Antimicrobial resistance: molecular approaches to track antimicrobial resistance gene spread from decentralised septic tank wastewater

Valentine Obinna Okonkwo

BSc (Hons), MSc

Submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy (PhD) by Research

Supervised by Prof. Cindy J. Smith, Dr Stephanie Connelly, Prof. William Sloan

James Watt School of Engineering

College of Science and Engineering

University of Glasgow

Submitted: May 2023

Abstract

Antimicrobial resistance (AMR) is a major global public health and wastewater treatment (WWT), including septic tanks, are now recognised as hotspots and potential sources of AMR genes to the environment. However, compared to centralised WWT (e.g., municipal WWT), an in-depth understanding of the contributions of septic tanks in the dissemination of AMR and mobile genetic elements (MGEs) remains scarce. Nonetheless, effective AMR gene monitoring from polluted settings such as WWT to the environment remains challenging primarily due to multiple AMR genes found in WWT. The class 1 integron-integrase (*intI*1) gene was proposed as a proxy for inferring potential AMR pollution elevates challenges associated with multiple monitoring. Yet, the suitability of this gene as an adequate and reliable proxy for inferring AMR pollution remains unclear.

To this end, this thesis focused on using state-of-the-art molecular tools to:- 1) Evaluate and validate the suitability of the *intI*1 gene as a proxy for inferring overall AMR abundance using wastewater samples from septic tanks from Thailand treating household and healthcare wastewater; and 2) Evaluate the contributions of conventional septic tanks (CST) associated with household and healthcare usage, and the newly developed solar septic tank (SST) associated with household usage in the dissemination of AMR genes and MGE to the environment.

The results from this study proposed one primer set (F3-R3) for *intI*1 quantification of genes and transcripts. However, it found that none of the current *intI*1 primers can distinguish between *intI*1 (highly conserved *intI*1 variant; >98% protein similarity to pVS1 *intI*1 reference protein) and *intI*1-like (lesser conserved *intI*1 variant; <98% protein similarity to pVS1), therefore, potentially contributes to an overestimation of quantified *intI*1 gene abundance. Furthermore, the relative abundance (relative to the *16S rRNA* gene) of a fewer number of AMR genes correlated positively and significantly to the abundance of *intI*1 compared to *intI*3. Therefore, taken together, indicates that *intI*1 cannot serve as a proxy for overall AMR gene abundance.

The septic tanks were found significant source of AMR gene subtypes and abundance. However, depending on the molecular method used the tank posing the highest risk of AMR or integrase, dissemination to the environment differed.

HT-QPCR, after careful validation of the array, identified the CST-household tank, among the three reactors, as potentially the higher contributor of AMR and integrases (*intI*1 and *intI*3) gene abundance to the environment via its sludge and effluent. In contrast, shotgun metagenomics identified the CST-healthcare septic tank, among the three reactors, as potentially the highest contributor of ARG abundance to the environment through its effluent and sludge (if applied directly to the environment). Therefore, emphasises the trade-off that must be considered when selecting a molecular tool for effective AMR monitoring. This study has provided valuable insights into contributions from septic tanks in disseminating AMR and integrases (*intI*1, *intI*2, *intI*3) genes to the environment.

Table of content

List of Tables

List of Figures

Acknowledgement

Firstly, I would like to take this opportunity to express my immense gratitude to my supervisor, Professor Cindy Smith, for her invaluable guidance, support, encouragement and feedback throughout my PhD journey. Without her help, the completion of this thesis would have not been possible. I am indeed grateful.

I am also grateful to my other supervisors, Dr Stephanie Connelly, and Professor William Sloan, for their support and valuable contributions during my PhD. I am also grateful to our collaborators at the Asian Institute for Technology in Thailand for their contributions in ensuring the success of the project.

I would also like to thank Dr Umer Ijaz and Dr Uzma Uzma for their support and guidance in bioinformatics, which, in no doubt, enabled the completion of this thesis.

I would like to extend a special thank you to my friend and colleague- Dr Fabien Cholet- for his invaluable support and friendship throughout this PhD Journey. I am grateful.

I am also grateful and appreciative of the support and friendship of my colleague- Dr Anastasiia Kostrytsia, particularly in constructing my metagenomic library for sequencing during the COVID-19 lockdown.

To my friends and colleagues and friends- Dr Dominic Quinn, Dr Ayo Ogundero, Dr Kelly Stewart, Clare Thom and William Barr- I am also grateful for their support, laughter and companionship throughout this PhD journey, which have made this journey a memorable experience.

I would also like to thank the technicians within the Water and Environment group for their invaluable support in the labs.

I am immensely grateful to my parents, Cyril and Beatrice Okonkwo, siblings, sister-in-law and niece (Jessica), for your constant love and support. Their unwavering belief in me and constant encouragement to strive for the best, even at my lowest moments throughout the PhD journey is sincerely appreciated.

I would like to extend a final thank you to all of my extended family and friends whose love and support also made this journey memorable.

Thank you all once again for making this journey a wonderful one. I will forever cherish it

Declaration of originality

I, Valentine Okonkwo, certify that the thesis presented here for examination for a PhD degree from the University of Glasgow is solely my own work, other than where I have clearly indicated the contributions or collected data of others. The copyright of this thesis rests with the author. No quotation from it is permitted without full acknowledgement.

I declare that the thesis does not include work forming part of a thesis presented successfully for another degree. I also declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

All references used in this thesis have been appropriately cited in both the text and the bibliography of references, aligning with the University of Glasgow's guidance on plagiarism. Sources of data for all the tables and figures that are not my own are provided. No third-party professional services of any kind have been employed to produce this work.

