Monitoring and modelling antibiotic resistance in Southeast Asian rivers

Amelie Ott

Submitted for a Doctor of Philosophy

School of Engineering

Newcastle University

June 2021

Abstract

Pinpointing environmental antibiotic resistance (AR) hotspots in rivers in low-and-middle income countries (LMICs) is hindered by a lack of available and comparable AR monitoring data relevant to such settings. Addressing this problem, a comprehensive spatial and seasonal assessment of water quality and AR conditions in a Malaysian river catchment was preformed to identify potential 'simple' surrogates that mirror elevated AR. This included screening for β-lactam resistant coliforms, 22 antibiotics, 287 AR genes and integrons, and routine water quality parameters, covering absolute concentrations and mass loadings. Novel approaches were developed and applied to advance environmental microbiome and resistome research. To investigate relationships, standardised 'effect sizes' (Cohen's D) were introduced for AR monitoring to improve comparability of field studies. Quantitative microbiome profiling (QMP) was applied to overcome biases caused by relative taxa abundance data. In addition, Hill numbers were introduced as a unified diversity framework for environmental microbiome research. Overall, water quality generally declined, and environmental AR levels increased as one moved downstream the catchment without major seasonal variations, except total antibiotic concentrations that were higher in the dry season (Cohen's D > 0.8, $P < 0.05$). Among simple surrogates, dissolved oxygen (DO) most strongly correlated (inversely) with total AR gene concentrations (Spearman's ρ 0.81, $P < 0.05$). This is suspected to result from minimally treated sewage inputs, which also contain AR bacteria and genes, depleting DO in the most impacted reaches. Thus, although DO is not a measure of AR, relatively lower DO levels reflect wastewater inputs, flagging possible AR hot spots. Furthermore, DO is easy-to-measure and inexpensive, already monitored in many catchments, and exists in many numerical water quality models (e.g., oxygen sag curves). Therefore, combining DO data and prospective modelling (e.g., with the watershed model HSPF) could guide local interventions, especially in LMIC rivers with limited data.

Title photo by Amelie Ott.

For Roland, Isabel and Gerald.

Acknowledgements

An enormous thank you goes to my supervisors. To David Graham for your excellent guidance, curiosity and enthusiasm over the years. To Greg O'Donnell for your outstanding modelling help and for introducing me into the world (and terminologies) of hydrology. To Michaela Goodson for supporting the field work and providing valuable medical perspectives throughout this project.

To research staff and students at Newcastle University Malaysia for providing technical and social support during my field work in Malaysia, including Jia Yee Ho, Sylvia Sue Xian Liew, Mardhiah Zainal Abidin, Muhammad Noor Hazwan Jusoh and my flatmate Esther. To colleagues at Universiti Teknologi Malaysia for providing valuable knowledge, data and sharing the Skudai model, including Mohd Ridza Mohd Haniffah and Al-Almin Danladi Bello. To researchers at National University of Singapore for performing the antibiotic analysis, including Ngoc Han Tran and Karina Yew-Hoong Gin. To colleagues at Chinese Academy of Science for performing the HT-qPCR assay, including Xinyuan Zhou, Jian-Qiang Su and Yong-Guan Zhu. To the Malaysian Department of Environment for sharing the historic water quality data. To Marcos Quintela-Baluja for assisting in the 16S rRNA qPCR analysis and providing guidance in planning the field monitoring. To Andrew Zealand for providing help with bioinformatics. To Mui-Choo Jong for your support during field work.

To friends and colleagues at Newcastle University for your help, much needed distractions during lunch breaks and great memories. To my research group for the social activites and putting things into perspective. To my friends and family for your support, love and for only sometimes asking when I will get a 'proper job'. To my friend and flatmate Holly for sharing my excitement when I discovered my final R colour palette and for engaging in my endless science monologues. To my mum for paving the way and for teaching me to stay curious & learn new things. To Oli for a lot of proof-reading, only sometimes wanting to be listed as a co-author and for your support throughout – am I now a proper doctor?

Declaration

I hereby certify that the work presented in this thesis is my original research work with the exception of below analysis:

- Ultra-high performance liquid chromatography-tandem mass spectrometry to measure antibiotics in Chapter 3, performed by Ngoc Han Tran at National University of Singapore.
- High-throughput qPCR to measure antibiotic resistant genes in Chapter 3, performed by Xinyuan Zhou at Chinese Academy of Science.
- Baseline HSPF model calibration and validation in Chapter 4, performed by Al-Amin Danladi Bello at Universiti Teknologi Malaysia.
- HSPF model simulation in Chapter 4 for Figure 4-7 and Figure 4-8, performed by Greg O'Donnell at Newcastle University.
- Illumina DNA sequencing in Chapter 5, performed by Andrew Nelson at Northumbria University.

Due reference is given to literature and any other research collaborations where appropriate. The research presented in this thesis has not been submitted for any other degree or professional qualification.

Statement of publications

Chapter 3 is based on a publication with some edits and further details added.

Ott, A., O'Donnell, G., Tran, N.H., Haniffah, M.R.B.M., Su, J.-Q., Zealand, A.M., Gin, K.Y.-H., Goodson, M., Zhu, Y.-G., Graham, D.W., 2021. Developing surrogate markers for predicting antibiotic resistance "hot spots" in rivers where limited data are available. Environ. Sci. Technol. 55, 7466–7478. https://doi.org/10.1021/acs.est.1c00939

Chapter 5 is based on a publication with some edits and further details added.

Ott, A., Quintela-Baluja, M., Zealand, A.M., O'Donnell, G., Haniffah, M.R.B.M., Graham, D.W., 2021. Improved quantitative microbiome profiling for environmental antibiotic resistance surveillance [submitted].

Table of contents

List of equations

List of figures

List of tables

Abbreviations

16S GCN	16S rRNA gene copy number
AR	Antibiotic resistance
ARBs	Antibiotic resistant bacteria
ARGs	Antibiotic resistant genes
ASV	Amplicon sequence variant
BASINS	Better assessment science integrating point and nonpoint sources
BOD	Biochemical oxygen demand
CARD	Comprehensive antibiotic resistance gene database
CFUs	Colony forming units
COD	Chemical oxygen demand
CPB	Carbapenem resistant bacteria
DDDs	Defined daily doses
DEM	Digital elevation model
DO	Dissolved oxygen
DoE	Department of Environment
EPA	Environmental Protection Agency
ESBL	Extended-spectrum β-lactamase
FCA	Fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs
GTAs	Gene transfer agents
GDP	Gross domestic product
HGT	Horizontal gene transfer
HICs	High-income countries
HSPF	Hydrological simulation program - FORTRAN
HT-qPCR	High-throughput qPCR
ILISs	Isotopically labelled internal standards
LC MS/MS	Liquid chromatography-tandem mass spectrometry
LMICs	Low and middle-income countries
MGEs	Mobile genetic elements
MIC	Minimum inhibitory concentration
MLSB	Macrolide-lincosamide-streptogramin B ARGs
MQL	Method quantification limits
MyAP-AMR	Malaysian action plan on antimicrobial resistance
NGS	Next-generation sequencing
NH_3-N	Ammonia
NSAR	National surveillance of antibiotic resistance
NSE	Nash–Sutcliffe model efficiency coefficient
NWQS	National water quality standards for Malaysia
OTU	Operational taxonomic unit
PNEC	Predicted no effect concentrations
^qD	Hill number
QMP	Quantitative microbiome profiling
qPCR	Quantitative PCR

RBO	Rank-biased overlap
RMP	Relative microbiome profiling
SARG	Structured antibiotic resistance gene
SDGs	Sustainable development goals
SE	Southeast
SPE	Solid phase extraction
TN	Total nitrogen
TP	Total phosphorus
UHPLC-MS/MS	Ultra-high performance liquid chromatography-tandem mass spectrometry
VGT	Vertical gene transfer
WQI	Water quality index
WWTPs	Wastewater treatment plants
WEPA	Water environment partnership in Asia
UNEP	United Nations environment programme
HRUs	Hydrologic response units
NO_3	Nitrate
PO_4	Phosphate
QMRA	Quantitative microbial risk assessment

Chapter 1. Introduction

Alexander Fleming's discovery of penicillin in 1929 revolutionized medicine, making common community-acquired, life-threatening infections such as pneumonia and gonorrhoea readily treatable[1]. However, in his Nobel Prize speech 16 years later, Flemming (1945) warned that bacteria could become resistant to these remarkable drugs. Indeed, after the clinical use of penicillin began in the 1940s, the first penicillin-resistant pathogens were soon reported[3] and by the late 1960s, penicillin treatment was ineffective for more than 80% of community and hospital-acquired *Staphylococcus aureus* infections[4].
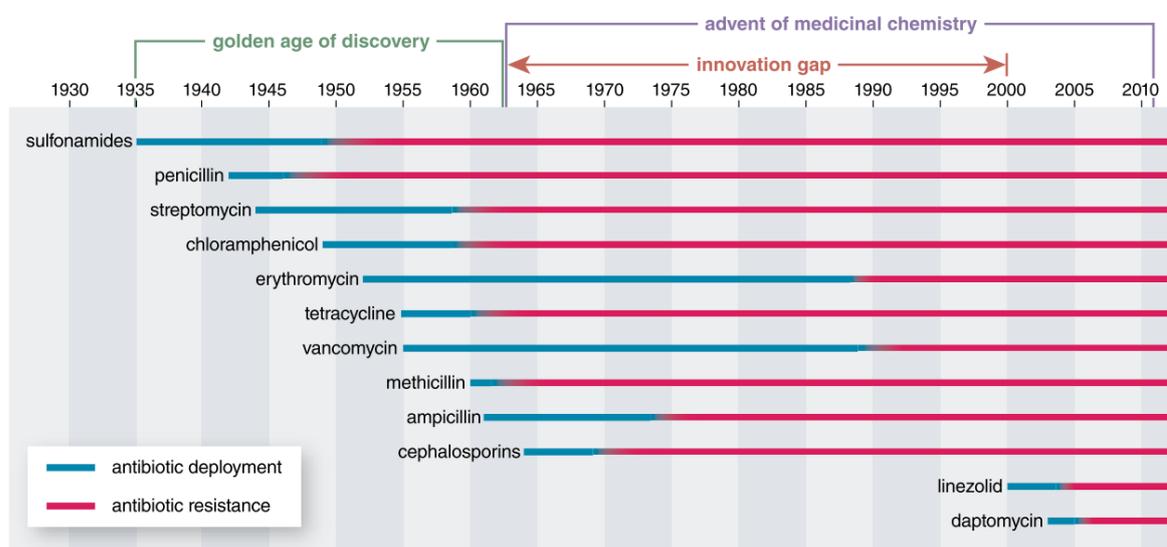
Figure 1-1. Bacterial resistance to new antibiotics is soon detected after their deployment[1].

In the golden age of antibiotic discovery (1930s to 1960s), the rapid spread of resistance was temporarily balanced by the ongoing discovery and clinical introduction of new classes of antibiotics (Figure 1-1)[1]. During this time, researchers discovered more than 150 types of antibiotics[1]. Many of these were easily isolated from soil bacteria (actinomycetes), but after this initial discovery, the pharmaceutical industry struggled to find new antibacterial drugs[5].

During the 'innovation gap', drug approvals declined with companies focussing on modifying existing classes of antibiotics[6]. The majority of currently prescribed antibiotics are derived from a few antibiotic classes that had been discovered by the mid-1980s[5]. Some of the latest registered representatives of novel antibacterial classes (linezolid and daptomycin), were introduced in early 2000 (Figure 1-1), but their related chemical classes (oxazolidinones and lipopeptides) were first reported or

patented in the 1970-80s[6–8]. This is concerning as resistance to one antibiotic often results in resistance to multiple antibiotics within the same class[5].

The development pipeline for new antibiotics is now virtually exhausted and research on alternative treatments to replace antibacterial drugs is still at an early stage[9]. The risk of a post-antibiotic era has to be considered, where common infections and minor injuries could be lethal again[9]. Recent predictions suggest that resistance to drugs including antibiotic resistance (AR) will result in more annual deaths by 2050 than cancer[10].

In their global action plan on antimicrobial resistance, the WHO embraces a One Health approach[11]. One Health describes a transdisciplinary effort to address health problems involving humans, animals and the environment[12]. Health is seen as the product of all exposures, including the environment[13]. While human AR exposure has been well documented[3,14], comprehensive environmental AR exposure assessments are lacking to explicitly support a One Health approach to AR. This is often due to lack of data. For example, Kelly et al. (2016) showed in a systematic analysis for 21 countries that only 3% of AR research projects from 2007 to 2013 focussed on the role of the environment. However, 80% of sewage worldwide is not treated sufficiently[16] and the environment plays an important role in receiving and spreading residues of antibiotics, antibiotic resistant genes (ARGs) and antibiotic resistant bacteria (ARBs)[17–19].

AR is particularly a problem in low and middle-income countries (LMICs). Insufficient waste treatment combined with weak healthcare systems, a high prevalence of over-the-counter sales of antibiotics and insufficiently hygienic living conditions contribute to a high burden of disease from resistant pathogens[20,21]. Southeast (SE) Asia with its rapid economic development has particularly been proposed as an epicentre for emerging infectious diseases and AR[22–25]. For example, Malaysia (study site for this thesis) was one of the first regions were plasmid-mediated resistance to the last-resort antibiotic colistin (*mcr-1*) was detected[26]. Last-resort antibiotics are drugs which are reserved to treat bacteria resistant to all other antibiotics. This means that infectious diseases are going to be more difficult to manage with increasing AR.

Despite the majority of AR burden falling on LMICs, the dynamics of AR have more often focused on high-income countries (HICs)[27]. Monitoring AR in environmental samples can be difficult and usually requires expensive equipment[28]. Many LMICs

monitor river water quality through national programmes, but these only include standard physical and chemical parameters (such as chemical oxygen demand, ammonia, or dissolved oxygen), missing AR release and human exposure data. Where environmental AR data exists, it is difficult to translate findings from one study to other regions and even within the same region across seasons[28].

To address AR in surface waters in LMICs (Figure 1-2), more cost-effective monitoring is required to develop predictive models that can ultimately, guide policy such as deciding where to improve existing or build new wastewater treatment plants. While water quality models have existed for decades, they typically do not include AR processes, lack 'real' data for parametrisation in LMICs such as representing watersheds with inadequate wastewater treatment and do not adequately consider key aquatic transport processes[29].

Figure 1-2. Orang Asli children swimming in a river in Johor, Malaysia[30].

1.1 Thesis aim and tasks

This study aimed to develop alternate approaches to identify environmental AR hotspots in rivers in LMICs. This aim was met by fulfilling the following tasks:

1. Perform a comprehensive spatial and seasonal assessment of water quality and AR conditions in a Southeast Asian river catchment.
2. Characterise water quality systematically by assessing both, pollutant concentration and pollutant mass loading data.
3. Introduce standardised 'effect sizes' to better understand relationships for AR monitoring and improve comparability of field studies.
4. Identify potential 'simple' water quality surrogates that mirror elevated AR.

5. Utilise national water quality datasets and existing surface water quality models to estimate AR fate with 'simple' AR surrogates.

6. Introduce quantitative microbiome profiling and unified Hill number diversities to enhance environmental microbiome and resistome research by providing more quantitative and representative data analyses.

1.2   Thesis structure

The thesis consists of six chapters:

Chapter 1 provides the motivation for this study and introduces the thesis aim and objectives.

Chapter 2 reviews the literature on monitoring and modelling environmental antibiotic resistance, focussing on rivers in Malaysia and the Southeast Asia region.

Chapter 3 assesses spatial and seasonal variations in water quality and AR parameters in a Malaysian river catchment to determine which simple markers best mirror locations of elevated AR. This contributes to the tasks 1, 2, 3 and 4.

Chapter 4 evaluates the use of LMIC monitoring data and surface water quality models to pinpoint AR hot spots in rivers where limited data are available. This contributes to the tasks 4 and 5.

Chapter 5 introduces the novel quantitative microbiome profiling approach and unified Hill number diversity framework to advance environmental microbiome and resistome research. This contributes to the tasks 1, 3 and 6.

Chapter 6 summarises the thesis findings and recommends future work.

Chapter 2. Literature review

2.1 Development and transfer of antibiotic resistance

Antibiotic resistance is a natural, widespread phenomenon in the environment and has existed since the dawn of our species, with ARGs detected in permafrost samples[31,32] and isolated from cave microbiomes[33,34]. Naturally occurring AR is called 'intrinsic AR'. Over the last decades, however, the presence and abundance of resistant genes has increased drastically, closely matching the era of large-scale antibiotic consumption[1]. Knapp et al. (2010) examined archived soil samples from 1940 to 2008 and found significant increases for all tested resistant gene classes since 1940, with tetracycline ARGs being more than 15 times more abundant now than in the 1970s.

Bacteria can acquire resistance either by mutation (vertical gene transfer, VGT, from parent to offspring) or by exchanging genetic material (horizontal gene transfer, HGT, from cell to cell)[27]. This resistance due to VGT or HGT is called 'acquired AR'. Typically, mutation rates are relatively low with on average one in a thousand genome replications incorporating a mutation[1]. Out of these mutations, one in a billion will generate mutants that are more resistant to antibiotics than the predecessors[1].

For clinical pathogens, there is evidence that many, if not all, ARGs originated from previously mutated environmental bacteria and did not evolve de novo[36]. For example, the CTX-M resistance gene, an extended-spectrum β-lactamase (ESBL) often detected in clinical pathogens, has high similarities with chromosomally encoded β-lactamases from the genus *Kluyvera*, an environmental species with little or no pathogenic activity against humans[37–40]. It is inconclusive how many genes in bacteria and archaea genomes have been affected by HGT. Depending on the method and genome analysed, estimates vary greatly between 0.05-80%[41–44].

HGT is facilitated by mobile genetic elements (MGEs). MGEs are a type of genetic material that can move around the genome, or that can be exchanged between species. MGEs play a major role in spreading resistance between a) live cells i.e. conjugation[45], b) via assimilation of extracellular DNA i.e. transformation[46], or c) via bacteriophage infection i.e. transduction[27,47,48] (Figure 2-1a-c).
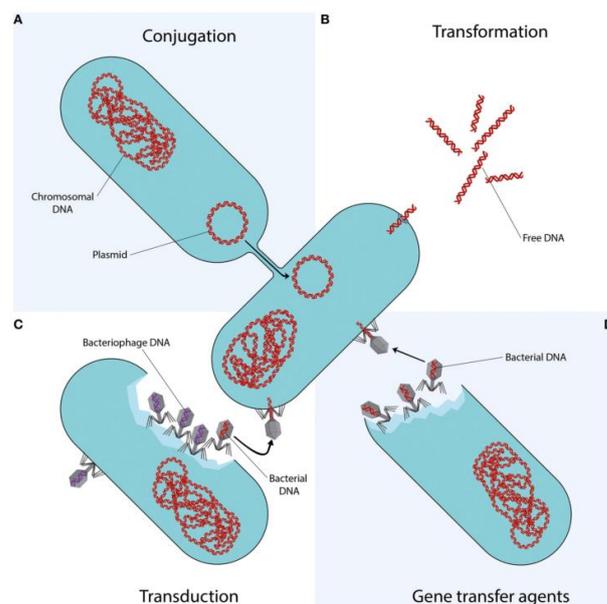
Figure 2-1. Mechanisms of horizontal gene transfer[34].

These HGT mechanisms are well known but novel HGT processes are continuously identified such as gene transfer agents (GTAs)[44] (Figure 2-1d). GTAs are host-cell produced particles that resemble viral structures[34] with ARG transfers demonstrated for pure cultures[49].

Antibiotics and other stressors such as metals have been observed to increase the concentrations of ARBs, ARGs and MGEs in controlled exposure experiments [50,51] and hospital settings[52] but results from environmental field studies are considerably less conclusive[28,53,54]. In surface water, antibiotic concentrations are typically very low and selective effects are often weak, not observed, or results are conflicting[36,55]. In addition, surface waters are often polluted by both, antibiotics and faecal matter, making it difficult to distinguish between the effect of the chemical stressor and the spread of ARGs and MGEs from faeces[56].

2.2   Antibiotic modes of action and antibiotic resistance

There are ten major classes of antibiotics currently in use (Table 2-1). While many antibiotics are of synthetic origin, the majority of antibiotic classes are still derived naturally from bacteria or fungi products[27]. Typically, human medicine prescribes last-generation antibiotics such as carbapenems, cephalosporins and fluoroquinolones while 'older' antibiotics such as penicillins, sulphonamides and tetracyclines tend to be used in animal production[57].

Table 2-1. Ten major classes of antibiotics currently in use[27].

Synthetic	Example	Natural	Example
1) Sulfonamides	Sulfamethoxazole	4) β-lactam	
2) Fluoroquinolones	Ciprofloxacin	– penicillins	Penicillin G
3) Oxazolidinones	Linezolid	– cephalosporins	Cephalosporin C
		– carbapenems	Meropenem
		5) Aminoglycosides	Kanamycin A
		6) Macrolides	Erythromycin A
		7) Tetracyclines	Tetracycline
		8) Glycopeptides	Vancomycin
		9) Lipopeptides	Colistin
		10) Rifamyacins	Rifamyacin SV

Antibiotics generally target bacteria via a few well-defined pathways such as disrupting cell membranes or inhibiting protein synthesis (Figure 2-2), with specifics varying depending on the drug[27].
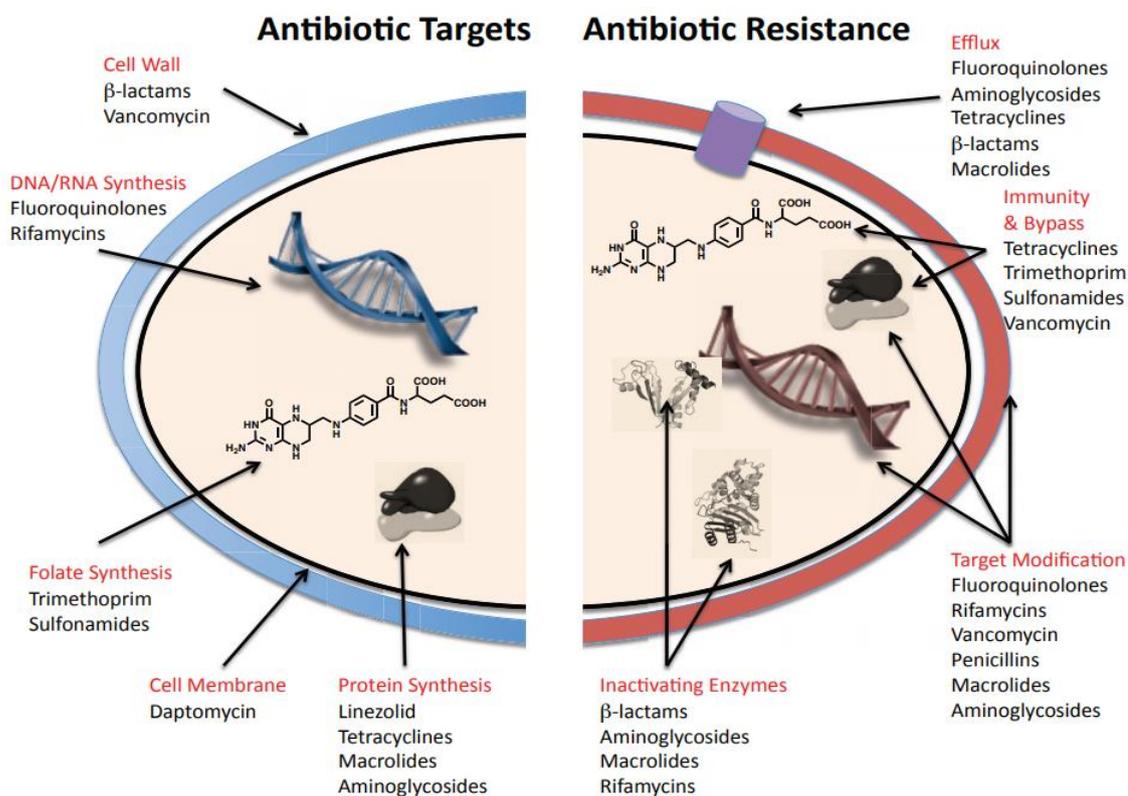
Figure 2-2. Antibiotic targets and mechanisms of resistance[58].

Bacteria can become resistant to antibiotics through multiple and complex mechanisms such as inactivating the drug by enzymatic degradation or expressing efflux pumps to lower the antibiotic concentration in the cell[27] (Figure 2-2). Often, resistance towards an antibiotic is a combination of several mechanisms as observed for tetracycline resistance where antibiotic modification, efflux mechanisms and target

modification can occur[1]. Some resistance genes do not only confer resistance against antibiotics, but also against other selective agents such as metals and biocides[48].

2.3   Exposure and spread of antibiotic resistance

Antibiotic resistant bacteria can infect individuals by many pathways, but three appear to be most important – (i) ARBs can be selected for by antibiotic consumption in the gut itself, ARBs can disseminate via (ii) exposed food and, or (iii) water[21] (Figure 2-3). In LMICs where sewage treatment is insufficient or non-existing, the latter two represent pathways of spread and exposure for AR, both rooted in contaminated water supplies[21]. However, even in HICs, conventional wastewater treatment plants (WWTPs) cannot always effectively reduce the burden of antibiotics, ARBs and ARGs[19,59–62]. Conventional wastewater treatment set-ups vary greatly but typically include pre-treatment (screening of large solids), primary treatment (settlement of suspended solids), secondary treatment (biological treatment through bacterial breakdown such as activated sludge) and sometimes tertiary treatment (such as filter membranes or disinfection by UV light)[63].

Humans and animals excrete a large fraction of consumed antibiotics in their biologically active form via urine and faeces[64–66]. Despite major advancements, biological WWTPs only moderately (50-80%) remove these drugs[18]. Depending on the WWTP process applied, Hiller et al., 2019 observed a vast range in removal efficiencies for tetracycline resistant faecal coliforms, ranging from 0.6 to 5.4 log removal. Log removals measure the ability of a treatment process to remove microorganisms[68]. A log removal of 1 and 2 are equivalent to 90% and 99% removal of the microorganisms at a specific treatment step, respectively. Narciso-da-Rocha et al. (2018) observed secondary treatment to reduce ARG abundance by 2 log units but no further significant reductions were recorded after the subsequent UV disinfection stage. The same study found UV treatment to reduce viable culturable *Enterobacteriaceae* by 2 log units[69]. Quintela-Baluja et al. (2019) showed that while concentrations of most ARG classes significantly reduced from the influent to effluent, WWTPs did not significantly reduce ARG richness and numbers of ARGs per genome. Using genetic source tracking, they found ARGs associated with human faecal pollution often directly passing through the WWTP and discharge in the environment.

Only a part of the ARBs released in WWTP effluent are able to cause diseases in humans[71]. However, their contribution to the environmental resistome and subsequently to the emergence of newly resistant pathogenic bacteria should not be neglected[71].

Polluted water sources can pose an AR exposure risk to humans, for example by swimming in impacted surface waters[72]. To holistically assess AR exposure, water and sanitation sources need to be considered together with other important factors such as hospital visits, international travel, meat consumption, agriculture and aquaculture[36,71] (Figure 2-3).
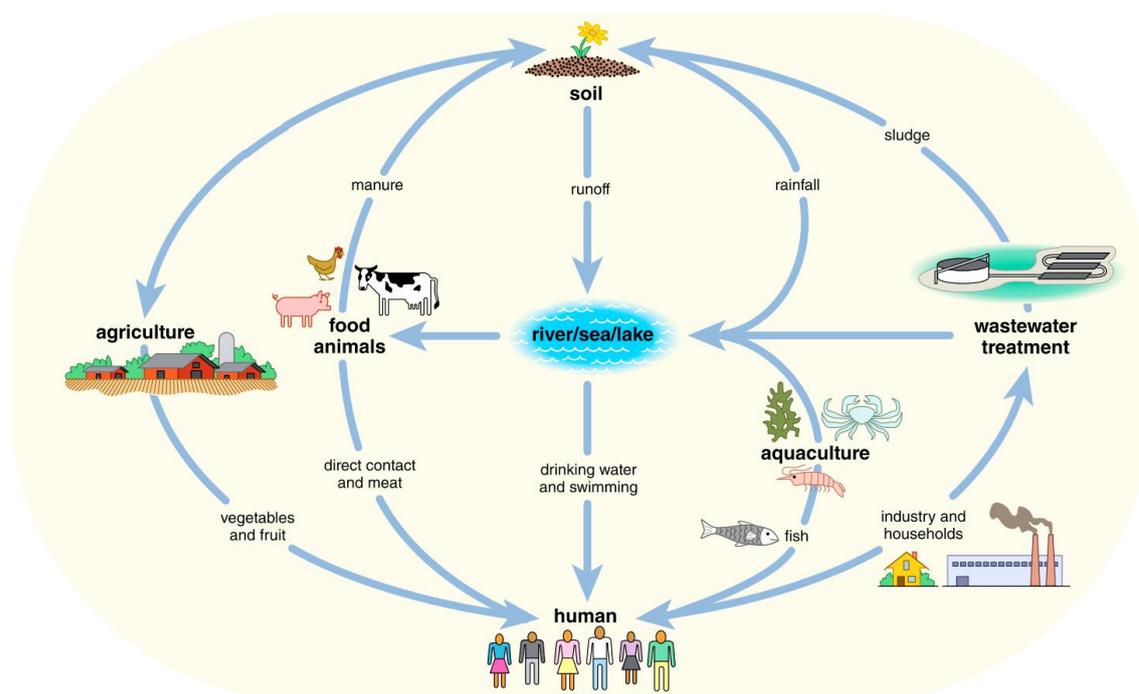
Figure 2-3. Antibiotic resistant bacteria can easily spread between people, animals, and the environment[1].

2.4 Antibiotic consumption and resistance in LMICs

Between 2000 to 2015, global antibiotic consumption per capita rose by 39% with the majority of this increase explained by growing demands in LMICs (Figure 2-4A)[73]. While antibiotic consumption rates in LMICs are converging to HIC standards, they have not yet reached them, despite the higher bacterial disease burden in LMICs[73]. Even though total antibiotic consumption has increased significantly in LMICs (114% from 2000 to 2015) (Figure 2-4B)[73], access to these life-saving drugs is not yet warranted for all income segments of the population[20].
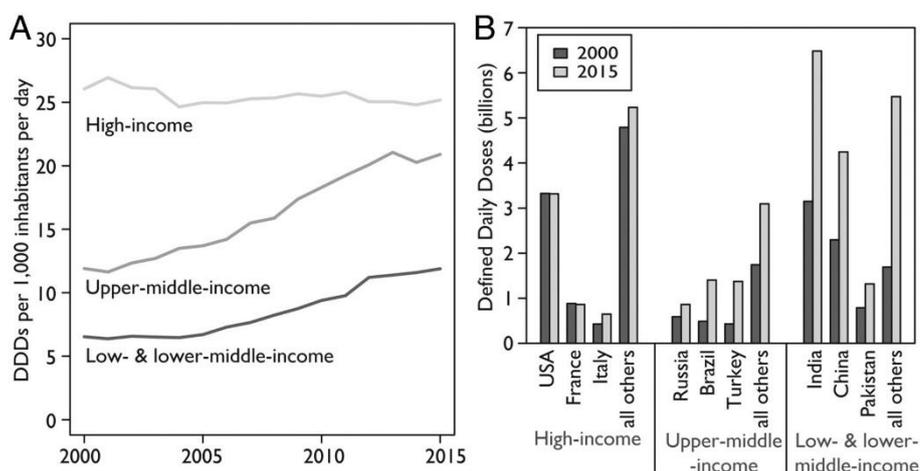
Figure 2-4. Comparison of the change in global antibiotic consumption from 2000 to 2015 as a function of country income classification, considering antibiotic consumption rate in defined daily doses (DDDs) per 1,000 inhabitants per day (A) and total antibiotic consumption in DDDs (B)[73].

Common challenges in LMICs are over-the-counter sales of antibiotics and self-medication[20]. This, combined with weak healthcare systems and unsafe or poorly maintained sanitation systems, has contributed to a disproportionally high prevalence of resistant pathogens in LMICs[20] (Figure 2-5). It is alarming that second- and third-line antibiotics required to treat these resistant bacteria are not widely available in LMICs, where more than three quarter of the world population live[20,27].
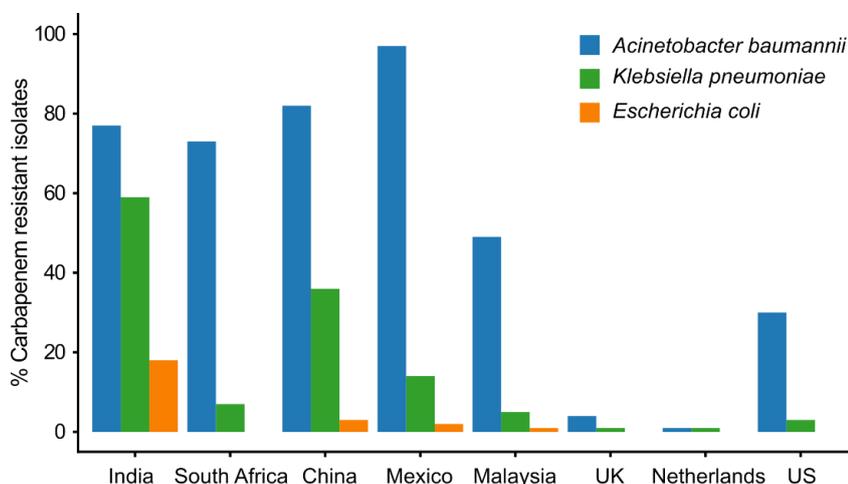
Figure 2-5. Cross-country comparison of carbapenem resistance patterns in blood and cerebrospinal fluid isolates for WHO priority pathogens *A. baumanii, K. pneumonia* and *E. coli*[74]. Country order based on 2017 gross domestic product (GDP) per capita[75]. Data source: The Center for Disease Dynamics Economics & Policy, 2019. Data for India, China, Malaysia, UK, Netherlands and US from 2017. Data for US and South Africa from 2016. Data for Mexico for *K. pneumonia* and *E. coli* from 2015 and for *A. baumanii* from 2014.

2.5 Antibiotic resistance and sanitation in Southeast Asia – Malaysia

2.5.1 Situation in Southeast Asia

A WHO conducted risk assessment found SE Asia to be one of the most at-risk parts of the world for the emergence and spread of AR[25]. For this thesis, SE Asia is defined to include Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste and Vietnam (Figure 2-6a)[77].

Within SE Asia, wealth, cultural traditions, sanitation levels and AR prevalence vary greatly across and within countries[77]. Where data is available, infection rates in adults and children with ESBL producing bacteria vary from 0 − 20% (Brunei) to 20 − 40% (Malaysia, Thailand, Philippines) to 60 − 80% (Myanmar and Vietnam)[78]. ESBL bacteria produce an extended spectrum enzyme that breaks down the majority of β-lactam antibiotics such as penicillin[78]. Infections with ESBL bacteria are treated with the remaining β-lactam antibiotics, called carbapenems[78]. High rates of mortality occur when pathogens become resistant to these last-resort antibiotics[78] (Figure 2-5).

Available antibiotic consumption rates in 2015, based on imported and produced data, differ from lowest defined daily doses (DDDs) per 1,000 inhabitants in Brunei, Philippines and Indonesia (6 − 8) to increased rates in Malaysia, Singapore and Thailand (12 − 18) and highest rates observed for Vietnam (31)[79]. The gross domestic product (GDP) per capita correlates with the percentage of people connected to the sewer network (Figure 2-6b). While almost all sewage in the high-income countries Brunei and Singapore is treated in WWTPs, sewer connection rates are much lower in the upper middle-income countries (Malaysia, Thailand) and do not commonly exist in the lower middle-income countries (Cambodia, Indonesia, Laos, Myanmar, Philippines, Timor- Leste, Vietnam)[80] (Figure 2-6b). Even within a country, sanitation levels can vary greatly depending on the region. For instance, while in urban Malaysia, 42% of the population are connected to sewers, this is only the case for 12% of the rural population[80].
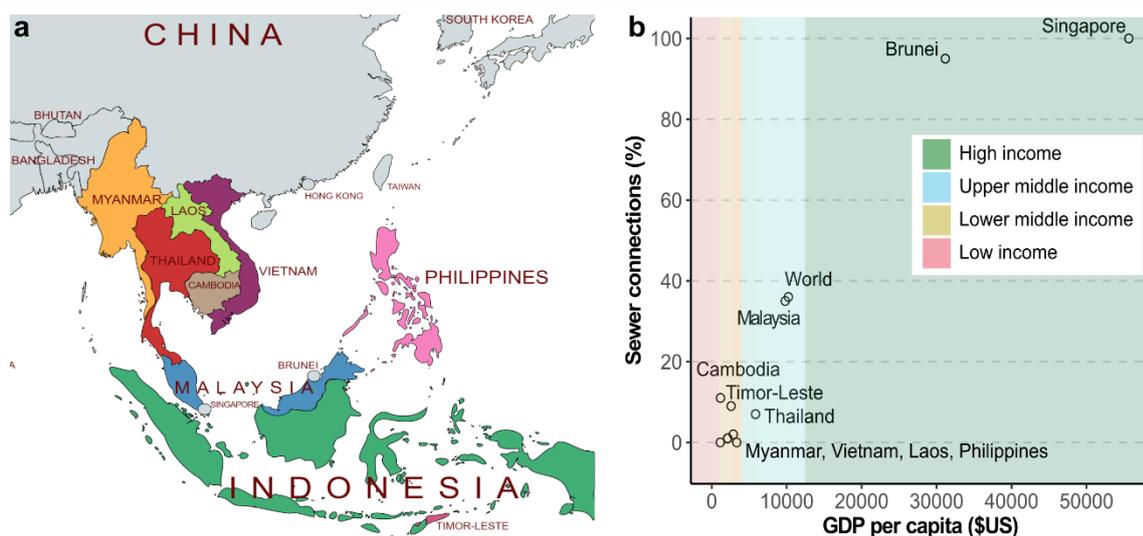
Figure 2-6a. Map of Southeast Asia (created with mapchart.net). b: Relationship between population percentages connected to sewers (combined rural and urban, data for 2015) and gross domestic product (GDP) per capita (data for 2015) in Southeast Asia. Data source: WHO-UNICEF, 2017; World Bank, 2019.

2.5.2 Antibiotic resistance in Malaysia

Malaysia, the study site, has one of the strongest, fastest-growing economies in SE Asia (Table 2-2)[81]. Increased wealth has allowed more Malaysians to access healthcare, including antibiotics. A national study in 2014 found that antibiotics were prescribed in 21% of patient encounters, although 46% of these were for upper respiratory tract viral infections, where antibiotics are often not suitable[82].

In 2000, a National Surveillance of Antibiotic Resistance (NSAR) programme was initiated to monitor AR bacteria in hospitals[83]. Particular local concern are increasing ESBL producing *Enterobacteriaceaea* and carbapenem resistant pathogens[84]. The NSAR programme recorded an increase in *Acinetobacter baumanii* resistant to carbapenem from 49% in 2008 to 61% in 2017[84]. Carbapenem-resistant *Klebsiella pneumoniaea* rates have increased from 0.3% in 2011 to 2.9% in 2017[84]. Spreading AR means that infectious diseases are more difficult to manage, ultimately potentially resulting in more morbidity and mortality in Malaysia[85].

In 2015, the World Health Assembly endorsed its global action plan on antimicrobial resistance (including AR[11]). This led Malaysia to establish their Malaysian Action Plan on antimicrobial resistance (MyAP-AMR) in 2017. The MyAP-AMR follows a One Health approach, combining human, animal and environmental health interventions[83].

Table 2-2. Country information for Malaysia.

Area	330,000 km^2 [86]
Population	32 million [87]
Ethnic groups	69% Bumniputera or indigenous (Malay, Orang Asli and other bumiputera), 23% Chinese, 7% Indian, 1% others [87]
Languages	Malay (official), Chinese, Tamil, English [86]
Religions	61% Islam (official, Government of Malaysia, 2010), 20% Buddhism, 9% Christianity, 6% Hinduism, 4% others [89]
Key industries	Manufacturing (electric appliances), agriculture and forestry (natural rubber, palm oil, and timber), and mining (tin, crude oil, and LNG) [86]
Gross domestic product (GDP) per capita (2018)	11,239 $ [75]

2.5.3 Sanitation and surface water quality in Malaysia

Waste treatment is not sufficiently developed in Malaysia to remove residues of antibiotics, ARGs and ARBs effectively[80,90]. National data from 2017 show 79% of the Malaysian population connected to sewers with 20% serviced by septic tanks and <1% relying on latrines and other[91]. Improvement to secondary (biological) treatment has taken place in some areas[90] but many locations still rely on septic tanks. Despite substantial investments and increasing awareness, surface water pollution, mostly through sewage, is still a major problem in Malaysia[92] (Figure 2-7). This is critical as surface waters supply 98% of fresh water for the country[92].

Malaysia has an established, wide-ranging river water quality program[93]. In 2015 (latest publicly available dataset), 477 rivers were monitored at regular intervals (usually every 1-3 months) for standard water quality parameters such as dissolved oxygen (DO), temperature, but also metal and coliform levels[93,94].

River contamination data is assessed by calculating a Water Quality Index (WQI) to define suitable terms of water use following the National Water Quality Standards for Malaysia (NWQS). The WQI incorporates parameter values for DO, biochemical oxygen demand (BOD), chemical oxygen demand (COD), ammonia (NH$_3$-N), suspended solids and pH[92]. From 2005 to 2015, water quality based on the WQI sub-index BOD decreased substantially in Malaysian rivers while the overall WQI classification ratio remained relatively stable (Figure 2-8)[94].

Figure 2-7. Two children standing on the edge of a plank walkway overlooking a garbage- and debris-strewn area of water in Borneo, Malaysia[30].

Similarly to environmental monitoring programs in other countries, the Malaysian river quality programme does not measure any AR parameters[94]. Despite the NSAR programme documenting the increase of AR in Malaysian hospitals, almost no research has been conducted on the occurrence and spread of AR in surface water. Only a few Malaysian studies have monitored selected ARGs, ARBs or antibiotics in surface waters[79,95–97], but no study has yet monitored combined seasonal and spatial effects for ARGs, ARBs and antibiotic concentrations and mass loadings.
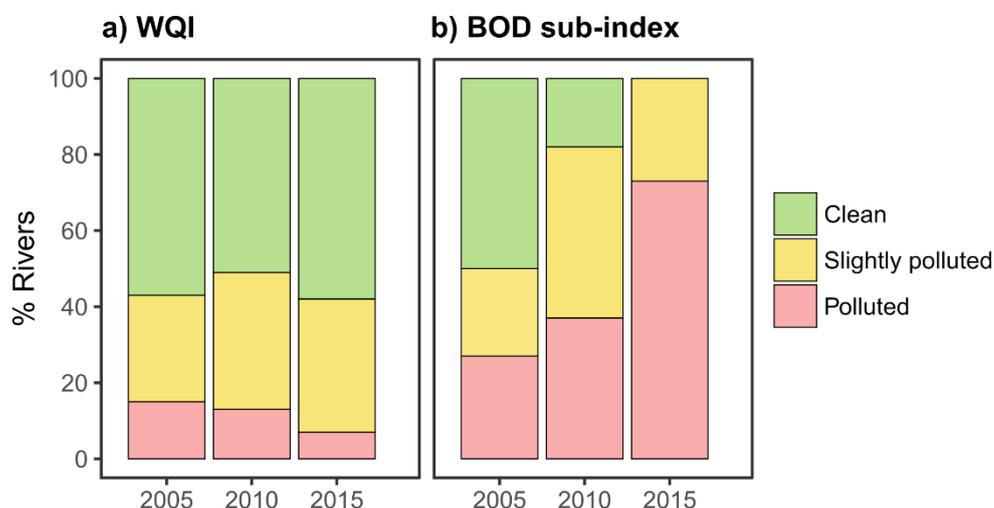
Figure 2-8. Change of river water quality in Malaysia from 2005 to 2015 based on the water quality index (WQI) index (a) and with a focus on the biochemical oxygen demand (BOD) sub-index (b) for over 450 annually monitored rivers. Data source: Department of Environment, 2015.

2.6 Monitoring antibiotic resistance in surface waters

Systematic surveillance of antibiotic use and antibiotic resistance is required to tackle AR through a One Health approach[11,98]. While long-standing clinical surveillance exists for AR in several countries[9], including Malaysia[84], environmental monitoring is

still in its infancy[28]. Comprehensive environmental AR exposure assessments are required to support holistic approaches to reducing AR. Various AR endpoints can be monitored, each with its advantages and disadvantages[19,36]. Ultimately, the choice of markers and sampling sites should be chosen based on the surveillance objective[28,71].

Given the limited available environmental AR data in LMICs[27], results need to be reported in a consistent way to allow comparison to other studies and potential extrapolation to regions with no data. Unfortunately, most studies only report statistical significance (P values) instead of calculating more meaningful statistics such as standardised effect sizes. Dimensionless effect sizes describe the size of a difference and allow to compare studies from different settings with different variables[99]. Despite the popularity of effect sizes in meta-analysis and psychological studies, they are not yet commonly used in AR/water quality monitoring studies.

In the following, the most relevant endpoints for AR monitoring are discussed with a focus on their feasibility for LMICs.

2.6.1 Antibiotic residuals and selective agents

Antibiotic concentrations in surface water are generally low with weak or no effects observed for AR selection[36,100]. Examples for highly polluted sites can be surface waters receiving effluents of pharmaceutical manufacturing companies. High drug releases in rivers have been linked to increased ARG and MGE levels[101,102]. Probably one of the most worst-case scenarios for environmental antibiotic pollution was observed by Larsson et al. (2007). They analysed effluents of a WWTP in India serving over 90 bulk drug manufacturers and detected several antibiotics in high concentrations over 100 µg/L with very high ciprofloxacin levels ranging up to 31 mg/L.

Typically, antibiotic concentrations in surface waters range from non-detectable to ng/L and lower µg/L figures, but can vary vastly depending on the region[79,104]. A review of antibiotics in surface water of East and SE Asia found antibiotic concentrations in surface waters to range from <1 ng/L to hundreds of µg/L with median concentrations from 10 to 100 ng/L[79]. The same study found wider ranges and higher maximum concentrations for certain antibiotics in surface water of East Asian countries such as China and South Korea than in the SE Asian nations such as Malaysia and the Philippines[79]. A meta-analysis observed average

sulfamethoxazole levels in surface waters to be 100-times higher in Africa (4.7 ± 8.9 μg/ L) than in Europe (41 ± 87 ng/ L)[104]. Nevertheless, while surface water antibiotic levels are usually augmented in LMICs in comparison to HIC, other factors such as water quality and sanitation levels are suggested to play more critical roles in reducing AR burden[56,100,105].

Other aspects, especially relevant for LMICs, are the cost and expertise required to maintain and operate analytical instruments for antibiotic measurements such as liquid chromatography-tandem mass spectrometry (LC MS/MS)[36]. Once antibiotic concentrations have been measured it can also be challenging to understand the environmental risk associated with the observed levels. There have been attempts to define threshold levels above which an antibiotic might exert selection for resistant bacteria[106,107]. These 'predicted no effect concentrations' (PNECs) can help to classify pollution levels, but ultimately remain estimates. For instance, PNECs do not account for additive effects of antibiotics and other selective agents such as metals or biocides. Another challenge is that there is no environmental guidance available on which antibiotics to target, except for the clinically orientated WHO ranking of critically important antimicrobials for human medicine[36,108].

Where antibiotic usage data is difficult to obtain in a LMICs context[109], surface water concentration data can help to estimate drug consumption rates. However, rapidly degradable antibiotics, such as many β-lactams, might not be detected representatively and could skew the observations[110–112].

2.6.2 Culture-based techniques

Growing bacteria on selective antibiotic agar allows quantification of viable ARBs with well-known traits[36]. AR monitoring in clinical isolates is well established and standardised, but specific methods for environmental samples were never developed, rather adapted[113].

Recently, the WHO proposed ESBL producing *E. coli* as an indicator organism for global AR surveillance in humans, animals and the environment[114]. The advantages of monitoring this organism are that it

(1) is based on established and accessible methods in HICs and LMICs,

(2) is included in the WHO list of antibiotic resistant 'priority pathogens' posing the greatest health threat[74],

(3) can describe AR levels comparatively for food, clinical and environmental samples and

(4) allows quantifying variations between different regions in HICs and LMICs[27].

ESBL *E. coli* monitoring as part of the WHO Tricycle Project has now been pilot-tested (2018-2020) in six LMICs (Ghana, Indonesia, Madagascar, Malaysia, Pakistan and Senegal) with plans to further expand the projects to three others (India, Jordan and Nepal)[115]. For this, training workshops and technical assistance are provided to the countries willing to implement ESBL *E. coli* monitoring[115].

However, limitations of using culture based indicator organisms for AR monitoring are that (1) a single indicator organism might not reliably represent other enteric pathogens[116], (2) complex microbial community compositions are not assessed, and (3) culturing misses the majority of (especially unculturable environmental) bacteria[36,117]. For the latter, depending on the environment, less than 1-10% of total bacteria present might be culturable[118,119].

