.N.Q. for *sul1* (36.0), *intI1* (Cq of 39.5), *tetW* (Cq of 37.7), and 16S rRNA gene (Cq of 33.2) (Table S5).

Metagenome and 16S rRNA amplicon sequencing analysis

The Illumina Novaseq6000 with a paired-end (2 × 150) strategy was used for shotgun sequencing. The basic filtering of metagenome raw data was performed by Novogene as follows: 1) Removing the reads that contained adapters, 2) Removing reads that contained undetermined base calls N > 10 %, and 3) Removing the low quality reads (quality score ≤ 5). The quality of reads was doubled-checked by the authors using FastQC v0.11.4 (Andrews, 2010). Reads with mean quality scores below 20 were removed by Prinseq-lite v0.20.4 (Schmieder and Edwards, 2011). The filtered reads were *de-novo* assembled using MEGAHIT v1.1.3 (Li et al., 2015) to produce a de Bruijn graph with multiple k-mer sizes (21, 47, 71, 95, and 121). Assembly statistics are given in Table S3. Open reading frames (ORFs) were predicted from the assembled contigs using Prodigal v2.6.3 with the default parameters (Hyatt

et al., 2010). As a global protein search algorithm, BLASTP v2.2.30 was utilized to annotate the predicted ORFs against two curated protein databases for ARGs, the Structured Antibiotic Resistance Genes (SARG) v2.0 (Yin et al., 2018) and Comprehensive Antibiotic Resistance Database (CARD) v3.1.0 (Alcock et al., 2020) with the e-value of 1×10^{-5}. ORFs matching ARGs with > 85.0 % identity and > 100 bp alignment lengths were retained for downstream analysis. We compared the results obtained with both databases in terms of the alpha diversity. Even though we observed agreement in many samples, the ARGs identified with each database could be either higher or lower than the other for some samples, implying each database might have both pros and cons (Fig. S2.12). We finally favored SARG v2.0 over CARD v3.1.0 for analyzing our datasets, because some gene homologues that we were interested in and that originated from the NCBI-nr database (i.g., *bacA*, *mexT* homologues) were identified only with SARG v2.0. Therefore, all the downstream data analysis reported in the manuscript was performed using the SARG v2.0 results.

Read mapping to contigs and ORFs was performed to calculate coverages. For read mapping, Bowtie2 was used (Langmead and Salzberg, 2012), and the depth information was generated using Samtools (Li et al., 2009). The coverages for contigs and ORFs were calculated from the depth information according to Albertsen et al. (2013). The coverage information was used as a basis for calculating abundance metrics of metagenome data. We primarily report relative abundance as GPM (genes per million) and absolute abundance as GPL (genes per liter), as previously established (Ju et al., 2019;Katz et al., 2010;Li and Dewey, 2011) (Dataset S5 & S6). The description of all metrics are given in Table S4. The Python scripts used in our workflow and a training data set are available from the first author's personal GitHub page (https://github.com/myjackson).

For analyzing the details of ARG-containing contigs, all the ORFs located within the same contig in which ARGs were found were annotated against the NCBI nr Protein Database (retrieved at March 4, 2020) (Coordinators, 2017) using DIAMOND v0.9.3 (Buchfink et al., 2015). Gene arrangement on the contigs was visualized using the genome viewer SnapGene® v5.1.3.1 (from GSL Biotech; available at https://www.snapgene.com/).

To assign taxonomy to selected contigs, two taxonomy classifiers for metagenomics were used – Kaiju v1.7.2 (Menzel et al., 2016), and Kraken2 (Wood et al., 2019). Kaiju was operated with the e-value of 0.05 using the NCBI nr Protein Database (Coordinators, 2017), and Kraken2 was operated using the default confidence score (0.0) and the Standard Kraken 2 Database (O'Leary et al., 2016). Even though both pipelines were shown to be accurate for assigning taxonomy to metagenome reads, the reliability for analyzing assembled contigs has not been established. Therefore, all contigs which had a consensus between Kaiju and Kraken2 results were additionally subjected to BLASTN searches against the nucleotide collection (nt) database (Coordinators, 2017) using the entire

sequences of contigs. Results from all three approaches were checked for consistency of the taxonomic assignements.

Amplicon sequencing of the V3 – 4 region (466 bp) of 16S rRNA gene was carried out by Novogene (https://en.novogene.com/). Briefly, 16S rRNA genes were amplified using Primers 341F – CCTAYGGGRBGCASCAG and 806R – GGACTACNNGGGTATCTAAT, and sequenced using the Illimina HiSeq2500 with a paired-end (2 × 250) strategy. The PCR amplification was performed together with a distilled water negative control, and the absence of contamination during amplification was confirmed by gel electrophoresis. For analyzing sequencing data, barcodes and primer sequences were removed from the raw sequences by Novogene, and the de-multiplexed sequences were used as input for downstream analysis. The DADA2 pipeline implemented in R was used for modelling and correcting the amplicon data (Callahan et al., 2016). We followed the work-flow suggested by the developers (https://benjjneb.github.io/dada2/tutorial). Firstly, the sequences were filtered with the following quality parameters: Quality score 2 (truncQ=2), the maximum expected error rate 2 (maxEE=2), and default values for the others. Then, the error rates were inferred by DADA2 algorithm, and the sample sequences were inferred using the previously calculated error rates to produce Amplicon Sequence Variants (ASVs). The paired reads were merged, and bimeras (two-parents mis-merged chimeras) removed. Taxonomy was assigned to ASVs using the DADA2-formatted Silva v132 database (Quast et al., 2013), the bacterial sequences were sub-selected, and used for further statistical analysis. The Shannon alpha-diversity index was calculated from ASVs abundance data using the R package Vegan (Oksanen et al., 2013). The ASV data was normalized at 50,000 sequences across samples (originally varied from 46,940 to 67,822) prior to the beta-diversity analysis. Non-metric multidimensional scaling (stress < 0.15) followed by Procrustes analysis was also performed using the normalized ASV sequences and relative abundance (GPM) of ARGs. The statistical significance (P-value) for Procrustes analysis was calculated using permutation test (number of permutation = 999).

Micropollutants analysis

All the water samples from VIL1 ~ 2 and MUE1 ~ 3, and selected samples from VIL3 (US, EF, D1, 2, 5, 8) were stored at 4 °C in the dark as soon as those were obtained on-site, and transported to the laboratory on the same day. Upon arrival at the laboratory, the 150 mL samples were stored at -20 °C in the dark until analyzed. The following 26 micropollutants were measured afterwards: 13 antibacterials and their derivatives (amoxicillin C and F, azithromycin, ciprofloxacin, clarithromycin, metronidazole, N4-acetylsulfamethoxazole, sulfamethoxazole, norfloxacin, sulfamethazine, sulfapyridine, triclosan, trimethoprim), 10 other pharmaceuticals (amisulpride, candesartan, carbamazepine, citalopram, diclofenac, hydrochlorothiazide, irbesartan, metoprolol, tramadol,

venlafaxine), and 3 industrial and biocidal chemicals (benzotriazole, mecoprop, 4/5-methylbenzotriazole).

After thawing the samples at room temperature for 6 h, samples were vigorously shaken. After 10 min settling, 1 mL supernatant was transferred into a 1.5 mL sample vial and spiked with labeled internal standards (IS, addition of 40 µL IS mixture for a final conc. of 400 ng/L). Large volume direct injection of a 100 µL sample was performed on an Agilent 1290 UHPLC equipped with an Acquity HSS T3 column (1.8 µm, 3.0x100 mm, Waters) for chromatographic separation. The HPLC was operated at a flow rate of 500 µL min^{-1}, an oven temperature of 30°C, with a gradient of 100% eluent A (nanopure water plus 0.1% formic acid) to 95% eluent B (methanol plus 0.1% formic acid) in 18.5 min, hold for 3.5 min, rise to 100% eluent A in 0.5 min, and hold for 4.5 min (including 2 min post run time). A triple quadrupole MS/MS (Agilent TQ6495B) was used for detection (electrospray ionization in switching mode, 3.5 kV in positive and 3.0 kV in negative mode, dynamic MRM with 650 ms cycle time and mass resolution of 0.7 Da, cell accelerator voltage 5 V; further MS/MS seetings in Table S6).

The calibration range was between 0.5 - 7500 ng/L. A separate calibration was done for ciprofloxacin and norfloxacin, where solutions were acidified with 0.1% formic acid. The calibration standards were prepared in a matrix water similar to tap water, containing 75 mg/L Ca2+, 11 mg/L Mg2+, 9.2 mg/L Na+, 1.7 mg/L K+, and 4 mg/L urea (resulting in a DOC of 0.8 mg/L), with a conductivity of around 600 µS/cm.

Quantification was done using labeled internal standards (IS). For calculation of the limit of quantification (LOQ), the lowest calibration standard with S/N > 10:1 was divided by the matrix factor. The matrix factor was calculated from the area of the IS in the sample divided by the average of the areas of the IS in the calibration row. The highest LOQ value calculated for all 65 samples was determined for each matrix ("worst case" LOQ). For recovery calculation, 17 samples were spiked with analytes (100 ng/L for downstream / upstream samples and 250 ng/L for WWTP effluent samples) and the relative recovery calculated (concentration of spiked samples minus concentration of sample, divided by the concentration of the spiked standard). For amoxicillin C and amoxicillin F, where no own labeled IS was available, concentrations were corrected by the recovery. LOQ and recoveries are list in the Dataset S10 & S11. For quality control, two transitions and retention time deviation were recorded. A tolerance of 80-120% was allowed for the qualifier recovery calculated as the average of ratios between the quantifier and qualifier transition over all calibration standards compared with the quantifier/qualifier ratio in each sample. The tolerance of the retention time deviation, i.e. the difference between the retention time of the analyte and the corresponding IS in the same sample, was 0.15 min.

Ions, nutrients, and heavy metals analysis

Cations (Na^+, NH_4^+, K^+, Ca^{2+}, Mg^{2+}), anions (Cl^-, NO_2^-, NO_3^-, PO_4^{3-}, SO_4^{2-}), dissolved nutrients ($P\text{-}PO_4^{3-}$, total dissolved phosphorus, $N\text{-}NH_4^+$), and dissolved organic carbon (DOC) were measured for all the collected water samples. Dissolved metals (Na, Mg, K, Ca, Cr, Mn, Co, Ni, Cu, As, Ag, Cd, Pb) were analyzed for selected samples (i.e., VIL3, MUE1 ~ 3). The analysis was performed according to the protocol described in Ju et al. (2019). In short, an aliquot of each water sample was filtered on-site using 0.45 µm cellulose acetate filters for measuring cations, anions, and nutrients. For measuring metals and DOC, an aliquot of each water sample was filtered using GF-filters (pre-burnt at 450 °C for 15 mins), and acidified (pH < 2.0) using 65% HNO_3 on-site and cooled at 4 °C in the dark during transportation. All the pretreated samples were stored at 4 °C in the dark upon arrival at the laboratory until analyzed. Cations and anions were measured using ion chromatography (IC), and metals were measured using high-resolution inductively coupled plasma mass spectrometry (ICP-MS). The results for Na, Mg, K, Ca were cross-compared between IC and ICP-MS, and no significant differences between them were observed (Dataset S12). DOC was measured using a total organic carbon analyzer (TOC-L, Shimadzu, Japan), and nutrients were measured by flow-injection analysis (ISO 13395, SAN++, Skalar, The Netherlands). The detailed protocols were described in the previous study (Ju et al., 2019). A distilled water negative control was also measured to confirm absence of contamination (< LOQ for all parameters).

Estimating the dilution effect on downstream levels of resistance determinants

As stated in the main text, we derived the Eq.1 according to the following procedure. Assuming there are multiple routes of water inflows (i.e., groundwater and/or tributary inputs) in between the upstream location A and downstream location B. Under the mass-conservation assumption, the following relationship should meet:

$$\boldsymbol{C_A Q_A + \sum_{k=1}^{n} C_k Q_k = C_B (Q_A + \sum_{k=1}^{n} Q_k)}$$

$$\boldsymbol{\therefore C_B = \frac{C_A Q_A + \sum_{k=1}^{n} C_k Q_k}{Q_A + \sum_{k=1}^{n} Q_k}} \quad \textbf{(Eq.5)}$$

Where, C_A: Concentration (mass or copies/volume) of a pollute at the upstream location A, C_B: Concentration of a pollute at the downstream location B, C_n = Concentration of a pollute at one of the additional routes for water inflows, Q_A: River discharge (volume/day) at the upstream location A, Q_B: River discharge at the downstream location B, Q_n = River discharge at one of the additional routes for water inflows

On the other hand, the average rate of concentration change of a pollute between the locations A and B "(C_B-C_A) / C_A" could be expressed as the following equation by substituting :

$$\frac{\boldsymbol{C_B - C_A}}{\boldsymbol{C_A}} = \frac{\boldsymbol{Eq.5 - C_A}}{\boldsymbol{C_A}}$$

$$= \frac{\boldsymbol{C_A Q_A + C_1 Q_1 + C_2 Q_2 + \cdots C_n Q_n - C_A Q_A - C_A \sum_{k=1}^{n} Q_k}}{\boldsymbol{C_A (Q_A + \sum_{k=1}^{n} Q_k)}} \qquad \textbf{(Eq.6)}$$

On the other hand, we assume most of additional water inflows are groundwater inputs because huge tributaries were not expected in our study distance in both sites by mean runoff (m^3/d) models run by BAFU, the Swiss Federal Office for the Environment (BAFU, 2000), nor by our observation while performing samplings. We assume the quality of groundwater inflows might not be significantly different over study distance. This assumption leads to:

$$\boldsymbol{\frac{1}{n}\sum_{k=1}^{n}(C_k - \bar{C})^2 \approx 0, \therefore C_1 Q_1 + C_2 Q_2 \cdots + C_n Q_n \approx \bar{C} Q_E,}$$

Where, Q_E indicates the total quantity of additional water inflows (= $\sum_{k=1}^{n} Q_k$)

Therefore, the Eq.6 could be reduced to:

$$\boldsymbol{\frac{C_B - C_A}{C_A} = \frac{\bar{C} - C_A}{C_A}\left(\frac{Q_E}{Q_A + Q_E}\right)} \quad \textbf{(Eq.7)}$$

Where, $\bar{C}$ indicates the representative concentration value for additional water inflows

The Eq.7 has following two major implications:

1) The rate of change for the downstream concentration of a pollute depends on its concentration in additional water inflows ($\bar{C}$).

2) If $\frac{C_B - C_A}{C_A}$ is normalized by $\frac{\bar{C} - C_A}{C_A}$ (i.e., $\frac{C_B - C_A}{C_A} \times \frac{C_A}{\bar{C} - C_A} = \frac{C_B - C_A}{\bar{C} - C_A}$), the remained term on the right [$\frac{Q_E}{Q_A + Q_E}$, hereafter termed as $DP_{A \to B}$] indicates the proportion of additional inflow quantity to the whole river discharge at the downstream location B.

After rearranging we obtain equation Eq.1 in the main text. We used DP for conservative tracers as a parameter for assessing dilution effects as this term for those parameters might not be over- or under-represented for those tracers. The derivation procedure above was graphically abstracted in Fig. 2.1.

Our first choice for conservative tracer was sodium. In order to confirm the mass-conservation assumption, sodium input loadings (EF + US) were compared to downstream loadings. The loadings were calculated by multiplying the concentration (mass/volume) of sodium by measured or

estimated river discharge (*Q*, volume/day). For instance, there were two gauging stations in VIL, one near US, and the other near D8, and we obtained *Q* from there (AWA, 2019, Jan. 11;FOEN, 2020, Sept. 02), also EF discharge (Q_{EF}) from the WWTP-VIL (Dataset S1). Then, those *Q* values were multiplied with corresponding concentrations to finally produce the sodium loadings in US, EF, and D8, VIL. In MUE, there was only one gauging station near D4 (FOEN, 2020, Sept. 02). The *Q* at US (Q_{US}) was estimated by the following equation using the dilution factor suggested by Ort and Siegrist (2009):

$$\boldsymbol{Q_{US} = Q_{EF} \times (DF - 1)} \qquad \textbf{(Eq.8)}$$

Where, $DF = \frac{Na_{EF} - Na_{US}}{Na_{D1} - Na_{US}}$; $Na_{EF,US,or\ D1}$ = Concentration of sodium at EF, US, or D1

The calculated sodium loadings are given in Fig. S2.9c. As shown in Fig. S2.9c, the sodium loadings increased somewhat with downstream distance, which indicates the sodium mass discharged from EF is conserved and some additional sodium inputs accumulate over the river continuum. Thus, we calculated DP over short distance (D1 to D2 for VIL; D1 to D3 for MUE) for sodium, and compared with the values for micropollutants. The somewhat increased sodium loadings might indicate there were significant inputs of sodium from additional water inflows as evidenced also by the measured sodium concentrations in upstream locations and tributaries. Sodium salts are widely distributed over the terrestrial environments, and highly soluble. Thus, sodium easily leaches to the groundwater and surface water.

In order to calculated DP, the representative concentration of water inflows ($\bar{C}$ in Eq.1) should be estimated so that DP could not be underestimated. For this purpose, the sodium levels for water inflows ($\overline{Na}_{in}$) were inferred as shown in Eq.2 in the main text. The calculated DP value for sodium was compared with the values for micropollutants except for benzotriazole which was consistently present in significant concentrations in upstream locations and tributaries, especially in VIL (Dataset S9). There were two additional micropollutants which were detected in upstream locations and tributaries, i.e. 4/5-methylbenzotriazole and diclofenac (Dataset S10 & 11). However, the upstream levels were not significant compared to their levels in downstream locations (< 5%) in most samplings, except for VIL3 and MUE2 for 4/5-methylbenzotriazole, and MUE2 for diclofenac. Thus, we considered those levels at US generally as insignificant, and set their $\bar{C} \approx 0$. According to the cluster analysis based on Bray-Curtis dissimilarity distance, 4/5-methylbenzotriazole and carbamazepine showed the most similar values compared to sodium (Fig. S2.9a), which indicates that these might be two most conservative tracers among the micropollutants. Indeed, their loadings did not change between D1 compared to distant downstream locations (D8 for VIL; D4 for MUE) (Fig. S2.9d). In contrast, diclofenac and hydrochlorthiazide showed the most dissimilar concentration

profiles compared to sodium, and the highest values of DP over short distance in general (Fig. S2.9a & S2.9b). Their loadings at D8 (VIL) or D4 (MUE) were decreased in most samplings, which indicates significant additional sinks (i.e., degradation via water chemistry, solar radiation, microbes, and/or sorption to organics/inorganics in the river system), as also known from literature. This loss of mass leads to an over-estimation of DP.

The estimated *Q* according to Eq.4 was cross-compared with the values that we obtained from gauging stations, and it was confirmed that the estimated *Q* values were in agreement with actual measured data from gauging stations (Fig. S2.9e). Among Q values estimated using different conservative tracers, sodium yielded the best prediction of measured *Q*, so we relied on this tracer in our calculation in Fig. S2.9e. This estimated *Q* was used when we calculate the ARG and *intI1* loadings over short distances (D2 for VIL; D3 for MUE) and far downstream distance (D8 for both sites) (Fig. 2.3b, c). The calculation of ARG and *intI1* loadings is analogous to the way we calculated the loadings for conservative tracers (i.e., loadings = concentration of an indicator × river discharge).

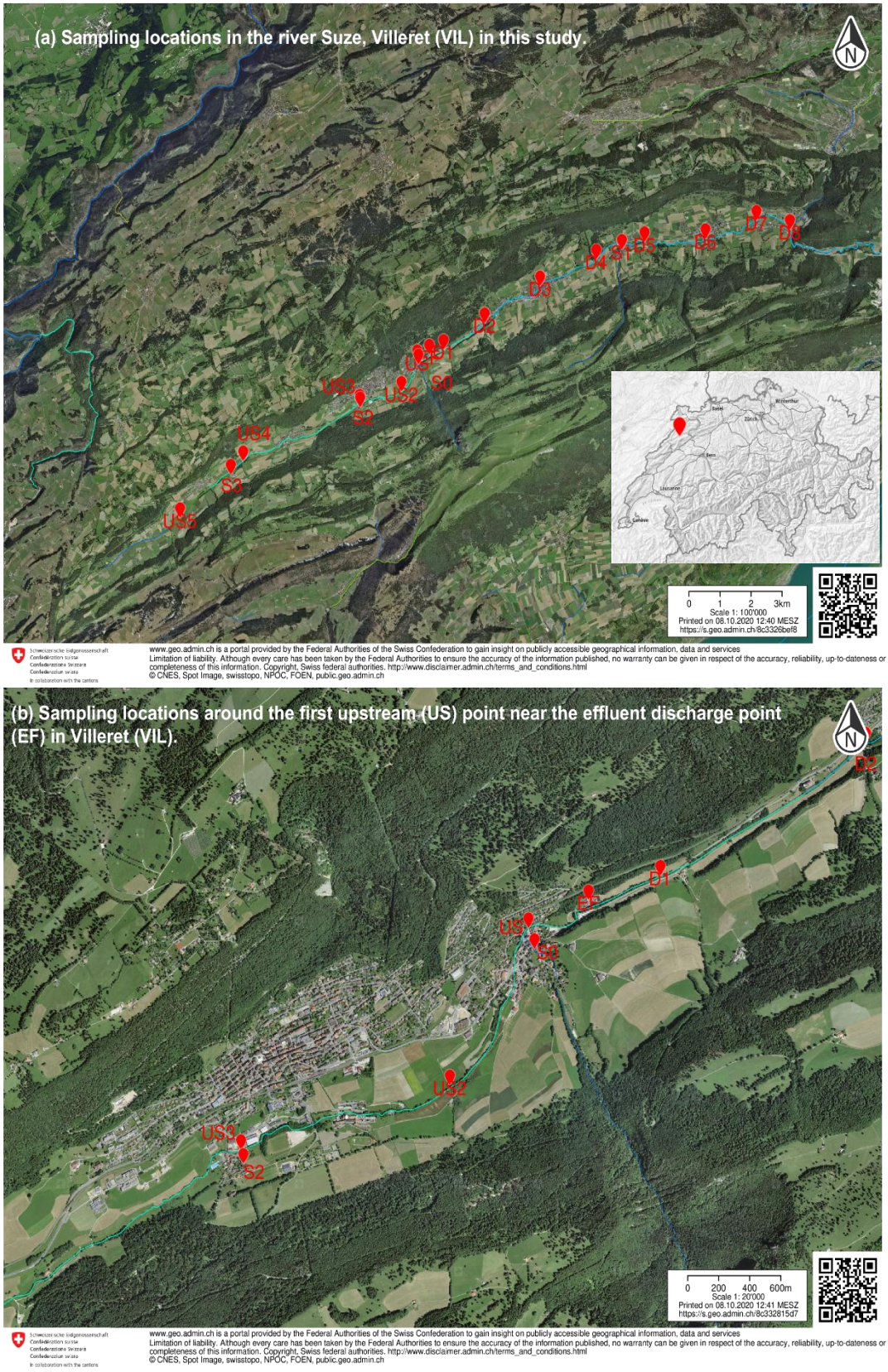


Figure S2.1. Sampling locations in the River Suze, Villeret (VIL) in Switzerland. BAFU (2019) Swiss water topographical catchment areas: Main river section, Bundesamt für Umwelt (BAFU), CH-3003 Bern.

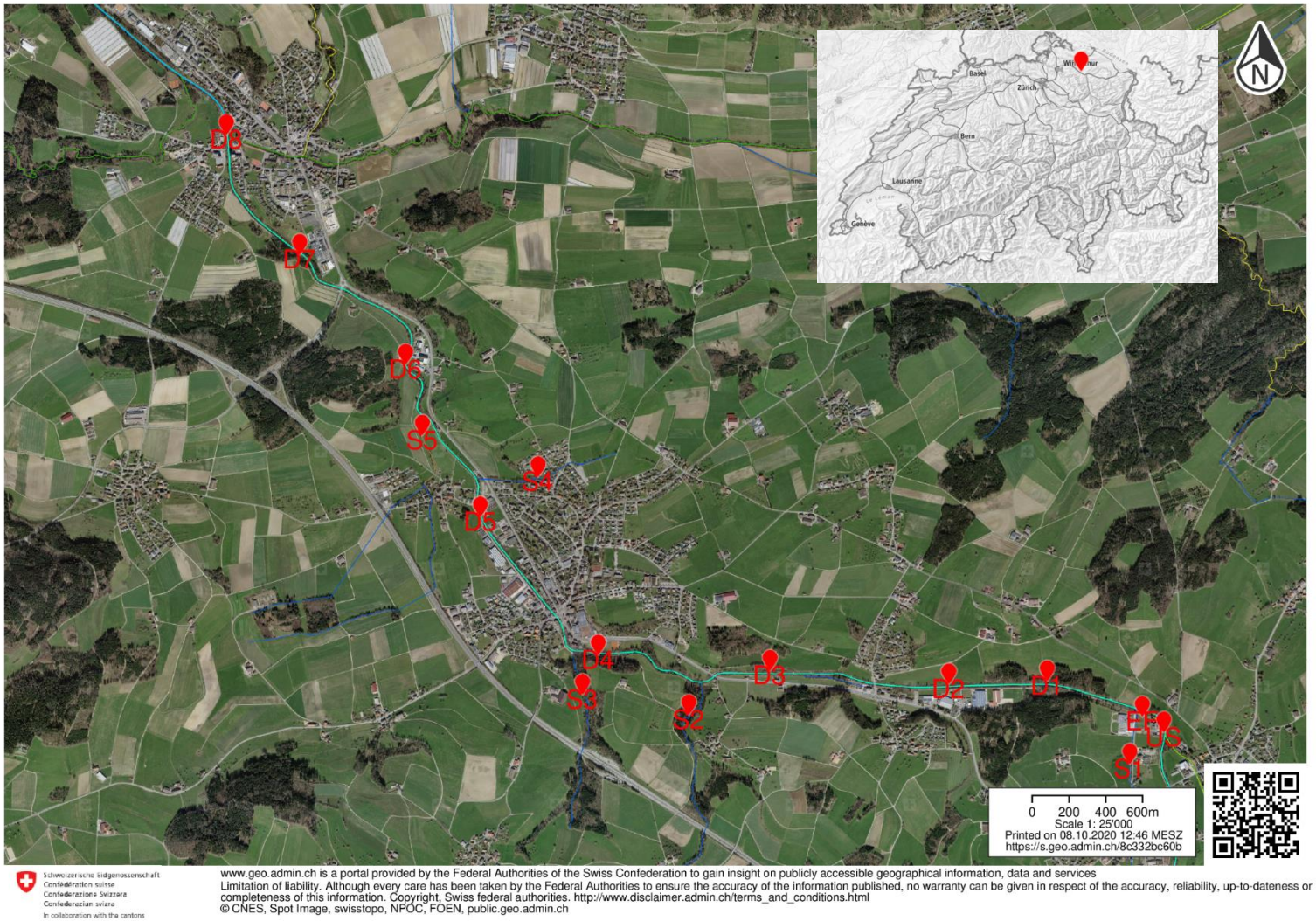


Figure S2.2. Sampling locations in the River Murg, Münchwilen (MUE) in Switzerland. For side streams, only 1 representative location (S5) was selected for sampling in MUE. BAFU (2019) Swiss water topographical catchment areas: Main river section, Bundesamt für Umwelt (BAFU), CH-3003 Bern.

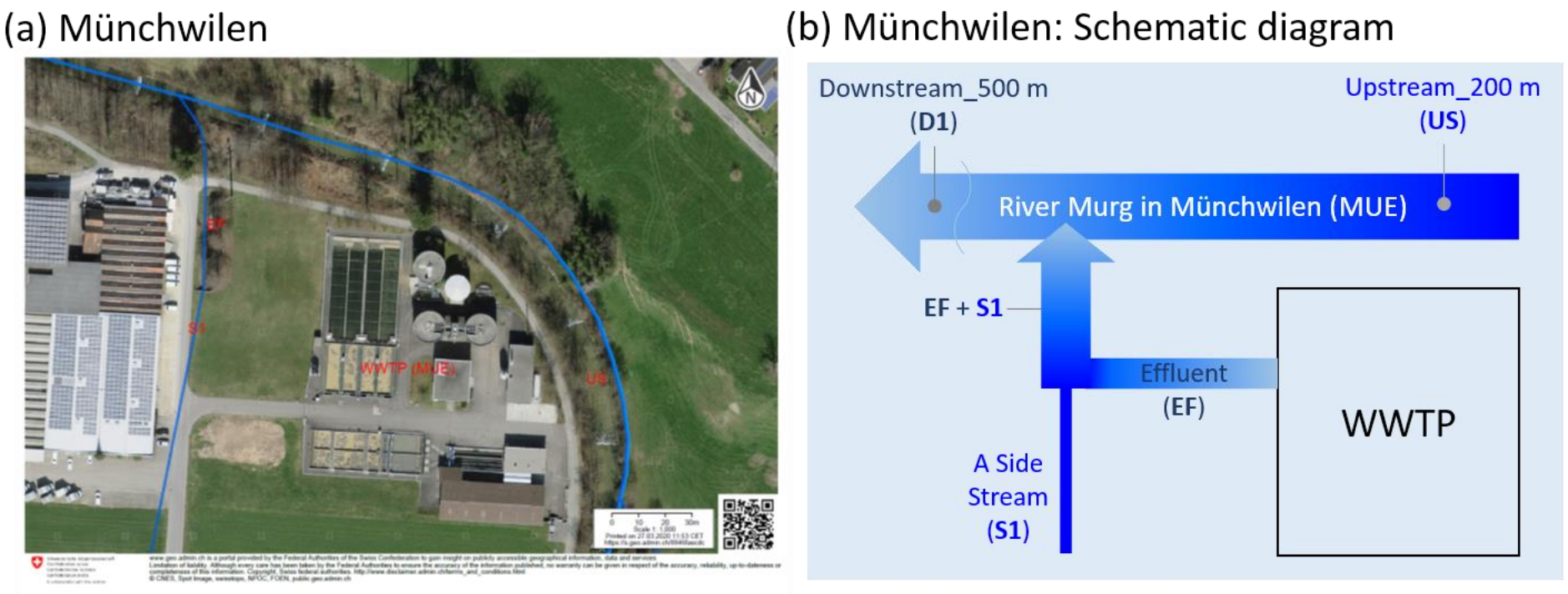


Figure S2.3. Sampling locations around the treated wastewater discharge point in the river Murg in Münchwilen, Switzerland. (a) Map obtained from the Swiss Federal Office for the Environment (https://map.geo.admin.ch/). (b) A schematic diagram explaining the sampling locations for US, EF, and D1. The proportion of EF to S1 was 8:2 – the EF + S1 portion was approximated with EF. BAFU (2019) Swiss water topographical catchment areas: Main river section, Bundesamt für Umwelt (BAFU), CH-3003 Bern.

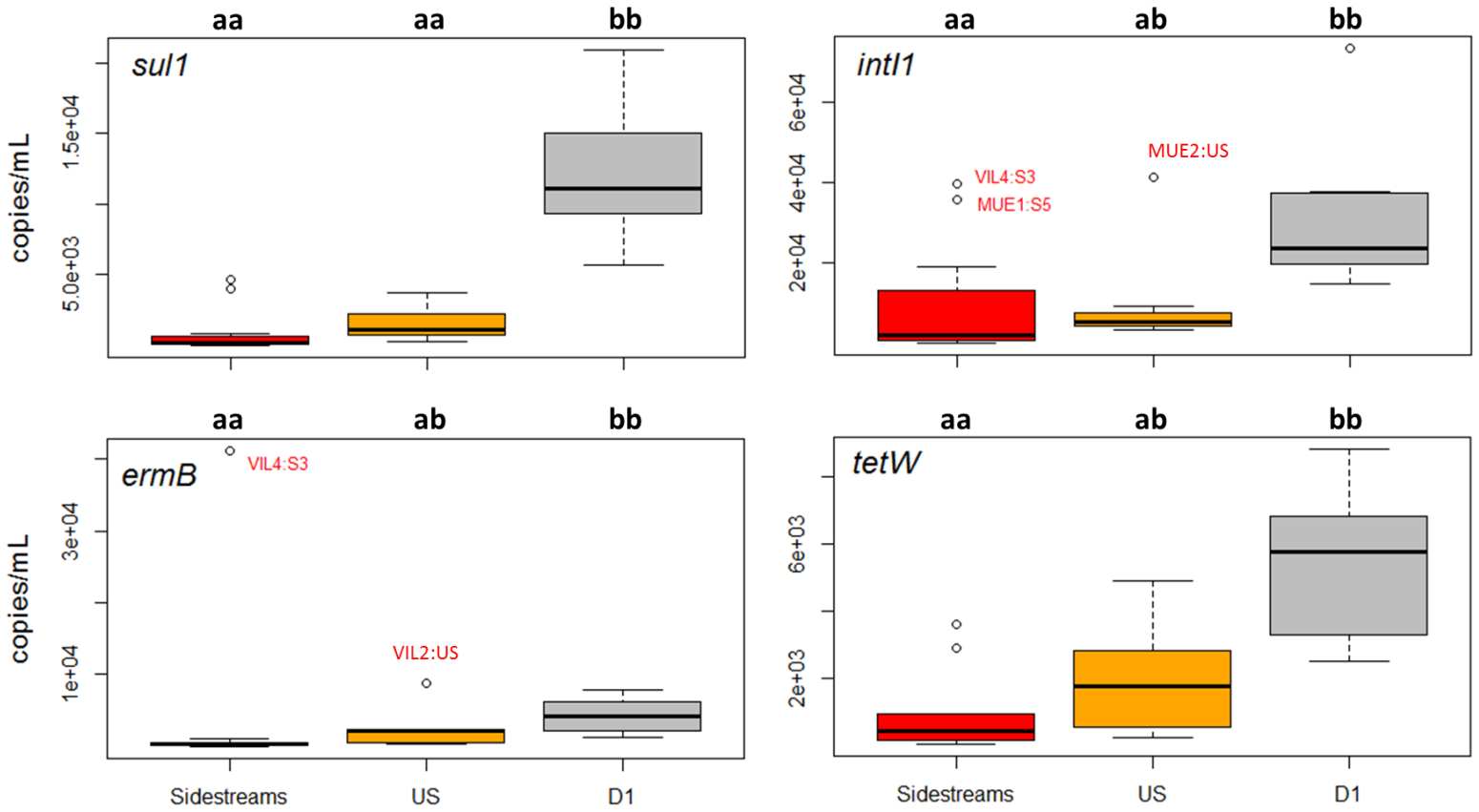


Figure S2.4. The comparison of levels of ARGs and *intI1* between sidestreams (S1 ~ 3 in VIL, and S5 in MUE), upstream near EF (US), and 500m downstream locations (D1). The treatments (Sidestreams, US, D1) that share the same alphabet are not significantly different each other according to the Kruskal–Wallis test, and the following post-hoc test (Dunn's test, p-adjustment by the Benjamini-Hochberg method) at 5% significance level.