I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Name: Valentine Okonkwo

Registration number: XXXXXXXX

Abbreviations

attC	Second integron recombination site
attI	Site-specific recombination site
AMR	Antimicrobial resistance
ARBs	Antimicrobial resistance bacteria
ARGs	Antibiotic resistance genes
CARD	Comprehensive antibiotic resistance database
cDNA	Complementary DNA
CL1-integron	Class 1 integron
CL2-integron	Class 2 integron
CL3-integron	Class 3 integron
CST	Conventional septic tanks
DDDs	Defined Daily Doses
Ds DNA	Double-stranded DNA
GCs	Gene cassettes
HGT	Horizontal gene transfer
HT-QPCR	High-throughput QPCR
IntI	Integrase protein
*IntI*1	Class 1 integron-integrase
*IntI*2	Class 2 integron-integrase
*IntI*3	Class 3 integron-integrase
MGB	Minor grove binding
MGE	Mobile genetic element
mRNA	Messenger RNA
MRIs	Mobile resistant integrons
NGS	Next-generation sequencing
Q-PCR	Real-time Quantitative PCR
Pc	integron Promoter
16S rRNA	16S Ribosomal ribonucleic acid
SD	Standard deviation
SST	Solar septic tank

TB	Tuberculosis
VGT	Vertical gene transfer
WWT	Wastewater treatment

Chapter 1

General Introduction and Literature Review

1.1 Antibiotic discovery and global significance

The discovery of antibiotics is still widely recognised as among the most significant discoveries that underpin advances in modern medicine (Hutchings et al. 2019; Cook and Wright 2023), giving rise to improved overall health, well-being and an increased life-expectancy in both humans and animals (Adedeji 2016). Since the first natural antibiotic (Penicillin) discovery by Sir Alexander Fleming in 1928 (Bennett and Chung 2001), a suite of antibiotics have been developed and characterised into various classes based on their mode of action and disease target (Table 1.1).

Recently, the World Health Organisation (WHO) has classified current antibiotics into three distinct groups: Access, Watch and Reserve (designated the acronym: AWaRe) (WHO 2021a). Antibiotics belonging to the Access group (e.g., penicillin, sulphonamides) are recommended as the first or second choice for the treatment of bacterial infection due to their antimicrobial activity against a range of commonly encountered susceptible pathogens (Jackson et al. 2019; WHO 2021a). The Watch list group of antibiotics (e,g., tetracyclines, rifamycins) are antibiotics that have a higher risk of resistance emerging from their use and are recommended for use as first or second-line treatment for specific bacterial infections (WHO 2017; Jackson et al. 2019). Antibiotics within the Reserve group are deemed as antibiotics of last resort when all other antibiotics fail. This group of antibiotics (e.g., colistin) are recommended for use only in suspected or confirmed cases of infections caused by multi-drug resistance bacterial pathogens(WHO 2021a).

Table 1.1: Classes of antibiotics and their mode of action adapted from (Coates et al. 2011; Oliphant and Eroschenko 2015; Hutchings et al. 2019)

Antibiotic class	Example of antibiotics	Mode of action
Aminoglycosides	Amikacin, Spectinomycin, Tobramycin, Streptomycin, Gentamicin, Neomycin, Netilmicin, Paromomycin, Dibekacin, Kanamycin, Isepamicin, Sisomicin.	Protein synthesis inhibitors
Tetracyclines	Demeclocycline, Tigecycline, Oxytetracycline, Minocycline, Tetracycline, Methacycline, Chlortetracycline, Doxycycline.	
Amphenicols	Florfenicol, Chloramphenicol, Thiamphenicol	
Macrolides	Spiramycin, Midecamycin, Erythromycin, Roxithromycin, Azithromycin, Clarithromycin	
Tuberactinomycins	viomycin	
Lincosamides	Lincomycin, Clindamycin	
Pleuromutilins	Tiamulin, Ratapamulin	
Oxazolidinones	Linezolid	
Streptogramins	Quinupristin, Pristinamycin, Dalfopristin	
Mupirocin	Mupirocin	
Fusidic acid	Fusidic acid	
Beta-lactams	<u>Penicillins</u> Penicillin G, Cloxacillin, Penicillin V, Azlocillin, Ampicillin, Piperacillin and Carbenicillin, Methicillin, dicloxacillin, Benzathine penicillin G, Amoxicillin, Oxacillin, Ticarcillin, Nafcillin, Temocillin, Mezlocillin.	Inhibits/ disrupts cell wall synthesis
	<u>Cephalosporins</u> 1st generation- Cephaloridine, Cephapirin, Cefazolin, Cephalothin, Cephradine. 2nd generation- Cefaclor, Cefuroxime, Cefmetazole, Cephalexin, Loracarbef, Cefamandole, Cefprozil, Cefoxitin. 3rd generation- Cefoperazone, Cefdinir, Cefpodoxime, Cefotaxime, Ceftibuten, Ceftazidime, Cefixime, Ceftizoxime 4th generation- Cefepime, Cefpirome. <u>Carbapenems</u> Doripenem, Imipenem, Ertapenem, Meropenem	
Beta-lactamase inhibitors	Sulbactam, Tazobactam, Clavulanic acid	
Monobactams	Aztreonam	
Lipopeptides	Daptomycin	
Polymyxins	Colistin	
Glycopeptides	Vancomycin, Teicoplanin, Telavancin	
Cycloserines	Seromycin	
Phosphonates	Fosfomycin	
Polypeptides	Gramicidin A	
Bacitracin	Bacitracin A	
Enniatins[a]	Fusafungine	
Pyridinamides	Isoniazidz	
Ethambutol	Ethambutol	
Thioamides	Ethionamide	
Fluoroquinolones	Sparfloxacin, Norfloxacin, Levofloxacin, Nalixidic acid, Ciprofloxacin, Oxolinic, Trovafloxacin, Grepafloxacin, Moxifloxacin, Temafloxacin, Enoxacin, Clinafloxacin, Fleroxacin, Lomefloxacin, Gatifloxacin, Sitafloxacin.	DNA synthesis inhibitors
Others	Novobiocin	
Nitrofurans	Furazolidone, Nitrofurantoin	Anaerobic DNA synthesis inhibitors
Nitroimidazole	Ornidazole, Metronidazole	
Rifamycins	Rifabutin, Rifampicin, Rifaximin, Rifapentine	RNA synthesis inhibitors
Sulphonamides	Sulfadiazine, Sulfamerazine, Sulfamethazine, Sulphanilamide, Sulfapyridine, Sulfamethoxazole, *Para*-aminobenzoic acid	Inhibitors of Folic acid synthesis

Sulphone	Dapsone	
Salicylates	4-Aminosalicylic acid	
Diaminopyrimidines	Trimethoprim	
Anthracyclines	Doxorubicin, Epirubin, Idarubicin	DNA replication (intercalators)
Others	Mithramycin, Tetracenomycin, Actinomycin D	

[a] fusafungine recommended for market withdrawal in February 2016 by the European Medicines Agency (Hutchings et al. 2019).