2.6.3 *Culture-independent techniques*

Culture-independent, nucleic-acid-based approaches (such as DNA sequencing or quantitative PCR (qPCR)) allow to monitor ARBs, ARGs and MGEs in the environment[27]. While these costly methods are often not suitable for routine monitoring in LMICs, they provide valuable insights in environmental resistomes, which cannot be obtained via culture-based techniques[36]. Culture-independent techniques have allowed the characterisation of environmental resistomes in many diverse samples, both within LMICs and HICs[27,120,121]. ARGs and MGEs are typically monitored by qPCR (conventional qPCR or high-throughput qPCR (Ht-qPCR)), but metagenomic DNA sequencing can also be applied.

With sequencing costs constantly falling, metagenomic sequencing is becoming more widely accessible[122]. DNA sequencing allows to profile microbial communities[70], including thousands of ARGs and MGEs[121], without prior target selection. The challenge of correctly annotating ARGs from sequencing data remains, but the robustness of publicly available ARG databases such as Comprehensive Antibiotic Resistance Gene Database (CARD)[123] and the Structured Antibiotic Resistance Gene (SARG) database[124] is improving[36]. Another challenge is to comprehensively analyse the increasing volumes of datasets to obtain global,

meaningful insights about AR. For this, machine and deep learning methods are being developed[125] together with tools to rank the risk of ARGs for metagenomes[126]. Field-deployable sequencing devices such as Nanopore[127] could provide widely accessible AR analysis for LMICs[100,128,129], but these devices are currently still too cost-intensive and can be experimentally challenging[27].

In comparison to sequencing, qPCR is fairly accessible and provides a more quantitative and sensitive assay[36]. Previously, only a small fraction of ARGs could be targeted in environmental samples, but increasing levels of whole genome sequences now provides a vast array of ARG targets from databases[130]. Ht-qPCR arrays allow the simultaneous quantification of hundreds of ARGs, while providing lower detection limits than sequencing approaches. Waseem et al. (2019) estimated ARG detection limits for Ht-qPCR in the range of 10^{-4} ARGs/16S rRNA, while the same detection limit for sequencing would require at least 10^8 reads during metagenomic analysis.

While analysing environmental samples with Ht-qPCR provides a quick overview of the presence/absence and quantity of several hundreds of ARGs and MGEs, the approach is expensive and typically, less accurate than performing individual qPCR assays. The cost of analysing one sample with Ht-qPCR depends on the amount of genes included, but typically ranges from £100 to £300 per sample[131]. Only a few specialised research institutions and companies offer Ht-qPCR analysis to quantify ARGs and MGEs in environmental samples. For LMICs (and most HICs), one strategy can be to firstly analyse representative samples of selected environments/regions with Ht-qPCR to characterise the resistome and then perform follow up monitoring studies on a few selected ARGs/MGEs with individual qPCRs.

However, both approaches, sequencing and qPCR, cannot easily link ARGs and MGEs to their hosts in complex environmental samples. Developments of epicPCR represent a step forward, but the methodology cannot yet differentiate between many genera and species[28,132]. In addition, it can be difficult to distinguish if the detected ARGs and MGEs are present in either dead or live cells[133]. Some studies have proposed to link microbiome data from DNA sequencing with (quantitative) ARG data to study environmental AR[134,135]. However, next-generation sequencing (NGS) data is inherently compositional, providing relative abundance information[136]. This makes it difficult to apply statistical tools to correlate the presence of ARBs with concentrations of ARGs[137].

It remains challenging to reliably assess the risk associated with measured ARG and MGE values as defined targets and threshold concentrations are missing. ARGs encoding resistance to clinically important antibiotics could be prioritised, but these highly relevant genes are often rare in the environment and might not always be detectable with current methods[36]. Analysing the sum of all analysed ARGs can inform about selection pressures, but might be dominated by a few highly abundant ARGs, not capturing differences in more clinically relevant genes[28].

Alternatively, MGEs have been proposed as AR markers as they can carry a large number of ARGs. In particular, the clinically relevant class 1 integron (*int1*) is promising[138]. However, although class 1 integron is ubiquitous in human and animal microbiomes, it is not associated with particular bacteria or pathogens, or specific to AR[28].

In conclusion, each AR marker has its advantages and disadvantages but where feasible, monitoring various endpoints can provide the most complete AR assessment[19,36]. For LMICs, routine surface water monitoring for ARBs such as ESBL *E. coli* can be a useful, easy to apply and cost-effective approach to capture environmental AR levels. However, where equipment and expertise are available, additional targeted in-depth analyses of environmental resistomes with culture-independent techniques such as qPCR or DNA sequencing provide valuable data to better understand AR burden[19].

2.7   Modelling antibiotic resistance in watersheds

2.7.1   Status of antibiotic resistance modelling

Next to field monitoring and laboratory experiments, modelling can provide an important tool to identify and manage the risk of AR[139,140]. Models can help to understand and predict the burden of AR by characterising influencing factors and validating intervention methods[29]. In addition, AR models can elucidate knowledge gaps and direct research to identify missing parameters and processes in the modelled system[29].

While there has been a growing interest in modelling AR[141], the majority of models only focus on human related AR transmission in either hospital or community settings[142]. However, to address the One Health dimension of AR[143], it is essential to also model transmission routes within and between populations (microbiota, animal and human populations) and across systems (communities, farms, rivers) (Figure

2-9). Considering the delayed consideration of monitoring AR in the environment, it is not surprising that environmental AR modelling is still rudimentary[142]. Birkegård et al. (2018) found that out of 38 assessed mathematical AR models, only four considered the environment, with one model describing AR bacteria growth in slurry[144] and three models considering AR spread and survival in rivers[145–147].
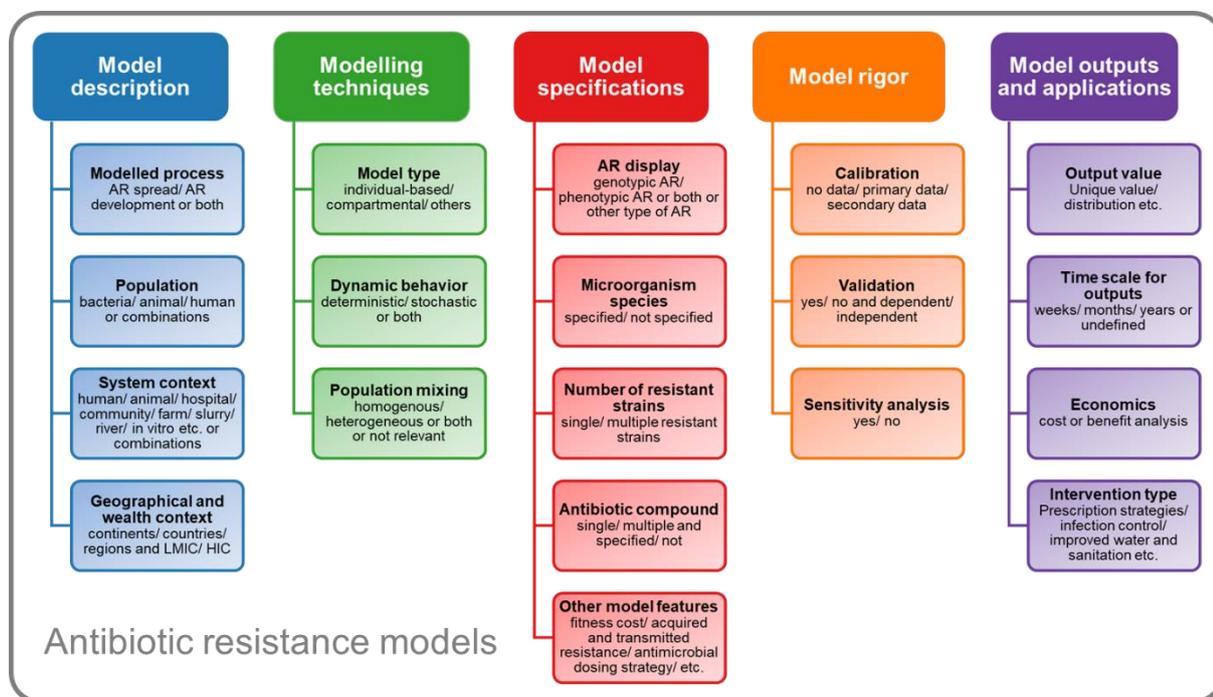
Figure 2-9. Characteristics of AR models (based on [29,142,148]).

Current AR models vary in structure and complexity, but often there is little justification given for the chosen model type, dynamic behaviour state or population mixing assumption (Figure 2-9)[29]. In addition, the majority of models ignore the uncertainty and stochasticity in the emergence and fate of AR, for instance modelling one single strain of a pathogen in one unit assuming homogenous mixing[29]. While the difficulty to model the complexity of AR is acknowledged, especially in view of knowledge gaps, more research initiatives should aim to consider external factors and multifaceted interactions in their models[29].

When single strains of pathogens are modelled, current studies often concentrate on diseases with long-established resistance, while few studies address recent rises in resistance in new pathogens[74,142]. For instance, Ramsay et al. (2018) found that only two out of 81 AR models examined the urgent threat of ESBL *Enterobacteriaceae*[149,150] with no model studying carbapenem-resistant *Enterobacteriaceae*.

Another aspect is that current AR models mostly assess 'micro' scale interventions such as the impact of different drug regimens or infection control strategies (e.g. isolation, hand hygiene), but rarely consider 'macro' scale interventions such as improving waste treatment or sanitation[142]. And while recent studies have linked climate change to potentially increase AR levels[151,152], no AR model has yet considered this environmental factor, except in general and non-specific terms[142].

In many studies, AR models are based on hypothetical situations in hospitals or communities with no supporting data[29]. Niewiadomska et al. (2019) found in their review only 43% of mathematical AR models calibrated against data with only 14% validating against independent datasets. Validation is an essential component of model development to reduce the risk of inaccurate model outcomes and conclusions[29]. The absence of data-based calibration and validation in many AR models could be due to a lack of sufficient data[29]. Data scarcity for AR models is a major concern for environmental studies, especially in LMICs with comprehensive AR monitoring being resource- and time-intensive[142].

2.7.2 Antibiotic resistance river models

Operational watershed models cover a wide array of water quality parameters and pollutants for river management (see 2.7.3)[153]. However, none of these off-the-shelf watershed models currently includes an AR component. To date, only a few AR research models have been published for specific locations, as summarized below[145,146,154,155].

Lawrence et al. (2010) developed one of the first mathematical models to describe ARB concentrations in rivers. Their relatively simple, hypothetical approach focussed on modelling the influx of bacteria from the shore (such as from WWTPs), transfer and loss of ARGs, and the river carrying capacity.

In contrast to this purely theoretical model, Hellweger et al. (2011, 2013) developed and applied a tetracycline resistance model to the Poudre River in Colorado. Their more complex model included variables for antibiotics, metal toxicants, susceptible and resistant bacteria, and organic matter. They calibrated their model with field data for tetracycline, ARG and ARB concentrations at five locations along the river. The model incorporated a basic streamflow element and accounted for some point (WWTPs, agriculture and livestock) and non-point pollution sources (based on land-

use). Depending on the assumptions, the model was able to explain some observed levels of tetracycline resistant bacteria in the Poudre River.

Gothwal and Thatikonda (2017) developed and applied a model for the transport of fluoroquinolone and its resistant bacteria for the Musi River in India. They accounted for the same variables as Hellweger et al. (2011, 2013), but applied the model to a LMIC setting to predict pollution conditions under different management interventions. They calibrated their model with a homogenous dataset for fluoroquinolone and fluoroquinolone resistant bacteria concentrations at 16 points along the river.

Such studies are promising but lack elements of comprehensive watershed modelling to describe water quantity and quality. The section below describes the mechanistic of watershed models and which models could be considered to incorporate an AR element to describe AR fate and exposure.

2.7.3 Watershed models

A watershed model quantifies the fluxes and storage of water within a catchment. Such models are typically based on the water cycle (Figure 2-10) and help to understand, predict and manage water resources[156].

The water cycle describes the transfer of water from precipitation to surface water and groundwater, to storage and runoff, and ultimately back to the atmosphere through evaporation[157]. The water mass balance describes these processes in a simplified form, forming the basis of watershed models (Equation 2-1).

Equation 2-1. Water mass balance.

$$\Delta S = P - E - Q$$

ΔS: Change in storage of soil/groundwater
P: Precipitation
E: Total evaporation
Q: Discharge

Based on the spatial discretisation of the catchment, watershed models can be grouped into lumped, semi-distributed and distributed (Table 2-3). Semi-distributed models represent important features in catchments while being less data intense than distributed models (Figure 2-10b). One of the most comprehensive semi-distributed watershed models is the freely available Hydrological Simulation Program – FORTRAN (HSPF).
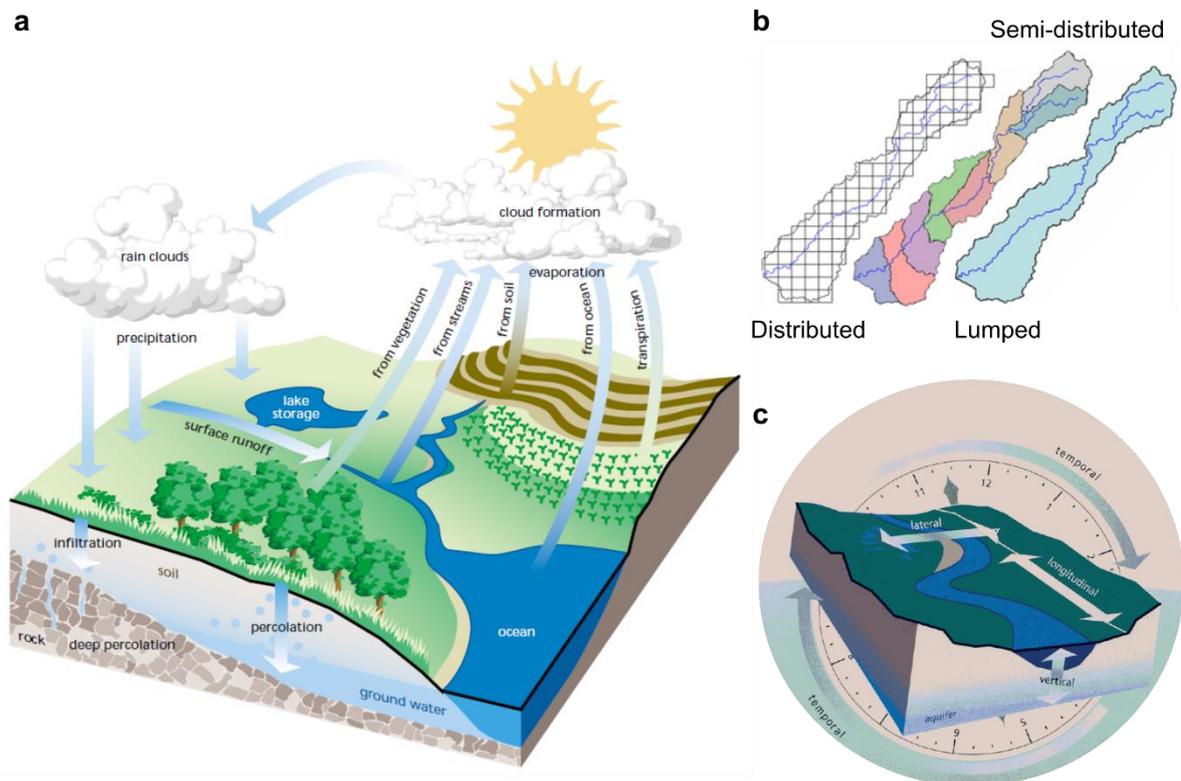
Figure 2-10. a) The water cycle[157]. b) Distributed, semi-distributed and lumped modelling of a catchment[158]. c) Three dimensions (longitudinal, vertical and lateral) of the stream corridor[157].

HSPF is one of only a very few models which enables the integrated simulation of land and soil contamination runoff processes with in-stream hydraulic and sediment-chemical interactions[159]. HSPF was developed by US Environmental Protection Agency (EPA) and is incorporated into the multipurpose environmental analysis system Better Assessment Science Integrating Point and Nonpoint Sources (BASINS)[160]. HSPF can simulate nonpoint and point-source pollutant loadings at any point in the watershed[161]. Runoff qualities are calculated with both simple relationships (such as empirical build up/wash off and constant concentrations) and detailed soil process options (such as leaching, sorption, and soil nutrient transformations)[162]. HSPF simulates organic chemical transfer and reaction processes such as biodegradation, photolysis, oxidation, hydrolysis, volatilization and sorption [162]. Stream nutrient processes cover DO, BOD, nitrogen and phosphorus reactions, pH and microorganisms[162]. HSPF can simulate detailed interactions between 15 or more water quality constituents within a river reach[159]. Despite its widespread use and applicability[163–166], an environmental AR component is not yet available for HSPF or any other watershed model.

Table 2-3. Comparison of the basic structure for rainfall-runoff models[167].

	Lumped	Semi-Distributed	Distributed
Method	Spatial variability is disregarded, entire catchment is modelled as one unit	Series of lumped homogenous units, with semi-distributed parameters	Spatial variability is accounted for, for example through the use of a grid
Inputs	Spatially averaged data for the catchment	Spatially averaged for the units	Data specified by grid cell
Strengths	Fast computational time, good at simulating average conditions	Represents the important features in catchment	Represents hydrological processes, often using physically based equations
Weaknesses	A lot of assumptions, loss of spatial resolution, not ideal for large areas	Averages data into sub-catchment areas, loss of spatial resolution	Data intense, high computational requirements
Examples	Empirical and conceptual models, machine learning	Conceptual and some physical models, TOPMODEL, SWAT, HSPF	Physically distributed models, MIKESHE, VELMA

2.8   Conclusion

The spread of AR pathogens is a global threat to human, animal, and environmental health. While the burden of AR has been studied intensely in clinical settings, the environmental dimension has long been neglected. Environmental AR is particularly a problem in LMICs, where insufficiently treated waste pollutes surface waters with residues of antibiotics, ARGs and ARBs. Contaminated waters enrich the environmental resistome and can pose a health risk to locals using surface waters for fishing, irrigation or recreational use.

SE Asia has been proposed as a hotspot for emerging diseases and AR spread. Clinical β-Lactam resistance is spreading in many countries, including in Malaysia. Malaysia is a middle-income country with reported over- and misuse of antibiotics while wastewater treatment cannot sufficiently remove residues of AR. Despite these alarming factors, comprehensive surface water AR monitoring has not yet been performed in Malaysia or many other SE Asian countries. Providing a comprehensive picture of environmental exposure could provide a strong starting point for One Health as part of the health protection system for Malaysia. Such a One Health approach to AR is crucial to reduce increasing morbidity and mortality caused by infections with AR pathogens in Malaysia and elsewhere.

Most AR monitoring data sets are only available for HICs as the required field work is resource- and time-intensive. In addition, agreements and guidance are missing on

which AR parameters to monitor with which methods. Many LMICs, including Malaysia, operate national river water quality programmes, but these do not include AR parameters. Where research studies on environmental AR are available, limited statistical analyses often hinder the comparison and extrapolation of findings to other studies and/or settings, respectively.

Existing watershed models such as HSPF could help to pinpoint AR hotspots when only limited/no environmental AR data are available. However, available river models do not include an AR element and are often not particularly well suited for LMICs.

AR models are still in their infancy and are generally only used within research settings. Challenges include establishing a consensus on an appropriate model structure, identifying which processes, and in how much detail need to be included, and how to quantify uncertainties in predictions.

Therefore, to address AR in Malaysia, SE Asia and LMICs in general, new AR monitoring data is required to develop predictive river water AR models. These AR models can then help to understand and predict the burden of AR by characterising influencing factors and validation intervention methods.

Chapter 3. Developing surrogate markers for predicting antibiotic resistance 'hot spots' in rivers where limited data are available

3.1 Introduction

Increasing resistance in microorganisms to antibiotics and other drugs poses a global health threat[10]. When a pathogen becomes resistant to critical drugs, formerly easy-to-treat infections can become lethal[168,169]. Consequently, scientists and policy makers must better understand drivers of AR to reduce its global spread. The number of peer-reviewed AR papers has quadrupled during the last ten years (title or abstract containing 'antibiotic resistance' web of science from 2010 to 2020) with more than 9,000 papers published in 2020 alone. However, our understanding of environmental AR spread lags behind other contexts[36]. When insufficiently treated wastewater enters rivers, residues of antibiotics, ARBs, and ARGs can radiate through the environment, potentially posing an exposure risk[170,171]. However, mitigating environmental AR spread is hindered by many factors, including: (1) inadequate data to make decisions about environmental AR exposures; (2) the complexity and diversity of environmental matrices; (3) conflicting definitions of AR and inconsistency in measuring methods; (4) reliance on overly expensive detection methods; (5) limited agreement on AR thresholds of possible concern; and (6) a limited understanding of how environmental AR levels translate to human health risk[28].

Limited data and expensive AR detection methods are especially problematic in LMICs, particularly comparing places and times, and identifying sites of greatest concern[172]. This is partly because most studies are more academic rather than practical, but also because researchers overly focus on testing statistical significance (P values) to report spatial or temporal differences. A lower P value is often interpreted as meaning a bigger difference between two settings, but statistical significance only means that it is unlikely for the null hypothesis to be true (such as H_0 = no difference in antibiotic concentration between up- and downstream river locations)[173], which often has limited value in quantifying the scale of differences.

In contrast to P values, we feel 'standardised dimensionless effect sizes' better describe the size of differences and allow comparison of studies from different

settings with different variables[99]. Effect sizes are easy to calculate and, unlike P values, provide a comparison independent of sample size[99]. Surprisingly, despite the popularity of effect size in meta-analysis and psychological studies, they have not been used in AR/water quality studies. We argue that to effectively interpret and compare AR levels, both statistical significance (P value) and substantive significance (standardised effect size with confidence intervals) should be reported[99].

Increasing the informative value of monitoring data is especially critical in LMICs. While antibiotic use per person is increasing in LMICs to HIC rates, sewage treatment often lags behind[27,73,80]. SE Asia with its rapid economic development has been proposed as an epicentre for emerging infectious diseases and AR[22–24]. In particular, ESBL producing and carbapenem resistant pathogens pose major health threats in the region[78,174].

Despite LMICs carrying a higher burden of AR, including Malaysia, environmental AR surveillance more often takes place in HICs[27]. As such, there is a chronic shortage of data in most LMICs, especially the relative susceptibility of local populations to the effects of AR due to limited accurate health surveillance data[172,175]. AR transmission models have been proposed to estimate the risk of AR [29], but environmental AR modelling, which might help fill in data gaps in LMICs, lags far behind[142]. While surface water quality models have existed for decades [176], few attempts have been made to model AR spread in watersheds [145,154,155], often hindered by limited knowledge of AR fate processes in the environment, and missing AR and-or hydrological calibration/validation data.

The aim of work within this chapter is to identify 'simple' easy-to-measure water quality surrogates that would aid monitoring and modelling of AR in locations with limited data. For this, we examined the Skudai river catchment in Malaysia, using simple AR culturing methods and routine water quality markers in parallel to more sophisticated methods. Furthermore, we show the value of effect sizes for environmental AR studies, which better account for spatial, seasonal and dilution effects, and improve comparability of monitoring studies in LMICs and HICs.

3.2   Material and methods

3.2.1   Catchment description

The Skudai river catchment in southern Malaysia (total drainage area 288 km^2 [166], see Figure 3-1) is comprised of urban/developed, agriculture (80% oil palm, 20%

rubber plantations), and forest land in roughly equal proportions[177]. The Skudai catchment lays within the Johor Bahru district (1,865 km^2 with 1.4 million inhabitants [178]).
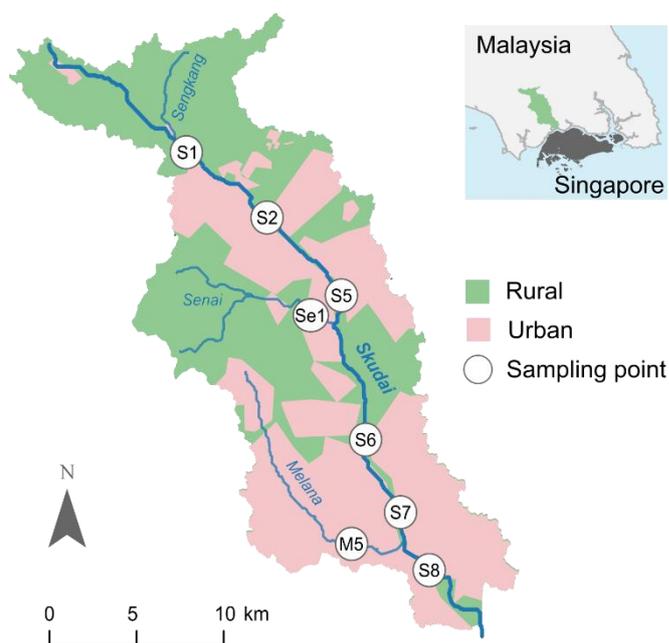
Figure 3-1. Skudai catchment in Malaysia with sampling points (102∘59'54.19" E and 104∘11'8.54" E longitude and 1∘56'31.67" N and 1∘22'41.16" N latitude).

Similar to many LMIC settings, sewage treatment in the Skudai catchment is inconsistent, sometimes with poorly defined discharge locations[97,179]. To our knowledge, no major pharmaceutical production facilities are located in the catchment[179].

The main Skudai river (42.8 km) passes rural and urban areas before it discharges into the sea (tidal influences downstream of S7). The Skudai has several tributaries, including the mostly rural Senai (11.8 km) and urbanized Melana (18.7 km)[166]. Malaysia has a humid tropical climate with two monsoon seasons, the relatively dry Southwest Monsoon from May to September and the wetter Northeast Monsoon from November to March, but substantial rainfall also occurs in the transitional periods[180,181].

3.2.2 Sample collection and processing

River water samples were collected from eight sampling points in the Skudai catchment (Figure 3-2, Appendix Table A-1) across four seasonal sampling campaigns: two in March 2018 (trips I,II) in the 'wet season' and two in July 2018 (trips III, IV) during the 'dry season'. Eight sampling points were chosen based on

land use and preliminary sampling data from 15 sites (results not shown), and included six locations on the Skudai itself (S1, S2, S5, S6, S7 and S8 during trips I-IV) and two sites on Senai and Melana tributaries (Se1 and M5; sampled during trip I, III, IV), respectively. This campaign resulted in a total of 30 sample sites from which technical triplicates were collected.

Figure 3-2. Skudai catchment photos showing Skudai river (a, b), a drain leading into Skudai river (c) and river water sampling (d).

Sampling events always were conducted over a single day in the morning, from up-to downstream, only at low tide, and on days when rainfall had not occurred within 24 h. Each sampling location was at a bridge, which allowed water collection from mid-river. Samples were collected in a pre-rinsed clean bucket (on a rope), waiting 2 min between taking each replicate. Sample water was stored in autoclaved glass bottles on ice in the dark (3 x 1 L, except 4 x 1L for S1 to assure sufficient DNA yield for downstream processes). In the laboratory, technical replicates were processed separately, splitting 1 L of sample into 15 mL for chemical analysis, 2 mL for coliform plating, 500 mL for antibiotic analysis, and 80-250 mL for DNA extraction.

River water temperature, DO, pH and conductivity were measured on-site with a HQ40D portable multi-meter (Hach). Conductivity was temperature corrected (NaCl non-linear with reference temperature 25 ºC). River volumetric flowrates were

estimated using the float method with an estimated accuracy of ± 20%[182,183]. River width and depth were used to calculate cross-sectional area at the time of sampling at each point, which was multiplied by the measured surface velocity to obtain the flowrate. A factor of 0.85 corrected for surface velocity data to provide an average across the vertical profile[182,183].

3.2.3 Chemical analysis

Water samples were filtered through 0.2 µm PES syringe filters (VWR) and stored for a maximum of 24 h at 4 ℃ prior to chemical analysis. Ammonia (NH_3-N, salicylate method[184]), chemical oxygen demand (COD, USEPA reactor digestion method[185]), total phosphorus (TP, USEPA PhosVer 3 with acid persulfate digestion method[186]), and total nitrogen (TN, persulfate digestion method[187]) were measured using commercial colorimetric test kits with a UV-vis spectrophotometer DR5000 (all Hach). Where necessary, samples were diluted with Milli-Q water prior to analysis.

The Malaysian Department of Environment (DoE) applies a Water Quality Index with three classifications ('clean', 'slightly polluted' and 'polluted') and the National Water Quality Standards for Malaysia (classes I-V) to evaluate river water quality based on selected parameters[188]. Combining both approaches, three water quality categories were created based on COD, NH_3-N and DO concentrations in the catchment: 'clean' (class I), 'slightly polluted' (class II) and 'polluted' (class III-V) (Appendix Table A-2). We compiled chemical data from S1 and S8 with national DoE river water quality data collected for the same locations throughout 2018 (Appendix Table A-3).

3.2.4 Coliform and other plating

Coliform ChromoSelect agar was used to quantify colony forming units (CFUs) of total coliform (TC), ESBL coliform (addition of ESBL supplement to agar), and carbapenem resistant bacteria (CPB-0.5 and CPB-2; addition of meropenem in dimethyl sulfoxide (DMSO) to agar at final concentrations of 0.5 µg/mL and 2 µg/mL) (all Sigma Aldrich). Each ESBL plate contained following antibiotics in final concentrations: ceftazidime 3 µg/mL, cefotaxime 3 µg/mL, ceftriaxone 2 µg/mL, aztreonam 2 µg/mL, fluconazole 10 µg/mL[189]. Meropenem concentrations were selected based on preliminary screening experiments[190] and the intermediate meropenem CLSI minimum inhibitory concentration (MIC) breakpoint for *Enterobacteriaceae* of 2 µg/mL[191]. ChromoSelect agar allowed visual differentiation of presumptive *E. coli* (subsequently referred to as *E. coli*[192]) versus other coliforms.

Where necessary, water samples were diluted with sterile phosphate buffered saline (VWR) to achieve 30 to 300 CFU per plate in three technical replicates[193]. Each plate contained 100 µL or 200 µL of sample and was incubated at 37ºC for 24 h. Negative controls and blanks were intermittently tested to verify in-lab contamination was minimised. CPB-2 were only measured for trips II-IV.

3.2.5 Antibiotic analysis

Solid phase extraction (SPE) coupled with ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) was used to quantify 22 antibiotics belonging to seven classes: β-Lactams, lincosamides, macrolides, quinolones/fluoroquinolones, sulfonamides, tetracyclines and others (Appendix Table A-4).

Duplicate 500 mL river water samples were filtered with glass microfiber and 0.45 µm cellulose acetate filter paper (both VWR). Samples were adjusted to pH 3.0 with hydrochloric acid, 2.5 mL of Na_4EDTA (100 g/L) was added, samples were spiked with 50 µL of isotopically labelled internal standards (ILISs at 100 ng) and stored for a maximum of 48 h at 4ºC in the dark prior to SPE. ILISs used in this study included ceftazidime-d_5, meropenem-d_6, ciprofloxacin-d_8, lincomycin-d_3, clindamycin-d_3, azithromycin-d_3, clarithromycin-d_3, erythromycin-d_6, sulfamethazine-d_4, sulfamethoxazole-d_4, trimethoprim-d_3, tetracycline-d_6, and chloramphenicol-d_5[194–196].

The SPE cartridges (chromabond HR-X cartridges 6 mL, 500 mg, Macherey-Nagel) were preconditioned with 5 mL of methanol, followed by 5 mL of acidified Milli Q water (pH 3) at a flow rate of 3 mL/min. Subsequently, 500 mL of spiked and acidified surface water samples were loaded onto the cartridges at a flow rate of 5 mL/min. After all water samples were passed through SPE cartridges, the cartridges were rinsed with 5 mL of acidified Milli-Q water (pH 3.0) to remove weakly bound impurities and Na_4EDTA. Then the SPE cartridges were dried for 30 min under vacuum. Elution of the target analytes from the SPE cartridges were performed using 5 mL of methanol at a flow rate of 1 mL/min. The resulting extracts containing the target analytes were dried under a gentle stream of nitrogen at 35 °C. The dried extracts were finally dissolved again with 1 mL of a mixture of methanol and Milli-Q water (50:50, v/v). The final aliquots were transferred into 2 mL amber vials and stored at -20 °C until UHPLC-MS/MS analyses.

Quantification of target antibiotics in water samples was performed using 13 ILISs, which corrects for losses during sample preparation and matrix effects during HPLC-MS/MS. The relative SPE recoveries for target antibiotics in the water samples varied from 85.6 to 117% (Appendix Table A-5). Method quantification limits (MQL) of target antibiotics ranged from 0.1 to 50 ng/L, depending on the compound (Appendix Table A-5). River antibiotic concentrations were compared to PNECs[107].

3.2.6 Antibiotic resistant gene quantification

River samples were analysed using HT-qPCR for 283 ARGs (36 aminoglycosides, 52 β-lactams, nine FCA (fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs), 46 MLSB (macrolide-lincosamide-streptogramin B ARGs), 51 non-specific efflux pumps, seven sulfonamides, 39 tetracycline, 32 vancomycin, 11 others), 12 MGEs (eight transposases, four integrases) and one 16S rRNA gene (Appendix Table A-6). For this analysis, we define the sum of transposase genes plus integron genes as MGEs, although we recognise that this is only an estimate based on the limited number of genes we quantified.

The water samples were filtered onto 0.22 µm cellulose-nitrate filters (Sartorius) to extract DNA with the FastDNA SPIN kit for soil (MP Biomedicals). Filtration volume varied depending on the sampling point (3 technical replicates of 80-250 mL each) with more water filtered for upstream location S1 to collect sufficient DNA. The product DNA was cleaned with the QIAquick Nucleotide Removal Kit (Qiagen). DNA quality and quantity were measured with the NanoDrop and Qubit dsDNA HS assay (both Thermo Fisher Scientific), respectively. DNA absorbance ratios were 260/280 > 1.8 and 260/230 > 1.5. Replicate samples were pooled in equal DNA aliquots to reach 2 µg DNA and freeze-dried prior to further analysis. Between analysis steps, DNA was stored at -20 ºC.

HT-qPCR was performed with SmartChip Real-Time PCR (Wafergen) as previously described[197,198]. Amplification efficiency always was between 90% and 110% and detection only was confirmed when all three technical replicates were positive. Relative copy number of ARGs and MGEs were calculated and transformed to absolute copy numbers by multiplying with 16S rRNA concentration for each sample. ARG and MGE cell concentrations were estimated by dividing the 16S rRNA concentration by 4.1, the estimated average 16S rRNA gene copy number per bacterium[199].

3.2.7 Statistical analysis and data visualisation

Data from this study can be accessed through the Center for Open Science repository, OSF (Ott, Amelie. 2021. 'Monitoring and Modeling of Antibiotic Resistance in Southeast Asian Rivers. https://osf.io/gcpky/?view_only=90e614c2c6b64483aa503694af113789). Statistical analysis was performed in R (v 4.0.5)[200]. Graphics were created using R package *ggplot2* (v 3.3.3)[201] and finalised in Inkscape (v 1.0.2)[202]. Box-plot elements are defined as centre line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range) and points (outliers).

The Skudai catchment map was composed in ArcGIS (v 10.6.1)[203]. The river catchment was extracted through digital elevation model (DEM) slope analysis[204]. Mass loading data were calculated by multiplying concentration data with the corresponding measured discharge (m^3/s) for each sampling site and trip.

The substitution method R2D was used to allow statistical analysis of left-censored data (e.g. antibiotic and coliform data)[205]. For this, measurements under detection limit were substituted with √2/2 times the limit of detection, but only if less than 40% of all data points were under the detection limit[205]. Parameters with higher rates of 'non-detects' were excluded from statistical analyses. Averages are reported as the mean with ± standard deviation (based on three or four biological replicates) throughout the chapter.

Equation 3-1. Cohen's D effect size for paired t-test[206].

$$Cohen's\ D\ effect\ size\ for\ paired\ t-test = \frac{Sample\ mean\ difference}{Sample\ standard\ deviation}$$

Equation 3-2. Cohen's D effect size for Welch's t-test[206].

$$Cohen's\ D\ effect\ size\ for\ Welch's\ t-test = \frac{mean_A - mean_B}{\sqrt{(variance_A + variance_B)/2}}$$

Statistical significance testing employed P values and calculated Cohen's D effect sizes[207,208] to assess spatial and seasonal differences in water quality and AR parameters. Large statistically significant spatial or seasonal effects were defined for values of Cohen's D < -0.8 or > 0.8 and P < 0.05[207]. Effect sizes can be negative or positive, depending on which mean is greater. Wet vs. dry season data were

compared with paired t-tests and corresponding Cohen's Ds (Equation 3-1). Up- (S1) vs. downstream (S8) data were compared with Welch's t-tests[209] and corresponding Cohen's Ds (Equation 3-2).

Benjamini-Hochberg P-adjustment was applied to correct for multiple testing[210]. Cohen's Ds were calculated with the 'cohen.d' function in the R package *effsize* (v 0.8.1)[211]. Normality was assessed with the Shapiro-Wilk test. Where necessary, parameters were transformed using the Box-Cox transformation[212], as implemented in the 'boxcox' function in the R package *MASS* (v 7.3-53.1)[213] (Appendix Table A-7). To visualise spatial and seasonal effects, Cohen's D effect sizes were plotted against P values for each parameter in volcano plots, using the R package *EnhancedVolcano* (v 1.8.0)[214]. To analyse water quality and AR parameter associations, Spearman's correlations were calculated with Benjamini-Hochberg multiple testing correction, using R packages *psych* (v 2.1.3)[215] and *corrplot* (v 0.84)[216].

3.3   Results

3.3.1   *Water quality and microbiology*

Water quality conditions in the catchment were characterised by generally low DO, high COD and very high NH_3-N concentrations based on national Malaysian thresholds (Figure 3-3a-c, Appendix Table A-8, Appendix Table A-9, Appendix Table A-10). Water quality declined in the Skudai from upstream being 'clean'/'slightly polluted' (S1: 7.5 ± 0.5 DO mg/L, 0.05 ± 0.03 NH_3-N mg/L, 5.8 ± 4.8 COD mg/L) to downstream being 'slightly polluted'/'polluted' (S8: 1.3 ± 0.3 mg DO/L, 4.9 ± 2 NH_3-N mg/L, 25.3 ± 16 COD mg/L). Measurements for DO, COD and NH_3-N aligned well with the national 2018 DoE monitoring data (Figure 3-3a-c).

Total coliform and beta-lactam resistant coliform concentrations all increased from upstream S1 (($1.1 ± 0.5$) x 10^3 TC CFU/mL, ($1.5 ± 1.3$) x 10^2 ESBL coliform CFU/mL, ($3.1 ± 4.1$) x 10^1 CRB-2 CFU/mL) to downstream S8 (($4.1 ± 3.3$) x 10^4 TC CFU/mL, ($4 ± 4.1$) x 10^3 ESBL coliform CFU/mL, ($1.1 ± 0.2$) x 10^2 CRB-2 CFU/mL) in the Skudai river (Appendix Figure A-1, Appendix Table A-11). Across the catchment, we observed an approximately one $\log_{10}$ difference between TC > ESBL coliform > CRB-2 concentrations, meaning that ~ 10% of total coliform produced ESBL and ~ 1% of total coliform were resistant to 2 μg/mL meropenem. *E. coli* and ESBL *E. coli* concentrations increased from upstream S1 ($3.5 ± 2$) x 10^1 CFU/mL and (<0.5 − 2) x

10^1 CFU/mL, respectively to downstream S8 (2.8 ± 2.1) x 10^3 CFU/mL and (<0.1 – 5) x 10^2 CFU/mL, respectively (Appendix Table A-11).

Volumetric flowrate in the Skudai increased greatly from upstream (S1: 0.5 ± 0.3 m^3/s) to downstream (S8: 82.7 ± 30.7 m^3/s) with small variations observed across seasons (Appendix Table A-8). Mass loading data showed much greater transport of chemical and microbial pollutants along the Skudai river from the rural to urban locations. NH_3-N concentrations increased almost 100-fold from up- (S1) to downstream (S8), but increases were > 14,000-fold greater based on NH_3-N mass loading data (Appendix Table A-9, Appendix Table A-10). Similarly, TC, ESBL coliform, CRB-0.5 and CRB-2 concentrations increased from up- (S1) to downstream (S8) 10^0 – 10^1-fold while their mass loadings increased 10^2 – 10^3-fold (Appendix Table A-11, Appendix Table A-12).
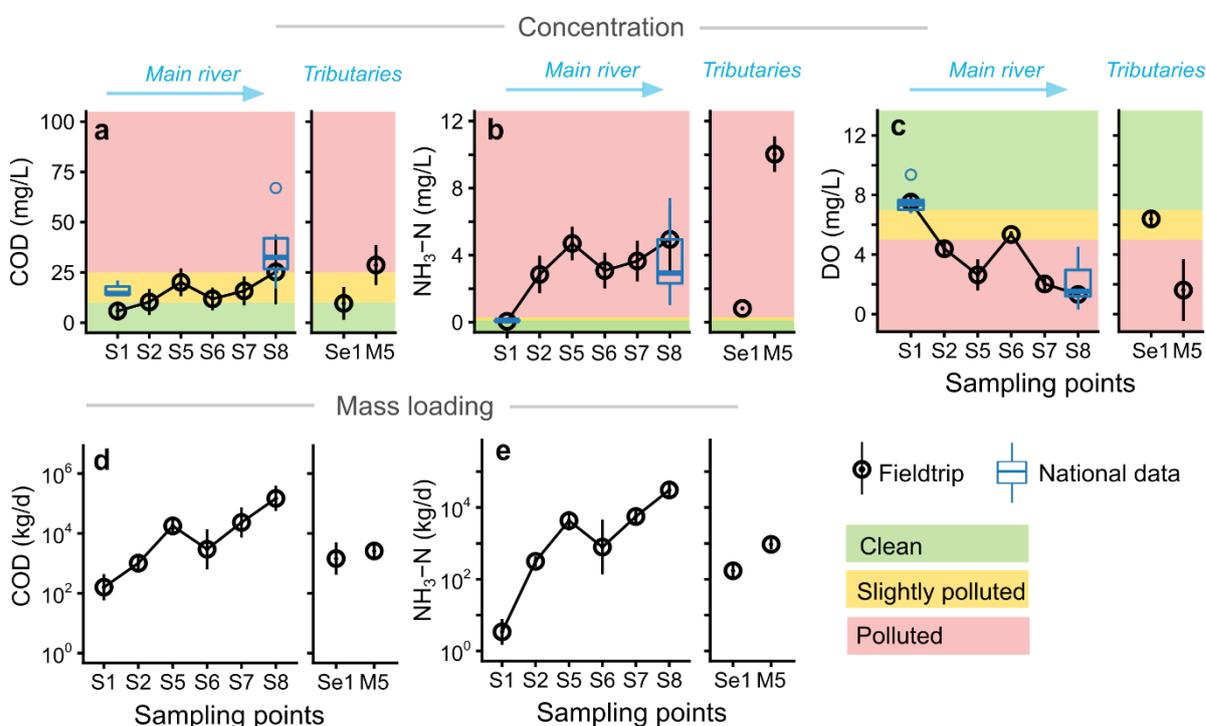
Figure 3-3. Chemical oxygen demand (COD, a and d), ammonia (NH_3-N, b and e) and dissolved oxygen (DO, c) concentrations (a-c) and mass loadings (d and e) in the river catchment. Data represented is based on four biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on three biological replicates for the tributaries (Se1, M5). Concentrations were compared to Malaysian water quality thresholds and Department of Environmental (DoE) monitoring data for S1 (DoE sampling point 3SI09) and S8 (DoE sampling point 3SI05). d: day.

Interestingly, water and microbial quality improved slightly mid-stream at S6 for most parameters in concentration and mass loading data (Figure 3-3, Appendix Figure A-1). Water and microbial quality concentrations were much poorer in the heavily urbanized Melana tributary (M5), both relative to the Skudai itself and the

predominantly rural Senai tributary (Se1) (Appendix Table A-8, Appendix Table A-9, Appendix Table A-11). As indicator of conditions, elevated CRB-2 and CRB-0.5 *E. coli* levels only were found in the Melana tributary across the catchment (Appendix Table A-11).

3.3.2 Antibiotic levels

Out of 22 antibiotics tested (Appendix Table A-4), eight antibiotics (meropenem, cefixime, ceftazidime, erythromycin, chlortetracycline, minocycline, oxytetracycline, tetracycline) were not detected in the Skudai catchment. Six antibiotics/antibiotic derivates (clindamycin, lincomycin, azithromycin, clarithromycin, dehydrated erythromycin, trimethoprim) were detected in all river samples. Highest concentrations were observed for amoxicillin (all samples 510 ± 906 ng/L; max 3336 ng/L at S2), sulfamethoxazole (all samples 181 ± 383 ng/L; max 1933 ng/L at S8) and ciprofloxacin (all samples 131 ± 162 ng/L; max 705 ng/L at M5) with maximum values always detected in dry season samples (Figure 3-4). Only amoxicillin and ciprofloxacin were detected above PNEC values[107] with all ciprofloxacin and 50% of amoxicillin measurements in the dry season exceeding the PNEC thresholds.

a) Detected antibiotics for all sampling trips and sampling points
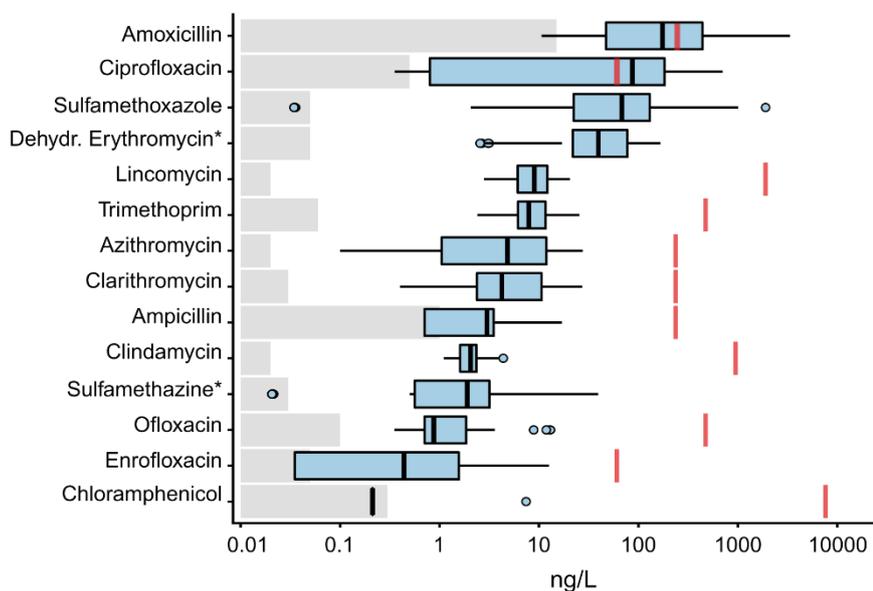
b) Differentiation by season
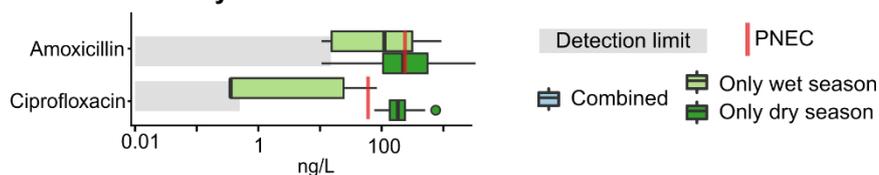
Figure 3-4. Antibiotic concentrations detected in the river catchment (n = 30) (a) with seasonal differentiation for amoxicillin and ciprofloxacin (b), compared to Predicted No Effect Concentrations (PNEC) (for * no PNEC defined)[107].

In more than 40% of the samples, ampicillin and chloramphenicol concentrations were under the detection limit. Consequently, only 14/16 detected antibiotics were summarized into 'total antibiotics' (Appendix Table A-9, Appendix Table A-10). Total antibiotics concentrations increased from up- (S1: 0.07 ± 0.05 mg/L) to downstream (S8: 1.27 ± 0.98 mg/L) and were higher in the dry than wet season.