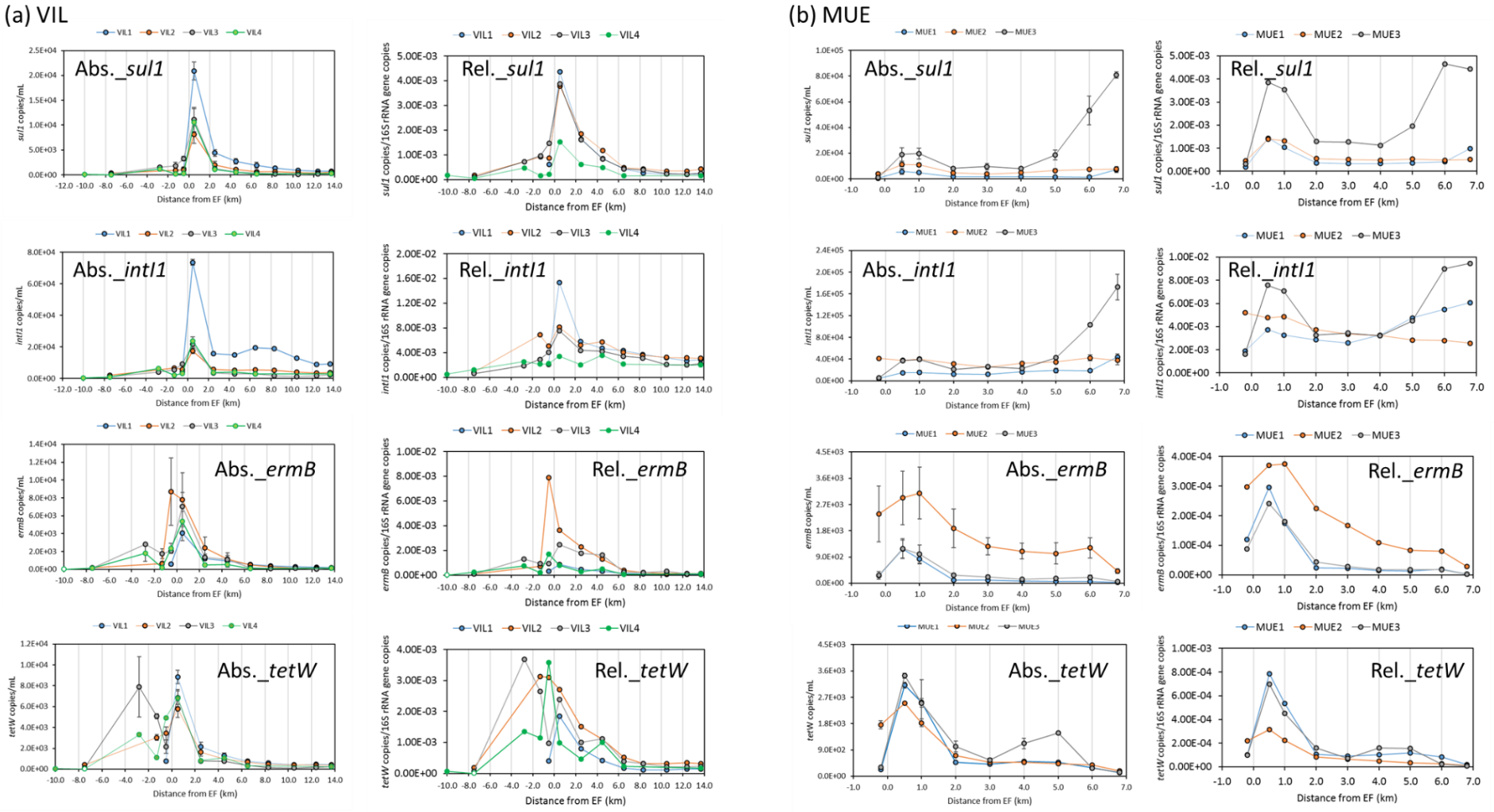
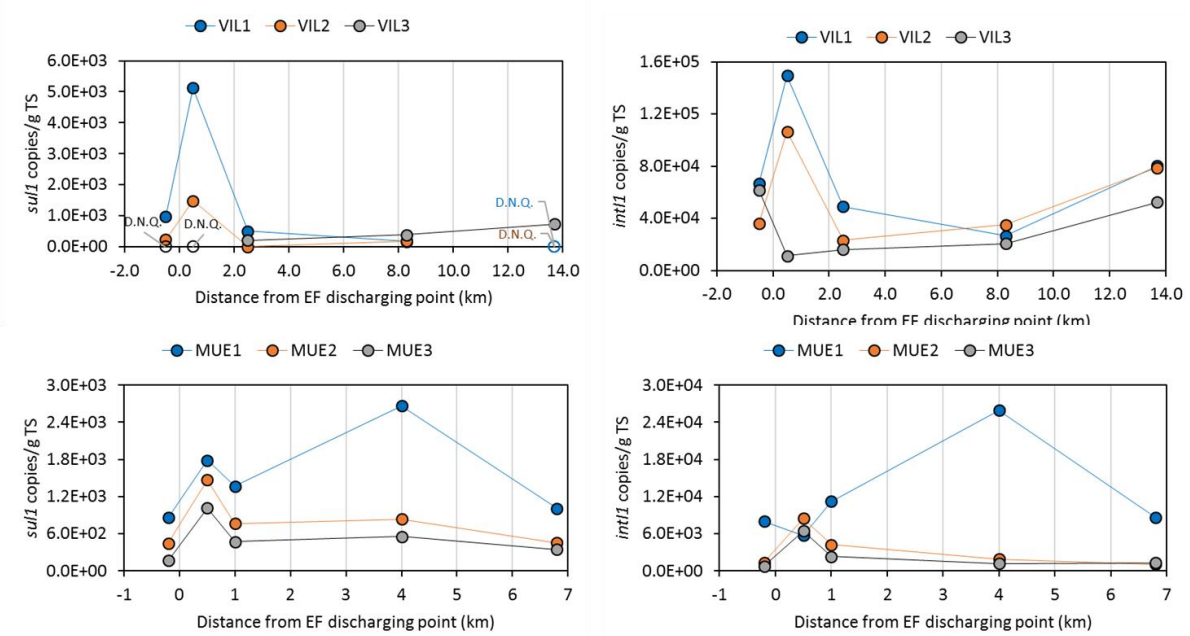


Figure S2.5. Levels (gene copies/mL) and relative abundance (copies/16S rRNA gene copies) of ARGs (*sul1*, *ermB*, *tetW*) and *intI1* in the upstream, and downstream river water quantified by qPCR. The levels (Abs.) and relative abundance (Rel.) of ARGs and *intI1* in Villeret (VIL) up to 13.7 km (a), and in Münchwilen (MUE) up to 6.8 km (b). The tips of error bars for levels (Abs.) indicate lower and upper values of biological duplicates.

(a) Level (copies/g TS)



(b) Relative abundance (copies/16S rRNA gene copies)

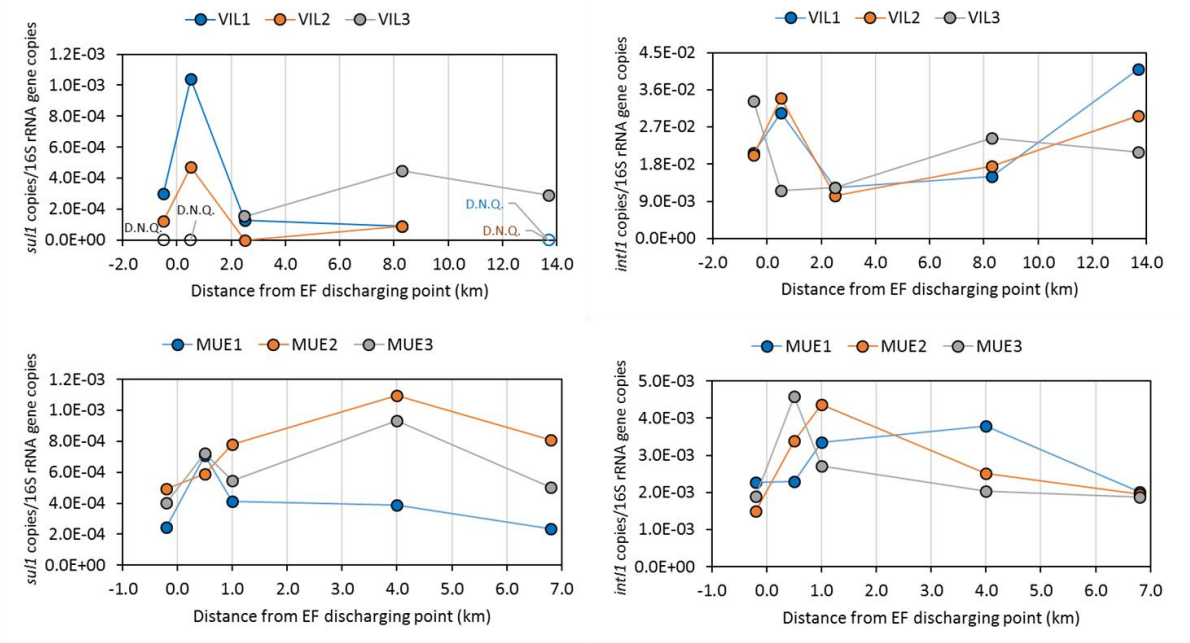


Figure S2.6. Levels (copies/g TS) – (a), and relative abundance (copies/16S rRNA gene copies) – (b) of *sul1*, and *intI1* in river sediment of the Suze near Villeret, and the Murg near Muenchwilen. The g TS indicates the mass (g) of total dried solids (TS). D.N.Q. indicates 'Detected but not quantifiable' data points.

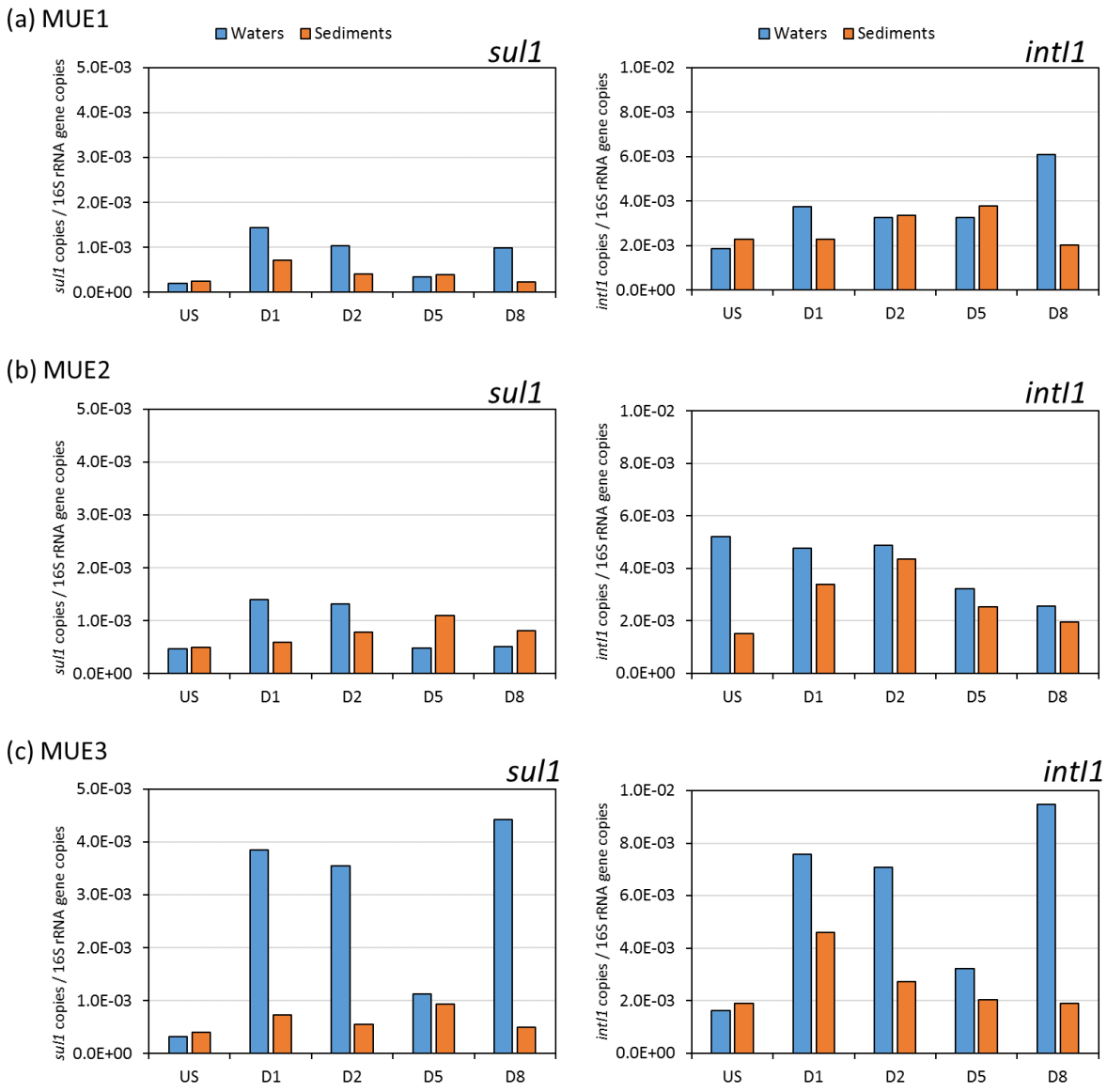


Figure S2.7. Comparison of relative abundance (gene copies / 16S rRNA gene copies) of *sul1* and *intI1* between river waters and sediments for the samplings in Münchwilen (MUE1 ~ 3).

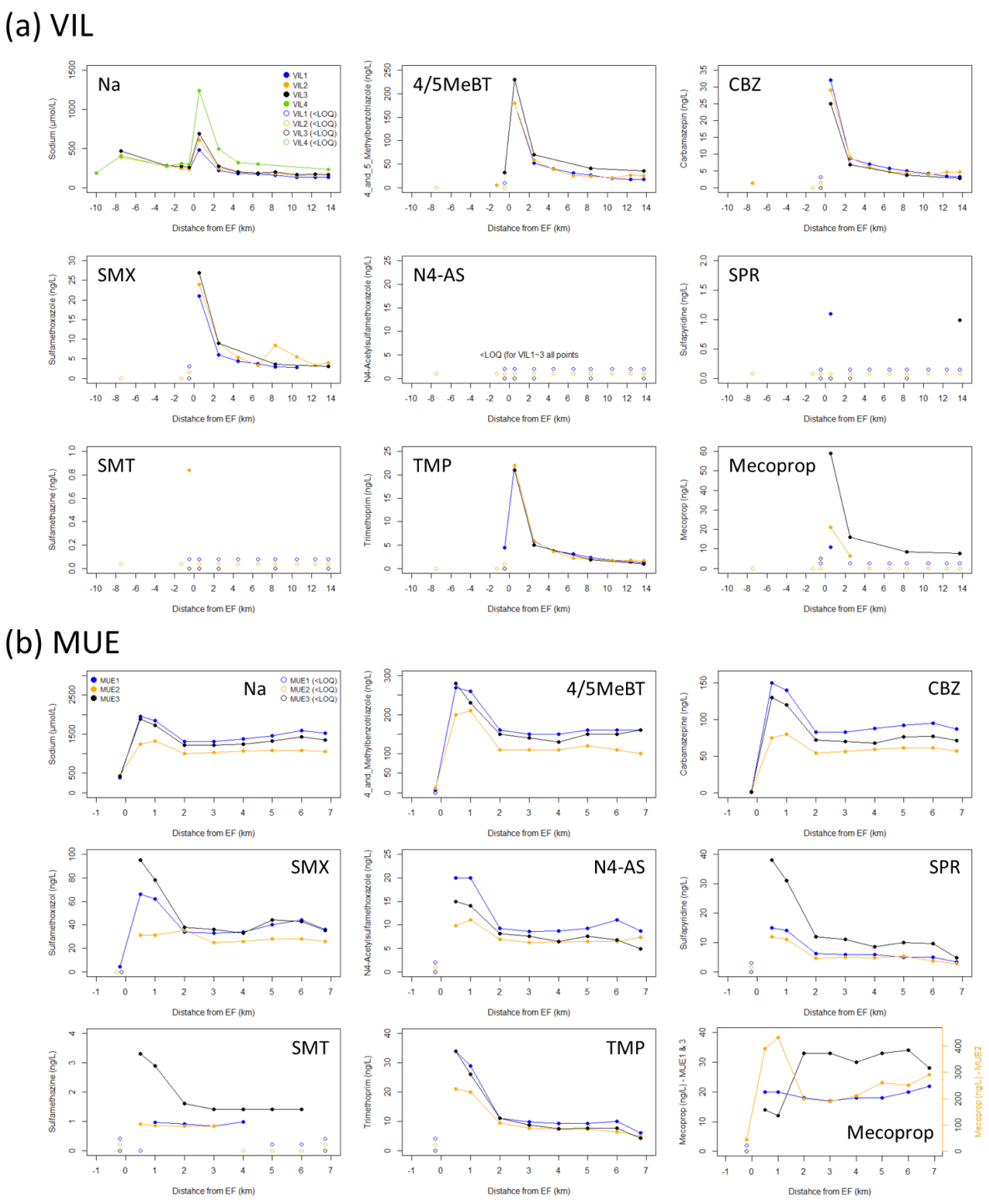


Figure S2.8. Concentrations of sodium, and selected micropollutants in the upstream, and downstream river water in Villeret (a), and Münchwilen (b). LOQ indicates the limit of quantification. The following abbreviations were used: 4/5MeBT – 4/5-methylbenzotriazole, CBZ – carbamazepine, SMX – sulfamethoxazole, N4-AS – N4-acetylsulfamethoxazole, SPR – sulfapyridine, SMT – sulfamethazine, TMP – trimethoprim.

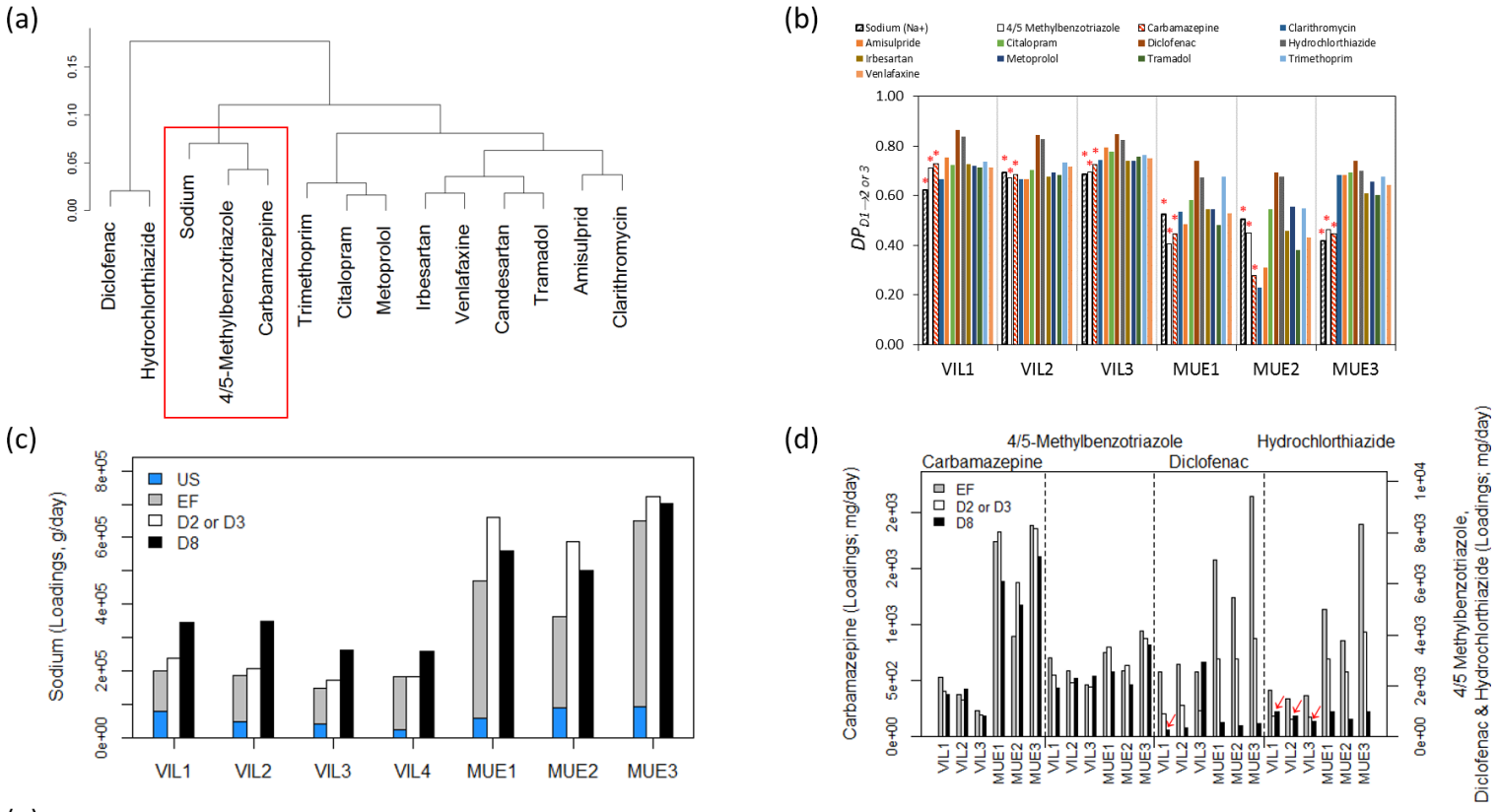


Figure S2.9. Comparison among different micropollutant indicators using $DP_{D1 \rightarrow 2 \text{ or } 3}$ in terms of cluster analysis based on Bray-Curtis distance – (a), and bar plot – (b). The short distance (D2 or 3) and long distance (D8) loadings of sodium – (c), and two other conservative tracers (i.e., carbamazepine, 4/5-methylbenzotriazole) and other micropollutants not behaving as conservative tracers (i.e., diclofenac, and hydrochlorothiazide) – (d). The measured or estimated river discharges at D8 (VIL) and D4 (MUE) were suggested in (e). The conservative tracers with the most similar $DP_{D1 \rightarrow 2 \text{ or } 3}$ were highlighted in red box (a), and asterisked in red (b). For diclofenac and hydrochlorothiazide in (d), loadings at D8 were calculated using the concentration values for the limit of quantification (highlighted with red arrows; the actual loadings could be even lower).

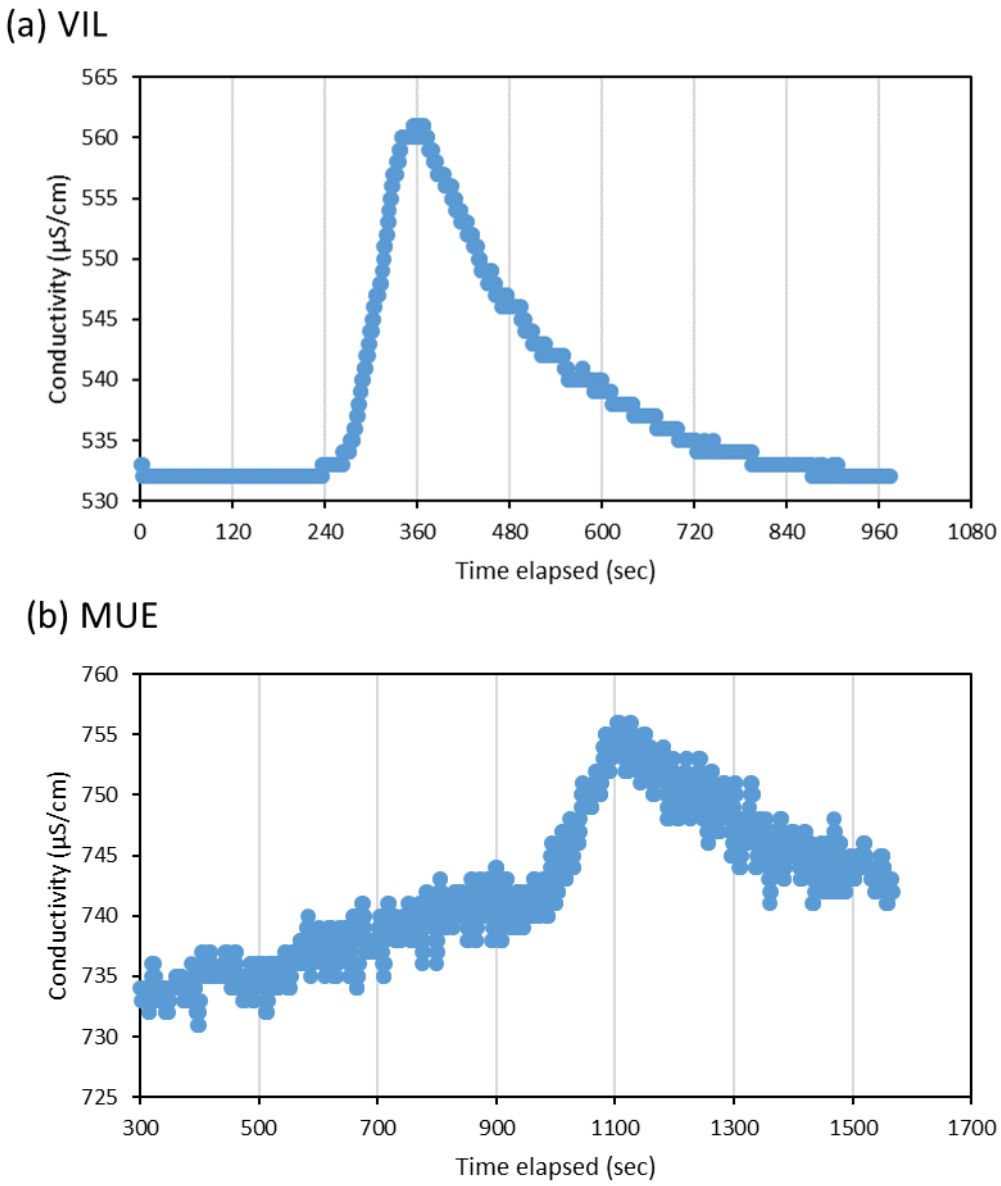


Figure S2.10. Conductivity profiles recorded at 150 m downstream distance in VIL (a), and 226 m downstream distance in MUE (b). X-axis of each diagram indicates the time elapsed after a high concentration of NaCl was spiked into the effluent discharge point. The measured-peak arrival time was 355 sec in VIL, and 1,103 sec in MUE, of which flow rates correspond to 0.42 m/s for VIL, and 0.20 m/s for MUE.

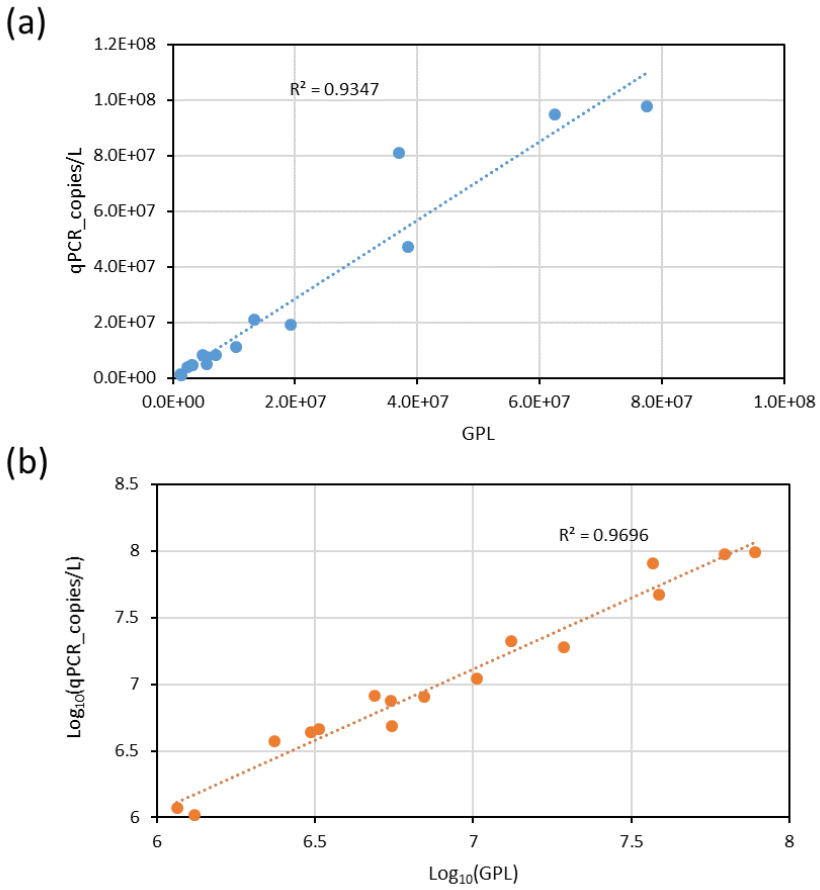


Figure S2.11. Correlation between GPL (gene per liter) estimated by metagenomics and absolute abundance (copies per liter) measaured by qPCR using *sul1*. The analysis was performed both in original unit (a), and in log_{10}-transformed unit (b). Significant correlation between two metrics was observed in both units.

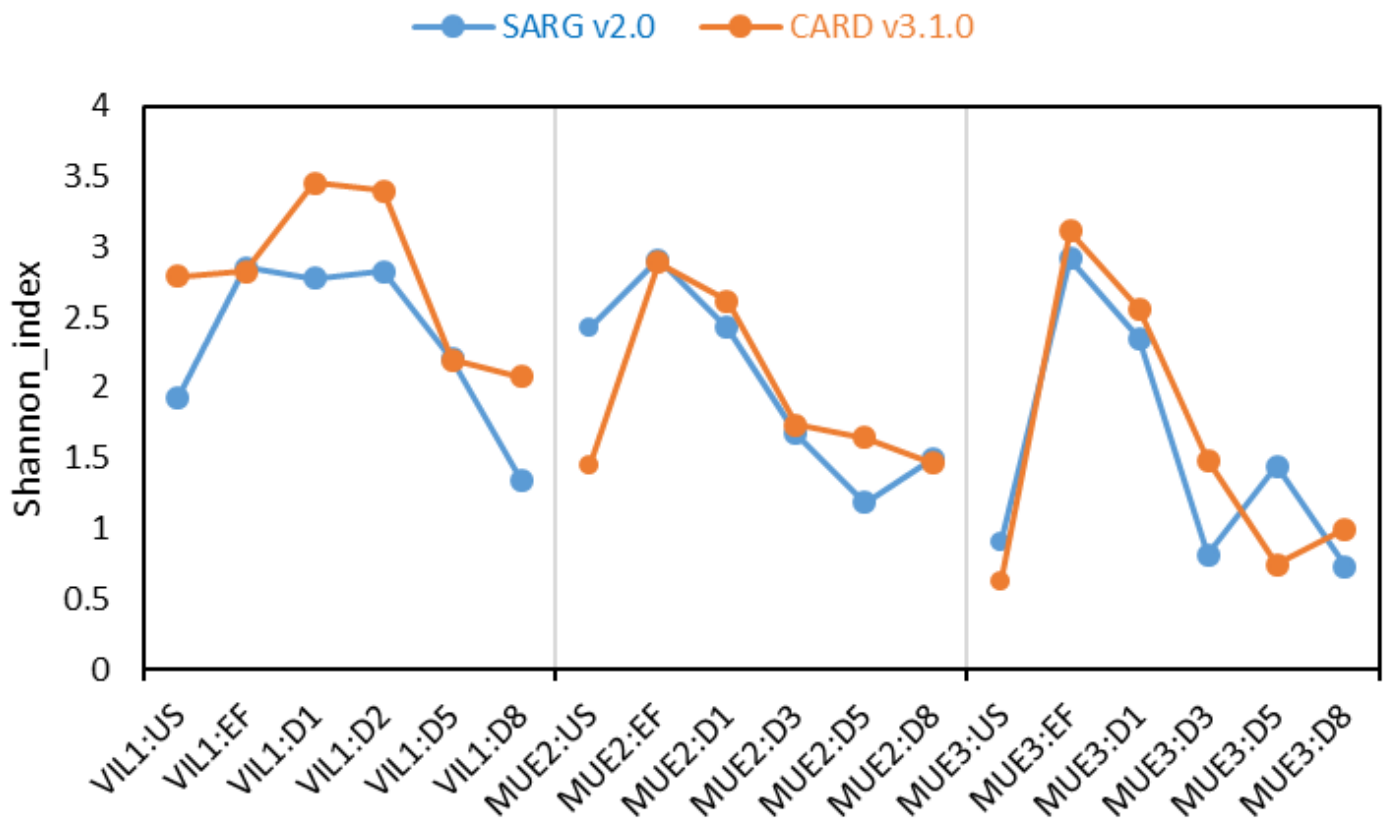
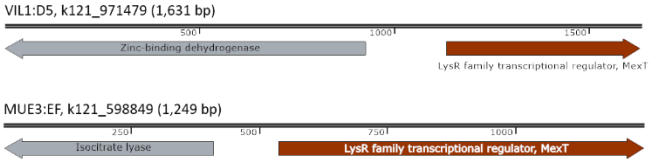


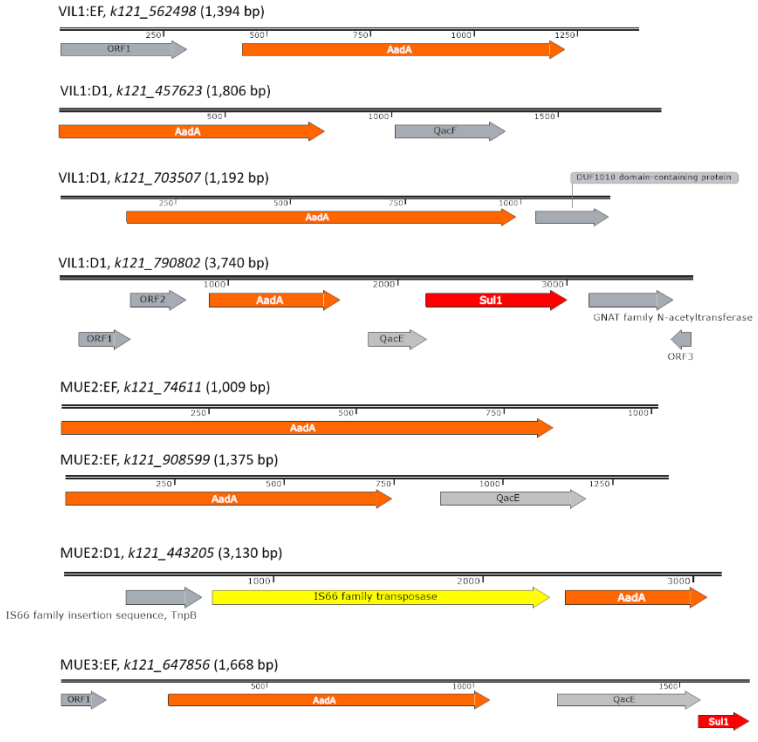
Figure S2.12. The comparison between the Structured Antibiotic Resistance Genes (SARG) v2.0 and Comprehensive Antibiotic Resistance Database (CARD) v3.1.0 in terms of the Shannon alpha diversity of identified ARG profiles. The ARG profiles were prepared in the unit of genes per million.

(a) *tetQ* containing contigs (> 1,000 bp)

(b) *mexT* containing contigs (> 1,000 bp)



(c) *aadA* containing contigs (> 1,000 bp)



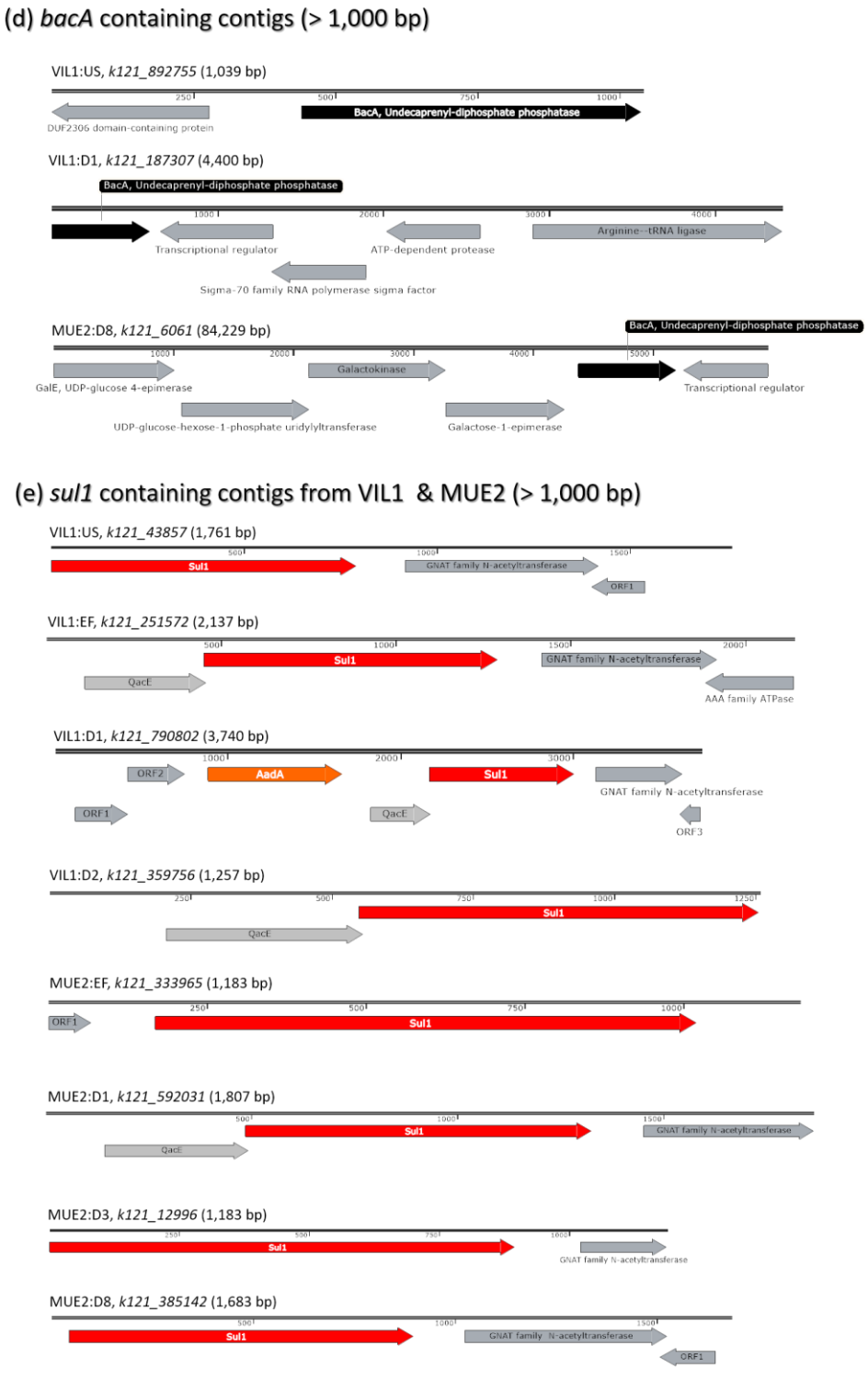


Figure S2.13. *tetQ*, *mexT*, *aadA*, *bacA* and *sul1* gene-containing contigs identified in this study (pp. 68 – 69). Contigs are from all analyzed sampling campaigns (VIL1, MUE2 & 3) unless noted otherwise. Contigs containing (a) *tetQ,* (b) *mexT,* (c) *aadA,* (d) *bacA* (Only 6 ORFs around *bacA* were shown for the contig *k121_6061*), (e) *sul1* (from VIL1 and MUE2). The proteins sequences corresponding to the genes showed > 90.0 % percent identity (P_{Ident}) to reference genes using DIAMOND protein search against NCBI nr protein database. The name of contigs was italicized (e.g. *k121_XXX*). Only contigs longer than 1,000 bp are shown.

Table S2.1. GPS coordinates (Swiss-coordinate system) for sampling locations.

Site	Location	GPS coordinates
VIL	US5	560524, 218622
VIL	US4	562557, 220280
VIL	US3	566310, 221933
VIL	US2	567657, 222317
VIL	US	568164, 223236
VIL	EF	568550, 223401
VIL	D1	569013, 223541
VIL	D2	570334, 224318
VIL	D3	572119, 225396
VIL	D4	573934, 226170
VIL	D5	575490, 226692
VIL	D6	577448, 226773
VIL	D7	579092, 227312
VIL	D8	580168, 227051
VIL	S0	568131, 223208
VIL	S1	574759, 226468
VIL	S2	566310, 221933
VIL	S3	562126, 219973
MUE	US	717378, 261105
MUE	EF	717262, 261193
MUE	D1	716747, 261384
MUE	D2	716209, 261365
MUE	D3	715236, 261440
MUE	D4	714301, 261527
MUE	D5	713664, 262274
MUE	D6	713256, 263098
MUE	D7	712681, 263693
MUE	D8	712284, 264353
MUE	S5	713398, 262795

Table S2.2. qPCR primers and probes used in this study.