1.2 Global consumption of antibiotics and other antimicrobial agents with emphasis on the consumption in the global south

Globally, the consumption of antibiotics and other antimicrobials (i.e., heavy metals, biocides, fungicides, antiviral agents and parasiticides (Coque et al. 2023) has substantially increased over the years (Jackson et al. 2019). In a prior study, Klein et al., (2018) used antibiotic sales data from 76 countries between 2000 and 2015 to estimate that global antibiotic consumption increased by 65% from 21.1 billion defined daily doses (DDDs) to 34.8 billion DDDs during that period. Based on the same data, Klein et al., also estimated that the global antibiotic consumption in 2015 was 42.3 billion DDDs and projected that if current policy/strategies remain unchanged, global antibiotic consumption would increase to 128 billion DDDs by 2030, an increase of 202% from 2015.

Additionally, the recent COVID-19 global pandemic caused a further increase in the already high global antibiotic consumption, especially within clinical settings (Satria et al. 2022; Sulayyim et al. 2022). Although a viral infection, prior published guidelines recommended administering antibiotics to COVID-19 patients, including those without confirmed bacterial co-infection (Langbehn et al. 2021; Satria et al. 2022; Sulayyim et al. 2022).

Indeed, the type of antibiotic consumed varies on a spatial-temporal scale. However, four antibiotic classes- Penicillin's (especially amoxicillin), Cephalosporins, Fluoroquinolones and Macrolides (Table 1.1) are commonly consumed globally (Klein et al. 2018). This trend is also true in the Global South, particularly in Thailand, where tetracycline is frequently prescribed in addition to the aforementioned antibiotics (Siltrakool et al. 2021).

1.3 Global challenges and subsequent impacts of increased global consumption

The increase in global consumption of antimicrobials, including antibiotics, is primarily attributed to mismanagement and extensive use, particularly in clinical and agri-and aqua-cultural settings (Holmes et al. 2016; Von Wintersdorff et al. 2016; Bassetti et al. 2022). This mismanagement and overuse have primarily been facilitated by 1) increased availability and easy accessibility, 2) relatively low cost 3) poor regulations on usage, especially in low-middle-income countries (Nepal and Bhatta 2018), and 4) overpopulation, amongst other factors.

Consequently, this gave rise to the rapid spread of antimicrobial resistance (AMR) across microbial taxa in both clinical settings and the environment. AMR is defined, by WHO, as the ability of microbes to adapt and survive exposure to antimicrobials, which renders the antimicrobial ineffective for treating infections caused by these microbes. Therefore, AMR poses a significant global threat to public health (Holmes et al. 2016) and AMR is currently classified as "*one of the top ten global public health threats facing humanity*" by WHO (2021b).

Bacterial infections from AMR were directly attributed to 1.27 million global deaths in 2019, while 4.95 million global deaths were associated with bacterial AMR infection in the same year (Murray et al. 2022). The death toll from AMR infection is expected to rise to 10 million global deaths per year by 2050, surpassing deaths from cancer (O'Neill, 2014). In addition, the cumulative global economic burden is projected to reach 3.4 trillion US dollars in the next decade (Coque et al. 2023). In light of the recent increase in AMR, WHO, in addition to categorising antibiotics based on importance, identified and classified 12 antibiotic-resistant bacterial pathogens (Table 1.2) that pose the greatest health risk to humans (WHO 2017), highlighting the urgency to develop new antibiotics to combat these pathogens (Tagliabue and Rappuoli 2018).

Table 1.2: List of "priority pathogens" and antibiotics they are resistant towards. Adapted from WHO published list of priority pathogens (WHO 2017)

Priority classification	Organism/ family	Resistance
Priority 1: Critical	*Acinetobacter baumannii*	Carbapenem-resistant
	Pseudomonas aeruginosa	Carbapenem-resistant
	Enterobacteriaceae	Carbapenem-resistant, ESBL-producing
Priority 2: High	*Enterococcus faecium*	Vancomycin-resistant
	Staphylococcus aureus	Methicillin-resistant, vancomycin-intermediate and resistant
	Helicobacter pylori	Clarithromycin-resistant
	Campylobacter spp	Fluoroquinolone-resistant
	Salmonellae	Fluoroquinolone-resistant
	Neisseria gonorrhoeae	Cephalosporin-resistant, Fluoroquinolone-resistant
Priority 3: Medium	*Streptococcus pneumoniae*	Penicillin-non-susceptible
	Haemophilus influenzae	Ampicillin-resistant
	Shigella spp	Fluoroquinolone-resistant

1.4 Other contributing challenges exacerbating the global AMR crisis

In addition to the extensive use and misuse of current antimicrobials, other global challenges have further exacerbated the global AMR crisis. These include 1) a rapid decline in the discovery, development and production of new antimicrobial drugs by pharmaceutical companies due to poor economic return due to the fast rate at which resistance emerges (Shlaes and Bradford 2018; Hutchings et al. 2019; Cook and Wright 2023). 2) increased incidence of new and re-emerging infectious bacterial diseases due to increased urbanisation and climate change among others (Mukherjee 2017); and 3) limited success of policies implemented to limit the spread of AMR (Di Cesare et al. 2016). Furthermore, on a genetic level, factors such as the co-selection of AMR genes are increasingly recognised to further exacerbate the global AMR burden (Baker-Austin et al. 2006; Bürgmann et al. 2018).

1.4.1 Co-selection

Co-selection is the selection for genes conferring resistance to both antibiotics and other antimicrobials (e.g., heavy metal) following exposure to either antibiotics or heavy metal (Pal et al. 2017) and it is achieved via co-and-cross-resistance (Baker-Austin et al. 2006; Pal et al. 2017).

Co-resistance occurs when exposure to a specific antimicrobial not only causes selection for the gene conferring resistance to the specific antimicrobial but also for other resistance genes because the genes are on the same mobile genetic element (Di Cesare et al., 2016; Pal et al., 2017; Bürgmann et al., 2018; Zhang et al., 2018). For example, exposure of microbes harbouring class 1 integron (CL1-integron) to quaternary ammonium (a biocide), would not only drive selection for quaternary ammonium resistance gene(s) but also cause selection of other resistance gene carried by the CL1-integron. This is owing to their physical linkage on the same CL1-integron structure. As such, an increase of all co-selected genes will be observed alongside the quaternary ammonium resistance gene in the presence of quaternary ammonium selection pressure.