3.3.3 Antibiotic resistant gene abundances

We detected 210 different ARGs (74% of assay) in the river catchment with 78 ARGs (28% of those assayed) shared between all river water samples (n = 30). All 12 MGEs assayed were detected in the catchment with nine MGEs (75% of assay) shared across all samples (n = 30).
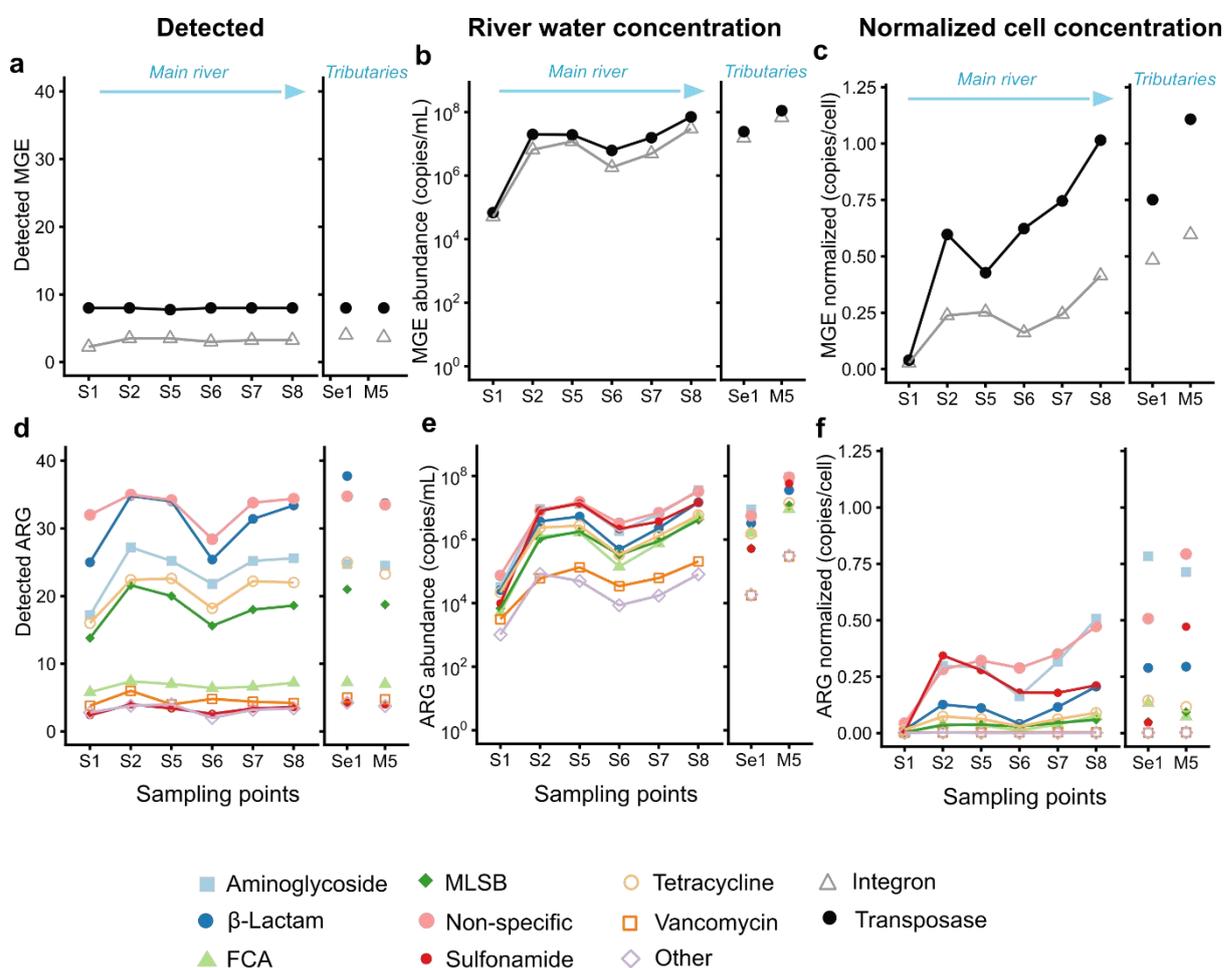
Figure 3-5. Antibiotic resistant gene (ARG) and mobile genetic element (MGE) detected (a, d), river water concentrations (b, e) and normalised cell concentration (c, f) measured with HT-qPCR per sampling point in the Skudai catchment. Mean data represented is based on four biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on three biological replicates for the tributaries (Se1, M5). For standard deviations, see SI Tables S14 to S17. FCA: fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs. MLSB: macrolide-lincosamide-streptogramin B ARGs.

ARG and MGE levels increased from up- to downstream in the Skudai river (Appendix Table A-13 to Appendix Table A-17), except for lower levels found mid-

stream at S6 (Figure 3-5), which parallels improved water quality conditions based on other measured parameters.

The number of detected ARGs increased from 119 ± 14 at S1 to 150 ± 8 at S8 (Appendix Table A-13). The increase in ARG diversity was most apparent at the headwaters of the river. The most upstream site, rural S1, and the next site, semi-urban sampling point S2, shared a core resistome of 157 ARGs and MGEs (Figure 3-6). However, only 5 unique ARGs were detected at S1, whereas 41 unique ARGs (such as *bla*CTX-M and *vanA*) and 1 MGE were detected at S2.

On a wider scale, ARG and MGE concentrations increased more than 10^2-fold from up- to downstream (S8: $1.2 ± 0.9 \times 10^8$ ARG copies/mL and $1.1 ± 0.9 \times 10^8$ MGE copies/mL), while ARG and MGE mass loadings increased more than 10^5-fold from up- to downstream (S8: $8.6 ± 7.2 \times 10^{20}$ ARG copies/d and $8.1 ± 7.3 \times 10^{20}$ MGE copies/d; SI Table S13). The normalised copy number of ARGs and MGEs per cell increased from 0.1 ± 0.1 and 0.1 ± 0 upstream to 1.7 ± 0.6 and 1.6 ± 0.6 downstream, respectively. Detected numbers, concentrations, and normalised copy numbers for ARGs and MGEs were higher in both tributaries (M5 and Se1) than downstream in the Skudai river (S8) (Appendix Table A-13 to Appendix Table A-17).

	Aminoglycoside	β-Lactam	FCA	MLSB	Non-specific	Sulfonamide	Tetracycline	Vancomycin	Other	MGE
S1		bla_L1 bla1					tetPB_01 tetPB_04	vanTC_01		
S2	aac_6_l1 aacA_aphD aacC1 aph_2_ld_02 spcN_02 str	ampC/blaDHA blaCTX_M_05 blaOKP blaPAO blaROB blaTLA blaZ cepA cfiA penA	cfr qnrA	ermT_01 lnuC mdtA msrA_01 msrC_01 vatE_01 vatE_02 vgb_01	marR_01 mtrD_03	folA sul1	tetB_01 tetC_01 tetC_02 tetD_01 tetH tetT tetR_03	vanA vanHD	fabK nimE	intI2

Venn diagram: S1 = 5, shared = 157, S2 = 42.

Figure 3-6. Differences in antibiotic resistant gene (ARG) and mobile genetic element (MGE) detection between the most upstream rural sampling point S1 and the next, semi-urban sampling point S2 on the Skudai. The Venn diagram indicates the number of ARGs and MGEs only detected at S1 (5), the number of shared ARGs and MGEs between S1 and S2 (157) and the number of ARGs and MGEs only detected at S2 (42). Data based on four biological replicates. FCA: fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs. MLSB: macrolide-lincosamide-streptogramin B ARGs.

3.3.4 Assessing seasonal and spatial effects

Dimensionless Cohen's D effect sizes were calculated to inform the magnitude of spatial (up- vs. downstream) and seasonal effects on water quality and AR levels. Reporting standardised effect sizes in concert with P values allows one to better compare study findings within and across studies[217]. This is particularly important for LMIC settings where limited data availability hinders the identification of environmental AR 'hotspots'.
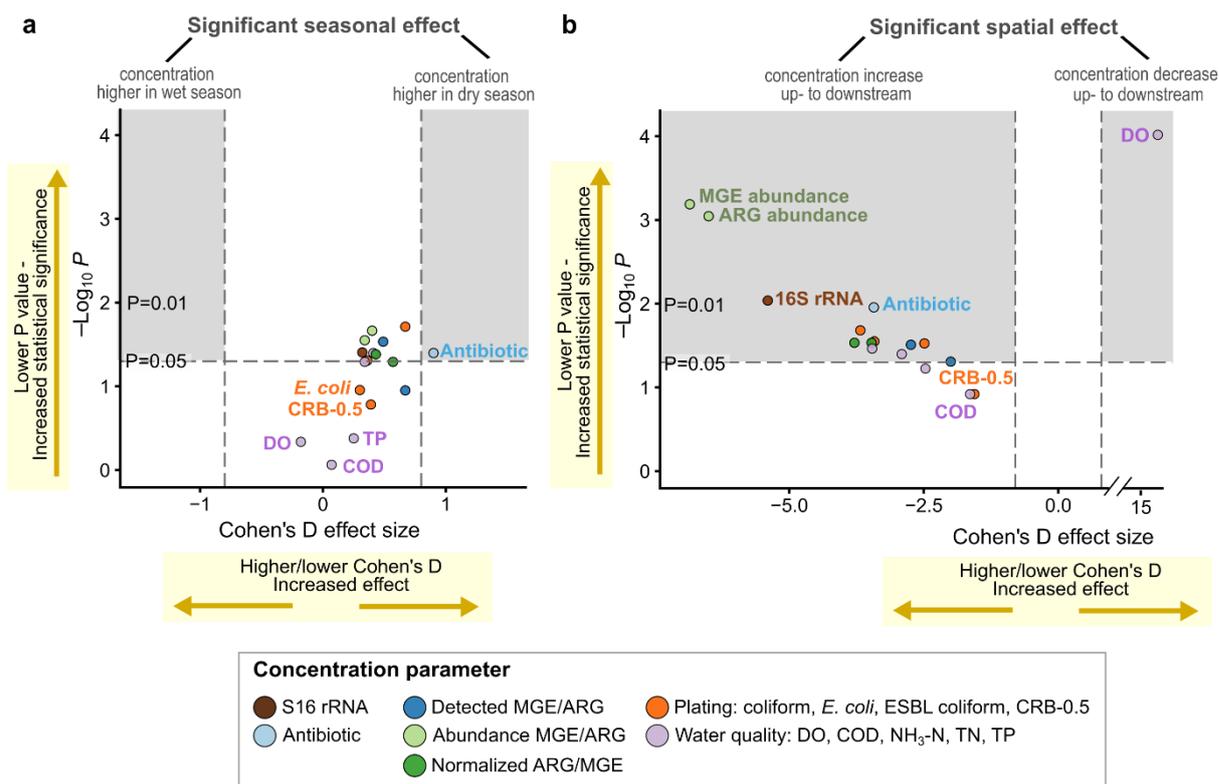
Figure 3-7. Comparing the effect of seasonality (a) and spatial variation between up- (S1) and downstream (S8) (b) for concentration parameters, based on statistical significance and Cohen's D effect size. Statistical comparisons performed with the paired t-test or Welch's t-test with Benjamini-Hochberg multiple testing correction. A high $-Log_{10}P$ value indicates high statistical significance with $-Log_{10}P(2) = P (0.01)$ and $-Log_{10}P(3) = P(0.001)$. A Cohen's D effect size over 0.8 or under -0.8 indicates a large seasonal or spatial effect on the parameter. Only selected parameters are labeled, for more detail see SI Tables S18, S19. ARG: antibiotic resistant genes. COD: chemical oxygen demand. CRB-0.5: carbapenem resistant bacteria selected for with 0.5 µg/mL meropenem. ESBL: extended-spectrum β-lactamase. DO: dissolved oxygen. MGE: mobile genetic elements. NH$_3$-N: ammonia. TC: total coliform. TN: total nitrogen. TP: total phosphorus.

Seasonality only significantly affected observed total antibiotic concentrations (paired t-test with $P < 0.05$ and large Cohen's D effect size > 0.8, Appendix Table A-18). For all other parameters, season did not have any significant effects on concentration and mass loading data (Figure 3-7a, Appendix Table A-18). Conversely, spatial effects (up- vs. downstream) were significantly greater for all parameters, and more apparent in mass loading data (Cohen's D range -13.9 for S16 rRNA to -6.8 for ESBL

coliform, Appendix Table A-19) versus concentration data (Cohen's D range MGE -6.85 to -1.6 for CRB-0.5, Appendix Table A-19). For concentrations, the largest Cohen's D effect sizes were observed for DO (Cohen's D 15.6), MGE and ARG river water concentrations (Cohen's D - 6.5 for ARG and -6.85 for MGE) (Figure 3-7b).

3.3.5 *Defining a surrogate marker for antibiotic resistance*

Spearman correlation analysis was performed between all monitored parameters to identify possible 'easy-to-measure' surrogates that associated with elevated AR in this catchment (Figure 3-8). For this, we focussed on correlations between AR indicators (ESBL coliform, ESBL *E. coli*, total antibiotics, total ARGs, total MGEs, int1) and physico-chemical water quality parameters (temperature, pH, DO, conductivity, NH_3-N, COD, TN, TP). These standard water quality parameters also are included in the Malaysian river water quality monitoring program[92].
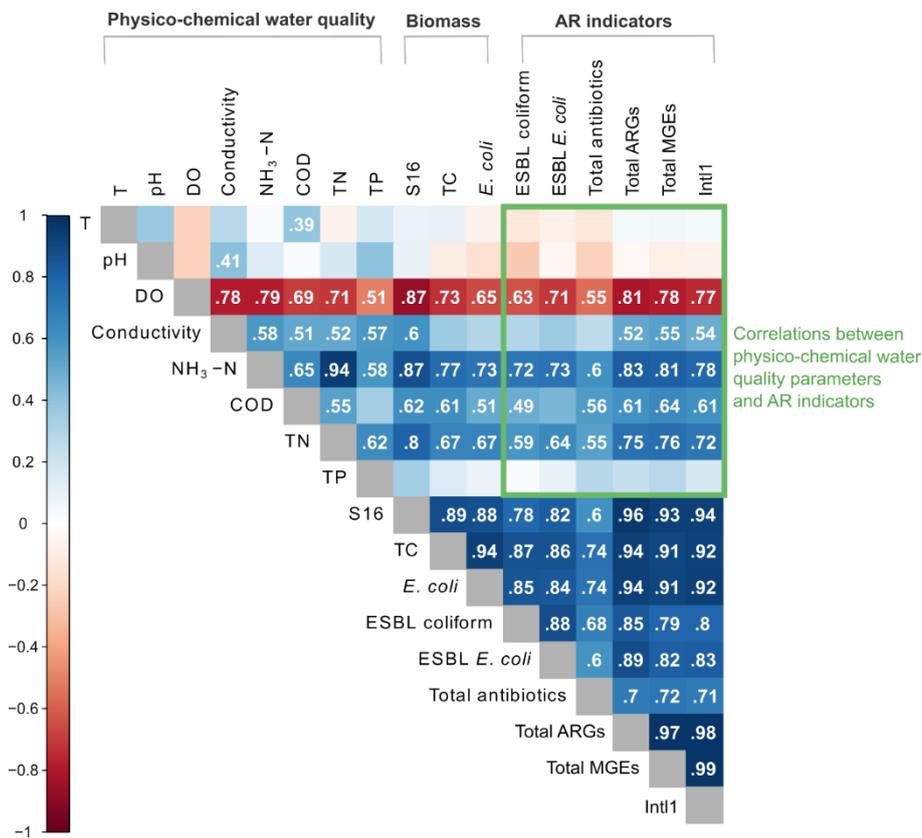
Figure 3-8. Spearman correlations between selected physico-chemical water quality, biomass and antibiotic resistance (AR) concentrations for the river catchment (n=30). Correlation values only shown for P < 0.05 with P values corrected for multiple testing with the Benjamini-Hochberg approach. ARG: antibiotic resistant genes. COD: chemical oxygen demand. DO: dissolved oxygen. MGE: mobile genetic elements. NH_3-N: ammonia. S16: S16 rRNA gene. T: temperature. TC: total coliform. TN: total nitrogen. TP: total phosphorus.

Out of the physico-chemical water quality parameters, DO and NH_3-N correlated strongest with total ARGs, the sum of all ARG copy number concentrations in river

water (Spearman's ρ 0.81 and 0.83 with P < 0.05, respectively). Within the AR indicators, total ARGs correlated strongly with *intI1* (Spearman's ρ 0.98, P < 0.05) but less so with total antibiotics (Spearman's ρ 0.7, P < 0.05). When comparing correlations between total ARGs and each ARG class with total versus individual antibiotic concentrations, the strongest correlations always were between total ARGs and total antibiotics (Appendix Table A-20). This was also true when comparing amoxicillin and ciprofloxacin with their ARG class, suggesting specific selection by individual antibiotics is not evident, even the detected antibiotics near their PNEC levels.

3.4   Discussion

3.4.1   Comprehensive environmental antibiotic resistance monitoring

Discharge and mass loadings are rarely estimated in environmental AR monitoring studies. However, we show that both concentration and load provide valuable complementary information to understand the processes occurring in a river catchment. In the Skudai, river health improved mid-stream at the semi-urban sampling point S6 despite worse water quality conditions further up- and downstream. Considering the combination of lower pollutant concentrations and mass loadings, this was likely caused by a combination of reduced wastewater entering the river in this more agricultural reach (in comparison to more urbanized reaches up- and downstream, Figure 3-1) while simultaneously, groundwater and/or cleaner tributaries (e.g. Senai) continued to dilute the river water with pollutants degrading and/or settling to the sediment[70]. More accurate methods exist to estimate flow than the applied float method; e.g., Acoustic Doppler Current Profilers. However, the easy and cost-effective application makes the float method particularly suitable for countries with limited resources[218].

Accounting for volumetric flow is particularly important for countries with dry and wet seasons. Comparing total antibiotic concentrations and mass loadings, we demonstrate that while antibiotic releases into the catchment likely do not consequentially vary across seasons for this catchment, reduced rainfall during the dry season resulted in increased river antibiotic concentrations and slightly increased exposures. Seasonality is expected to have a much larger effect on water quality/AR parameters in other SE Asian regions with more pronounced dry/wet seasons than here for southern peninsular Malaysia.

The highest antibiotic concentrations in the catchment based on both, maximum and mean, were recorded for amoxicillin, ciprofloxacin, and sulfamethoxazole. A 2014 study found amoxicillin to be the most prescribed antibiotic in Malaysia[219]. Mean amoxicillin river concentrations were higher than previously recorded for European treated wastewater treatment plant effluents and surface waters[220–222]. Sulfamethoxazole and ciprofloxacin concentrations were higher than previously recorded for Malaysian surface waters[223,224], but comparable to some other East and SE Asian surface water studies[79]. There is limited knowledge on which environmental antibiotic concentrations select for resistant bacteria[106,107]. In this study, only amoxicillin and ciprofloxacin exceeded the PNEC thresholds[107], specifically during the dry season.

Comparing the Skudai ARG concentrations to other ARG HT-qPCR studies based on the same primer sets and analytical methods (Appendix Figure A-2), we found the upstream ARG levels to be comparable to previous findings in upstream Chinese river reaches ($10^5 – 10^6$ ARG copies/mL[135,225]). Downstream Skudai ARG concentrations (~ 10^8 ARG copies/mL, this study and[135]) were similar to wastewater treatment effluent ARG concentrations ($10^7 – 10^9$ ARG copies/mL) recorded in Spain and China but lower than influent ARG concentrations (10^9-10^{10} ARG copies/mL) from the same studies[70,226]. The detected number of ARGs upstream in the Skudai was higher than in any other of the reported upstream river water, upstream river sediment, lake, or soil samples. The number of detected ARGs downstream in this study also was the highest across all cited studies.

Movement from the rural (S1) to semi-urban (S2) locale added over 40 additional genes, many associated with faecal matter and possible multidrug resistance, such as *bla*CTX-M and *vanA*. *Bla*CTX-M encodes for high resistance to β-lactam antibiotics[227]. *VanA* is a plasmid borne gene which confers high resistance to vancomycin and is most commonly associated with *E. faecium* and *E. faecalis*[228]. The S1 to S2 reach has limited wastewater treatment which likely introduced these ARGs into the river, suggesting limited local wastewater treatment may be the dominant source of AR genes in this part of the river, which also was seen in an AR estuary study in southern Malaysian[97].

3.4.2 *Reporting standardised effect sizes*

Effect sizes are commonly applied in bioinformatics, medical drug trials, and meta-analysis[229]. However, to the best of our knowledge, this is the first work to apply the

principle of standardised effect sizes to AR/river water quality monitoring. While unstandardised effect size statistics such as mean differences are important, additional reporting of standardised, dimensionless effect sizes such as Cohen's D effect size allows one to more easily compare seasonal and spatial effects on various parameters[217]. This is particularly crucial for understanding and comparing results from environmental AR monitoring studies where analysis costs are high, resulting in little available data, mostly existing for HICs[27]. Routine reporting of effect sizes will encourage researchers to view their results in the context of previous studies and facilitate the incorporation of results into future meta-analysis[217]. We support Nakawaga and Cuthill (2007) in their encouragement to report effect size statistics and their confidence intervals in all biological journals.

Using volcano plots, we provide an easy way to visualise seasonal and spatial effects together with P values to compare different water quality and AR parameters. For concentration data, we observed the largest statistically significant spatial effects (up vs. downstream) for ARG, MGE and DO concentrations. Spatial effects were even larger for all parameters based on their mass loadings than concentrations. This is not surprising when considering that the Skudai river increases in depth and particularly, width from 5 m at the most upstream sampling point to 75 m at the most downstream sampling point. For this study, we applied the Cohen's D threshold of over 0.8 or under -0.8 to define a large effect size as originally proposed by Cohen for behavioural studies[207]. However, depending on the study design, this threshold can be adapted.

3.4.3 Surrogate marker for predicting antibiotic resistance.

River water in more urban areas of the Skudai catchment were characterised by higher NH_3-N and lower DO levels, both indicators of faecal pollution[230]. When sewage enters a river, the organic matter and nitrogen containing components are oxidized, decreasing DO drastically[231]. This process has been known for many years and is mathematically described by DO sag curves[232]. Based on our data and local water quality thresholds, the Skudai catchment is classified in the slightly polluted to polluted range, which aligns with the Malaysian DoE classification[94]. Our DO and NH_3-N data aligns well with the long-term national Malaysian dataset (Figure 3-3b,c), suggesting our correlations between these parameters and AR markers might be used to extend existing Malaysian datasets to AR prediction, in theory suggesting places of potentially elevated AR using modelling where no current AR data exists.

For the Skudai catchment, we found DO and NH_3-N exhibit the strongest correlations with high total ARG concentrations. This does not mean that low DO or high NH_3-N cause high total ARG concentrations (or vice versa), although recent work shows that HGT frequency can be much higher under low oxygen conditions[233], suggesting lower DO may increase local ARG transmission. In this catchment, lower DO and higher NH_3-N are likely associated with insufficiently treated sewage entering the river, which hints it is also a major route for ARGs entering the river.

Given the above, DO is particularly well-suited as a surrogate for AR as it can easily be measured with a hand-held probe, relative differences often mirror sewage inputs, and DO potentially impacts in situ HGT frequency. DO is also one of the most commonly modelled indicators of stream, river and lake health with a vast array of models available[159]. Consequently, we propose that for this catchment, DO concentrations are a useful surrogate to understand previous AR levels and model future AR levels. Future work should evaluate the applicability of this surrogate for other catchments in Malaysia and SE Asia. However, for such surrogates to have greatest value, they should be coupled with other predictive AR approaches that do not heavily rely on directly monitored data, such as genomic and other modelling tools for AR bacteria[234–236].

Interestingly, within the AR indicators, total antibiotic concentrations exhibited the lowest correlations with other AR parameters. The weaker correlation of total antibiotics with the other AR parameters might be due to the fact that many antibiotics quickly degrade in the environment while some ARGs and ARBs persist for longer[237]. However, even in the Skudai river that has relatively high antibiotic levels, any selective effect of antibiotics is probably minor (Figure 3-4) compared with the greater load of ARGs entering the river through less treated wastewater (Figure 3-5). This is best exemplified by the many 'new' ARGs entering the river between S1 and S2 (Figure 3-6), which dwarfs any effect of antibiotics themselves. This does not mean low levels of antibiotics are incapable of influencing ARG selection in aquatic systems[238], but data here suggest untreated sewage inputs have much greater immediate impact on in situ AR than antibiotic releases in a catchment like the Skudai.

Taken together, this work shows that simple water quality markers, like DO and NH_3-N, can be valuable surrogates for local stakeholders to identify AR hotspots in rivers

and propose social and/or engineering interventions. This does not mean that they are universally applicable, such as near major non-sewage organic waste inputs. However, DO and NH_3-N clearly mirror sewage, which often dominates ARG and AR bacteria inputs, especially in LMIC rivers. DO and NH_3-N also are inexpensive to measure and already exist in current monitoring programmes. Therefore, we propose DO and NH_3-N as the 'first point of call' surrogates for AR in rivers. They clearly can be coupled with parameters such as ESBL *E. coli* for environmental AR monitoring, which the WHO is already using to monitor AR across environments (Tricycle project[115]). However, DO and NH_3-N are more amenable to water quality modelling, which might ultimately be the best and most affordable way of identifying AR 'hot spots' in places with limited existing data.

Chapter 4. Predicting antibiotic resistance with national water quality monitoring data and watershed modelling

4.1 Introduction

Many LMICs maintain longstanding national water quality monitoring programs[92]. Sustainable Development Goal (SDG) Indicator 6.3.2 recommends oxygen, salinity, nitrogen, phosphorus and acidification as core water quality parameters, but often LMICs measure additional parameters[239]. For example, the Water Environment Partnership in Asia (WEPA) states Indonesia and Malaysia monitoring over 40 and 70 parameters, respectively (Figure 4-1)[92].
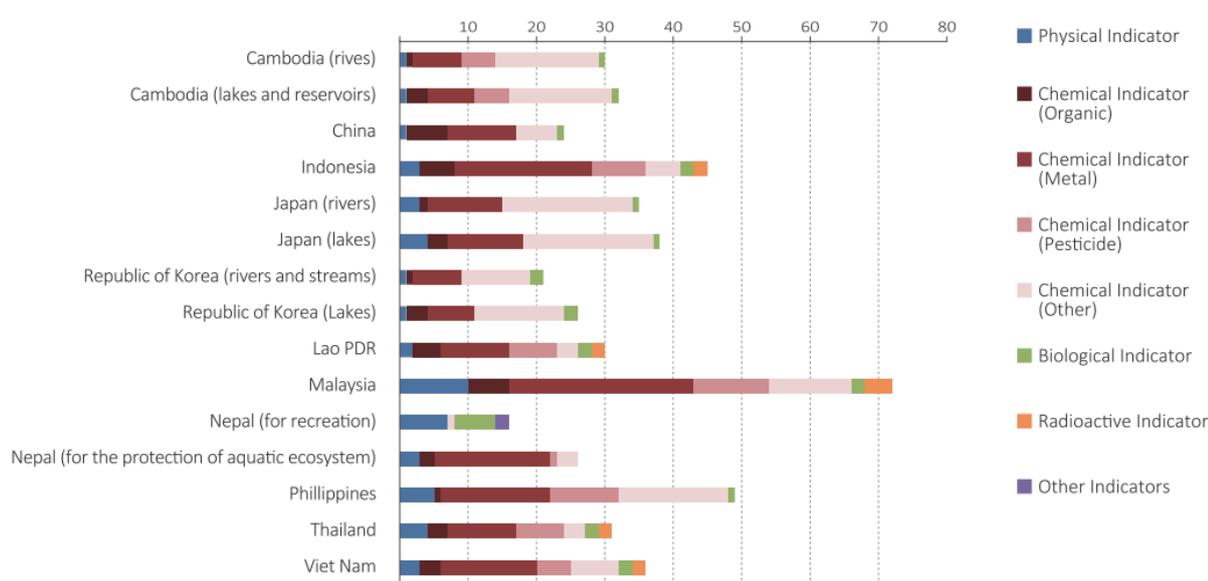
Figure 4-1. Number of indicators included in national surface water quality monitoring programs for selected Asian countries. All countries except Lao PDR and Nepal conduct regular national water quality monitoring of public water bodies. In Lao PDR, water quality is monitored on an ad-hoc basis as necessary. In Nepal, water quality is monitored by different ministries and agencies, but no systematic surface water quality monitoring is performed[92].

The number of water quality monitoring stations differs depending on the country and data source consolidated. The Global Freshwater Quality Database GemStat, part of the UN Environment Programme (UNEP), states average monitoring station densities of 0.3, 0.02 and 0.08 stations per 10,000 km^2 for Latin America, Africa and Asia, respectively[240,241]. As comparison, the EU Water Framework Directive recommends a station density for water quality surveillance of 4 per 10,000 km^2 [242]. However, often more national monitoring stations exist than are reported for in GemStat. For Malaysia, seven river water quality stations are included in GemStat while the

Malaysian Department of Environment (DoE) states 904 stations in 477 rivers, covering 146 river basins (Figure 4-2a,c)[92,240].

The well-established Malaysian river water quality monitoring program started in 1978 to record baselines and detect water quality changes[93]. Water samples are collected at regular intervals from designated stations for in-situ and laboratory analysis to determine physico-chemical and biological parameters (Figure 4-2b)[93]. Typically, manual water quality monitoring is performed with a few selected continuous water quality monitoring stations at sensitive locations, including upstream of water abstraction points[93].

a) Peninsular Malaysia

b) River water quality parameters
DO, temperature, conductivity, salinity, turbidity, pH, BOD, COD, solids, NH_3-N, NO_3, PO_4, heavy metals, anions, Coliform and *E. coli*

c) Sarawak and Sabah, Malayisa

DoE river water quality stations (2006)
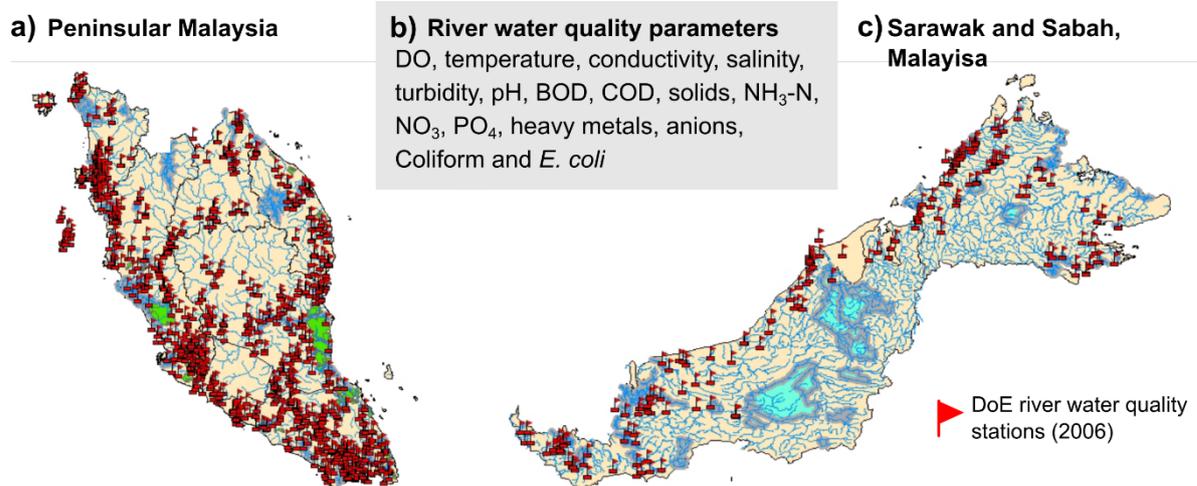
Figure 4-2. National Malaysian Department of Environment (DoE) river water quality stations for peninsular Malaysia (a) and Sarawak, Sabah (c) with river water quality parameters stated (b)[93].

Consequently, a vast amount of data are available not only in Malaysia but also other SE Asian countries to identify and manage water pollution[92]. In the absence of environmental AR monitoring and modelling in LMICs, such national water quality datasets i) can help pinpoint augmented AR levels through identifying pollution hotspots via easy-to-measure surrogates captured in the datasets (Chapter 3), and ii) are crucial for developing watershed models to predict AR levels for different pollution scenarios[243].

The watershed model HSPF is well suited to incorporate an AR element for several reasons: i) HSPF is often used to investigate the effects of land-use changes and point or diffuse source treatment alternatives, which are key research questions for environmental AR[162]; ii) HSPF is composed of a set of modules to simulate hydrology and water quality in the watershed and water body, allowing sufficient flexibility to model AR[244,245]; and, iii) HSPF allows to simulate key processes, covering soil

contaminant runoff processes together with in-stream hydraulics, water temperature, sediment transport, nutrients and sediment-chemical interactions, capturing important river processes to describe environmental AR (Appendix Table B-1)[159].

The previous chapter demonstrated that lower DO and higher NH₃-N are likely associated with insufficiently treated sewage entering the Skudai river, which is probably also the major route for ARGs entering the river. This link is now further explored with the national Malaysian DoE dataset for this catchment. In the absence of AR parameters for this (and any other LMIC) national dataset, the relationships and predictive power between abiotic water quality parameters (such as NH₃-N and DO) and biotic parameters (such as coliform and *E. coli*) are assessed. In addition, we propose to amend existing abiotic water quality components in HSPF (such as DO) to model AR fate and exposure in surface waters.

4.2 Material and methods

4.2.1 National water quality monitoring data

Water quality data for the Skudai catchment was available for 2002 to 2006 and 2010 to 2018. Nine DoE sampling points were included in the analysis, seven on the Skudai river (3SI18, 3SI09, 3SI10, 3SI13, 3SI07, 3SI06, 3SI05) and two on the Melana river (3SI15, 3SI16) (Appendix Figure B-1). Out of these, three DoE sampling points aligned with sampling points from the previous study (Chapter 3): S1 (3SI09), S7 (3SI06) and S8 (3SI05). From 2002 to 2016, water quality measurements were performed every month, but reduced to every two months in 2017 and 2018 with less regular monitoring at the most upstream Skudai (3SI18) and Melana (3SI15) sampling locations (Appendix Table B-2). Consequently, these two sampling points were excluded from annual overview trend analysis to not skew the results.

Equation 4-1. Malaysian water quality index (WQI) calculation based on sub-indices (SI). For subindex equations, see [246].

$$WQI = 0.22 \times SI_{DO} + 0.19 \times SI_{BOD} + 0.16 \times SI_{COD} + 0.15 \times SI_{AN} + 0.16 \times SI_{SS} + 0.12 \times SI_{pH}$$

SI_{DO} Subindex dissolved oxygen
SI_{BOD}:   Subindex biochemical oxygen demand
SI_{COD}:   Subindex chemical oxygen demand
SI_{AN} Subindex ammonium nitrogen
SI_{SS} Subindex suspended solids
SI_{pH} Subindex pH

Following DoE parameters were included into the analysis: DO (mg/L), temperature (ºC), conductivity (us), salinity (ppt), COD (mg/L), NH₃-N (mg/L), NO₃ (mg/L), PO₄

(mg/L), coliform (CFU/100 mL) and *E. coli* (CFU/100 mL). Following the Malaysian DoE approach, the WQI was calculated as described in Equation 4-1[246]. The WQI classifies rivers into polluted (0-59), slightly polluted (60-80) or clean (81-100)[246]. Details on DoE water quality monitoring methodologies such as test kits, agars, detection limits, replicates and controls were not available.

4.2.2 Statistical analysis and data visualisation

Statistical analysis was performed in R (v 4.0.5)[200]. Graphics were created using R package *ggplot2* (v 3.3.3)[201] and finalised in Inkscape (v 1.0.2)[202]. Box-plot elements are defined as centre line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range) and points (outliers). DoE water quality data for 2018 was analysed with Spearman's correlations with $P < 0.05$ (Benjamini-Hochberg multiple testing corrected), using R packages *psych* (v 2.1.3)[215] and *corrplot* (v 0.84)[216]. Linear regression modelling assessed the relationship between NH_3-N and $\log_{10}$ coliform concentrations. Model residuals were analysed for error mean, normality (histogram and Q-Q plot) and independence (Durbin-Watson test in R package *lmtest* (v. 0.9-38))[247]. The predictive power of the linear regression model was tested by comparing predicted 2017 coliform data (based on NH_3-N) with measured 2017 coliform data, calculating the Pearson correlation coefficient.

4.2.3 HSPF Skudai model

The watershed model HSPF was chosen to be assessed for use in providing information to detect AR hot spots. A previously developed HSPF Skudai baseline model[166,177,179,248–250] was run to model streamflow and DO from 2002 to 2015, covering the original calibration and validation periods used in split-sample testing of model performance. R^2 and Nash–Sutcliffe model efficiency coefficient (NSE) are reported to assess model performance. HSPF is a semi-distributed model (Table 2-3). Consequently, the Skudai catchment was divided into 33 hydrologic response units (HRUs)[179]. Each HRU is assumed to have a similar hydrologic response based on land use/cover, soil, slope, and land management practices[245]. The various hydrologic processes in HSPF are mathematically represented as flows and storages[161]. In each HRU, the soil layer is vertically divided into three layers of storage called upper-zone, lower-zone, and active groundwater storage (Figure 4-3). The moisture conditions in these three storages impact the water flow and evapotranspiration in each HRU[162]. Each HRU has three types of flow components contributing to the streamflow, called surface flow, interflow and groundwater flow,

each of which is associated with one of the three soil layers[162]. The streamflow is routed downstream from reach to reach using the kinematic wave model, an approach widely used in hydrological models to describe channel and overland flow[244,251]. Land segments in HSPF are either pervious or impervious, allowing water to infiltrate or not, respectively[161].

Figure 4-3. Hydrologic processes simulated in HSPF pervious land segment (PERLND) module (amended from [252]). ET: evapotranspiration.

Figure 4-4. Conceptual model of dissolved oxygen (DO) processes represented in HSPF. BOD: biochemical oxygen demand[161,253]. For details on the code, see HSPF user manual[161].

HSPF simulates DO through separate subroutines as a function of water temperature, downstream transport of DO and BOD, settling of BOD material, SOD,

resuspension of benthic BOD, reaeration, and oxygen depletion caused by the decay of oxygen-demanding materials (Figure 4-4)[159,244]. DO sources and sinks are simulated as first-order reactions, except for SOD which is assigned by the user to each reach[159]. HSPF accounts for the absence or deficiency of oxygen by reducing oxidation reactions at low DO concentrations[159].

4.3 Results

4.3.1 National water quality data

While WQI, DO and NH_3-N concentrations varied annually for the Skudai catchment, no clear improvement or decrease in river water quality was observable when comparing measurements for all DoE stations from 2002 to 2018 (Appendix Figure B-2a,c,e). Slight decreases in river water quality only became apparent when analysing the most downstream Skudai sampling point (3SI05 = S8), for example for NH_3-N concentrations (Appendix Figure B-2b).

The predictive power of easy-to-measure abiotic surrogates for more complex biotic parameters (here coliform/*E. coli* but ultimately environmental AR, see Chapter 3) was investigated for the DoE monitoring dataset. For 2018 data, Spearman's correlation were strongest for NH_3-N with coliform and *E. coli* (Spearman's ρ 0.77 and 0.79, respectively, with $P < 0.05$, n = 54, Figure 4-5) and slightly weaker for DO with coliform and *E. coli* (Spearman's ρ 0.64 and 0.56, respectively with $P < 0.05$, n = 54, Figure 4-5). Despite higher NH_3-N correlations with *E. coli* than coliforms, coliforms were chosen for linear regression analysis as several *E. coli* data points were under the detection limit.
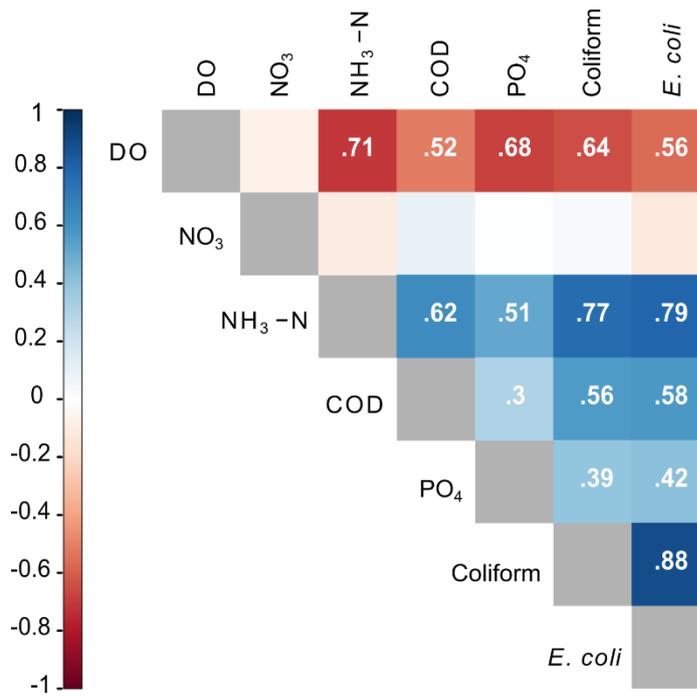
Figure 4-5. Spearman correlations between selected physico-chemical water quality and microbial parameters for the Skudai catchment based on the 2018 Department of Environment (DoE) dataset (n = 54). Correlation values shown for P < 0.05 with P values corrected for multiple testing with the Benjamini-Hochberg approach. COD: chemical oxygen demand. DO: dissolved oxygen. NH_3-N: ammonia. NO_3: nitrate. PO_4: phosphate.

The linear regression model (NH_3-N, log_{10} coliform) explained 52% of the data variability at P < 0.0001 (Figure 4-6). The mean of the residuals approximated zero with the residuals being serially uncorrelated (Durbin-Watson test P >0.05) and normally distributed (Appendix Figure B-3). The linear regression model was applied to predict 2017 coliform concentrations with 2017 NH_3-N data. The measured and predicted coliform concentrations correlated moderately (Pearson's r 0.48, P < 0.0001, n = 73).

a) Linear regression model

$\log_{10}$ coliform (CFU/100 mL) =
4.3874 + 0.2681 x NH$_3$-N (mg/L)

R^2 = 0.52
$P < 0.0001$

b) Predicted 2017 coliform data

Pearson's r = 0.48
$P < 0.0001$
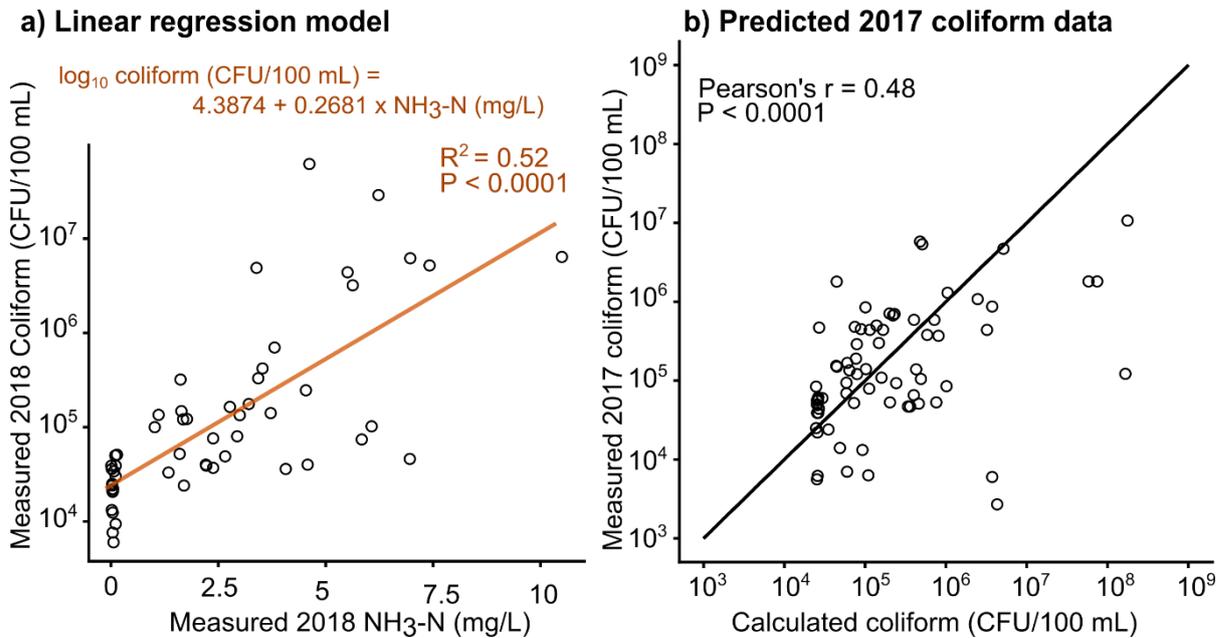
Figure 4-6. a) Linear regression model for NH$_3$-N and coliform ($\log_{10}$ transformed) concentrations based on the 2018 Department of Environment (DoE) dataset (n = 54). b) Predicted coliform concentrations based on measured NH$_3$-N 2017 data (see equation in panel a) plotted against measured coliform data for 2017 (n = 73).

4.3.2 HSPF simulation of streamflow and dissolved oxygen

HSPF was tested for its reliability to model streamflow and DO concentrations in the Skudai catchment (Appendix Table B-3) with the future goal of predicting relative ARG concentrations using DO and-or ammonia as abiotic surrogates. HSPF model calibration and validation followed the split-sample calibration/validation procedure (see [179] for a full list of calibrated parameters).
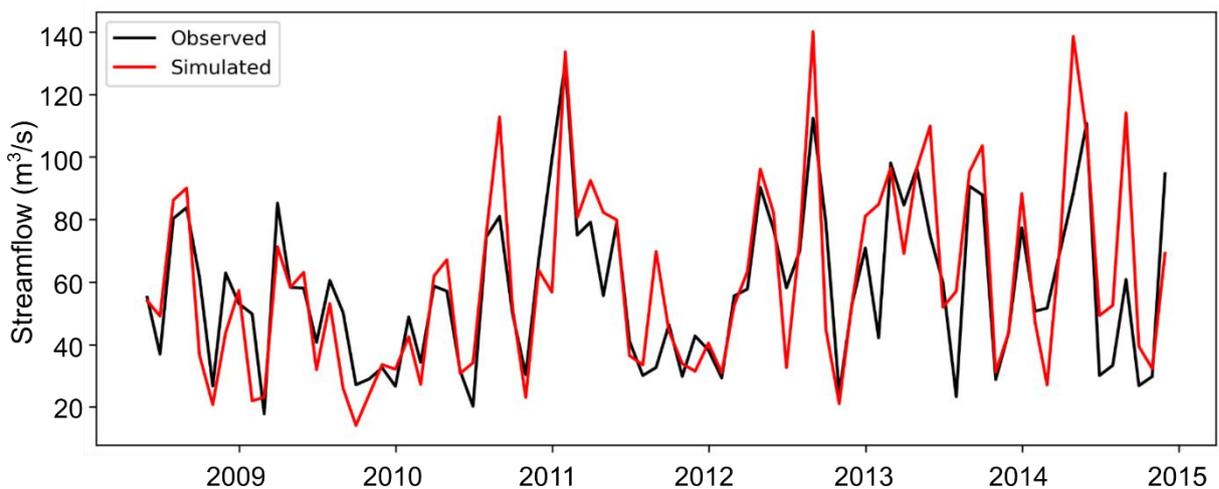
Figure 4-7. Comparison of the observed and simulated streamflow mid-stream Skudai at Kampung Separa (model validation).

HSPF simulated streamflow in the Skudai catchment (Figure 4-7) well for the calibration (2002 – 2008, R^2 = 0.89, NSE = 0.88) and validation period (2008 – 2014, R^2 = 0.83, NSE = 0.82)[179].
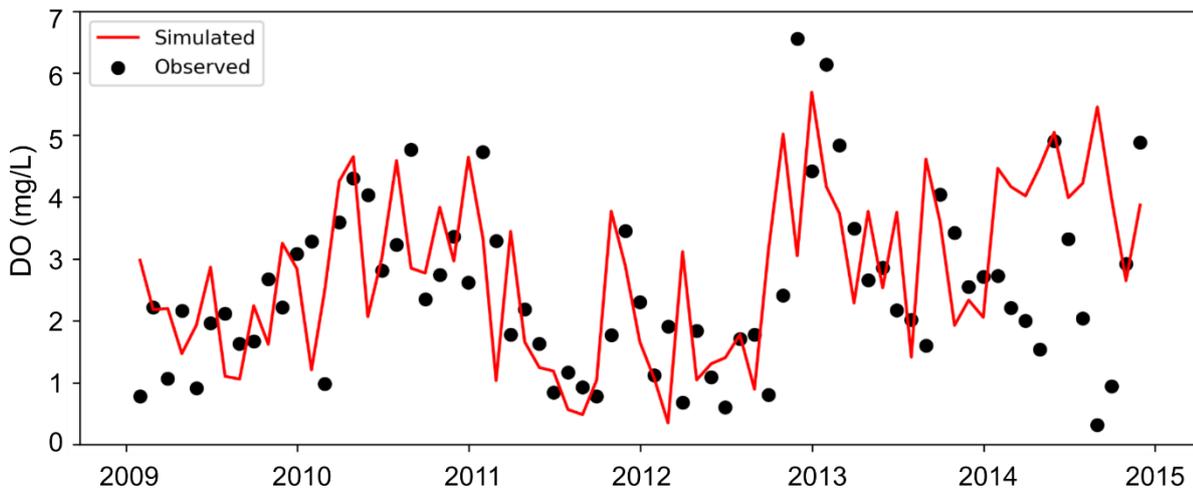
Figure 4-8. Comparison of the observed and simulated dissolved oxygen (DO) concentration mid-stream Skudai (3SI17; model validation).

The Skudai HSPF baseline model simulated DO (Figure 4-8) well for the calibration period (2002 – 2009, R^2 = 0.79, NSE = 0.79) and moderately for the validation period (2009 – 2015, R^2 = 0.44, NSE = 0.39)[179].

For both, streamflow and DO concentrations, the HSPF baseline model well represented the catchment in capturing seasonality and interannual variability.

4.4   Discussion

This study supports the hypothesis that easy-to-measure abiotic data from national water quality monitoring in LMICs could be used to predict more complex biotic parameters such as AR.

For the Malaysian DoE dataset, abiotic parameters (DO and NH_3-N) correlated strongest with *E. coli* and coliform levels. With a crude simple linear regression approach, we were able to predict coliform concentrations with statistical significance, in this case, solely based on NH_3-N levels. Clear relationships between abiotic and biotic factors suggests that a hydrological model such as HSPF should be able to predict ARG concentrations based on DO or NH_3-N concentrations.

Prior to amending an abiotic component in HSPF to describe AR, HSPF was first investigated for its ability to accurately simulate streamflow and DO in the Skudai catchment. HSPF well captured the seasonality of river flows and DO together with

inter annual variability (Figure 4-7). This may be taken to infer that the dominant flow pathways are reasonably captured by the model (i.e., faster runoff, slower interflow in the upper soil layers and groundwater; note that the model has 3 layers which correspond to these 3 runoff processes). A prerequisite for water quality modelling at the catchment scale is a good representation of the hydrological cycle.