Primer	Genes	Assay type	Sequences (5' to 3')	Annealing temp. (°C)*	References
ermB-F	*ermB*	SYBR Green	GATACCGTTTACGAAATTGG	63	(Chen et al., 2007)
ermB-R			GAATCGAGACTTGAGTGTGC		
bla_{CTX}-F	bla_{CTX}	SYBR Green	CTATGGCACCACCAACGATA	60	(Marti et al., 2013)
bla_{CTX}-R			ACGGCTTTCTGCCTTAGGTT		
tetW-F	*tetW*	TaqMan	CGGCAGCGCAAAGAGAAC	60	(Walsh et al., 2011)
tetW-R			CGGGTCAGTATCCGCAAGTT		
tetW-Probe			*FAM*-CTGGACGCTCTTACG-*TAMRA*		
qSUL653f	*sul1*	TaqMan	CCGTTGGCCTTCCTGTAAAG	60	(Heuer and Smalla, 2007)
qSUL719			TTGCCGATCGCGTGAAGT		
tpSUL1-Probe			*FAM*-CAGCGAGCCTTGCGGCGG-*TAMRA*		
intI1-F	*intI1*	TaqMan	GCCTTGATGTTACCCGAGAG	60	(Barraud et al., 2010)
intI1-R			GATCGGTCGAATGCGTGT		
intI1-Probe			*FAM*-ATTCCTGGCCGTGGTTCTGGGTTTT-*BHQ1*		
BAC349-F	16S rRNA	TaqMan	AGGCAGCAGTDRGGAAT	53	(Takai and Horikoshi, 2000)
BAC806-R			GGACTACYVGGGTATCTAAT		
BAC516F-Probe			*FAM*-TGCCAGCAGCCGCGGTAATACRDAG-*TAMRA*		

* These values were modified from the original references, and previously optimized in our laboratory settings.

Table S2.3. Statistics for *de-novo* assembly of metagenomes in this study.

Seq_ID	Sample	N50 (bp)	Average (bp)	Min_size (bp)	Max_size (bp)	Total_reads
M17	VIL1:US	592	593	200	466,585	75,210,034
M18	VIL1:EF	922	778	200	850,734	75,907,548
M19	VIL1:D1	779	721	200	495,260	78,803,830
M20	VIL1:D2	798	730	200	391,042	83,042,550
M23	VIL1:D5	673	654	200	350,937	87,969,544
M26	VIL1:D8	654	640	200	318,279	95,726,670
M54	MUE2:US	783	699	200	503,974	90,222,038
M55	MUE2:EF	821	746	200	326,854	84,037,346
M56	MUE2:D1	888	761	200	895,036	99,669,088
M58	MUE2:D3	854	741	200	895,047	89,883,398
M60	MUE2:D5	891	754	200	893,940	83,842,740
M63	MUE2:D8	891	748	200	894,872	78,390,602
M66	MUE3:US	788	694	200	893,351	79,266,602
M67	MUE3:EF	890	766	200	319,588	72,612,446
M68	MUE3:D1	981	802	200	903,965	76,994,406
M70	MUE3:D3	912	773	200	895,081	80,602,956
M72	MUE3:D5	905	761	200	895,035	96,339,350
M76	MUE3:D8	895	751	200	893,549	87,042,100

Table S2.4. Description of quantitative metagenomic metrics used in this study. The Python scripts used for their calculation and accompanying tutorials can be found on the main author's personal GitHub page (https://github.com/myjackson).

Metric	Definition	Calculation method	Reference
RPK	Reads per kilobase	Count the number of reads assigned to a gene (ORF), and normalize by the length of the gene.	(Katz et al., 2010)
PMSF	Per million scaling factor	Sum all the RPKs in a sample, and divide by one million.	(Li and Dewey, 2011)
GPM	Genes per million	Divide the RPK of each gene by PMSF	(Ju et al., 2019)
RPK-16S*	Reads per kilobase for 16S rRNA genes	Sum all the RPKs of 16S rRNA genes in a sample	Modified from (Ju et al., 2019)
GP16S**	Genes per 16S rRNA gene	Divide the RPK of each gene by RPK-16S	(Ju et al., 2019)
GPL	Genes per liter	Multiply the GP16S of each gene with the 16S rRNA gene copies per liter quantified by qPCR and normalized to sample volume	(Ju et al., 2019)

* The RPKs of 16S rRNA genes were calculated as follows: 1) Taxonomy was assigned to each ORF using the SILVA v138 database (Quast et al., 2013), and the ORFs to which taxonomy was successfully assigned were considered as 16S rRNA gene fragments. 2) The number of reads for each identified 16S rRNA gene fragments were counted, and it was divided by the reference length. Only the 16S rRNA gene fragments with > 85.0 % identity and > 200 bp alignment length were considered.

** The accuracy of GPL was confirmed by comparing those values with qPCR absolute abundance using *sul1* as shown in the Fig. S2.11.

Table S2.5. Information on quantitative PCR validation. Cq indicates quantification cycle, LOD denotes the limit of detection, S.D. indicates standard deviation, NTC means non-template control. TaqMan probe was termed as "Hydrolysis probe" according to Bustin et al. (2009). An extraction control was prepared by filtrating sterilized nanopure distilled water followed by DNA extraction to control potential contamination during water filtration and DNA extraction processes.

qPCR primer	Sample	Specificity verification method	Slope	qPCR-efficiency	Y-intercept	Linear dynamic range (copies - copies)	LOD-copies (Cq)	S.D. at LOD (min - max)	Cq of the NTC	Cq of the experiment control
sul1	Wat-Rep1	Hydrolysis probe	-3.42 ~ -3.46	1.944 ~ 1.962	40.1 ~ 40.7	50 - 50,000,000	50 (33.8 - 34.3)	0.37 - 0.41	N.D.	36.0
	Wat-Rep2	Hydrolysis probe	-3.28 ~ -3.42	1.962 ~ 2.018	39.2 ~ 40.1	50 - 50,000,000	50 (33.8 - 34.6)	0.37 - 0.81	N.D.	Not Analyzed
	Sediment	Hydrolysis probe	-3.38	1.976	39.52	50 - 50,000,000	50 (34.0)	0.46	N.D.	Not Analyzed
intI1	Wat-Rep1	Hydrolysis probe	-3.47 ~ -3.49	1.935 ~ 1.942	41.5 ~ 41.6	50 - 50,000,000	50 (34.9 - 35.4)	0.52 - 0.68	N.D.	39.5
	Wat-Rep2	Hydrolysis probe	-3.46 ~ -3.49	1.935 ~ 1.945	41.6 ~ 41.9	50 - 50,000,000	50 (35.4 - 36.1)	0.50 - 0.68	N.D.	Not Analyzed
	Sediment	Hydrolysis probe	-3.39	1.974	40.66	50 - 50,000,000	50 (35.0)	0.44	N.D.	Not Analyzed
tetW	Wat-Rep1	Hydrolysis probe	-3.73	1.854	42.8	50 - 50,000,000	50 (35.4)	0.34	N.D.	37.7
	Wat-Rep2	Hydrolysis probe	-3.75 ~ -3.79	1.836 ~ 1.849	42.9 - 43.3	50 - 50,000,000	50 (35.3 - 35.7)	0.45 - 0.48	40	Not Analyzed
ermB	Wat-Rep1	Melt curve analysis	-3.66 ~ -3.70	1.863 ~ 1.875	39.8 ~ 40.0	50 - 50,000,000	50 (33.4 - 34.1)	0.36 - 0.88	45	N.D.
	Wat-Rep2	Melt curve analysis	-3.47 ~ -3.66	1.877 ~ 1.942	38.2 ~ 38.6	50 - 50,000,000	50 (32.6 - 32.8)	0.17 - 0.31	N.D.	Not Analyzed
bla_{CTX}	Wat-Rep1	Melt curve analysis	-3.60	1.895	41.2	500 - 50,000,000	500 (31.5)	0.08	N.D.	Not Analyzed
	Wat-Rep2	Melt curve analysis	-3.58	1.902	40.9	50 - 50,000,000	50 (34.8)	0.22	40	Not Analyzed
16S rRNA gene	Wat-Rep1	Hydrolysis probe	-3.48	1.939	41.4	50 - 50,000,000	50 (32.3)	0.24	33.5	33.2
	Wat-Rep2	Hydrolysis probe	-3.47	1.941	41.4	50 - 50,000,000	50 (32.6)	0.27	33.8	Not Analyzed
	Sediment	Hydrolysis probe	-3.52	1.922	41.2	50 - 50,000,000	50 (32.3)	0.18	35.7	Not Analyzed

Table S2.6. MS/MS settings for measuring micropollutants. F1= Quantifier, F2= Qualifier.

Compound	Precursor mass [m/z]	Product mass F1/F2 [m/z]	Polarity	Collision Energy F1/F2 [V]	Retention time [min]	Dwell time [ms]
4/5-Methylbenzotriazole	134.1	79.1/106.1	Positive	20/18	11.6	25.7
4/5-Methylbenzotriazole-d6	140.1	85.1/112.1	Positive	22/18	11.6	25.7
Amisulpride	370.2	242.0/112.1	Positive	30/34	8.2	18.1
Amisulpride-d5	375.2	242.0/117.1	Positive	34/34	8.1	18.4
Amoxicillin C	366.1	159.9/114.0	Positive	14/54	9.7	14.5
Amoxicillin F	189.1	170.8/119.8	Positive	18/38	8.7	16.0
Azithromycin	749.5	594.1/158.1	Positive	32/38	12.2	31.9
Azithromycin-D3	752.5	594.4/158.0	Positive	30/38	12.2	31.9
Benzotriazole	120.1	65.2/92.2	Positive	20/16	9.5	13.2
Benzotriazole-d4	124.2	69.1/96.1	Positive	26/18	9.4	13.0
Candesartan	441.2	263.1/235.2	Positive	12/24	15.6	31.5
Candesartan-d5	446.2	268.1/240.1	Positive	16/24	15.6	31.5
Carbamazepine	237.1	194.0/193.1	Positive	22/38	14.3	64.8
Carbamazepine-d8	245.2	202.1/201.2	Positive	30/34	14.2	75.2
Ciprofloxacin	332.1	314.2/231.0	Positive	22/42	9.4	12.9
Ciprofloxacin-d8	340.2	322.2/235.1	Positive	22/42	9.4	12.9
Citalopram	325.2	108.8/262.0	Positive	38/18	12.7	52.0
Citalopram-d6	331.2	108.9/262.1	Positive	30/18	12.7	52.0
Clarithromycin	748.5	158.1/590.4	Positive	26/18	15.7	30.0
Clarithromycin-d3	751.5	161.3/593.4	Positive	30/14	15.7	30.0
Diclofenac	296.0	215.0/214.1	Positive	30/20	17.6	80.8
Diclofenac-d4	300.1	219.1/218.0	Positive	34/20	17.6	70.5
Hydrochlorothiazide	295.9	269.0/204.9	Negative	18/22	6.6	19.9
Hydrochloro-thiazide-13C,d2	299.0	269.8/206.0	Negative	18/26	6.6	19.9
Irbesartan	429.2	207.1/180.1	Positive	26/46	15.8	30.6
Irbesartan-d4	433.3	211.1/184.1	Positive	30/50	15.8	30.6
Mecoprop	213.0/215.0	141.0/143.1	Negative	12/8	16.6	42.4
Mecoprop-d6	219.1/221.1	147.3/149.2	Negative	12/8	16.5	34.6
Metoprolol	268.2	116.1/72.2	Positive	20/24	10.2	19.1
Metoprolol-d7	275.2	123.0/79.1	Positive	18/22	10.2	19.1
Metronidazole	172.1	128.0/82.1	Positive	10/20	6.5	22.0
Metronidazole-d4	176.1	128.0/82.1	Positive	14/30	6.5	24.3
N4-Acetylsulfamethoxazol	296.1	134.0/65.0	Positive	22/60	11.2	23.1
N4-Acetylsulfamethoxazole-d5	301.1	139.1/69.3	Positive	26/50	11.2	23.0
Norfloxacin	320.1	302.0/231.0	Positive	22/50	9.2	13.2
Norfloxacin-d5	325.2	307.0/231.0	Positive	22/50	9.2	13.2
Sulfamethazine	279.1	92.0/186.0	Positive	36/16	8.8	14.7
Sulfamethazine-d4	283.1	96.1/186.0	Positive	36/16	8.8	14.7
Sulfamethoxazole	254.1	108.0/92.0	Positive	30/20	9.7	13.9
Sulfamethoxazole-d4	258.1	112.1/96.1	Positive	26/34	9.6	14.0
Sulfapyridine	250.1	108.0/92.0	Positive	30/20	7.5	21.0
Sulfapyridine-d4	254.1	112.1/96.1	Positive	30/34	7.5	21.0
Tramadol	264.2	58.2/42.2	Positive	16/80	10.1	16.1
Tramadol-d6	270.2	64.2/45.2	Positive	24/80	10.1	16.1
Triclosan	286.9	35.2/142.1	Negative	20/32	18.8	178.6
Triclosan-d3	290.0	35.2	Negative	8	18.8	131.3
Trimethoprim	291.2	229.8/261.0	Positive	22/30	8.1	18.4
Trimethoprim-d9	300.2	234.1/264.0	Positive	26/30	8.0	19.3
Venlafaxine	278.2	260.2/115.0	Positive	4/62	12.0	29.0
Venlafaxine-d6	284.3	266.3/121.0	Positive	10/30	12.0	29.0

CHAPTER 3.

Wastewater bypass is a major temporary point-source of antibiotic resistance genes and multi-resistance risk factors in a Swiss river

Jangwoo Lee[1,2], Karin Beck[1], and Helmut Bürgmann[1,*]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Department of Surface Waters – Research and Management, Kastanienbaum, Switzerland

[2]Department of Environmental Systems Science, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

* Corresponding Author: Helmut Bürgmann (Helmut.Buergmann@eawag.ch).

This chapter was submitted to a journal as a research article, and is now under review:

Lee et al., Wastewater bypass is a major temporary point-source of antibiotic resistance genes and multi-resistance risk factors in a Swiss river. Under review at Water Research.

Abstract

Untreated combined sewage (i.e., bypass) is often discharged by wastewater treatment plants to receiving rivers during stormwater events, where it may contribute to increased levels of antibiotic resistance genes (ARGs) and multi-resistance risk factors (i.e., multi-resistant bacteria and multi-resistance genomic determinants (MGDs)) in the receiving water. Other contamination sources, such as soil runoff and resuspended river sediment could also play a role during stormwater events. Here we report on stormwater event-based sampling campaigns to determine temporal dynamics of ARGs and multi-resistance risk factors in bypass, treated effluent, and the receiving river, as well as complimentary data on catchment soils and surface sediments. Both indicator ARGs (qPCR) and resistome (ARG profiles revealed by metagenomics) indicated bypass as the main contributor to the increased levels of ARGs in the river during stormwater events. Furthermore, we showed for the first time that the risk of exposure to bypass-borne multi-resistance risk factors increase under stormwater events and that many of these MGDs were plasmid associated and thus potentially mobile. In addition, elevated resistance risk factors persisted for some time (up to 22 hours) in the receiving water after stormwater events, likely due to inputs from distributed overflows in the catchment. This indicates temporal dynamics should be considered when interpreting the risks of exposure to resistance from event-based contamination. We propose that reducing bypass from wastewater treatment plants may be an important intervention option for reducing dissemination of antibiotic resistance.

Keywords

Antimicrobial Resistance; Stormwater Events; Wastewater Bypass; Metagenomics; River

3.1. Introduction

Antibiotic resistance has been considered one of the biggest challenges to public health, and its global increase has been recognized as an impending public health crisis by intergovernmental entities. Antibiotic resistance is transmitted not only via direct human-to-human interaction, but also via multi-sectoral routes, such as interdependent routes among humans, food animals, and environments (McEwen and Collignon, 2018).

Wastewater treatment plants (WWTPs) are a known route through which sewage-borne resistance genes are discharged into the environment (Bürgmann et al., 2018;Rizzo et al., 2013). Even though many studies showed that the level (genes per volume) of antibiotic resistance genes (ARGs) decreases during wastewater treatment processes (Ju et al., 2019;Marano et al., 2020;Rodriguez-Mozaz et al., 2015), profound levels of ARGs remain in treated effluents, which leads to profound impacts of WWTPs on receiving rivers (Lee et al., 2021;Rodriguez-Mozaz et al., 2015).

The abovementioned studies have generally been conducted under conditions of normal WWTP operation. However, many sewer systems or WWTPs will regularly experience high flow events caused e.g. by high intensity precipitation. If such flows exceed the capacity of the WWTP or sewer system, it may lead to wastewater bypass – i.e. the release of untreated combined sewage into receiving waters (Weyrauch et al., 2010). During stormwater events, the level of ARGs in wastewater-receiving waters can increase due to the input of wastewater bypass, but high volume flows from precipitation may reduce or compensate. Resistance levels may also increase during stormwater events due to inputs from other non-point sources. Soil runoff from river catchment, and sediment resuspension are considered as potential non-point sources under stormwater events (Tsihrintzis and Hamid, 1997). Storm drains and urban runoff (Baral et al., 2018) may represent temporary point sources at various points in a catchment. However, bypass is expected to contain the highest abundance of resistance determinants (i.e., ARGs and resistant bacteria) and multi-resistance risk factors (i.e., multi-resistant bacteria and multi-resistance genomic determinants (MGDs)) because untreated sewage contains resistance determinants (Ju et al., 2019;Rodriguez-Mozaz et al., 2015), and multi-resistant bacteria at considerably higher levels than treated effluent (Czekalski et al., 2012). The potential combined effects of different contributors on receiving waters during stormwater events were investigated in the Stroubles Creek, Virginia, USA by monitoring surface water qualities (Garner et al., 2017). Two other studies monitored combined-sewage overflows and the receiving surface waters noted elevated levels of ARGs in the receiving waters during stormwater events in the Hudson River, Raritan Bay, and Passaic River, USA (Eramo et al., 2017), and North Shore Channel, USA (Chaudhary et al., 2018). A study performed in the Antelope Creek in Lincoln,

Nebraska, USA (Baral et al., 2018) that does not receive wastewater, the authors found that inflows from storm drains had the strongest impact on the riverine resistome (i.e., total ARGs identified using shot-gun sequencing) compared to impacts from sewage leakage, street sweepings, soils, and sediments. While wastewater-receiving river waters have been studied quite intensively so far, there have thus been comparatively few studies systemically examining the combined effects among potential contamination sources. For instance, the contribution of 'bypass' has not been systematically compared with other potential sources (i.e., treated effluent, surface soil runoff, and sediment resuspension), thus whether those other sources also contribute significantly to the increase of riverine ARGs or not remains unanswered. Furthermore, the potential existence of bypass-borne 'multi-resistance' risk factors in the river has not been known.

In this study we therefore investigated the effect of stormwater events leading to bypass discharge on a highly wastewater-impacted river, the River Murg near the WWTP Münchwilen, Switzerland. The main goals of this study were 1) to assess the impact of stormwater-related disturbance and resilience of the resistome of the Murg river by monitoring temporal dynamics of the resistome during stormwater events that lead to combined-sewage bypass, 2) to identify the key source(s) which contribute the most to the increase of riverine resistance levels, and 3) to assess the impact of stormwater events in terms of multi-resistance risk factors. We used various molecular biological, microbiological, and ecological parameters to identify the key factor among different pollution sources. Those parameters include absolute and relative abundance of well-known anthropogenic antibiotic resistance markers (i.e., *sul1*, and *intI1*) by quantitative PCR (qPCR) and alpha- and beta-diversity analysis of the resistome based on environmental shotgun metagenomics. This approach was motivated by the hypothesis that different sources would be distinguishable by the different (relative) abundance and/or the composition and diversity of resistance indicators. To analyze multi-resistance risk factors, we used metagenome-assembled contigs and heterotrophic cultivation – presumptive multi-resistant bacteria and MGDs, which are defined as contigs where ≥ 2 ARGs conferring resistance to different antibiotic classes are co-located in this study. The underlying hypothesis was that untreated combined-sewage (i.e., bypass) is more likely to contain high levels of multi-resistance risk factors, and thus their abundance might increase in receiving waters during stormwater events.

3.2. Materials and Methods

3.2.1. Site description and samplings

Samples were obtained from the river Murg near Münchwilen, Thurgau, in Switzerland which is one of the most wastewater-impacted rivers in Switzerland,. This site was well studied under dry-

flow condition in 2018 (Lee et al., 2021), and has the following characteristics: 1) High proportion of effluent discharge to receiving water discharge (33.0 – 38.0 % under base-flow condition), 2) No known point-source inputs exist other than one WWTP, 3) The WWTP receives combined sewage (community sewage + storm drain) during stormwater events, and the bypass (i.e., untreated combined sewage) is discharged after primary sedimentation to the receiving river when the treatment and rainwater storage capacity of the WWTP is exceeded during heavy rainfall, 4) Other than one discharge point for bypass near the WWTP (Fig. 3.1 – left side), there are additional upstream points where combined-sewage may be discharged: The nearest one in 800 m, and the farthest one in 11.0 km upstream of the WWTP (H. Zbinden, personal communication to authors, June 29, 2021), 5) the river catchment is widely utilized as agricultural area (i.e., pasture and meadow for livestock farms) (BAFU, 2013). Coordinates and key information on the sampling sites as well as information on the WWTP Münchwilen are summarized in Dataset S1.

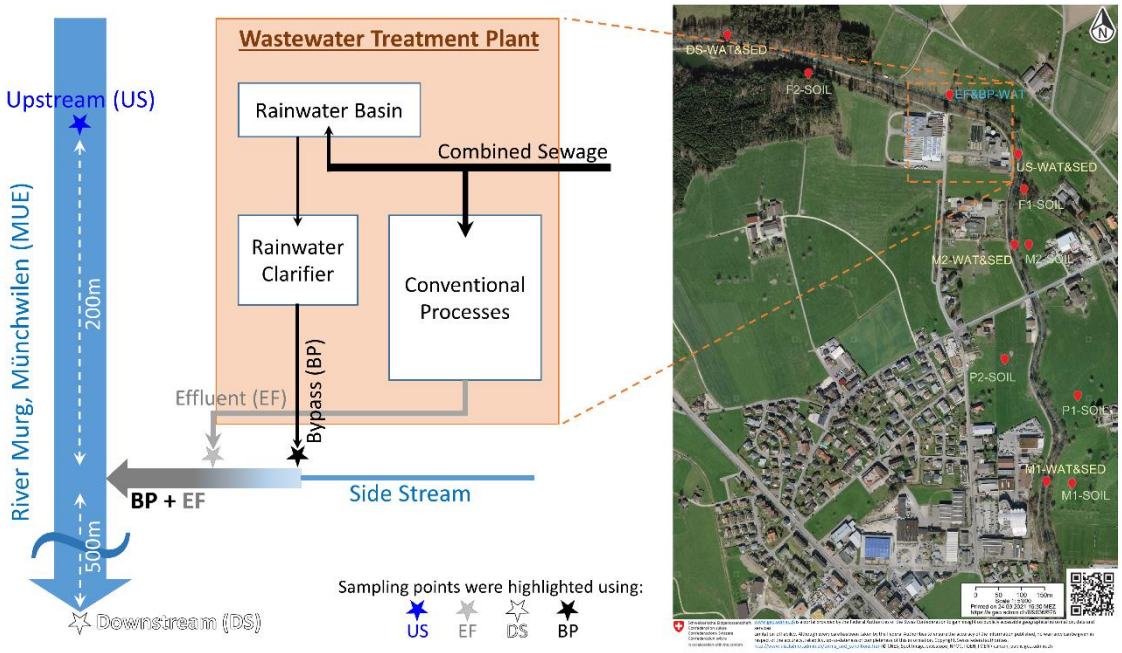


Figure 3.1. Left side: Schematic diagram of wastewater and bypass flow and treatment at the wastewater treatment plant of Münchwilen (CH) and discharge locations in the receiving waters. Right side: satellite image map showing location of the WWTP and sampling locations in the River Murg, and for adjacent soils near the city of Münchwilen. WAT: water sampling in the river upstream (US, 200 m) and downstream (DS 500 m) of the wastewater discharge point and of effluent (EF) and bypass (BP). SED: sites for sediment sampling. SOIL: sites for soil sampling.

On June 25 (DRY1) and August 27 (DRY2) in 2019, water samples were obtained under base-flow condition (no precipitation in the previous 24 hours) over 24 hours in hourly 1 L batches using autosamplers (ISCO, USA) installed at three sampling points: river Murg 200 m upstream (US) and 500 m downstream (DS) of the wastewater discharge point, and from effluent (EF) (prior to discharge). Water samples were cooled with ice and cooling packs inside the autosamplers. Surface

soils and sediments (< 5 cm) from the river catchment were obtained in a separate campaign under base-flow condition on May 28, 2019 (DRY0) as shown in Fig. 3.1 – right side: Two soil samples each were obtained from forests (F1, 2), meadows (M1, 2), and pastures (P1, 2). River sediments were sampled once at the DS and US locations, and at two sites (M1, 2) further upstream (0.45 and 1.0 km). Both surface soil and sediment samples were obtained by subsampling and pooling 5 subsamples from each location into sterile containers on site.

Event-based sampling campaigns were carried out over the summer of 2019. Weather forecasts (provided by MeteoSwiss) for the Münchwilen region were monitored from July to September 2020. When heavy rainfall was predicted for the catchment area of the Murg, autosamplers were deployed in the same way as described for base-flow sampling for US, DS, and EF sites. An additional autosampler was installed at the wastewater bypass line (BP) of the WWTP prior to discharge (Fig. 3.1 – left side). Two event-based sampling campaigns were eventually performed on August 12-13 (13:00 – 12:00) and September 23 (01:00 – 24:00), 2019 when there was a total of 10.5 mm and 20.0 mm of precipitation during 48 hours (24 hours on the sampling date plus the previous 24 hours, refer to Fig. S3.1) (MeteoSwiss, 2019, December 20). During these events, BP could only be sampled when there was sufficient bypass flow while the autosampler was operating.

All samples were cooled at 4 °C in the dark while being transported to the laboratory within 32 hours (from the starting time of auto-samplings).

3.2.2. Sample pooling strategy

For DRY1 and DRY2 sampling campaigns, upon arrival at the laboratory, hourly water samples were pooled over 6 consecutive hours to produce 4 samples for further analysis (i.e., morning (MOR, 06:00-11:00), afternoon (AFT, 12:00-17:00), evening (EVE, 18:00-23:00), and night (NIG, 24:00-05:00)) for every sample type (US, DS, EF). Exceptions were made for MOR and NIG samples in DRY1-EF where we could not obtain samples during 01:00 – 07:00 due to too low water levels in the outflow (Table S3.1).

For RAIN1 and RAIN2 sampling campaigns, equal volumes of hourly water samples (i.e., each 1.0 L) were pooled from time intervals before, during, between, and/or after stormwater events for 2 – 7 consecutive hours as shown in Table S3.1. Those pooled samples were stored at 4 °C in the dark, and further analysis (i.e., cultivation and biomass filtration) was performed the next day (after < 16 hours).

The information on BP flow condition and sample pooling was encoded into the sample labeling scheme, which denotes samples as BP or no BP (nBP) depending on BP flow at the time of sampling. The sampling times of pooled samples is also given. For example, “BP-2(+8-9h)” denotes the second

sample taken during active bypass flow, and is a composite sample taken from 8 to 9 am. All sample designations are listed in Table S3.1.

3.2.2. Heterotrophic plate counts

Colony counts of presumptive clarithromycin and tetracycline resistant bacteria (CLR/TET) was shown to be an useful indicator for anthropogenic resistance inputs in our previous studies (Czekalski et al., 2012;Lee et al., 2021). The CLR/TET colonies were cultivated for water samples using biomass concentrated on 0.2 µm pore size cellulose nitrate filters followed by incubation on R2A agar plates in the presence of clarithromycin (4.0 mg/L) and tetracycline (16.0 mg/L) as outlined in our previous publications (Czekalski et al., 2012;Lee et al., 2021). Samples were diluted before filtration using 0.85% NaCl according to the previously optimized ranges (i.e., 10 mL loading volumes of 10^{-3} ~ 10^{-1} diluted samples) (Lee et al., 2021). Contamination controls using a blank solution (i.e., sterile 0.85% NaCl) were performed for each sampling campaign, and no growth of colonies was confirmed. Technical triplicates were incubated for each sample, and standard errors of triplicates are shown as error bars.

3.2.3. Biomass filtration, DNA extraction, and quantitative PCR

To obtain concentrated suspended biomass for DNA extraction, water samples were filtered through a sterilized 0.2 µm pore size cellulose-nitrate filter (Sartorius, Germany) using autoclaved Nalgene™ filter units (Thermo Fisher Scientific, USA). The maximum filtration volume was up to 1.0 L for river waters, 0.5 L for EF, and 0.1 L for bypass, but the exact volume varied by sample because filters were clogged at different volumes (Dataset S2). After biomass filtration, the filters were stored at -20 °C until processing for DNA extraction. Soil and sediment samples were frozen immediately after arrival at the laboratory and stored at -20 °C until processed further for DNA extraction.

DNA extraction was performed using DNeasy PowerWater Kit (Qiagen, Germany) for water samples, and DNeasy PowerMax Soil Kit (Qiagen, Germany) for soils and sediments according to the manufacturer's instruction. After extraction, DNA quality indicators (i.e., 260/280 and 260/230 absorbance ratios) were checked using NanoDrop One spectrophotometer (Thermo Fisher Scientific, USA), and concentrations were analyzed using both NanoDrop and Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific, USA) (Dataset S2). The extracted DNA samples were stored at -20 °C until analyzed.

qPCR targeting two resistance indicators (i.e., *sul1* and *intI1*) and the 16S rRNA gene were performed as outlined in our previous studies (Czekalski et al., 2014;Ju et al., 2019;Lee et al., 2021). The two resistance indicator genes (i.e., *sul1* and *intI1*) used in our study are two of the most widely used genetic markers for tracking anthropogenic sources of resistance (Berendonk et al.,

2015;Gillings et al., 2015). Primers are given in Table S3.2, and key information for qPCR validation is given in Table S3.3. Standards were run in quintuplicates, samples in triplicate. Measurements with Cp values above negative controls but below the limit of detection (LOD; average Cp of lowest valid standard), or which had a standard deviation of Cp values of triplicates > 0.5 were not quantified, and labeled as 'Detected but not quantifiable (D.N.Q.)'. Samples without valid Cp or with average Cp values ≥ smallest Cp value from the non-template controls were indicated as 'Not-detected (N.D.)'. Standard errors of triplicate measurements are displayed as error bars in the figures.

To calculate absolute abundances from qPCR results, we calculated the total number of gene copies per biomass filter and normalized by filtration volumes for water samples (i.e., copies per volume). For soils and sediments, the total copies measured in wet mass were normalized to dry mass (i.e., copies per g dried soils or sediments). The dry mass of soils and sediments was determined according to Standard Methods (APHA-AWWA-WPCF., 1981).

3.2.4. Shotgun metagenomic sequencing and downstream data analysis

The following DNA samples were selected for shotgun metagenomics sequencing: 1) 6 samples from two DRY samplings (i.e., the same DNA mass (ng) of different time points (MOR-AFT-EVE-NIG) were pooled together for each location; 3 locations (US, EF, DS) from each campaign are included) and 2) 4 time points (BP-1(+1h), nBP(+2h), nBP(+6-11h), and nBP(+24h)) for US and DS, 2 time points (BP-1(+1h), nBP(+24h)) for EF, and 1 time point for bypass (BP-1(+1h)) from RAIN1, 3) 4 time points (nBP(+1-3h), BP-3(+10-11h), nBP(+12-14h), and nBP(+18-24h)) for US, DS, EF, and 3 time points (BP-1(+4h), BP-2(+8-9h), BP-3(+10-11h)) from RAIN2. Information on the selected samples and general information on the obtained metagenomics libraries is given in Dataset S3. In addition to our own samples, raw metagenome reads of soils from two other studies (i.e., grazed and ungrazed grassland soils from Sinkiang/Inner Mongolia in China; compost amended, and non-amended greenhouse soils from microcosm experiments in Virginia, USA) were downloaded from the Sequence Read Archive (SRA) (Chen et al., 2019;Zheng et al., 2021), and subjected to bioinformatics analysis (Dataset S3). By comparing with this data, we expected to draw more general conclusion, e.g., by ruling out whether or not the chance that the soil resistome from other regions with similar or different land-usages could be similar to the resistome of stormwater-disturbed waters.

The selected DNA samples from our study were sequenced using the Illumina Novaseq6000 with a paired-end (2 × 150) strategy by Novogene Europe (Cambridge, UK). The reads containing adapters and low quality reads (N > 10 % and quality score ≤ 5) were removed by Novogene, and the read qualities were double checked by the authors using FastQC v0.11.4 (Andrews, 2010).

For resistome analysis, we used a read-based annotation approach because assembly efficiencies (i.e., percentage aligned sequences calculated using eq.1) were relatively low for soil and sediment samples. As a result, resistance gene profiles could not be properly represented using the *de-novo* assembly based annotation approach used in our previous publication (Lee et al., 2021) (Fig. S3.2).

$$\text{Assembly Efficiency } (\%) = \frac{\sum_{i=1}^{n}(Contig\ Lengths_i \times Average\ Coverage_i)}{Total\ metagenome\ read\ bases\ in\ a\ sample} \times 100 \quad (\text{eq. } 1)$$

Where, n denotes the total number of assembled contigs in a sample and average coverage indicates the average sequencing depths per contig calculated according to Albertsen et al. (2013).

Read-based ARG annotation was performed using DeepARG short-reads pipeline v1.0.2 with default parameters. This pipeline was developed to process short-reads and to find and annotate resistance genes using a deep learning algorithm after incorporating several ARG databases publicly available (Arango-Argoty et al., 2018). In short, this pipeline finds and quantifies ARG-like and 16S rRNA gene-like reads, and normalizes the reads assigned to ARGs to 16S rRNA gene-like reads (Arango-Argoty et al., 2018). Both our samples and downloaded metagenomes (from SRA) were analyzed using this approach.

In water samples only (where assembly efficiencies were relatively high, > 51.0 %), a *de-novo* assembly based approach was used for the limited purpose of analyzing MGDs. We followed the work-flow described in our previous publications (Ju et al., 2019;Lee et al., 2021). A short summary of the bioinformatics work-flow is also suggested in the supporting information (SI). Identification of ARGs was based on two published databases - CARD v3.1.0 protein homolog model (Alcock et al., 2020) for ARGs, and INTEGRALL v1.2 (Moura et al., 2009) for *intI1*.

3.2.5. Analysis of multi-resistance genomic determinants

In order to list-up MGDs, all contigs containing ≥ 2 ARGs that confer resistance to different classes of antibiotics were sub-selected (Dataset S6). For defining resistance classes, we strictly followed the classification of antibiotics shown in CARD v3.1.0. The occurrence of MGDs was quantitatively assessed in terms of the relative abundance defined as follows in this study (eq.2):

$$\text{Relative Abundance of total MGDs} = \sum_{i=1}^{m} Cov_{MGD-i} \Big/ \sum_{j=1}^{n} Cov_{Contig-j} \quad (\text{eq. } 2)$$

Where, Cov_{MGD} indicates the average coverage of a MGD; Cov_{Contig} denotes the average coverage of a contig; m, n indicate the total number of MGDs and contigs, respectively, in a sample.

The ARGs physically co-located within MGDs were further visualized in a directed network using R (igraph). The network analysis was performed using an adjacency matrix produced from two vertices

(two co-located ARGs) and an edge (co-occurrence frequency). The edge information was produced according to the following equation (eq.3):

A = [The MGDs containing ARG-A in all samples], B = [The MGDs containing ARG-B in all samples]

$$E_{ARG-A \rightarrow ARG-B} = n(A \cap B) / n(A) \quad (\text{eq. } 3)$$

A detailed example of the derivation process for edges is given in the SI.

We tried to assign taxonomy to MGDs using Kraken2 (Wood et al., 2019), Kaiju v1.7.2 (Menzel et al., 2016) and BLAST against NCBI-nt database (Sayers et al., 2019) as outlined in our previous publication (Lee et al., 2021). In short, we considered the assigned taxonomy as a valid information only if all three approaches yielded a consensus classification at genus level.

3.2.6. Statistical analysis and visualization

All statistics and graphs were produced using R. The Shannon-index was used as a parameter for alpha-diversity, and calculated using Vegan (Oksanen et al., 2013). The dissimilarity of resistome among metagenome samples (i.e., beta-diversity analysis) were analyzed using non-metric multi-dimensional scaling (NMDS) in Vegan. Pairwise t-test (p-adjustment method: Benjamini-Hochberg) was performed to test for significant differences among potential contamination sources (i.e., wastewaters, surface sediments, soils) in terms of alpha-diversity index.

3.3. Results

3.3.1. Absolute abundance of resistance indicators (*sul1*, *intI1*, and CLR/TET) in waters

Two sampling campaigns performed under dry weather (baseflow) conditions were used to establish reference values for resistance indicator levels in US, DS, and EF under normal operation, as shown by horizontal lines in Fig. 3.2. The results from these campaigns also showed the expected elevated absolute and relative abundance of resistance indicators in EF and the expected general increase of the abundance in DS compared to US as a result of EF discharge. We observed little intra-day variability (Fig. S3.3).

Analysis of the resistance indicators in water samples from stormwater event-based sampling revealed that bypass (BP) contained high levels of resistance. For instance, the levels of *sul1*, *intI1*, and CLR/TET in BP were higher than in effluent (EF) by up to 2.1 order of magnitude (Fig. 3.2). Accordingly, the levels of those indicators in up- and downstream water of the receiving river (US and DS) also increased when bypass flow was active. It is important to note that we learned from the operators after the campaign that the US sampling point also receives combined-sewage overflows from discharge points further upstream. The levels of *sul1*, *intI1*, and CLR/TET in US and DS were

highest during bypass flow events (i.e., BP-1 in RAIN1, and BP-3 in RAIN2). The peak levels in bypass-affected DS samples exceeded the DS levels at dry conditions by up to 2.4 orders of magnitude, and even exceeded the levels in treated effluent (EF). This indicates that aquatic riverine resistance at receiving points was profoundly influenced by bypass from WWTP and upstream points of discharge (at 0.8 and 2.6 km upstream locations) during bypass events. In contrast, levels in EF were not affected or increased only slightly during or after the bypass event (Fig. 3.2). While the levels of CLR/TET in EF at RAIN2 was higher than for baseflow conditions (DRY) by 0.4 – 0.9 order of magnitude, this was also the case for the pre-event EF (nBP(+1-3h)) and DS levels from this campaign, thus probably reflecting a change in the effluent not related to the bypass event (Fig. 3.2).