In contrast, **cross-resistance** occurs when a single resistance trait (e.g., efflux pump) confers resistance to a range of antimicrobial classes simultaneously (Baker-Austin et al., 2006; Di Cesare et al., 2016; Bürgmann et al., 2018; Zhang et al., 2018). Thus, rendering an entire antimicrobial class ineffective in treating bacteria harbouring genes that confer efflux pump resistance mechanism.

1.5 Transmission of resistance traits between microbes

Exposure to antimicrobials exerts selection pressure on microbes, causing microbes with inherited (intrinsic) resistance or acquired resistance genes to survive and thrive in the presence of these antimicrobial(s) (Holmes et al. 2016; Lerminiaux and Cameron 2019). Genes conferring AMR are generally acquired via vertical gene transfer (VGT) or horizontal gene transfer (HGT) (Cox and Wright 2013).

1.5.1 Intrinsic resistance

Intrinsic resistance occurs as a result of an inherent structure (i.e., outer cell membrane in gram-negative bacteria or efflux pumps) that enables a bacterial species or genus to naturally resist the deleterious effect of a specific type/ group of antimicrobial agents following exposure (Fernández and Hancock 2012; Arzanlou et al. 2017). For example, the antibiotic teixobactin, which is the first member of a novel class of lipid II binding antibiotics, discovered from uncultured soil bacteria in 2015, is ineffective against gram-negative bacteria (Ling et al. 2015). This is because it cannot penetrate the outer membrane of gram-negative bacteria (Ling et al. 2015; Hussein et al. 2020).

1.5.2 Vertical Gene Transfer (VGT)

VGT mechanism involves the acquisition of genes conferring resistance to a particular/ group of antimicrobial agents via spontaneous mutation, which are then passed on vertically during bacterial replication (Cox and Wright 2013). In natural environments, this form of gene transfer mechanism co-exists alongside horizontal gene transfer (HGT) mechanism thus highlighting the various mechanisms at the disposal of microbes to acquire and disseminate AMR genes (Li et al. 2019).

1.5.3 Horizontal gene transfer (HGT)

HGT is commonly associated with the rapid dissemination of AMR genes within microbial communities compared to VGT and is achieved through three main mechanisms: 1) transformation, 2) transduction, and 3) conjugation (Figure 1.1) (Soucy et al. 2015; Von Wintersdorff et al. 2016; Lerminiaux and Cameron 2019).

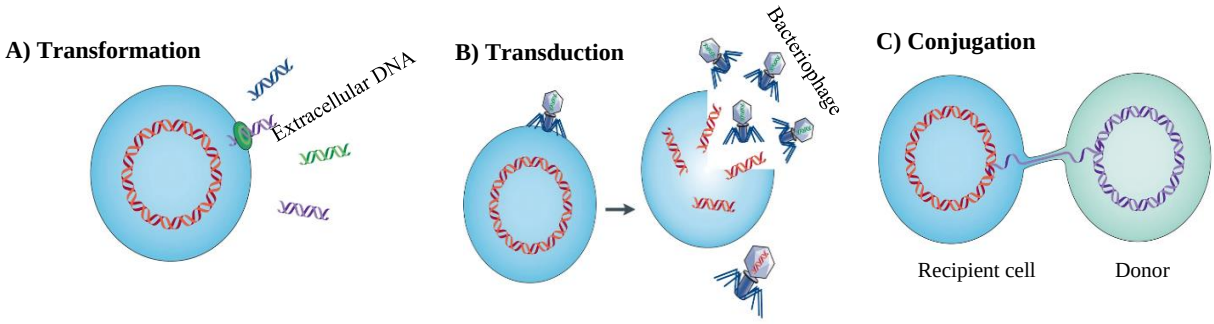

Figure 1.1: Horizontal mechanisms of gene transfer. A) Transformation, B) Transduction and C) Conjugation. Figure copied and adapted from Soucy et al., (2015).

1.5.3.1 Transformation

Transformation involves the uptake of naked DNA (extracellular DNA derived from lysed bacterial cells) from the environment by bacteria (Figure 1.1A) (Von Wintersdorff et al. 2016). For a successful transformation to occur, the recipient bacterial cell must be in a competent state (i.e., a state where the cells can uptake extracellular DNA) and extracellular DNA must be present within the environment. Additionally, integration of translocated DNA (i.e., the naked extracellular DNA) into the recipient genome or re-circularisation (in

plasmids only) must also occur once uptake of extracellular DNA occurs. This ensures the stabilisation of the translocated DNA (Von Wintersdorff et al. 2016).

1.5.3.2 Transduction

Transduction is a phage (viruses that infect bacteria) mediated mechanism of gene transfer (Figure 1.1B) and it is recognised as a contributor to the dissemination of AMR genes (Lerminiaux and Cameron 2019). The process of transduction involves the transfer of DNA, such as AMR genes, contained within the bacteriophage capsid following infection of a new bacterial host (Gillings 2017b; Lerminiaux and Cameron 2019). The transduction process is successful if the transferred DNA is recombined into the genome of the newly infected bacterial host (Lerminiaux and Cameron 2019). This process not only transfers important genes (e.g., AMR genes) beneficial to the newly infected host but also promotes phage survival and dissemination (Von Wintersdorff et al. 2016).

1.5.3.3 Conjugation

Conjugation (Figure 1.1C) is by far the most studied and efficient form of HGT among the three HGT mechanisms mentioned (Von Wintersdorff et al. 2016). During conjugation, genetic materials (e.g., AMR genes) are acquired from donor-to-recipient bacteria through physical cell-to-cell contact via the pilus or adhesin (Figure 1.1C) (Soucy et al. 2015; Lopatkin et al. 2016; Von Wintersdorff et al. 2016). Bacterial exposure to antibiotics can trigger the conjugation process (Lerminiaux and Cameron 2019) causing a rapid dissemination of ARGs between microbes (Cai et al. 2022). Conjugation typically occurs via plasmid transfer or through integrated conjugation elements located on chromosomes (Lopatkin et al. 2016).

1.6 Resistance mechanism of AMR genes

Inherent or acquired AMR genes can confer resistance to antimicrobials, in particular antibiotics, through five major mechanisms namely: 1) drug inactivation, 2) drug target alteration, 3) drug target protection, 4) drug target replacement, and 5) drug efflux.

1.) **Drug Inactivation**

Antibiotics can be inactivated by bacteria either through the production of chemicals that alter the antibiotics or by destroying the antibiotics with degradative enzymes

such as β-lactamases (Figure 1.2) (Gupta and Birdi 2017; Uddin et al. 2021; Darby et al. 2022). Inactivation via chemical alteration occurs by the attachment of bacterial enzymes to various chemical groups on the antibiotic; thus, preventing binding of the antibiotic to the target on the bacterial cell (Uddin et al. 2021).