The intension was to then extend the model to 2018 to compare simulated DO with measured ARG data and ultimately amend the DO HSPF component to accurately predict ARG levels. However, data issues due to incomplete flow record access delayed model development. Nevertheless, the envisioned concept is demonstrated below where measured DoE DO data and linear regression modelling was applied to estimate 2010 Skudai ARG concentrations (Figure 4-9).

a) Measured 2010 DO concentrations   **b) Estimated 2010 ARG concentrations**
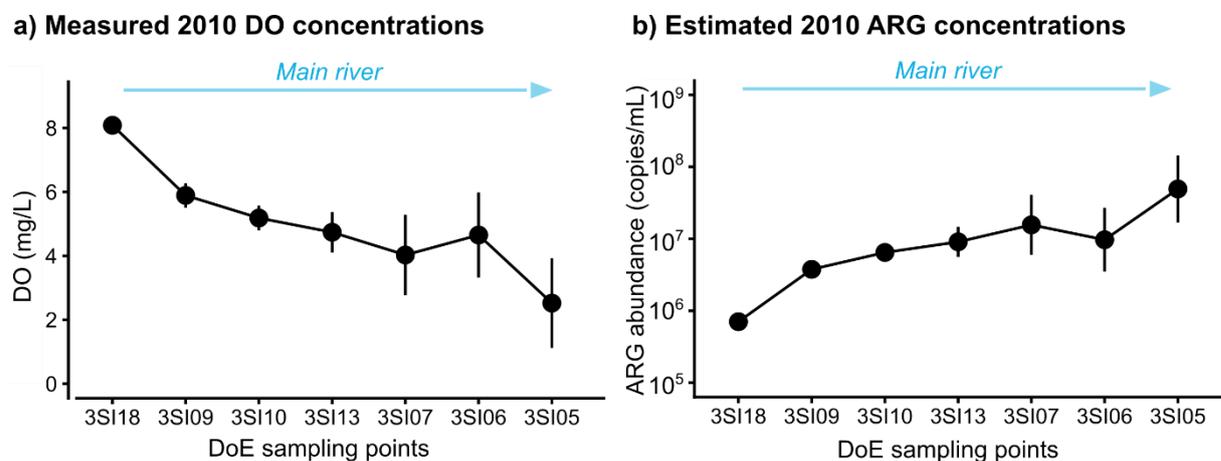
Figure 4-9. Demonstrating the concept to estimate antibiotic resistance gene (ARG) levels through easy-to-measure surrogates such as dissolved oxygen (DO). Monthly measured 2010 Department of Environment (DoE) DO data (a) and predicted ARG data (b) for the Skudai river, represented as means with standard deviations. ARG levels were estimated with a linear regression model, based on ARG-DO relationships observed in the 2018 field study (Chapter 3, $\log_{10}$ ARG = -0.3319*DO + 8.5305).

No AR datasets are available to validate these predictions, but the figure visualises how easy-to-measure surrogates might predict AR 'hot spots' in LMIC catchments. The above analyses provide an indication of the capabilities of modelling, and the future promise. By capturing key hydrological and water quality processes, catchment models such as HSPF will further improve AR simulations via abiotic surrogates. Work is ongoing to modify HSPF to predict relative ARG concentrations using DO and/or ammonia as abiotic surrogates. Such surrogate-based predictive monitoring approaches will not substitute for detailed local analysis, but they can be used to triage catchments for limited expense, allowing LMICs to focus resources on AR studies on places with potentially greater exposure risk.

Chapter 5. Improved quantitative microbiome profiling for environmental antibiotic resistance surveillance

5.1 Introduction

The huge impact of COVID-19 on human health and global economies provides a glimpse of what might occur if AR continues to increase[9]. Between 2014 and 2016, more than one million people died due to drug resistant pathogen infections and increasing death tolls are expected in the future[10]. AR pathogens not only spread through hospitals, but also enter the environment via insufficiently treated sewage[121,254]. This is especially a problem in emerging countries. Increased economic wealth permits greater access to antibiotics while waste management often lags behind[21]. However, quantifying the extent of environmental AR over space and time is difficult because methods are not standardised, with researchers using different measures of AR (e.g. antibiotics, ARGs; ARBs; and MGEs) across studies[28]. Ideally, bacterial hosts of ARGs should be tracked[255], but reliable molecular methods that couple bacteria species and ARG abundances (e.g. epicPCR[132], Hi-C[256]) are still in their infancy. Further, linking microbiome characteristics from DNA sequencing with quantitative ARG data is an unfulfilled aspiration for studying environmental AR[134,135]. This restricts our ability to perform realistic Quantitative Microbial Risk Assessments (QMRA) needed to quantify true risks of environment AR exposures[257,258]. Correlation-based methods can develop hypotheses to guide future experimental work but they are restricted due to technical biases introduced from DNA sequencing[229,255,259].

Next-generation sequencing (NGS) data are inherently compositional, providing relative abundance information at best[136]. It is impossible to measure absolute growths or declines of particular microorganisms solely with relative abundances as, for example, the increase of one taxon leads to the concurrent decrease of other(s)[137]. Analysing relative abundance data using inappropriate statistical tools can yield up to 100% false detection rates and their application contributes to a general lack of reproducibility among microbiome studies[260,261].

While compositional approaches are available[136], the gold standard requires cell count estimates to calculate absolute abundances[255,261]. Such a quantitative approach can also correct sequencing data for sampling intensity to account for varied microbial loads across samples[137] (Figure 5-1). Despite environmental studies

routinely providing cell count estimates, these data are rarely used to calculate absolute microbial taxon abundances[262], with no studies correcting for sampling intensity[137]. We contend that environmental researchers should use quantitative microbiome profiling (QMP)[137] instead of relative microbiome profiling (RMP) to represent a more accurate picture of relationships between microbiomes and metadata (such as ARG concentrations) and guide future QMRA applications.
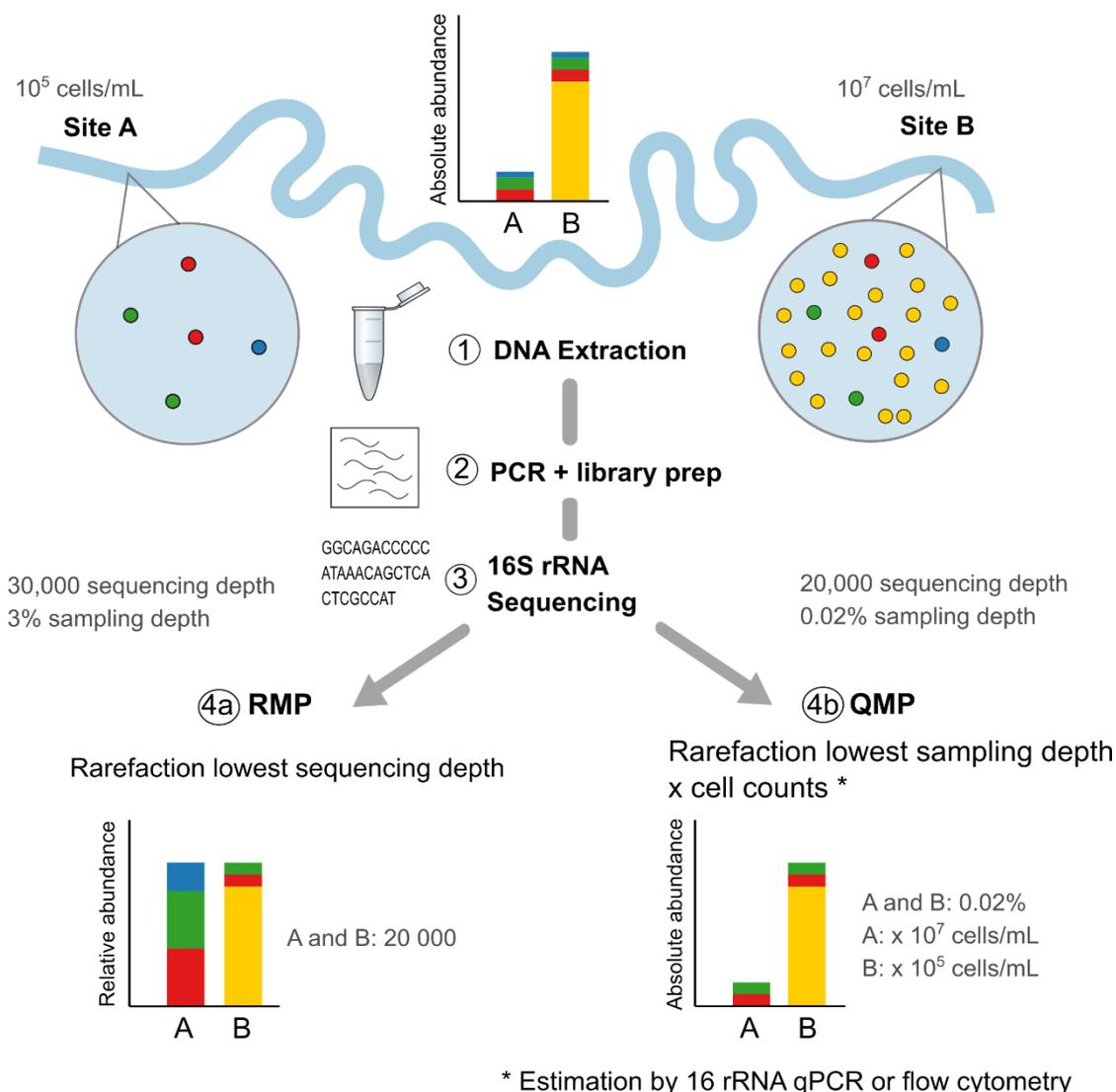
Figure 5-1. Schematic explaining relative (RMP) and quantitative (QMP) environmental microbiome profiling. Both, the RMP and QMP approach do not correct for biases introduced by sample collection, DNA extraction, PCR or library preparation. QMP approach based on Vandeputte et al. (2017). While cell counts vary 100-fold between river water samples A and B, sequencing depth (= reads) per sample is independent of cell counts in next-generation sequencing. The RMP approach rarefies to lowest sequencing depth per sample, calculating relative abundance (%), which results in sample A being sequenced more intensively than sample B. The relative abundance profile poorly reflects the real environmental taxa distribution. The QMP approach corrects for sampling intensity by rarefying to the lowest sampling depth (= sequencing depth divided by cell counts) and then multiplies the rarefied taxon abundance with estimated cell counts to obtain absolute abundances (here per mL river water). As the blue taxon was equally abundant in A and B, the fact that it is included for RMP sample A can be considered an artefact of uneven sampling intensity.

Characterising and comparing anthropogenic impacts on environmental microbiomes (e.g. sewage entering rivers, waste leaching etc.) is generally hindered by the use of varying microbial diversity indices across studies[263–265]. For a more meaningful quantification, 'diversity' needs to be unambiguously defined and applied in microbiome research[266]. Common diversity indices such as the Shannon and Simpson index do not measure diversity, but uncertainty and probability, respectively[265]. In contrast, Hill numbers (Figure 5-2) provide a statistical framework that unifies and generalizes popular indices, and are intuitive and flexible enough to address a wide range of scientific questions[265,267,268]. Hill numbers were first proposed almost 50 years ago[268], but despite their continued appraisal[265–267], their use in microbiome research is rare[269,270], especially for environmental microbiomes. Hill numbers ^qD also have several additional advantages over other common diversity indices (Table 5-1).
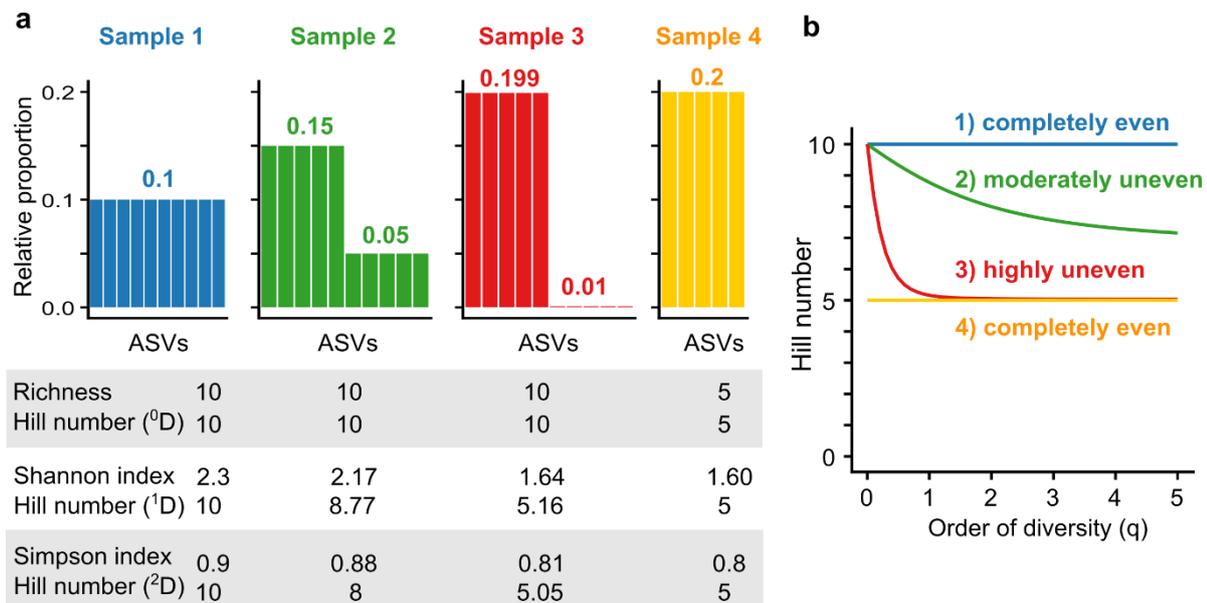
Figure 5-2. Schematic explaining the relationship between microbiome composition, diversity indices (richness, Shannon index and Simpson index), Hill numbers ^qD (a) and diversity profiles for four theoretical systems (b). Figure adapted from Alberdi and Gilbert (2019a). For sample 1 and sample 4, all amplicon sequence variants (ASVs) are evenly distributed, so Hill numbers of all orders of diversity (q) stay the same within sample 1 and sample 4. As sample 4 has half the amount of equally abundant ASVs to sample 1, Hill numbers also half, in contrast to the Shannon index or Simpson index. At q = 0, only richness is considered, ignoring relative abundance. Consequently, for q = 0, Hill numbers for samples 1, 2 and 3 are the same. For q > 0, Hill numbers decrease as the importance attributed to abundant ASVs increases. As sample 3 is dominated by 5 ASVs, Hill numbers ^1D and ^2D approximate 5. The diversity profile (b) shows the number of ASVs and evenness of the four theoretical systems. A flat profile indicates evenness.

Despite clear advantages in using Hill numbers[271] and the QMP approach[137] for improving reliability and comparability of environmental microbiomes, their application is rare[137,269], and to our knowledge, has never been combined. In this chapter, a

workflow is provided that combines QMP (based on parallelization of amplicon sequencing and 16S rRNA qPCR data to estimate cell counts) with absolute resistome profiling (based on high-throughput qPCR for almost 300 ARGs and MGEs) to better qualify AR in an impacted river. Such absolute microbiome profiling bypasses compositional effects in the reconstruction of microbiota interaction networks, allowing one to investigate correlations of taxa with ARGs and MGEs essential for QMRA. We also show the benefits of using the unified Hill number diversity framework to compare microbial community dynamics over space and time and confirm how misleading RMP approaches are for interpreting environmental microbiome and resistome data.

Table 5-1. Advantages of Hill numbers in comparison to standard diversity indices.

1	Interpretation of the measure and its **measurement unit is always the same** in 'effective numbers of species', i.e. the number of equally abundant species (or for DNA based approaches operational taxonomic unit (OTU)/amplicon sequence variant (ASV) [271]) required to generate an identical diversity[268].
2	Hill numbers **double** as the amount of equally common species doubles (called the 'doubling principle'), which allows more meaningful **calculations of statistical significant changes**[265].
3	The **sensitivity** towards abundant and rare species can be **modulated** with a single parameter with Hill numbers (order of diversity – q).
4	Hill numbers can be computed taking into account **phylogenetic or functional relationships** among species (e.g. similar to Faith's Phylogenetic Diversity[271]).
5	Hill numbers were originally developed for **abundance data**, but can also be applied to **incidence data**[267].
6	Within the Hill framework, the **diversity of a system can be partitioned**, so α-diversity (average diversity of subsystems) multiplied by β-diversity (difference between subsystems) gives γ-diversity (entire diversity of the system)[272,273].
7	**Multiple (dis)similarity measurements** derived from β-diversities can be calculated from Hill numbers with some being equal to other popular indices e.g. Unifrac[267].
8	The calculation of Hill numbers is **straight-forward** and can easily be **implemented into existing bioinformatic pipelines**[274].

5.2   Methods

5.2.1   Sample collection and DNA extraction

We collected river water samples (3 x 1 L) from the Skudai catchment, Malaysia (288 km^2, Figure 3-1) at eight sampling points (6x main river and 2x tributaries) during five sampling trips to capture seasonality (1x November 2017, 2x March 2018 and 2x July 2018). In total, 38 samples were collected with five biological replicates for the main Skudai river (S1, S2, S5, S6, S7, S8) and four biological replicates for the tributaries Melana (M5) and Senai (Se1). For more details on sample collection, see 3.2.2.

River water was filtered onto 0.22 µm cellulose-nitrate filters to extract DNA with the FastDNA SPIN kit for soil (MP Biomedicals). DNA was cleaned with the QIAquick Nucleotide Removal Kit (Qiagen). DNA quality and quantity were measured with NanoDrop and the Qubit dsDNA HS assay (both Thermo Fisher Scientific), respectively. The three technical replicates were pooled to have sufficient DNA for downstream processes. DNA was stored at -20 ⁰C. For more details on DNA extraction, see 3.2.6.

5.2.2 16S rRNA qPCR to estimate cell concentration

16S rRNA qPCR assays were performed in triplicate with 16S rRNA 1055f-1392r primers[275] and SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) on the BioRad CFX C1000 System (Bio-Rad) following thermocycle program: (i) 2 min of initial denaturation at 98 ⁰C, and 40 cycles of (ii) 5 s denaturation and 98 ⁰C, and (iii) 5 s annealing/extension at 60 ⁰C[70]. DNA samples were diluted to a working solution of 5 ng/µL and an internal control DNA (gfp_qPCR_f: TCGGTTATGGTGTTCAATGC; gfp_qPCR_R: GACTTCAGCACGTGTCTTGTAG) was used as inhibition controls for the qPCR. Standard curves of each set of primers were constructed using plasmid clones of the target sequences of between 10^2 and 10^8 copy numbers, used in parallel with each qPCR run. Cell concentration was estimated by dividing the 16S rRNA concentration by 4.1, the estimated average 16S rRNA GCN per bacterium[199]. We did not incorporate individual 16S GCN adjustments on the sequencing reads[137,276] as current correction approaches were found to introduce rather than reduce biases[277]. The resolution of Illumina MiSeq often only allows ASV characterisation to genus level, but already within species, 16S GCN can vary widely (e.g. 6 to 11 16S GCN for *Escherichia coli*[278]).

5.2.3 High-throughput qPCR to quantify the resistome

HT-qPCR of ARGs and MGEs was performed using SmartChip Real-Time PCR (Wafergen). A total of 296 primer sets (Supplementary Table 8) were used to detect 283 ARGs (52 β-lactams, 51 non-specific efflux pumps, 46 MLSBs, 39 tetracyclines, 36 aminoglycosides, 32 vancomycins, 11 others, 9 FCA, 7 sulfonamides), 12 MGEs (8 transposases , 4 integrases) and one 16S rRNA gene as previously described[197,198]. For more details on the Ht-qPCR assay and analysis, see 3.2.6.

5.2.4 16S rRNA sequencing and bioinformatics

The hypervariable V4 region 515F-806R[279] of the 16S rRNA gene was sequenced on the Illumina MiSeq platform with V2 500 cycle chemistry at NU-OMICS, Northumbria University, UK. Sample preparation and sequencing followed the Schloss MiSeq Wet Lab SOP[280] with the only deviation of spiking a 4.5 pM library, as opposed to 4 pM. Sequencing included a positive control (mock community, ZymoBIOMICS Microbial Community DNA Standard, Zymo Research), negative control (water), and extraction control (extracted water). Raw sequences were processed with QIIME2 v.2019.4[281]. Reads were denoised into ASVs with DADA2[282,283], assigning ASVs to genus level with the SILVA reference database (v 138)[284–286]. The V4 primer region 515F-806R was extracted from the SILVA 138 SSU NR99 dataset to retain more sequences within this region as opposed to using primer sequence to find and remove the corresponding region in the QIIME2 environment[287]. The SILVA 138 V4 classifier was trained with the machine learning software library scikit-learn v.0.20.0 using Naïve Bayes methods (*fit-classifier-naive-bayes*[288]) through the *feature-classifier* plugin[289]. The taxonomy was assigned through the same plugin, using the sklearn-based taxonomy classifier (*classify-sklearn*[288]). Accounting for MiSeq bleed-through between runs[290], rare ASVs of less than 0.1% of the mean sample depth were removed. The taxonomy and ASV table biom file[283] were produced for downstream analysis in R[200] with the phyloseq (v 1.34.0)[291] and vegan (v 2.5-7)[292] package. ASVs not classified at phylum level were removed, resulting in a total of 2,735 taxa for 38 samples with minimum 12,712 and maximum 83,570 reads.

5.2.5 Quantitative and relative microbiome profiling

For QMP, we rarefied samples to an equal sampling depth (ratio between sequencing depth and cell counts (Appendix Figure C-1)) with the R function *rarefy_even_sampling_depth* (seed 711)[137]. Reads were not corrected for individual 16S rRNA GCN. The resulting rarefied abundances were multiplied with the estimated cell concentration per sample to obtain absolute microbial taxa abundance per mL of river water. For RMP, we rarefied sampled to an equal sequencing depth of 12,712 (seed 711), resulting in relative microbial abundances.

5.2.6 Rank-based RMP and QMP comparisons

We analysed ASV rank order changes between the RMP and QMP approach with the rank-biased overlap (RBO) measure and a genus co-occurrence network based

on Spearman's correlation. RBO is a similarity measure on ranked lists, developed to measure the expected overlap of indefinite rankings[293]. RBO does not require every item to appear in both rankings, is not tied to a particular prefix length and its top-weightedness can be adjusted. For the latter, parameter p determines the strength of the weighting to top ranks. Raising p increases the depth of comparison, e.g. for p = 0.9, p = 0.95 or p = 0.97, 85% of the RBO measure focus on the first ten, first 20 or first 50 results, respectively[293]. We calculated RBO on the most abundant 100 ASVs with p = 0.95 to top-weigh the first 20 results in R with the package gespeR (v 1.23.0)[294].

For the co-occurrence patterns, we first removed unclassified or ambiguously defined ASVs at genus level and then selected ASVs present in at least 85% of samples based on the QMP data (=24 ASVs). The same 24 ASVs were also selected in the RMP data. We defined and visualised taxon-taxon associations by Spearman's correlations between pairs of taxa with Benjamini-Hochberg multiple testing correction in R with the packages psych (v 2.1.3)[215] and corrplot (v 0.84)[216].

5.2.7 Resistome volcano plot

We assessed the difference in $\log_{10}$ ARG and MGE river water concentrations between up- and downstream (S1 to S8) with the Welch's t-test, applying Benjamini-Hochberg P adjustment to correct for multiple testing. We plotted the $\log_{10}$ fold change against statistical significance in a volcano plot with the R package EnhancedVolcano (v 1.8.0)[214].

5.2.8 Network analysis for microbiome and resistome correlations

We investigated microbiome and resistome co-occurrence by calculating all possible pairwise Spearman's rank correlations among bacterial orders, ARGs and MGEs present in the river water samples (n = 38). Only statistically robust correlations with Spearman's $\rho > 0.8$ and Benjamini-Hochberg multiple testing corrected $P < 0.01$[295] were included in the network. Network analysis was performed in R with visualisation including topological property calculations in Gephi (v 0.9.2)[296].

5.2.9 Hill diversity analysis

Abundance-based Hill numbers and diversity profiles for RMP and QMP were calculated and plotted with the hilldiv R package (v 1.5.1)[274]. The Sørensen-type overlap dissimilarity measure for q = 1 was used to quantify the effective average proportion of nonshared ASVs in the catchment and visualised in a NMDS plot. As

the Hill number ^qD equation[265,271] is not defined for q = 1, the R package hilldiv calculated ^qD for this case with q= 0.99999 (Equation 5-1).

Equation 5-1. Hill number ^qD equation.

$$^qD = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)}$$

^qD: Hill number
q: Order of diversity
S: Species richness
p_i: Proportional abundance of species i

5.2.10 Statistical analysis and graphics

Raw amplicon sequencing data that support the findings of this study have been deposited in European Nucleotide Archive with study accession number PRJEB42314. All other data can be accessed through the Center for Open Science repository, OSF ((Ott, Amelie. 2021. 'Monitoring and Modeling of Antibiotic Resistance in Southeast Asian Rivers. https://osf.io/gcpky/?view_only=90e614c2c6b64483aa503694af113789).

We performed all statistical analysis in R (v 4.0.5)[200]. We composed graphics using ggplot2 (v 3.3.3)[201] with finalisations in Inkscape (v 1.0.2)[202] except for where stated differently. The Skudai catchment map was composed in ArcGIS (v 10.6.1)[203]. To assess statistically significant difference in microbiomes and resistomes between upstream (S1) and downstream (S8), we tested for normality with the Shapiro-Wilk test, followed by comparisons with the Welch's-test[209]. Effect size was measured with Cohen's D with the R package effsize (v 0.8.1)[211]. Box-plot elements are defined as centre line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range) and points (outliers).

5.3 Results

5.3.1 Relative and absolute microbial taxa abundances

For this study, we collected river water samples in a Malaysian rural-to-urban catchment from eight sampling points over five field trips in different seasons. A previous study for this catchment found no large statistically significant seasonal effects for water quality and resistome data (see Chapter 3). Consequently, mean concentrations with standard deviations are reported per sampling point across seasons. We estimated river water cell concentrations with 16S rRNA qPCR,

correcting for multiple 16S rRNA gene copies per cell. In the catchment, cell counts varied more than 100-fold across samples with mean upstream concentrations of $(9 \pm 3) \times 10^5$ cells/mL (S1) and mean downstream concentrations of $(2 \pm 1) \times 10^7$ cells/mL (S8) (Appendix Figure C-1).

River water microbiomes were assessed by 16S rRNA sequencing with Illumina MiSeq, classifying ASVs to genus level. After data quality filtering, reads varied from 12,712 to 83,570 (median 28,187, Appendix Figure C-1). Sampling depth (i.e., reads/cell count) was highest in upstream samples (S1; mean 3.4%), with lower sampling depths obtained elsewhere in the catchment (mean 0.16% - 0.59%, Appendix Figure C-1). The lower cell counts upstream resulted in S1 samples being 21x more intensely sampled in the microbiome analysis than the most downstream site, S8 (Appendix Figure C-1).
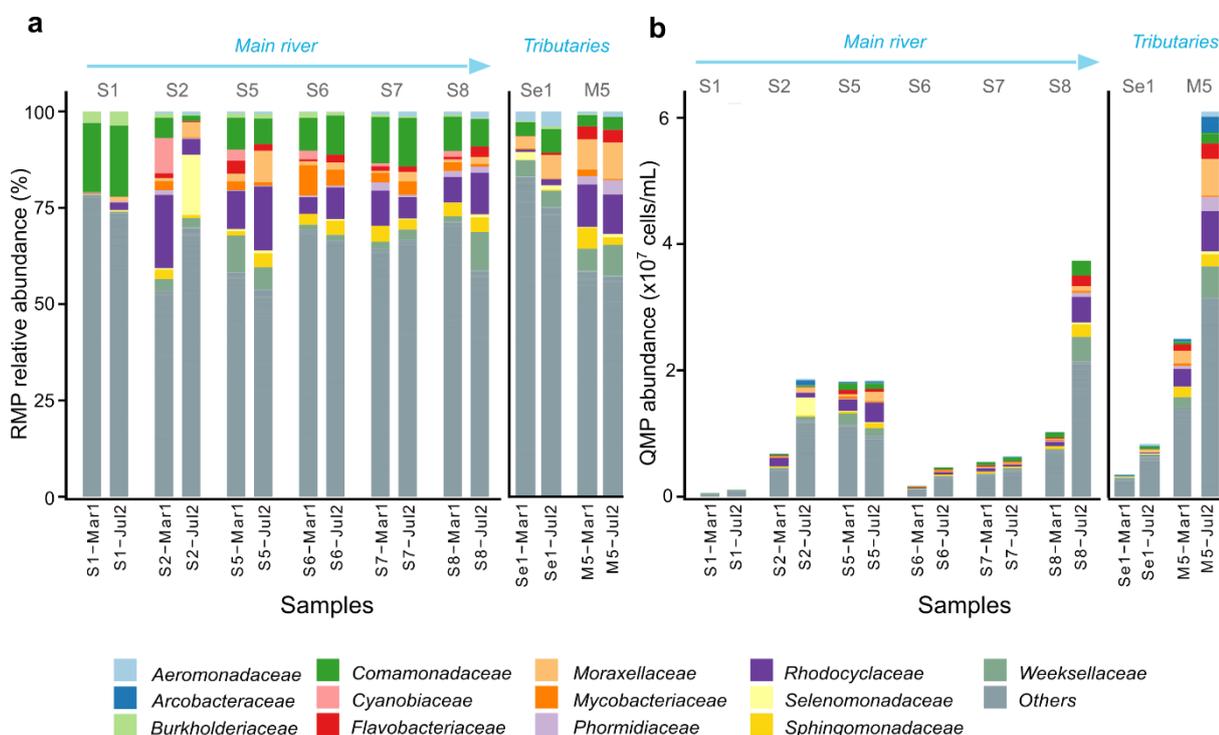
Figure 5-3. Barplots showing the 20 most abundant ASVs grouped into families with remain pooled into 'Other' for the relative (RMP; a) and quantitative (QMP; b) microbiome profiling approach, analysing river water samples from eight sampling points for two sampling campaigns (March and July 2018, n = 16)). See Appendix Figure C-2 for all 38 samples.

For RMP normalisation, samples were rarefied to equal sequencing depth (i.e., number of reads per sample; here 12,712 reads, Appendix Figure C-3). Despite known problems[297], the RMP approach remains the common practice in environmental microbiome research to calculate relative abundances of taxa (Figure 5-3)[298]. For QMP[137], samples were rarefied to equal sampling depth (here 0.05%)

and multiplied with the estimated cell counts per sample to obtain absolute abundance of taxa per mL river water (Figure 5-3). In contrast to Vandeputte et al. (2017), individual 16S rRNA gene copy number (16S GCN) adjustment was not performed because related methods are imprecise, introducing additional bias[277].

The most abundant ASVs (based on QMP, Appendix Table C-1) were *Cloacibacterium, Acinetobacter, C39* (genus level), and *Comamonadaceae* (family level). When comparing taxa changes across the catchment, the RMP barplot (Figure 5-3a) provides misleading results. For example, it suggests *Comamonadaceae* decrease as one moves downstream (S1 → S8), whereas when one takes cell counts into consideration (Figure 5-3b), *Comamonadaceae* increases from up- to downstream, which might be linked to progressive waste inputs.
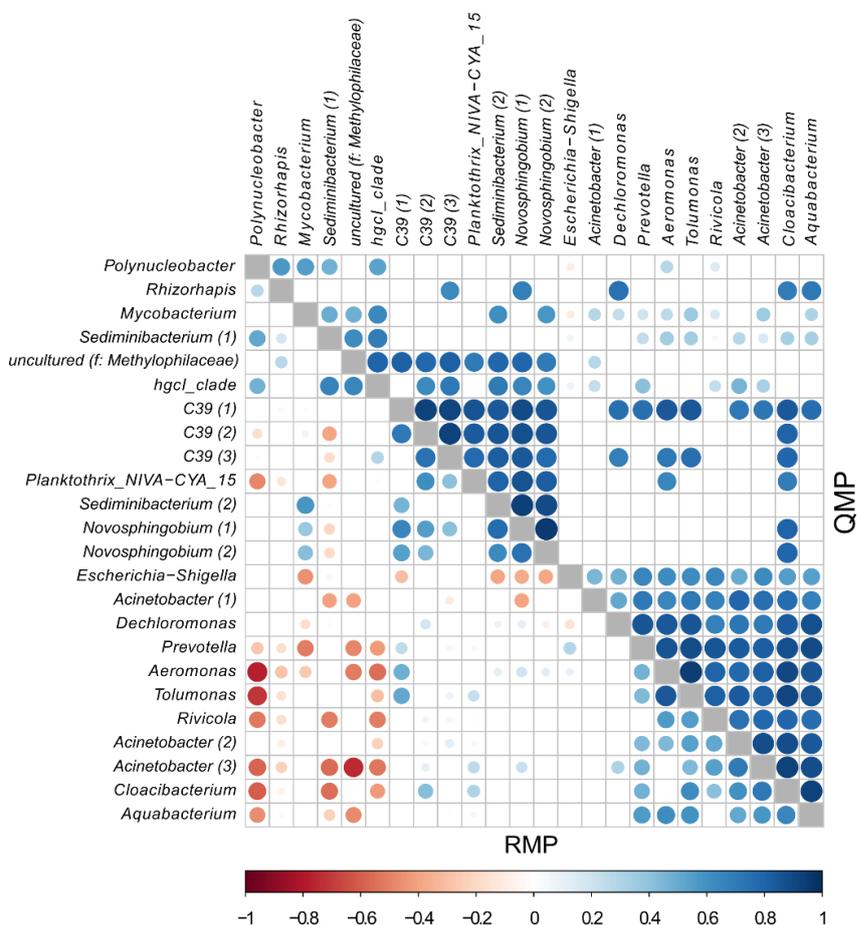
Figure 5-4. Co-occurrence patterns for genus detected in at least 85% of the samples based on relative (RMP) and quantitative (QMP) microbiome profiling. Pairwise correlations between taxon abundances were calculated, and significant correlations (Benjamini-Hochberg adjusted test, $P < 0.05$) are represented by circles, the colour and size of each circle represent the correlation coefficient (Spearman's ρ). f: family.

As relationships between microbiomes and metadata are often explored using non-parametric rank-based methods, we assessed whether the ASV rank order was

conserved in the QMP vs RMP approaches. Out of the 20 most abundant ASVs determined with QMP, 16 also were present in the top 20 ASVs from the RMP approach, but only three ASVs were at the same rank order in both listings (Appendix Table C-1). Assessing the similarity of the rank order of the 100 most abundant ASVs with the rank-biased overlap for top-weightedness[293], we found that only 32% of the QMP and RMP results were in common (score 0.32 with p = 95, focussing 86% of the weight on top 20 ASVs), suggesting the two methods providing different pictures of the system - RMP only provides composition, whereas QMP provides composition and abundance in tandem.

Correlation analyses are often used to infer taxon-taxon interactions[259]. Constructing RMP and QMP genus co-occurrence networks (Figure 5-4), we detected a much larger number of significant co-varying genus pairs in the QMP than RMP network (249 versus 116). The RMP network also was dominated by negative correlations. None of the moderate to strong RMP correlations (P < 0.05, Spearman's ρ -0.5 to -1) were detected in the QMP correlation matrix (Figure 5-4).

5.3.2 *Hill numbers for microbial diversity*

Within the Hill framework, microbial diversity can be calculated for subsystems (α-diversity; the sampling locations), the entire system (γ-diversity; the river catchment), and the difference between subsystems (β-diversity; between sampling points), all expressed using one unit, the effective number of ASVs[271]. The importance of 'richness' (ASV count in a community) and 'evenness' (equality of ASV frequency in a community) to the overall diversity can be modulated with the parameter q[299]. For diversity of order zero (q = 0), the Hill number is a 'richness' value because it becomes insensitive to ASV frequency, which overweighs rare ASVs. At q = 1 (exponential of Shannon index), ASVs are weighed by their frequency without favouring rare or abundant ASVs. For q = 2 (inverse of Simpson index), abundant ASVs are overweighted[265]. While specific q values can be selected to calculate diversity, using α-diversities at q = 0, q = 1 and q = 2 together allows one to assess the degree of dominance in a community (Appendix Figure C-4). This information can be summarized in a 'diversity profile', a graph of diversity versus q, visualising the contributions of richness and evenness to a community's diversity (Figure 5-5). The richer a community (higher ASV count), the higher the graph starts, whereas the more uneven the community (few dominant ASVs), the steeper the slope of the graph[265].
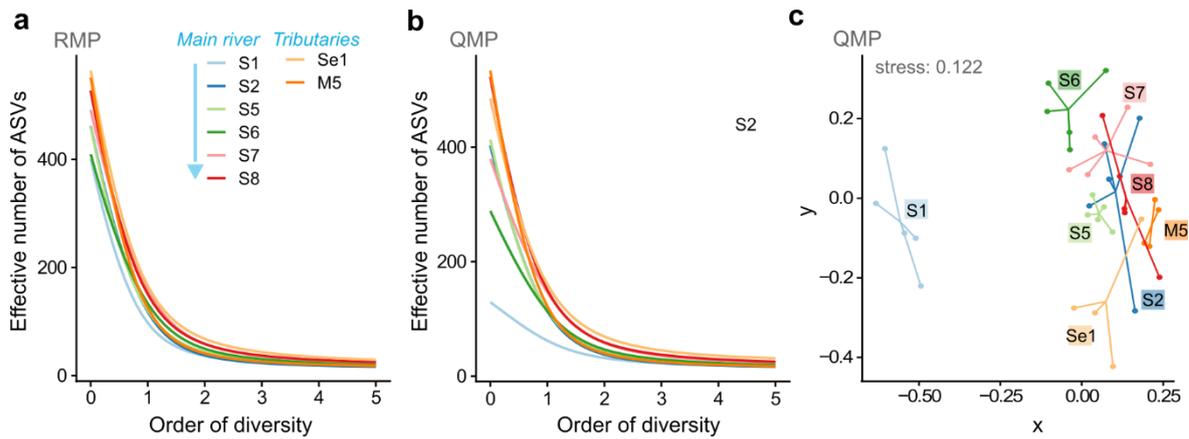
Figure 5-5. Microbial diversity calculated within the Hill framework across the river catchment. Hill diversity plots represent α-diversities per sampling point based on the relative (a) and quantitative (b) microbiome profiling approach for varying q values. NMDS Sørensen-type overlap dissimilarity plot (c) is based on β-diversity, calculated for the QMP data with q = 1. Data represented (n=38) for the eight sampling points is based on five biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on four biological replicates for the tributaries (Se1, M5).

Microbial diversities at each sampling point in the RMP diversity profile were closely aligned, with clearer differentiation seen for the QMP data (Figure 5-5a, b). Both approaches showed microbial diversity was lower upstream (S1) than elsewhere in the catchment, but spatial differences were smaller using RMP (Appendix Figure C-4). This trend also was observed when calculating the Shannon and Simpson index (Appendix Figure C-3). Further, γ-diversity of the catchment was higher using the RMP versus the QMP approach, but the values for two approaches converged for q > 0. For RMP, γ-diversity in effective numbers of ASVs was 2721 (q = 0), 338 (q = 1) and 96 (q = 2) and for QMP, the values were 2428 (q = 0), 328 (q = 1) and 96 (q = 2).

Results from the RMP and QMP approach differed most in their diversity calculations for the least impacted upstream sampling point S1 (mean difference α at q =0 was 272 effective number of ASVs, Appendix Figure C-4) with the QMP approach better correcting for varying sampling depths (Appendix Figure C-1), thus avoiding 'over-sequencing'. For the QMP approach (Figure 5-5b), the upstream microbial community (S1) was significantly less diverse for q = 0 and q = 1 than the farthest downstream (S8) (Welch's t-test with P < 0.05 and large Cohen's D effect size < -0.8, Appendix Table C-2). At S1, the microbial community also was more even than at any other sampling point downstream (Figure 5-5b).

Comparing the α-diversities for the tributaries Se1 and M5 (Appendix Figure C-4) further shows the benefit of reporting Hill numbers at varying q values. While the tributaries have similar diversities at q = 0 (richness), the diversities for q > 0 (taking

frequency into account) decrease more rapidly for the heavily polluted M5, showing a more uneven microbial community in comparison to the less polluted Se1 (Figure 5-5b, Appendix Figure C-4).

Within the Hill framework, dissimilarity matrices are based on β-diversities[272,273]. We used the Sørensen-type overlap dissimilarity measure for q = 1 to quantify the effective average proportion of nonshared ASVs in the catchment[274] (Figure 5-5c). The NMDS plot shows the changing community structure as one moves from rural upstream (S1) to more urbanised downstream (Figure 5-5c).

5.3.3 Characterising the river resistome

We quantified the river water resistome by applying high-throughput qPCR with 283 ARG, eight transposase and four integron primers. In total, 211 ARGs (~75% of those assayed) were detected in the river catchment with 70 ARGs (25% of assay) shared between all river water samples (n = 38 samples). All 12 MGEs were measured at least once in the sample with eight MGEs (75% of assay) shared across all samples (n = 38) (Appendix Table C-3). Detected ARGs encoded resistance to eight classes of antibiotics, with β-lactam resistance being the most common (45 detected/52 in the assay) (Appendix Table C-3).

Summarizing ARGs and MGEs, their detected numbers (number of ARGs or MGEs), river water concentrations ($\log_{10}$ ARG or MGE copies/mL) and cell concentrations (ARG or MGE copies/cell) all significantly increased from upstream (S1) to downstream (S8) (Welch's t-test with P < 0.05 and large Cohen's D effect size < -0.8, Appendix Table C-4) with the tributaries frequently having the highest ARG and MGE concentrations (Appendix Figure C-5, Appendix Table C-5). River water ARG concentrations increased more than two $\log_{10}$ steps along the catchment with ARG copy numbers per cell increasing from 0.1 copies/cell upstream to 2.2 copies/cell downstream (Appendix Figure C-5, Appendix Table C-5).).

The most abundant ARGs in the catchment encoded resistance against sulphonamides (*sul2*), aminoglycosides (*aadA1, aadA2,*), β-lactams (*blaOXA10*) and for non-specific efflux pumps (*qacEdelta1*, *qacH*) with their mean concentrations ranging between 1×10^7 to 2×10^6 gene copies/mL river water (Appendix Table C-6).
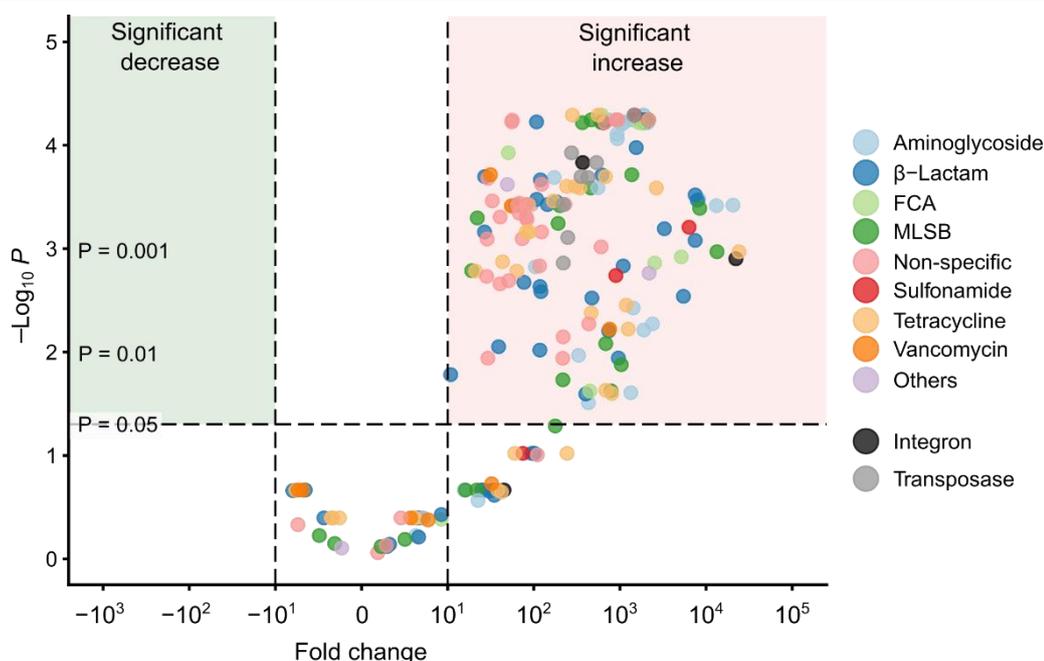
Figure 5-6. Volcano plot displaying ARG and MGE $\log_{10}$ fold river water concentration changes between upstream (S1) and downstream (S8). Statistical significance calculated with the Welch's t-test, applying Benjamini-Hochberg P adjustment. Fold change calculated by subtracting mean ARG or MGE river water concentration for S1 (n = 5) from S8 (n = 5).

To assess the resistome changes along the river, we plotted ARG and MGE $\log_{10}$ fold river water concentration changes from up- to downstream (S1 to S8) against statistical significance in a volcano plot (Figure 5-6). 146 ARG and MGE concentrations increased significantly at least 10-fold between up- and downstream (Welch's t-test, Benjamini-Hochberg adjusted P < 0.05). Four ARGs encoding for aminoglycoside, MLSB and tetracycline resistance and integron 3 increased more than four $\log_{10}$ steps from up- to downstream (Figure 5-6).

5.3.4 Network analysis of microbiomes and resistomes

Network analysis has been proposed to explore the associations between microbiomes and resistomes, but to date, such networks have been either based on relative values[134] or semi-quantitative data (relative NGS data for microbiomes and absolute HT-qPCR for resistomes[135], see Figure 5-7a). Combining QMP (rather than RMP) with HT-qPCR data allows one to more fully compose the quantitative networks (Figure 5-7b), overcoming negative correlation biases and spurious associations reported for relative abundance co-occurrence networks[136]. Based on the absolute taxa abundance data, the QMP network had a higher number of nodes and edges with a higher average node connectivity (=average degree) than the RMP network (Figure 5-7, Appendix Table C-7). While for the QMP network, 36 taxa at order level had strong correlations (Spearman's ρ > 0.8 and P < 0.01) with at least

three other nodes, this was only the case for 13 taxa in the RMP network (Figure 5-7, Appendix Table C-7).

Order				Transposases
Abs - Absconditabacteriales SR1	Clo - Clostridia	Lac - Lachnospirales	Rhi - Rhizobiales	tp1 - tnpA 01
Aci - Acidaminococcales	Cor - Coriobacteriales	Met - Methanobacteriales	Rho - Rhodobacterales	tp2 - tnpA 02
Aer - Aeromonadales	Cya - Cyanobacteriales	Mic - Micrococcales	SJA - SJA 15	tp3 - tnpA 03
Bac - Bacteroidales	Des - Desulfovibrionales	Mit - Microtrichales	Sph - Sphingomonadales	tp4 - tnpA 04
Bif - Bifidobacteriales	Ent - Enterobacterales	Osc - Oscillospirales	Spb - Sphingobacteriales	tp5 - tnpA 05
Bur - Burkholderiales	Fla - Flavobacteriales	PeM - PeM15	Syn - Synergistales	tp6 - Tp614
Cam - Campylobacterales	Fra - Frankiales	Pep - Peptostreptococcales Tissierellales	Vei - Veillonellales Selenomonadales	tp7 - tnpA 07
Chi - Chitinophagales	HOC - HOC36	Pro - Propionibacteriales	Ver - Verrucomicrobiales	**Integrons**
Chl - Chloroflexales	KIS - KIST.JJY010	Pse - Pseudomonadales	Xan - Xanthomonadales	in1 - intl 1LC
				in1c - clintl1
				in3 - intl3

Figure 5-7. Network analysis based on relative (RMP; a) and quantitative (QMP; b) microbiome profiling, revealing co-occurrence patterns among ARGs (blue circles), MGEs (red and orange circles) and taxa at order level (green circles). A connection represents a strong (Spearman's ρ > 0.8) and significant (P < 0.01, adjusted with Benjamini Hochberg) correlation. The size of each node is proportional to the number of connections (=degree). Only nodes with at least three other connections are shown.

For the QMP network, the most connected ARGs, transposases, and integrons were *blaOXA10* (152 degrees), *tnpA 02* (147 degrees) and clinical integron 1 (*clintl1*; 146 degrees), respectively (Figure 5-7b). The most correlating taxa belonged to the order of *Burkholderiales* (141 degrees), *Flavobacteriales* (135 degrees) and *Campylobacterales* (134 degrees), indicating that these bacteria might be frequent hosts of ARGs, and/or that these bacteria came from a similar source to the ARGs and MGEs (Figure 5-7b). While these correlations do not replace further monitoring, they help in hypothesis formulation, addressing better-grounded research questions[259].

5.4 Discussion

Our understanding of complex environmental microbiomes has been hindered by overly relying on relative abundance data and inconsistent definitions of diversity in describing microbial changes. This hampers the ability of environmental researchers to reliably link microbiome and resistome changes in the investigation of AR fate and spread, and other practical questions[255], such as providing quantitative data for QMRAs – a crucial knowledge gap for assessing environmental AR exposure risk.

To date, few papers have reported absolute taxa abundances[262,276,300] and, to our knowledge, only one human study[137] used rarefaction to make sampling depths equal prior to multiplying the relative taxa abundances with cell concentrations. While this normalisation step removes sequencing information for 'over-sequenced' samples (here upstream S1), it is necessary to allow a reliable comparison of microbial diversity, especially when cell counts vary widely across samples (here 100-fold). Only after sampling depth correction in QMP, did we find diversity to have increased significantly in the catchment from rural up- to urban downstream; a critical observation was not possible using the RMP approach.