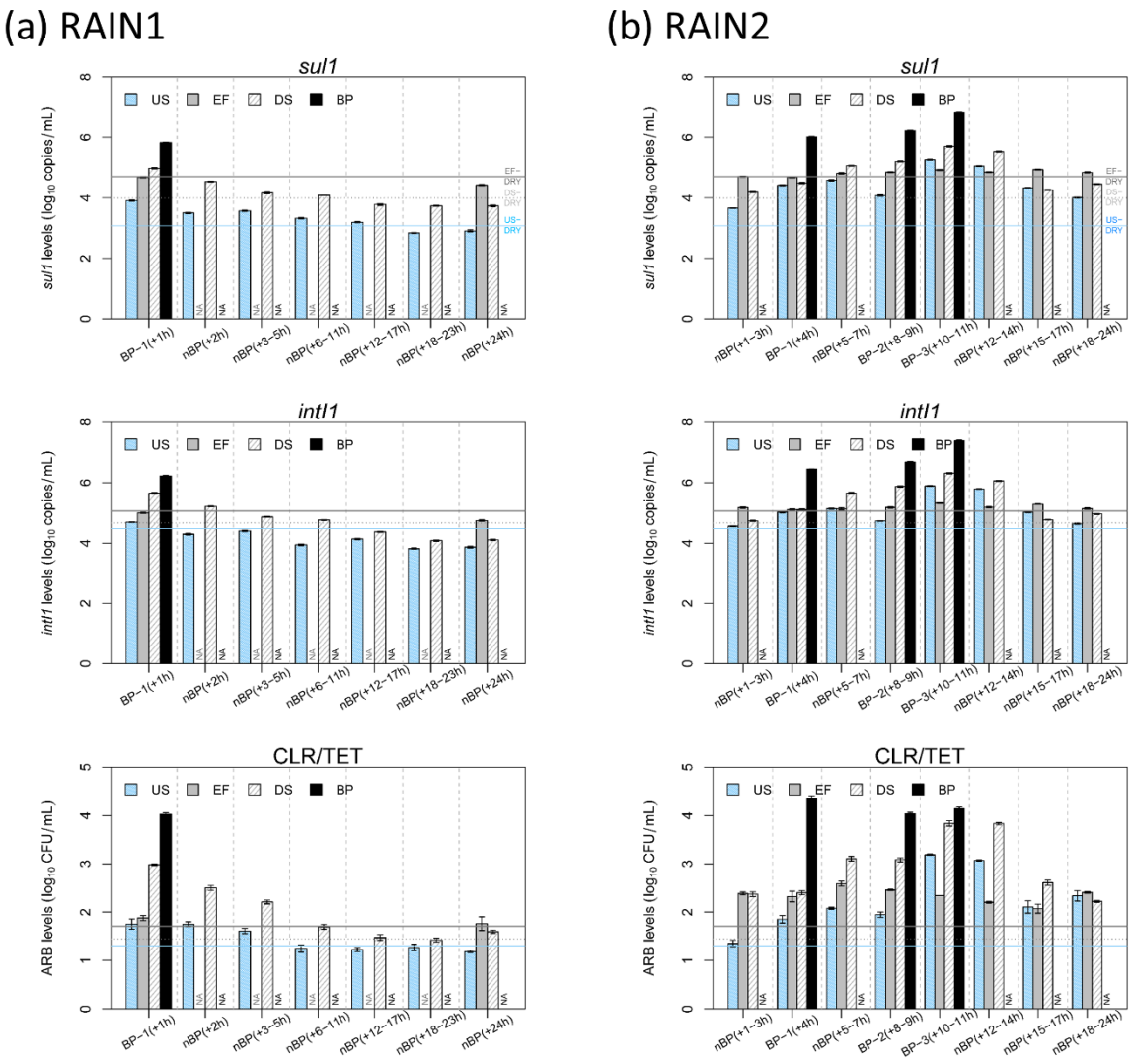


Figure 3.2. Absolute abundance (gene copies or colony forming unit per volume) of resistance indicators in water samples during bypass events: resistance genes (*sul1*, *intI1*), and multi-resistant bacterial counts (CLR/TET, presumptive clarithromycin and tetracycline resistant bacteria). (a) Results from the RAIN1 sampling campaign, and (b) RAIN2 sampling campaign. The averaged abundance during the DRY sampling campaigns (DRY1 + 2) for EF (effluents), DS (downstream), and US (upstream) samples are indicated by colored lines (dark grey for EF; light grey for DS; blue for US). BP indicates bypass samples, which are only obtained while bypass flow is active during or after precipitation events. NA indicates not available for BP and not analyzed for EF.

We observed that elevated levels of resistance indicators at US and DS temporarily persisted for some time after BP flow had stopped (Fig. 3.2). The levels returned to the pre-disturbance status (the levels at DRY) only after 17 – 22 hours in RAIN1. The temporal persistence of resistance was even more profound in the campaign RAIN2 when there were heavier rains. The levels of *sul1*, *intI1*, and CLR/TET at US and DS were still higher by up to1.4 order of magnitude compared to the pre-disturbance status (the levels at DRY) after 13 hours from the last bypass event (BP-3(+10-11h)).

For most samples, the DS levels of resistance indicators were above the US level (except for *intI*1 and *sul*1 levels during RAIN2, nBP(+15-17h) sample) showing a persistent effect of the WWTP (EF and BP). Notably however, the US levels increased as well especially during RAIN2. While the existence of further combined-sewage discharge points upstream may explain this observation, we could not a priori rule out the possibility of other contributions (i.e., sediment resuspension and surface soil runoffs). In order to quantitatively assess the potential contributions of different sources, we analyzed relative abundances of resistance indicator genes in the water samples and potential sources (bypass, effluent, sediment and soil) by normalizing to 16S rRNA gene levels.

3.3.2. Relative abundance of *sul1* and *intI1* in waters, soils, and sediments

Relative abundance of *sul1* and *intI1* at US and DS increased dramatically during bypass events (BP-1 in RAIN1, and BP-1 to -3 in RAIN2) (Fig. 3.3). Similar to the pattern that we observed in absolute abundance analysis, the increased relative abundance also temporarily persisted in receiving waters, and gradually decreased over time (Fig. 3.3). The abundance of *sul1* and *intI1* was below the LOD (i.e., D.N.Q. or N.D.) for most soil samples. Where it could be determined, the relative abundance of these indicators in soil was far below the values for water samples (Fig. 3.3). The sediment samples showed higher values compared to soils, especially for the sediment obtained from DS. However, all relative abundances determined for sediment were lower than in river water with exception of *sul1* in US in some pre- and post-bypass samples (i.e., nBP(+3-5h) to nBP(+24h) in RAIN1, and nBP(+1-3h) in RAIN2) (Fig. 3.3) (M01-43 in Dataset S3). To broaden the dataset for soil, 19 published soil metagenomes were downloaded from NCBI-SRA. We analyzed the ARG content using a read-based approach (see Dataset S3 for key information on raw reads). Alpha- and beta-diversity analysis of resistomes was performed using relative abundance data of ARG identified at resistance subtype-level (Dataset S4).

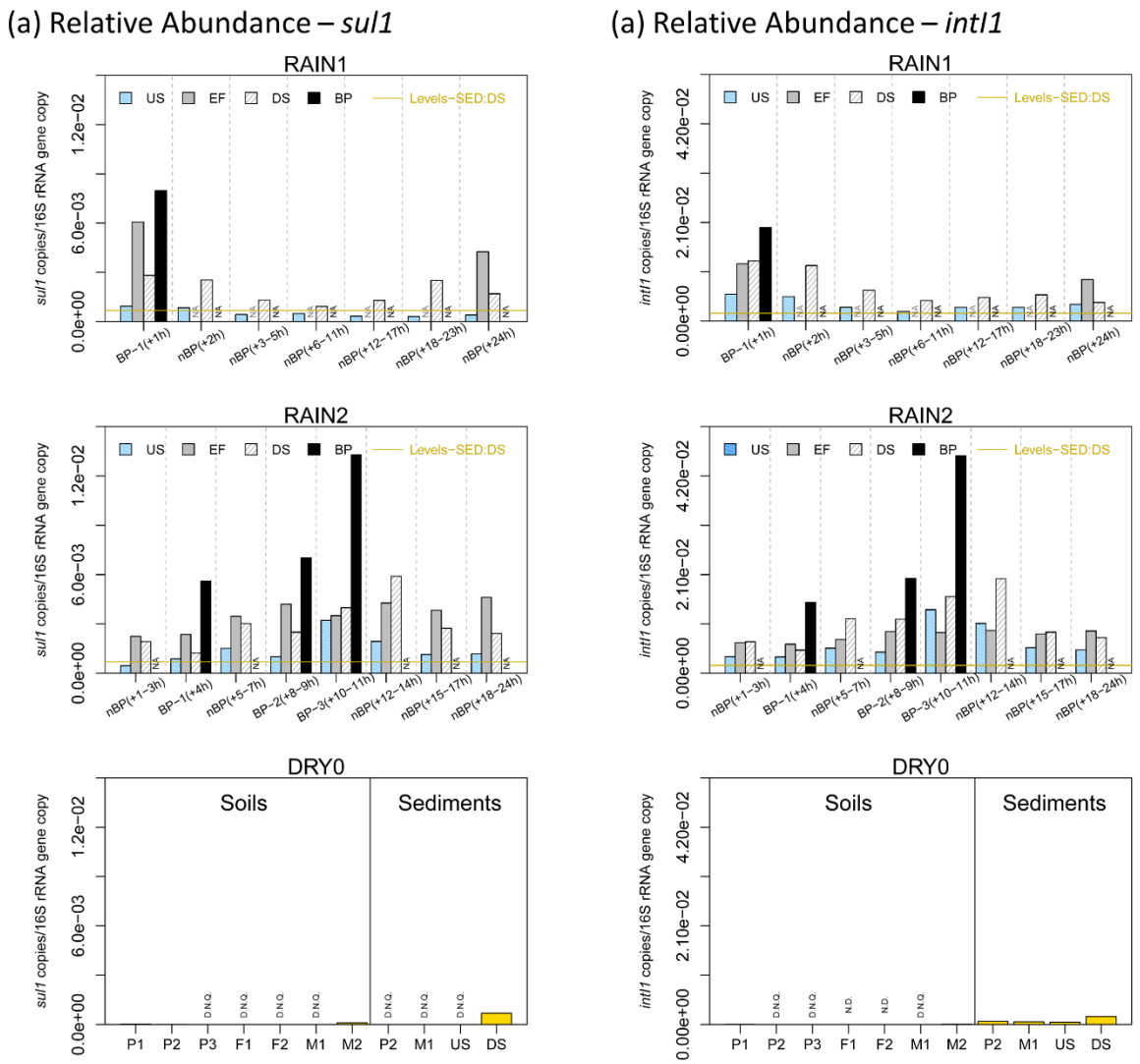


Figure 3.3. Relative abundance (gene copies per 16S rRNA gene copy) of resistance indicators (*sul1*, *intI1*) in water samples during stormwater events (RAIN1 and 2), and in sediments and soils obtained under dry weather condition (DRY0). The relative abundance of indicators for downstream sediments (Levels-SED:DS) was displayed in a gold line in water samples (RAIN1 and 2). D.N.Q. indicated 'Detected but not quantifiable', and N.D. denotes 'Not detected'. Soils were from pasture (P), forests (F) or meadow (M). NA indicates not available for BP and not analyzed for EF.

ANOVA and the post-hoc pairwise t-test of Shannon diversity index values of the retrieved resistomes from the three potential contamination sources in the River Murg catchment (i.e., wastewater, sediments, and soils-CH) revealed that the Shannon-index of wastewater resistomes was significantly higher than of soil or sediment (Fig. 3.4a). The Shannon-index values of soil resistomes from other studies (i.e., soils-CN and soils-USA) were similar to or lower than for soils-CH. The two highest Shannon index values in US and DS were observed during and right after bypass discharges (i.e., BP-3 and nBP(+12-14h) in RAIN2 for US; BP-3 in RAIN2 and BP-1 in RAIN1 for DS), which were higher than sediments and soils.

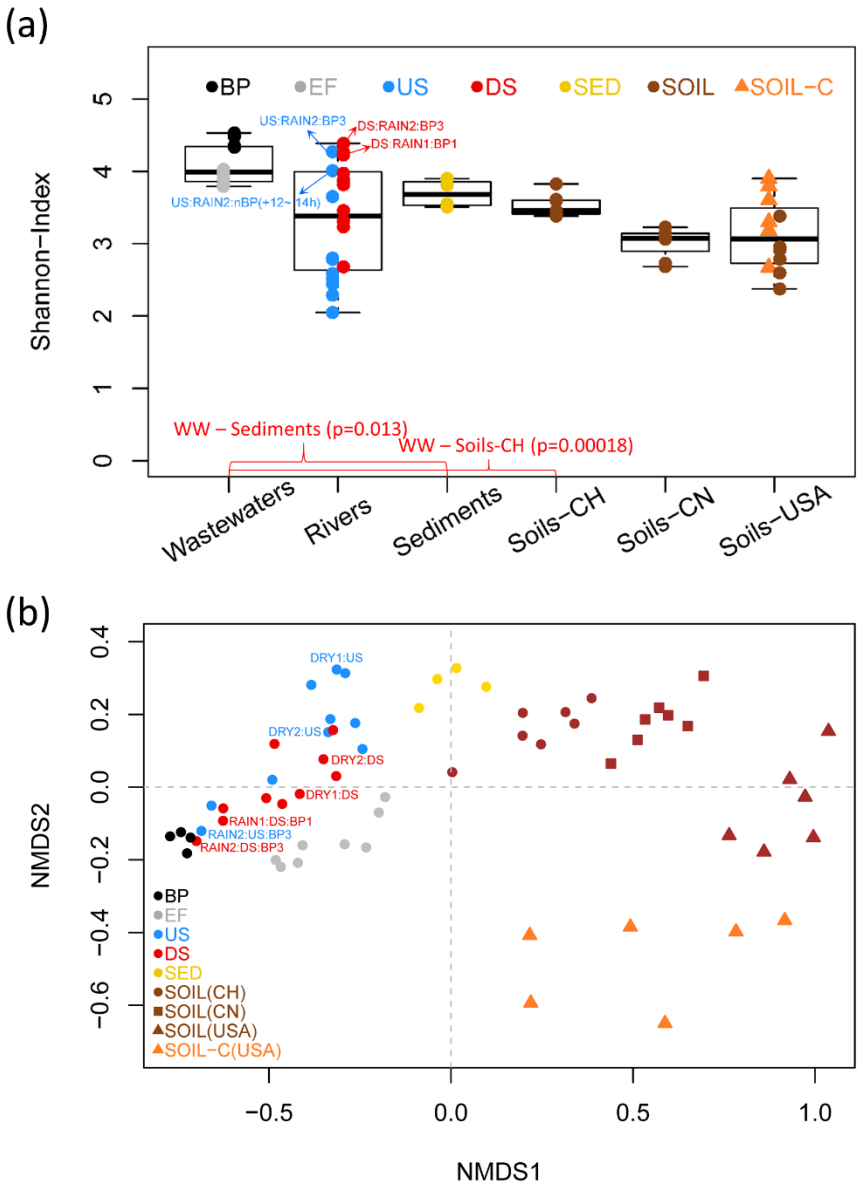


Figure 3.4. Alpha- and beta-diversity analysis of sample resistomes. (a) Boxplot of Shannon-index values as an indicator of resistome alpha-diversity for each compartment. Boxplots show the median (central line), 25th and 75th percentile (box) and minimum and maximum value (whiskers), actual values are overlayed (markers, see legend in (b)). Statistically significant differences (t-test) among potential contamination sources (i.e., wastewaters, sediments, and surface soils) are indicated by red brackets. (b) Non-metric multi-dimensional scaling (NMDS) analysis of resistome structure (resistance gene relative abundance data) in different compartments. BP, EF, US, DS indicates bypass, effluents, upstream, and downstream waters, respectively. Soils-CH are from this study. Soils-CN (Sinkiang/Inner Mongolia, China), Soils-USA (compost unamended; Virginia, USA), and Soils-C (compost amended; Virginia, USA) are from other studies (Zheng et al., 2021; Chen et al., 2019).

Beta-diversity analysis of ARG subtypes using NMDS based on Bray-Curtis distance (stress = 7.49) showed that dissimilarities between resistomes of water and other compartments were profound (Fig. 3.4b). Water samples clustered together mostly in the 2nd and 3rd quadrants on the ordination plot, and soils were mostly located in the 1st and 4th quadrants of the plot. The dissimilarities between sediments and water resistomes (especially for US samples) were low compared to those

between the soils and water, but still clearly distinct from highly stormwater-influenced river water samples (e.g., BP-3 in RAIN2 for US; BP-1 in RAIN1 and BP-3 in RAIN2 for DS). The resistomes of high stormwater-influenced water samples were however most similar to the bypass samples.

3.3.3. Riverine resistome during stormwater events

Resistance classes with high explanatory power for the resistome beta-diversity (and thus differing between resistomes in different habitats or with different influence of bypass) were selected as follows: 1) resistome profiles were aggregated by class of antibiotic resistance in Dataset S5 and analyzed using NMDS (stress = 5.67) (Fig. S3.4), 2) the correlation between resistance classes and the ordination, and 3) the resistance classes showing high statistical significance (p-value < 0.0001) of the correlation were selected. A total of 16 resistance types were selected with these criteria, conferring resistance to macrolide-lincosamide-streptogramin (MLS), aminoglycoside, bacitracin, beta-lactam, diaminopyrimidine, fluoroquinolone, fosmidomycin, glycopeptide, multidrug, nitroimidazole, peptide, pleuromutilin, rifamycin, sulfonamide, tetracenomycin-C, and tetracycline.

The relative abundances of those 16 resistance classes of resistance in water, soil, and sediment samples are shown in Fig. 3.5. Bypass and effluent samples contained in high abundance MLS, aminoglycoside, beta-lactam, diaminopyrimidine, fluoroquinolone, sulfonamide, and tetracycline resistance genes. Those 7 resistance classes showed high relative abundance also in high stormwater-influenced US and DS samples (i.e., BP-1 in RAIN1, and BP-1 to -3 in RAIN2). In soils and sediments, the following 7 types of ARGs were prevalent: fosmidomycin, glycopeptide, multidrug, nitroimidazole, pleuromutilin, rifamycin, and tetracenomycin-C resistance genes. In spite of being selected as resistance classes with high explanatory value, bacitracin resistance genes were highly prevalent in waters in general but also occurred in soil and sediment, and the “multidrug” resistance class was abundant in all sample types but especially abundant in the soil and sediment. Both are therefore shown separately in Fig. 3.5 and were not considered further.

The relative abundance of ARGs in the classes not selected by our criteria are shown in Fig. S3.5.

3.3.4. Analysis of multi-resistance genomic determinants

The results shown above provide strong evidence that resistance levels increased profoundly in receiving river waters during stormwater events, and that the majority of the increase in resistance determinants in river waters originates from bypass according to the results presented in the sections 3.3.1 ~ 3.3.3.

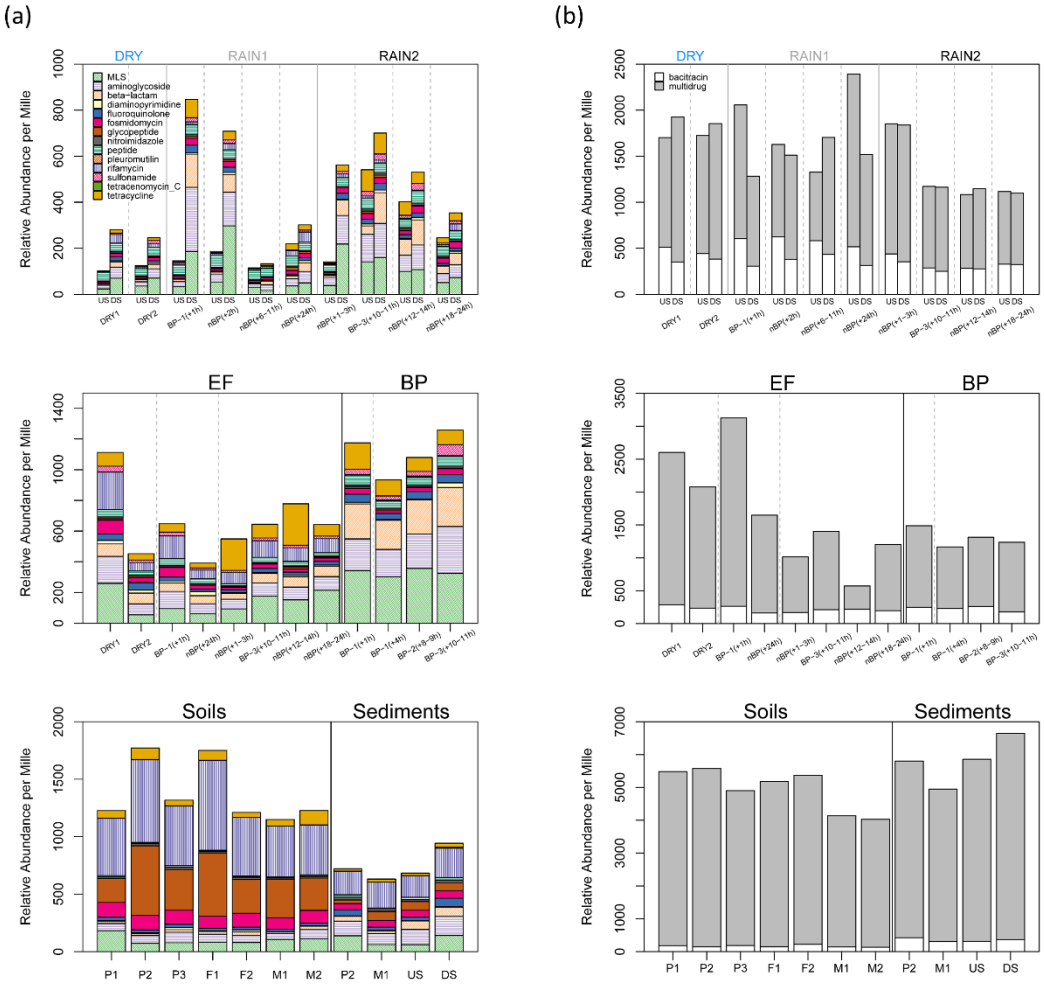


Figure 3.5. Resistome profiles by class of antibiotic resistance in river waters (upstream and downstream waters), wastewaters (bypass and effluents), soils and sediments. Total 16 types of resistance classes were selected using biplot analysis – the 14 classes of antibiotic resistance which significantly correlated with the ordination (Fig. S3.4) ($p<0.0001$) are shown in (a); the others (i.e., bacitracin and multidrug resistance genes) were displayed separately in (b) because those had much higher values of relative abundance than the previous 14 classes. The order of stacked bars is same as the order of resistance classes in the legend. Unit: Relative abundance per Mille (i.e., ARG reads per 1000 16S rRNA gene reads).

Based on these findings, we hypothesized that the risk of exposure to various 'multi-resistance' risk factors would also increase. While this was already partly supported by our CLR/TET analysis in Fig. 3.2, we aimed to obtain a more comprehensive picture by screening *de-novo* assembled contigs for evidence of multi-resistance. Contigs containing ≥ 2 ARGs were classified as MGDs, and their relative abundances in each sample were calculated according to eq.2. The temporal dynamics of MGDs in terms of relative abundance (Fig. 3.6) followed a similar pattern to that of other resistance indicators (Fig. 3.2 & 3.3). Highest abundances were found in all BP samples. High abundances of MGDs were also observed in the river during BP1 (RAIN1) and BP-1 ~ 3 (RAIN2). Elevated relative abundance of MGDs persisted temporarily, but decreased over time (Fig. 3.6). Notably the relative abundance of MGDs in EF was low compared to BP and to BP-affected river water.

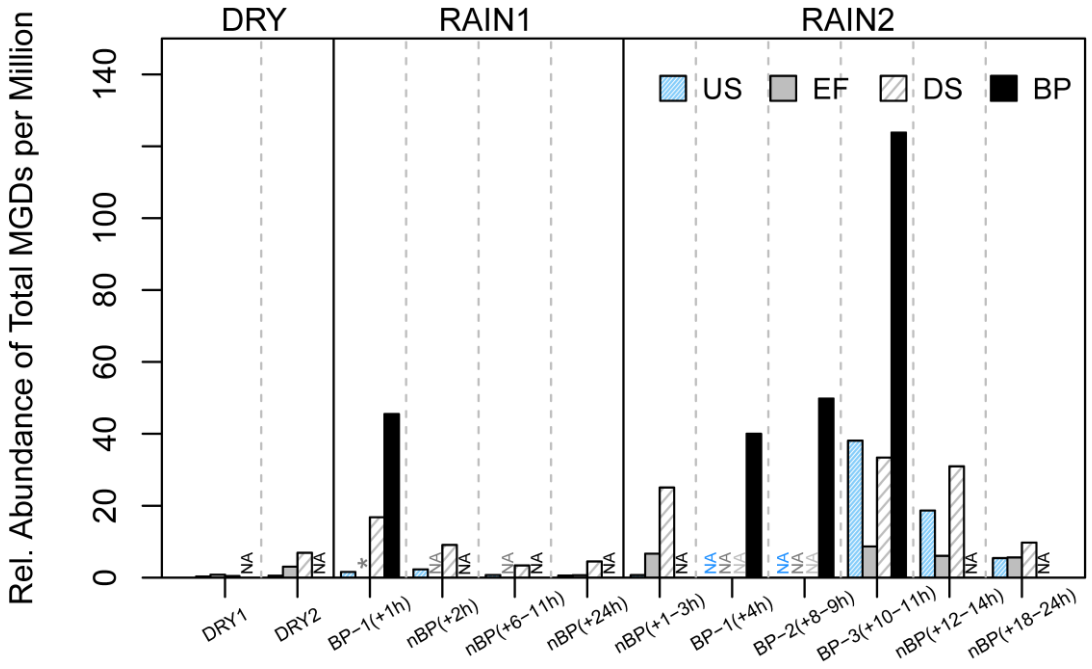


Figure 3.6. Relative abundance of total multi-resistance genomic determinants (calculated according to eq.2) in water samples. US, EF, DS, and BP indicate upstream, effluent, downstream, and bypass water samples. NA indicates not available for BP and not analyzed for EF, and the asterisk (*) denotes analyzed but not-detected.

To provide a better understanding of the nature of the resistances contained in the MGDs, we visualized the co-localization of ARGs on MGDs with a network analysis. The directed network shows the frequency with which each gene co-occurs on MGDs with each other gene (Fig. 3.7a). According to references, many of those genes are expected to be associated with plasmids. For instance, various *aac(6')-Ib*, *aadA*, *catB*, *cmlA*, *dfrA*, *sul1*, and *intI1* homologues were found to be located in bacterial plasmids isolated from wild animals (Dolejska and Papagiannitsis, 2018), activated sludges and treated wastewaters (Tennstedt et al., 2003), sediments, and various water environments (i.e., river waters, drinking and wastewaters) (Ma et al., 2017). Among genes associated with animal-, human- and environmental-origin plasmids, the following subtypes also occurred in MGDs of our study: *aac(6')-Ib*, *aadA*, *aadA5*, *catB2*, *catB3*, *cmlA5*, *dfrA14*, OXA-2, OXA-10, OXA-129, *sul1*, and *intI1*. Assuming that the ARGs that are directly linked to the subtypes listed above are also associated with plasmids, a large portion of the ARGs (35 out of 72 subtypes; Fig. 3.7b) identified on MGDs are likely plasmid-located. Those potentially plasmid-associated ARGs confer resistance to aminoglycoside, beta-lactam, diaminopyrimidine, fluoroquinolone, phenicol, and sulfonamide antibiotic classes. The ARGs that are previously reported to be associated with gene cassettes of *intI1* (Tennstedt et al., 2003) are indicated by larger nodes in Fig. 3.7b.

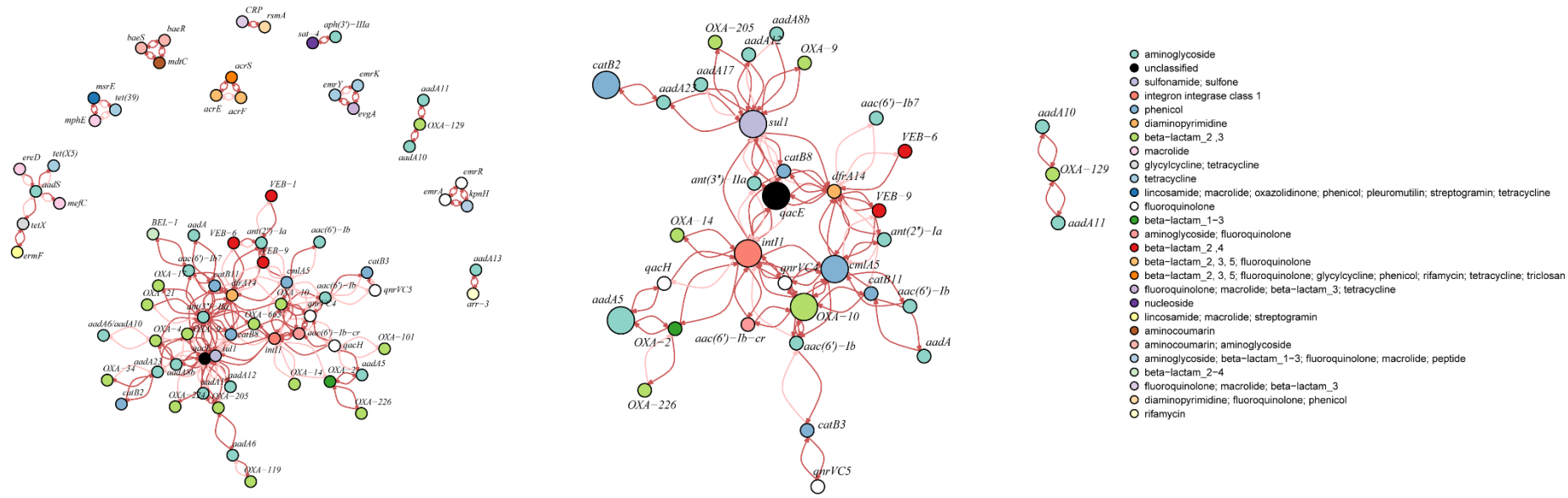


Figure 3.7. Directed networks visualizing physical association among resistance genes located in multi-resistant genomic determinants (MGDs) from water samples. (a) The entire networks visualizing total ARGs associated with MGDs, (b) Sub-networks visualizing the ARGs potentially associated with plasmids. The classification of antibiotic resistance is according to CARD v3.1.0. The edges (co-occurrence frequencies calculated according to eq.3) with > 0.5 (50 %) were colored in a thick red, and those with ≤ 0.5 were colored in a light red. The direction of the arrow indicates co-occurrence of the gene of origin with the target gene. The beta-lactam_1 ~ 6 indicate the classes conferring resistance to carbapenem, cephalosporin, penam, monobactam, cephamycin, and penem antibiotics, respectively. The large nodes in (b) are potentially associated with gene cassettes of *intI1*.

Table 3.1. The multi-resistance genomic determinants (MGDs) with successfully assigned genus level taxonomies. Only MGDs with consensus assignment from all three assignment methods used (Kaiju, Kraken2, BLASTn) were shown. Pident indicates the percentage of identification (%), and Q_{hsp} denotes the query coverage for high-scoring segment pair (%).

Contig Information					**Assigned Taxonomy**		**BLASTn Results (P_{ident} > 85.0 & Q_{hsp} > 60.0)**				**Containing-ARGs**
Contig-ID	**Sample**	**Average Coverage**	**Relative Abundance (to total contigs)**	**Length (bases)**	**Kaiju**	**Kraken2**	**Matched Fragments**	**Sequence ID**	**P_{ident}**	**Q_{hsp}**	
k121_300880	M03	5.2	4.4E-05	1,864	*Pseudomonas aeruginosa*	*Pseudomonas aeruginosa*	*Pseudomonas aeruginosa* strain FDAARGOS_571 chromosome	gb|CP033833.1|	90.3	93.0	*aadA6* OXA-119
k121_1708054	M13	8.9	6.5E-05	2,246	*Escherichia coli*	*Escherichia coli*	*Escherichia coli* strain SCU-107 chromosome	gb|CP053384.1|	99.6	99.0	*evgA* *emrK* *emrY*
k121_852659	M13	7.4	5.3E-05	10,689	*Citrobacter* spp.	*Citrobacter* sp. CFNIH10	*Citrobacter* sp. Y3 chromosome	gb|CP050009.1|	99.5	100.0	*kpnH* *emrA* *emrR*
k121_1425685	M14	4.5	3.8E-05	1,067	*Escherichia coli*	*Escherichia coli*	*Escherichia coli* F070 DNA	dbj|AP023237.1|	99.9	100.0	*acrS* *acrE*
k121_588822	M19	17.5	1.6E-04	52,811	*Tolumonas auensis* DSM 9187	*Tolumonas auensis* DSM 9187	*Tolumonas auensis* DSM 9187	gb|CP001616.1|	98.8	84.0	CRP *rsmA*

The MGDs retrieved from water samples were then subjected to contig-based taxonomy assignment in order to infer potential bacterial hosts (Dataset S7). After annotating taxonomy using two read-based taxonomy assignment tools (i.e., Kaiju and Kraken2), a total of 12 out of 126 MGDs showed a consensus at genus level. Those 12 MGDs were additionally subjected to BLAST analysis against NCBI-nt DB. Finally, a total of 5 MGDs showed a consensus among all three approaches (Table 3.1), and we considered the assigned taxonomy as potential hosts to those MGDs. Four of the five assigned hosts were potential pathogens, including *Pseudomonas aeruginosa*, *Escherichia coli*, and *Citrobacter* sp. On the other hand, *Tolumonas auensis* was described as an isolate from freshwater sediment capable of toluene production (Fischer-Romero et al., 1996).

3.4. Discussion

3.4.1. Wastewater bypass as a major contamination source for resistance during stormwater events

Various lines of evidence (i.e., relative abundance of resistance indicators and alpha- and beta-diversity analysis of various compartments of river systems) revealed that wastewater bypass was a major source of resistance during stormwater events in river Murg. This indicates WWTPs and sewer systems are important intervention points for tackling dissemination of ARGs in the aquatic environment during stormwater events. Without proper interventions, the risk of public exposure to those genes will remain high.

The risk of exposure to MLS, aminoglycoside, beta-lactam, diaminopyrimidine, fluoroquinolone, sulfonamide, and tetracycline resistance genes through contact with river water is considerably elevated during such events (Fig. 3.5). Furthermore, bypass events occur frequently in the Murg catchment near Münchwilen. For instance, they occurred on a total of 118 calendar dates in 2019, the year of our study. The bypass events occurred especially frequently during August ~ September, occurring on 54 out of 61 days. The elevated exposure risk also persists for several hours after the stormwater event (discussed in detail below). Considering that beta-lactam antibiotics are among the most commonly prescribed antibiotics in Switzerland for both in- and out-patients (FOPH and FSVO, 2020), the discharge and potential for public exposure to bacteria with those resistance groups could be potentially problematic. While the actual risks of exposure and likelihood of spreading resistance determinants to the population through this route remain unknown, our findings may provide some justification to recommend caution against exposure to river water during and after strong precipitation events in rivers receiving high levels of bypass. Obtaining a quantitative overview of the frequency and magnitude of bypass discharge would be an important asset in this sense.

Considering that both relative and absolute abundances of resistance indicators were relatively stable over time for EF and considerably lower than those for bypass (i.e. incoming combined

sewage) reducing combined sewer overflows or bypass by increasing the treatment or retention capacities of WWTP and retention basins could be a way to reduce the amount of resistance factors released to receiving waters. Scaling-up existing treatment facilities could be an option, or at least, this aspect could be considered in the early stage of WWTP installation or renovation, i.e., making capacities of new WWTPs be high enough to handle large quantities of incoming combined sewage.

3.4.2. A temporal persistence of bypass-born resistance at receiving waters after stormwater events

Our observation that bypass-borne resistance at US and DS sites persisted for an unexpectedly long time (i.e., for 22 hours until it fell to pre-disturbance levels in RAIN1) indicates that the temporal dynamics of resistance deserves further investigation and should be considered when interpreting the fate of event-based resistance inputs in the river and associated risks. The downstream transport of event-based pollutant inputs has been widely studied (Jamieson et al., 2005;Nevers and Boehm, 2010;Parsaie and Haghiabi, 2017), and hydraulic processes (i.e., advection, and dispersion) are regarded as main drivers. Our previous study showed that EF is fully mixed cross-sectionally after 500 m downstream distance at the Münchwilen study site (Lee et al., 2021). Thus, the prolonged increase of bypass-borne resistance genes at both DS and US sampling sites appears to be likely due to advective transport and longitudinal dispersion of upstream inputs from combined sewer overflows. Several such potential discharge points exist in the catchment, however data on their contribution during the observed events is not available.

Given our current data and study design, it is not possible to quantitatively explain the temporal dynamics of resistance genes from bypass and combined-sewer overflows in the receiving waters. Therefore, further studies will be required to better understand the factors that influence the total loadings, peak levels and duration of event-based inputs of resistance factors from bypass and sewer overflows. Catchment-wide measurements of bypass properties (e.g., discharge, and representative values of resistance levels in BP, etc) and inputs during stormwater events combined with hydraulic modelling could provide a clearer picture.