2.) Drug Target Alteration

Drug target alteration is among the most common in bacterial pathogens and it usually involves permanent alterations to drug target sites via mutation (Gupta and Birdi 2017; Wilson et al. 2020; Darby et al. 2022). This, in turn, decreases the binding of antibiotics to the target site reducing antibiotic efficacy (Figure 1.2) (Darby et al. 2022).

3.) Drug Target Replacement

Similar to target alteration, target replacement occurs as a result of bacteria replacing molecules on their cells that are the target of antibiotics; thus, preventing the binding of the antibiotics to the bacteria cell (Figure 1.2) (Gupta and Birdi 2017).

4.) Drug Target Protection

Target protection is when a resistance protein physically protects the antibiotic target (i.e., bacterial antibiotic target site), thus protecting the target from the effect of antibiotics (Figure 1.2) (Wilson et al. 2020). Unlike target alteration, this form of resistance mechanism does not cause permanent change to the antibiotic target. Instead, it may persist or be reattached to the antibiotic target in the case of repeated exposure to antibiotics (Wilson et al. 2020).

5.) Drug Efflux

Antimicrobials, such as antibiotics, can be exported out of bacterial cells via efflux pumps, which are basically transmembrane proteins (Figure 1.2) (Kapoor et al. 2017; Darby et al. 2022). Immediately after entering the bacterial cells, the antimicrobial is quickly pumped out of the cell via the efflux pump at the same time; thus, protecting the cell from the deleterious effect of the antimicrobial due to intracellular accumulation (Kapoor et al. 2017). Efflux pumps can be found on the outer cell membrane and in the cytoplasmic membrane (Kapoor et al. 2017; Darby et al. 2022). Except for the polymyxin antibiotic class (Table 1.1), all antibiotics can be exported out of bacteria cells via the efflux pump; thus indicating how easily bacteria cells can rapidly develop multi-drug resistance (Kapoor et al. 2017)

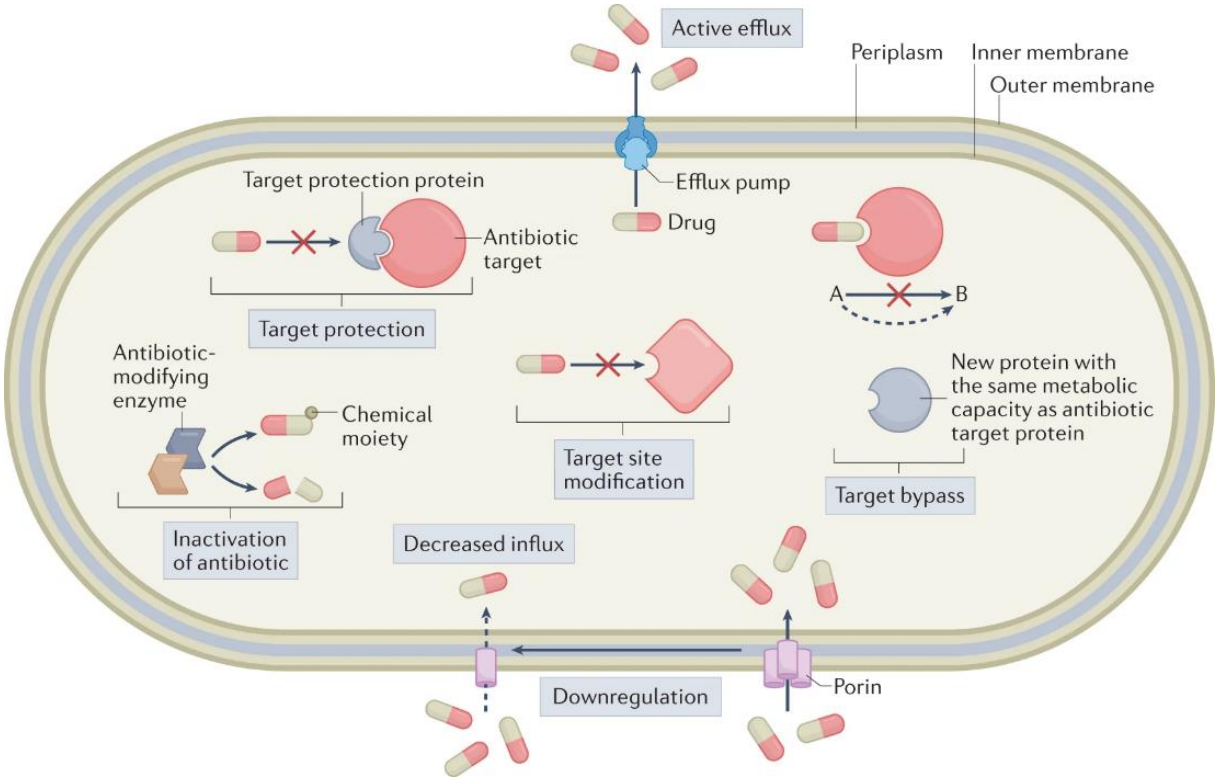

Figure 1.2: Schematic illustration of bacterial resistance mechanisms. Figure copied from Darby et al.,(2022).

1.7 Vectors for the transmission of AMR between or within microbial taxa

Mobile genetic elements (MGEs), such as plasmids, transposons and integrons act as vectors and play a crucial role in the acquisition and rapid dissemination of AMR genes between or within microbes.

1.7.1 Plasmids

Plasmids (circular/linear double-stranded DNA) are integral components of the bacterial genome that are capable of self-replication independent of chromosomes (Carattoli 2011; Lerminiaux and Cameron 2019). Plasmids mostly carry accessory genes (i.e., AMR genes) that are beneficial to the bacteria host but do not contain core genes crucial to the growth and replication of bacterial cells (Bennett 2008). Plasmids harbouring AMR genes are predominately conjugative plasmids, which basically means that these plasmids encode

functions that facilitate cell-cell transfer of DNA either via the pilus or adhesin (Bennett 2008). As a result, these conjugative plasmids harbouring AMR genes can easily disseminate resistance genes via HGT across a wide range of microbial taxa (Bennett 2008). In addition, resistance plasmids can persist in any environment including environments without antibiotic pressure, which in turn, increases the risk of dissemination (Lerminiaux and Cameron 2019), thus, highlighting the complexities in tackling the global AMR crisis.