Despite environmental QMP not addressing all known biases in microbiome research, it allows more accurate and easier absolute quantification of microbiota variation. In environmental studies, cell counts are routinely measured and QMP can be conducted at no extra cost, requiring little bioinformatic workflow adjustments. In this study, absolute taxa abundance data allowed to explore environmental microbiome and resistome interactions, overcoming biases related to relative taxa abundance data.

Researchers can choose between several methods to estimate cell counts for absolute taxa abundance calculation, such as 16S rRNA qPCR[276] (as applied here) or flow cytometry[137,262]. Flow cytometry protocols are available for almost all environmental compartments (e.g. wastewater[301], biofilms[302] or seawater[303]) to optimise cell detection. However, a qPCR method quantifying the same product as NGS might reduce bias of using different methods (flow cytometry of cells vs. qPCR of 16S rRNA). Ideally, the same primer region should be targeted to estimate cell counts and assess the microbiome. Recent advances now also allow the quantification of viable cells with digital PCR[304,305].

Diversity has been defined in so many different ways that its ability to transfer accurate information on microbial community changes, e.g. due to human impact, is compromised[266]. Jost and Chao (2020) introduced the analogy that diversity indices (e.g. Shannon or Simpson index) are connected to diversity in the same manner as a sphere's diameter is connected to its volume. While the diameter is an index of the sphere's volume, it is not the volume itself. They state that using the diameter instead of volume in engineering calculations would result in chaos, but this is what biologists are currently doing with diversity indices[299]. Shannon and Simpson index are useful diversity indices with an important role in ecology, but their values provide information on uncertainty and probability, respectively, rather than measuring diversity[265]. The Hill number framework provides a better and more unified approach to calculate and compare microbial diversities across environmental compartments, especially where the parameter q can be used to modulate the sensitivity towards abundant versus rare ASVs.

Depending on the study purpose, scientists might choose to calculate Hill numbers for several q for an in-depth diversity analysis (as performed here) or for one q value only. To define a core microbiome or when rare ASVs are considered untrustworthy due to technical bias (e.g. PCR or sequencing errors), q = 2 could be chosen to put more weight on abundant ASVs and results could be interpreted as effective number of dominant ASVs in the system[267,271]. In contrast, when the rarest ASVs are as important as the most abundant ASVs, for example for conservation purposes, q = 0 could be chosen[271]. The recently published R hilldiv package[274] enables DNA-based diversity calculations with Hill numbers.

In this study, we observed an increase in diversity and decrease in evenness along the river from a less polluted upstream to a more polluted downstream. Environmental AR increased along the river as indicated by the enrichment of ARGs and MGEs. The ARG concentrations measured up- and downstream in this Malaysian catchment are comparable to levels previously reported for two Chinese rivers, using the same HT-qPCR assay[135,225]. The downstream ARG concentrations in this study and Peng et al. (2020) are in the same range as HT-qPCR ARG concentrations reported for effluents from a Spanish and Chinese wastewater treatment plant[70,226].

The increase in diversity, together with the increasing levels of cell counts, ARGs and MGEs in this rural-to-urban catchment are likely caused by insufficiently treated

sewage entering the river (see Chapter 3). The most abundant ASVs for this catchment were *Cloacibacterium*, *Acinetobacter*, C39 (genus level) and *Comamonadaceae* (family level), also common in wastewater-impacted water bodies in China and India[306–308]. Comparing co-occurrence networks of absolute taxa with absolute ARG and MGE data allowed proposing hypothesis of possible taxa harbouring AR to be further investigated in experimental studies.

This study shows the straightforward and easy implementation of a quantitative microbial profiling approach and intuitive diversity characterisation with Hill numbers is superior to RMP approaches. We recommend our new combined approach become the norm for future environmental microbiome (and resistome) research, especially to underpin improved QMRAs. Only when such methods are employed will environmental AR studies become more quantitative and truly comparable.

Chapter 6. Conclusions and recommendations

6.1 Conclusions

AR is a global threat but the extend and implications are larger for LMICs. In LMICs, a combination of weak healthcare systems, a high prevalence of over-the-counter antibiotic sales, lacking sanitation and insufficiently treated waste accelerate the spread of diseases and AR. For the later, contaminated waters can enrich the environmental resistome, posing a health risk to locals relying on surface waters for fishing, washing clothes and irrigation. The WHO has proposed SE Asia as a hotspot for AR. In particular, clinical β-Lactam resistance is spreading in many SE Asian countries, including in Malaysia. Malaysia, the study site here, has one of the strongest, fastest-growing economies in SE Asia which has allowed more Malaysians to access healthcare, including antibiotics. Despite the majority of AR burden falling on LMICs, most environmental AR monitoring data sets are only available for HICs as the required field work is resource- and time-intensive.

Comprehensive environmental AR exposure assessments are crucial to investigate links between environmental and human health, and ultimately understand which types of exposure translate into health consequences. One Health thinking about AR is crucial to effectively reduce increasing numbers of deaths and disabilities caused by infections with AR pathogens in Malaysia and elsewhere. Where sporadic environmental AR data exists, it is difficult to translate findings from one study to other regions and even within the same region across seasons. Modelling represents an efficient, cost-effective tool for LMICs to identify AR hotspots in rivers and propose engineering and/or social interventions. However, while many watershed models exist, no standardised, hydrological model yet includes an AR component.

Conversely, many LMICs, including Malaysia, operate long-standing national river monitoring programmes, but these do not capture AR either. To address AR in surface waters in LMICs, novel cost-effective monitoring is required to develop predictive models that can ultimately, guide policy such as deciding where to improve existing or build new wastewater treatment plants.
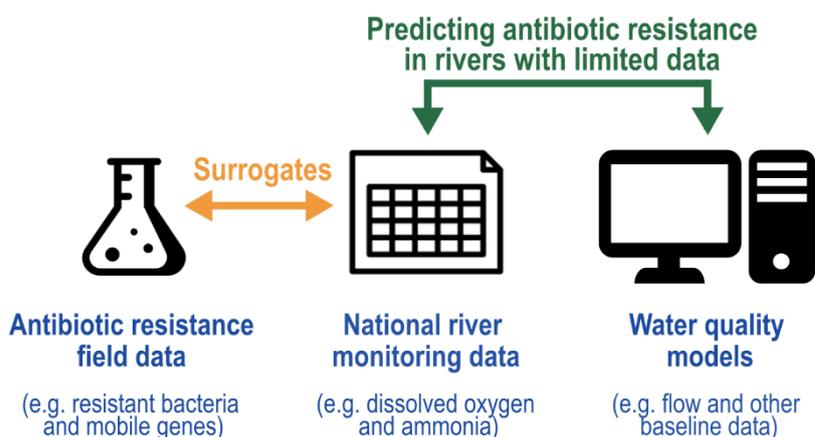
Figure 6-1. Graphical abstract describing the thesis concept.

Six main tasks were fulfilled by the work presented in Chapters 3-5.

1. Perform a comprehensive spatial and seasonal assessment of water quality and AR conditions in a Southeast Asian river catchment

Pinpointing environmental AR hotspots in LMICs is hindered by a lack of available and comparable AR monitoring data relevant to such settings. Addressing this problem, a comprehensive spatial and seasonal assessment of water quality and AR conditions in the Malaysian Skudai river catchment was performed to identify potential 'simple' surrogates that mirror elevated AR. Screening as conducted for β-lactam resistant coliforms, 22 antibiotics, 287 AR genes and integrons, and routine water quality parameters, covering absolute concentrations and mass loadings at eight sampling points across five sampling trips (n = 38).

Water quality conditions in the catchment were characterised by generally low DO, high COD and very high NH_3-N concentrations based on national Malaysian thresholds. Overall, water quality generally declined, and environmental AR levels increased as one moved downstream the catchment without major seasonal variations, except total antibiotic concentrations that were higher in the dry season (Cohen's D > 0.8, paired t-test, P < 0.05). Out of 14 measured antibiotics, only amoxicillin (all samples mean 510 ± SD 906 ng/L; max 3336 ng/L), and ciprofloxacin (all samples mean 131 ± SD 162 ng/L; max 705 ng/L) were detected above PNEC values. All ciprofloxacin and 50% of amoxicillin measurements in the dry season exceeded the PNEC thresholds.

Across the catchment, we observed an approximately one log10 difference between total coliform > ESBL coliform > carbapenem resistant coliform, meaning that ~ 10%

of total coliform produced ESBL and ~ 1% of total coliform were resistant to 2 µg/mL meropenem. ESBL *E. coli* concentrations increased from upstream $(<0.5 − 2) \times 10^1$ CFU/mL to downstream $(<0.1 – 5) \times 10^2$ CFU/mL.

211 different ARGs (~75% of assay) were detected in the river catchment with 70 ARGs (25% of assay) shared between all river water samples. Detected ARGs encoded resistance to eight classes of antibiotics, with β-lactam resistance being the most common (45 genes detected out of 52 assayed). River water ARG concentrations increased more than two $\log_{10}$ steps along the catchment with ARG copy numbers per cell increasing from 0.1 copies/cell upstream to 2.2 copies/cell downstream. 146 ARG and MGE concentrations increased significantly at least 10-fold between up- and downstream (Welch's t-test, $P < 0.05$). Four ARG concentrations encoding for aminoglycoside, MLSB and tetracycline resistance and integron 3 increased more than 1000-fold from up- to downstream.

The increase in ARG diversity was most apparent in the headwaters of the river. Movement from the rural (S1) to semi-urban (S2) locale added over 40 additional genes, many associated with faecal matter and multidrug resistance, such as *bla*CTX-M and *vanA.*

This study provides one of the most comprehensive assessments of relationships between 'routine' water quality monitoring data and AR markers in a river catchment in a LMIC. Understanding environmental AR exposures in LMICs such as Malaysia is crucial to holistically tackle AR to protect humans, animals and the environment.

2. Characterise water quality systematically by assessing both, pollutant concentration and pollutant mass loading data

Discharge and mass loadings are rarely estimated in environmental AR monitoring studies. However, this thesis shows that both concentration and load provide valuable complementary information to understand the processes occurring in a river catchment.

Mass loading data showed much greater transport of chemical and microbial pollutants along the Skudai river from the rural to urban locations. NH_3-N concentrations increased almost 100-fold from up- to downstream, but increases were > 14,000-fold greater based on NH_3-N mass loading data. Similarly, total coliform, ESBL coliform, carbapenem resistant coliform concentrations increased

from up- to downstream $10^0 - 10^1$-fold while their mass loadings increased $10^2 - 10^3$-fold.

Accounting for volumetric flow is particularly important for countries with dry and wet seasons. Comparing total antibiotic concentrations and mass loadings, we demonstrate that while antibiotic releases into the catchment likely do not vary across seasons for this catchment, reduced rainfall during the dry season resulted in increased river antibiotic concentrations and consequently, augmented exposure. Seasonality is expected to have a much larger effect on water quality/AR parameters in other SE Asian regions with more pronounced dry/wet seasons than here for southern peninsular Malaysia.

Measuring discharge in AR river studies is also crucial to allow a better use of these datasets for future AR watershed models.

3. **Introduce standardised 'effect sizes' to better understand relationships for AR monitoring and improve comparability of field studies**

Given the limited available environmental AR data in LMICs, results need to be reported in a consistent way to allow comparison to other studies and potential extrapolation to regions with no data. This thesis presented and applied a better way of determining significant effects in environmental AR studies. This was based on using 'standardised effect sizes' (here Cohen's D), rather than just P values, which not only determines whether significant differences exist between samples over space or time, but also weighs the scale of differences.

Using volcano plots provides an easy way to visualise seasonal and spatial effects together with P values to compare different water quality and AR parameters. For concentration data, the largest statistically significant spatial effects (up vs. downstream) were observed for ARG, MGE and DO concentrations (Cohen's D 15.6, -6.85, -6.5, respectively at Welch's t-test $P < 0.05$). Spatial effects were even larger for all parameters based on their mass loadings than concentrations.

To the best knowledge, this is the first study to apply the principle of standardised effect sizes to AR/river water quality monitoring. Effect sizes are easy to calculate and, unlike P values, provide a comparison independent of sample size. Moving forward, researchers are encouraged to routinely report effect sizes together with P values. This would allow comparison of studies different settings with different variables, which is especially critical for data scare LMICs.

4. Identify potential 'simple' water quality surrogates that mirror elevated AR

Limited data and expensive AR detection methods hinder LMICs to identify environmental sites of greatest AR concern. The aim of this study was to identify 'simple' easy-to-measure water quality surrogates that would aid monitoring and modelling of AR in locations with limited data.

For the Skudai catchment, the study found DO (Spearman's ρ 0.81, $P < 0.05$) and NH_3-N (Spearman's ρ 0.83, $P < 0.05$) to exhibit the strongest correlations with high total ARG concentrations. In this catchment, lower DO and higher NH_3-N are likely associated with insufficiently treated sewage entering the river, which is probably also the major route for ARGs entering the river.

Interestingly, within the AR indicators, total antibiotic concentrations exhibited the lowest correlations with other AR parameters. The weaker correlation of total antibiotics with the other AR parameters might be due to the fact that many antibiotics quickly degrade in the environment while some ARGs and ARBs persist for longer. This is not to say that low levels of antibiotics are incapable of influencing ARG selection in aquatic systems, but data here suggest untreated sewage inputs have much greater impact than in situ antibiotics on AR in catchments like the Skudai.

This study shows that simple water quality markers, like DO and NH_3-N, can be valuable surrogates for local stakeholders to identify AR hotspots in rivers and propose social and/or engineering interventions. This does not mean that they are universally applicable, such as near major non-sewage organic waste inputs. However, DO and NH_3-N clearly mirror sewage, which often dominates ARG and AR bacteria inputs, especially in LMIC rivers. DO and NH_3-N also are inexpensive to measure, exist in many numerical water quality models and already exist in current monitoring programmes.

This study provides a comprehensive picture of environmental exposures, including a surrogate way of identifying locations of potential increased exposure (i.e., exposure hot spots). This provides a strong starting point for One Health as part of the health protection system for Malaysia.

5. Utilise national water quality datasets and existing surface water quality models to estimate AR fate with 'simple' AR surrogates

Most LMICs do not have the resources to perform extensive field monitoring to identify AR hotspots. This thesis hypothesised that if one could link water quality markers often measured in regular national sampling programs and determine their relative value as surrogates for AR, one could massively expand the capacity of LMICs to identify points of intervention for AR mitigation. Surrogates would be especially valuable if they were amenable to numerical catchment modelling, which could extend predictions to places with limited or no data at all.

The field study found DO and NH_3-N to mirror ARG levels in the Skudai catchment. The DO and NH_3-N data also aligned well with the long-term national Malaysian dataset for the Skudai, suggesting correlations between these and AR markers could be used to extend existing Malaysian datasets to AR prediction.

A crude simple linear regression approach was able to well predict coliform concentrations with statistical significance, solely based on NH_3-N levels. This clear relationships between abiotic and biotic factors suggests that a hydrological model such as HSPF should be able to predict ARG concentrations based on DO or NH_3-N concentrations. HSPF is well suited to incorporate an environmental AR element as it allows to simulate plenty of hydrological and water quality processes, describing point and diffuse pollution sources.

A Skudai HSPF model up well captured seasonality of river flows and DO together with inter annual variability up to 2015. The intension was to expand the Skudai HSPF model to 2018, matching the available field AR data, but data issues due to incomplete flow record access delayed model development. Nevertheless, the concept to simulate ARG through a DO surrogate was demonstrated with a linear regression model. By capturing key hydrological and water quality processes, catchment models such as HSPF will further improve AR simulations via abiotic surrogates.

Such surrogate-based predictive monitoring approaches will not substitute for detailed local analysis, but they can be used to triage catchments for limited expense, allowing LMICs to focus resources on AR studies on places with potentially greater exposure risk.

6. Introduce quantitative microbiome profiling and unified Hill number diversities to enhance environmental microbiome and resistome research by providing more quantitative and representative data analyses

Understanding environmental microbiomes and AR is hindered by over reliance on relative abundance data from next-generation sequencing. Relative microbiome profiling (RMP) limits our ability to quantify changes in microbiomes and resistomes over space and time because sequencing depth and cell counts are not considered. Analysing relative abundance data using inappropriate statistical tools can yield up to 100% false detection rates and their application contributes to a general lack of reproducibility among microbiome studies.

This study combined quantitative microbiome profiling (QMP; parallelisation of amplicon sequencing and 16S rRNA qPCR to estimate cell counts) and absolute resistome profiling (based on high-throughput qPCR) to quantity AR along the anthropogenically impacted Skudai river. It was shown that QMP overcomes biases caused by relative taxa abundance data. Assessing the similarity of the rank order of the 100 most abundant ASVs, only 32% of the QMP and RMP results were found to be in common (score 0.32 with p = 95, focussing 86% of the weight on top 20 ASVs), suggesting the two methods providing different pictures of the system - RMP only provides composition, whereas QMP provides composition and abundance in tandem. Correlation analyses are often used to infer taxon-taxon interactions. Constructing RMP and QMP genus co-occurrence networks, a much larger number of significant co-varying genus pairs were detected in the QMP than RMP network (249 versus 116). The RMP network was dominated by negative correlations. None of the moderate to strong RMP correlations (P < 0.05, Spearman's ρ -0.5 to -1) were detected in the QMP correlation matrix. In environmental studies, cell counts are routinely measured and QMP can be conducted at no extra cost, requiring little bioinformatic workflow adjustments. The study contends that environmental researchers should use QMP instead of RMP to represent a more accurate picture of relationships between microbiomes and metadata (such as ARG concentrations) and guide future quantitative microbiological risk assessment (QMRA) applications. With these methods, QMRAs will become more precise and researchers might finally provide legitimate predictions of AR exposure in environmental settings.

The study also shows the benefits of using unified Hill number diversities to characterise environmental microbial communities. In this study, an increase in

diversity and decrease in evenness was observed along the river from less polluted upstream to more polluted downstream. The thesis also introduced diversity profiles to visualise the contributions of richness and evenness to a community's diversity. Characterising and comparing anthropogenic impacts on environmental microbiomes is generally hindered by the use of varying microbial diversity indices across studies. For a more meaningful quantification, 'diversity' needs to be unambiguously defined and applied in microbiome research. Hill numbers provide a statistical framework that unifies and generalizes popular indices, and are intuitive and flexible enough to address a wide range of scientific questions.

6.2 Recommendations for future work

1. A One Health approach is required to tackle AR in SE Asia and elsewhere. This study contributes through a comprehensive assessment of environmental AR exposure to provide a better platform for developing links between human and environmental health. In the future, more research is needed to understand which types of exposure lead to health consequences.

2. This study demonstrates a surrogate way of identifying locations of potential increased AR exposure. Future work should evaluate the applicability of easy-to-measure surrogates such as DO and NH_3-N in other catchments in Malaysia and SE Asia. More in-depth AR field studies are required to validate the use of existing national water quality datasets to predict AR hotspots. Collaborations with local stakeholders are crucial to identify the most applicable surrogate(s) for specific regions.

3. The modelling of AR at the catchment scale is in its infancy. There are issues with determining which processes need to be incorporated and how much detail, and how uncertainty can be quantified. This study contributes to the database that can be used for future model development and validation through the provision of open access data sets. However, there is an urgent need, if progress is to be made, for additional data sets to be collected. This should not only consider concentrations, but also river flow to allow calculation of mass loading.

4. Realistic QMRA is needed to quantify true risks of environmental AR exposures. This study shows the straightforward implementation of quantitative microbiome and resistome profiling to provide the required quantitative data for QMRAs. This new combined approach is recommended for future environmental microbiome (and resistome) research, especially to underpin improved QMRAs.

Figure 6-2. Final Skudai river water sample (S1) collected for this thesis.

References

1. Dantas, G. & Sommer, M. How to Fight Back Against Antibiotic Resistance. *Am. Sci.* **102**, 42 (2014). DOI:10.1511/2014.106.42.

2. Flemming, A. *Nobel Lecture Penicillin.* https://www.nobelprize.org/uploads/2018/06/fleming-lecture.pdf (1945).

3. Rammelkamp, C. H. & Maxon, T. Resistance of Staphylococcus aureus to the Action of Penicillin. *Exp. Biol. Med.* **51**, 386–389 (1942). DOI:10.3181/00379727-51-13986.

4. Lowy, F. D. Antimicrobial resistance : the example of Staphylococcus aureus. **111**, 1265–1273 (2003). DOI:10.1172/JCI200318535.In.

5. Urahn, S. K. *et al. A Scientific Roadmap for Antibiotic Discovery. The Pew Charitable Trusts* http://www.pewtrusts.org/en/research-and-analysis/reports/2016/05/a-scientific-roadmap-for-antibiotic-discovery (2016).

6. Silver, L. L. Challenges of Antibacterial Discovery. *Clin. Microbiol. Rev.* **24**, 71–109 (2011). DOI:10.1111/j.1749-6632.2010.05828.x.

7. Fugitt, R. B. & Luckenbaugh, R. W. United States Patent: 5-Halomethyl-3-phenylu- 2-oxazolidinones. 2–6 (1978).

8. Debono, M. *et al.* A21978C, a complex of new acidic peptide antibiotics. Isolation, chemistry, and mass spectral structure elucidation. *J. Antibiot. (Tokyo).* **40**, 761–777 (1987). DOI:10.7164/antibiotics.40.761.

9. WHO. *Antimicrobial resistance: global report on surveillance.* https://www.who.int/drugresistance/documents/surveillancereport/en/ (2014).

10. O'Neill, J. *Tackling drug-resistant infections globally: final report and recommendations.* https://wellcomecollection.org/works/thvwsuba (2016).

11. WHO. *Global Action Plan on Antimicrobial Resistance.* www.wpro.who.int/entity/drug_resistance/resources/global_action_plan_eng.pdf (2015).

12. A., M. S. *et al.* Antimicrobial Resistance: a One Health Perspective. *Microbiol. Spectr.* **6**, 6.2.10 (2018). DOI:10.1128/microbiolspec.ARBA-0009-2017.

13. Collignon, P. J. & McEwen, S. A. One health-its importance in helping to better control antimicrobial resistance. *Trop. Med. Infect. Dis.* **4**, (2019). DOI:10.3390/tropicalmed4010022.

14. McGowan, J. E. J. Antimicrobial Resistance in Hospital Organisms and Its Relation to Antibiotic Use. *Rev. Infect. Dis.* **5**, 1033–1048 (1983). DOI:10.1093/clinids/5.6.1033.

15. Kelly, R., Zoubiane, G., Walsh, D., Ward, R. & Goossens, H. Public funding for research on antibacterial resistance in the JPIAMR countries, the European Commission, and related European Union agencies: A systematic observational analysis. *Lancet Infect. Dis.* **16**, 431–440 (2016). DOI:10.1016/S1473-3099(15)00350-3.

16. UNESCO. *Wastewater: the untapped resource*. https://unesdoc.unesco.org/ark:/48223/pf0000247153 (2017).

17. Xu, J. *et al.* Occurrence of antibiotics and antibiotic resistance genes in a sewage treatment plant and its effluent-receiving river. *Chemosphere* **119**, 1379–1385 (2015). DOI:10.1016/j.chemosphere.2014.02.040.

18. Oberoi, A. S., Jia, Y., Zhang, H., Khanal, S. K. & Lu, H. Insights into the Fate and Removal of Antibiotics in Engineered Biological Treatment Systems: A Critical Review. *Environ. Sci. Technol.* **53**, 7234–7264 (2019). DOI:10.1021/acs.est.9b01131.

19. Manaia, C. M. *et al.* Antibiotic resistance in wastewater treatment plants: Tackling the black box. *Environ. Int.* **115**, 312–324 (2018). DOI:10.1016/j.envint.2018.03.044.

20. Anderson, M. *et al. Health systems and policy analysis - Averting the AMR crisis : What are the avenues for policy. European Observatory on Health Systems and Policies* https://www.ncbi.nlm.nih.gov/books/NBK543406/ (2019).

21. Graham, D. W., Collignon, P., Davies, J., Larsson, D. G. J. & Snape, J. Underappreciated role of regionally poor water quality on globally increasing antibiotic resistance. *Environ. Sci. Technol.* **48**, 11746–11747 (2014). DOI:10.1021/es504206x.

22. Coker, R. J., Hunter, B. M., Rudge, J. W., Liverani, M. & Hanvoravongchai, P. Emerging infectious diseases in southeast Asia: Regional challenges to control. *Lancet* **377**, 599–609 (2011). DOI:10.1016/S0140-6736(10)62004-1.

23. Walther, B. A. *et al.* Biodiversity and health: Lessons and recommendations from an interdisciplinary conference to advise Southeast Asian research, society and policy. *Infect. Genet. Evol.* **40**, 29–46 (2016). DOI:10.1016/j.meegid.2016.02.003.

24. Zellweger, M., Carrique-mas, J., Limmathurotsakul, D., Day, N. P. J. & Thwaites, G. E. A current perspective on antimicrobial resistance in Southeast Asia. *Antimicrob. Chemother.* **72**, 2963–2972 (2017). DOI:10.1093/jac/dkx260.

25. Chereau, F., Opatowski, L., Tourdjman, M. & Vong, S. Risk assessment for antibiotic resistance in South East Asia. *BMJ* **358**, 2–8 (2017). DOI:10.1136/bmj.j3393.

26. Liu, Y. Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *Lancet Infect. Dis.* **16**, 161–168 (2016). DOI:10.1016/S1473-3099(15)00424-7.

27. Vikesland, P. *et al.* Differential Drivers of Antimicrobial Resistance across the World. *Acc. Chem. Res.* **52**, 916–924 (2019). DOI:10.1021/acs.accounts.8b00643.

28. Huijbers, P. M. C., Flach, C.-F. & Larsson, D. G. J. A conceptual framework for the environmental surveillance of antibiotics and antibiotic resistance. *Environ. Int.* **130**, 104880 (2019). DOI:10.1016/j.envint.2019.05.074.

29. Birkegård, A. C., Halasa, T., Toft, N., Folkesson, A. & Græsbøll, K. Send more data: a systematic review of mathematical models of antimicrobial resistance. *Antimicrob. Resist. Infect. Control* **7**, 117 (2018). DOI:10.1186/s13756-018-0406-1.

30.   UNICEF. Malaysia. http://cop21.unicef.fr/country/malaysia/ (2019).

31.   D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011). DOI:10.1038/nature10388.

32.   Perron, G. G. *et al.* Functional characterization of bacteria isolated from ancient arctic soil exposes diverse resistance mechanisms to modern antibiotics. *PLoS One* **10**, 1–19 (2015). DOI:10.1371/journal.pone.0069533.

33.   Bhullar, K. *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS One* **7**, 1–11 (2012). DOI:10.1371/journal.pone.0034953.

34.   von Wintersdorff, D. of A. R. in M. E. through H. G. T. J. H. *et al.* Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Front. Microbiol.* **7**, 173 (2016). DOI:10.3389/fmicb.2016.00173.

35.   Knapp, C. W., Dolfing, J., Ehlert, P. A. I. & Graham, D. W. Evidence of Increasing Antibiotic Resistance Gene Abundances in Archived Soils since 1940. *Environ. Sci. Technol.* **44**, 580–587 (2010). DOI:10.1021/es901221x.

36.   Bürgmann, H. *et al.* Water and sanitation: An essential battlefront in the war on antimicrobial resistance. *FEMS Microbiol. Ecol.* **94**, 1–14 (2018). DOI:10.1093/femsec/fiy101.

37.   Rossolini, G. M., D'Andrea, M. M. & Mugnaioli, C. The spread of CTX-M-type extended-spectrum β-lactamases. *Clin. Microbiol. Infect.* **14**, 33–41 (2008). DOI:10.1111/j.1469-0691.2007.01867.x.

38.   Poirel, L., Kämpfer, P. & Nordmann, P. Chromosome-encoded ambler class a β-lactamase of Kluyvera georgiana, a probable progenitor of a subgroup of CTX-M extended-spectrum β-lactamases. *Antimicrob. Agents Chemother.* **46**, 4038–4040 (2002). DOI:10.1128/AAC.46.12.4038-4040.2002.

39.   Poirel, L., Rodriguez-Martinez, J.-M., Mammeri, H., Liard, A. & Nordmann, P. Origin of Plasmid-Mediated Quinolone Resistance Determinant QnrA. *Antimicrob. Agents Chemother.* **49**, 3523–3252 (2005). DOI:10.1128/AAC.49.8.3523.

40.   Humeniuk, C. *et al.* β-lactamases of Kluyvera ascorbata, probable progenitors of some plasmid-encoded CTX-M types. *Antimicrob. Agents Chemother.* **46**, 3045–3049 (2002). DOI:10.1128/AAC.46.9.3045-3049.2002.

41.   Lukjancenko, O., Wassenaar, T. M. & Ussery, D. W. Comparison of 61 Sequenced Escherichia coli Genomes. *Microb. Ecol.* **60**, 708–720 (2010). DOI:10.1007/s00248-010-9717-3.

42.   Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. T. Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Res.* **16**, 1099–1108 (2006). DOI:10.1101/gr.5322306.

43.   Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. Biased biological functions of horizontally-transferred genes in prokaryotic genomes. *Nat. Genet.* **36**, 760–766

(2004). DOI:10.1038/ng1381.

44. Lang, A. S., Zhaxybayeva, O. & Beatty, J. T. Gene transfer agents: Phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* **10**, 472–482 (2012). DOI:10.1038/nrmicro2802.

45. Lopatkin, A. J. *et al.* Antibiotics as a selective driver for conjugation dynamics. *Nat. Microbiol.* **1**, 1–8 (2016). DOI:10.1038/nmicrobiol.2016.44.

46. Davison, J. Genetic exchange between bacteria in the environment. *Plasmid* **42**, 73–91 (1999). DOI:10.1006/plas.1999.1421.

47. Brown-Jaque, M., Calero-Cáceres, W. & Muniesa, M. Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid* **79**, 1–7 (2015). DOI:10.1016/j.plasmid.2015.01.001.

48. Singer, A. C., Shaw, H., Rhodes, V. & Hart, A. Review of antimicrobial resistance in the environment and its relevance to environmental regulators. *Frontiers in Microbiology* vol. 7 (2016). DOI:10.3389/fmicb.2016.01728.

49. Stanton, T. B., Humphrey, S. B., Sharma, V. K. & Zuerner, R. L. Collateral effects of antibiotics: Carbadox and metronidazole induce VSH-I and facilitate gene transfer among Brachyspira hyodysenteriae strains. *Appl. Environ. Microbiol.* **74**, 2950–2956 (2008). DOI:10.1128/AEM.00189-08.

50. Lundström, S. V. *et al.* Minimal selective concentrations of tetracycline in complex aquatic bacterial biofilms. *Sci. Total Environ.* **553**, 587–595 (2016). DOI:10.1016/j.scitotenv.2016.02.103.

51. Murray, A. K. *et al.* Novel insights into selection for antibiotic resistance in complex microbial communities. *MBio* **9**, 1–12 (2018). DOI:10.1128/mBio.00969-18.

52. Almagor, J., Temkin, E., Benenson, I., Fallach, N. & Carmeli, Y. The impact of antibiotic use on transmission of resistant bacteria in hospitals: Insights from an agent-based model. *PLoS One* **13**, 1–14 (2018). DOI:10.1371/journal.pone.0197111.

53. Bengtsson-Palme, J. *et al.* Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. *Sci. Total Environ.* **572**, 697–712 (2016). DOI:10.1016/j.scitotenv.2016.06.228.

54. Karkman, A. *et al.* High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS Microbiol. Ecol.* **92**, 1–7 (2016). DOI:10.1093/femsec/fiw014.

55. Gao, P., Munir, M. & Xagoraraki, I. Correlation of tetracycline and sulfonamide antibiotics with corresponding resistance genes and resistant bacteria in a conventional municipal wastewater treatment plant. *Sci. Total Environ.* **421**–**422**, 173–183 (2012). DOI:10.1016/j.scitotenv.2012.01.061.

56. Karkman, A., Pärnänen, K. & Larsson, D. G. J. Fecal pollution explains antibiotic resistance gene abundances in anthropogenically impacted environments. *Nat.*

Commun. 341487 (2018) doi:10.1101/341487. DOI:10.1101/341487.

57. Manaia, C. M., Macedo, G., Fatta-Kassinos, D. & Nunes, O. C. Antibiotic resistance in urban aquatic environments: can it be controlled? *Appl. Microbiol. Biotechnol.* **100**, 1543–1557 (2016). DOI:10.1007/s00253-015-7202-0.

58. Wright, G. D. Q&A: Antibiotic resistance: Where does it come from and what can we do about it? *BMC Biol.* **8**, (2010). DOI:10.1186/1741-7007-8-123.

59. Czekalski, N., Gascón Díez, E. & Bürgmann, H. Wastewater as a point source of antibiotic-resistance genes in the sediment of a freshwater lake. *ISME J.* **8**, 1381–90 (2014). DOI:10.1038/ismej.2014.8.

60. Rodriguez-Mozaz, S. *et al.* Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. *Water Res.* **69**, 234–242 (2015). DOI:10.1016/j.watres.2014.11.021.

61. Pärnänen, K. M. M. *et al.* Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Sci. Adv.* **5**, (2019). DOI:10.1126/sciadv.aau9124.

62. Vaz-Moreira, I., Nunes, O. C. & Manaia, C. M. Bacterial diversity and antibiotic resistance in water habitats: Searching the links with the human microbiome. *FEMS Microbiol. Rev.* **38**, 761–778 (2014). DOI:10.1111/1574-6976.12062.

63. Defra. *Sewage treatment in the UK*. vol. 2 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/69582/pb6655-uk-sewage-treatment-020424.pdf (2002).

64. Singer, A. C. *et al.* Assessing the ecotoxicologic hazards of a pandemic influenza medical response. *Environ. Health Perspect.* **119**, 1084–1090 (2011). DOI:10.1289/ehp.1002757.

65. Zhang, Q. Q., Ying, G. G., Pan, C. G., Liu, Y. S. & Zhao, J. L. Comprehensive evaluation of antibiotics emission and fate in the river basins of China: Source analysis, multimedia modeling, and linkage to bacterial resistance. *Environ. Sci. Technol.* **49**, 6772–6782 (2015). DOI:10.1021/acs.est.5b00729.

66. Verlicchi, P. & Zambello, E. Predicted and measured concentrations of pharmaceuticals in hospital effluents. Examination of the strengths and weaknesses of the two approaches through the analysis of a case study. *Sci. Total Environ.* **565**, 82–94 (2016). DOI:10.1016/j.scitotenv.2016.04.165.

67. Hiller, C. X., Hübner, U., Fajnorova, S., Schwartz, T. & Drewes, J. E. Antibiotic microbial resistance (AMR) removal efficiencies by conventional and advanced wastewater treatment processes: A review. *Sci. Total Environ.* **685**, 596–608 (2019). DOI:10.1016/j.scitotenv.2019.05.315.

68. Water Research Australia. *Fact sheet: Log removal values in wastewater treatment*. https://www.waterra.com.au/publications/document-search/?download=949 (2014).

69. Narciso-da-Rocha, C. *et al.* Bacterial lineages putatively associated with the dissemination of antibiotic resistance genes in a full-scale urban wastewater treatment plant. *Environ. Int.* **118**, 179–188 (2018). DOI:10.1016/j.envint.2018.05.040.

70. Quintela-Baluja, M. *et al.* Spatial ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. *Water Res.* **162**, 347–357 (2019). DOI:10.1016/j.watres.2019.06.075.

71. Manaia, C. M. Assessing the Risk of Antibiotic Resistance Transmission from the Environment to Humans: Non-Direct Proportionality between Abundance and Risk. *Trends Microbiol.* **25**, 173–181 (2017). DOI:10.1016/j.tim.2016.11.014.

72. Leonard, A. F. C. *et al.* Exposure to and colonisation by antibiotic-resistant E. coli in UK coastal water users: Environmental surveillance, exposure assessment, and epidemiological study (Beach Bum Survey). *Environ. Int.* **114**, 326–333 (2018). DOI:10.1016/j.envint.2017.11.003.

73. Klein, E. Y. *et al.* Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proc. Natl. Acad. Sci.* **115**, E3463–E3470 (2018). DOI:10.1073/pnas.1717295115.

74. WHO. WHO publishes list of bacteria for which new antibiotics are urgently needed. http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/ (2017).

75. World Bank. GDP per capita (current US$). https://data.worldbank.org/indicator/NY.GDP.PCAP.CD (2019).

76. The Center for Disease Dynamics Economics & Policy. ResistanceMap: Antibiotic resistance. https://resistancemap.cddep.org/AntibioticResistance.php (2019).

77. Graham, D. *et al.* Strategic Approach for Prioritising Local and Regional Sanitation Interventions for Reducing Global Antibiotic Resistance. *Water* **11**, 27 (2018). DOI:10.3390/w11010027.

78. Hu, Y. J., Ogyu, A., Cowling, B. J., Fukudaa, K. & Panga, H. Available evidence of antibiotic resistance from extended-spectrum ß-lactamase-producing enterobacteriaceae in paediatric patients in 20 countries: A systematic review and meta-analysis. *Bull. World Health Organ.* **97**, 486-501B (2019). DOI:10.2471/BLT.18.225698.

79. Anh, H. Q. *et al.* Antibiotics in surface water of East and Southeast Asian countries: A focused review on contamination status, pollution sources, potential risks, and future perspectives. *Sci. Total Environ.* **764**, 142865 (2021). DOI:10.1016/j.scitotenv.2020.142865.

80. WHO-UNICEF. *Launch version July 12 2017 Progress on Drinking Water , Sanitation and Hygiene.* https://www.who.int/mediacentre/news/releases/2017/launch-version-report-jmp-water-sanitation-hygiene.pdf (2017).

81. World Bank. The World Bank in Malaysia.
https://www.worldbank.org/en/country/malaysia/overview (2020).

82. Ab Rahman, N., Teng, C. L. & Sivasampu, S. Antibiotic prescribing in public and private practice: a cross-sectional study in primary care clinics in Malaysia. *BMC Infect. Dis.* **16**, 208 (2016). DOI:10.1186/s12879-016-1530-2.

83. Ministry of Health Malaysia. *Malaysian Action Plan on Antimicrobial Resistance (MyAP-AMR) 2017-2021. Ministry of Health Malaysia* https://www.moh.gov.my/moh/resources/Penerbitan/Garis Panduan/Garis panduan Umum (Awam)/National_Action_Plan_-_FINAL_29_june.pdf (2017).

84. Antibiotic Resistance Surveillance Reference Laboratory. *National Antibiotic Resistance Surveillance Report 2019*.
https://www.imr.gov.my/images/uploads/NSAR/2019/NSAR_report_2019_03092020.pdf (2019).

85. Institute for Health Metrics and Evaluation (IHME). GBD Compare.
http://vizhub.healthdata.org/gbd-compare (2019).

86. Japan Sanitation Consortium. *Country sanitation assessment in Malaysia report*.
http://www.jsanic.org/publications/Country_Survey_Reports/Malaysia/JSC_Malaysia_ Sanitation_Assessment_Report.pdf (2011).

87. Department of Statistics Malaysia. Current Population Estimates, Malaysia 2016-2017.
https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul_id=a1d 1UTFZazd5ajJiRWFHNDduOXFFQT09&menu_id=L0pheU43NWJwRWVSZklWdzQ4 TlhUUT09 (2017).

88. Government of Malaysia. *Federal Constitution 1957 (Reprint 2010)*.
http://www.agc.gov.my/agcportal/uploads/files/Publications/FC/Federal Consti (BI text).pdf (2010) doi:JW516221 18-09-2010. DOI:JW516221 18-09-2010.

89. Department of Statistics Malaysia. Population Distribution and Basic Demographic Characteristic Report 2010 (Updated: 05/08/2011).
https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=117&bul_id=MD MxdHZjWTk1SjFzTzNkRXYzcVZjdz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4Tl hUUT09 (2011).

90. IWK. Sewage. https://www.iwk.com.my/do-you-know/sewage-treatment-methods (2017).

91. WHO & UNICEF. *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities. Launch version July 12 Main report Progress on Drinking Water , Sanitation and Hygiene* (2019).

92. WEPA. *Outlook on water environmental management in Asia*. http://wepa-db.net/en/publication/2018_outlook/wepa_outlook_report_2018_en.pdf (2018).

93. Daud, H. *Water Quality (River) Monitoring System/ Programme and Pollution Control*.

Department of Environment http://www.wepa-db.net/pdf/0810malaysia/f.pdf (2006).

94. Department of Environment. *Malaysia environmental quality report- river water quality*. https://www.doe.gov.my/portalv1/en/info-laporan/laporan-jabatan-alam-sekitar/laporan-kualiti-alam-keliling/324265 (2015).

95. Alhaj, N., Mariana, N. S., Raha, A. R. & Ishak, Z. Prevalence of Antibiotic Resistance among Escherichia coli from Different Sources in Malaysia. *Int. J. Poult. Sci.* **6**, 293–297 (2007). DOI:10.3923/ijps.2007.293.297.

96. Ghaderpour, A. *et al.* Diverse and abundant multi-drug resistant E. coli in Matang mangrove estuaries, Malaysia. *Front. Microbiol.* **6**, (2015). DOI:10.3389/fmicb.2015.00977.

97. Ho, J. Y. *et al.* Multidrug-resistant bacteria and microbial communities in a river estuary with fragmented suburban waste management. *J. Hazard. Mater.* **405**, 124687 (2020). DOI:10.1016/j.jhazmat.2020.124687.

98. JPIAMR. *Strategic research and innovation agenda on antimicrobial resistance*. https://www.jointprogramming.nl/en/show/IN_JPIAMR_SRIA.htm (2019).

99. Sullivan, G. M. & Feinn, R. Using Effect Size—or Why the P Value Is Not Enough . *J. Grad. Med. Educ.* **4**, 279–282 (2012). DOI:10.4300/jgme-d-12-00156.1.

100. Thongsamer, T. *et al.* Environmental antimicrobial resistance is associated with faecal pollution in Central Thailand's coastal aquaculture region. *J. Hazard. Mater.* **416**, 125718 (2021). DOI:10.1016/j.jhazmat.2021.125718.

101. González-Plaza, J. J. *et al.* Antibiotic-manufacturing sites are hot-spots for the release and spread of antibiotic resistance genes and mobile genetic elements in receiving aquatic environments. *Environ. Int.* **130**, 104735 (2019). DOI:10.1016/j.envint.2019.04.007.

102. Bielen, A. *et al.* Negative environmental impacts of antibiotic-contaminated effluents from pharmaceutical industries. *Water Res.* **126**, 79–87 (2017). DOI:10.1016/j.watres.2017.09.019.

103. Larsson, D. G. J., de Pedro, C. & Paxeus, N. Effluent from drug manufactures contains extremely high levels of pharmaceuticals. *J. Hazard. Mater.* **148**, 751–755 (2007). DOI:10.1016/j.jhazmat.2007.07.008.

104. Fekadu, S., Alemayehu, E., Dewil, R. & Van der Bruggen, B. Pharmaceuticals in freshwater aquatic environments: A comparison of the African and European challenge. *Sci. Total Environ.* **654**, 324–337 (2019). DOI:10.1016/j.scitotenv.2018.11.072.

105. Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *Lancet Planet. Heal.* **2**, e398–e405 (2018). DOI:10.1016/S2542-5196(18)30186-4.

106. Tell, J. *et al.* Science-based Targets for Antibiotics in Receiving Waters from Pharmaceutical Manufacturing Operations. *Integr. Environ. Assess. Manag.* **15**, 312–319 (2019). DOI:10.1002/ieam.4141.

107. Bengtsson-Palme, J. & Larsson, D. G. J. Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. *Environ. Int.* **86**, 140–149 (2016). DOI:10.1016/j.envint.2015.10.015.

108. WHO. *Critically Important Antimicrobials for Human Medicine-5th Revision. World Health Organization* https://www.who.int/foodsafety/publications/antimicrobials-fifth/en/ (2017).

109. Lamba, M., Graham, D. W. & Ahammad, S. Z. Hospital Wastewater Releases of Carbapenem-Resistance Pathogens and Genes in Urban India. *Environ. Sci. Technol.* **51**, 13906–13912 (2017). DOI:10.1021/acs.est.7b03380.

110. Berendsen, B. J. A., Wegh, R. S., Memelink, J., Zuidema, T. & Stolker, L. A. M. The analysis of animal faeces as a tool to monitor antibiotic usage. *Talanta* **132**, 258–268 (2015). DOI:10.1016/j.talanta.2014.09.022.

111. Dinh, Q. T. *et al.* Fate of antibiotics from hospital and domestic sources in a sewage network. *Sci. Total Environ.* **575**, 758–766 (2017). DOI:10.1016/j.scitotenv.2016.09.118.

112. Lindberg, R. H., Wennberg, P., Johansson, M. I., Tysklind, M. & Andersson, B. A. V. Screening of human antibiotic substances and determination of weekly mass flows in five sewage treatment plants in Sweden. *Environ. Sci. Technol.* **39**, 3421–3429 (2005). DOI:10.1021/es048143z.

113. Berendonk, T. U. *et al.* Tackling antibiotic resistance: The environmental framework. *Nat. Rev. Microbiol.* **13**, 310–317 (2015). DOI:10.1038/nrmicro3439.

114. Matheu, J., Aidara-Kane, A. & Andremont, A. The ESBL tricycle AMR surveillance project: a simple, one health approach to global surveillance. http://resistancecontrol.info/2017/the-esbl-tricycle-amr-surveillance-project-a-simple-one-health-approach-to-global-surveillance/ (2017).

115. WHO. *Global Antimicrobial Resistance Surveillance System (GLASS) Report: Early Implementation 2020. World Health Organization* https://apps.who.int/iris/bitstream/handle/10665/332081/9789240005587-eng.pdf?ua=1 (2020).

116. Savichtcheva, O. & Okabe, S. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. *Water Res.* **40**, 2463–2476 (2006). DOI:10.1016/j.watres.2006.04.040.

117. Pruden, A. *et al.* An Environmental Science and Engineering Framework for Combating Antimicrobial Resistance. *Environ. Eng. Sci.* **35**, 1005–1011 (2018).

DOI:10.1089/ees.2017.0520.

118. Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995). DOI:10.1128/MMBR.59.1.143-169.1995.

119. Vaz-Moreira, I., Egas, C., Nunes, O. C. & Manaia, C. M. Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. *FEMS Microbiol. Ecol.* **83**, 361–374 (2013). DOI:10.1111/1574-6941.12002.

120. Su, J.-Q. *et al.* Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. *Microbiome* **5**, 84 (2017). DOI:10.1186/s40168-017-0298-y.

121. Hendriksen, R. S. *et al.* Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **10**, (2019). DOI:10.1038/s41467-019-08853-3.

122. Wetterstrand, K. A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data (2019).

123. Jia, B. *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017). DOI:10.1093/nar/gkw1004.

124. Yang, Y. *et al.* ARGs-OAP: Online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. *Bioinformatics* **32**, 2346–2351 (2016). DOI:10.1093/bioinformatics/btw136.

125. Arango-Argoty, G. *et al.* DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 1–15 (2018). DOI:10.1186/s40168-018-0401-z.

126. Zhang, A. *et al.* Choosing Your Battles: Which Resistance Genes Warrant Global Action? *bioRxiv - Prepr.* (2019) doi:10.1101/784322. DOI:10.1101/784322.

127. Van Der Helm, E. *et al.* Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res.* **45**, 1–8 (2017). DOI:10.1093/nar/gkw1328.

128. Acharya, K. *et al.* A comparative assessment of conventional and molecular methods, including MinION nanopore sequencing, for surveying water quality. *Sci. Rep.* **9**, 1–11 (2019). DOI:10.1038/s41598-019-51997-x.

129. Acharya, K. *et al.* Metagenomic water quality monitoring with a portable laboratory. *Water Res.* **184**, 116112 (2020). DOI:10.1016/j.watres.2020.116112.

130. Waseem, H. *et al.* Contributions and Challenges of High Throughput qPCR for Determining Antimicrobial Resistance in the Environment: A Critical Review. *Molecules* **24**, 163 (2019). DOI:10.3390/molecules24010163.

131. Resistomap. Antibiotic resistance gene profiling service. https://resistomap.com/services (2020).

132. Hultman, J. *et al.* Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. *FEMS Microbiol. Ecol.* **94**, 1–10 (2018). DOI:10.1093/femsec/fiy038.

133. Girones, R. *et al.* Molecular detection of pathogens in water - The pros and cons of molecular techniques. *Water Res.* **44**, 4325–4339 (2010). DOI:10.1016/j.watres.2010.06.030.

134. Li, B. *et al.* Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J.* **9**, 2490–2502 (2015). DOI:10.1038/ismej.2015.59.

135. Peng, F. *et al.* Urbanization drives riverine bacterial antibiotic resistome more than taxonomic community at watershed scale. *Environ. Int.* **137**, 105524 (2020). DOI:10.1016/j.envint.2020.105524.

136. Gloor, G., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **8**, 1–6 (2017). DOI:10.3389/fmicb.2017.02224.

137. Vandeputte, D. *et al.* Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017). DOI:10.1038/nature24460.

138. Gillings, M. R. *et al.* Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J.* **9**, 1269–1279 (2015). DOI:10.1038/ismej.2014.226.