3.4.3. The risk of exposure not only to resistance, but also to ‘multi-resistance’ increases at bypass-receiving waters

To the best of our knowledge, we showed for the first time that WWTP bypass of untreated sewage leads to high relative abundance of potentially problematic multi-resistance factors in bypass-receiving rivers, thus increasing the risk of exposure to multi-resistant bacteria during and shortly after stormwater events.

Untreated sewage contains a resistome much more closely related to the resistome of the human gut compared to effluent (Ju et al., 2019). Thus, the risk of these inputs could be even higher. Many

MGDs containing the ARGs shown in Fig. 3.7b are potentially associated with plasmids, and could be transmitted to previously susceptible cells via horizontal gene transfer (HGT) in the presence of antibiotic-mediated selection. The plasmids containing those MGDs could also persist or evolve in environmental microbial communities, for instance, in the form of attached growths near wastewater-receiving points where continuous anthropogenic disturbance (i.e., EF), and also event-based disturbance (i.e., stormwater events) exist.

Recent studies provide evidence that ARGs could persist even in the absence of antibiotic-driven selection. Plasmids are a key to mediate these processes. For instance, under laboratory condition, it has been shown that ARG-encoding non-mobile plasmids persist in *E.coli* over long timescales under non-selective conditions (Wein et al., 2019). Another study showed that conjugal plasmids encoding ARGs were transmitted to donor cells via HGT at high rates in *E.coli* even without antibiotic-mediated selection (Lopatkin et al., 2017). The ubiquity of HGT in highly dense population (e.g., biofilms), positive selection coupled with other compensations, and/or population dynamics could be contributing factors (Lopatkin et al., 2017). These findings indicate that there is a risk of environmental persistence of disseminated MGDs even in locations where antibiotic levels are not high, such as downstream river locations, and reservoirs (i.e., lakes).

Contig-based taxonomy assignment results in Table 3.1 suggest that three MGDs (k121_1708054, k121_852659, and k121_1425685) retrieved from bypass (M13) and bypass-receiving DS water (M14) are hosted by potential opportunistic pathogens (i.e., *E.coli*, and *Citrobacter* sp.). The risk of potential exposure to multi-resistant *E.coli*, and *Citrobacter* sp. in bypass-receiving water under stormwater events therefore has to be considered. Even though these organisms are often found in the human gut and do not normally lead to serious infections in healthy individuals, the risk of infection still exists once exposed to a high dose of pathogenic strains (i.e., > infectious dose), such as diarrhea in the case of *E.coli* (Hunter, 2003), and urinary tract infection for *Citrobacter* sp. (Abbott, 2011). Even though their relative abundances in our analysis do not appear high (i.e., 0.0038 – 0.0065 % of contigs in each sample), care should be taken to minimize human exposure to these pathogens. Using our current approach, it was not possible to assign taxonomy to many other MGDs, so a comprehensive overview of MGD-host relationships is not provided. One of the ways to increase success rates for contig-based taxonomy assignment in the future could be to obtain longer contigs so that those sequences could have a higher likelihood to include taxonomic markers (e.g., housekeeping genes). While it would still be difficult to assign taxonomy to highly mobile fragments (i.e., plasmid sequences) that are usually shared by many taxa, it might be possible to assign taxonomy at least to chromosomal associated sequences in this way. Practically speaking, applying long-read sequencing, or combining short- and long-read sequencings could help to obtain longer

sequenced or *de-novo* assembled fragments, thus increase the success rate of taxonomic assignment.

Conclusions

- WWTPs are potentially important intervention points for preventing or minimizing discharges of bypass-borne ARGs. Future interventions could be made e.g., by increasing the proportion of treated wastewater discharges to incoming combined-sewages during stormwater events.
- Temporal persistence of bypass-borne ARGs in the receiving water should be considered when interpreting the fate of aquatic ARGs during and after stormwater events. Transport of upstream combined-sewage inputs could be the reason – future study involving hydraulic aspects is required.
- The risk of exposure to multi-resistance risk factors increased profoundly in the bypass-receiving river due to bypass inputs during stormwater events.
- Large portion of bypass-borne MGDs were expected to be associated with plasmids – proper interventions for tackling discharges of bypass-borne MGDs are required to prevent potential persistence and evolution of those factors in the environment.

Data availability

The raw sequencing data are deposited in NCBI-Sequence Read Archive under BioProject: PRJNA733009 (Reviewer link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA733009?reviewer=ee6tqcr189e0ft1ujlbc389e4t), and will be publicly available upon acceptance of this manuscript to the journal. All the other datasets (that were not shown in this publication) and R codes will be available at Eawag Research Data Institutional Collection (https://opendata.eawag.ch/) upon acceptance of this manuscript to the journal.

Author contributions

J.L and H.B designed this study, and participated in all stages of the work as main authors. K.B participated in planning and performing field/laboratory works.

Supplementary Information

Bioinformatics work-flow

For contig-based analysis using metagenomic assembly, we used the work-flow suggested in our previous studies. In brief, 1) *de-novo* assembly was performed for the filtered reads using MEGAHIT v1.2.9 (Li et al., 2015), 2) open reading frames (ORFs) were predicted using Prodigal v2.6.3 with default parameters (Hyatt et al., 2010), 3) annotation was performed using BLAST v2.9.0 (Altschul et al., 1990) (cutoffs: E-value ≤ 1E-07; percentage of identical matches (P_{ident}) ≥ 85.0 %; query coverage per high-scoring segment pair (Q_{HSP}) ≥ 60.0 %)) against CARD v3.1.0 protein homolog model (Alcock et al., 2020) for ARGs, and INTEGRALL v1.2 (Moura et al., 2009) for *intI1*, 4) read mapping against assembled contigs and ORFs was performed using Bowtie2 v2.3.2 (Langmead and Salzberg, 2012), and 5) depths were calculated using Samtools v1.9 (Li et al., 2009), and average coverages for contigs and ORFs were calculated according to Albertsen et al. (2013).

Network analysis of co-located genes within the same contig

Given that two genes are co-located in the same assembled contig, a directed network could be produced from two vertices (two co-located genes) and the edge information which could link the relationship between them. In this study, we defined the edge as "Co-occurrence frequency", specifically, the proportion of the number of contigs that contain both genes (A and B) to the number of contigs containing the gene A ($E_{contigA \rightarrow contigB}$ in the *eq.1*).

First of all, the contigs containing ≥ 2 genes of concern (i.e., ≥ 2 co-located ARGs in our case) should be listed-up beforehand, and this will be the starting point. From the information on two vertices (i.e., the list of co-located ARGs and the corresponding contig IDs), an adjacency matrix could be produced. Then, a topology matrix could be produced according to the following operation (Cormen et al., 2001;Ort et al., 2009).

$$T = (I - Adj)^{-1} \quad (eq.3)$$

Where, T indicates topology matrix; I denotes identity matrix; Adj indicates adjacency matrix

The structure of topology matrix (T) is suggested below:

$$T_{(n+m)\,(n+m)} = \begin{array}{c|cccccc} (n+m)\times(n+m) & \boldsymbol{ARG_1} & \cdots & \boldsymbol{Contig_1} & \cdots & \boldsymbol{Contig_m} \\ \boldsymbol{ARG_1} & a_{1\,1} & \cdots & a_{1\,(n+1)} & \cdots & a_{1\,(n+m)} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \boldsymbol{ARG_n} & a_{n\,1} & \cdots & a_{n\,(n+1)} & \ddots & a_{n\,(n+m)} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \boldsymbol{Contig_m} & a_{(n+m)\,1} & \cdots & a_{(n+m)(n+1)} & \cdots & a_{(n+m)(n+m)} \end{array}$$

Where, n indicates the number of vertices for ARGs; m indicates the number of vertices for contigs; $a_{i\,j}$ indicates the element of i^{th} row and j^{th} column, and has the binary value of 0 or 1 (0 for absence, and 1 for presence)

The sub-matrix (shaded in red above, defined hereby as ${T^*}_{n\,m}$) could be selected, and this sub-matrix has following properties:

1) The value of element "1" indicates "Presence" , and "0" indicates "Absence" of the ARG (labeled in its corresponding row).
2) The column sum indicates the total number of ARGs that the corresponding contig (labeled in its corresponding column) contains. For instance, $\sum_{i=1}^{n} a_{i\,1}$ (shaded in green in ${T^*}_{n\,m}$ below) indicates the total number of ARGs that the *Contig*$_1$ contains.
3) The row sum indicates the total number of contigs with which the corresponding ARG (labeled in its corresponding row) is associated. For example, $\sum_{i=1}^{m} a_{1\,i}$ (shaded in yellow in ${T^*}_{n\,m}$ below) indicates the total number of contigs that the *ARG*$_1$ is associated with.

$$
{T^*}_{n\,m} = \begin{array}{c|ccc}
n \times m & \boldsymbol{Contig_1} & \cdots & \boldsymbol{Contig_m} \\
\boldsymbol{ARG_1} & a_{1\,1} & \cdots & a_{1\,m} \\
\boldsymbol{ARG_2} & a_{2\,1} & \cdots & a_{2\,m} \\
\vdots & \vdots & \ddots & \vdots \\
\boldsymbol{ARG_n} & a_{n\,1} & \cdots & a_{n\,m}
\end{array}
$$

Where, n indicates the number of vertices for ARGs; m indicates the number of vertices for contigs; $a_{i\,j}$ indicates the element of i^{th} row and j^{th} column, and has the binary value of 0 or 1 (0 for absence, and 1 for presence)

Therefore, the total number of contigs containing *ARG*$_1$ ($n(set1)$) could be defined as:

$$n(set1) = \sum_{j=1}^{m} a_{1\,j}$$

Furthermore, the total number of contigs containing *ARG*$_1$ and *ARG*$_2$ [$n(set1 \cap set2)$] could be defined as:

$$set1 \cap set2 = \{a_{2\,J}\}, \qquad J \in \{x \mid a_{1\,x} = 1\}$$

$$\therefore n(set1 \cap set2) = \text{ sum of all the elements in } \{a_{2\,J}\}$$

Finally, $E_{ARG_1 \rightarrow ARG_2}$ could be calculated by $n(set1 \cap set2)/n(set1)$ according to the *eq.2*. The whole algorithm was written in R, and all codes and datasets are available in the first author's GitHub page (https://github.com/myjackson).

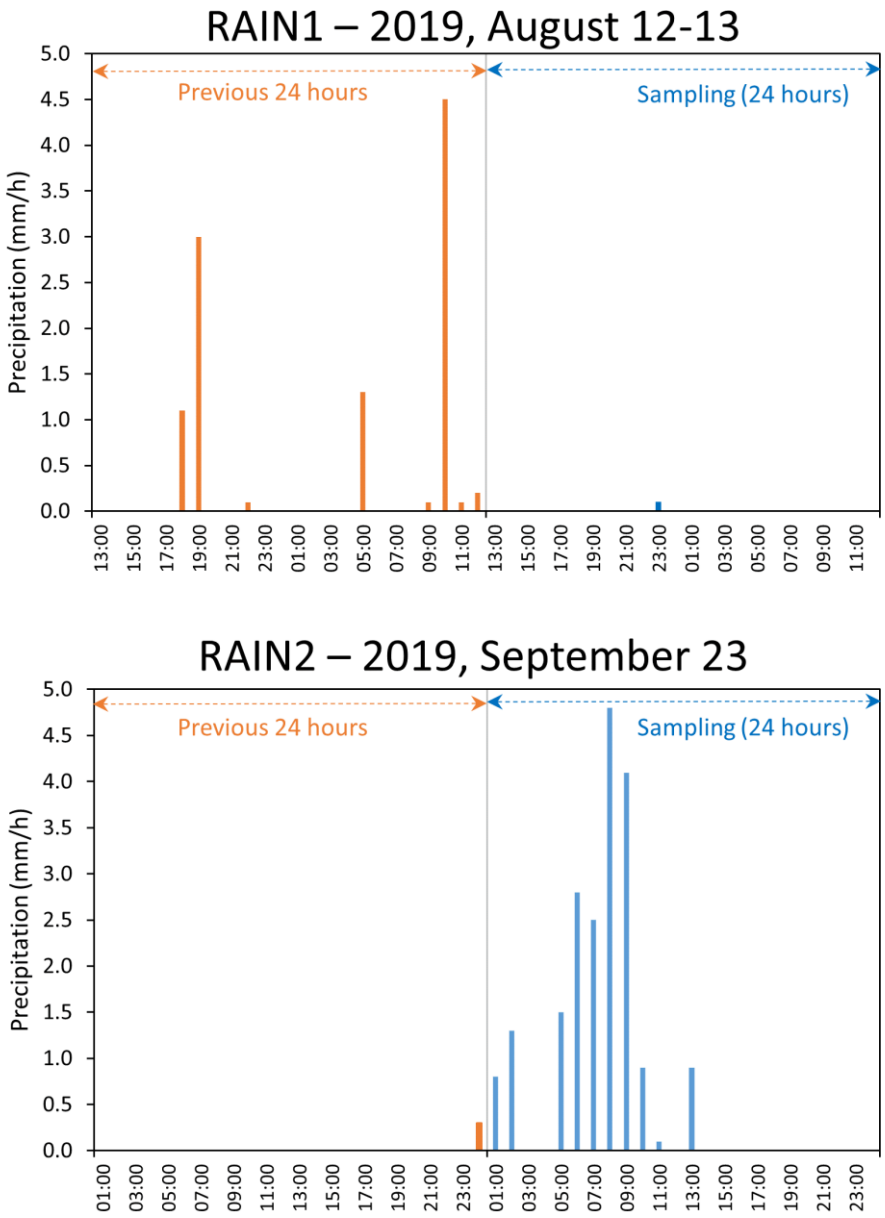


Figure S3.1. Precipitation data during stormwater samplings on August 12-13 (RAIN1), and September 23, 2019 (RAIN2) in Münchwilen, Switzerland. Data obtained from MeteoSwiss – IDAWEB (https://gate.meteoswiss.ch/idaweb/).

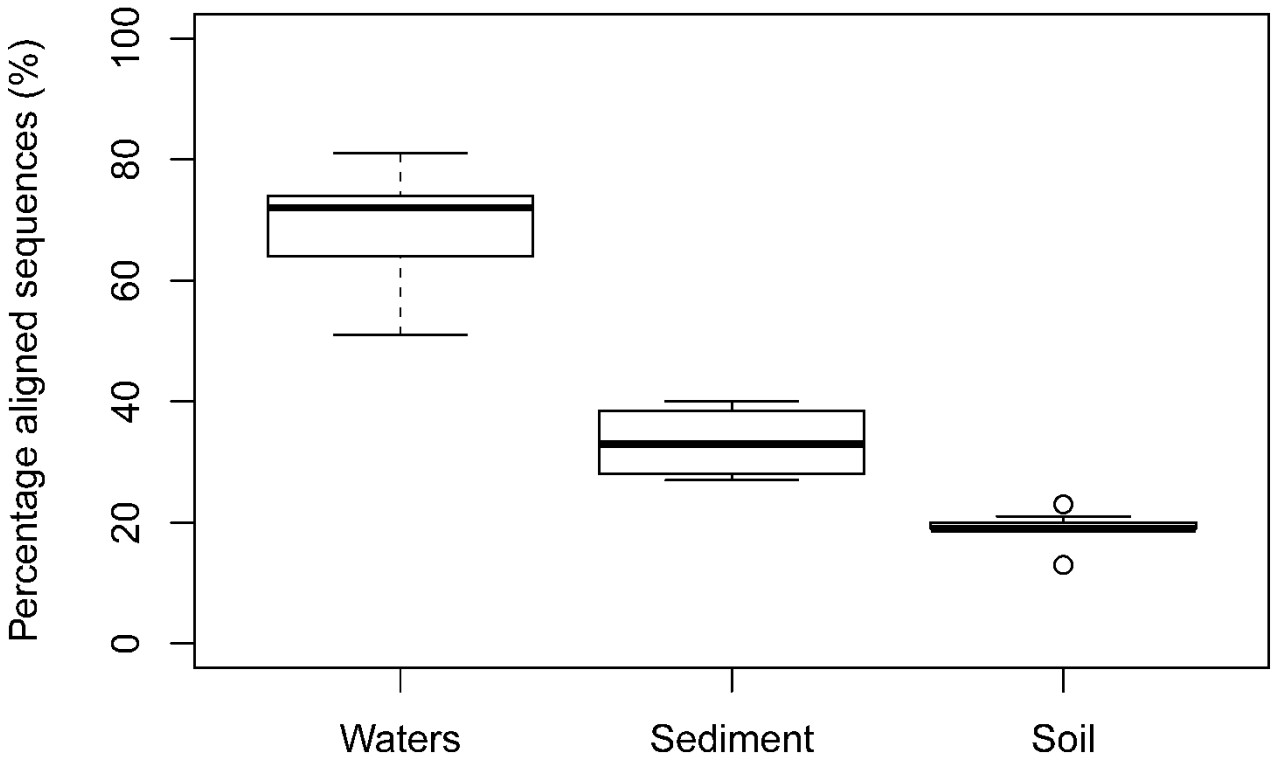


Figure S3.2. Assembly efficiencies (i.e., percentage aligned sequences, calculated using eq.1) by each sample type. N (sample size) = 32 for waters (incl. bypass, treated effluents, and river waters), 4 for sediments, and 7 for soils.

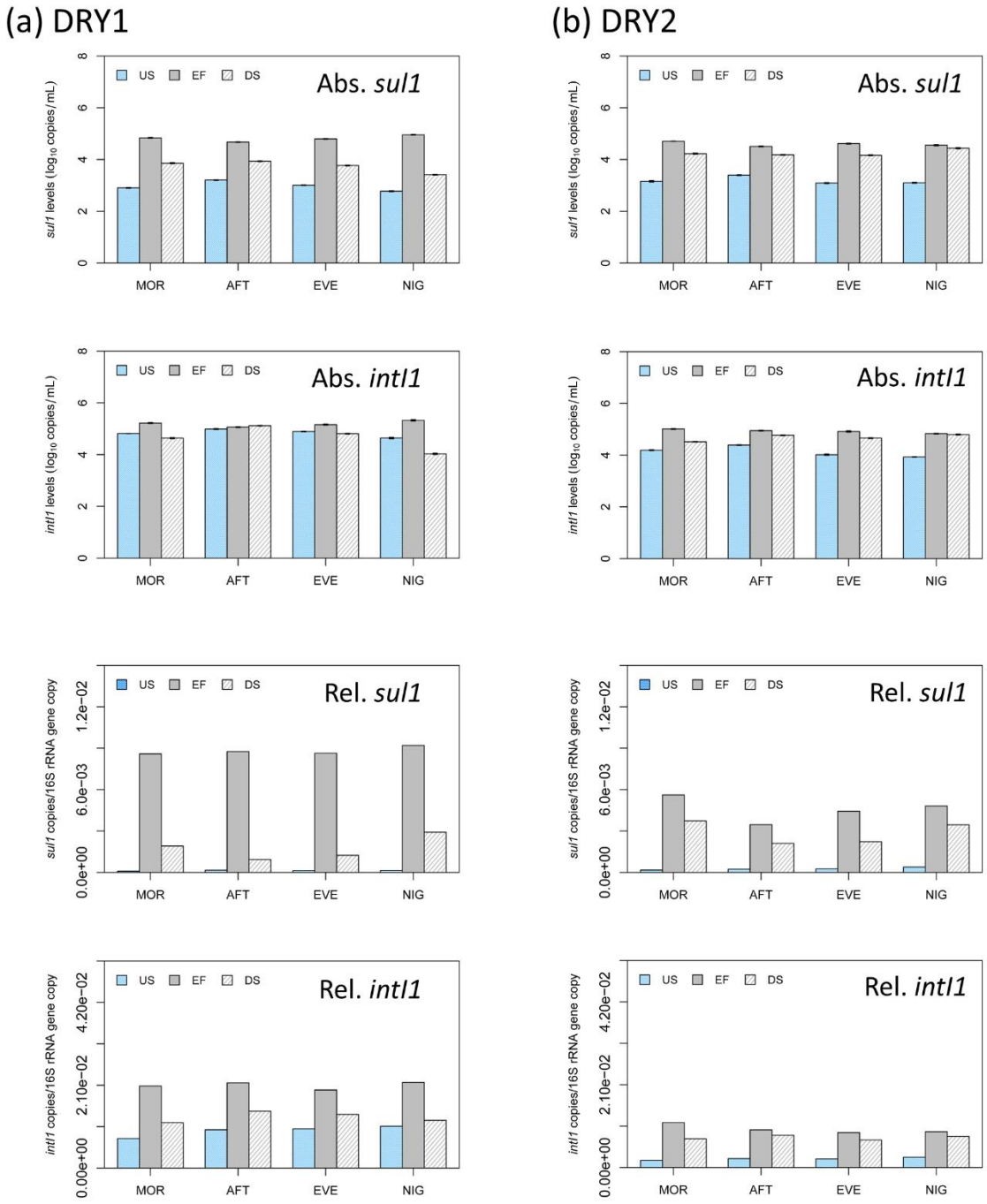


Figure S3.3. Absolute (Abs.) and relative abundance (Rel.) of two resistance indicators (*sul1* and *intI1*) in two dry-weather samplings (DRY1 on the left – a, and DRY2 on the right – b) in the River Murg in Münchwilen, Switzerland.

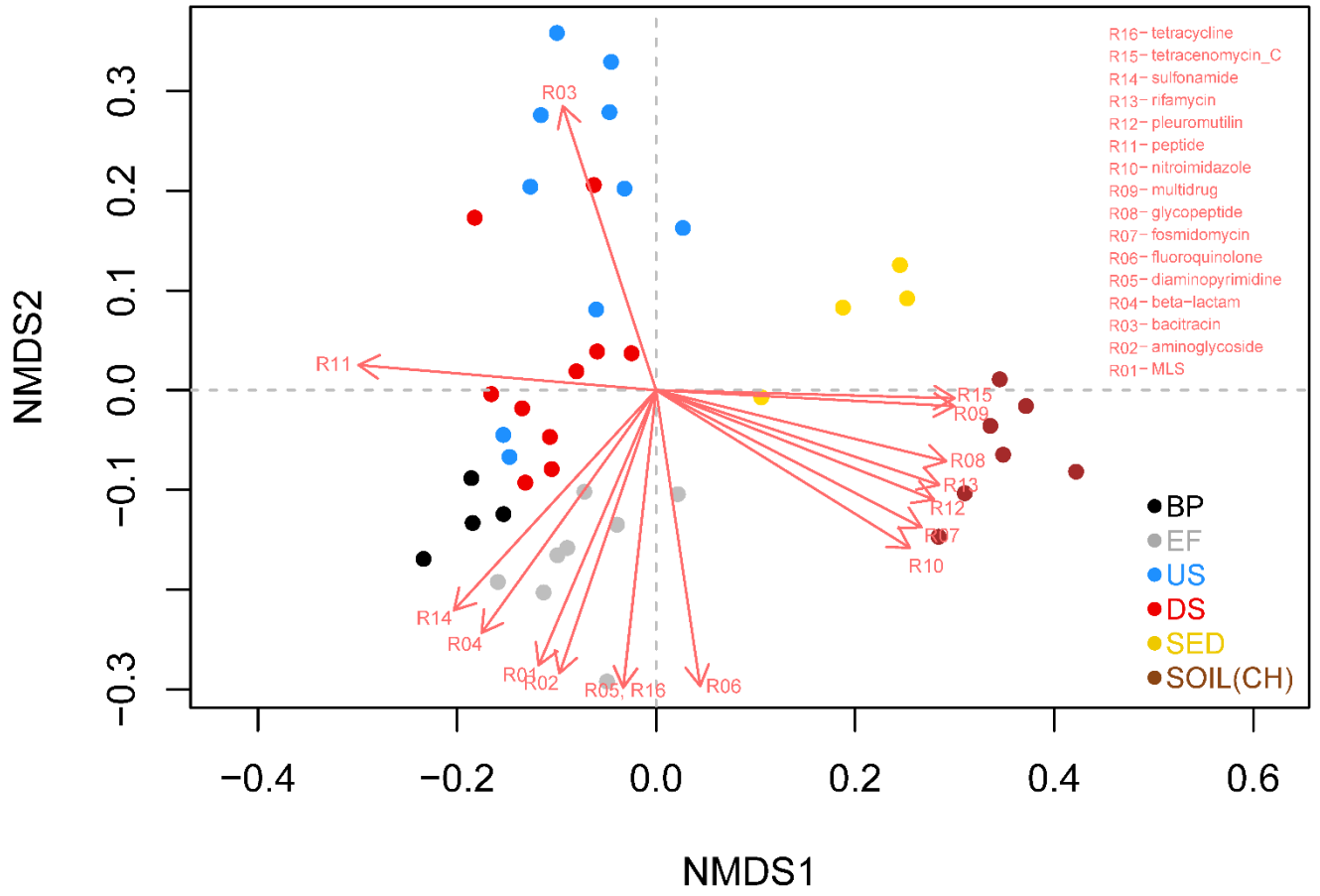


Figure S3.4. Ordination and bioplot analysis using non-metric multi-dimensional scaling (NMDS) for resistome. The analysis was performed in terms of resistance classes. A total of 16 major classes of resistance that were significantly correlated with the ordination (p-value < 0.0001) were displayed as vectors (in red).

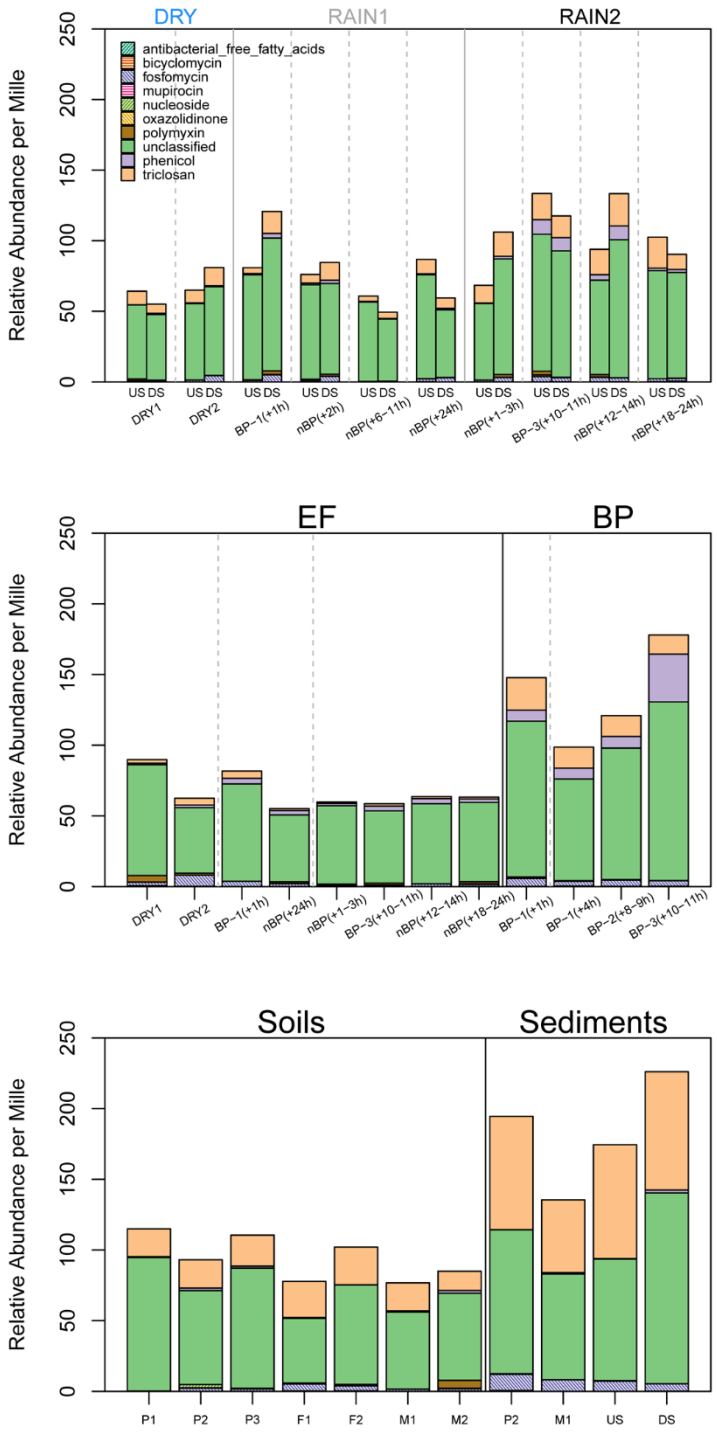


Figure S3.5. Resistome profiles by class of antibiotic resistance in river waters (upstream and downstream waters), wastewaters (bypass and effluents), soils and sediments. A total of 10 classes were shown: antibacterial-free-fatty-acids, bicyclomycin, fosfomycin, mupirocin, nucleoside, oxazolidinone, polymyxin, unclassified, phenicol, triclosan resistance genes. Unit: Relative abundance per Mille (i.e., ARG reads per 1000 16S rRNA gene reads).

Table S3.1. Sample pooling strategy for stormwater samplings. Samples were taken hourly, and those with the same label (e.g., (+4 hrs) for 3 samples from 15:00 to 17:00 in RAIN1) were pooled together. The time points when bypass events were happened were labeled as 'BP (i.e., BP-1, BP-2, BP-3)'.

Campaign	Sample	Sampling time (24 hours)																						
		01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	00:00
DRY1	US	NIG					MOR						AFT						EVE						NIG
	DS	NIG					MOR						AFT						EVE						NIG
	EF	NA	NA	NA	NA	NA	NA	NA	MOR				AFT						EVE						NIG
DRY2	US	NIG					MOR						AFT						EVE						NIG
	DS	NIG					MOR						AFT						EVE						NIG
	EF	NIG					MOR						AFT						EVE						NIG
RAIN1	US	nBP(+12~17h)					nBP(+18~23h)						nBP (+24h)	BP-1 (+1h)	nBP (+2h)	nBP(+3~5h)			nBP(+6~11h)						
	DS	nBP(+12~17h)					nBP(+18~23h)						nBP (+24h)	BP-1 (+1h)	nBP (+2h)	nBP(+3~5h)			nBP(+6~11h)						
	EF	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	nBP (+24h)	BP-1 (+1h)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	BP	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	BP-1 (+1h)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
RAIN2	US	nBP(+1~3h)			BP-1 (+4h)		nBP(+5~7h)		BP-2 (+8~9h)		BP-3 (+10~11h)		nBP(+12~14h)			nBP(+15~17h)			nBP(+18~24h)						
	DS	nBP(+1~3h)			BP-1 (+4h)		nBP(+5~7h)		BP-2 (+8~9h)		BP-3 (+10~11h)		nBP(+12~14h)			nBP(+15~17h)			nBP(+18~24h)						
	EF	nBP(+1~3h)			BP-1 (+4h)		nBP(+5~7h)		BP-2 (+8~9h)		BP-3 (+10~11h)		nBP(+12~14h)			nBP(+15~17h)			nBP(+18~24h)						
	BP	NA	NA	NA	BP-1 (+4h)	NA	NA	NA	BP-2 (+8~9h)		BP-3 (+10~11h)		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Sampling starts		Sampling ends																							

Table S3.2. Quantitative PCR (qPCR) primers and probes used in this study.

Primer	Genes	Assay type	Sequences (5' to 3')	Annealing temp. (°C)	References
qSUL653f			CCGTTGGCCTTCCTGTAAAG		
qSUL719	*sul1*	TaqMan	TTGCCGATCGCGTGAAGT	60	(Heuer and Smalla, 2007)
tpSUL1-Probe			*FAM*-CAGCGAGCCTTGCGGCGG-*BHQ1*		
intI1-F			GCCTTGATGTTACCCGAGAG		
intI1-R	*intI1*	TaqMan	GATCGGTCGAATGCGTGT	60	(Barraud et al., 2010)
intI1-Probe			*FAM*-ATTCCTGGCCGTGGTTCTGGGTTTT-*BHQ1*		
BAC349-F			AGGCAGCAGTDRGGAAT		
BAC806-R	16S rRNA	TaqMan	GGACTACYVGGGTATCTAAT	53*	(Takai and Horikoshi, 2000)
BAC516F-Probe			*FAM*-TGCCAGCAGCCGCGGTAATACRDAG-*BHQ1*		

* These values were modified from the original references, and previously optimized in our laboratory settings.

Table S3.3. Selected information for quantitative PCR (qPCR) validation according to the MIQE guideline (Bustin et al., 2009). LOD indicates the limit of detection; NTC denotes non-template control (i.e., molecular grade H_2O replacing DNA templates). Technical quintuplicates were applied to the qPCR standard curves.

qPCR primer	Slope	Specificity verification method	qPCR-efficiency	Y-intercept	Linear dynamic range (copies - copies)	LOD-copies (Avr. Cp)	S.D. (of Cp) at LOD	Cp of the NTC
sul1	-3.667	Hydrolysis probe (TaqMan)	1.874	41.14	5.0E+07 - 5.0E+01	50 (34.0)	0.42	N.D.
intI1	-3.473	Hydrolysis probe (TaqMan)	1.941	41.43	5.0E+07 - 5.0E+01	50 (35.8)	0.45	N.D.
16S rRNA gene	-3.517	Hydrolysis probe (TaqMan)	1.925	41.35	5.0E+07 - 5.0E+01	50 (32.3)	0.30	32.6

CHAPTER 4.

Model-based prediction of hotspots for antimicrobial resistance contamination in Swiss river networks

Jangwoo Lee[1,2], Christoph Ort[1], and Helmut Bürgmann[1]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, CH-6047 Kastanienbaum or CH-8600 Dübendorf, Switzerland

[2]Department of Environmental Systems Science, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

This chapter is in preparation, and will be submitted to a journal in the near future:

Lee et al., Model-based prediction of hotspots for antimicrobial resistance contamination in Swiss river networks. In preparation.

Abstract

Rivers are considered as one of the major routes by which antimicrobial resistance (AMR) is transmitted in natural environments. Due to the many uses of river water for recreation, irrigation and as a source of drinking water it may be a potential exposure pathway for riparian populations. This motivates development of a tool for predicting aquatic AMR contamination in rivers. In this study, we developed a predictive model for public exposure to aquatic AMR in Swiss river networks using two AMR indicators (i.e., *sul1* and *intI1*) under Graph Theory. The following hypothesis were tested: 1) Wastewater treatment plants (WWTPs) are the main source of anthropogenic AMR, 2) AMR loadings decrease over downstream distance, 3) Background levels of AMR indicators present in non-impacted river water need to be considered. A model employing hypothesis 1) and 3) but rejecting hypothesis 2) provided the best fit with measured validation data. The final model calculates *sul1* and *intI1* loadings (i.e., copies per time) and levels (i.e., copies per volume) at WWTP effluents receiving waters in the entire Swiss river network. We demonstrate potential uses of such a model to predict e.g. a public exposure index for riverine aquatic AMR for each Swiss canton. Our model provides a theoretical framework with which hotspots for riverine aquatic AMR could be prioritized, and thus could be a potentially used as a tool for decision making in future interventions for tackling environmental dissemination of AMR.

Keywords

Antimicrobial Resistance; Exposure; River; Modeling; Wastewater; Antibiotic Resistance Genes

4.1. Introduction

Antimicrobial resistance (AMR) has emerged as one of the main issues in public health. From a One Health and Global Health perspective, various factors and sectors are intertwined with the development and dissemination of AMR (Hernando-Amado et al., 2019). For instance, not only the evolution and spread of AMR in human clinical settings is relevant, but also development and spread of AMR from other anthropogenic (e.g., farms, wastewater treatment plants (WWTPs), pharmaceutical industries, etc.), and natural environments (Hernando-Amado et al., 2019).

Among various natural environments, rivers receive particularly large anthropogenic inputs of AMR from various sources that can subsequently be transmitted to other regions and environments to which the rivers are connected. For instance, evidences of urban, and agricultural inputs along the Poudre and South Platte Rivers in Colorado, USA were identified using quantitative PCR (qPCR) using 2 sulfonamide and 11 tetracycline antibiotic resistance genes (ARGs), and phylotype and phylogenetic analysis of *tetW* as indicators (Storteboom et al., 2010). More recently, anthropogenic resistomes (i.e., entire profiles of ARGs identified by metagenomics) were identified along the Han River which passes through Seoul, Korea (Lee et al., 2020). These examples show that populations living in downstream regions could be exposed to riverine ARGs introduced to the upstream river. This necessitates comprehensive surveillance tool for riverine ARGs, to support planning and prioritizing possible interventions for tackling dissemination of anthropogenic inputs.

One such tool would be models that can predict resistance contamination levels in river systems at the landscape scale. One example of this was developed for the Thames River, UK (Amos et al., 2015). The authors developed a model predicting the prevalence of environmental integron integrase class 1 genes (*intI1*) (i.e., ratio of *intI1* genes to 16S rRNA genes) in sediments at a catchment level (Amos et al., 2015). Their final model was based on a multiple linear regression model taking into account landcover properties in the catchment and seasonal effects, as well as the separately modeled impact of upstream WWTP on each site. This study indicates that landscape scale predictions are possible, and that WWTP are a major driver of AMR levels. These assumptions were also supported by Czekalski et al. (2015). In that study, a similar approach was used – multiple regression modelling between catchment integrated land-use data and *sul1* levels in surface waters from 21 Swiss lakes. The authors found that WWTPs are likely to be drivers of *sul1* levels. However, both studies used a data fitted model, and the transferability to other rivers systems or conditions beyond the fitted data remains unclear. Generally the model-based assessment of anthropogenic antibiotic resistance contamination in aquatic systems remains in its infancy, and connecting contamination level to exposure risk in riverine ‘aquatic’ environments has not been attempted.