1.7.2 Transposons

At its simplest, transposons are referred to as *"jumping gene"*(Babakhani and Oloomi 2018). These distinct DNA segments are capable of relocating themselves, along with any AMR genes carried, from one location on a DNA molecule to another on the same molecule or a different DNA molecule (Babakhani and Oloomi 2018; Partridge et al. 2018). This process of transposition is achieved via two distinct mechanisms referred to as "*cut and paste* " or "*copy and paste*", which allows transposons to move to a new location without requiring significant homology between their sequence and the new location's DNA (Hickman et al. 2010).

Furthermore, transposons can move from plasmid to plasmid, chromosome to chromosome, plasmid to chromosome or vice versa (Babakhani and Oloomi 2018). Among the families of transposons characterised thus far, the Tn3 and Tn7-like families are associated with AMR resistance (Partridge et al. 2018). Within the Tn3 family transposons, the Tn21 sub-family and close relatives are known to sometimes entrain mercury (*mer*) resistance genes and may sometimes carry CL1-integrons meaning they are considered potentially important in disseminating AMR genes (Partridge et al. 2018). Similarly, within the Tn*7*-like family, the Tn7 transposons are known to carry class 2 integrons (CL2-integrons), while the Tn*402* (and members of the Tn*5053* family), also within the Tn7-like family, are known to carry CL1-integrons or *mer* gene and are flanked by a 25 bp inverted repeats (Partridge et al. 2018).

1.7.3 Integrons

Integrons are genetic platforms that facilitate the capture, integration, assembly and accurate expression of exogenous genes embedded within compact structures known as gene cassettes, located at the variable region of the platform (Hall 2012; Gillings 2014).

An estimated 10% (Zheng et al. 2020) to 15% (Gillings 2014) of sequenced bacteria genomes have been reported to contain integrons, and many of these integrons are located on chromosomes (Gillings 2017a; Ghaly et al. 2021). These chromosomally located integrons tend to entrain gene cassettes that encode currently unknown functions (Ghaly et al. 2021).

Nonetheless, analysis of integron sequences based on their *intI* (integrase) homology, led to the characterisation of over 100 types of integrons (Hall 2012). Irrespective of the type, all integrons are known to consist of three core features at their 5'conserved region (Gillings 2014), which are essential for the integration of gene cassettes (GCs) onto the integron platform and expression of integrated GCs. These features include- 1) an *IntI* gene, which encodes a site-specific integrase protein; 2) a site-specific recombination site, where genomic sequences are inserted; and 3) a promoter that monitors incorporated gene cassette transcription and ensures their accurate expression.

Similar to integrons, integrases (*intI*) also have different types, such as integron-integrase, λ integrase, Cre, Flp, and XerC-XerD integrase (Messier and Roy 2001). Integron-integrase, which belongs to the tyrosine recombinases family, differs from all other characterised integrases in that it possesses an additional unique16-amino-acid conserved motif, which is crucial for its activity (Messier and Roy 2001; Gillings 2014).

In addition to being located on chromosomes, integrons can also be found associated with other mobile elements such as transposons and plasmids (herein referred to as mobile integrons). These mobile integrons typically entrain AMR genes but in fewer numbers (up to six gene cassettes can be carried on mobile integron) especially in anthropogenic environments (Gillings 2014). Among the different mobile integron types characterised, only five types (class 1 - 5) have been associated with disseminating AMR (Mazel 2006). Among these, classes 1, 2 and 3 are the most commonly studied and associated with multiple AMR phenotypes (Mazel 2006; Quintela-Baluja et al. 2021).

1.7.3.1 Class 1 integrons

Among the mobile integrons associated with the spread of AMR across diverse bacteria taxa, the CL1-integron is by far the most abundant and most commonly surveyed in both clinical settings and polluted environments, such as wastewater treatment (WWT). CL1-integrons are essentially non-mobile but their association with the *Tn402* transposons or plasmids ensures their potential mobility (Gillings 2014). *Tn402* transposons target the resolution (res) site of plasmids; thus, enabling the Tn402-CL1-integron hybrid to move into a diverse

plasmid type, which in turn, facilitates dissemination into a wider bacteria taxa via HGT (Gillings et al. 2015). As a result, these mobile CL1-integrons are found present on various other mobile elements that can be transferred freely between pathogenic and non-pathogenic bacteria commonly associated with both humans and animals (Gillings et al. 2015).

A typical mobile CL1-integron found in these clinical and anthropogenic polluted environmental settings usually consists of three segments: 1) a 5'conserved region, 2) a variable region, and 3) a 3'conserved region (Figure 1.1). The 5'conserved region contains the integron-integrase (*intI*1) gene (complete length: 337 amino acids) (Hansson et al. 2002), a recombination site (*attI*1) and a promoter (Pc) (Figure 1.3) (Gillings et al. 2015). The *intI*1 gene encodes the integrase enzyme that catalyses site-specific recombination events allowing the integration of captured GCs onto the CL1-integron platform. Furthermore, within the *intI*1 coding sequence sits the Pc promoter (downstream of the *intI*1 gene), which is responsible for the transcription and expression of the integrated cassette (Cambray et al. 2010; Gillings et al. 2015). Thirteen Pc promoter variants, with different transcription levels, have been characterised (Domingues et al. 2012). These variations in the Pc promoter introduce variability to the *intI*1 gene sequence, resulting in 10 different *intI*1 gene variants characterised (Domingues et al. 2012). Despite this, the *intI*1 gene isolated within the clinical context and human-impacted systems such as WWT, tend to exhibit high sequence similarity (≥98% protein identity) to each other (Roy et al. 2021).

The variable region of these highly conserved CL1-integrons predominantly entrain cassettes (Figure 1.3), while the 3'conserved region comprises a truncated *qacE*Δ*1* gene (encode quaternary ammonium compound), a *sul*1 gene (encode sulphonamides resistance), and an open reading frame 5 (orf5) of currently unknown function (Domingues et al. 2012).