139. Opatowski, L., Guillemot, D., Boëlle, P. & Temime, L. Contribution of mathematical modeling to the fight against bacterial antibiotic resistance. *Curr Opin Infect Dis.* **24**, 279–287 (2011). DOI:10.1097/QCO.0b013e3283462362.

140. Spicknall, I. H., Foxman, B., Marrs, C. F. & Eisenberg, J. N. S. A modeling framework for the evolution and spread of antibiotic resistance: Literature review and model categorization. *Am. J. Epidemiol.* **178**, 508–520 (2013). DOI:10.1093/aje/kwt017.

141. Temime, L., Hejblum, G., Setbon, M. & Valleron, A. J. The rising impact of mathematical modelling in epidemiology: Antibiotic resistance research as a case study. *Epidemiol. Infect.* **136**, 289–298 (2008). DOI:10.1017/S0950268807009442.

142. Niewiadomska, A. M. *et al.* Population-level mathematical modeling of antimicrobial resistance: A systematic review. *BMC Med.* **17**, 1–20 (2019). DOI:10.1186/s12916-019-1314-9.

143. WHO, FAO & OIE. *Antimicrobial Resistance: A Manual for Developing National Action Plans.* https://apps.who.int/iris/handle/10665/204470 (2016).

144. Baker, M., Hobman, J. L., Dodd, C. E. R., Ramsden, S. J. & Stekel, D. J. Mathematical modelling of antimicrobial resistance in agricultural waste highlights importance of gene transfer rate. *FEMS Microbiol. Ecol.* **92**, 1–10 (2016).

DOI:10.1093/femsec/fiw040.

145. Hellweger, F. L. Simple Model of Tetracycline Antibiotic Resistance in Aquatic Environment: Accounting for Metal Coselection. *J. Environ. Eng.* **139**, 913–921 (2013). DOI:10.1061/(ASCE)EE.1943-7870.0000696.

146. Hellweger, F. L., Ruan, X. & Sanchez, S. A simple model of tetracycline antibiotic resistance in the aquatic environment (with application to the Poudre River). *Int. J. Environ. Res. Public Health* **8**, 480–497 (2011). DOI:10.3390/ijerph8020480.

147. Mostefaoui, I. M. Mathematical analysis of a model describing the number of antibiotic resistant bacteria in a polluted river. *Math. Methods Appl. Sci.* **37**, 1956–1973 (2014). DOI:10.1002/mma.2949.

148. Ramsay, D. E. *et al.* Application of dynamic modelling techniques to the problem of antibacterial use and resistance: A scoping review. *Epidemiol. Infect.* **146**, 2014–2027 (2018). DOI:10.1017/S0950268818002091.

149. Talaminos, A. *et al.* Modelling the epidemiology of Escherichia coli ST131 and the impact of interventions on the community and healthcare centres. *Epidemiol. Infect.* **144**, 1974–1982 (2016). DOI:10.1017/S0950268816000030.

150. Domenech de Cellès, M., Zahar, J. R., Abadie, V. & Guillemot, D. Limits of patient isolation measures to control extended-spectrum beta-lactamase-producing Enterobacteriaceae: model-based analysis of clinical data in a pediatric ward. *BMC Infect. Dis.* **13**, (2013). DOI:10.1186/1471-2334-13-187.

151. Caminade, C., McIntyre, K. M. & Jones, A. E. Impact of recent and future climate change on vector-borne diseases. *Ann. N. Y. Acad. Sci.* **1436**, 157–173 (2019). DOI:10.1111/nyas.13950.

152. MacFadden, D. R., McGough, S. F., Fisman, D., Santillana, M. & Brownstein, J. S. Antibiotic resistance increases with local temperature. *Nat. Clim. Chang.* **8**, 510–514 (2018). DOI:10.1038/s41558-018-0161-6.

153. Benedini, M., Tsakiris, G. & Water. *Water Quality Modelling for Rivers and Streams.* (Springer, 2013).

154. Gothwal, R. & Thatikonda, S. Mathematical model for the transport of fluoroquinolone and its resistant bacteria. *Environ. Sci. Pollut. Res. Vol.* **25**, 20439–20452 (2018). DOI:10.1007/s11356-017-9848-x.

155. Lawrence, B., Mummert, A. & Somerville, C. A Model of the Number of Antibiotic Resistant Bacteria in Rivers. *Preprint* 1–18 (2010).

156. Becker, A. & Serban, P. *Hydrological Models for Water-Resources System Design and Operation.* https://library.wmo.int/doc_num.php?explnum_id=1696 (1990).

157. FISRWG. *Stream Corridor Restoration: Principles, Processes, and Practices. Federal Interagency Stream Restoration Working Group* https://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/national/water/?cid=stelprdb10432

44 (1998).

158. The COMET Program. Runoff Processes: International Edition. http://download.comet.ucar.edu/memory-stick/hydro/basic_int/runoff/print.htm (2010).

159. Vellidis, G., Barnes, P., Bosch, D. D. & Cathey, A. M. Mathematical simulation tools for developing dissolved oxygen TMDLs. *ASABE* **49**, 1003–1022 (2006).

160. Daniel, E. B. *et al.* Watershed Modeling and its Applications: A State-of-the-Art Review. *Open Hydrol. J.* **5**, 26–50 (2011). DOI:10.2174/1874378101105010026.

161. Bicknell, B. R., Imhoff, J. C., Kittle, J. L. J., Jobes, T. H. & Donigian, A. S. J. *HSPF Version 12 User's Manual*. *EPA* vol. 18 https://www.casqa.org/sites/default/files/hspf-12.pdf (2001).

162. Yang, Y. & Wang, L. A Review of Modelling Tools for Implementation of the EU Water Framework Directive in Handling Diffuse Water Pollution. *Water Resour. Manag.* **24**, 1819–1843 (2010). DOI:10.1007/s11269-009-9526-y.

163. Houston Engineering Inc. *Buffalo River Watershed HSPF Modeling*. https://www.pca.state.mn.us/sites/default/files/wq-iw5-06q.pdf (2013).

164. Fonseca, A., Botelho, C., Boaventura, R. A. R. & Vilar, V. J. P. Integrated hydrological and water quality model for river management: A case study on Lena River. *Sci. Total Environ.* **485**–**486**, 474–489 (2014). DOI:10.1016/j.scitotenv.2014.03.111.

165. Shirinian-Orlando, A. A. & Uchrin, C. G. Modeling the hydrology and water quality using BASINS/HSPF for the upper Maurice River watershed, New Jersey. *J. Environ. Sci. Heal. - Part A Toxic/Hazardous Subst. Environ. Eng.* **42**, 289–303 (2007). DOI:10.1080/10934520601134254.

166. Bello, A. D., Hashim, N. & Mohd Haniffah, M. Predicting Impact of Climate Change on Water Temperature and Dissolved Oxygen in Tropical Rivers. *Climate* **5**, 58 (2017). DOI:10.3390/cli5030058.

167. Sitterson, J. *et al. An Overview of Rainfall-Runoff Model Types An Overview of Rainfall-Runoff Model Types*. *EPA* https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=339328&Lab=NERL (2017).

168. Michael, C. A., Dominey-Howes, D. & Labbate, M. The Antimicrobial Resistance Crisis: Causes, Consequences, and Management. *Front. Public Heal.* **2**, 1–8 (2014). DOI:10.3389/fpubh.2014.00145.

169. Abushaheen, M. A. *et al.* Antimicrobial resistance, mechanisms and its clinical significance. *Disease-a-Month* **66**, 100971 (2020). DOI:10.1016/j.disamonth.2020.100971.

170. WHO. *Antimicrobial Resistance: An Emerging Water, Sanitation and Hygiene Issue Briefing Note WHO/FWC/WSH/14.07*. www.who.int/about/licensing/copyright_form/en/index.html (2014).

171. Davis, B. C. *et al.* Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. *Environ. Sci. Technol.* **54**, 15108–15119 (2020). DOI:10.1021/acs.est.0c05567.

172. Dunachie, S. J., Day, N. P. & Dolecek, C. The challenges of estimating the human global burden of disease of antimicrobial resistant bacteria. *Curr. Opin. Microbiol.* **57**, 95–101 (2020). DOI:10.1016/j.mib.2020.09.013.

173. Nuzzo, R. Scientific method: Statistical errors. *Nature* **506**, 150–152 (2014). DOI:10.1038/506150a.

174. Malchione, M. D., Torres, L. M., Hartley, D. M., Koch, M. & Goodman, J. L. Carbapenem and colistin resistance in Enterobacteriaceae in Southeast Asia: Review and mapping of emerging and overlapping challenges. *Int. J. Antimicrob. Agents* **54**, 381–399 (2019). DOI:10.1016/j.ijantimicag.2019.07.019.

175. Kakkar, M. *et al.* Antimicrobial resistance in South East Asia: time to ask the right questions. *Glob. Health Action* **11**, (2018). DOI:10.1080/16549716.2018.1483637.

176. Wang, Q., Li, S., Jia, P., Qi, C. & Ding, F. A review of surface water quality models. *Sci. World J.* 1–7 (2013) doi:10.1155/2013/231768. DOI:10.1155/2013/231768.

177. Bello, A. D., Haniffah, M. R. M., Hanapi, M. N. & Usman, A. B. Identification of critical source areas under present and projected land use for effective management of diffuse pollutants in an urbanized watershed. *Int. J. River Basin Manag.* **0**, 1–14 (2018). DOI:10.1080/15715124.2018.1461108.

178. Ranhill Consulting Sdn Bhd. *Review of the national water resources study (2000-2050) and formulation of national water resources policy - Final report*. vol. 17 http://www.water.gov.my/jps/resources/PDF/Hydrology Publication/Vol17Johor.pdf (2011).

179. Bello, A. D. Modelling the impacts of land-use and climate change in Skudai river watershed. (UTM, 2018).

180. Wong, C. L., Venneker, R., Uhlenbrook, S., Jamil, A. B. M. & Zhou, Y. Variability of rainfall in Peninsular Malaysia. *Hydrol. Earth Syst. Sci. Discuss.* **6**, 5471–5503 (2009). DOI:10.5194/hessd-6-5471-2009.

181. WEPA. *Water Environmental Management in Asia 2015*. http://wepa-db.net/en/publication/2015_outlook/index.html (2015).

182. Michaud, J. P. & Wierenga, M. *Estimating Discharge and Stream Flows. A Guide for Sand and Gravel Operators. Ecology Publication* https://apps.ecology.wa.gov/publications/documents/0510070.pdf (2005).

183. Hydromatch. *Flow estimation for streams and small rivers*. http://www.hydromatch.com/sites/default/files/downloads/DIY-flow-measurement-guide.pdf (2014).

184. Hach. *Nitrogen, Ammonia - Method 10031*. https://www.hach.com/asset-

get.download-en.jsa?id=7639983747 (2015).

185. Hach. *Chemical Oxygen Demand - Method 8000.* http://www.hach.com/asset-get.download.jsa?id=7639983816 (2014).

186. Hach. *Phosphorus - Method 8190.* https://www.hach.com/asset-get.download-en.jsa?id=7639983838 (2014).

187. Hach. *Nitrogen, Total - Method 10071.* https://www.hach.com/asset-get.download-en.jsa?id=7639983804 (2014).

188. Huang, Y. F., Ang, S. Y., Lee, K. M. & Lee, T. S. Quality of Water Resources in Malaysia. in *Research and Practices in Water Quality* (2015). doi:10.5772/58969. DOI:10.5772/58969.

189. Sigma Aldrich. *Product Information 61471 ESBL ChromoSelect Agar Supplement.* https://www.sigmaaldrich.com/catalog/product/sial/61471?lang=de&region=DE (2013).

190. Hands, C. L. Relationships between Zinc and Meropenem Resistance in the Natural Environment and Experimental Bioreactors. (Newcastle University, 2016).

191. CLSI. *Performance Standards for Antimicrobial Susceptibility Testing. 31st ed. CLSI supplement M100.* https://clsi.org/standards/products/microbiology/documents/m100/ (2021).

192. Maheux, A. F., Bouchard, S., Bérubé, È. & Bergeron, M. G. Comparison of MI, Chromocult® coliform, and Compass CC chromogenic culture-based methods to detect Escherichia coli and total coliforms in water using 16S rRNA sequencing for colony identification. *J. Water Health* **15**, 353–359 (2017). DOI:10.2166/wh.2017.174.

193. Breed, R. S. & Dotterrer, W. D. The number of colonies allowable on satisfactory agar plates. *J. Bacteriol.* **1**, 321–331 (1916).

194. Le, T. H., Ng, C., Tran, N. H., Chen, H. & Gin, K. Y. H. Removal of antibiotic residues, antibiotic resistant bacteria and antibiotic resistance genes in municipal wastewater by membrane bioreactor systems. *Water Res.* **145**, 498–508 (2018). DOI:10.1016/j.watres.2018.08.060.

195. Tran, N. H., Chen, H., Reinhard, M., Mao, F. & Gin, K. Y. H. Occurrence and removal of multiple classes of antibiotics and antimicrobial agents in biological wastewater treatment processes. *Water Res.* **104**, 461–472 (2016). DOI:10.1016/j.watres.2016.08.040.

196. Tran, N. H. *et al.* Simultaneous analysis of multiple classes of antimicrobials in environmental water samples using SPE coupled with UHPLC-ESI-MS/MS and isotope dilution. *Talanta* **159**, 163–173 (2016). DOI:10.1016/j.talanta.2016.06.006.

197. Zhu, Y. G. *et al.* Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat. Microbiol.* **2**, (2017). DOI:10.1038/nmicrobiol.2016.270.

198. Zhu, Y. G. *et al.* Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3435–3440 (2013).

DOI:10.1073/pnas.1222743110.

199. Klappenbach, J. A. rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Res.* **29**, 181–184 (2001). DOI:10.1093/nar/29.1.181.

200. R Core Team. R: A Language and Environment for Statistical Computing. (2020).

201. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (2016).

202. Inkscape Project. Inkscape. (2017).

203. ESRI. ArcGIS Desktop. (2018).

204. Attal, M. *Structural Analysis of Rocks and Regions 2017 - Topographic analysis.* https://blogs.ed.ac.uk/mattal/ (2017).

205. Antweiler, R. C. Evaluation of Statistical Treatments of Left-Censored Environmental Data Using Coincident Uncensored Data Sets. II. Group Comparisons. *Environ. Sci. Technol.* **49**, 13439–13446 (2015). DOI:10.1021/acs.est.5b02385.

206. Kassambara, A. T-test essentials: definition, formula and calculation. *Data Novia* https://www.datanovia.com/en/lessons/t-test-effect-size-using-cohens-d-measure/ (2018).

207. Cohen, J. A power primer. *Psychol. Bull.* **112**, 155–159 (1992). DOI:10.1037/0033-2909.112.1.155.

208. Kassambara, A. T-test Effect Size using Cohen's D Measure. *Datanovia* https://www.datanovia.com/en/lessons/t-test-effect-size-using-cohens-d-measure/#cohens-d-for-welch-test (2020).

209. Delacre, M., Lakens, D. & Leys, C. Why psychologists should by default use welch's t-Test instead of student's t-Test. *Int. Rev. Soc. Psychol.* **30**, 92–101 (2017). DOI:10.5334/irsp.82.

210. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995). DOI:10.1111/j.2517-6161.1995.tb02031.x.

211. Torchiano, M. effsize:Efficient Effect Size Computation. (2020).

212. Box, G. E. P. & Cox, D. R. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B* **26**, 211–243 (1964). DOI:10.1111/j.2517-6161.1964.tb00553.x.

213. Riplley, B. *et al.* Package 'MASS'. Support Functions and Datasets for Venables and Ripley's MASS. (2018).

214. Blighe, K., Rana, S. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. (2020).

215. Revelle, W. psych: Procedures for Personality and Psychological Research. (2019).

216. Wei, T. & Simko, V. R package 'corrplot': Visualization of a Correlation Matrix. (2017).

217. Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biol. Rev.* **82**, 591–605 (2007). DOI:10.1111/j.1469-185X.2007.00027.x.

218. Dobriyal, P., Badola, R., Tuboi, C. & Hussain, S. A. A review of methods for monitoring streamflow for sustainable water resource management. *Appl. Water Sci.* **7**, 2617–2628 (2017). DOI:10.1007/s13201-016-0488-y.

219. Sivasampu, S., Yasmin Farhana, A., Ong, S., Goh, P. & Noor Hisham, A. *National Medical Care Statistics Primary Care 2014*. *Ministry of Health Malaysia* http://www.crc.gov.my/nhsi/wp-content/uploads/publications/nmcs2014/NMCS_2014_fullreport.pdf (2016).

220. Andreozzi, R. *et al.* Antibiotics in the Environment: Occurrence in Italian STPs, Fate, and Preliminary Assessment on Algal Toxicity of Amoxicillin. *Environ. Sci. Technol.* **38**, 6832–6838 (2004). DOI:10.1021/es049509a.

221. Castiglioni, S., Bagnati, R., Calamari, D., Fanelli, R. & Zuccato, E. A multiresidue analytical method using solid-phase extraction and high-pressure liquid chromatography tandem mass spectrometry to measure pharmaceuticals of different therapeutic classes in urban wastewaters. *J. Chromatogr. A* **1092**, 206–215 (2005). DOI:10.1016/j.chroma.2005.07.012.

222. Kasprzyk-Hordern, B., Dinsdale, R. M. & Guwy, A. J. Multi-residue method for the determination of basic/neutral pharmaceuticals and illicit drugs in surface water by solid-phase extraction and ultra performance liquid chromatography–positive electrospray ionisation tandem mass spectrometry. *J. Chromatogr. A* **1161**, 132–145 (2007). DOI:https://doi.org/10.1016/j.chroma.2007.05.074.

223. Low, K. *et al.* Prevalence and risk assessment of antibiotics in riverine estuarine waters of Larut and Sangga Besar River, Perak. *J. Oceanol. Limnol.* **39**, 122–134 (2021). DOI:10.1007/s00343-020-9246-y.

224. Shimizu, A. *et al.* Ubiquitous occurrence of sulfonamides in tropical Asian waters. *Sci. Total Environ.* **452**–**453**, 108–115 (2013). DOI:10.1016/j.scitotenv.2013.02.027.

225. Zheng, J. *et al.* High-throughput profiling of seasonal variations of antibiotic resistance gene transport in a peri-urban river. *Environ. Int.* **114**, 87–94 (2018). DOI:10.1016/j.envint.2018.02.039.

226. An, X. L. *et al.* Tracking antibiotic resistome during wastewater treatment using high throughput quantitative PCR. *Environ. Int.* **117**, 146–153 (2018). DOI:10.1016/j.envint.2018.05.011.

227. Ramadan, A. A., Abdelaziz, N. A., Amin, M. A. & Aziz, R. K. Novel blaCTX-M variants and genotype-phenotype correlations among clinical isolates of extended spectrum beta lactamase-producing Escherichia coli. *Sci. Rep.* **9**, 1–12 (2019). DOI:10.1038/s41598-019-39730-0.

228. Crank, C. & O'Driscoll, T. Vancomycin-resistant enterococcal infections: epidemiology, clinical manifestations, and optimal management. *Infect. Drug Resist.* **8**, 217–230 (2015). DOI:10.2147/IDR.S54125.

229. Debelius, J. *et al.* Tiny microbes, enormous impacts: What matters in gut microbiome studies? *Genome Biol.* **17**, 1–12 (2016). DOI:10.1186/s13059-016-1086-x.

230. U.S. EPA. Causal Analysis/Diagnosis Decision Information System (CADDIS): About Stressors. *Office of Research and Development* https://www.epa.gov/caddis-vol2/learn-about-stressors (2017).

231. Weiner, R. & Matthews, R. *Environmental Engineering.* (Elsevier Science, 2003).

232. Fair, G. M. The Dissolved Oxygen Sag: An Analysis. *Sewage Work. J.* **11**, 445–461 (1939).

233. Jong, M.-C., Harwood, C. R., Blackburn, A., Snape, J. R. & Graham, D. W. Impact of Redox Conditions on Antibiotic Resistance Conjugative Gene Transfer Frequency and Plasmid Fate in Wastewater Ecosystems. *Environ. Sci. Technol.* **54**, 14984–14993 (2020). DOI:10.1021/acs.est.0c03714.

234. Chowdhury, A. S., Call, D. R. & Broschat, S. L. PARGT: a software tool for predicting antimicrobial resistance in bacteria. *Sci. Rep.* **10**, 11033 (2020). DOI:10.1038/s41598-020-67949-9.

235. Her, H.-L. & Wu, Y.-W. A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the Escherichia coli strains. *Bioinformatics* **34**, i89–i95 (2018). DOI:10.1093/bioinformatics/bty276.

236. Li, X. *et al.* Antimicrobial Resistance Risk Assessment Models and Database System for Animal-Derived Pathogens. *Antibiotics* **9**, 829 (2020). DOI:10.3390/antibiotics9110829.

237. Gothwal, R. & Shashidhar, T. Antibiotic Pollution in the Environment: A Review. *Clean - Soil, Air, Water* vol. 43 479–489 (2015). DOI:10.1002/clen.201300989.

238. Knapp, C. W. *et al.* Indirect Evidence of Transposon-Mediated Selection of Antibiotic Resistance Genes in Aquatic Systems at Low-Level Oxytetracycline Exposures. *Environ. Sci. Technol.* **42**, 5348–5353 (2008). DOI:10.1021/es703199g.

239. UN Water. *Step-by-step monitoring methodology for indicator 6.3.2.* https://www.unwater.org/publications/step-step-methodology-monitoring-water-quality-6-3-2/ (2018).

240. UNEP. Global Freshwater Quality Database GEMStat. https://gemstat.org/ (2021).

241. UNEP. *A Snapshot of the World ' s Water Quality : Towards a global assessment. United Nations Environment Programme* https://uneplive.unep.org/media/docs/assessments/unep_wwqa_report_web.pdf (2016).

242. Metcalfe, C., Guppy, L. & Qadir, M. *Global Barriers to Improving Water Quality: A Critical Review. United Nations University Institute for Water, Environment and Health* https://inweh.unu.edu/wp-content/uploads/2018/01/Global-Barriers-To-Improving-Water-Quality-A-Critical-Review_web.pdf (2017).

243. Strokal, M. *et al.* Urbanization: an increasing source of multiple pollutants to rivers in the 21st century. *Urban Sustain.* **1**, (2021). DOI:10.1038/s42949-021-00026-w.

244. Heathcote, I. W. *A Review and Comparison of Dissolved Oxygen Models Appropriate for the Grand River Watershed. Wyndham Research I* https://www.grandriver.ca/en/our-watershed/resources/Documents/Water_Quality_GRSM_Comparison.pdf (2009).

245. Johnson, M. S. *et al.* Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: A comparison of HSPF and SMR. *J. Hydrol.* **284**, 57–76 (2003). DOI:10.1016/j.jhydrol.2003.07.005.

246. Naubi, I., Zardari, N. H., Shirazi, S. M., Ibrahim, N. F. B. & Baloo, L. Effectiveness of water quality index for monitoring Malaysian river water quality. *Polish J. Environ. Stud.* **25**, 231–239 (2016). DOI:10.15244/pjoes/60109.

247. Hothorn, T. *et al.* lmtest: Testing Linear Regression Models.

248. Bello, A. A. D., Hashim, N. B. & Haniffah, R. M. Impact of urbanization on the sediment yield in tropical watershed using temporal land-use changes and a GIS-based model. *J. Water L. Dev.* **34**, 33–45 (2017). DOI:10.1515/jwld-2017-0036.

249. Bello, A.-A. D., Haniffah, M. R. M., Hashim, N. B. & Anuar, K. M. Estimation of Hydrological Changes in a Tropical Watershed Using Multi-Temporal Land-Use and Dynamic Modelling. *J. Teknol.* **80**, 123–136 (2018). DOI:10.11113/jt.v80.11179.

250. Hashim, N. B., Bello, A. D., Ridza, M. & Haniffah, M. Application of remote sensing in land use characterization and modelling hydrology of a coastal watershed. *11TH Int. Civ. Eng. POST Grad. Conf. - 1ST Int. Symp. Expert. Eng. Des.* (2015).

251. Reggiani, P., Todini, E. & Meißner, D. Analytical solution of a kinematic wave approximation for channel routing. *Hydrol. Res.* **45**, 43–57 (2014). DOI:10.2166/nh.2013.157.

252. Mining, D. O. F. *Calibration Parameters Used to Simulate Streamflow from Application of the Hydrologic Simulation Program-FORTRAN Model (HSPF) to Mountainous Basins Containing Coal Mines in West Virginia Scientific Investigations Report 2005 – 5099.* https://pubs.usgs.gov/sir/2005/5099/ (2005).

253. Flynn, A., Dilks, D. & Weintraub, L. *Truckee River HSPF Model Water Quality Algorithms.* http://www.truckeeriverinfo.org/files/truckee/12-14-2011_6_TRHSPF_Algorithms_Memo_DRAFT.pdf (2011).

254. Hassoun-Kheir, N. *et al.* Comparison of antibiotic-resistant bacteria and antibiotic resistance genes abundance in hospital and community wastewater: A systematic review. *Sci. Total Environ.* **743**, 140804 (2020). DOI:10.1016/j.scitotenv.2020.140804.

255. Rice, E. W., Wang, P., Smith, A. L. & Stadler, L. B. Determining Hosts of Antibiotic Resistance Genes: A Review of Methodological Advances. *Environ. Sci. Technol. Lett.* **7**, 282–291 (2020). DOI:10.1021/acs.estlett.0c00202.

256. Belton, J. M. *et al.* Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012). DOI:10.1016/j.ymeth.2012.05.001.

257. Amarasiri, M., Sano, D. & Suzuki, S. Understanding human health risks caused by antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARG) in water environments: Current knowledge and questions to be answered. *Crit. Rev. Environ. Sci. Technol.* **50**, 2016–2059 (2020). DOI:10.1080/10643389.2019.1692611.

258. Ashbolt, N. J., Amézquita, A., Backhaus, T., Borriello, P. & Brandt, K. K. Review Human Health Risk Assessment (HHRA) for Environmental Development and Transfer of Antibiotic Resistance. **121**, 993–1001 (2013). DOI:10.1289/ehp.1206316.

259. Carr, A., Diener, C., Baliga, N. S. & Gibbons, S. M. Use and abuse of correlation analyses in microbial ecology. *ISME J.* **13**, 2647–2655 (2019). DOI:10.1038/s41396-019-0459-z.

260. Knight, R. *et al.* Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* **16**, 1–13 (2018). DOI:10.1038/s41579-018-0029-9.

261. Morton, J. T. *et al.* Establishing microbial composition measurement standards with reference frames. *Nat. Commun.* **10**, 2719 (2019). DOI:10.1038/s41467-019-10656-5.

262. Props, R. *et al.* Absolute quantification of microbial taxon abundances. *ISME J.* **11**, 584–587 (2017). DOI:10.1038/ismej.2016.117.

263. Köchling, T., Sanz, J. L., Galdino, L., Florencio, L. & Kato, M. T. Impact of pollution on the microbial diversity of a tropical river in an urbanized region of northeastern Brazil. *Int. Microbiol.* **20**, 11–24 (2017). DOI:10.2436/20.1501.01.281.

264. Li, Y. *et al.* How did the bacterial community respond to the level of urbanization along the Yangtze River? *Environ. Sci. Process. Impacts* **22**, 161–172 (2020). DOI:10.1039/c9em00399a.

265. Jost, L. Entropy and diversity. *OIKOS* **113**, (2006). DOI:10.1111/j.2006.0030-1299.14714.x.

266. Tuomisto, H. Commentary: Do we have a consistent terminology for species diversity? Yes, if we choose to use it. *Oecologia* **167**, 903–911 (2011). DOI:10.1007/s00442-011-2128-4.

267. Chao, A., Chiu, C.-H. & Jost, L. Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. *Annu. Rev. Ecol. Evol. Syst.* **45**, 297–324 (2014). DOI:10.1146/annurev-ecolsys-120213-091540.

268. Hill, M. O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **54**, 427–432 (1973). DOI:10.2307/1934352.

269. Props, R., Monsieurs, P., Mysara, M., Clement, L. & Boon, N. Measuring the biodiversity of microbial communities by flow cytometry. *Methods Ecol. Evol.* **7**, 1376–1385 (2016). DOI:10.1111/2041-210X.12607.

270. Alberdi, A. *et al.* DNA metabarcoding and spatial modelling link diet diversification with distribution homogeneity in European bats. *Nat. Commun.* **11**, 1–8 (2020). DOI:10.1038/s41467-020-14961-2.

271. Alberdi, A. & Gilbert, M. T. P. A guide to the application of Hill numbers to DNA-based diversity analyses. *Mol. Ecol. Resour.* **19**, 804–817 (2019). DOI:10.1111/1755-0998.13014.

272. Chao, A., Chiu, C. H., Hsieh, T. C. & Inouye, B. D. Proposing a resolution to debates on diversity partitioning. *Ecology* **93**, 2037–2051 (2012). DOI:10.1890/11-1817.1.

273. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439 (2007). DOI:10.1890/06-1736.1.

274. Alberdi, A. & Gilbert, M. T. P. hilldiv: an R package for the integral analysis of diversity based on Hill numbers. *bioRxiv* **1**, 545665 (2019). DOI:10.1101/545665.

275. Harms, G. *et al.* Real-time PCR quantification of nitrifying bacteria in a municipal wastewater treatment plant. *Environ. Sci. Technol.* **37**, 343–351 (2003). DOI:10.1021/es0257164.

276. Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A. & Korpela, K. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS One* **15**, 1–10 (2020). DOI:10.1371/journal.pone.0227285.

277. Louca, S., Doebeli, M. & Parfrey, L. W. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* **6**, 1–12 (2018). DOI:10.1186/s40168-018-0420-9.

278. Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K. & Schmidt, T. M. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **43**, D593–D598 (2014). DOI:10.1093/nar/gku1201.

279. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4516–4522 (2011). DOI:10.1073/pnas.1000080107.

280. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013). DOI:10.1128/AEM.01043-13.

281. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019). DOI:10.1038/s41587-019-0209-9.

282. Callahan, B. J. *et al.* DADA2: High resolution sample inference from Illumina amplican data. *Nat. Methods* **13**, 1–6 (2016). DOI:10.1038/nmeth.3869.DADA2.

283. McDonald, D. *et al.* The Biological Observation Matrix (BIOM) format or: how I learned

to stop worrying and love the ome-ome. *Gigascience* **1**, 7 (2012). DOI:10.1186/2047-217X-1-7.

284. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013). DOI:10.1093/nar/gks1219.

285. Yilmaz, P. *et al.* The SILVA and 'all-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res.* **42**, 643–648 (2014). DOI:10.1093/nar/gkt1209.

286. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* (eds. van der Walt, S. & Millman, J.) 51–56 (2010).

287. Robeson, M. S. make_SILVA_db. https://github.com/mikerobeson/make_SILVA_db (2020).

288. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2012).

289. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 1–17 (2018). DOI:10.1186/s40168-018-0470-z.

290. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012). DOI:10.1038/nbt.2198.

291. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217 (2013). DOI:10.1371/journal.pone.0061217.

292. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.5-6. *Cran R* (2019).

293. Webber, W., Moffat, A. & Zobel, J. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.* **28**, 1–34 (2010). DOI:10.1145/1852102.1852106.

294. Schmich, F. gespeR: Gene-Specific Phenotype EstimatoR. R package version 1.20.0. 1–7 (2020).

295. Junker, B. & Schreiber, F. *Analysis of Biological Networks.* (Wiley, 2008).

296. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. (2009).

297. McMurdie, P. J. & Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, (2014). DOI:10.1371/journal.pcbi.1003531.

298. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 1–18 (2017). DOI:10.1186/s40168-017-0237-y.

299. Jost, L. & Chao, A. *Diversity analysis. Chapman & Hall/CRC Applied Environmental Statistics* (Taylor & Francis, 2020).

300. Guo, X. *et al.* Host-Associated Quantitative Abundance Profiling Reveals the Microbial Load Variation of Root Microbiome. *Plant Commun.* **1**, 100003 (2020). DOI:10.1016/j.xplc.2019.100003.

301. Brown, M. R. R. *et al.* A flow cytometry method for bacterial quantification and biomass estimates in activated sludge. *J. Microbiol. Methods* **160**, 73–83 (2019). DOI:10.1016/j.mimet.2019.03.022.

302. Vignola, M., Werner, D., Hammes, F., King, L. C. & Davenport, R. J. Flow-cytometric quantification of microbial cells on sand from water biofilters. *Water Res.* **143**, 66–76 (2018). DOI:10.1016/j.watres.2018.05.053.

303. Ott, A. *et al.* Multi-laboratory Validation of a New Marine Biodegradation Screening Test for Chemical Persistence Assessment. *Environ. Sci. Technol.* **54**, 4210–4220 (2020). DOI:10.1021/acs.est.9b07710.

304. Dong, L., Wang, S., Fu, B. & Wang, J. Evaluation of droplet digital PCR and next generation sequencing for characterizing DNA reference material for KRAS mutation detection. *Sci. Rep.* **8**, (2018). DOI:10.1038/s41598-018-27368-3.

305. Santander, R. D., Meredith, C. L. & Aćimović, S. G. Development of a viability digital PCR protocol for the selective detection and quantification of live Erwinia amylovora cells in cankers. *Sci. Rep.* **9**, (2019). DOI:10.1038/s41598-019-47976-x.

306. Zhou, Z. C. *et al.* Antibiotic resistance genes in an urban river as impacted by bacterial community and physicochemical parameters. *Environ. Sci. Pollut. Res.* **24**, 23753–23762 (2017). DOI:10.1007/s11356-017-0032-0.

307. Marathe, N. P. *et al.* Untreated urban waste contaminates Indian river sediments with resistance genes to last resort antibiotics. *Water Res.* **124**, 388–397 (2017). DOI:10.1016/j.watres.2017.07.060.

308. Chen, Q. L. *et al.* An underappreciated hotspot of antibiotic resistance: The groundwater near the municipal solid waste landfill. *Sci. Total Environ.* **609**, 966–973 (2017). DOI:10.1016/j.scitotenv.2017.07.164.

309. Zhao, Y. *et al.* Evidence for co-selection of antibiotic resistance genes and mobile genetic elements in metal polluted urban soils. *Sci. Total Environ.* **656**, 512–520 (2019). DOI:10.1016/j.scitotenv.2018.11.372.

310. Xiang, Q. *et al.* Spatial and temporal distribution of antibiotic resistomes in a peri-urban area is associated significantly with anthropogenic activities. *Environ. Pollut.* **235**, 525–533 (2018). DOI:10.1016/j.envpol.2017.12.119.

311. Liu, L. *et al.* Large-scale biogeographical patterns of bacterial antibiotic resistome in the waterbodies of China. *Environ. Int.* **117**, 292–299 (2018). DOI:10.1016/j.envint.2018.05.023.

312. Duda, P. B., Hummel, P. R., Donigian, A. S. & Imhoff, J. C. BASINS/HSPF: Model use, calibration and validation. *Am. Soc. Agric. Biol. Eng.* **55**, 1523–47 (2012).

Appendix

List of appendix figures

List of appendix tables

Appendix A

Appendix Figure A-1. Total coliform (TC), ESBL coliform and carbapenem resistant bacteria screened for with 2 μg/mL meropenem (CRB-2) concentrations (a) and mass loadings (b) in the river catchment. Data represented is based on four biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on three biological replicates for the tributaries (Se1, M5).
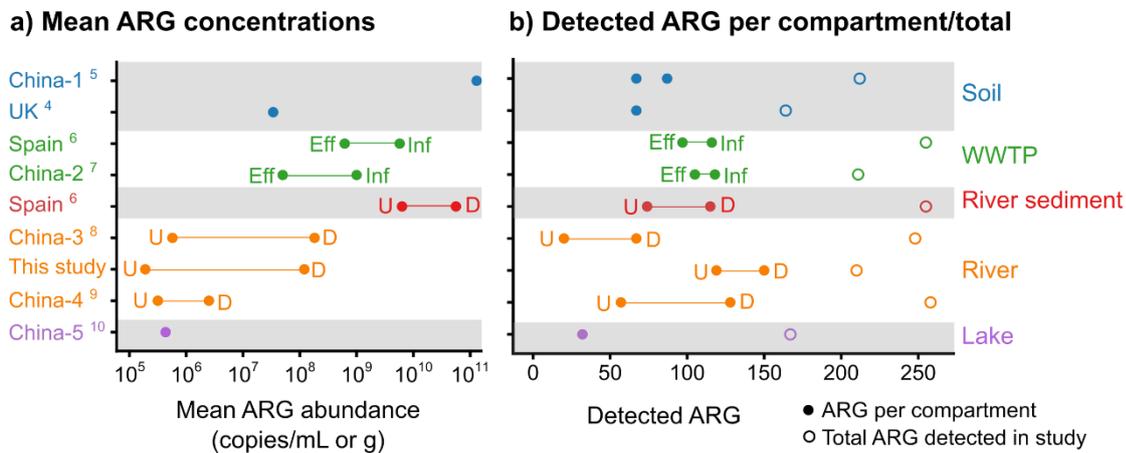
Appendix Figure A-2. Comparison of high-throughput (HT) qPCR antibiotic resistance gene (ARG) concentrations (a) and detections (b; out of 283 ARGs) across different environmental compartments (e.g., soil, sewage, river, lake). Selected studies monitored ARG levels in soil (UK [309]; China [310]), wastewater treatment plant (WWTP) influent (Inf) and effluent (Eff) (Spain [70]; China [226]), river sediment upstream (U) and downstream (D) (Spain [70]); river water upstream (U) and downstream (D) (China [135,225]; Malaysia – this study) and in lake water phases (China [311]). To the best of our knowledge, all studies applied the same HT-qPCR assay, screening for 283 ARGs with the same primer sets. For all studies, 3/3 technical qPCR replicates had to amplify for a positive result, except for the Spanish study, where 2/3 technical qPCR replicates were sufficient for a positive result. For ARG abundance data (a), mean concentrations per environmental compartment are reported. For detected ARG (b), mean numbers of ARG per environmental compartment and total ARG detected per study are reported.

Appendix Table A-1. Sampling point coordinates.

Sampling point	Latitude	Longitude
S1	1°41'05.9"N	103°34'22.0"E
S2	1°38'58.8"N	103°36'58.1"E
S5	1°36'22.6"N	103°39'04.5"E
S6	1°32'54.6"N	103°39'41.8"E
S7	1°31'11.3"N	103°40'41.4"E
S8	1°29'57.2"N	103°40'59.0"E
Se1	1°36'09.3"N	103°38'45.7"E
M5	1°30'13.5"N	103°38'59.1"E

Appendix Table A-2. Modified water quality classes for Malaysia based on the Malaysian Department of Environment (DoE) Water Quality Index (WQI, classifications clean - slightly polluted - polluted) and National Water Quality Standards for Malaysia (NWQS, classes I-V, see below). To unify both approaches, we translated the thresholds from the five NWQS classes to match the three WQI groups[188].

DoE NWQS	Class I	Class II	Class III	Class IV	Class V
Assigned classes (based on DoE WQI)	Clean	Slightly polluted	Polluted		
NH_3-N	0.1	0.3	0.9	2.7	>2.7
BOD	1	3	6	12	>12
COD	10	25	50	100	>100
DO	7	5	3	1	<1
pH	7	6	5	<5	>5
TSS	25	50	150	300	>300
WQI	92.7	76.5	51.9	31	>31

BOD: biochemical oxygen demand. COD: chemical oxygen demand. DO: dissolved oxygen. NH_3-N: ammonia. TSS: total suspended solids.

Appendix Table A-3. Department of Environment (DoE) data for the Skudai catchment (2018). DoE sampling point 3SI09 = S1 and DoE sampling point 3SI05 = S8.

Sampling point	Month	DO	NH_3-N	COD
3SI09	January	7.49	0.11	19
3SI05	January	1.11	2.21	17
3SI09	March	9.36	0.05	14
3SI05	March	0.31	3.21	36
3SI09	May	6.902	0.11	14
3SI05	May	1.461	2.66	26
3SI09	July	7.719	0.07	15
3SI05	July	3.424	5.51	44
3SI09	September	6.771	0.14	14
3SI05	September	1.596	7.42	67
3SI09	November	7.329	0.1	21
3SI05	November	4.534	1.02	29

COD: chemical oxygen demand. DO: dissolved oxygen. NH_3-N: ammonia.

Appendix Table A-4. Monitored antibiotics (ng/L) in aqueous phase of the Skudai river catchment, Malaysia during four sampling campaigns (two March 2018, two July 2018) with n=30.

Antibiotic class	Antibiotic name	Abbr.	Detected[1]	DL	Min - Max	Mean [1]	PNEC [2]
β-Lactam	Meropenem	MER	0/30 *	1	n.a.	n.a.	64
	Cefixime	CEFX	0/30 *	1	n.a.	n.a.	64
	Ceftazidime	CFZ	0/30 *	15	n.a.	n.a.	500
	Amoxicillin	AMX	24/30	15	<15 - 3336	510 ± 906	250
	Ampicillin	AMP	17/30 *	1	<1 - 17	n.a.	250
Lincosamides	Clindamycin	CLI	30/30	0.02	1 - 4	2 ± 1	1000
	Lincomycin	LIN	30/30	0.02	3 - 21	10 ± 5	2000
Macrolide	Azithromycin	AZT	30/30	0.02	0.1 - 28	8 ± 8	250
	Clarithromycin	CLAR	30/30	0.03	0.4 -27	7 ± 7	250
	Dehydrated Erythromycin	ERY-H2O	30/30	0.05	3 - 167	54 ± 48	n.a.
	Erythromycin	ERY	0/30 *	0.05	n.a.	n.a.	1000
Quinolones/ Fluoroquinolones	Ciprofloxacin	CIPX	22/30	0.5	<0.5 - 705	131 ± 162	64
	Enrofloxacin	ENFLX	18/30	0.05	<0.5 - 13	2 ± 3	64
	Ofloxacin	OFLX	25/30	1	<1 - 13	2 ± 3	500
Sulfonamides	Sulfamethazine	SMZ	24/30	0.03	<0.03 - 39	5 ± 9	n.a.
	Sulfamethoxazole	SMX	27/30	0.05	<0.05 - 1933	181 ± 383	16000
Tetracyclines	Chlortetracycline	CTC	0/30 *	1	n.a.	n.a.	n.a.
	Minocycline	MIN	0/30 *	10	n.a.	n.a.	1000
	Oxytetracycline	OXY	0/30 *	7.5	n.a.	n.a.	500
	Tetracycline	TET	0/30 *	4.5	n.a.	n.a.	1000
Others	Chloramphenicol	CAP	1/30 *	0.3	<0.3 - 8	n.a.	8000
	Trimethoprim	TMP	30/30	0.06	2 - 26	10 ± 6	500

DL: detection limit. n.a.: not available. PNEC: predicted no effect concentration where available[107]. * no mean was calculated as over 40% of values were under detection limit. [1] only calculated for when over 60% of values were detected[2].

Appendix Table A-5. Relative SPE recovery with mean and relative standard deviation (RSD) and method quantification limit (MQL) for antibiotics in surface water. For antibiotic abbreviations, see Appendix Table A-4.

Target antibiotics	Method validation data	
	SPE recovery, Mean ± RSD (%)	MQL (ng/L)
AMP	117.0 ± 9.1	1.5
AMX	106.2 ± 13.4	40
AZT	103.2 ± 0.8	0.1
CAP	101.1 ± 4.4	1.0
CEFX	111.5 ± 9.1	0.3
CFZ	104.7 ± 8.5	50
CIPX	97.6 ± 9.7	1.5
CLAR	99.6 ± 13.8	0.1
CLI	100.8 ± 4.3	0.1
CTC	107.7 ± 10.9	3.0
ENFLX	102.5 ± 14.9	0.5
ERY	98.3 ± 9.3	0.2
ERY-H_2O	99.7 ± 8.4	0.2
LIN	101.5 ± 13.1	0.1
MER	102.7 ± 2.3	4.0
MIN	85.6 ± 11.4	30
OFLX	107.1 ± 7.7	0.6
OXY	109.2 ± 10.6	25
SMX	100.7 ± 5.7	0.15
SMZ	103.3 ± 7.6	0.1
TET	106.3 ± 3.6	15
TMP	100.2 ± 8.0	0.2

Appendix Table A-6. Antibiotic resistant gene (ARG) and mobile genetic element (MGE) primer list for high-throughput qPCR.