As a starting point for our work, we referred to published models predicting micropollutant loadings (i.e., mass per unit time) and levels (i.e., mass per volume) in the Swiss river network, considering WWTPs (with their connected populations and healthcare services) as major sources (Kuroda et al., 2016;Ort et al., 2009). These models could not be directly applied to predict ARG levels for various reasons. Unlike e.g. pharmaceutical micropollutants that can often be sourced to few input sources like hospitals or pharmaceutical industry, or antibiotics that can be scaled to population but that are all channeled through WWTP, the drivers of ARG inputs from WWTP were initially unclear. Further, ARGs are biological indicators, thus their fates in the river could be more dynamic than micropollutants, e.g., resistant bacteria could proliferate in the river system via horizontal and/or vertical gene transfer (Vikesland et al., 2017) or could degrade due to cell death, predation or other mechanisms. Also intrinsic, or natural existence of ARGs was known in natural environments (Cox and Wright, 2013), as it were, anthropogenic inputs are not the only source.

Therefore, in this study, we developed a predictive model for riverine AMR contamination using two of the most widely used anthropogenic AMR indicators – *sul1* and *intI1* (Berendonk et al., 2015;Gillings et al., 2015). We hypothesized that WWTP effluents are the major sources of *sul1* and *intI1*. Cumulative loadings of those indicators in the entire river networks were calculated using Graph Theory similarly to the previously established micropollutant models (Kuroda et al., 2016;Ort et al., 2009). The theoretical framework provides a simple mathematical representation of the river network and is thus the model is easily transferable to other river networks. We built the model to allow testing the following hypothesis:

1) WWTP inputs are the main sources of wastewater-borne AMR indicators (i.e., *sul1* and *intI1*) and suffice to model their loadings and levels on the landscape scale.

2) Removal processes of wastewater-borne AMR indicators (i.e., *sul1* and *intI1*) exist along the downstream transport distance. To parameterize the removal processes, we referred to our previous publication where downstream fates of AMR indicators were studied in the River Suze catchment near Villeret, Switzerland (Lee et al., 2021).

3) The background levels of *sul1* and *intI1* intrinsically occurring in surface waters is non-negligible and needs to be taken into account (Czekalski et al., 2016b;Lee et al., 2021).

The final goal of this study was to estimate loadings and levels of AMR indicators at wastewater-receiving waters in the entire river networks in Switzerland with a non-fitted model. We then explored the potential of such a model to predict potential exposure of the population, as a tool to prioritize hotspots of exposure risk.

4.2. Materials and Methods

4.2.1. The Core Algorithm – River Network Analysis Using Graph Theory

Total 745 WWTPs which serve 97 % of Swiss population were included in this Model. Those WWTPs discharge treated effluents to inland surface waters (i.e., rivers or lakes). A large portion of WWTPs (i.e., total 631 WWTPs serving 83 % of total population) are connected to rivers. Directed networks were produced for those river-connected WWTPs. Based on the assumption that the long residence times in Lakes would result in a "reset" of the resistance indicator load, sub-networks were disconnected from each other when they are connected via a lake. The network was calculated using the following matrix operations as previously done for micropollutant models (Kuroda et al., 2016;Ort et al., 2009). For instance, all the edges between two connected nodes could be summarized in a topology matrix. A graphical abstract for this calculation was shown in Fig. S4.1.

$$T_{ij} = (I - adj)^{-1} \quad \text{(eq.1)}$$

Where, T indicates the topology matrix, I denotes the identity matrix, and adj indicates the adjacency matrix.

Assuming edges as loadings and nodes as WWTPs, the accumulated loadings from all connected WWTPs upstream at each WWTP can be calculated by summing up the loading values in each column of topology matrix.

4.2.2. Model Assumption 1 – ARG Loadings from WWTPs

Cumulative loadings from upstream WWTPs were calculated using the topology matrix (shown in eq.1) at each wastewater-receiving node. Loadings (ARG copies per time) are calculated by effluent discharge (volume per time) multiplied by ARG level in the effluent (ARG copies per volume). Therefore, variation of ARG loadings among WWTPs could originate either from ARG level ($\bar{C}_{EF}$) or from WWTP discharge (Q_{EF}). We compared the variance potential between those two factors in terms of coefficient of variation (CV, i.e., standard deviation divided by mean) using the data obtained from total 42 samplings at 30 WWTPs for *sul1*, and 30 samplings at 21 WWTPs for *intI1* referring to our previous projects (Fig. S4.2) (Czekalski et al., 2016b;Ju et al., 2019;Ju et al., In Preparation;Lee et al., 2021). As a result, the CV was much higher for Q_{EF} than ARG levels (Fig. S4.2) – variation of ARG loadings among WWTPs thus largely derives from Q_{EF}. Statistical analysis of the ARG levels in WWTP effluent did not provide a suitable model to predict ARG levels by, e.g. the type of treatment process. In the present study we therefore neglected variance in ARG levels in effluents and chose the median of the 42 available effluent values for *sul1* (30 values for *intI1*) as a single representative value. To calculate ARG loadings from WWTPs, the representative values (i.e., 9.9E+03 copies/mL for $\bar{C}_{EF\_sul1}$; 1.5E+04 copies/mL for $\bar{C}_{EF\_intI1}$) were multiplied by the corresponding Q_{EF}. The Q_{EF} values for WWTPs were calculated in the same way that was previously published (Ort et al.,

2009), based on the assumption of 400 L day^{-1} per capita multiplied by population equivalent for each site.

4.2.3. Model Assumption 2 – Downstream Fate of Wastewater-borne ARGs

We studied two scenarios: 'No Decay', i.e. conservative transport of the resistance indicators and 'Decay', where a decay function describes the decrease of the resistance indicator with downstream distance. The No Decay scenario assumes no decreases of *sul1* loadings with downstream distance (Model-1). Cumulative loadings at WWTP$_i$ (CL_i) could be calculated as follows:

$$CL_i = L_i + \sum_{k=1}^{j} L_k \qquad \text{(Model-1: No Decay)}$$

Where, CL_i indicates the cumulative loading at WWTP$_i$; L_i denotes the ARG loading discharged from WWTP$_i$; *j* indicates the total number of upstream WWTPs connected to WWTP$_i$

We previously observed in the River Suze catchment near Villeret, Switzerland, that wastewater-borne *sul1* loadings decreased with downstream distance from the point of discharge under base-flow conditions (Lee et al., 2021). Hypothesizing that this observation can be generalized, we derived a decay function for *sul1* from the River Suze catchment data (Fig. S4.3). This function was applied to upstream WWTP loadings, which leads to the following equation for calculating cumulative loadings below each WWTP node:

$$CL_i = L_i + \sum_{k=1}^{j} f(D_k) \cdot L_k \qquad \text{(Model-2: Decay)}$$

Where, $f(D_k) = e^{1.5e-04 \times D_k}$ indicates the value of the decay function for the cumulative distance (D_k) between WWTP$_i$ and WWTP$_k$

For *intI1*, we observed the loading did not profoundly decrease or increase over downstream distance (Fig. S4.3), so we only analyzed the No Decay case in Model-1.

4.2.4. Model Assumption 3 – Natural Background Levels of *sul1* and *intI1*

Another observation from our previous projects was that *sul1* and *intI1* genes exist even in the headwater sections of rivers where no point-sources are known (Fig. S4.4) (Czekalski et al., 2016b;Ju et al., In Preparation;Lee et al., 2021). Therefore, we assumed that certain natural background levels (i.e., median values in Fig. S4.4; 3.6E+02 copies/mL for *sul1*, and 1.5E+03 for *intI1*) of *sul1* and *intI1* are always present in incoming waters upstream of each wastewater-receiving point. The background loading (BL, i.e., discharge of incoming water upstream × background level) of *sul1* and *intI1* was calculated for each point from the estimated natural background level and discharge, and added to loadings according to Model-1 and Model-2 as follows:

$$CL_i = BL_i + L_i + \sum_{k=1}^{j} L_k \quad \text{(Model-3: No Decay + Background Loading)}$$

$$CL_i = BL_i + L_i + \sum_{k=1}^{j} f(D_k) \cdot L_k \quad \text{(Model-4: Decay + Background Loading)}$$

Where, $(BL_i = \bar{C}_{US} \cdot Q_{US\_i})$ and $(Q_{US\_i} = Q_{95\%\_i} - Q_{EF\_i})$; $\bar{C}_{US}$ indicates the background concentration of *sul1* or *intI1*; Q_{US_i} indicates incoming upstream discharge at $WWTP_i$; $Q_{95\%\_i}$ denotes the flow exceeded 95 % (i.e., 347 out of 365 days) of the year at the effluent-receiving point of $WWTP_i$ predicted according to Ort et al. (2009); *j* indicates the total number of upstream WWTPs connected to $WWTP_i$.

4.2.5. Validation Samplings and Molecular Biological Analysis

Samples were taken from September 22 to November 01, 2020 at 46 locations in 17 river catchments located in 11 different cantons in Switzerland (Table S4.1). Samplings were mostly performed under dry-weather (i.e., no precipitation in the previous 24 hours), or in a few cases after minor precipitation (i.e., < 5.0 mm per day in the previous 24 hours) (Source MeteoSwiss). We thus assume that the impact of stormwater-related pollution inputs (e.g., surface runoff, sediment resuspension, wastewater bypass under stormwater events, etc) (Lee et al., in Preparation) were negligible. Under these conditions treated effluents from WWTPs were hypothesized to be the dominant source of ARG pollution. Samples were taken up- and down-stream at ≥ 260 m distance from effluent-discharge points (Table S4.1). Assuming that cross-sectional mixing was completed after 150 m (Ort et al., 2009), this ensures complete mixing of discharged wastewater and river water at the point of sampling. At each location, 2 L of surface water sample was obtained, stored at 4 °C in sterile water containers, and transported to the laboratory within 4 hours. After storage at 4 °C overnight, water samples were processed as outlined in our previous publication (Lee et al., 2021). In short, samples were filtered through 0.2 µm pore-size filters using sterile filtration units, and the concentrated biomass on the filters was subjected to DNA filtration using DNeasy Power-Water Kit (Qiagen, Germany) following manufacturer's instruction. DNA qualities and concentrations were measured using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, USA) (Data not shown). Two ARG markers (i.e., *sul1* and *intI1*) were quantified using qPCR as outlined in our previous study (Lee et al., 2021), using the primers given in (Lee et al., 2021) Validation information for qPCR according to MIQE guidelines (Bustin et al., 2009) is given in Table S4.2. Samples were measured in triplicate, and quantification standards were measured in quintuplicate. The absence of qPCR inhibition was confirmed by comparing the *sul1* levels (copies/mL) analyzed from 10 (d10) and 100 (d100) times diluted samples – no significant differences were observed, and the values obtained from 10 times diluted samples were used for further analysis . The standard deviation of Cp values among technical triplicates of samples were ≤ 0.5, and all reported values were above the limit of

detection (LOD) and successfully quantified according to the definition described in Czekalski et al. (2015). An extraction control was prepared as follows: 500 mL of nanopure distilled water was filtered, and DNA was extracted, as described above. The Cp values of *sul1* and *intI1* from extraction controls were always below the corresponding LOD (Table S4.2). The total copies quantified in the extracted DNA sample were divided by the filtered sample volume to calculate ARG levels with the unit 'copies per volume'.

4.2.6. Model Validation and Selection

Linear regression analysis ($y = ax + b$) between measured and predicted loadings was performed for each model. The best model with highest R^2 and lowest P-value was selected, and used in the further downstream analysis.

To obtain measured loadings, the measured levels of *sul1* and *intI1* were multiplied by either predicted river discharges for the locations where the estimated base-flow discharges (i.e., $Q_{95\%}$) are available (i.e., 'E' category in Table S4.1 – the last column) according to Ort et al. (2009), or measured river discharges ($Q_{Measured}$) provided by federal or cantonal gauging stations nearby (i.e., 'M' category in Table S4.1 – the last column). The predicted river discharges (i.e., $Q_{95\%}$) were derived under the 'base-flow' assumption, so this could be somewhat different from the actual values at our sampling dates. To compensate for this variation, we derived the 'flow-correction factor' (as follows), and multiplied this to measured loadings where $Q_{95\%}$ was used (e.g., $ML_i = C_i \cdot Q_{95\%} \cdot F_i$, where, ML_i indicates the measured loading at i, C_i denotes the levels of AMR indicators at i, F_i indicates the flow-correction factor at i). For instance, 1) The baseflow statistics (the flow exceeded, on average, on 347 out of 365 days for cantonal stations; annual minimum flow rate averaged over 7 days for federal stations) was obtained from the nearest upstream or downstream gauging station, 2) the measured river discharge (by the gauging station) at the sampling date was divided by the base-flow statistics, and this ratio was defined as 'flow-correction factor (F_i)'.

Model-1 to -4 were applied to the matrices representing the river network in the graph-theoretical algorithm, which provides predicted loadings for each node (WWTP) in the network. As the measurements for our validation campaign were taken at various distances from the WWTP, and loadings are not conserved in the models considering decay, we need to calculate the actual loadings at the sampling points. Therefore, 'distance between wastewater discharge points and validation sampling points' and the 'flow-correction factor' were considered e.g. for downstream validation points (i.e., 'DS' category in Table S4.1 – 'Categories' column), as follows:

$$CL_i = f(D_i) \cdot L_i + \sum_{k=1}^{j} f(D_k + D_i) \cdot L_k \qquad \text{(Model-2.2: Decay)}$$

$$CL_i = BL_i + L_i + \sum_{k=1}^{j} L_k \quad \text{(Model-3.2: No Decay + Background Loading)}$$

$$CL_i = BL_i + f(D_i) \cdot L_i + \sum_{k=1}^{j} f(D_k + D_i) \cdot L_k \quad \text{(Model-4.2: Decay + Background Loading)}$$

Where, D_i indicates the downstream distance between $WWTP_i$ and actual sampling location; $BL_i = \bar{C}_{US} \cdot Q_{Measured\_i}$ or $\bar{C}_{US} \cdot Q_{US\_i} \cdot F_i$. *No modification was applied for Model-1 where neither background loadings nor downstream decay function were assumed.

For upstream validation points (i.e., 'US' category in Table S4.1 – 'Categories' column), the loadings from the nearest downstream WWTPs were not considered (i.e., $L_i = 0$).

After selecting the best model by the abovementioned procedures, 'predicted levels' were also calculated, e.g., dividing the predicted loading by the river discharge. This predicted levels were also validated using linear regression analysis.

4.2.7. Uncertainty Propagation

We assumed that uncertainty in our model derives mostly from prediction errors originating from variability of *sul1* and *intI1* levels in treated wastewaters ($\bar{C}_{EF\_sul1\ or\ intI1}$) (Fig. S4.4), and in natural background levels of those ARGs ($\bar{C}_{US\_sul1\ or\ intI1}$) (Fig. S4.4) because we do not currently have a model to describe their variation among sites. Therefore, prediction errors from these sources were propagated using a Monte-Carlo randomization simulation. Two variables (i.e., $\bar{C}_{EF\_sul1\ or\ intI1}$, and $\bar{C}_{US\_sul1\ or\ intI1}$) were randomized under uniform distribution assumption within the first (25 %) and third (75 %) interquartile ranges (IQR) which were empirically referring to Fig. S4.4. The simulation was repeated for 999 times, and the first and third IQR for each prediction value were visualized as error bars in the final model.

4.2.8. Prediction of Public Exposure to Aquatic ARGs

Potential for public exposure to those indicators by canton was estimated as follows:

Indicator of Exposure Potential (by canton) $= \sum_{i=1}^{n} (sul1_{WWTP_i}\ or\ intI1_{WWTP_i} \times RESID_{WWTP_i})$ (eq.2)

Where, $sul1_{WWTP_i}\ or\ intI1_{WWTP_i}$ indicates the predicted *sul1* or *intI1* level at receiving point of the $WWTP_i$; $RESID_{WWTP_i}$ denotes the resident population connected to the $WWTP_i$; n is total number of WWTPs in a canton

4.2.9. Statistics and Visualization

All statistics and calculations (i.e., regression analysis, river network analysis, and Monte-Carlo simulation) were performed in R. Built-in functions were used for regression analysis (i.e., nls() for non-linear analysis, and lm() for linear analysis). The river network analysis was performed with a

newly developed script relying on the R package 'igraph' for core graph algorithms. A Monte-Carlo randomization simulation was also performed in R. All the scripts and datasets will be available in the Eawag institutional open data repository (ERIC, https://opendata.eawag.ch/) upon publication of this chapter in an academic journal. Selected model algorithms will also be uploaded to the first author's GitHub page (https://github.com/myjackson). All the figures were produced using built-in R functions (e.g., plot(), boxplot(), etc), except for the maps visualizing hotspots for AR indicators which were produced using ArcGIS.

4.3. Results and Discussion

4.3.1. Choosing the best model for describing aquatic AMR

The best model among different assumptions (Model-1 to 4) was selected based on the results (R^2 and P-values) of linear regression analysis between expected (i.e., modeled) and measured loadings.

Whichever assumptions we tested, all the models were highly significantly correlated with the measured loadings (P≤ 1.4E-05) (Table 4.1 and Fig. 4.1). Since WWTP input is the only source considered in this model, this indicates that WWTPs are an important sources of ARGs in Swiss river networks. Even the models without additional assumptions (i.e., without downstream decay and background level assumptions) were statistically significant for both *sul1* and *intI1* (Model-1 in Table 4.1). However, under Model-1, measured values were higher than values predicted by the model in most cases for both *sul1* and *intI1* (Fig. 4.1). This suggests that additional sources of those indicators should be taken into account to improve model predictions. Accordingly, the R^2 for the model decreased when downstream decay was considered (Model-2 in Table 4.1, and Fig. 4.1). This may indicate that the decay function we observed in the River Suze catchment near Villeret cannot be generalized to the Swiss river network. More work will be required to determine if incorporating a decay function would improve a model that considers additional sources more accurately, if the decay function we derived generally overestimates the decrease, or if downstream fates of aquatic ARGs might differed by site, i.e. are a function of unknown environmental or biological factors. Model-3, which takes into account background ARG levels under no-decay assumption achieved the best overall prediction for both indicators (Fig. 4.1). For this reason we selected Model-3 for further detailed analysis. Model-3 showed reasonably good correlation with measured loadings (R^2=0.72, P = 1.8E-13 for *sul1*, and R^2=0.76, P=5.0E-15 for *intI1*) (Fig. 4.1). The linear regression was close to the 1:1 line for *sul1*, indicating that the model captured the overall trends of this indicator in the Swiss river network. For *intI1* the linear regression was always above the 1:1 line indicating that the model underestimated the true loadings, although the offset was not large compared to the range of measured loading.

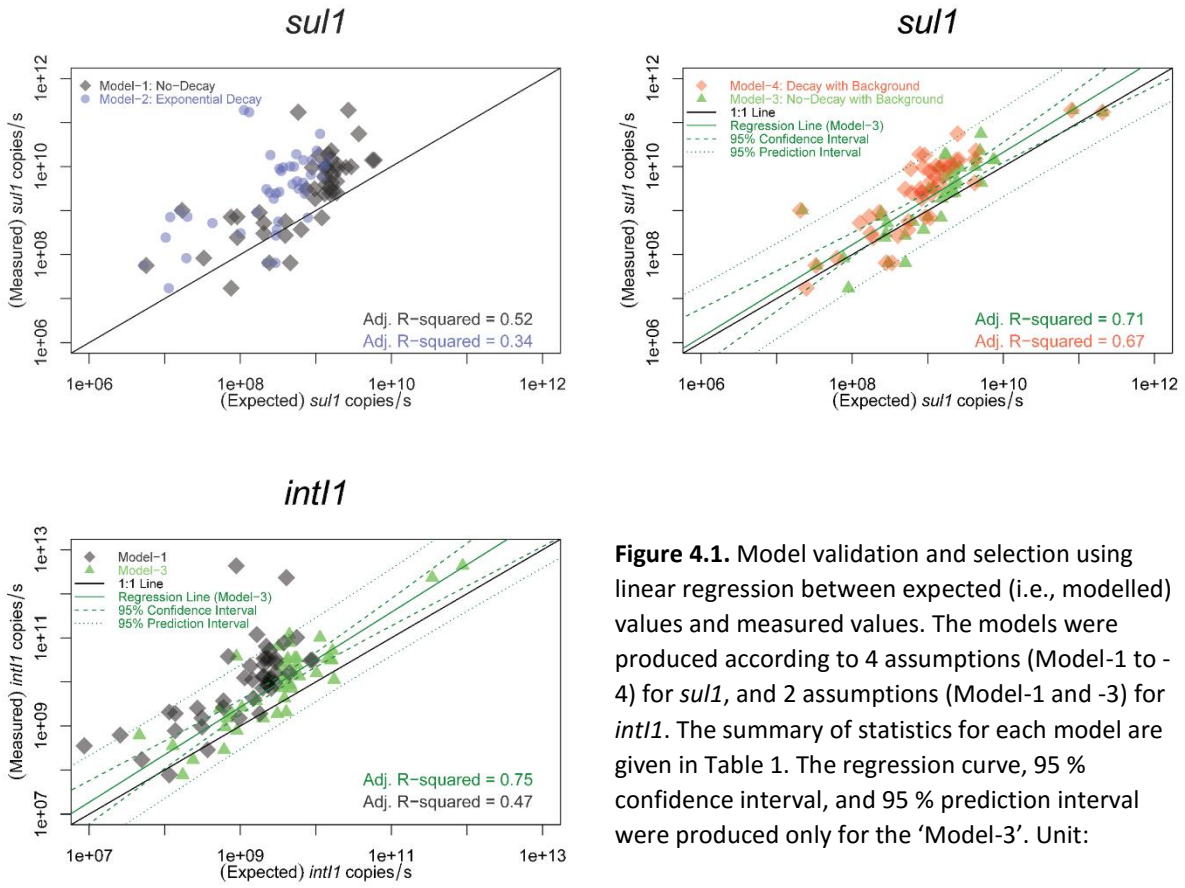


Figure 4.1. Model validation and selection using linear regression between expected (i.e., modelled) values and measured values. The models were produced according to 4 assumptions (Model-1 to -4) for *sul1*, and 2 assumptions (Model-1 and -3) for *intI1*. The summary of statistics for each model are given in Table 1. The regression curve, 95 % confidence interval, and 95 % prediction interval were produced only for the 'Model-3'. Unit:

Table 4.1. Summary table for model validation and selection. Validation was performed using linear regression (Y = aX + b) between expected (i.e., modelled) and measured values. The finally selected models were asterisked (*).

AMR Indicator	Model	Description	Adj. R^2	P-value	Slope (a)	Y-intercept (b)
sul1	Model-1	None	0.52	1.0E-08	1.0	0.5
	Model-2	None + Decay	0.34	1.4E-05	0.8	2.5
	Model-3*	None + Background	0.71	1.8E-13	1.1	-0.2
	Model-4	None + Decay + Background	0.67	1.7E-12	1.0	0.3
intI1	Model-1	None	0.47	8.7E-08	1.0	1.0
	Model-3*	None + Background	0.75	5.0E-15	1.1	-0.3

For Model-1, -2 and -4 a majority of values were located above the 1:1 line (Fig. 4.1). This means the modeled values underestimated actual loadings, thus additional sources that we did not consider in the model might exist. This was not the case for Model-3, including background loading and assuming no decay. It is possible that these assumptions compensate for sources not considered in our model. Indeed, contamination sources other than WWTP effluents are known, such as industrial wastewaters, and non-point sources such as urban and/or agricultural runoff. Many river catchments are utilized agriculturally in Switzerland (BfS and SFSO, 2013) and agricultural sources, e.g. ARGs in

manure or animal feces transported into rivers by surface runoff (Barrios et al., 2020), might influence the ARG loadings of the river to some extent. Determining the impact of additional, especially non-point sources will require further study. First steps could be made by testing statistical relationships among aquatic ARG loadings or levels and non-point source indicators such as percentage of agricultural utilization, and number of livestock by catchment, etc (Amos et al., 2015;Czekalski et al., 2015).

While residual errors (i.e., deviation of observed values from modeled values; parameterized by R^2) were small enough to yield p-values in the statistically significant range, the errors were not trivial considering that our models were validated at log_{10}-scale. We assumed that those errors might originate, especially from uncertainty of two factors – the levels of *sul1* and *intI1* in EF and natural backgrounds. According to the Monte-Carlo randomization simulation, the propagated errors cannot account for the entire deviation of the model from measured loadings, but they are also not negligible, especially for *sul1* (Fig. 4.2a). This could result from relatively large variance of *sul1* levels in EF among sites compared to *intI1* (Fig. S4.2). If factors that can explain the variance could be found, the predictability of our models could be improved. As one of the possibilities, we hypothesized that *sul1* and *intI1* levels vary by the type of treatment process. Therefore, ANOVA was performed by setting wastewater process types (A ~ I) as treatments (Fig. S4.5). Treatments for which the number of samples was < 4 were excluded, i.e., only the processes E ~ I were considered for both *sul1* and *intI1* while performing ANOVA. Differences among treatments were not significant at 5 % level (i.e., P-value = 0.07 for *sul1*, and 0.86 for *intI1*) for both *sul1* and *intI1*. Given that sample sizes were not high enough to represent each process in our analysis (e.g., n = 4 ~ 9 for E, F, and I), we cannot completely rule out the possibility that ARG levels in EF could vary by process types after increasing sample sizes. In the future, more comprehensive comparison is required to unravel potential quantitative relationships between them, if the relationship is valid, this could be applied to our models, and improve predictability.

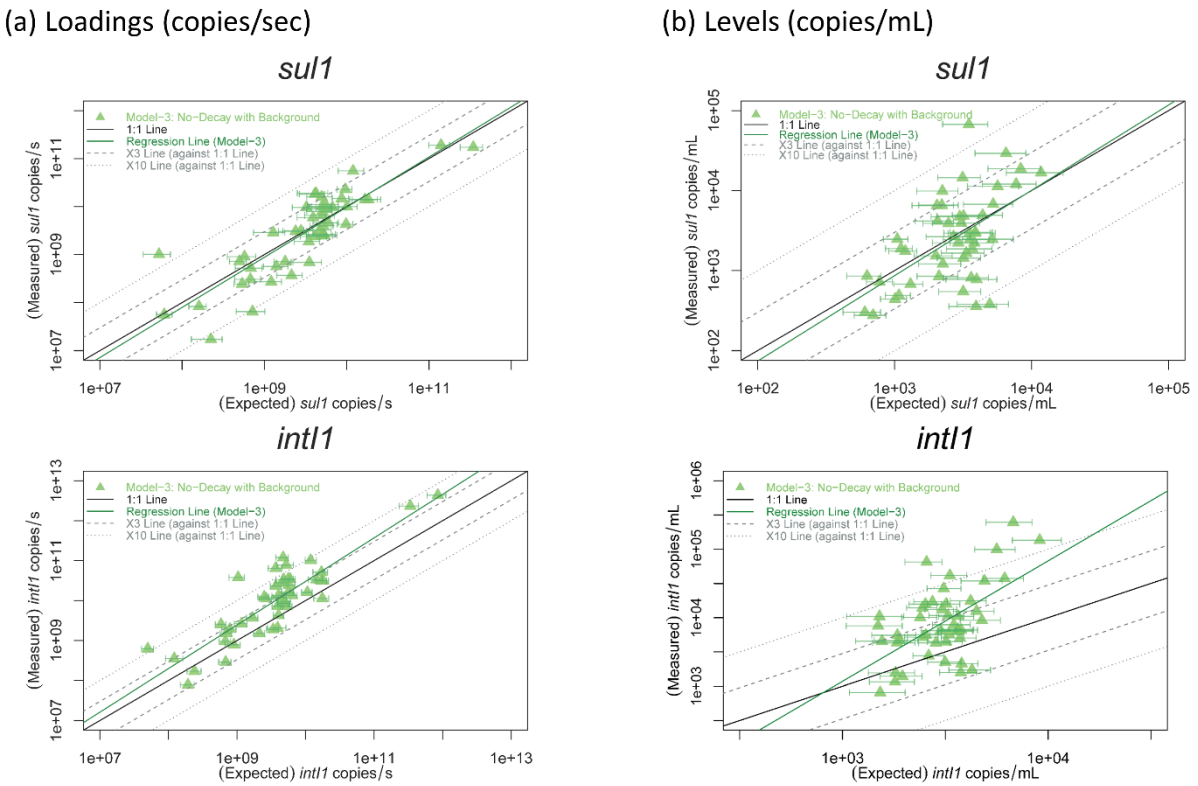


Figure 4.2. Validation of Model-3 for *sul1* and *intI1* as AMR indicators in terms of (a) loadings, and (b) levels. Error bars indicate uncertainty (i.e., prediction error) propagated from variability in measured indicator levels in effluent and unpolluted rivers (background resistance) using Monte-Carlo randomization simulation (n=999). The values with triangular symbol are median values among 999 randomized sets. The first (i.e., 25 %), and third (i.e., 75 %) interquartile ranges of prediction errors were visualized as an error bar.

The model validity was also tested in terms of 'levels' using the finally selected model (Model-3). For instance, 'predicted levels' were calculated as described in the section 3.2.6, and regression analysis between predicted and measured levels was performed as mentioned above. Uncertainty analysis was performed, and visualized as error bars following the same way above (Fig. 4.2b). Even though Model-3 was still statistically valid (R^2 = 0.201 and P = 0.0011 for *sul1*; R^2 = 0.326 and P = 2.042e-05 for *intI1*) in terms of levels, the predictabilities were much lower than the loading-based validation. This means our model do not allow to predict the level with a high resolution. To see if this model is non-parametrically valid, we tested Spearman rank-sum test. In terms of both *sul1* and *intI1*, the model was statistically significant (P = 0.0088 for *sul1*; P = 0.002449 for *intI1*). Therefore, we mainly used our model non-parametrically, for instance, to prioritize 'hotspots' where a high AMR contamination is expected.

4.3.2. Prioritizing 'hotspots' for monitoring AMR pollution in rivers

In the following, we explore some aspects of how our model can be leveraged for guiding future AMR resistance monitoring efforts or for guiding interventions. Both the expected *sul1* and *intI1*

loadings According to Model-3, and levels at wastewater-receiving points were derived, and displayed on the map in Fig. 4.3.

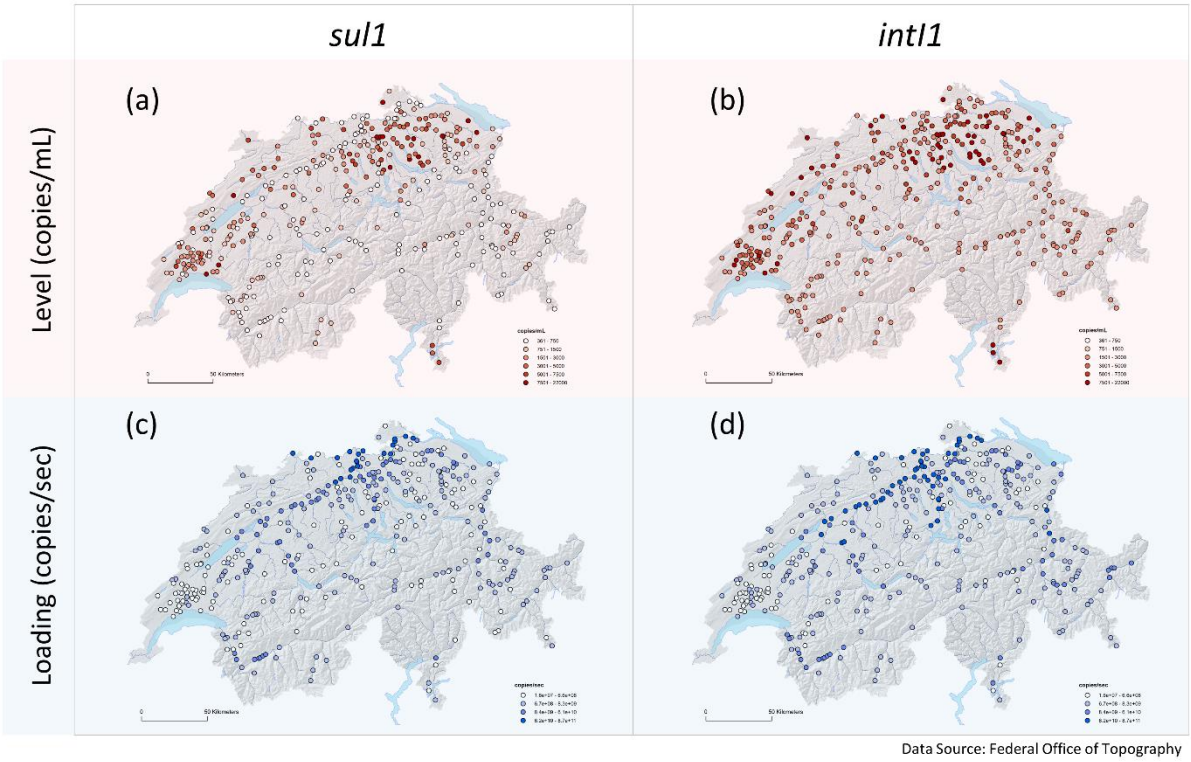


Figure 4.3. Modeled *sul1* (left) and *intI1* (right) levels (a, b; red) and loadings (c, d; blue) at treated wastewater-receiving waters in Swiss river networks according to Model-3. Darker color indicates higher values (see in-figure color scales).

In terms of loadings, and under assumptions of no decay large downstream rivers are expected to have high loadings. For instance, the top 10 sites with the highest *sul1* and *intI1* loadings are located in the rivers Rhein and Aare, that have discharges (i.e., Q95%) ≥ 143 m^3/s at the simulated sites. The contributions of the local WWTP effluent is only up to 5.5 % for *sul1*, and 2.1 % for *intI1*. The cumulative upstream EF-origin loadings comprised up to 44.3 % (for *sul1*), and 22.6 % (for *intI1*) of total loadings in those sites indicating that background loadings took up the largest proportion.

On the other hand, small upstream rivers are predicted in our model to be occasionally high in terms of levels. For example, the top 10 sites with the highest *sul1* and *intI1* levels are predicted in the River Furtbach (ZH, Zurich), Wildbach (ZH), Jonen (ZH), Wissenbach (ZH), Steinach (SG, St.Gallen), Alpbach (SG), Klingengraben (SH, Schaffhausen), Paudèze (VD, Vaud), and Seyon (NE, Neuchâtel). All of these rivers discharge ≤ 0.09 m^3/s. The high levels of *sul1* and *intI1* in those sites are due to the high proportions (i.e., ≥ 37.0 % for *sul1*; ≥ 36.1 % for *intI1*) of the nearest EF-origin loadings to the total expected loadings. This is supported by studies in rivers a with high proportion of wastewater, where high levels of ARG have been reported. According to Lee et al. (2021) performed at the river Murg and Suze where the proportion of EF was up to 38.0 % (Murg) and 35.9 % (Suze), the levels of

sul1 in receiving waters were much higher than in the upstream waters by 2 ~ 37 times. Furthermore, another study performed in a high EF-impacted German river (22 – 55 %), the levels of ARGs at the receiving water increased by approximately 0.5 – 1.5 order of magnitude (Brown et al., 2019).

In general, the levels of ARGs are expected to be high in small-sized, and high wastewater-impacted rivers. In large-sized rivers, discharged ARGs could be diluted by natural waters, and this could make the levels of aquatic ARGs relatively low. In light of a possible future establishment of environmental guidelines or limits for environmental AMR levels, our model may be useful for prioritizing monitoring efforts.

4.3.3. Prediction of public exposure to river aquatic AMR

Another aspect that can be explored with our model is the extent to which the population could be exposed to aquatic AMR which may be an additional consideration when prioritizing potential interventions. To demonstrate this in principle, we considered 'resident population' connected to the WWTP to be potentially exposed to the local river water. From this assumption we calculated a simple index of potential exposure incidence according to eq.2, which sums up all WWTP sites in each canton. This index thus provides a rough indication of the exposure potential in each canton of Switzerland. Results are plotted in Fig. 4.4. According to this analysis, Zurich (ZH) is expected to have the highest risk of exposure among the entire Swiss cantons, followed by St.Gallen (SG), Ticino (TG), Bern (BG), and Aargau (AG) (the order among those four cantons (i.e., SG, BG, TG, and AG) is slightly different by indicator; refer to Fig. 4.4). ZH is the largest canton in Switzerland in terms of population, so has the highest number of resident population connected to WWTPs in Switzerland. Furthermore, impact of EF (i.e., the proportion of EF discharge to the river discharge, 29.4 (± 4.9) % in average) for WWTPs in ZH is higher than in any other canton. These features result in ZH having the highest potential exposure indicator value. Such analyses could guide decisions on interventions or monitoring efforts: given limited resources, ZH could be prioritized in terms of governmental interventions for tackling dissemination of wastewater-borne ARGs. However, decision making will also have to take into account various other aspects, such as socio- and economic-dimensions, so in-depth discussion is required by involving multi-sectoral stakeholders.