In contrast to the mobile CL1-integrons, class 1 integrons can also be found on chromosomes (Gillings 2014). These chromosomal CL1-integrons are commonly found in non-pathogenic *Betaproteobacteria,* such as those belonging to the *Hydrogenophaga, Aquabacterium, Acidovorax, Imtechium, Azoarcus, and Thauera* genera (Gillings 2014). In addition, the *intI*1 gene of these chromosomal integrons tends to exhibit greater sequence diversity (<98% protein identity) when compared to the highly conserved *intI*1 gene of mobile CL1-integron (Gillings 2014). Furthermore, the GCs entrained on these chromosomal CL1-integrons encode currently unknown functions rather than AMR (Gillings 2014). Of note, these lesser conserved *intI*1 genes are also found in various environments including anthropogenic polluted environments (Gillings et al. 2008b).

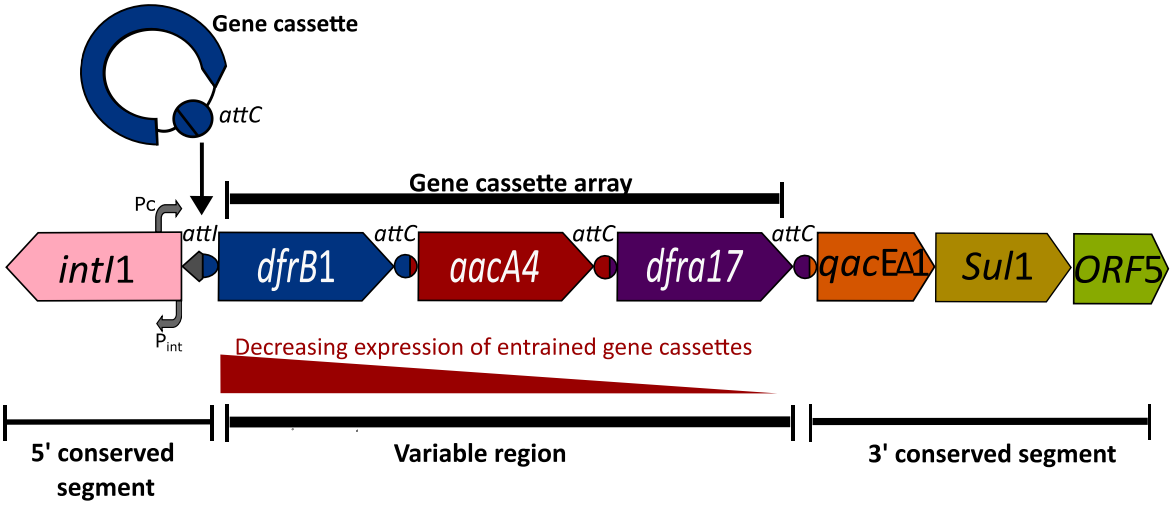

Figure 1.3: Basic structure of a typical CL1-integron consisting of an integron-integrase gene (*intI*1), two promoters and a site-specific recombination site (*attI*) at the 5' conserved segment; a variable region where gene cassettes are integrated and expressed; and a 3' conserved segment that includes a truncated *qacEΔ*1 gene, *sul*1 gene and an open reading frame (ORF5) of currently unknown function. Figure adapted from Mazel,(2006) and Ma et al., (2017).

1.7.3.2 Class 2 and 3 mobile integrons

CL2-integrons are commonly associated with the Tn7 transposon family and its derivatives (including the Tn4132 and Tn1825) (Ramírez et al. 2010; Stalder et al. 2012; Sultan et al. 2018). The CL2-integron-integrase (*intI*2) gene typically contains a stop codon at position 179 (Hansson et al. 2002), which leads to an internal disruption rendering the integrase protein non-functional and truncated (Stalder et al., 2012). As a result, the entrained GCs remain stable and consist of *dfrA*1 (confers trimethoprim resistance), *sat*2 (confers streptothricin resistance), *aadA*1 (confers streptomycin and spectinomycin resistance), and *orfX* (confers a currently unknown function) (Stalder et al., 2012). Interestingly, *intI*2 share a 46% protein identity to the *intI*1 gene and is 325 amino acids in length (Hansson et al. 2002; Hall 2012).

Class 3 mobile integrons (CL3-integrons) are associated with the *Tn402* transposons (Deng et al., 2015) and are detected increasingly frequently compared to CL2-integrons (Quintela-Baluja et al. 2021). Additionally, the CL3-integron-integrase (*intI*3) gene is more closely related to the *intI*1 and shares a 60% protein identity with *intI*1 (Hall 2012; Roy et al. 2021).

1.7.3.3 Gene cassettes

Gene cassettes (GCs) are compact mobile vectors (vary from 0.5 to 1kb in length) that typically contain a single gene, a recombination site (*attC*) and are promoterless (Labbate et al. 2009; Stalder et al. 2012; Partridge et al. 2018). The *attC* recombination site (formerly referred to as 59-base elements) associated with mobile integrons differs in sequence and length (57 to 141 bp) considerably (Messier and Roy 2001; Partridge et al. 2009; Gillings 2014), but share a conserved region and short imperfect inverted repeats at their flanking ends (Partridge et al. 2018). A GC is non-replicative and is typically found in its free circular form (Partridge et al. 2018). Once captured by an integron, the GC is transformed from its original circular form to a linear form and then integrated into the integron platform (Figure 1.3). This integration process onto the integron platform is by site-specific recombination catalysed by the *intI* gene (encodes the integrase enzyme), between the *attI* recombination site and *attC* site (GC recombination site) (Partridge et al. 2018). Moreover, each GC is inserted independently and the insertions of multiple GCs encoding AMR result in the expression of multidrug resistance in the microorganism harbouring the integron (Labbate et al. 2009; Partridge et al. 2018). Conversely, integrated cassettes can be excised from the platform via site-specific recombination between two *attC* sites, catalysed by the *intI* enzyme. (Stalder et al. 2012)

Integrated GCs closest to the integron Pc promoter will have the most prominent expression (Figure 1.3) (Stalder et al. 2012). Remarkably, over 130 different cassettes conferring resistance to most current antibiotics classes (including but not limited to Beta-lactams, lincomycin, erythromycin and all aminoglycosides) have been described (Cambray et al., 2010). Moreover, all GCs found entrained on CL2-and-CL3-integrons have also been found on mobile CL1-integrons.