Assay ID	Forward Primer	Reverse Primer	Classification	Mechanism
16S rRNA	GGGTTGCGCTCGTTGC	ATGGYTGTCGTCAGCTCGTG	16S	Not applicable
catA1	GGGTGAGTTTCACCAGTTTTGATT	CACCTTGTCGCCTTGCGTATA	FCA	deactivate
catB3	GCACTCGATGCCTTCCAAAA	AGAGCCGATCCAAACGTCAT		deactivate
catB8	CACTCGACGCCTTCCAAAG	CCGAGCCTATCCAGACATCATT		deactivate
Cfr	GCAAAATTCAGAGCAAGTTACGAA	AAAATGACTCCCAACCTGCTTTAT		deactivate
cmlA1-01	TAGGAAGCATCGGAACGTTGAT	CAGACCGAGCACGACTGTTG		efflux
cmlA1-02	AGGAAGCATCGGAACGTTGA	ACAGACCGAGCACGACTGTTG		efflux
cmx(A)	GCGATCGCCATCCTCTGT	TCGACACGGAGCCTTGGT		efflux
floR	ATTGTCTTCACGGTGTCCGTTA	CCGCGATGTCGTCGAACT		efflux
qnrA	AGGATTTCTCACGCCAGGATT	CCGCTTTCAATGAAACTGCAA		unknown
Aac	CCCTGCGTTGTGGCTATGT	TTGGCCACGCCAATCC	Aminoglycoside	deactivate
aac(6')I1	GACCGGATTAAGGCCGATG	CTTGCCCTTGATATTCAGTTTTTATAACCA		deactivate
aac(6')-Ib(aka aacA4)-02	CGTCGCCGAGCAACTTG	CGGTACCTTGCCTCTCAAACC		deactivate
aac(6')-Ib(aka aacA4)-01	GTTTGAGAGGCAAGGTACCGTAA	GAATGCCTGGCGTGTTTGA		deactivate
aac(6')-Ib(aka aacA4)-03	AGAAGCACGCCCGACACTT	GCTCTCCATTCAGCATTGCA		deactivate
aac(6')-II	CGACCCGACTCCGAACAA	GCACGAATCCTGCCTTCTCA		deactivate
aac(6')-Iy	GCTTTGCGGATGCCTCAAT	GGAGAACAAAAATACCTTCAAGGAAA		deactivate
aacA/aphD	AGAGCCTTGGGAAGATGAAGTTT	TTGATCCATACCATAGACTATCTCATCA		deactivate
aacC	CGTCACTTATTCGATGCCCTTAC	GTCGGGCGCGGCATA		deactivate
aacC1	GGTCGTGAGTTCGGAGACGTA	GCAAGTTCCCGAGGTAATCG		deactivate
aacC2	ACGGCATTCTCGATTGCTTT	CCGAGCTTCACGTAAGCATTT		deactivate
aacC4	CGGCGTGGGACACGAT	AGGGAACCTTTGCCATCAACT		deactivate
aadA-01	GTTGTGCACGACGACATCATT	GGCTCGAAGATACCTGCAAGAA		deactivate
aadA-02	CGAGATTCTCCGCGCTGTA	GCTGCCATTCTCCAAATTGC		deactivate
aadA1	AGCTAAGCGCGAACTGCAAT	TGGCTCGAAGATACCTGCAA		deactivate
aadA-1-01	AAAAGCCCGAAGAGGAACTTG	CATCTTTCACAAAGATGTTGCTGTCT		deactivate
aadA-1-02	CGGAATTGAAAAAACTGATCGAA	ATACCGGCTGTCCGTCATTT		deactivate
aadA2-01	ACGGCTCCGCAGTGGAT	GGCCACAGTAACCAACAAATCA		deactivate
aadA2-02	CTTGTCGTGCATGACGACATC	TCGAAGATACCCGCAAGAATG		deactivate
aadA2-03	CAATGACATTCTTGCGGGTATC	GACCTACCAAGGCAACGCTATG		deactivate
aadA5-01	ATCACGATCTTGCGATTTTGCT	CTGCGGATGGGCCTAGAAG		deactivate
aadA5-02	GTTCTTGCTCTTGCTCGCATT	GATGCTCGGCAGGCAAAC		deactivate
aadA9-01	CGCGGCAAGCCTATCTTG	CAAATCAGCGACCGCAGACT		deactivate
aadA9-02	GGATGCACGCTTGGATGAA	CCTCTAGCGGCCGGAGTATT		deactivate
aadD	CCGACAACATTTCTACCATCCTT	ACCGAAGCGCTCGTCGTATA		deactivate
aadE	TACCTTATTGCCCCTTGGAAGAGTTA	GGAACTATGTCCCTTTTAATTCTACAATCT		deactivate
aph	TTTCAGCAAGTGGATCATGTTAAAAT	CCAAGCTGTTTCCACTGTTTTTC		deactivate
aph(2')-Id-02	TAAGGATATACCGACAGTTTTGGAAA	TTTAATCCCTCTTCATACCAATCCATA		deactivate
aph(2')-Id-01	TGAGCAGTATCATAAGTTGAGTGAAAAG	GACAGAACAATCAATCTCTATGGAATG		deactivate
aph6ia	CCCATCCCATGTGTAAGGAAA	GCCACCGCTTCTGCTGTAC		deactivate
aphA1(aka kanR)	TGAACAAGTCTGGAAAGAAATGCA	CCTATTAATTTCCCCTCGTCAAAAA		deactivate
spcN-01	AAAAGTTCGATGAAACACGCCTAT	TCCAGTGGTAGTCCCCGAATC		deactivate
spcN-02	CAGAATCTTCCTGAAAAGTTTGATGAA	CGCAGACACGCCGAATC		deactivate
str	AATGAGTTTTGGAGTGTCTCAACGTA	AATCAAAACCCCTATTAAAGCCAAT		deactivate
strA	CCGGTGGCATTTGAGAAAAA	GTGGCTCAACCTGCGAAAAG		deactivate
strB	GCTCGGTCGTGAGAACAATCT	CAATTTCGGTCGCCTGGTAGT		deactivate
ampC/blaDHA	TGGCCGCAGCAGAAAGA	CCGTTTTATGCACCCAGGAA	β-Lactam	deactivate
ampC-01	TGGCGTATCGGGTCAATGT	CTCCACGGGCCAGTTGAG		deactivate
ampC-02	GCAGCACGCCCCGTAA	TGTACCCATGATGCGCGTACT		deactivate
ampC-04	TCCGGTGACGCGACAGA	CAGCACGCCGGTGAAAGT		deactivate
ampC-05	CTGTTCGAGCTGGGTTCTATAAGTAAA	CAGTATCTGGTCACCGGATCGT		deactivate
ampC-06	CCGCTCAAGCTGGACCATAC	CCATATCCTGCACGTTGGTTT		deactivate
ampC-07	CCGCCCAGAGCAAGGACTA	GCTCGACTTCACGCCGTAAG		deactivate
ampC-09	CAGCCGCTGATGAAAAAATATG	CAGCGAGCCCACTTCGA		deactivate

Assay ID	Forward Primer	Reverse Primer	Classification	Mechanism
bla1	GCAAGTTGAAGCGAAAGAAAAGA	TACCAGTATCAATCGCATATACACCTAA		deactivate
bla-ACC-1	CACACAGCTGATGGCTTATCTAAAA	AATAAACGCGATGGGTTCCA		deactivate
blaCMY	CCGCGGCGAAATTAAGC	GCCACTGTTTGCCTGTCAGTT		deactivate
blaCMY2-01	AAAGCCTCATGGGTGCATAAA	ATAGCTTTTGTTTGCCAGCATCA		deactivate
blaCMY2-02	GCGAGCAGCCTGAAGCA	CGGATGGGCTTGTCCTCTT		deactivate
blaCTX-M-04	CTTGGCGTTGCGCTGAT	CGTTCATCGGCACGGTAGA		deactivate
blaCTX-M-01	GGAGGCGTGACGGCTTTT	TTCAGTGCGATCCAGACGAA		deactivate
blaCTX-M-02	GCCGCGGTGCTGAAGA	ATCGGATTATAGTTAACCAGGTCAGATTT		deactivate
blaCTX-M-03	CGATACCACCACGCCGTTA	GCATTGCCCAACGTCAGATT		deactivate
blaCTX-M-05	GCGATAACGTGGCGATGAAT	GTCGAGACGGAACGTTTCGT		deactivate
blaCTX-M-06	CACAGTTGGTGACGTGGCTTAA	CTCCGCTGCCGGTTTTATC		deactivate
blaGES	GCAATGTGCTCAACGTTCAAG	GTGCCTGAGTCAATTCTTTCAAAG		deactivate
bla-L1	CACCGGGTTACCAGCTGAAG	GCGAAGCTGCGCTTGTAGTC		deactivate
blaMOX/blaCMY	CTATGTCAATGTGCCGAAGCA	GGCTTGTCCTCTTTCGAATAGC		deactivate
blaIMP-02	AAGGCAGCATTTCCTCTCATTTT	GGATAGATCGAGAATTAAGCCACTCT		deactivate
blaIMP-01	AACACGGTTTGGTGGTTCTTGTA	GCGCTCCACAAACCAATTG		deactivate
blaOCH	GGCGACTTGCGCCGTAT	TTTTCTGCTCGGCCATGAG		deactivate
blaOKP	GCCGCCATCACCATGAG	GGTGACGTTGTCACCGATCTG		deactivate
blaOXA1/blaOXA30	CGGATGGTTTGAAGGGGTTTATTAT	TCTTGGCTTTTATGCTTGATGTTAA		deactivate
blaOXA10-01	CGCAATTATCGGCCTAGAAACT	TTGGCTTTCCGTCCCATTT		deactivate
blaOXA10-02	CGCAATTATCGGCCTAGAAACT	TTGGCTTTCCGTCCCATTT		deactivate
blaOXY	CGTTCAGGCGGCAGGTT	GCCGCGATATAAGATTTGAGAATT		deactivate
blaPAO	CGCCGTACAACCGGTGAT	GAAGTAATGCGGTTCTCCTTTCA		deactivate
blaPER	TGCTGGTTGCTGTTTTTGTGA	CCTGCGCAATGATAGCTTCAT		deactivate
blaPSE	TTGTGACCTATTCCCCTGTAATAGAA	TGCGAAGCACGCATCATC		deactivate
blaROB	GCAAAGGCATGACGATTGC	CGCGCTGTTGTCGCTAAA		deactivate
blaSFO	CCGCCGCCATCCAGTA	GGGCCGCCAAGATGCT		deactivate
blaSHV-01	TCCCATGATGAGCACCTTTAAA	TTCGTCACCGGCATCCA		deactivate
blaSHV-02	CTTTCCCATGATGAGCACCTTT	TCCTGCTGGCGATAGTGGAT		deactivate
blaTEM	AGCATCTTACGGATGGCATGA	TCCTCCGATCGTTGTCAGAAGT		deactivate
blaTLA	ACACTTTGCCATTGCTGTTTATGT	TGCAAATTTCGGCAATAATCTTT		deactivate
blaVEB	CCCGATGCAAAGCGTTATG	GAAAGATTCCCTTTATCTATCTCAGACAA		deactivate
blaVIM	GCACTTCTCGCGGAGATTG	CGACGGTGATGCGTACGTT		deactivate
blaZ	GGAGATAAAGTAACAAATCCAGTTAGATATGA	TGCTTAATTTTCCATTTGCGATAAG		deactivate
cepA	AGTTGCGCAGAACAGTCCTCTT	TCGTATCTTGCCCGTCGATAAT		deactivate
cfiA	GCAGCGTTGCTGGACACA	GTTCGGGATAAACGTGGTGACT		deactivate
cfxA	TCATTCCTCGTTCAAGTTTTCAGA	TGCAGCACCAAGAGGAGATGT		deactivate
cphA-01	GCGAGCTGCACAAGCTGAT	CGGCCCAGTCGCTCTTC		deactivate
cphA-02	GTGCTGATGGCGAGTTTCTG	GGTGTGGTAGTTGGTGTTGATCAC		deactivate
fox5	GGTTTGCCGCTGCAGTTC	GCGGCCAGGTGACCAA		deactivate
mecA	GGTTACGGACAAGGTGAAATACTGAT	TGTCTTTTAATAAGTGAGGTGCGTTAATA		protection
NDM1	ATTAGCCGCTGCATTGAT	CATGTCGAGATAGGAAGTG		deactivate
pbp	CCGGTGCCATTGGTTTAGA	AAAATAGCCGCCCCAAGATT		protection
pbp2x	TTTCATAAGTATCTGGACATGGAAGAA	CCAAAGGAAACTTGCTTGAGATTAG		protection
Pbp5	GGCGAACTTCTAATTAATCCTATCCA	CGCCGATGACATTCTTCTTATCTT		protection
penA	AGACGGTAACGTATAACTTTTTGAAAGA	GCGTGTAGCCGGCAATG		protection
intl-1(clinic) = clintl1	CGAACGAGTGGCGGAGGGTG	TACCCGAGAGCTTGGCACCCA	MGEs/ Integron	integrase
intl-1LC (int1)	GGCATCCAAGCAGCAAG	AAGCAGACTTGACCTGA		integrase
intl2	TGCTTTTCCCACCCTTACC	GACGGCTACCCTCTGTTATCTC		integrase
intl3	GCCACCACTTGTTTGAGGA	GGATGTCTGTGCCTGCTTG		integrase
IS613	AGGTTCGGACTCAATGCAACA	TTCAGCACATACCGCCTTGAT		transposase
tnpA-01	CATCATCGGACGGACAGAATT	GTCGGAGATGTGGGTGTAGAAAGT		transposase
tnpA-02	GGGCGGGTCGATTGAAA	GTGGGCGGGATCTGCTT		transposase
tnpA-03	AATTGATGCGGACGGCTTAA	TCACCAAACTGTTTATGGAGTCGTT	MGEs/ Transposase	transposase
tnpA-04	CCGATCACGGAAAGCTCAAG	GGCTCGCATGACTTCGAATC		transposase
tnpA-05	GCCGCACTGTCGATTTTTATC	GCGGGATCTGCCACTTCTT		transposase
tnpA-07	GAAACCGATGCTACAATATCCAATTT	CAGCACCGTTTGCAGTGTAAG		transposase
Tp614	GGAAATCAACGGCATCCAGTT	CATCCATGCGCTTTTGTCTCT		transposase
carB	GGAGTGAGGCTGACCGTAGAAG	ATCGGCGAAACGCACAAA	MLSB	efflux

Assay ID	Forward Primer	Reverse Primer	Classification	Mechanism
ereA	CCTGTGGTACGGAGAATTCATGT	ACCGCATTCGCTTTGCTT		deactivate
ereB	GCTTTATTTCAGGAGGCGGAAT	TTTTAAATGCCACAGCACAGAATC		deactivate
erm(34)	GCGCGTTGACGACGATTT	TGGTCATACTCGACGGCTAGAAC		protection
erm(35)	TTGAAAACGATGTTGCATTAAGTCA	TCTATAATCACAACTAACCACTTGAACGT		protection
erm(36)	GGCGGACCGACTTGCAT	TCTGCGTTGACGACGGTTAC		protection
ermA	TTGAGAAGGGATTTGCGAAAAG	ATATCCATCTCCACCATTAATAGTAAACC		protection
ermA/ermTR	ACATTTTACCAAGGAACTTGTGGAA	GTGGCATGACATAAACCTTCATCA		protection
ermB	TAAAGGGCATTTAACGACGAAACT	TTTATACCTCTGTTTGTTAGGGAATTGAA		protection
ermC	TTTGAAATCGGCTCAGGAAAA	ATGGTCTATTTCAATGGCAGTTACG		protection
ermF	CAGCTTTGGTTGAACATTTACGAA	AAATTCCTAAAATCACAACCGACAA		protection
ermJ/ermD	GGACTCGGCAATGGTCAGAA	CCCCGAAACGCAATATAATGTT		protection
ermK-01	GTTTGATATTGGCATTGTCAGAGAAA	ACCATTGCCGAGTCCACTTT		protection
ermK-02	GAGCCGCAAGCCCCTTT	GTGTTTCATTTGACGCGGAGTAA		protection
ermT-01	GTTCACTAGCACTATTTTTAATGACAGAAGT	GAAGGGTGTCTTTTTAATACAATTAACGA		protection
ermT-02	GTAAAATCCCTAGAGAATACTTTCATCCA	TGAGTGATATTTTTGAAGGGTGTCTT		protection
ermX	GCTCAGTGGTCCCCATGGT	ATCCCCCCGTCAACGTTT		protection
ermY	TTGTCTTTGAAAGTGAAGCAACAGT	TAACGCTAGAGAACGATTTGTATTGAG		protection
lmrA-01	TCGACGTGACCGTAGTGAACA	CGTGACTACCCAGGTGAGTTGA		efflux
lnuA-01	TGACGCTCAACACACTCAAAAA	TTCATGCTTAAGTTCCATACGTGAA		deactivate
lnuB-01	TGAACATAATCCCCTCGTTTAAAGAT	TAATTGCCCTGTTTCATCGTAAATAA		deactivate
lnuB-02	AAAGGAGAAGGTGACCAATACTCTGA	GGAGCTACGTCAAACAACCAGTT		deactivate
lnuC	TGGTCAATATAACAGATGTAAACCAGATTT	CACCCCAGCCACCATCAA		deactivate
matA/mel	TAGTAGGCAAGCTCGGTGTTGA	CCTGTGCTATTTTAAGCCTTGTTTCT		efflux
mdtA	CCTAACGGGCGTGACTTCA	TTCACCTGTTTCAAGGGTCAAA		efflux
mefA	CCGTAGCATTGGAACAGCTTTT	AAACGGAGTATAAGAGTGCTGCAA		efflux
mphA-01	CTGACGCGCTCCGTGTT	GGTGGTGCATGGCGATCT		deactivate
mphA-02	TGATGACCCTGCCATCGA	TTCGCGAGCCCCTCTTC		deactivate
mphB	CGCAGCGCTTGATCTTGTAG	TTACTGCATCCATACGCTGCTT		deactivate
mphC	CGTTTGAAGTACCGAATTGGAAA	GCTGCGGGTTTGCCTGTA		deactivate
msrA-01	CTGCTAACACAAGTACGATTCCAAAT	TCAAGTAAAGTTGTCTTACCTACACCATT		efflux
msrC-01	TCAGACCGGATCGGTTGTC	CCTATTTTTTGGAGTCTTCTCTCTAATGTT		efflux
oleC	CCCGGAGTCGATGTTCGA	GCCGAAGACGTACACGAACAG		efflux
pikR1	TCGACATGCGTGACGAGATT	CCGCGAATTAGGCCAGAA		protection
pikR2	TCGTGGGCCAGGTGAAGA	TTCCCCTTGCCGGTGAA		protection
vatB-01	GGAAAAAGCAACTCCATCTCTTGA	TCCTGGCATAACAGTAACATTCTGA		deactivate
vatB-02	TTGGGAAAAAGCAACTCCATCT	CAATCCACACATCATTTCCAACA		deactivate
vatC-01	CGGAAATTGGGAACGATGTT	GCAATAATAGCCCCGTTTCCTA		deactivate
vatC-02	CGATGTTTGGATTGGACGAGAT	GCTGCAATAATAGCCCCGTTT		deactivate
vatE-01	GGTGCCATTATCGGAGCAAAT	TTGGATTGCCACCGACAAT		deactivate
vatE-02	GACCGTCCTACCAGGCGTAA	TTGGATTGCCACCGACAATT		deactivate
vgaA-01	CGAGTATTGTGGAAAGCAGCTAGTT	CCCGTACCGTTAGAGCCGATA		efflux
vgaA-02	GACGGGTATTGTGGAAAGCAA	TTTCCTGTACCATTAGATCCGATAATT		efflux
vgb-01	AGGGAGGGTATCCATGCAGAT	ACCAAATGCGCCCGTTT		deactivate
vgbB-01	CAGCCGGATTCTGGTCCTT	TACGATCTCCATTCAATTGGGTAAA		efflux
vgbB-02	ATACGAGCTGCCTAATAAAGGATCTT	TGTGAACCACAGGGCATTATCA		deactivate
acrA-01	CAACGATCGGACGGGTTTC	TGGCGATGCCACCGTACT		efflux
acrA-02	GGTCTATCACCCTACGCGCTATC	GCGCGCACGAACATACC		efflux
acrA-03	CAGACCCGCATCGCATATT	CGACAATTTCGCGCTCATG		efflux
acrA-04	TACTTTGCGCGCCATCTTC	CGTGCGCGAACGAACAT		efflux
acrB-01	AGTCGGTGTTCGCCGTTAAC	CAAGGAAACGAACGCAATACC		efflux
acrR-01	GCGCTGGAGACACGACAAC	GCCTTGCTGCGAGAACAAA		efflux
acrR-02	GATGATACCCCTGCTGTGAGA	ACCAAACAAGAAGCGCAAGAA	Non-specific	efflux
adeA	CAGTTCGAGCGCCTATTTCTG	CGCCCTGACCGACCAAT		efflux
acrA-05	CGTGCGCGAACGAACA	ACTTTGCGCGCCATCTTC		efflux
acrF	GCGGCCAGGCACAAAA	TACGCTCTTCCCACGGTTTC		efflux
ceoA	ATCAACACGGACCAGGACAAG	GGAAAGTCCGCTCACGATGA		efflux
cmeA	GCAGCAAAGAAGAAGCACCAA	AGCAGGGTAAGTAAAACTAAGTGGTAAATCT		efflux
cmr	CGGCATCGTCAGTGGAATT	CGGTTCCGAAAAAGATGGAA		efflux

118

Assay ID	Forward Primer	Reverse Primer	Classification	Mechanism
emrD	CTCAGCAGTATGGTGGTAAGCATT	ACCAGGCGCCGAAGAAC		efflux
marR-01	GCGGCGTACTGGTGAAGCTA	TGCCCTGGTCGTTGATGA		efflux
mdetl1	ATACAGCAGTGGATATTGGTTTAATTGT	TGCATAAGGTGAATGTTCCATGA		efflux
mdtE/yhiU	CGTCGGCGCACTCGTT	TCCAGACGTTGTACGGTAACCA		efflux
mepA	ATCGGTCGCTCTTCGTTCAC	ATAAATAGGATCGAGCTGCTGGAT		efflux
mexA	AGGACAACGCTATGCAACGAA	CCGGAAAGGGCCGAAAT		efflux
mexD	TTGCCACTGGCTTTCATGAG	CACTGCCGGAGAACTGTCTGTAGA		efflux
mexE	GGTCAGCACCGACAAGGTCTAC	AGCTCGACGTACTTGAGGAACAC		efflux
mexF	CCGCGAGAAGGCCAAGA	TTGAGTTCGGCGGTGATGA		efflux
mtrC-01	GGACGGGAAGATGGTCCAA	CGTAGCGTTCCGGTTCGAT		efflux
mtrC-02	CGGAGTCCATCGACCATTTG	ATCGTCGGCAAGGAGAATCA		efflux
mtrD-02	GGTCGGCACGCTCTTGTC	TGAAGAATTTGCGCACCACTAC		efflux
mtrD-03	CCGCCAAGCCGATATAGACA	GGCCGGGTTGCCAAA		efflux
oprD	ATGAAGTGGAGCGCCATTG	GGCCACGGCGAACTGA		efflux
oprJ	ACGAGAGTGGCGTCGACAA	AAGGCGATCTCGTTGAGGAA		efflux
pmrA	TTTGCAGGTTTTGTTCCTAATGC	GCAGAGCCTGATTTCTCCTTTG		efflux
putative multidrug	AATTTTGCCGATTATTGCTGAAA	GATTGTCATCATTCGTTTATCACCAA		efflux
qac	CAATAATAACCGAAATAATAGGGACAAGTT	AATAAGTGTTCCTAGTGTTGGCCATAG		efflux
qacA	TGGCAATAGGAGCTATGGTGTTT	AAGGTAACACTATTTTCGGTCCAAATC		efflux
qacA/qacB	TTTAGGCAGCCTCGCTTCA	CCGAATCCAAATAAAACCCAATAA		efflux
qacEdelta1-01	TCGCAACATCCGCATTAAAA	ATGGATTTCAGAACCAGAGAAAGAAA		efflux
qacEdelta1-02	CCCCTTCCGCCGTTGT	CGACCAGACTGCATAAGCAACA		efflux
qacH-01	GTGGCAGCTATCGCTTGGAT	CCAACGAACGCCCACAA		efflux
qacH-02	CATCGTGCTTGTGGCAGCTA	TGAACGCCCAGAAGTCTAGTTTT		efflux
rarD-02	TGACGCATCGCGTGATCT	AAATTTTCTGTGGCGTCTGAATC		efflux
sdeB	CACTACCGCTTCCGCACTTAA	TGAAAAAACGGGAAAAGTCCAT		efflux
tolC-01	GGCCGAGAACCTGATGCA	AGACTTACGCAATTCCGGGTTA		efflux
tolC-02	CAGGCAGAGAACCTGATGCA	CGCAATTCCGGGTTGCT		efflux
tolC-03	GCCAGGCAGAGAACCTGATG	CGCAATTCCGGGTTGCT		efflux
ttgA	ACGCCAATGCCAAACGATT	GTCACGGCGCAGCTTGA		efflux
ttgB	TCGCCCTGGATGTACACCTT	ACCATTGCCGACATCAACAAC		efflux
yceE/mdtG-01	TGGCACAAAATATCTGGCAGTT	TTGTGTGGCGATAAGAGCATTAG		efflux
yceE/mdtG-02	TTATCTGTTTTCTGCTCACCTTCTTTT	GCGTGGTGACAAACAGGCTTA		efflux
yceL/mdtH-01	TCGGGATGGTGGGCAAT	CGATAACCGAGCCGATGTAGA		efflux
yceL/mdtH-02	CGCGTGAAACCTTAAGTGCTT	AGACGGCTAAACCCCATATAGCT		efflux
yceL/mdtH-03	CTGCCGTTAAATGGATGTATGC	ACTCCAGCGGGCGATAGG		efflux
yidY/mdtL-01	GCAGTTGCATATCGCCTTCTC	CTTCCCGGCAAACAGCAT		efflux
yidY/mdtL-02	TGCTGATCGGGATTCTGATTG	CAGGCGCGACGAACATAAT		efflux
fabK	TTTCAGCTCAGCACTTTGGTCAT	AAGGCATCTTTTTCAGCCAGTTC		deactivate
imiR	CCGGACTAGAGCTTCATGTAAGC	CCCACGCGGTACTCTTGTAAA		unknown
nisB	GGGAGAGTTGCCGATGTTGTA	AGCCACTCGTTAAAGGGCAAT		unknown
speA	GCAAGAGGTATTTGCTCAACAAGA	CAGGGTCACCCTCATAAAGAAAA		unknown
bacA-01	CGGCTTCGTGACCTCGTT	ACAATGCGATACCAGGCAAAT		deactivate
bacA-02	TTCCACGACACGATTAAGTCATTG	CGGCTCTTTCGGCTTCAG		deactivate
fosB	TCACTGTAACTAATGAAGCATTAGACCAT	CCATCTGGATCTGTAAAGTAAAGAGATC	Other	deactivate
fosX	GATTAAGCCATATCACTTTAATTGTGAAAG	TCTCCTTCCATAATGCAAATCCA		deactivate
nimE	TGCGCCAAGATAGGGCATA	GTCGTGAATTCGGCAGGTTTA		unknown
pncA	GCAATCGAGGCGGTGTTC	TTGCCGCAGCCAATTCA		unknown
sat4	GAATGGGCAAAGCATAAAAACTTG	CCGATTTTGAAACCACAATTATGATA		deactivate
dfrA1	GGAATGGCCCTGATATTCCA	AGTCTTGCGTCCAACCAACAG		deactivate
dfrA12	CCTCTACCGAACCGTCACACA	GCGACAGCGTTGAAACAACTAC		deactivate
folA	CGAGCAGTTCCTGCCAAAG	CCCAGTCATCCGGTTCATAATC		deactivate
sul1	CAGCGCTATGCGCTCAAG	ATCCCGCTGCGCTGAGT	Sulfonamide	protection
sul2	TCATCTGCCAAACTCGTCGTTA	GTCAAAGAACGCCGCAATGT		protection
sulA/folP-01	CAGGCTCGTAAATTGATAGCAGAAG	CTTTCCTTGCGAATCGCTTT		protection
sulA/folP-03	CACGGCTTCGGCTCATGT	TGCCATCCTGTGACTAGCTACGT		protection
tet(32)	CCATTACTTCGGACAACGGTAGA	CAATCTCTGTGAGGGCATTTAACA		protection
tet(34)	CTTAGCGCAAACAGCAATCAGT	CGGTGATACAGCGCGTAAACT	Tetracycline	unknown
tet(35)	ACCCCATGACGTACCTGTAGAGA	CAACCCACACTGGCTACCAGTT		unknown
tet(36)	AGAATACTCAGCAGAGGTCAGTTCCT	TGGTAGGTCGATAACCCGAAAAT		protection

Assay ID	Forward Primer	Reverse Primer	Classification	Mechanism
tet(36)	TGCAGGAAAGACCTCCATTACAG	CTTTGTCCACACTTCCACGTACTATG		protection
tet(37)	GAGAACGTTGAAAAGGTGGTGAA	AACCAAGCCTGGATCAGTCTCA		unknown
tetA-01	GCTGTTTGTTCTGCCGGAAA	GGTTAAGTTCCTTGAACGCAAACT		efflux
tetA-02	CTCACCAGCCTGACCTCGAT	CACGTTGTTATAGAAGCCGCATAG		efflux
tetB-01	AGTGCGCTTTGGATGCTGTA	AGCCCCAGTAGCTCCTGTGA		efflux
tetB-02	GCCCAGTGCTGTTGTTGTCAT	TGAAAGCAAACGGCCTAAATACA		efflux
tetC-01	CATATCGCAATACATGCGAAAAA	AAAGCCGCGGTAAATAGCAA		efflux
tetC-02	ACTGGTAAGGTAAACGCCATTGTC	ATGCATAAACCAGCCATTGAGTAAG		efflux
tetD-01	TGCCGCGTTTGATTACACA	CACCAGTGATCCCGGAGATAA		efflux
tetD-02	TGTCATCGCGCTGGTGATT	CATCCGCTTCCGGGAGAT		efflux
tetE	TTGGCGCTGTATGCAATGAT	CGACGACCTATGCGATCTGA		efflux
tetG-01	TCAACCATTGCCGATTCGA	TGGCCCGGCAATCATG		efflux
tetG-02	CATCAGCGCCGGTCTTATG	CCCCATGTAGCCGAACCA		efflux
tetH	TTTGGGTCATCTTACCAGCATTAA	TTGCGCATTATCATCGACAGA		efflux
tetJ	GGGTGCCGCATTAGATTACCT	TCGTCCAATGTAGAGCATCCATA		efflux
tetK	CAGCAGTCATTGGAAAATTATCTGATTATA	CCTTGTACTAACCTACCAAAAATCAAAATA		efflux
tetL-01	AGCCCGATTTATTCAAGGAATTG	CAAATGCTTTCCCCCTGTTCT		efflux
tetL-02	ATGGTTGTAGTTGCGCGCTATAT	ATCGCTGGACCGACTCCTT		efflux
tetM-01	CATCATAGACACGCCAGGACATAT	CGCCATCTTTTGCAGAAATCA		protection
tetM-02	TAATATTGGAGTTTTAGCTCATGTTGATG	CCTCTCTGACGTTCTAAAAGCGTATTAT		protection
tetO-01	ATGTGGATACTACAACGCATGAGATT	TGCCTCCACATGATATTTTTCCT		protection
tetW-01	ATGAACATTCCCACCGTTATCTTT	ATATCGGCGGAGAGCTTATCC		protection
tetPA	AGTTGCAGATGTGTATAGTCGTAAACTATCTATT	TGCTACAAGTACGAAAACAAAACTAGAA		efflux
tetPB-01	ACACCTGGACACGCTGATTTT	ACCGTCTAGAACGCGGAATG		protection
tetPB-02	TGATACACCTGGACACGCTGAT	CGTCCAAAACGCGGAATG		protection
tetPB-03	TGGGCGACAGTAGGCTTAGAA	TGACCCTACTGAAACATTAGAAATATACCT		protection
tetPB-05	CTGAAGTGGAGCGATCATTCC	CCCTCAACGGCAGAAATAACTAA		protection
tetQ	CGCCTCAGAAGTAAGTTCATACACTAAG	TCGTTCATGCGGATATTATCAGAAT		protection
tetR-02	CGCGATAGACGCCTTCGA	TCCTGACAACGAGCCTCCTT		efflux
tetR-03	CGCGATGGAGCAAAAGTACAT	AGTGAAAAACCTTGTTGGCATAAAA		efflux
tetS	TTAAGGACAAACTTTCTGACGCATC	TGTCTCCCATTGTTCTGGTTCA		protection
tetT	CCATATAGAGGTTCCACCAAATCC	TGACCCTATTGGTAGTGGTTCTATTG		protection
tetU-01	GTGGCAAAGCAACGGATTG	TGCGGGCTTGCAAAACTATC		unknown
tetV	GCGGGAACGACGATGTATATC	CCGCTATCTCACGACCATGAT		efflux
tetX	AAATTTGTTACCGACACGGAAGTT	CATAGCTGAAAAAATCCAGGACAGTT		unknown
vanA	AAAAGGCTCTGAAAACGCAGTTAT	CGGCCGTTATCTTGTAAAAACAT		protection
vanB-01	TTGTCGGCGAAGTGGATCA	AGCCTTTTTCCGGCTCGTT		protection
vanC-01	ACAGGGATTGGCTATGAACCAT	TGACTGGCGATGATTTGACTATG		protection
vanC-02	CCTGCCACAATCGATCGTT	CGGCTTCATTCGGCTTGATA		protection
vanC-03	AAATCAATACTATGCCGGGCTTT	CCGACCGCTGCCATCA		protection
vanC1	AGGCGATAGCGGGTATTGAA	CAATCGTCAATTGCTCATTTCC		protection
vanC2/vanC3	TTTGACTGTCGGTGCTTGTGA	TCAATCGTTTCAGGCAATGG		protection
vanG	ATTTGAATTGGCAGGTATACAGGTTA	TGATTTGTCTTTGTCCATACATAATGC		protection
vanHB	GAGGTTTCCGAGGCGACAA	CTCTCGGCGGCAGTCGTAT		protection
vanHD	GTGGCCGATTATACCGTCATG	CGCAGGTCATTCAGGCAAT		protection
vanRA-01	CCCTTACTCCCACCGAGTTTT	TTCGTCGCCCCATATCTCAT		protection
vanRA-02	CCACTCCGGCCTTGTCATT	GCTAACCACATTCCCCTTGTTTT	Vancomycin*	protection
vanRB	GCCCTGTCGGATGACGAA	TTACATAGTCGTCTGCCTCTGCAT		protection
vanRC	TGCGGGAAAAACTGAACGA	CCCCCCATACGGTTTTGATTA		protection
vanRC4	AGTGCTTTGGCTTATCTCGAAAA	TCCGGCAGCATCACATCTAA		protection
vanRD	TTATAATGGCAAGGATGCACTAAAGT	CGTCTACATCCGGAAGCATGA		protection
vanSA	CGCGTCATGCTTTCAAAATTC	TCCGCAGAAAGCTCAATTTGTT		protection
vanSB	GCGCGGCAAATGACAAC	TTTGCCATTTTATTCGCACTGT		protection
vanSC-01	ATCAACTGCGGGAGAAAAGTCT	TCCGCTGTTCCGCTTCTT		protection
vanSC-02	GCCATCAGCGAGTCTGATGA	CAGCTGGGATCGTTTTTCCTT		protection
vanTC-01	CACACGCATTTTTTCCCATCTAG	CAGCCAACAGATCATCAAAACAA		protection
vanTC-02	ACAGTTGCCGCTGGTGAAG	CGTGGCTGGTCGATCAAAA		protection
vanTE	GTGGTGCCAAGGAAGTTGCT	CGTAGCCACCGCAAAAAAT		protection
vanTG	CGTGTAGCCGTTCCGTTCTT	CGGCATTACAGGTATATCTGGAAA		protection

120

Assay ID	Forward Primer	Reverse Primer	Classification	Mechanism
vanWB	CGGACAAAGATACCCCCTATAAAG	AAATAGTAAATTGCTCATCTGGCACAT		protection
vanWG	ACATTTTCATTTTGGCAGCTTGTAC	CCGCCATAAGAGCCTACAATCT		protection
vanXA	CGCTAAATATGCCACTTGGGATA	TCAAAAGCGATTCAGCCAACT		protection
vanXB	AGGCACAAAATCGAAGATGCTT	GGGTATGGCTCATCAATCAACTT		protection
vanXD	TAAACCGTGTTATGGGAACGAA	GCGATAGCCGTCCCATAAGA		protection
vanYB	GGCTAAAGCGGAAGCAGAAA	GATATCCACAGCAAGACCAAGCT		protection
vanYD-01	AAGGCGATACCCTGACTGTCA	ATTGCCGGACGGAAGCA		protection
vanYD-02	CAAACGGAAGAGAGGTCACTTACA	CGGACGGTAATAGGGACTGTTC		protection

FCA: fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs). MLSB: macrolide-lincosamide-streptogramin B ARGs. *: For consistency, ARG categories are based on previous studies using the same Ht-qPCR assay. However, note that vancomycin is a specific antibiotic in the class 'glycopeptide'. The van operon is not confined to resistance to vancomycin and confers resistance to most glycopeptide antibiotics.

Appendix Table A-7. Box-cox transformations prior statistical analysis (significance testing and effect size calculations).

Parameter	Transformation
Physical parameters (river flow)	Log_{10}
Physical parameters (DO)	None
Chemical concentrations (DO, NH_3-N, TN, TP)	None
Chemical concentration (COD)	Log_{10}
Chemical mass loadings (NH_3-N, TN, TP, COD)	Log_{10}
Coliform concentrations and mass loadings (TC, ESBL coliform, CPB-0.5, CPB-2)	Log_{10}
Total antibiotic concentration and mass loading	Log_{10}
Detected ARGs and MGEs	None
ARG and MGE concentrations and mass loadings	Log_{10}
Normalised ARG and MGE cell concentrations	None

ARGs: antibiotic resistant genes. COD: chemical oxygen demand. CPB: carbapenem resistant bacteria. DO: dissolved oxygen. MGEs: mobile genetic elements. NH_3-N: ammonia. TC: total coliform. TN: total nitrogen. TP: total phosphate.

Appendix Table A-8. Physical parameters per sampling point (mean of biological replicates with standard deviations). Data for S1-S8 based on four biological replicates except for flow wet and dry season. Data for Se1 and M5 based on three biological replicates except for flow wet and dry season. Fold-change calculated by dividing the mean S8 with the mean S1 values.

	Flow (m³/s)			Temperature (ºC)	pH	Conductivity (µS/cm)	DO (mg/L)
	All (n= 3-4 per site)	Wet season (n= 1-2 per site)	Dry season (n=2 per site)				
S1	0.45 ± 0.29	0.59 ± 0.41	0.31 ± 0	27.18 ± 0.52	5.88 ± 0.44	75.53 ± 10.3	7.52 ± 0.5
S2	1.55 ± 0.96	1.88 ± 1.47	1.19 ± 0.16	27.88 ± 0.98	6.08 ± 0.59	297.78 ± 240.44	4.39 ± 0.62
S5	11.96 ± 6.47	7.83 ± 2.98	16.15 ± 6.92	27.45 ± 0.95	6.16 ± 0.39	357.25 ± 12.04	2.63 ± 1.05
S6	6.2 ± 4.16	8.16 ± 1.91	4.24 ± 5.73	28.1 ± 1.19	6.3 ± 0.4	299 ± 61.66	5.35 ± 0.22
S7	25.38 ± 24.94	40.39 ± 30.79	10.32 ± 2.84	28.12 ± 2.27	6.15 ± 0.35	1306.5 ± 1882.71	2.03 ± 0.59
S8	82.7 ± 30.66	78.07 ± 50.9	87.02 ± 9.46	28.67 ± 1.69	6.25 ± 0.36	3349.25 ± 4818.75	1.32 ± 0.26
Se1	2.64 ± 1.4	1.79	3.06 ± 1.68	27.03 ± 0.67	5.62 ± 0.26	97.07 ± 3.76	6.39 ± 0.55
M5	1.11 ± 0.26	0.81	1.23 ± 0.11	28.8 ± 2.33	6 ± 0.36	317.67 ± 23.29	1.61 ± 2.08
S8/S1	184	140	281	1	1	44	0.2 (S1/S8 = 6)

DO: dissolved oxygen.

Appendix Table A-9. Chemical concentrations per sampling point (mean of biological replicates with standard deviations). Total antibiotics summarize all antibiotics and antibiotic derivates detected in at least 60% of the samples above the detection limit (amoxicillin, ciprofloxacin, sulfamethoxazole, dehydrated erythromycin, lincomycin, trimethoprim, azithromycin, clarithromycin, clindamycin, sulfamethazine, ofloxacin, enrofloxacin). Data for S1-S8 based on four biological replicates. Data for Se1 and M5 based on three biological replicates. Fold-change calculated by dividing the mean S8 with the mean S1 values.

	COD (mg/L)	NH₃-N (mg/L)	TN (mg/L)	TP (mg/L)	Total antibiotics (mg/L)
S1	5.8 ± 4.77	0.05 ± 0.03	1.09 ± 0.51	0.43 ± 0.08	0.07 ± 0.05
S2	10.29 ± 6.46	2.85 ± 1.13	4.8 ± 2	0.95 ± 0.42	1.19 ± 1.67
S5	20.08 ± 6.92	4.7 ± 1.01	7.59 ± 0.63	2.25 ± 2.37	0.67 ± 0.25
S6	11.83 ± 5.67	3.08 ± 1.06	6.2 ± 1.27	6.94 ± 1.17	0.35 ± 0.16
S7	15.83 ± 7.17	3.65 ± 1.22	6.08 ± 1.71	5.25 ± 2.36	0.62 ± 0.16
S8	25.25 ± 16.04	4.94 ± 2	6.08 ± 2.36	3.95 ± 2.01	1.27 ± 0.98
Se1	9.61 ± 8.18	0.82 ± 0.09	1.92 ± 0.56	0.41 ± 0.18	0.88 ± 0.14
M5	28.67 ± 9.94	10.03 ± 1.06	12.11 ± 0.75	2.13 ± 0.46	2.79 ± 2.04
S8/S1	4	99	6	9	17

COD: chemical oxygen demand. NH₃-N: ammonia. TN: total nitrogen. TP: total phosphate.

Appendix Table A-10. Chemical mass loadings per sampling point (mean of biological replicates with standard deviations). Total antibiotics summarize all antibiotics and antibiotic derivates detected in at least 60% of the samples above the detection limit (amoxicillin, ciprofloxacin, sulfamethoxazole, dehydrated erythromycin, lincomycin, trimethoprim, azithromycin, clarithromycin, clindamycin, sulfamethazine, ofloxacin, enrofloxacin). Data for S1-S8 based on four biological replicates. Data for Se1 and M5 based on three biological replicates. Fold-change calculated by dividing the mean S8 with the mean S1 values.

	COD (kg/d)	NH$_3$-N (kg/d)	TN (kg/d)	TP (kg/d)	Total antibiotics (g/d)
S1	223 ± 180	2 ± 3	40 ± 22	17 ± 13	73 ± 46
S2	1,154 ± 633	329 ± 112	567 ± 232	113 ± 48	1,187 ± 1,669
S5	19,641 ± 8,009	4,914 ± 2666	7,924 ± 4,551	1,659 ± 878	668 ± 255
S6	5,026 ± 3,724	1,488 ± 1,142	3,234 ± 2,269	3,521 ± 2,273	348 ± 156
S7	42,955 ± 59,766	6,442 ± 4,378	11,045 ± 7,638	7,905 ± 2,650	622 ± 161
S8	205,902 ± 189,333	34,392 ± 17,377	42,610 ± 22,092	28,830 ± 17,632	1,266 ± 979
Se1	2,088 ± 1,516	183 ± 81	405 ± 124	90 ± 46	878 ± 141
M5	2,877 ± 1,584	973 ± 307	1,164 ± 306	210 ± 85	2,785 ± 2,039
S8/S1	923	14,270	1,060	1,650	17

COD: chemical oxygen demand. NH$_3$-N: ammonia. TN: total nitrogen. TP: total phosphate.

Appendix Table A-11. Coliform concentrations per sampling point in colony forming units (CFU) per mL river water (mean of biological replicates with standard deviation or minimum and maximum values). ESBL E. coli, CRB-0.5 *E. coli* and CRB-2 *E. coli* measurements were under detection limit in more than 40% of the sample, so the R2D substitution method was not applied. For these parameters we report minimum and maximum values instead of means. CRB-2 and CRB-2 *E. coli* was only measured for trips II-IV, so the mean and standard deviation are based on three biological replicates. For all other parameters, data for S1-S8 is based on four biological replicates and data for Se1 and M5 is based on three biological replicates. Fold-change calculated by dividing the mean S8 with the mean S1 values.

	TC	*E. coli*	ESBL coliform	ESBL *E. coli*	CRB-0.5	CRB-0.5 *E. coli*	CRB-2	CRB-2 *E. coli*
S1	(1.1 ± 0.5) x 10^3	(3.5 ± 2) x 10^1	(1.5 ± 1.3) x 10^2	(<0.5 − 2) x 10^1	(9 ± 3.8) x 10^1	<0.5 x 10^1	(3.1 ± 4.1) x 10^1	<0.5 x 10^1
S2	(1.7 ± 1.3) x 10^4	(3.4 ± 5.5) x 10^3	(3.5 ± 5.6) x 10^3	(<0.1 − 3.5) x 10^2	(0.9 ± 1.3) x 10^3	<0.5 / 1 x 10^1	(1.1 ± 0.9) x 10^2	<0.5 x 10^1
S5	(2.7 ± 0.6) x 10^4	(1.1 ± 0.2) x 10^3	(1.1 ± 0.8) x 10^3	(<1 − 8) x 10^1	(4.8 ± 3.8) x 10^2	<1 x 10^1	(7.1 ± 2.7) x 10^1	<0.5 x 10^1
S6	(1.4 ± 0.7) x 10^3	(3.3 ± 1.2) x 10^1	(5.4 ± 4.3) x 10^1	(<0.5 − 1) x 10^1	(4.3 ± 2) x 10^1	<0.5 x 10^1	(9 ± 8) x 10^0	<0.5 x 10^1
S7	(1.1 ± 0.1) x 10^4	(4.9 ± 1.3) x 10^2	(9.4 ± 5.4) x 10^2	(<0.05 − 1) x 10^2	(2.5 ± 1.9) x 10^2	<0.5 / 1 x 10^1	(4.2 ± 3.1) x 10^1	<0.5 x 10^1
S8	(4.1 ± 3.3) x 10^4	(2.8 ± 2.1) x 10^3	(4 ± 4.1) x 10^3	(<0.1 − 5) x 10^2	(3.8 ± 3.4) x 10^2	<0.5 / 1 x 10^1	(1.1 ± 0.2) x 10^2	<0.5 x 10^1
Se1	(1.1 ± 0.2) x 10^4	(6.4 ± 0.2) x 10^2	(7.2 ± 1.2) x 10^2	(<1 − 4) x 10^1	(4 ± 1.7) x 10^2	<1 x 10^1	(8.9 ± 1.5) x 10^1	<0.5 x 10^1
M5	(8 ± 2.7) x 10^4	(1.3 ± 0.3) x 10^4	(8.3 ± 5.9) x 10^3	(<0.1 − 6) x 10^2	(1.9 ± 1.5) x 10^3	(<1 − 1) x 10^1	(4 ± 1) x 10^2	(<0.5 − 1) x 10^1
S8/S1	37	80	27	NA	4	NA	3	NA

CRB: carbapenem resistant bacteria. CRB-0.5: CRB screened with 0.5 µg/mL meropenem. CRB-2: CRB screened with 2 µg/mL meropenem. ESBL: extended-spectrum β-lactamase. NA: not applicable. TC: total coliform (TC).

Appendix Table A-12. Coliform mass loading per sampling point in colony forming units (CFU) per river water per day (mean of biological replicates with standard deviation). ESBL *E. coli,* CRB-0.5 *E. coli* and CRB-2 *E. coli* mass loadings were not calculated as more than 40% of measurements were under detection limit. CRB-2 mass loadings were only measured for trips II-IV, so the mean and standard deviation data is based on three biological replicates. For all other parameters, data for S1-S8 is based on four biological replicates and data for Se1 and M5 is based on three biological replicates. Fold-change calculated by dividing the mean S8 with the mean S1 values. NA = not applicable.

	TC	*E. coli*	ESBL coliform	CRB-0.5	CRB-2
S1	$(4 \pm 2) \times 10^{13}$	$(1 \pm 0.5) \times 10^{12}$	$(4 \pm 3) \times 10^{12}$	$(4 \pm 4) \times 10^{12}$	$(0.8 \pm 1) \times 10^{12}$
S2	$(2 \pm 1) \times 10^{15}$	$(4 \pm 6) \times 10^{14}$	$(4 \pm 6) \times 10^{14}$	$(1 \pm 1) \times 10^{14}$	$(1 \pm 1) \times 10^{13}$
S5	$(3 \pm 0.9) \times 10^{15}$	$(1 \pm 0.7) \times 10^{15}$	$(1 \pm 0.9) \times 10^{15}$	$(5 \pm 4) \times 10^{14}$	$(1 \pm 0.8) \times 10^{14}$
S6	$(6 \pm 5) \times 10^{14}$	$(2 \pm 1) \times 10^{13}$	$(3 \pm 3) \times 10^{13}$	$(3 \pm 2) \times 10^{13}$	$(5 \pm 7) \times 10^{12}$
S7	$(2 \pm 2) \times 10^{16}$	$(9 \pm 7) \times 10^{14}$	$(2 \pm 0.9) \times 10^{15}$	$(4 \pm 2) \times 10^{14}$	$(6 \pm 5) \times 10^{13}$
S8	$(3 \pm 2) \times 10^{17}$	$(2 \pm 2) \times 10^{16}$	$(3 \pm 3) \times 10^{16}$	$(3 \pm 3) \times 10^{15}$	$(7 \pm 3) \times 10^{14}$
Se1	$(3 \pm 2) \times 10^{15}$	$(1 \pm 0.8) \times 10^{14}$	$(2 \pm 0.8) \times 10^{14}$	$(8 \pm 3) \times 10^{13}$	$(2 \pm 0.9) \times 10^{13}$
M5	$(8 \pm 3) \times 10^{15}$	$(1 \pm 0.5) \times 10^{15}$	$(9 \pm 8) \times 10^{14}$	$(2 \pm 2) \times 10^{14}$	$(4 \pm 0.6) \times 10^{13}$
S8/S1	7,500	20,000	7,500	750	875

CRB: carbapenem resistant bacteria. CRB-0.5: CRB screened with 0.5 µg/mL meropenem. CRB-2: CRB screened with 2 µg/mL meropenem. ESBL: extended-spectrum β-lactamase. NA: not applicable. TC: total coliform (TC).

Appendix Table A-13. Summarized antibiotic resistance gene (ARG) and mobile genetic element (MGE) levels in river water (mean of biological replicates with standard deviation). Data for S1-S8 based on four biological replicates. Data for Se1 and M5 based on three biological replicates. Fold-change calculated by dividing the mean S8 with the mean S1 values.

	Detected (number)		River water concentration (copies/mL)		River water mass loading (copies/d)		Normalised cell concentration (copies/cell)	
	ARG	MGE	ARG	MGE	ARG	MGE	ARG	MGE
S1	119 ± 14	10 ± 1	$(1.9 \pm 1.7) \times 10^5$	$(1.3 \pm 1.1) \times 10^5$	$(6.3 \pm 4.1) \times 10^{15}$	$(4.3 \pm 2.6) \times 10^{15}$	0.1 ± 0.1	0.1 ± 0
S2	161 ± 15	12 ± 1	$(3.9 \pm 3.3) \times 10^7$	$(3.1 \pm 3.6) \times 10^7$	$(4.5 \pm 3.5) \times 10^{18}$	$(3.6 \pm 4) \times 10^{18}$	1.3 ± 0.5	0.9 ± 0.7
S5	154 ± 7	11 ± 1	$(6.5 \pm 1.4) \times 10^7$	$(3.7 \pm 0.8) \times 10^7$	$(7.2 \pm 5.1) \times 10^{19}$	$(4.1 \pm 3.1) \times 10^{19}$	1.1 ± 0.2	0.7 ± 0.2
S6	120 ± 5	11 ± 0	$(7.5 \pm 4.6) \times 10^6$	$(6.9 \pm 2.7) \times 10^6$	$(4 \pm 3.7) \times 10^{18}$	$(3.7 \pm 3) \times 10^{18}$	0.7 ± 0.1	0.8 ± 0.2
S7	145 ± 3	11 ± 1	$(2.2 \pm 0.6) \times 10^7$	$(1.9 \pm 0.9) \times 10^7$	$(5.5 \pm 6.9) \times 10^{19}$	$(5.1 \pm 7.2) \times 10^{19}$	1.2 ± 0.2	1 ± 0.3
S8	150 ± 8	11 ± 1	$(1.2 \pm 0.9) \times 10^8$	$(1.1 \pm 0.9) \times 10^8$	$(8.6 \pm 7.2) \times 10^{20}$	$(8.1 \pm 7.3) \times 10^{20}$	1.7 ± 0.6	1.6 ± 0.6
Se1	165 ± 3	12 ± 0	$(2.2 \pm 1.5) \times 10^7$	$(1.4 \pm 0.8) \times 10^7$	$(4.4 \pm 7.1) \times 10^{18}$	$(2.9 \pm 1.3) \times 10^{18}$	2 ± 0.6	1.3 ± 0.3
M5	156 ± 9	12 ± 1	$(3.1 \pm 2.6) \times 10^8$	$(2.3 \pm 1.9) \times 10^8$	$(3.1 \pm 2.6) \times 10^{19}$	$(2.3 \pm 2) \times 10^{19}$	2.6 ± 1.2	1.9 ± 1
S8/S1	1.3	1.1	632	846	136,508	188,372	17	16

Appendix Table A-14. Detected antibiotic resistance genes (ARGs) per class in river water (mean of biological replicates with standard deviation). Data for S1-S8 based on four biological replicates. Data for Se1 and M5 based on three biological replicates. Numbers in brackets indicate total number of ARGs per class included in the high-throughput qPCR assay.