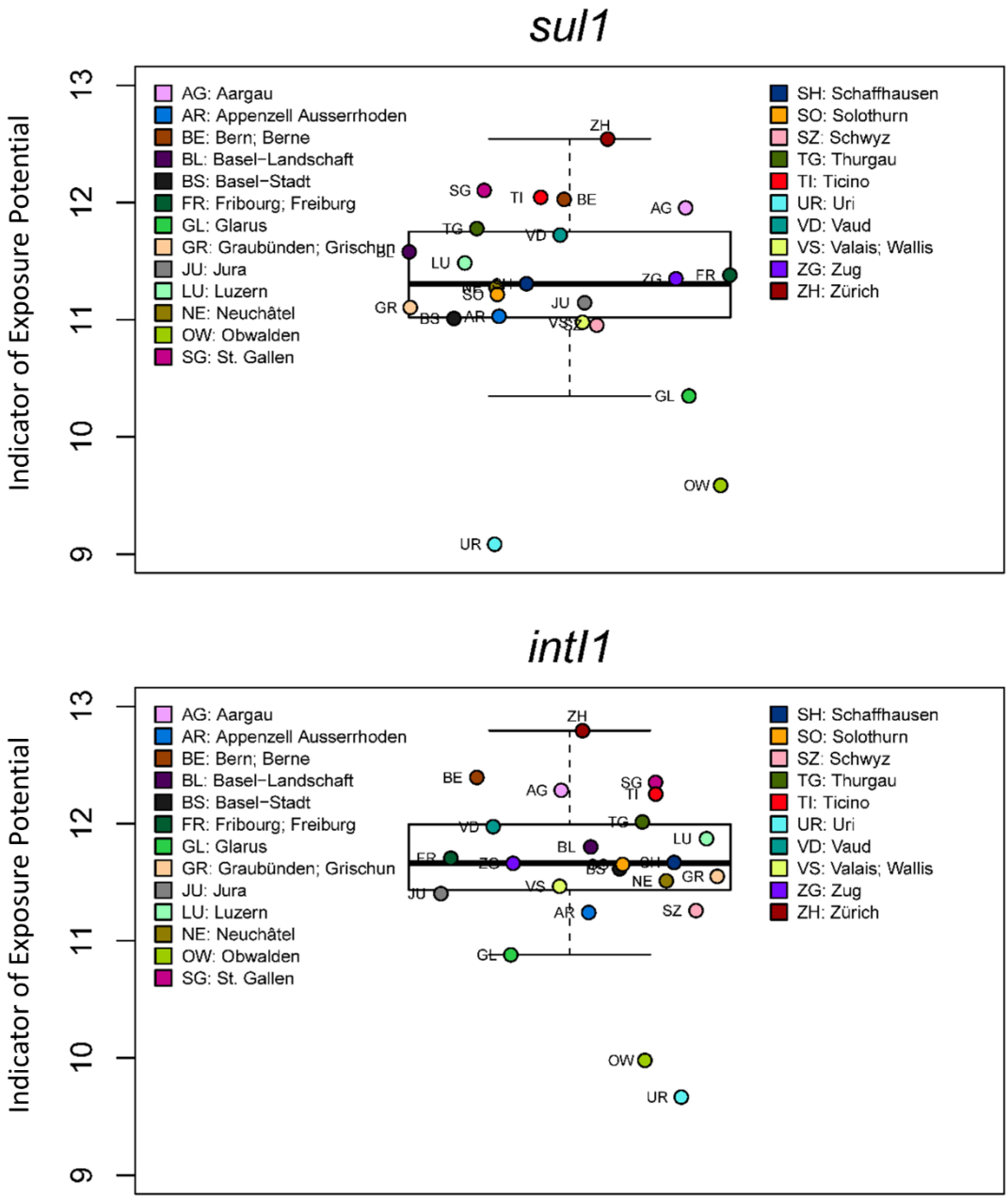


Figure 4.4. Modeled exposure units (copies L-1 · people) by Swiss canton for two antibiotic resistance indicators (i.e., *sul1* and *intI1*). The exposure units were calculated according to eq.2.

One of the ways to tackle dissemination of wastewater-borne ARGs could be to apply tertiary treatment processes in WWTPs. Ozonation could be one of the options. Ozone was proven to be effective against multiresistant bacteria and the *sul1* gene (Czekalski et al., 2016a). The latter was more recalcitrant, and did not decrease significantly under the typical condition for ozonation in the full-scale WWTP. The authors suggested that higher doses of ozone (i.e., > 0.55g O_3 g DOC^{-1}) are required for ARG removal. In Switzerland, the federal government aims to install advanced treatment processes in about 100 existing large WWTPs within the next 20 years since 2016 when a new action for water protection come into effect (Bourgin et al., 2018). Even though the main purpose of this act

is to remove 'micropollutants' from WWTPs, the aspect of wastewater-borne 'ARGs' could also be taken into account. Our models could provide a tool for prioritizing potential hotspots for interventions aiming to reduce AMR discharge into aquatic envrionments. Furthermore, this model could also be used to explore expected outcomes of future interventions based on 'As-is, To-be' analysis. After parameterizing the removal rate of ARGs by ozone referring to other studies, the public exposure could be assessed based on different scenarios of ozonation treatments (e.g., no-ozonation, low-doses, and high-doses), and compared. We intend to perform such analyses in the near future.

Data Availability

All the datasets and R codes will be available at Eawag Research Data Institutional Collection (https://opendata.eawag.ch/) upon acceptance of this manuscript to the journal.

Author Contributions

J.L and H.B designed this study. The model algorithms were developed by J.L referring to relevant references (Ort et al., 2009; Kuroda et al., 2016) under the supervision by H.B. C.O provided important advices while setting up datasets and developing models. Cyrill Dimitri Anderfuhren is acknowledged for a non-authored contribution to the development of early version of our models. Other non-authored contributions are acknowledged in the 'Acknowledgements' section of this thesis.

Supplementary Information

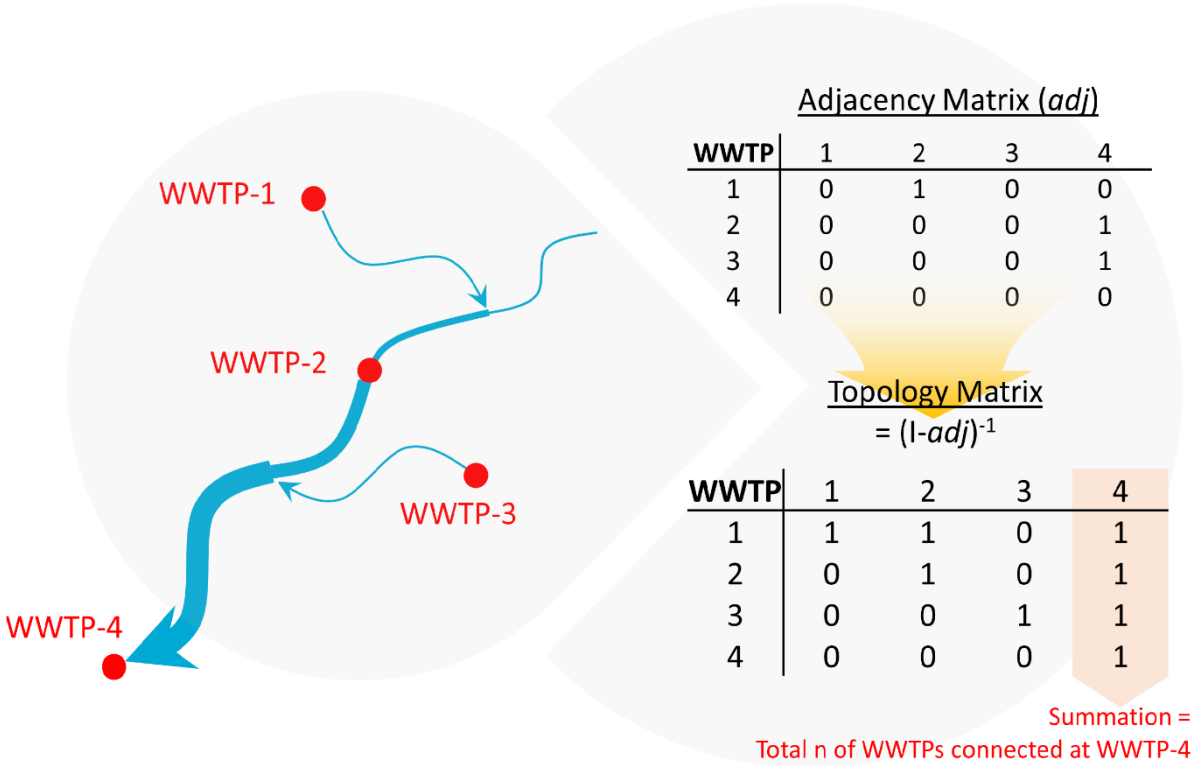


Figure S4.1. An exemplary diagram showing river network analysis. The edges shown here are 'presence (1)' and 'absence (0)' information. Summation of each column will give the total number of connected wastewater treatment plants (WWTPs) upstream at each site (i.e., effluent receiving point). The edges could be substituted, for instance by effluent loadings; then the summation of a column in topology matrix indicates total accumulated wastewater loadings at each site.

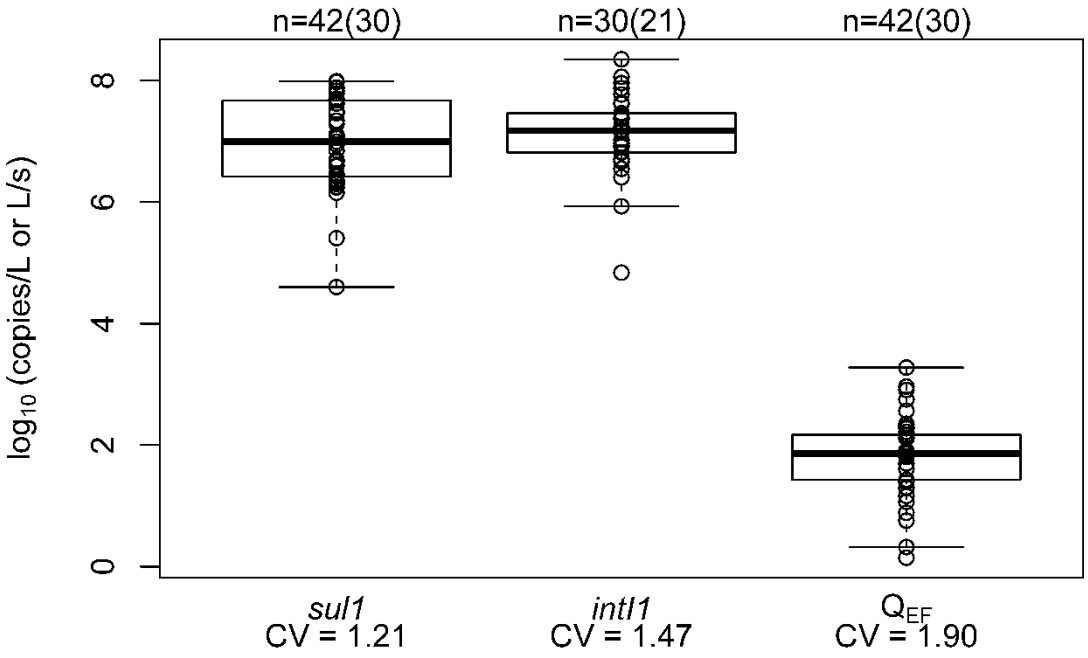


Figure S4.2. Log_{10}-transformed levels (copies/L) of *sul1* and *intI1* in treated effluent (EF), and EF-discharge (Q_{EF}, L/s) from 42 sampling campaigns in 30 different wastewater treatment plants (i.e., 42(30)) referring to our previous studies (Czekalski et al., 2016b;Ju et al., 2019;Ju et al., In Preparation;Lee et al., 2021). For *intI1*, only 30 data points in 21 different WWTPs (i.e., 30(21)) are available. CV indicates the coefficient of variation.

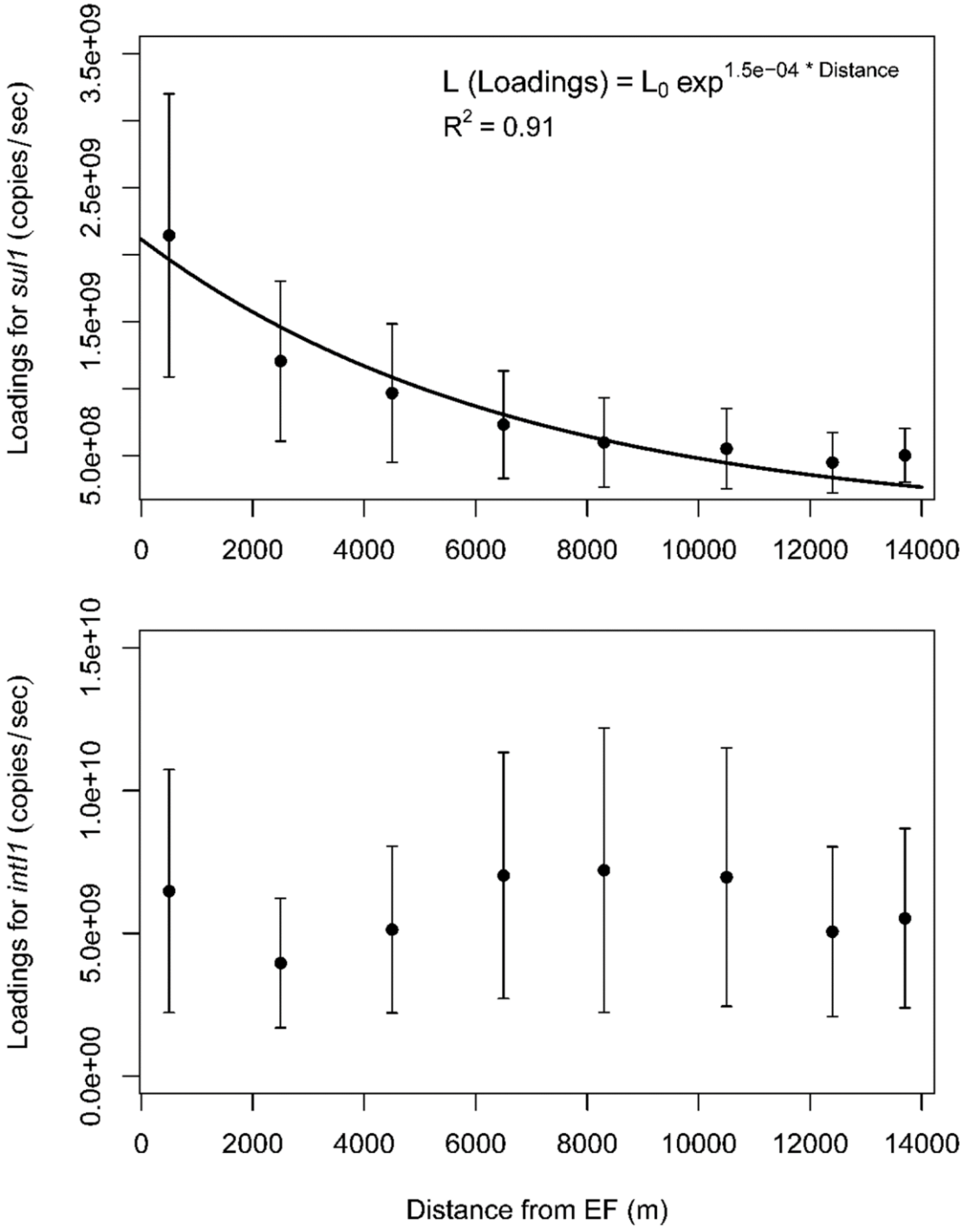


Figure S4.3. Downstream profiles of *sul1* and *intI1* loadings in the River Suze catchment near Villeret, Switzerland where one wastewater treatment plant discharge point exist (at distance = 0 m). The results were from Lee et al. (2021), and the figure was re-created in this study. The error bars indicate standard errors among the values from three sampling campaigns in July, 2019.

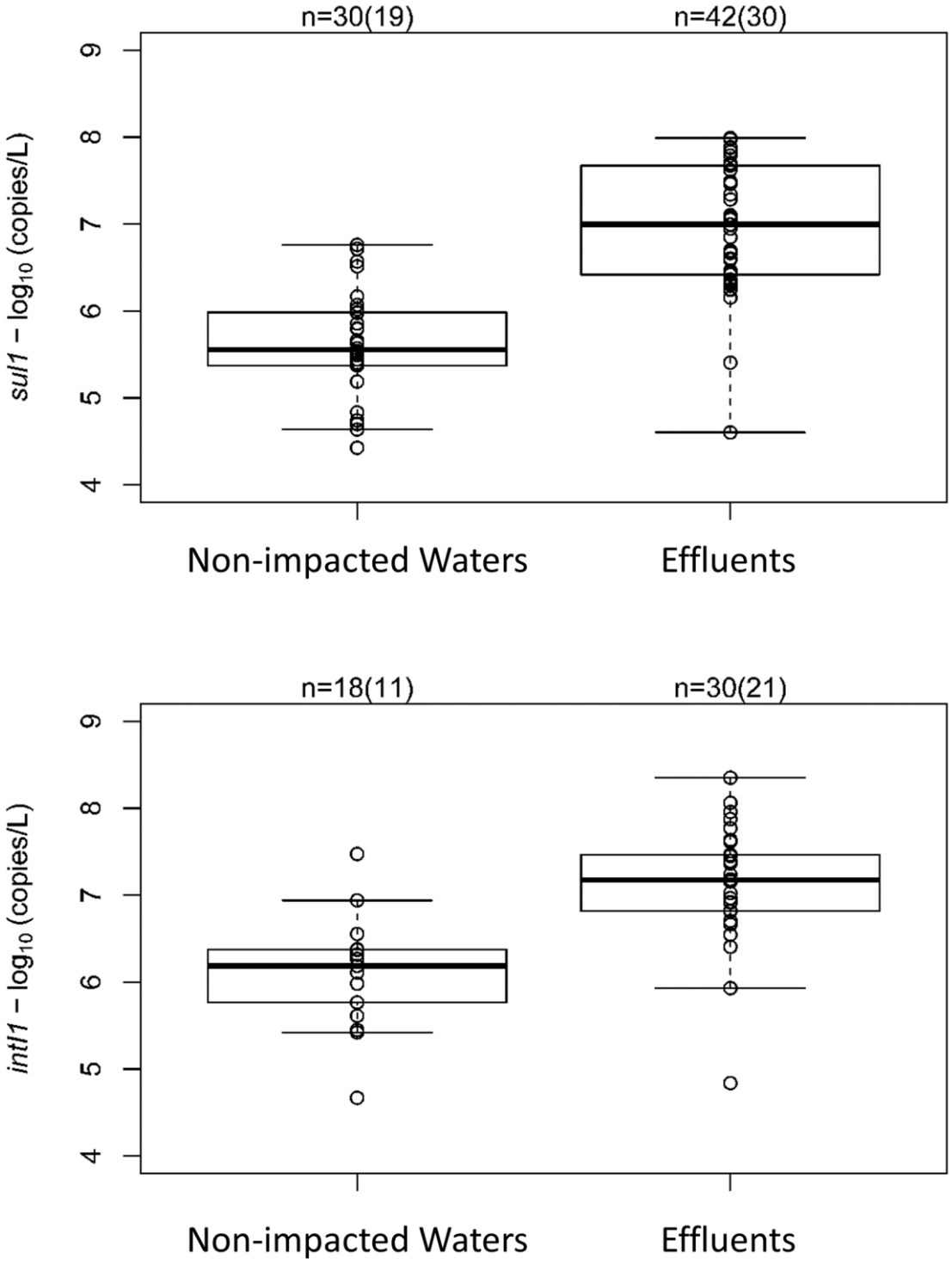


Figure S4.4. The *sul1* and *intI* levels (copies/L) in treated effluents (Effluents), and upstream locations where no known-point sources were located (Non-impacted waters). Total number of samplings were suggested as a number without brackets, and the number of sites were suggested a number within brackets (e.g., n = 30(21) indicates 30 sampling campaigns in 21 sites. The datasets were from our previous studies (Czekalski et al., 2016b;Ju et al., In Preparation;Lee et al., 2021).

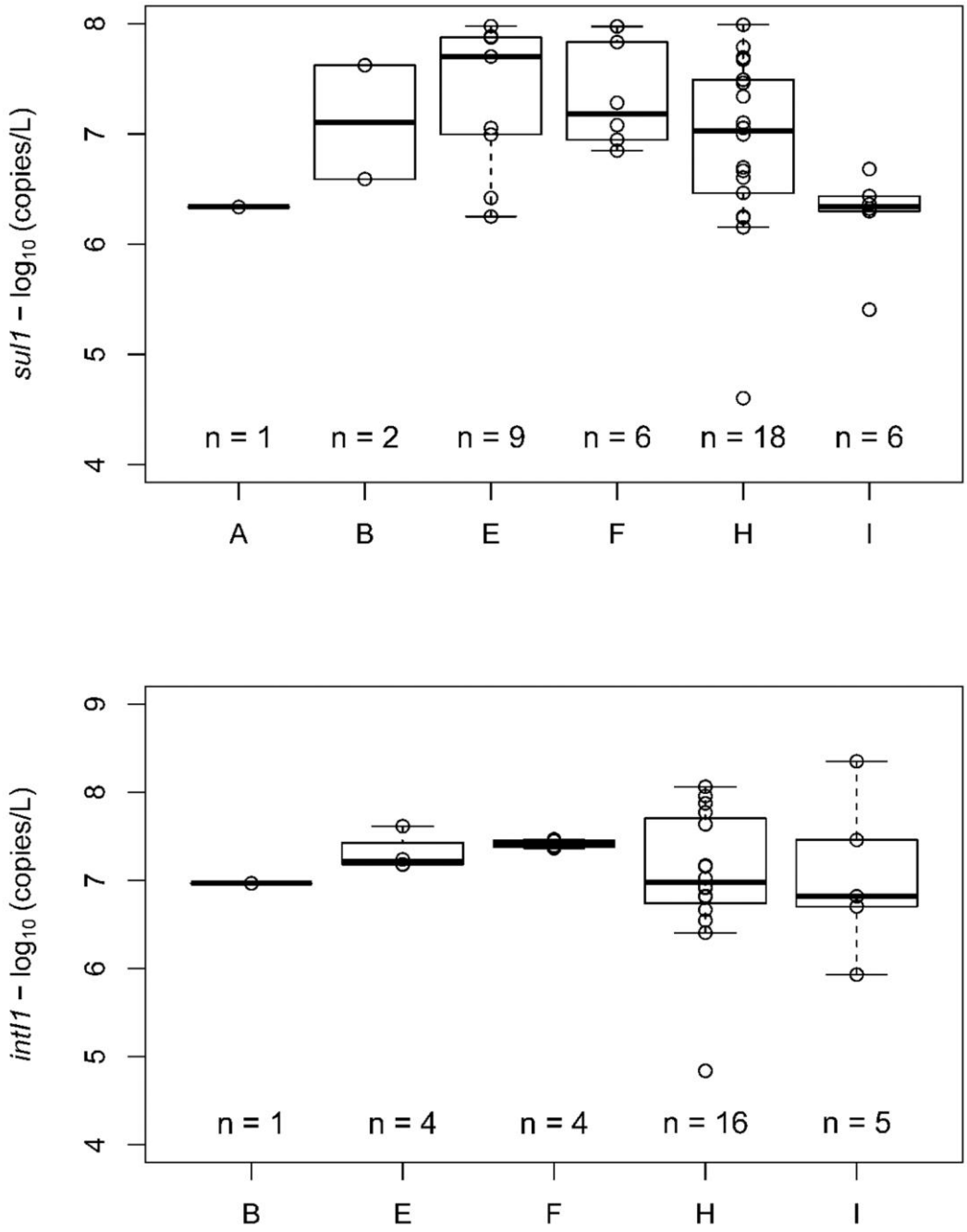


Figure S4.5. log_{10}-transformed levels of *sul1* and *intI1* in treated effluents by wastewater process type. The data shown here are the same as those in Fig. S4.2 and S4.4. Different letters represent different treatment processes, i.e., A = Mechanical-biological treatment (C; removal of carbon compounds); B = C + Phosphorus removal (P); E = C + P + Nitrification (NH_4); F = C + P + NH_4 + Sand filtration (Sand); H = MB + P + NH_4 + Denitrification (N); I = C + P + NH_4 + Sand + N.

Table S4.1. Site information for validation sampling campaigns. Samplings were performed from September 22 to November 01, 2020.

Site_ID	GPS-coordinates (Swiss-grid)		Categories	Nearest WWTP IDs	Distance from WWTP discharge points (m)	Sampling Dates	Cantons	River Catchments	Precipitation in the previous 24 hours (mm/day)	River Discharge Derivation Method (M for Measured; E for Estimated)
	X	Y								
FL_US	733218	252908	US	340200	300	22.09.2020	SG	Glatt	1.3	E
FL_D1	732790	253131	DS	340200	361	22.09.2020	SG	Glatt	1.3	E
FL_D2	729315	256373	DS	340200	9,400	22.09.2020	SG	Glatt	1.3	M
UZ_US	728025	257380	US	340802	577	22.09.2020	SG	Thur	1.3	E
UZ_D1	729087	257439	DS	340802	500	22.09.2020	SG	Thur	1.3	E
NIB_US	732658	259428	US	342200	488	22.09.2020	SG	Thur	1.3	E
NIB_D1	733382	260174	DS	342200	500	22.09.2020	SG	Thur	1.3	E
BAS_US	611063	269905	US	270101	500	22.09.2020	BS	Rhein	0	E
FUE_D1	621301	263083	DS	282500	500	22.09.2020	BL	Ergolz	0	E
SIS_D2	622557	259569	US	282500	3,500	22.09.2020	BL	Ergolz	0	M
BUD_D1	622459	257544	DS	282300	500	22.09.2020	BL	Frenke	0	E
TRI_D1	648161	230493	DS	110400	500	06.10.2020	LU	Suhre	0.3	E
TRI_D2	646975	233435	DS	110400	3,500	06.10.2020	LU	Suhre	0.3	M
ATT_D1	646225	235763	DS	427200	500	06.10.2020	AG	Suhre	0.3	E
ATT_D2	646033	240408	US	414400	550	06.10.2020	AG	Suhre	0.3	E
SCH_D1	646358	241469	DS	414400	800	06.10.2020	AG	Suhre	0.3	E
SCH_D2	646163	246039	DS	414400	5,660	06.10.2020	AG	Suhre	0.3	M
DAG_D1	640063	231638	DS	113400	500	08.10.2020	LU	Wigger	3	E
DAG_D2	637602	236997	DS	113400	6,000	08.10.2020	LU	Wigger	3	M
FUL_D1	629520	235144	DS	257500	15,000	08.10.2020	SO	Aare	3	M
TAV_D1	585583	231918	DS	69600	500	08.10.2020	BE	Birs	4.4	E
COU_D1	593129	232776	DS	69000	300	08.10.2020	BE	Birs	4.4	E
COU_D2	595732	236936	DS	69000	6,500	08.10.2020	BE	Birs	4.4	M
DMN_D1	777341	176314	DS	385104	500	14.10.2020	GR	Landwasser	0	E
DGS_D1	778640	178717	DS	385103	500	14.10.2020	GR	Landwasser	0	E

DGT_D1	779236	180604	DS	385102	500	14.10.2020	GR	Landwasser	0	E
RUE_D1	707049	233824	DS	11801	500	14.10.2020	ZH	Jona	0	E
RUE_D2	707017	234607	US	11801	500	14.10.2020	ZH	Jona	0	E
BUB_D1	705553	235686	DS	11301	500	14.10.2020	ZH	Schwarz	0	E
WAL_D2	709717	236550	DS	12001	1,450	14.10.2020	ZH	Jona	0	M
WAL_D1	710532	236707	DS	12001	500	14.10.2020	ZH	Jona	0	E
VIT_D1	543404	165191	DS	553800	500	19.10.2020	VD	Mentue	0	E
POL_D1	542183	165407	DS	553300	500	19.10.2020	VD	Coruz	0	E
PEY_D1	543388	168034	DS	568200	500	19.10.2020	VD	Mentue	0	E
DON_D1	544526	178042	DS	591300	500	19.10.2020	VD	Mentue	0	E
DON_D2	545388	180957	DS	591300	4,500	19.10.2020	VD	Mentue	0	M
LUC_D1	555172	174316	DS	567500	500	19.10.2020	VD	Broye	0	E
HEN_D1	557436	177724	DS	581900	500	19.10.2020	VD	Broye	0	E
GRA_D1	558903	179761	DS	581800	500	19.10.2020	VD	Broye	0	E
PAY_D1	561818	187554	DS	582200	500	19.10.2020	VD	Broye	0	E
BUS_D1	559147	187501	DS	200400	500	19.10.2020	FR	Petite Glâne	0	E
GRR_D1	562613	190856	DS	581700	500	19.10.2020	VD	Petite Glâne	0	E
GRR_D2	566323	194201	DS	581700	5,800	19.10.2020	VD	Petite Glâne	0	M
DAG2_D1	640075	231619	DS	113400	500	01.11.2020	LU	Wigger	0	E
BUT_D2	641008	226262	DS	108300	12,560	01.11.2020	LU	Buttisholzbach	0	M
BUT_D1	648011	218441	DS	108300	260	01.11.2020	LU	Buttisholzbach	0	E

Table S4.2. Key validation information for quantitative PCR standard curves. LOD indicates the limit of detection.

qPCR primer	Slope	Specificity verification method	qPCR-efficiency	Y-intercept	Linear dynamic range (copies - copies)	LOD-copies (Avr. Cq)	S.D. (of Cp) at LOD	Cp of the NTC
sul1	-3.22	Hydrolysis probe (TaqMan)	2.044	38.7	5.0E+07 - 5.0E+01	50 (34.8)	0.22	N.D.
intI1	-3.48	Hydrolysis probe (TaqMan)	1.938	40.8	5.0E+07 - 5.0E+01	50 (34.8)	0.29	N.D.
16S rRNA gene	-3.54	Hydrolysis probe (TaqMan)	1.917	41.8	5.0E+07 - 5.0E+01	50 (33.2)	0.38	34.8

CHAPTER 5. Conclusions

5.1. Summary and Discussion of Limitations

In this thesis, key factors governing the fate of anthropogenic inputs of AMR in the river were studied. Under dry-weather conditions, while the effluents from WWTPs were the most critical factors, other factors (i.e., dilution by additional water inputs, and other source/decay mechanisms) were shown to play important roles in driving the fate of riverine AMR. The weather itself was found to be a significant factor. Following heavy rains, the contamination by WWTPs became even more serious. The discharge of untreated sewage through bypass and combines sewers overflow increased the levels of AMR indicators at receiving waters, to values much higher than the baseline under dry-weather conditions. Based on key observations above, a nation-wide river network model for AMR contamination was developed. This model confirmed one of the key observations made in the case studies – namely that WWTPs are the major sources of AMR. However, results of model validation also indicate that there probably are 'additional sources of AMR' other than WWTPs. While this model could be further improved, it is still useful at least to 'prioritize' hotspots for AMR contamination. More detailed points were further discussed in the following paragraphs (i - v).

i. Dilution as a main mechanism for short-range decreases of AMR levels

A modelling based on mass-flow concepts and conservative tracers under the dry-flow revealed that sharp decreases of wastewater-borne AMR levels over short distance (2.0 – 2.5 km) appeared to be due to the dilution by additional water inputs, e.g., groundwater infiltration and/or tributary inputs. However, this could be different at other sites depending on the local hydrogeologic conditions. If there are no or low inputs of groundwaters or tributaries, the degree of decrease over downstream distance might be much less.

Considering that there were not profound differences between modeled- and measured-levels over 2.0 – 2.5 km distance, chapter 2 states that additional source/sink mechanisms were not found to be significant over such a short distance. The expected hydraulic retention time over 2.0 – 2.5 km was on the order of hours (99 mins/2 km for MUE; 128 mins/2.5 km for VIL). This might be too short for those mechanisms (e.g, increased mortality of discharged bacterial cells due to environmental conditions, sedimentation or predation) to be effective. Unless other acute sources (e.g., additional point/non-point sources other than the WWTP) exist in downstream locations, wastewater-borne resistance determinants are probably not likely to significantly increase or decrease over such short ranges. However actual rates for such mechanisms have not been determined in this work.

ii. Additional sources/decays become more apparent over longer downstream distance

The mass-flow (loadings) based analysis presented in chapter 2 revealed that additional source/sink mechanisms became more apparent over longer downstream distance (up to 13.7 km for VIL; up to 6.8 km for MUE).

In many cases, the loadings of most ARGs (except for *intI1*) decreased with downstream distance. This means wastewater-borne ARGs do not persist along the downstream waterway. However, the loadings of *intI1* at downstream locations did not significantly decrease or increase, and we assume this might be because *intI1* exists intrinsically in undisturbed freshwaters at relatively high levels, so does not work well as an indicator for anthropogenic AMR contamination in some situations.

In some cases, the loadings of ARGs also did not always decrease, as evidenced by one of our samplings (MUE3) where pronounced increases of *sul1* and *intI1* loadings were observed in far downstream locations. Various lines of evidences suggested that this increase might not originate from anthropogenic sources.

As a speculation, I suggest in Chapter 2 'stream biofilms' as a potential explanation for this. Biofilms are considered as hotspots for horizontal gene transfer (HGT) in aquatic environments (Abe et al., 2020). For instance, conjugation rarely occurs in planktonic phases because a direct interaction between donor and recipient cells is difficult (Abe et al., 2020). The situation becomes different after forming biofilms. Considering that cells are densely packed in this form, they are expected to achieve higher contact frequencies compared to planktonic life forms. Thus conjugation could occur more frequently (Abe et al., 2020). Indeed, many studies reported that biofilms promote HGT (Abe et al., 2020;Madsen et al., 2012;Molin and Tolker-Nielsen, 2003). Furthermore, a study on wastewater biofilms revealed that relative abundances of *sul1* and *intI1* were much higher than other genes (*ermB* and *qnrS*), especially in biofilm-fluid interfaces (Petrovich et al., 2019). The authors proposed a possibility that the cells on biofilm surfaces could be detached, released into the water, thus could selectively enrich *sul1* and *intI1* in the water. I speculate the aquatic *sul1* and *intI1* observed in MUE3 could have been enriched in a similar way – detached from stream biofilm surfaces where HGT could occur frequently. However, we have not made any analysis of biofilms in this system during our investigation, so this explanation remains speculative. Therefore, further studies linking biofilms to aquatic resistome would be required in the River Murg, Münchwilen, and the impact of biofilms on the aquatic resistome could be a field of study of some interest also more generally.

iii. Wastewater bypass is temporarily a major source of antimicrobial resistance

Among various stormwater-related disturbances (i.e., bypass, effluents, river surface sediments, catchment surface soils), bypass was found to be the main contributor to the elevation of aquatic

ARG levels in receiving river waters. The bypass-borne resistance determinants were persistently present for a while (up to 22 hours) in the water. This might be due to the advective transport and longitudinal dispersion of upstream inputs. This observation indicates that the fate of bypass inputs in receiving waters is temporarily dynamic, so a future study considering hydraulic aspects is required. Such studies face a number of challenges. First of all, the hydraulic modelling will require a better knowledge of bypass characteristics (e.g., levels and temporal dynamics of resistance determinants). The four bypass samples shown in this study might not be enough, more repetitive samplings of bypass are needed. Other quantitative information on active flow times, rates, and discharge locations for all the bypass pipelines are also required. To quantify hydraulic effects (advection and dispersion), the properties of river catchment are needed, such as flow rates during the period of monitoring, and site-specific dispersion coefficient, etc. Those properties could be empirically parameterized, then the advection-dispersion equation could be used to model the spatio-, and temporal-dynamics of resistance determinants in the Murg catchment.

Another observation was that ARG levels were relatively stable over time in treated effluents. This means that we could achieve certain degree of ARG removal from combined sewages even under high flow conditions is as long as those are treated by conventional wastewater processes. As briefly suggested in CH.3, increasing the proportion of incoming combined sewages that are treated would help to reduce the amount of bypass-borne ARG discharges during stormwater events. Upgrading currently existing WWTPs is an option, at least this aspect could be considered in the early stage of WWTP installation – having enough treatment capacities so that a large portion of combined sewage could be treated during stormwater events. At the same time, informing residents of the potential risk of exposure to bypass-borne AMR might also be required. If not, the risk of public exposure to aquatic AMR could remain high during and shortly after stormwater events.

iv. The risk of exposure to ‘multi-resistance risk factors’ increases during stormwater events

The risk of exposure not only to resistance, but also to ‘multi-resistance’ increased during and shortly after stormwater events. Using a network analysis visualizing ARG co-location in CH.3, a large portion of MGDs were shown to be potentially associated with plasmids. Considering plasmids could be transferred to other recipient cells, those MGDs could be persistent, or could even proliferate in the environment.

‘Where exactly could those MGDs persist or proliferate in the environment?’ is still an open question. As discussed earlier in the section-ii (p.133), water compartments are less likely to be the hotspots because contact frequencies among cells might not be high enough or planktonic stages to allow for significant conjugation rates (Abe et al., 2020). I assume that biofilms near wastewater

discharge points could be potential spots. Bypass-borne MGDs could be captured by those biofilms, and could further proliferate via HGT stimulated by chronic (i.e., treated wastewaters) or sporadic (i.e., bypass) disturbances. Considering that HGT could occur even without antibiotics-mediated selection pressures (Lopatkin et al., 2017), MGDs could persist even in biofilms from further downstream points or reservoirs (i.e., lakes or marine estuaries), once disseminated.

v. The river network model confirms that 'WWTPs' are the main anthropogenic AMR sources

The river network model for AMR contamination confirmed key observations from the previous chapter (CH.2) and other project (Ju et al., In Preparation) under dry-flow conditions: 1) WWTPs are major source of aquatic *sul1* and *intI1* in the river, 2) Intrinsic *sul1* and *intI1* exist in river waters. On the other hand, another hypothesis learnt from CH.2 (i.e., wastewater-borne *sul1* loadings decrease over downstream distance in the River Suze, Villeret) could not be generalized to the entire rivers – additional sources of AMR might exist. While the current model is still useful to prioritize 'hotspots' for AMR contamination, further improvements could be made by figuring out other significant sources of AMR, and by parameterizing those factors.

Among potential additional sources, I speculate that industrial wastewaters might play a role to a certain extent. While the exact figure and statistics are not available yet, profound amounts of industrial wastewaters are discharged to Swiss rivers (R. Gulde from Verband Schweizer Abwasser, personal communication to Jangwoo Lee, Apr. 30, 2021), thus certain amounts of ARG loadings from those might exist in the water. Furthermore, a diffusion from non-point sources could also potentially contribute to AMR contamination. A previous study performed in 21 Swiss lakes identified potential impacts of agricultural activities in some lakes (i.e., Lakes Baldegg and Greife), even though the main drivers of AMR contamination were still WWTPs in general (Czekalski et al., 2015). Given current analysis, it is not possible to figure out which source(s) except for WWTPs is likely to be significant driver(s) of AMR contamination in Swiss rivers. Therefore a future study is required, e.g., examining the relationship between AMR levels (or loadings) and land usages, etc.

5.2. Synthesis and Outlook

Under dry-flow condition, treated effluents are key contamination sources of AMR in rivers according to our dry-sampling studies in CH.2 and 4. Even though there could still be other sources, for instance other anthropogenic sources and/or in-system growths as evidenced by MU3 results in CH.1, our models assuming WWTPs as major contamination sources in CH.4 were still statistically valid – treated effluents contribute significantly to AMR pollution in rivers. Under stormwater events, bypass is a major source of AMR contamination in a high wastewater-impacted river of which catchment is utilized by agricultural practices (CH.3). My thesis suggests that WWTPs are key

contamination sources of AMR in both weather conditions. Therefore, WWTPs could be important potential intervention points for tackling dissemination of wastewater-borne AMR.