	Amino-glycoside (36)	β-Lactams (52)	FCA (9)	MLSB (46)	Non-specific (51)	Sulfon-amide (7)	Tetra-cycline (39)	Vanco-mycin (32)	Others (11)
S1	17 ± 3	25 ± 4	6 ± 1	14 ± 3	32 ± 1	3 ± 1	16 ± 3	4 ± 0	3 ± 1
S2	27 ± 2	34 ± 4	7 ± 1	21 ± 3	35 ± 1	4 ± 1	22 ± 4	7 ± 1	4 ± 1
S5	25 ± 1	34 ± 2	7 ± 1	20 ± 2	34 ± 2	4 ± 1	23 ± 2	4 ± 1	4 ± 0
S6	20 ± 1	24 ± 3	6 ± 1	15 ± 1	27 ± 2	3 ± 1	18 ± 2	5 ± 1	2 ± 1
S7	25 ± 1	31 ± 1	6 ± 1	18 ± 1	34 ± 1	4 ± 1	21 ± 1	4 ± 1	3 ± 1
S8	25 ± 1	33 ± 3	7 ± 1	18 ± 2	35 ± 1	4 ± 1	21 ± 2	4 ± 1	3 ± 0
Se1	25 ± 1	38 ± 4	7 ± 1	21 ± 2	35 ± 1	4 ± 1	26 ± 3	5 ± 1	4 ± 1
M5	25 ± 2	34 ± 2	7 ± 1	19 ± 3	34 ± 2	4 ± 0	24 ± 1	5 ± 3	4 ± 0

FCA: fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs. MLSB: macrolide-lincosamide-streptogramin B ARGs.

Appendix Table A-15. Antibiotic resistance gene (ARG) concentrations in river water (copies/mL) per class (mean of biological replicates with standard deviation). Data for S1-S8 based on four biological replicates. Data for Se1 and M5 based on three biological replicates.

	Amino-glycoside	β-Lactams	FCA	MLSB	Non-specific	Sulfon-amide	Tetra-cycline	Vanco-mycin	Others
S1	$(3.6 ± 3.7)$ x 10^4	$(3 ± 3.3)$ x 10^4	$(5.7 ± 4.7)$ x 10^3	$(7.5 ± 7.4)$ x 10^3	$(8 ± 5.7)$ x 10^4	$(1.1 ± 1.1)$ x 10^4	$(2.1 ± 1.4)$ x 10^4	$(3.4 ± 2.6)$ x 10^3	$(1 ± 0.9)$ x 10^3
S2	$(1.1 ± 1.2)$ x 10^7	$(4.3 ± 4.7)$ x 10^6	$(1.5 ± 1.8)$ x 10^6	$(1.2 ± 1.4)$ x 10^6	$(9 ± 7.6)$ x 10^6	$(9.2 ± 3.4)$ x 10^6	$(2.7 ± 3.5)$ x 10^6	$(6.8 ± 4)$ x 10^4	$(9.8 ± 15)$ x 10^4
S5	$(1.6 ± 0.3)$ x 10^7	$(6.3 ± 1.8)$ x 10^6	$(2 ± 0.6)$ x 10^6	$(2.1 ± 0.5)$ x 10^6	$(1.8 ± 0.32)$ x 10^7	$(1.6 ± 0.6)$ x 10^7	$(3.2 ± 0.8)$ x 10^6	$(1.6 ± 0.6)$ x 10^5	$(6 ± 1.2)$ x 10^4
S6	$(1.4 ± 0.7)$ x 10^6	$(3.7 ± 2.2)$ x 10^5	$(8.7 ± 4.4)$ x 10^4	$(2.8 ± 1.5)$ x 10^5	$(2.9 ± 1.7)$ x 10^6	$(2.1 ± 1.7)$ x 10^6	$(2.3 ± 0.9)$ x 10^5	$(2.9 ± 1.8)$ x 10^4	$(8.7 ± 7)$ x 10^3
S7	$(6.1 ± 1)$ x 10^6	$(2.2 ± 0.4)$ x 10^6	$(7.5 ± 1.6)$ x 10^5	$(9.2 ± 2.5)$ x 10^5	$(6.9 ± 2.1)$ x 10^6	$(3.7 ± 2.3)$ x 10^6	$(1.1 ± 0.4)$ x 10^6	$(6.7 ± 2.7)$ x 10^4	$(1.8 ± 0.7)$ x 10^4
S8	$(3.8 ± 2.7)$ x 10^7	$(1.6 ± 1.2)$ x 10^7	$(5.9 ± 4.3)$ x 10^6	$(4.8 ± 3.5)$ x 10^6	$(3.7 ± 2.5)$ x 10^7	$(1.7 ± 1.2)$ x 10^7	$(5.9 ± 3.4)$ x 10^6	$(2.4 ± 1.8)$ x 10^5	$(9 ± 5.4)$ x 10^4
Se1	$(8.8 ± 6.4)$ x 10^6	$(3.3 ± 2.5)$ x 10^6	$(1.6 ± 1.6)$ x 10^6	$(5.4 ± 2.4)$ x 10^5	$(5.6 ± 3.2)$ x 10^6	$(5.1 ± 2.9)$ x 10^5	$(1.5 ± 0.6)$ x 10^6	$(1.7 ± 0.9)$ x 10^4	$(1.8 ± 0.2)$ x 10^4
M5	$(9 ± 8.5)$ x 10^7	$(3.6 ± 2.8)$ x 10^7	$(9 ± 6.8)$ x 10^6	$(1.2 ± 1.2)$ x 10^7	$(9.1 ± 7.3)$ x 10^7	$(5.8 ± 4.1)$ x 10^7	$(1.4 ± 1.2)$ x 10^7	$(2.9 ± 1.4)$ x 10^5	$(3.1 ± 2.2)$ x 10^5

FCA: fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs. MLSB: macrolide-lincosamide-streptogramin B ARGs.

Appendix Table A-16. Normalised antibiotic resistance gene (ARG) cell concentrations in river water (copies/cell) per class (mean of biological replicates with standard deviation). Data for S1-S8 based on four biological replicates. Data for Se1 and M5 based on three biological replicates.

	Amino-glycoside	β-Lactams	FCA	MLSB	Non-specific	Sulfon-amide	Tetra-cycline	Vanco-mycin	Others
S1	0.02 ± 0.02	0.01 ± 0.01	0 ± 0	0 ± 0	0.04 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	0 ± 0	0 ± 0
S2	0.32 ± 0.22	0.13 ± 0.08	0.04 ± 0.04	0.04 ± 0.03	0.3 ± 0.11	0.37 ± 0.1	0.08 ± 0.07	0 ± 0	0 ± 0
S5	0.29 ± 0.07	0.11 ± 0.03	0.04 ± 0.01	0.04 ± 0.01	0.33 ± 0.03	0.29 ± 0.07	0.06 ± 0.01	0 ± 0	0 ± 0
S6	0.14 ± 0.05	0.04 ± 0.01	0.01 ± 0	0.03 ± 0.01	0.28 ± 0.07	0.19 ± 0.07	0.03 ± 0.01	0 ± 0	0 ± 0
S7	0.33 ± 0.05	0.12 ± 0.02	0.04 ± 0.01	0.05 ± 0.01	0.37 ± 0.05	0.19 ± 0.09	0.06 ± 0.01	0 ± 0	0 ± 0
S8	0.51 ± 0.24	0.22 ± 0.1	0.08 ± 0.04	0.07 ± 0.03	0.51 ± 0.15	0.24 ± 0.05	0.09 ± 0.02	0 ± 0	0 ± 0
Se1	0.78 ± 0.31	0.29 ± 0.12	0.13 ± 0.09	0.05 ± 0.01	0.51 ± 0.11	0.05 ± 0.01	0.15 ± 0.02	0 ± 0	0 ± 0
M5	0.71 ± 0.45	0.29 ± 0.13	0.07 ± 0.03	0.09 ± 0.07	0.79 ± 0.36	0.47 ± 0.18	0.12 ± 0.06	0 ± 0	0 ± 0

FCA: fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs. MLSB: macrolide-lincosamide-streptogramin B ARGs.

Appendix Table A-17. Mobile genetic element (MGE) levels in river water (mean of biological replicates with standard deviation). All measurements based on technical triplicates. Data for S1-S8 based on four biological replicates. Data for Se1 and M5 based on three biological replicates.

	Detected (number)		River water concentration (copies/mL)		Normalised cell concentration (copies/cell)	
	Int	Tran	Int	Tran	Int	Tran
S1	2 ± 1	8 ± 0	$(6 ± 4.1) \times 10^4$	$(7.4 ± 6.8) \times 10^4$	0.03 ± 0.02	0.04 ± 0.03
S2	4 ± 1	8 ± 0	$(7.6 ± 5.8) \times 10^6$	$(2.4 ± 3.1) \times 10^7$	0.26 ± 0.09	0.66 ± 0.61
S5	4 ± 1	8 ± 1	$(1.4 ± 0.4) \times 10^6$	$(2.3 ± 0.5) \times 10^7$	0.26 ± 0.07	0.41 ± 0.09
S6	3 ± 0	8 ± 0	$(1.7 ± 1) \times 10^6$	$(5.3 ± 1.8) \times 10^6$	0.16 ± 0.04	0.61 ± 0.22
S7	3 ± 1	8 ± 0	$(5.1 ± 2.5) \times 10^6$	$(1.4 ± 0.6) \times 10^7$	0.27 ± 0.09	0.73 ± 0.24
S8	3 ± 1	8 ± 0	$(3.5 ± 2.9) \times 10^7$	$(7.9 ± 5.8) \times 10^7$	0.47 ± 0.23	1.09 ± 0.38
Se1	4 ± 0	8 ± 0	$(5.7 ± 3.5) \times 10^6$	$(8.5 ± 4.7) \times 10^6$	0.51 ± 0.13	0.78 ± 0.2
M5	4 ± 1	8 ± 0	$(8.9 ± 7.6) \times 10^7$	$(1.4 ± 1.2) \times 10^8$	0.73 ± 0.37	1.19 ± 0.58

Int: integrons. Tran: transposases.

Appendix Table A-18. Seasonal effects. Paired t-tests and Cohen's D effect sizes comparing mass loadings (ML), concentrations (C) and detected numbers (D) for water quality and antibiotic resistant parameters between the dry and wet season. For applied transformations, see Appendix Table A-7.

				Paired t-test			Cohen's D effect size		
Type	Parameter	Type	Unit	P value	Degrees of freedom	t value	Value	95% confidence interval range	
Water quality	DO	C	mg/L	0.4200	7	-0.85662	-0.18	-0.64	0.27
	NH₃-N	C	mg/L	0.0328	7	2.65340	0.34	0.06	0.62
		ML	kg/d	0.6597	7	0.45969	0.03	-0.11	0.17
	COD	C	mg/L	0.8636	7	0.17825	0.07	-0.73	0.87
		ML	kg/d	0.5765	7	-0.58559	-0.08	-0.36	0.21
	TN	C	kg/d	0.0213	7	2.95380	0.41	0.1	0.72
		ML	mg/L	0.4207	7	0.85535	0.08	-0.12	0.27
	TP	C	kg/d	0.3615	7	0.97614	0.25	-0.3	0.79
		ML	mg/L	0.7489	7	-0.33296	-0.03	-0.22	0.16
Plating	TC	C	CFU/mL	0.0299	7	2.71750	0.36	0.07	0.66
		ML	CFU/d	0.2749	7	1.18430	0.15	-0.13	0.43
	E. coli	C	CFU/mL	0.0770	7	2.07240	0.3	-0.02	0.61
		ML	CFU/d	0.1381	7	1.67380	0.21	-0.06	0.47
	ESBL coliform	C	CFU/mL	0.0059	7	3.89890	0.67	0.26	1.07
		ML	CFU/d	0.0197	7	3.01030	0.4	0.1	0.7
	CRB-0.5	C	CFU/mL	0.1287	7	1.72210	0.39	-0.11	0.89
		ML	CFU/d	0.2891	7	1.14700	0.23	-0.21	0.67
Antibiotic	Total antibiotics	C	ng/L	0.0224	7	2.91970	0.9	0.12	1.69
		ML	g/d	0.0746	7	2.09300	0.35	-0.02	0.72
S16 rRNA	S16 rRNA	C	copies/mL	0.0200	7	2.99840	0.32	0.08	0.55
		ML	copies/d	0.3778	7	0.94143	0.12	-0.16	0.4
Detected MGEs/ARGs	Detected ARGs	D	NA	0.0124	7	3.34140	0.49	0.16	0.82
	Detected MGEs	D	NA	0.0796	7	2.04940	0.67	-0.1	1.44
Abundance MGEs/ARGs	Abundance ARGs	C	copies/mL	0.0105	7	3.46360	0.34	0.12	0.56
		ML	copies/d	0.1449	7	1.64040	0.19	-0.06	0.43
	Abundance MGEs	C	copies/mL	0.0073	7	3.73210	0.4	0.16	0.64
		ML	copies/d	0.1013	7	1.88570	0.22	-0.03	0.46
Normalised ARGs/MGEs	Normalised ARGs	C	copies/cell	0.0239	7	2.87330	0.43	0.09	0.77
		C	copies/cell	0.0329	7	2.65190	0.57	0.07	1.08

ARGs: antibiotic resistant genes. CFU: colony forming units. COD: chemical oxygen demand. CRB-0.5: carbapenem resistant bacteria selected for with 0.5 µg/mL meropenem. ESBL: extended-spectrum β-lactamase. DO: dissolved oxygen. MGE: mobile genetic elements. NA: not applicable. NH₃-N: ammonia. TC: total coliform. TN: total nitrogen. TP: total phosphorus.

Appendix Table A-19. Spatial effects. Welch's t-test and Cohen's D effect size effect-sizes comparing mass loadings (ML), concentrations (C) and detected numbers (D) for water quality and antibiotic resistant parameters between the up- (S1) and downstream (S8). For applied transformations, see Appendix Table A-7.

Type	Parameter	Type	Unit	Welch's t-test			Cohen's D effect-size		
				P value	Degrees of freedom	t value	Value	95% confidence interval range	
Water quality	DO	C	mg/L	0.0000	4.52	22.02	15.57	5.89	25.25
	NH3-N	C	mg/L	0.0163	3.00	-4.89	-3.46	-6.19	-0.73
		ML	kg/d	0.0000	4.84	-17.49	-12.37	-20.13	-4.61
	COD	C	mg/L	0.0895	3.53	-2.33	-1.64	-3.65	0.36
		ML	kg/d	0.0001	5.99	-9.77	-6.91	-11.47	-2.34
	TN	C	kg/d	0.0218	3.28	-4.12	-2.91	-5.39	-0.43
		ML	mg/L	0.0000	5.99	-17.33	-12.25	-19.95	-4.56
	TP	C	kg/d	0.0393	3.01	-3.50	-2.47	-4.77	-0.18
		ML	mg/L	0.0000	5.80	-14.70	-10.39	-16.98	-3.80
Plating	Coliform	C	CFU/mL	0.0067	3.95	-5.21	-3.68	-6.52	-0.84
		ML	CFU/d	0.0001	4.38	-13.79	-9.75	-15.96	-3.54
	E. coli	C	CFU/mL	0.0104	3.68	-4.83	-3.42	-6.13	-0.70
		ML	CFU/d	0.0004	3.63	-12.36	-8.74	-14.36	-3.12
	ESBL coliform	C	CFU/mL	0.0131	5.85	-3.52	-2.49	-4.79	-0.18
		ML	CFU/d	0.0001	5.66	-9.55	-6.75	-11.23	-2.27
	CRB-0.5	C	CFU/mL	0.0888	4.21	-2.20	-1.56	-3.53	0.42
		ML	CFU/d	0.0001	5.99	-9.93	-7.02	-11.66	-2.39
Antibiotic	Antibiotics	C	ng/L	0.0032	5.75	-4.86	-3.43	-6.16	-0.71
		ML	g/d	0.0001	4.55	-12.61	-8.92	-14.64	-3.19
S16 rRNA	S16 rRNA	C	copies/mL	0.0025	3.57	-7.63	-5.40	-9.12	-1.67
		ML	copies/d	0.0000	4.89	-19.67	-13.91	-22.59	-5.23
Detected MGE/ARG	Detected ARG	D	NA	0.0141	4.55	-3.87	-2.74	-5.14	-0.33
		D	NA	0.0300	6.00	-2.83	-2.00	-4.12	0.12
Abundance MGE/ARG	Abundance ARG	C	copies/mL	0.0002	5.17	-9.20	-6.50	-10.84	-2.17
		ML	copies/d	0.0000	5.13	-18.04	-12.75	-20.74	-4.76
	Abundance MGE	C	copies/mL	0.0001	5.31	-9.68	-6.85	-11.38	-2.32
		ML	copies/d	0.0000	4.99	-18.07	-12.78	-20.79	-4.77
Normalised ARG/MGE	Normalised ARG	C	copies/cell	0.0120	3.07	-5.36	-3.79	-6.68	-0.90
		C	copies/cell	0.0120	3.07	-5.36	-3.47	-6.21	-0.73

ARGs: antibiotic resistant genes. CFU: colony forming units. COD: chemical oxygen demand. CRB-0.5: carbapenem resistant bacteria selected for with 0.5 µg/mL meropenem. ESBL: extended-spectrum β-lactamase. DO: dissolved oxygen. MGE: mobile genetic elements. NA: not applicable. NH$_3$-N: ammonia. TC: total coliform. TN: total nitrogen. TP: total phosphorus.

Appendix Table A-20. Spearman correlations between river water concentrations of total antibiotics, amoxicillin and ciprofloxacin (both antibiotics detected in the catchment above their PNECs), total antibiotic resistant genes (ARGs) and ARGs reported by antibiotic class for the river catchment (n=30). Correlation values only shown for P < 0.05 with P values corrected for multiple testing with the Benjamini Hochberg approach.

	Total antibiotics	Amoxicillin	Ciprofloxacin
ARG	0.70*	0.39	0.56*
Aminoglycoside	0.78*	0.49*	0.6*
β-Lactam	0.76*	0.46	0.62*
FCA	0.79*	0.48*	0.61*
MLSB	0.74*	0.40	0.58*
Non-specific	0.68*	-	0.56*
Sulfonamide	0.51*	-	0.5*
Tetracycline	0.77*	0.48*	0.57*
Vancomycin	0.58*	-	0.55*
Other	0.71*	0.41	0.58*

*: P < 0.01. FCA: fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol ARGs. MLSB: macrolide-lincosamide-streptogramin B ARGs.

Appendix B
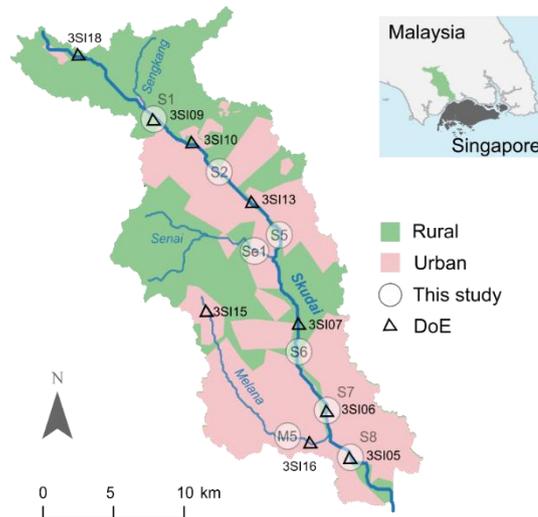
Appendix Figure B-1. Skudai catchment in Malaysia with sampling points from this study (for details see Chapter 3) and selected national Department of Environment (DoE) water quality monitoring sites (102°59'54.19" E and 104°11'8.54" E longitude and 1°56'31.67" N and 1°22'41.16" N latitude).
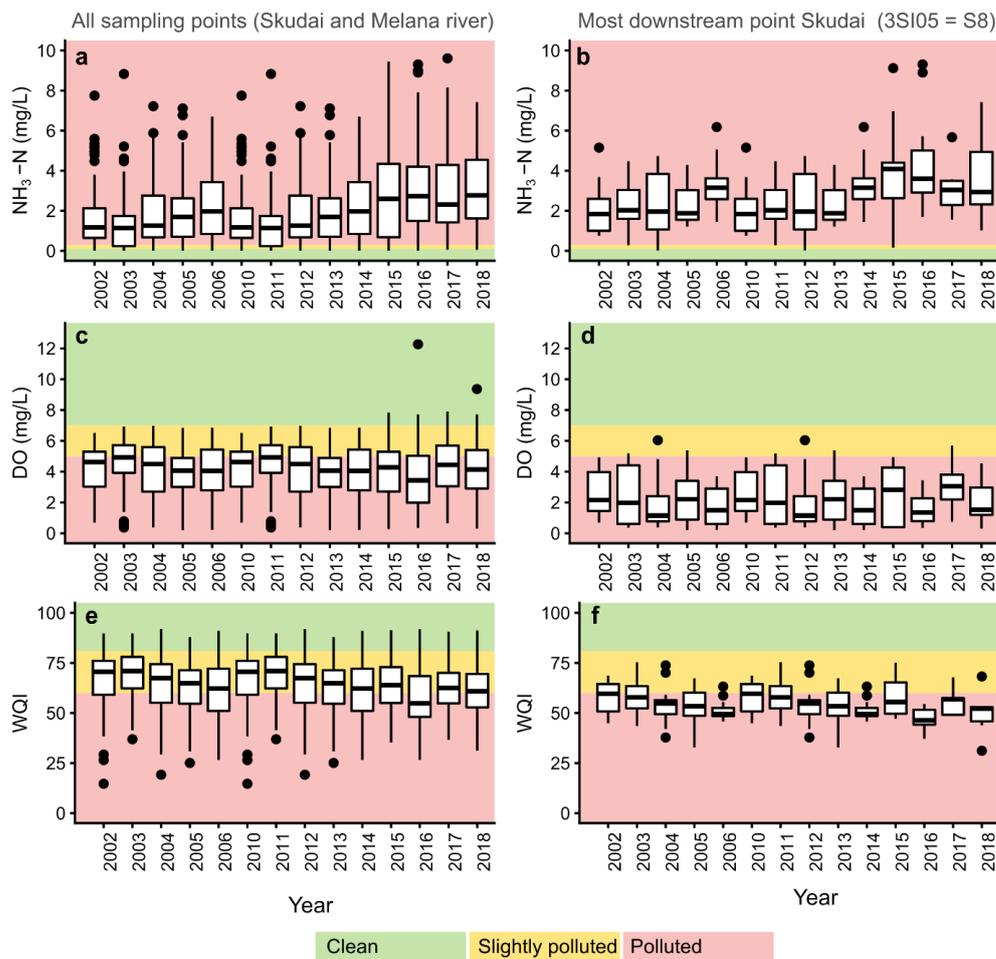
Appendix Figure B-2. Annual ammonia (NH_3-N), dissolved oxygen (DO) and water quality index (WQI) levels for the Skudai catchment (a,c,e based on national monitoring stations 3SI09, 3SI10, 3SI13, 3SI07, 3SI06, 3SI05 and 3SI16) and at the most downstream Skudai sampling point (b,d,f) for 2002-2006 and 2010-2018. For panels a, c and e, boxplots represent 84 measurements per year except for 2015 (91), 2017 (49) and 2018 (42). For panels b, d and f, boxplots represent 12 measurements per year except for 2015 (13), 2017 (7) and 2018 (6).

a) NH₃-N x residuals for simple regression

b) Histogram of residuals

c) Normal Q-Q Plot
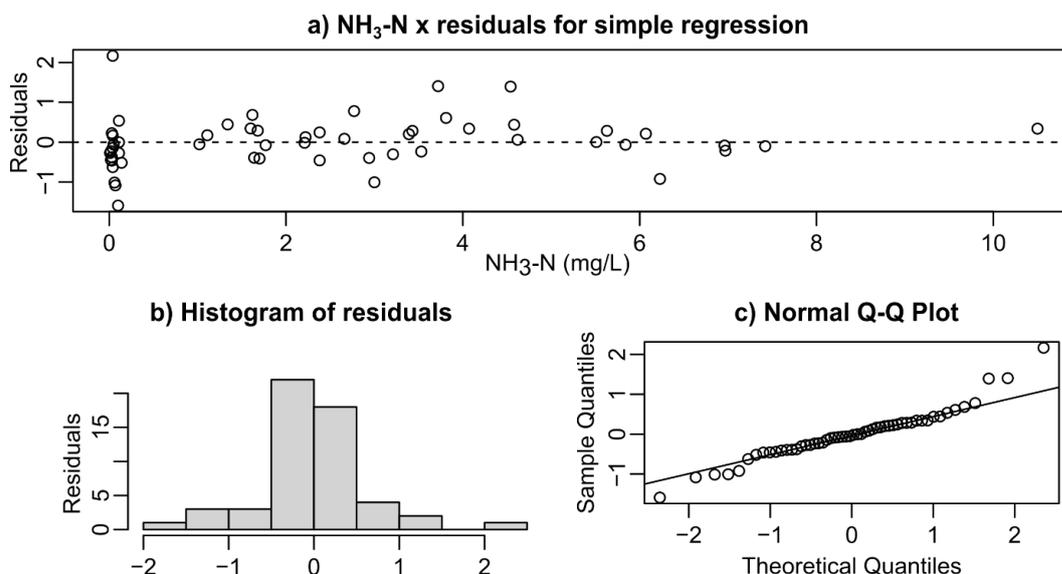
Appendix Figure B-3. Residual analysis of the linear regression model NH_3-N and coliform ($\log_{10}$ transformed) concentrations for the Department of Environment (DoE) Skudai catchment water quality dataset 2018.

Appendix Table B-1. HSPF modelling options (based on [312]).

Pervious land segments	• Air temperature as a function of elevation • Snow accumulation and melting • Hydrological cycle components (evapotranspiration, surface detention surface runoff, infiltration, interflow, base flow, percolation to deep groundwater) • Sediment production and removal • Soil temperature • Surface water temperature, dissolved oxygen and carbon dioxide concentrations in overland flow • Generalized water quality constituents modelled as accumulated storages removed by flow or potency factors associated with sediment • More detailed modelling of pesticide processes (runoff, leaching, adsorption/desorption, degradation) • Nutrient processes (transport by flow and sediment association, leaching, adsorption/desorption, denitrification, nitrification, plant uptake, immobilization, mineralization) • Tracer elements
Impervious land segments	• Air temperature as a function of elevation • Snow accumulation and melting • Water budget (surface components only) • Solids accumulation and removal including methods that are independent of storm events • Surface water temperature and gas concentrations • Generalized water quality constituents
Channel segments	• Hydraulic behaviour using the kinematic wave assumption • Longitudinal advection of dissolved and entrained constituents • Water temperature using a heat balance approach (absorption of shortwave radiation, longwave radiation, emission of longwave radiation, conduction-convection and evaporation) • Inorganic sediment deposition, scour and transport by particle size • Partitioning, hydrolysis, volatilization, oxidation, biodegradation, first-order decay, and parent chemical/metabolite transformations for generalized chemicals • Dissolved oxygen and BOD processes (decay, settling, benthal sinks and sources, re-aeration, sinks and sources related to plankton metabolism) • Nitrogen processes (ammonia volatilization, ammonification, denitrification, ammonia adsorption/desorption with suspended sediment) • Phosphate adsorption/desorption with suspended sediment; phytoplankton processes (growth, respiration, sinking, zooplankton predation, death) • Zooplankton process (growth, respiration, death) • Benthic algae processes (growth, respiration, death) • Carbon dioxide-bicarbonate system processes (carbon dioxide invasion, zooplankton respiration, BOD decay, net growth of algae, and benthal releases) that determine pH

Appendix Table B-2. National Department of Environment (DoE) water quality monitoring sampling campaigns per sampling point and year.

	Skudai upstream → downstream							Melana	
	3SI18	3SI09 (=S1)	3SI10	3SI13	3SI07	3SI06 (=S7)	3SI05 (=S8)	3SI15	3SI16
2002	6	12	12	12	12	12	12	6	12
2003	6	12	12	12	12	12	12	6	12
2004	7	12	12	12	12	12	12	6	12
2005	6	12	12	12	12	12	12	6	12
2006	6	12	12	12	12	12	12	7	12
2010	6	12	12	12	12	12	12	6	12
2011	6	12	12	12	12	12	12	6	12
2012	7	12	12	12	12	12	12	6	12
2013	6	12	12	12	12	12	12	6	12
2014	6	12	12	12	12	12	12	7	12
2015	5	12	12	12	12	12	12	5	12
2016	4	12	12	12	12	12	12	4	12
2017	5	7	7	7	7	7	7	5	7
2018	6	6	6	6	6	6	6	6	6
All	82	158	158	158	158	158	158	82	158

Appendix Table B-3. Overview of input data applied to operate the HSPF Skudai model.[179]

Data type	Period cover	Time step	Sources
Meteorological data			
Rainfall	1987/99 - 2015	Hourly from five rain gauges (not all covering the full period cover)	Department of Irrigation and Drainage and Malaysia Meteorological Department
Cloud cover	1999 - 2015	Hourly	National Climate Data Online, National Centers for Environmental Information
Dew temperature	1999 - 2015	Hourly	National Climate Data Online, National Centers for Environmental Information
Air temperature	1999 - 2015	Hourly	National Climate Data Online, National Centers for Environmental Information
Wind speed	1999 - 2015	Hourly	National Climate Data Online, National Centers for Environmental Information
Evaporation	1987 - 2015	Daily	Department of Irrigation and Drainage
Atmospheric Deposition			
Wet atmospheric deposition	2002 - 2015	Monthly	Malaysia Meteorological Department
Dry atmospheric deposition	2002 - 2015	Monthly	Department of Environment
Hydrological data			
Stream flow/water level, station 1636401 located at Kampong Separa	2002 – 2015 (missing data 2007 – 2008)	Monthly	Department of Irrigation and Drainage
Stream flow/water level, UTM station at Kampong Pertanian	2012 - 2014	Monthly/hourly	Department of Hydraulics and Hydrology, Faculty of Civil Engineering (FKA) of Universiti Teknologi Malaysia (UTM)
Topography, soil and land classification			
Digital elevation model (DEM)	2010	7.5 min, 1 arc sec interval and 30 m resolution	Global Data Explorer (Land Processes Distributed Active Archive Centre (LP DAAC), National Aeronautics and Space Administration (NASA), USGS data

			centre, George Mason University's Centre for Spatial Information Science and Systems)
Soil data	1970	1:250,000	Soil survey division of Ministry of Agriculture and Fisheries
Land use through remote sensing data (Landsat 4-5, with TM sensor Landsat 7 with ETM+ sensor and Landsat 8 with OLI sensor)	1989, 1999, 2009, 2013, 2015	30 x 30 m	Obtained from USGS EROS Data Centre (EDC) and accessed via USGS Global Visualization Viewer (GLOVIS)
Water quality data			
Water temperature, DO, BOD, nitrate nitrogen, ammonia nitrogen, orthophosphate at four stations	2002 - 2014	Monthly	Department of Environment
Point source data			
Wastewater treatment plants for 121 sources	Not applicable	Not applicable	Indah Water Konsortium (IKW)

Appendix C

Appendix Figure C-1. Cell concentration and reads per sample (a) and sampling depth per sampling point (b). For (a), Samples ranked according to decreasing cell concentration. Cell concentration and reads per sample correlated moderately (n = 38, Spearman ρ = 0.52, P = 0.0007528). For sampling depth (b), reads per sample (=sequencing depth) was divided by cell concentration. Data represented is based on five biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on four biological replicates for the tributaries (Se1, M5). Box-plot elements are defined as center line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range) and points (outliers).

Appendix Figure C-2. Barplots showing the 20 most abundant ASVs grouped into families with remain pooled into 'Other' for the relative (RMP; a) and quantitative (QMP; b) microbiome profiling approach. Data represented (n = 38) is based on five biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on four biological replicates for the tributaries (Se1, M5).

Appendix Figure C-3. Rarefaction curves (a) and alpha diversity indices (richness, Shannon index and Simpson index) for the QMP (b) and RMP (c) dataset. Data represented (n = 38) is based on five biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on four biological replicates for the tributaries (Se1, M5). Box-plot elements are defined as centre line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range) and points (outliers).
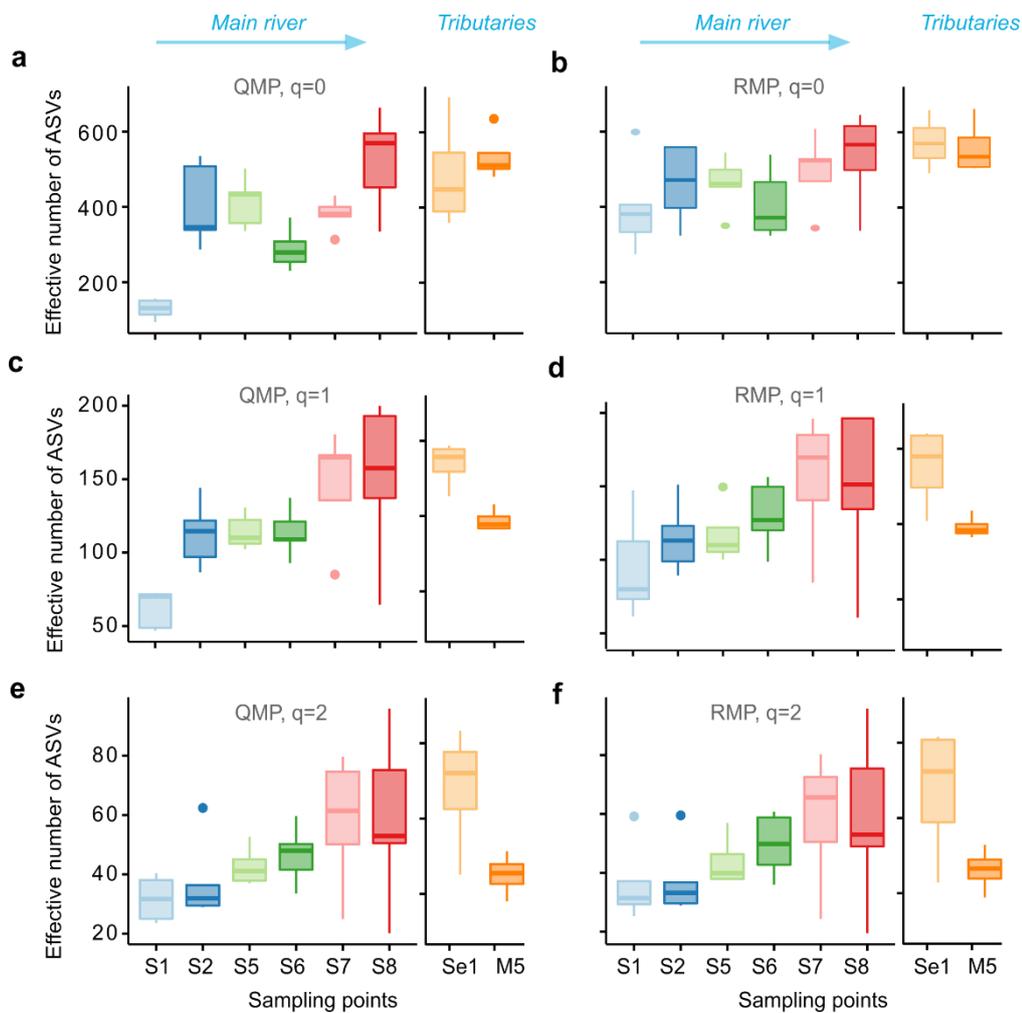
Appendix Figure C-4. Hill number α-diversities per sampling point for q = 0, q = 1 and q = 2 for the quantitative microbiome profiling (QMP) approach (a, c, e) and the relative microbiome profiling (RMP) approach (b, d, f). Data represented (n = 38) is based on five biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on four biological replicates for the tributaries (Se1, M5). Box-plot elements are defined as center line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range) and points (outliers).
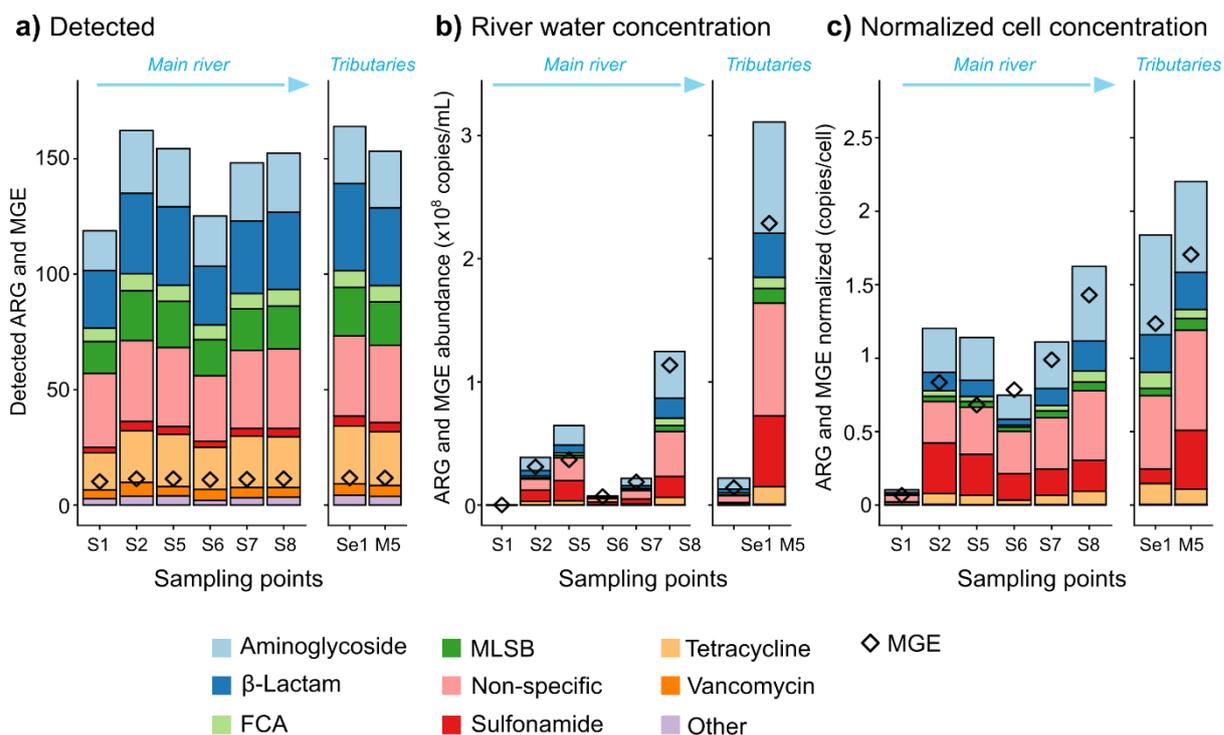
Appendix Figure C-5. ARG and MGE numbers (a), river water concentrations (b) and normalised cell concentration (c) detected with HT-qPCR per sampling point in the Skudai catchment. Mean data represented is based on five biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on four biological replicates for the tributaries (Se1, M5).

Appendix Table C-1. List of the 20 most abundant ASVs classified to genus level based on the QMP and RMP approach. For genus level, numbers in brackets were added to differentiate ASVs with the same genus. where grey colouring for the QMP or RMP listing highlights ASVs not present in the RMP or QMP listing, respectively. * indicates where ASV rank order is the same in the QMP and RMP listing. ND = not classified to genus level.

ASV	Phylum	Class	Order	Family	Genus
Top 20 ASVs based QMP approach					
1	Bacteroidota	Bacteroidia	Flavobacteriales	*Weeksellaceae*	*Cloacibacterium (1)*
2	Proteobacteria	Gammaproteobacteria	Pseudomonadales	*Moraxellaceae*	*Acinetobacter (1)*
3	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (1) **
4	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*ND (1)*
5	Proteobacteria	Alphaproteobacteria	Sphingomonadales	*Sphingomonadaceae*	*Novosphingobium (1)**
6	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (2)**
7	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (3)*
8	Bacteroidota	Bacteroidia	Flavobacteriales	*Flavobacteriaceae*	*Flavobacterium (1)*
9	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (4)*
10	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	*Phormidiaceae*	*Planktothrix_NIVA-CYA_15 (1)*
11	Actinobacteriota	Actinobacteria	Corynebacteriales	*Mycobacteriaceae*	*Mycobacterium (1)*
12	Proteobacteria	Gammaproteobacteria	Pseudomonadales	*Moraxellaceae*	*Acinetobacter (2)*
13	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*ND (2)*
14	Proteobacteria	Gammaproteobacteria	Aeromonadales	*Aeromonadaceae*	*Tolumonas (1)*
15	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	*Selenomonadaceae*	*ND (3)*

137

ASV	Phylum	Class	Order	Family	Genus
16	Campilobacterota	Campylobacteria	Campylobacterales	*Arcobacteraceae*	*Arcobacter (1)*
17	Campilobacterota	Campylobacteria	Campylobacterales	*Arcobacteraceae*	*Pseudarcobacter (1)*
18	Proteobacteria	Alphaproteobacteria	Sphingomonadales	*Sphingomonadaceae*	*Novosphingobium (1)*
19	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*Aquabacterium (1)*
20	Bacteroidota	Bacteroidia	Flavobacteriales	*Flavobacteriaceae*	*Flavobacterium (1)*
Top 20 ASVs based RMP approach

ASV	Phylum	Class	Order	Family	Genus
1	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*ND (1)*
2	Bacteroidota	Bacteroidia	Flavobacteriales	*Weeksellaceae*	*Cloacibacterium (1)*
3	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (1)**
4	Proteobacteria	Gammaproteobacteria	Pseudomonadales	*Moraxellaceae*	*Acinetobacter (1)*
5	Proteobacteria	Alphaproteobacteria	Sphingomonadales	*Sphingomonadaceae*	*Novosphingobium (1)**
6	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (2)**
7	Actinobacteriota	Actinobacteria	Corynebacteriales	*Mycobacteriaceae*	*Mycobacterium (1)*
8	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*ND (4)*
9	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*ND (5)*
10	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Burkholderiaceae*	*Polynucleobacter (1)*
11	Proteobacteria	Gammaproteobacteria	Pseudomonadales	*Moraxellaceae*	*Acinetobacter (2)*
12	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*ND (2)*
13	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (3)*
14	Bacteroidota	Bacteroidia	Flavobacteriales	*Flavobacteriaceae*	*Flavobacterium (1)*
15	Cyanobacteria	Cyanobacteriia	Synechococcales	*Cyanobiaceae*	*Cyanobium_PCC-6307 (1)*
16	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Rhodocyclaceae*	*C39 (4)*
17	Proteobacteria	Gammaproteobacteria	Aeromonadales	*Aeromonadaceae*	*Tolumonas (1)*
18	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	*Phormidiaceae*	*Planktothrix_NIVA-CYA_15 (1)*
19	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Comamonadaceae*	*ND (6)*
20	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	*Selenomonadaceae*	*ND (3)*

Appendix Table C-2. Results from the Welch's t-test and Cohen's D effect size calculations comparing Hill number diversities upstream (S1) and downstream (S8). Comparisons based on five biological replicates for each site. Statistical significance for $P < 0.05$ and large effect size for $D < -0.8$.

	Welch's t-test			Cohen's D effect size		
	P value	Degrees of freedom	t value	Value	95% confidence interval range	
QMP q=0	0.0021	4.31	-6.62	-4.19	-6.80	-1.58
QMP q=1	0.0203	4.45	-3.53	-2.23	-4.09	-0.38
QMP q=2	0.0993	4.56	-2.06	-1.31	-2.91	0.30
RMP q=0	0.1245	8.00	-1.72	-1.09	-2.65	0.48
RMP q=1	0.1175	6.71	-1.80	-1.14	-2.71	0.44
RMP q=2	0.1747	5.65	-1.55	-0.98	-2.53	0.56

Appendix Table C-3. ARG and MGE numbers included in the HT-qPCR assay, detected in each river water sample and maximum amount detected.

	Gene number included in HT-qPCR assay	Detected in each river water sample (n=38)		Maximum amount detected in catchment	
		Number	% of assay	Number	% of assay
ARGs	283	70	25	211	75
• Aminoglycoside	36	13	36	31	86
• β-Lactams	52	14	27	45	87
• FCA	9	4	44	9	100
• MLSB	46	9	20	29	63
• Non-specific	51	14	27	39	76
• Others	11	1	9	8	73
• Sulfonamide	7	1	14	6	86
• Tetracycline	39	12	31	31	79
• Vancomycin	32	2	6	13	41
MGEs	12	9	75	12	100
• Integrons	4	2	50	4	100
• Transposase	8	7	88	8	100
ALL	578	79	14	223	39

Appendix Table C-4. Results from the Welch's t-test and Cohen's D effect size calculations comparing $\log_{10}$ ARG and MGE river water concentrations (gene copies/mL), ARG and MGE detected numbers, and ARG and MGE normalised cell concentrations for up-(S1) and downstream (S8). Comparisons based on five biological replicates for each site. Statistical significance for $P < 0.05$ and large effect size for $D < -0.8$.

	Welch's t-test			Cohens D effect size		
	P value	Degrees of freedom	t value	Exact value	95% confidence interval range	
ARG $\log_{10}$ river water concentration (copies/mL)	<0.0001	6.91	-11.75	-7.43	-11.53	-3.33
MGE $\log_{10}$ river water concentration (copies/mL)	<0.0001	7.19	-12.07	-7.63	-11.83	-3.44
Detected ARGs (number)	0.0015	7.06	-5.01	-3.17	-5.36	-0.98
Detected MGEs (number)	0.0057	7.69	-3.79	-2.40	-4.31	-0.49
ARG normalised cell concentration (copies/cell)	0.0035	4.10	-6.06	-3.83	-6.29	-1.38
MGE normalised cell concentration (copies/cell)	0.0071	4.03	-5.06	-3.20	-5.40	-1.00

Appendix Table C-5. ARG and MGE detected numbers, river water concentrations and normalised cell concentrations per sampling point in the Skudai catchment. Mean and standard deviations based on five biological replicates for the main river (S1, S2, S5, S6, S7, S8) and on four biological replicates for the tributaries (Se1, M5).

	Detected (number)		River water concentration (copies/mL)		Normalised cell concentration (copies/cell)	
	ARGs	MGEs	ARGs	MGEs	ARGs	MGEs
S1	118.8 ± 12.4	10.2 ± 0.4	$(1.8 ± 1.5) \times 10^5$	$(1.2 ± 1) \times 10^5$	0.1 ± 0.06	0.07 ± 0.04
S2	162.2 ± 13.5	11.4 ± 0.5	$(3.4 ± 3.1) \times 10^7$	$(2.6 ± 3.3) \times 10^7$	1.2 ± 0.42	0.84 ± 0.61
S5	154.4 ± 6.2	11.4 ± 0.9	$(5.4 ± 2.6) \times 10^7$	$(3.1 ± 1.5) \times 10^7$	1.14 ± 0.14	0.68 ± 0.14
S6	125.2 ± 13.1	11.0 ± 0.0	$(8.6 ± 4.7) \times 10^6$	$(8 ± 3.4) \times 10^6$	0.75 ± 0.14	0.79 ± 0.19
S7	148.2 ± 8.6	11.2 ± 0.4	$(2.2 ± 0.6) \times 10^7$	$(2.1 ± 0.9) \times 10^7$	1.11 ± 0.19	0.99 ± 0.28
S8	152.4 ± 8.4	11.4 ± 0.5	$(1.1 ± 0.8) \times 10^8$	$(1 ± 0.8) \times 10^8$	1.62 ± 0.56	1.43 ± 0.6
Se1	164.0 ± 3.7	11.8 ± 0.5	$(5.7 ± 7) \times 10^7$	$(3.9 ± 5.1) \times 10^7$	1.84 ± 0.57	1.24 ± 0.27
M5	153.3 ± 9.3	11.8 ± 0.5	$(2.4 ± 2.5) \times 10^8$	$(1.8 ± 1.9) \times 10^8$	2.2 ± 1.22	1.7 ± 0.89

Appendix Table C-6. Ten most abundant ARGs in the Skudai catchment based on the mean river water concentration (n = 38).

Gene name	Classification	Mechanism	Mean ARG copies/mL	Standard deviation ARG copies/mL
sul2	Sulfonamide	Cellular protection	1.1×10^7	1.8×10^7
qacEdelta1_02	Non-specific	Efflux pump	7.2×10^6	1.3×10^7
qacEdelta1_01	Non-specific	Efflux pump	7×10^6	1.3×10^7
aadA2_03	Aminoglycoside	Antibiotic deactivate	4.2×10^6	8.1×10^6
aadA1	Aminoglycoside	Antibiotic deactivate	3.2×10^6	6×10^6
qacH_02	Non-specific	Efflux pump	2.3×10^6	3.6×10^6
aadA2_01	Aminoglycoside	Antibiotic deactivate	2.1×10^6	4.5×10^6
aadA2_02	Aminoglycoside	Antibiotic deactivate	$2. \times 10^6$	4.4×10^6
blaOXA10_01	β-Lactam	Antibiotic deactivate	$2. \times 10^6$	3.6×10^6
blaOXA10_02	β-Lactam	Antibiotic deactivate	1.7×10^6	2.9×10^6

Appendix Table C-7. Properties for the ARG, MGE and taxa (order level) networks based on the relative microbiome profiling (RMP) approach and quantitative microbiome profiling (QMP) approach. Only nodes with at least three other connections are shown.

	RMP	QMP
Nodes	153	176
• ARGs	130	130
• Transposase	7	7
• Integrons	3	3
• Taxa	13	36
Edges	7690	9455
Network diameter (maximum distance between edges)	5	3
Average path length	1.4	1.37
Average degree (node connectivity)	100.523	107.443
Graph density	0.661	0.614
Modularity index	0.072	0.094