As one of the ways to improve over the current situation, I proposed to increase the proportion of treated wastewater to incoming combined sewage during stormwater events in CH.3. In this way, we could obtain certain degree of removal of bypass-borne resistance determinants from wastewaters. The resistance determinants in treated wastewaters could be further removed, for instance by installing ozonation processes in WWTPs as mentioned in CH.4. As mentioned in CH.4, the Swiss government aims to install ozonation in the selected WWTPs. Our predictive model shown in CH.4 could provide insights on this selection process, for instance using a scenario based analysis, which will be implemented by me and collaborators in the near future.

Another possible intervention that I briefly mentioned earlier in the section-iii (p.134) is to inform residents about the risk of exposure to aquatic AMR, especially during and shortly after stormwater events when the levels of aquatic AMR increase profoundly. There could be two important points regarding informing the public – (1) raising public awareness in terms of AMR risks, and (2) informing 'where' and 'when' exactly the risk increases. The former should precede the latter so that people could be motivated enough to avoid of the risk, when informed. To achieve this, the efforts not only from scientists, but also from various social players (e.g., (public)educators, civic activists, WWTP operators, local/federal governments, etc) are required. For the latter point (2), more scientific studies, especially on hydraulic aspects are required. For instance, as shown in CH.3, bypass-borne resistance determinants are spatially, and temporally dynamic. Hydraulic processes (i.e., advective transport and longitudinal dispersion), and better knowledge about relevant sources (see above the section iii) might be the key factors. In future studies, hydraulic modeling of resistance determinants during and after stormwater events could be implemented to predict 'where' and 'when' the determinants profoundly exist in waterways.

While this thesis provides a theoretical framework with which the question 'how much are we exposed to riverine aquatic AMR?' could be at least partly answered e.g., under dry-flows (CH.4), the current study still does not tell how 'hazardous' those environmental ARGs or ARB are. This is because, to the best of our knowledge, there are still research gaps that need to be further studied, e.g., human health 'dose-response' relationship. For instance, we do not know 'how much environmental ARGs could cause symptoms once entered into human systems', or at least there has not been a solid framework with which the relationship could be quantitatively parameterized.

To link environmental ARGs to human health risks, it would be important to know 'what is proportion of ARG hosts that human pathogens account for?'. Since not all ARG hosts (i.e., ARB) are

pathogenic. Epidemiological studies should follow, for instance in contaminated river catchments where waterborne disease outbreaks occur. In this case, probably other low income countries where those outbreaks are frequently reported, rather than Switzerland may be better places to study. In the case of waterborne disease outbreaks, we should aim to establish datasets which link the patients (who suffer from waterborne diseases with AMR issues) to the levels of AMR pathogens in the river to which they were exposed. Then, we could statistically infer threshold levels at which infections occur. Various waterborne pathogens could be monitoring targets, such as *Vibro cholera*, *E. coli*, *Salmonella* spp., *Legionella* spp., which potentially cause cholera, hemolytic uremic syndrome, salmonellosis, Legionnaires' disease, etc (Leclerc et al., 2002). However, a direct exposure to water environments might not be the only route. There could be other exposure routes, for instance via contacts with contaminated wild or domestic animals (animal to human transmission), contacts with other patients (human to human transmission), eating irrigated vegetables (food to human transmission), etc. Therefore, not only the 'environmental-aspect' of resistance, but also other relevant fields, such as medical-, agricultural-, veterinary-aspects of resistance should be taken into account altogether, for instance, under the 'One-Health' concept in future exposure studies.

Acknowledgements

Most of all, I would like to thank to Dr. Helmut Bürgmann for supporting me as a supervisor. Thank you for giving me an opportunity to joining your group, also being there for me whenever I asked you advices and helps. I could only be able to bring out my full potential as a student-scientist thanks to your trusts, supports, and encouragements.

I also want to thank to Dr. Alex Hall, and Dr. Célia Manaia for supervising and examining my thesis works as PhD committee. I could be able to finalize and submit my thesis thanks to your inputs and encouragements. I owe thanks to Dr. Bernhard Wehrli for kindly offering me a moderation for my doctoral examination.

I appreciate very much for collaboration and supports from the project team and collaborators. I owe special thanks to Karin Beck. Thank you for helping me with field works, lab works, also supporting me personally. I also want to thank to Dr. Feng Ju for giving me many advices throughout this project, also encouragements as a colleague and friend. I am grateful to Dr. Christa McArdell and Andreas Maccagnan for the supports, discussion, and inputs, especially when performing micropollutant analysis. I also owe thanks to Dr. Christian Stamm, Marco Dal Molin, Dr. Fabrizio Fenicia for giving me feedbacks, and discussion, especially when studying hydraulic aspects of my study. I appreciate the helps, feedbacks, and suggestions from Dr. Christoph Ort, and Dr. Rebekka Gulde in the last chapter of my thesis. I also owe big thanks to Ayella Maile-Moskowitz who helped me with field works, lab works, also gave me many feedbacks when writing the first chapter paper. I thank to Dr. Amy Pruden, and Dr. Peter Vikesland for giving me many inputs, helps, and feedbacks. I also appreciate the technical helps from Severin Erb, and the helps in the initial phase of exposure modelling from Cyrill Dimitri Anderfuhren. I could be able to finish many samplings, lab- and data-analysis thanks to your supports in the last chapter of my thesis. I appreciate the helps from Rosi Siber, Patrick Kathriner, Irene Brunner, Thomas Rüttimann, and Birgit Beck for technical supports. I also thank to Dr. Cintia Ramón Casañas for kindly offering me a critical review of my first chapter paper.

I owe thanks to Dr. Giulia Gionchetta, Kathrin Baumann, Delaney Snead, Tomy Doda, and Karin Beck for giving me a review of this thesis. It was really helpful, and I think the quality was improved very much thanks to you all. I appreciate the supports and helps from current or former members of Microbial Ecology Group – Helmut, Karin, Feng, Giulia, Kathrin, Robert, Magdalena, Nina, Melea, and Lian, thank you all! I also appreciate such a vibrant atmosphere and friendly environment from Eawag-kb society. I really enjoyed my life here in Kastanienbaum thanks to you all for many events, including many parties, aperos, football games, and even lunch-talks at kb-terrace. I also want to

thank to Ron, Remika, and Atreya for all the good time we had for many parties, dinners, movie nights! Thanks for your encouragements and supports as friends. I also want to thank to my friends Julia and Delphine for greeting me in France whenever I visited, especially when you helped me translating an email into French! That really helped me to communicate with the wastewater-engineer in STEP-Villeret. I also want to thank to my former colleague and previous mentor Dr. Seung Gu Shin for warmly greeting, and encouraging me whenever I visited him in Korea.

I want to thank to my lovely families, my mom Ran, my dad Jong-gook, and my sister Nari. I could only be able to move forward, stand strong, and stay confident thanks to all your emotional supports. I also thank to my cousins Junghwa and Nampyo, and my uncle Gyusu for greeting me whenever I visit your places in Korea, and praying for me. I am very proud of being a member of my families, I will always be with you all spiritually wherever I am.

My thesis works were financially supported by the Swiss National Science Foundation (SNSF) under the National Research Program NRP72 “Antimicrobial Resistance”, grant 167116. I also thank to the Genetic Diversity Centre (GDC) at ETH Zurich for providing me with their resources, and many advices on bioinformatics.

Best wishes,

Jangwoo

Reference

Abbott, S.L. (2011) Manual of Clinical Microbiology, 10th Edition, pp. 639-657, American Society of Microbiology.

Abe, K., Nomura, N. and Suzuki, S. (2020) Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. FEMS microbiology ecology 96(5), fiaa031.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W. and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nature biotechnology 31(6), 533-538, 10.1038/nbt.2579.

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V. and McArthur, A.G. (2020) CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic acids research 48(D1), D517-D525, 10.1093/nar/gkz935.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. Journal of molecular biology 215(3), 403-410, 10.1016/S0022-2836(05)80360-2.

Amos, G.C., Gozzard, E., Carter, C.E., Mead, A., Bowes, M.J., Hawkey, P.M., Zhang, L., Singer, A.C., Gaze, W.H. and Wellington, E.M. (2015) Validated predictive modelling of the environmental resistome. The ISME journal 9(6), 1467-1476.

Andersson, D.I. and Hughes, D. (2012) Evolution of antibiotic resistance at non-lethal drug concentrations. Drug resistance updates 15(3), 162-172, 10.1016/j.drup.2012.03.005.

Andrews, S. (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

APHA-AWWA-WPCF. (1981) Standard methods for the examination of water and wastewater, APHA American Public Health Association.

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L.S., Vikesland, P. and Zhang, L. (2018) DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome 6(1), 23, 10.1186/s40168-018-0401-z.

AWA (2019, Jan. 11) Amt für Wasser und Abfall (AWA), Kanton Bern. Retrieved from https://www.bve.be.ch/bve/de/index/direktion/organisation/awa.html.

BAFU (2000) Mean runoff and flow regime types for the river network of Switzerland, Bundesamt für Umwelt (BAFU), CH-3003 Bern.

BAFU (2013) Land use statistics 2004/09 based on the standard nomenclature NOAS04, Bundesamt für Umwelt (BAFU), CH-3003 Bern.

Baral, D., Dvorak, B.I., Admiraal, D., Jia, S.G., Zhang, C. and Li, X. (2018) Tracking the Sources of Antibiotic Resistance Genes in an Urban Stream during Wet Weather using Shotgun Metagenomic Analyses. Environmental science & technology 52(16), 9033-9044, 10.1021/acs.est.8b01219.

Barraud, O., Baclet, M.C., Denis, F. and Ploy, M.C. (2010) Quantitative multiplex real-time PCR for detecting class 1, 2 and 3 integrons. Journal of antimicrobial chemotherapy 65(8), 1642-1645, 10.1093/jac/dkq167.

Barrios, R.E., Khuntia, H.K., Bartelt-Hunt, S.L., Gilley, J.E., Schmidt, A.M., Snow, D.D. and Li, X. (2020) Fate and transport of antibiotics and antibiotic resistance genes in runoff and soil as affected by the timing of swine manure slurry application. Science of the total environment 712, 136505.

Bartram, J., Cotruvo, J., Exner, M., Fricker, C. and Glasmacher, A. (2003) Heterotrophic plate counts and drinking-water safety, IWA publishing.

Bello-López, J.M., Cabrero-Martínez, O.A., Ibáñez-Cervantes, G., Hernández-Cortez, C., Pelcastre-Rodríguez, L.I., Gonzalez-Avila, L.U. and Castro-Escarpulli, G. (2019) Horizontal gene transfer and its association with antibiotic resistance in the genus Aeromonas spp. Microorganisms 7(9), 363.

Bengtsson-Palme, J., Kristiansson, E. and Larsson, D.J. (2018) Environmental factors influencing the development and spread of antibiotic resistance. FEMS microbiology reviews 42(1), fux053.

Bengtsson-Palme, J. and Larsson, D.G. (2016) Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. Environment international 86, 140-149, 10.1016/j.envint.2015.10.015.

Berendonk, T.U., Manaia, C.M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., Burgmann, H., Sorum, H., Norstrom, M., Pons, M.N., Kreuzinger, N., Huovinen, P., Stefani, S., Schwartz, T., Kisand, V., Baquero, F. and Martinez, J.L. (2015) Tackling antibiotic resistance: the environmental framework. Nature reviews microbiology 13(5), 310-317, 10.1038/nrmicro3439.

BfS and SFSO (2013) Land use in Switzerland: results of the Swiss land use statistics. Federal Statistical Office: Neuchatel, Switzerland.

Boerlin, P. and Reid-Smith, R.J. (2008) Antimicrobial resistance: its emergence and transmission. Animal health research reviews 9(2), 115-126.

Bourgin, M., Beck, B., Boehler, M., Borowska, E., Fleiner, J., Salhi, E., Teichler, R., von Gunten, U., Siegrist, H. and McArdell, C.S. (2018) Evaluation of a full-scale wastewater treatment plant upgraded with ozonation and biological post-treatments: Abatement of micropollutants, formation of transformation products and oxidation by-products. Water research 129, 486-498.

Brown, P.C., Borowska, E., Schwartz, T. and Horn, H. (2019) Impact of the particulate matter from wastewater discharge on the abundance of antibiotic resistance genes and facultative pathogenic bacteria in downstream river sediments. Science of the total environment 649, 1171-1178.

Buchfink, B., Xie, C. and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. Nature methods 12(1), 59-60, DOI 10.1038/nmeth.3176.

Bundesrat (2015) Strategy on Antibiotic Resistance Switzerland, Bundesrat (The Federal Council of Switzerland).

Bürgmann, H., Frigon, D., Gaze, W.H., Manaia, C.M., Pruden, A., Singer, A.C., Smets, B.F. and Zhang, T. (2018) Water and sanitation: an essential battlefront in the war on antimicrobial resistance. FEMS microbiology ecology 94(9), fiy101, 10.1093/femsec/fiy101.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W. and Shipley, G.L. (2009) The MIQE Guidelines: M inimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments. Clinical chemistry 55(4), 611–622, 10.1373/clinchem.2008.112797.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J. and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods 13(7), 581-583, 10.1038/nmeth.3869.

Cassini, A., Hogberg, L.D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G.S., Colomb-Cotinat, M., Kretzschmar, M.E., Devleesschauwer, B., Cecchini, M., Ouakrim, D.A., Oliveira, T.C., Struelens, M.J., Suetens, C., Monnet, D.L. and Burden of, A.M.R.C.G. (2019) Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. The Lancet infectious diseases 19(1), 56-66, 10.1016/S1473-3099(18)30605-4.

Chaudhary, A., Kauser, I., Ray, A. and Poretsky, R. (2018) Taxon-Driven Functional Shifts Associated with Storm Flow in an Urban Stream Microbial Community. MSphere 3(4), 10.1128/mSphere.00194-18.

Chen, C.Q., Pankow, C.A., Oh, M., Heath, L.S., Zhang, L.Q., Du, P., Xia, K. and Pruden, A. (2019) Effect of antibiotic use and composting on antibiotic resistance gene abundance and resistome risks of soils receiving manure-derived amendments. Environment international 128, 233-243, 10.1016/j.envint.2019.04.043.

Chen, J., Yu, Z., Michel, F.C., Jr., Wittum, T. and Morrison, M. (2007) Development and application of real-time PCR assays for quantification of erm genes conferring resistance to macrolides-lincosamides-streptogramin B in livestock manure and manure management systems. Applied and environmental microbiology 73(14), 4407-4416, 10.1128/AEM.02799-06.

Cockerill, F., Patel, J., Alder, J., Bradford, P., Dudley, M. and Eliopoulos, G. (2013) Performance standards for antimicrobial susceptibility testing: twenty-third informational supplement; M100-S23. Wayne, PA: CLSI.

Coordinators, N.R. (2017) Database resources of the national center for biotechnology information. Nucleic acids research 45(Database issue), D12, 10.1093/nar/gkx1095.

Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2001) Introduction to algorithms second edition. The Knuth-Morris-Pratt Algorithm.

Cox, G. and Wright, G.D. (2013) Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. International Journal of Medical Microbiology 303(6-7), 287-292.

Czekalski, N., Berthold, T., Caucci, S., Egli, A. and Burgmann, H. (2012) Increased levels of multiresistant bacteria and resistance genes after wastewater treatment and their dissemination into Lake Geneva, Switzerland. Frontiers in microbiology 3, 106, 10.3389/fmicb.2012.00106.

Czekalski, N., Diez, E.G. and Burgmann, H. (2014) Wastewater as a point source of antibiotic-resistance genes in the sediment of a freshwater lake. The ISME journal 8(7), 1381-1390, 10.1038/ismej.2014.8.

Czekalski, N., Imminger, S., Salhi, E., Veljkovic, M., Kleffel, K., Drissner, D., Hammes, F., Burgmann, H. and Von Gunten, U. (2016a) Inactivation of antibiotic resistant bacteria and resistance genes by ozone: from laboratory experiments to full-scale wastewater treatment. Environmental science & technology 50(21), 11862-11871.

Czekalski, N., Sigdel, R., Birtel, J., Matthews, B. and Bürgmann, H. (2015) Does human activity impact the natural antibiotic resistance background? Abundance of antibiotic resistance genes in 21 Swiss lakes. Environment international 81, 45-55.

Czekalski, N., Von Gunten, U. and Bürgmann, H. (2016b) Antibiotikaresistenzen im Wasserkreislauf. Aqua & Gas 96, 72-80.

Davis, M.A., Martin, K.A. and Austin, S.J. (1992) Biochemical activities of the parA partition protein of the P1 plasmid. Molecular microbiology 6(9), 1141-1147, 10.1111/j.1365-2958.1992.tb01552.x.

Dharmaraj, S. (2021) The Basics: RT-PCR. Thermo Fisher Scientific, Retrieved from https://www.thermofisher.com/ch/en/home/references/ambion-tech-support/rtpcr-analysis/general-articles/rt--pcr-the-basics.html.

Dolejska, M. and Papagiannitsis, C.C. (2018) Plasmid-mediated resistance is going wild. Plasmid 99, 99-111, 10.1016/j.plasmid.2018.09.010.

Eawag (2014) WWTP of Switzerland: überarbeitet auf der Basis des Projektes: Maurer M. und Herlyn A. (2007) Zustand, Kosten und Investitionsbedarf der schweizerischen Abwasserentsorgung. Eawag/Bafu Bericht.

Eramo, A., Delos Reyes, H. and Fahrenfeld, N.L. (2017) Partitioning of Antibiotic Resistance Genes and Fecal Indicators Varies Intra and Inter-Storm during Combined Sewer Overflows. Frontiers in microbiology 8, 2024, 10.3389/fmicb.2017.02024.

Fischer-Romero, C., Tindall, B.J. and Juttner, F. (1996) Tolumonas auensis gen. nov., sp. nov., a toluene-producing bacterium from anoxic sediments of a freshwater lake. The international journal of systematic and evolutionary microbiology 46(1), 183-188, 10.1099/00207713-46-1-183.

Fischer, H.B., List, J.E., Koh, C.R., Imberger, J. and Brooks, N.H. (1979) Mixing in inland and coastal waters, Academic press.

FOEN (2020, Sept. 02) Swiss Federal Office for the Environment (FOEN), Hydrological data and Forecasts. Retrieved from https://www.hydrodaten.admin.ch/en/.

FOPH and FSVO (2020) Swiss Antibiotic Resistance Report 2020. Usage of Antibiotics and Occurrence of Antibiotic Resistance in Switzerland. Federal Office of Public Health and Federal Food Safety and Veterinary Office. FOPH publication number: 2020-OEG-64.

Garner, E., Benitez, R., von Wagoner, E., Sawyer, R., Schaberg, E., Hession, W.C., Krometis, L.H., Badgley, B.D. and Pruden, A. (2017) Stormwater loadings of antibiotic resistance genes in an urban stream. Water research 123, 144-152, 10.1016/j.watres.2017.06.046.

Gerards, M. (2011) International Policy Overview: Antibiotic Resistance, Retrieved from http://apps.who.int/medicinedocs/en/m/abstract/Js19478en/.

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M. and Stokes, H.W. (2008) The evolution of class 1 integrons and the rise of antibiotic resistance. Journal of bacteriology 190(14), 5095-5100, 10.1128/JB.00152-08.

Gillings, M.R. (2014) Integrons: past, present, and future. Microbiology and Molecular Biology Reviews 78(2), 257-277.

Gillings, M.R., Gaze, W.H., Pruden, A., Smalla, K., Tiedje, J.M. and Zhu, Y.G. (2015) Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. The ISME journal 9(6), 1269-1279, 10.1038/ismej.2014.226.

Gosio, B. (1893) Contributo all'etiologia della pellagra; ricerche chimiche e batteriologiche sulle alterazioni del mais. Giornale della Reale Accademia di Medicina di Torino 61, 484-487.

Hernando-Amado, S., Coque, T.M., Baquero, F. and Martínez, J.L. (2019) Defining and combating antibiotic resistance from One Health and Global Health perspectives. Nature Microbiology 4(9), 1432-1442.

Heuer, H. and Smalla, K. (2007) Manure and sulfadiazine synergistically increased bacterial antibiotic resistance in soil over at least two months. Environmental microbiology 9(3), 657-666, 10.1111/j.1462-2920.2006.01185.x.

Hunter, P.R. (2003) Drinking water and diarrhoeal disease due to Escherichia coli. Journal of water and health 1(2), 65-72, 10.2166/wh.2003.0008.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics 11(1), 119, 10.1186/1471-2105-11-119.

Jamieson, R., Joy, D.M., Lee, H., Kostaschuk, R. and Gordon, R. (2005) Transport and deposition of sediment-associated Escherichia coli in natural streams. Water research 39(12), 2665-2675, 10.1016/j.watres.2005.04.040.

Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C.S., Singer, H.P., Johnson, D.R., Zhang, T. and Burgmann, H. (2019) Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. The ISME journal 13(2), 346-360, 10.1038/s41396-018-0277-8.

Ju, F. and Zhang, T. (2015) Experimental design and bioinformatics analysis for the application of metagenomics in environmental sciences and biotechnology. Environmental science & technology 49(21), 12628-12640.

Katz, Y., Wang, E.T., Airoldi, E.M. and Burge, C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods 7(12), 1009-1015, 10.1038/nmeth.1528.

Kitchin, J.E.S., Pomeranz, M.K., Pak, G., Washenik, K. and Shupack, J.L. (1997) Rediscovering mycophenolic acid: a review of its mechanism, side effects, and potential uses. Journal of the American Academy of Dermatology 37(3), 445-449.

Kuroda, K., Itten, R., Kovalova, L., Ort, C., Weissbrodt, D.G. and McArdell, C.S. (2016) Hospital-use pharmaceuticals in Swiss waters modeled at high spatial resolution. Environmental science & technology 50(9), 4742-4751.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9(4), 357-359, 10.1038/nmeth.1923.

Launay, M.A., Dittmer, U. and Steinmetz, H. (2016) Organic micropollutants discharged by combined sewer overflows–characterisation of pollutant sources and stormwater-related processes. Water research 104, 82-92.

Leclerc, H., Schwartzbrod, L. and Dei-Cas, E. (2002) Microbial agents associated with waterborne diseases. Critical reviews in microbiology 28(4), 371-409.

Lee, J., Jeon, J.H., Shin, J., Jang, H.M., Kim, S., Song, M.S. and Kim, Y.M. (2017) Quantitative and qualitative changes in antibiotic resistance genes after passing through treatment processes in municipal wastewater treatment plants. Science of the total environment 605, 906-914.

Lee, J., Ju, F., Maile-Moskowitz, A., Beck, K., Maccagnan, A., McArdell, C.S., Dal Molin, M., Fenicia, F., Vikesland, P.J., Pruden, A., Stamm, C. and Burgmann, H. (2021) Unraveling the riverine antibiotic

resistome: The downstream fate of anthropogenic inputs. Water research 197, 117050, 10.1016/j.watres.2021.117050.
Lee, K., Kim, D.W., Lee, D.H., Kim, Y.S., Bu, J.H., Cha, J.H., Thawng, C.N., Hwang, E.M., Seong, H.J., Sul, W.J., Wellington, E.M.H., Quince, C. and Cha, C.J. (2020) Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. Microbiome 8(1), 2, 10.1186/s40168-019-0774-7.
Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12(1), 323, 10.1186/1471-2105-12-323.
Li, D., Liu, C.M., Luo, R., Sadakane, K. and Lam, T.W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10), 1674-1676, 10.1093/bioinformatics/btv033.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16), 2078-2079, 10.1093/bioinformatics/btp352.
Li, J., Cheng, W., Xu, L., Jiao, Y., Baig, S.A. and Chen, H. (2016) Occurrence and removal of antibiotics and the corresponding resistance genes in wastewater treatment plants: effluents' influence to downstream water environment. Environmental science and pollution research 23(7), 6826-6835, 10.1007/s11356-015-5916-2.
Lopatkin, A.J., Meredith, H.R., Srimani, J.K., Pfeiffer, C., Durrett, R. and You, L. (2017) Persistence and reversal of plasmid-mediated antibiotic resistance. Nature communications 8(1), 1689, 10.1038/s41467-017-01532-1.
Ma, L.P., Li, A.D., Yin, X.L. and Zhang, T. (2017) The Prevalence of Integrons as the Carrier of Antibiotic Resistance Genes in Natural and Man-Made Environments. Environmental science & technology 51(10), 5721-5728, 10.1021/acs.est.6b05887.
Madsen, J.S., Burmølle, M., Hansen, L.H. and Sørensen, S.J. (2012) The interconnection between biofilm formation and horizontal gene transfer. FEMS Immunology & Medical Microbiology 65(2), 183-195.
Mansfeldt, C., Deiner, K., Machler, E., Fenner, K., Eggen, R.I.L., Stamm, C., Schonenberger, U., Walser, J.C. and Altermatt, F. (2020) Microbial community shifts in streams receiving treated wastewater effluent. Science of the total environment 709, 135727, 10.1016/j.scitotenv.2019.135727.
Mao, D., Yu, S., Rysz, M., Luo, Y., Yang, F., Li, F., Hou, J., Mu, Q. and Alvarez, P. (2015) Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. Water research 85, 458-466.
Marano, R.B., Fernandes, T., Manaia, C.M., Nunes, O., Morrison, D., Berendonk, T.U., Kreuzinger, N., Telson, T., Corno, G. and Fatta-Kassinos, D. (2020) A global multinational survey of cefotaxime-resistant coliforms in urban wastewater treatment plants. Environment international 144, 106035, 10.1016/j.envint.2020.106035.
Marti, E., Jofre, J. and Balcazar, J.L. (2013) Prevalence of antibiotic resistance genes and bacterial community composition in a river influenced by a wastewater treatment plant. PloS one 8(10), e78906, 10.1371/journal.pone.0078906.
McEwen, S.A. and Collignon, P.J. (2018) Antimicrobial Resistance: a One Health Perspective. Microbiology spectrum 6(2), 521-547, 10.1128/microbiolspec.ARBA-0009-2017.
Melnyk, A.H., Wong, A. and Kassen, R. (2015) The fitness costs of antibiotic resistance mutations. Evolutionary applications 8(3), 273-283.
Menzel, P., Ng, K.L. and Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature communications 7(1), 11257, 10.1038/ncomms11257.
MeteoSwiss (2019, December 20) IDAweb: The data portal of MeteoSwiss for research and teaching.
Molin, S. and Tolker-Nielsen, T. (2003) Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. Current opinion in biotechnology 14(3), 255-261.

Moura, A., Soares, M., Pereira, C., Leitao, N., Henriques, I. and Correia, A. (2009) INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics 25(8), 1096-1098, 10.1093/bioinformatics/btp105.

Myrold, D.D., Zeglin, L.H. and Jansson, J.K. (2014) The potential of metagenomic approaches for understanding soil microbial processes. Soil Science Society of America Journal 78(1), 3-10.

Nevers, M.B. and Boehm, A.B. (2010) Modeling fate and transport of fecal bacteria in surface water. The fecal bacteria, 165-188.

Nicolaou, K.C. and Rigol, S. (2018) A brief history of antibiotics and select advances in their synthesis. The Journal of antibiotics 71(2), 153-184.

Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P.A. (2017) metaSPAdes: a new versatile metagenomic assembler. Genome research 27(5), 824-834.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., McVeigh, D.H.R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y.M., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W.J., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H.Z., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D. and Pruitt, K.D. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research 44(D1), D733-D745, 10.1093/nar/gkv1189.

O'Neill, J. (2014) 'Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. 2014. The Wellcome Trust and UK Government.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R., Simpson, G.L., Solymos, P., Stevens, M.H.H. and Wagner, H. (2013) Package 'vegan'. Community ecology package, version 2(9), 1-295.

Ort, C., Hollender, J., Schaerer, M. and Siegrist, H. (2009) Model-based evaluation of reduction strategies for micropollutants from wastewater treatment plants in complex river networks. Environmental science & technology 43(9), 3214-3220.

Ort, C. and Siegrist, H. (2009) Assessing wastewater dilution in small rivers with high resolution conductivity probes. Water Science and Technology 59(8), 1593-1601, 10.2166/wst.2009.174.

Parsaie, A. and Haghiabi, A.H. (2017) Computational modeling of pollution transmission in rivers. Applied water science 7(3), 1213-1222, 10.1007/s13201-015-0319-6.

Peix, A., Ramirez-Bahena, M.H. and Velazquez, E. (2009) Historical evolution and current status of the taxonomy of genus Pseudomonas. Infection, genetics and evolution 9(6), 1132-1147, 10.1016/j.meegid.2009.08.001.

Peng, Y., Leung, H.C., Yiu, S.-M. and Chin, F.Y. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28(11), 1420-1428.

Petrovich, M.L., Rosenthal, A.F., Griffin, J.S. and Wells, G.F. (2019) Spatially resolved abundances of antibiotic resistance genes and intI1 in wastewater treatment biofilms. Biotechnology and bioengineering 116(3), 543-554.

Proia, L., Anzil, A., Subirats, J., Borrego, C., Farre, M., Llorca, M., Balcazar, J.L. and Servais, P. (2018) Antibiotic resistance along an urban river impacted by treated wastewaters. Sci Total Environ 628-629, 453-466, 10.1016/j.scitotenv.2018.02.083.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glockner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41(Database issue), D590-596, 10.1093/nar/gks1219.

Rafraf, I.D., Lekunberri, I., Sànchez-Melsió, A., Aouni, M., Borrego, C.M. and Balcázar, J.L. (2016) Abundance of antibiotic resistance genes in five municipal wastewater treatment plants in the Monastir Governorate, Tunisia. Environmental pollution 219, 353-358.

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., Michael, I. and Fatta-Kassinos, D. (2013) Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes

spread into the environment: A review. Science of the total environment 447, 345-360, 10.1016/j.scitotenv.2013.01.032.

Rodriguez-Mozaz, S., Chamorro, S., Marti, E., Huerta, B., Gros, M., Sanchez-Melsio, A., Borrego, C.M., Barcelo, D. and Balcazar, J.L. (2015) Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. Water research 69, 234-242, 10.1016/j.watres.2014.11.021.

Sabri, N., Schmitt, H., Van der Zaan, B., Gerritsen, H., Zuidema, T., Rijnaarts, H. and Langenhoff, A. (2018) Prevalence of antibiotics and antibiotic resistance genes in a wastewater effluent-receiving river in the Netherlands. Journal of environmental chemical engineering, 102245, 10.1016/j.jece.2018.03.004.

Sayers, E.W., Agarwala, R., Bolton, E.E., Brister, J.R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T., Holmes, J.B., Kim, S., Kimchi, A., Kitts, P.A., Lathrop, S., Lu, Z., Madden, T.L., Marchler-Bauer, A., Phan, L., Schneider, V.A., Schoch, C.L., Pruitt, K.D. and Ostell, J. (2019) Database resources of the National Center for Biotechnology Information. Nucleic acids research 47(D1), D23-D28, 10.1093/nar/gky1069.

Schmieder, R. and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6), 863-864, 10.1093/bioinformatics/btr026.

Seiler, C. and Berendonk, T.U. (2012) Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. Frontiers in microbiology 3, 399, 10.3389/fmicb.2012.00399.

Sharma, V.K., Johnson, N., Cizmas, L., McDonald, T.J. and Kim, H. (2016) A review of the influence of treatment strategies on antibiotic resistant bacteria and antibiotic resistance genes. Chemosphere 150, 702-714, 10.1016/j.chemosphere.2015.12.084.

SNSF (2019) Portrait of the National Research Programme (NRP 72): Antimicrobial Resistance. The Swiss National Science Foundation (SNSF), http://www.nfp72.ch/SiteCollectionDocuments/NFP72_programm_2019_en.pdf.

Soucy, S.M., Huang, J. and Gogarten, J.P. (2015) Horizontal gene transfer: building the web of life. Nature Reviews Genetics 16(8), 472-482.

Sriram, A., Kalanxhi, E., Kapoor, G., Craig, J., Balasubramanian, R., Brar, S., Criscuolo, N., Hamilton, A., Klein, E. and Tseng, K. (2021) State of the World's Antibiotics 2021: A Global Analysis of Antimicrobial Resistance and Its Drivers. Center for Disease Dynamics, Economics & Policy: Washington, DC, USA.

Stamm, C., Rasanen, K., Burdon, F.J., Altermatt, F., Jokela, J., Joss, A., Ackermann, M. and Eggen, R.I.L. (2016) Unravelling the Impacts of Micropollutants in Aquatic Ecosystems: Interdisciplinary Studies at the Interface of Large-Scale Ecology. Advances in ecological research 55, pp. 183-223, Academic press, 10.1016/bs.aecr.2016.07.002.

Stoob, K., Singer, H.P., Mueller, S.R., Schwarzenbach, R.P. and Stamm, C.H. (2007) Dissipation and transport of veterinary sulfonamide antibiotics after manure application to grassland in a small catchment. Environmental science & technology 41(21), 7349-7355, 10.1021/es070840e.

Storteboom, H., Arabi, M., Davis, J.G., Crimi, B. and Pruden, A. (2010) Tracking Antibiotic Resistance Genes in the South Platte River Basin Using Molecular Signatures of Urban, Agricultural, And Pristine Sources. Environmental science & technology 44(19), 7397-7404, 10.1021/es101657s.

Takai, K. and Horikoshi, K. (2000) Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Applied and environmental microbiology 66(11), 5066-+, Doi 10.1128/Aem.66.11.5066-5072.2000.

Tennstedt, T., Szczepanowski, R., Braun, S., Puhler, A. and Schluter, A. (2003) Occurrence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. FEMS microbiology ecology 45(3), 239-252, 10.1016/S0168-6496(03)00164-8.

Toronto (2021) Wastewater Treatment Plant Bypasses, The City of Toronto, Retrieved from https://www.toronto.ca/services-payments/water-environment/managing-sewage-in-toronto/wastewater-treatment-plant-bypasses/.

Tsihrintzis, V.A. and Hamid, R. (1997) Modeling and Management of Urban Stormwater Runoff Quality: A Review. Water resources management 11(2), 137-164.

Velísková, Y., Sokáč, M., Halaj, P., Bara, M.K., Dulovičová, R. and Schügerl, R. (2014) Pollutant spreading in a small stream: a case study in Mala Nitra Canal in Slovakia. Environmental processes 1(3), 265-276, 10.1007/s40710-014-0021-y.

Vikesland, P.J., Pruden, A., Alvarez, P.J., Aga, D., Bürgmann, H., Li, X.-d., Manaia, C.M., Nambi, I., Wigginton, K. and Zhang, T. (2017) Toward a comprehensive strategy to mitigate dissemination of environmental sources of antibiotic resistance, ACS Publications.

Walsh, F., Ingenfeld, A., Zampicolli, M., Hilber-Bodmer, M., Frey, J.E. and Duffy, B. (2011) Real-time PCR methods for quantitative monitoring of streptomycin and tetracycline resistance genes in agricultural ecosystems. Journal of microbiological methods 86(2), 150-155, 10.1016/j.mimet.2011.04.011.

Wein, T., Hulter, N.F., Mizrahi, I. and Dagan, T. (2019) Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. Nature communications 10(1), 2595, 10.1038/s41467-019-10600-7.

Weyrauch, P., Matzinger, A., Pawlowsky-Reusing, E., Plume, S., von Seggern, D., Heinzmann, B., Schroeder, K. and Rouault, P. (2010) Contribution of combined sewer overflows to trace contaminant loads in urban streams. Water research 44(15), 4451-4462, 10.1016/j.watres.2010.06.011.

WHO (2001) WHO global strategy for containment of antimicrobial resistance, World Health Organization.

WHO (2012) The evolving threat of antimicrobial resistance: options for action, World Health Organization.

WHO (2020) World Health Organization (WHO) newsroom: Antibiotic resistance. Retrieved from https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance.

Willems, A. (2014) The Prokaryotes, pp. 777-851, Springer.

Wittmer, I.K., Bader, H.P., Scheidegger, R., Singer, H., Luck, A., Hanke, I., Carlsson, C. and Stamm, C. (2010) Significance of urban and agricultural land use for biocide and pesticide dynamics in surface waters. Water research 44(9), 2850-2862, 10.1016/j.watres.2010.01.030.

Wood, D.E., Lu, J. and Langmead, B. (2019) Improved metagenomic analysis with Kraken 2. Genome biology 20(1), 257, 10.1186/s13059-019-1891-0.

Xu, Y., Guo, C., Luo, Y., Lv, J., Zhang, Y., Lin, H., Wang, L. and Xu, J. (2016) Occurrence and distribution of antibiotics, antibiotic resistance genes in the urban rivers in Beijing, China. Environmental pollution 213, 833-840.

Yin, X., Jiang, X.T., Chai, B., Li, L., Yang, Y., Cole, J.R., Tiedje, J.M. and Zhang, T. (2018) ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. Bioinformatics 34(13), 2263-2270, 10.1093/bioinformatics/bty053.

Zhang, X.X., Zhang, T. and Fang, H.H. (2009) Antibiotic resistance genes in water environment. Applied microbiology and biotechnology 82(3), 397-414, 10.1007/s00253-008-1829-z.

Zheng, Z., Li, L., Makhalanyane, T.P., Xu, C., Li, K., Xue, K., Xu, C., Qian, R., Zhang, B. and Du, J. (2021) The composition of antibiotic resistance genes is not affected by grazing but is determined by microorganisms in grassland soils. Science of the total environment 761, 143205, 10.1016/j.scitotenv.2020.143205.

Zurfluh, K., Hächler, H., Nüesch-Inderbinen, M. and Stephan, R. (2013) Characteristics of extended-spectrum β-lactamase-and carbapenemase-producing Enterobacteriaceae isolates from rivers and lakes in Switzerland. Applied and environmental microbiology 79(9), 3021-3026, 10.1128/AEM.00054-13.