1.8 Wastewater treatment plants and environmental spread of AMR genes

The deployment and widespread use of engineered systems such as WWT in sewage treatment have effectively reduced the global disease burden. WWT receive waste from various source (Figure 1.4), such as domestic or hospital waste(Gibson et al. 2023), and a significant amount of antibiotics from human and animal waste (30 to 90% of antibiotics are excreted unchanged via urine and faeces (Sarmah et al., 2006)) (Bürgmann et al. 2018). As such, this represents a unique interface where both human pathogenic and non-pathogenic microorganisms from humans, animals and the environment interact and exchange genes via

HGT (Che et al. 2022). The multiple antibiotics and antimicrobial agents such as metals and biocides entering the WWT from diverse sources, albeit sub-inhibitory concentration in nature (10-1000 times < than concentrations used in human and animal treatments (Coque et al. 2023)), exert selection pressure on the high-density, diverse and complex microbial communities within the system. This, in turn, drives the acquisition of AMR genes from the surrounding pool of AMR genes as well as enrichment of ARB (Liguori et al. 2022).

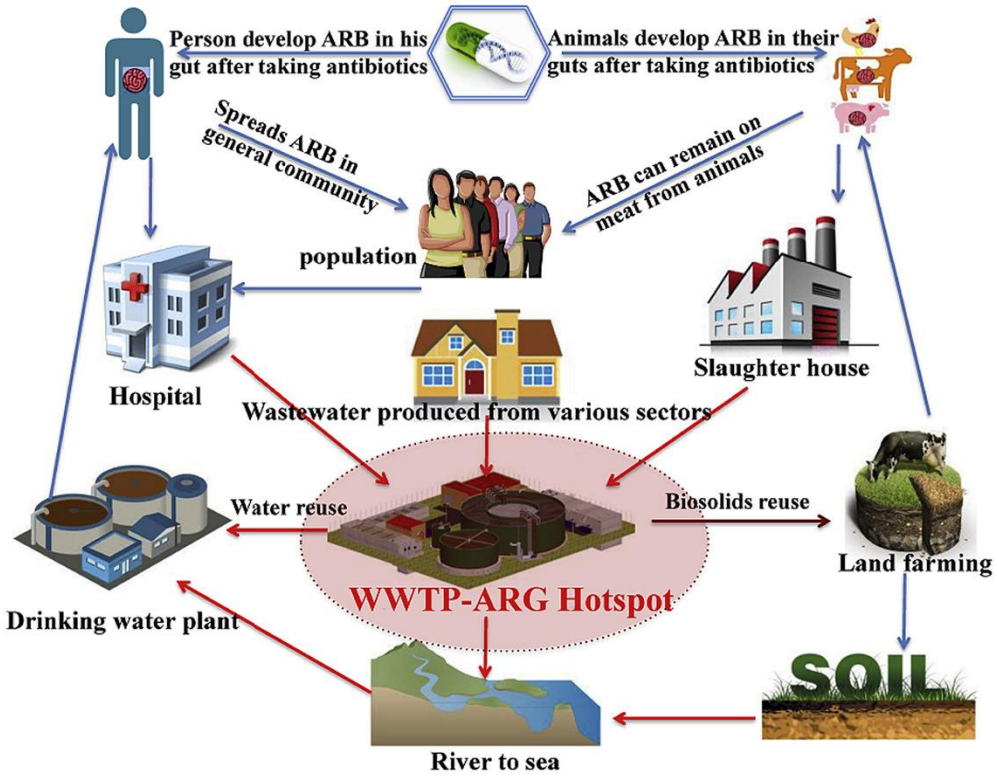

Figure 1.4: Schematic illustration highlighting the various source of waste, including antibiotics, received by wastewater treatment plant and environments receiving discharged WWT by-products (sludge and effluent). Figure copied from Guo et al.,(2017).

Typically, centralised WWT employs a combination of physico-chemical and biological processes, coupled with advanced oxidative processes (i.e., ozonation or UV disinfection) to degrade organic and chemical pollutants (i.e., antibiotics) and reduce pathogen, antimicrobial resistant bacteria (ARB) and AMR gene load from treated waste prior to its immediate discharge to the environment (Garrido-Cardenas et al. 2017; Bürgmann et al. 2018). The combination of these processes makes WWT system effective at significantly reducing AMR gene load, pathogens and other co-selecting agents from treated waste (effluent) (Coque et al. 2023). However, despite their effectiveness, WWT is unable to completely remove ARB, AMR genes (including clinically relevant AMR genes) and MGEs (particularly CL1-integrons) from the treated sludge and effluent. As such, recognised as

sources of ARB, AMR genes and MGEs to the wider environment (Karkman et al. 2016; Liguori et al. 2022)

1.8.1 Decentralised WWT with emphasis on septic tanks

Despite the widespread use of centralised WWT across the globe, a significant proportion of the global population (approximately 2.7 billion) is served by decentralised WWT including conventional septic tanks (Harada et al. 2016). Conventional septic tanks (CSTs), by design, are low-cost sewage treatment systems (Abbassi et al. 2018; Sharma et al. 2022). In addition, CSTs are easy to install and require no energy consumption for their operation, making them one of the most widely used household-scale decentralised WWT in off-grid areas with no access to centralised WWT (Connelly et al. 2019; Sharma et al. 2022).

Typically, a CST consists of two chambers that facilitate the separation of raw sewage components. In one chamber, oil and fats float to the surface while solids (sludge) from raw sewage settle to the bottom, creating an anaerobic environment that promotes anaerobic digestion of retained solids. Meanwhile, the second chamber facilitates the discharge of liquid. However, since the only form of wastewater treatment in the CST reactor is microbial degradation of retained solids under anaerobic conditions (Muralikrishna and Manickam 2017), which usually enters the reactor faster than they are degraded, CST is therefore typified by inadequate treatment performance (Ramage et al. 2019). Over time, the retained sludge accumulates, causing a decrease in the tank's volume coupled with a shorter effluent retention time (Connelly et al. 2019) which further worsens the quality of effluent discharged. Discharged effluent, in many cases, is released directly into the surrounding environment (Connelly et al. 2019). This poor-quality effluent, which contains pathogens from sewage waste, ARB and AMR genes, contaminates groundwater when discharged to the surface environment and sinks into the ground (Bijekar et al. 2022). Moreover, in many countries, surface water, containing the discharged effluent, is increasingly used in cooking, drinking, farming and bathing owing to an increasing global shortage of fresh water (Edokpayi et al. 2017; Garrido-Cardenas et al. 2017), or as surface water supplies for drinking water treatment. Thus, posing a significant health risk to humans and animals. Furthermore, the de-sludged septic tank solids are often subjected to no additional form of treatment and are released directly into the environment, often applied as manure to crops. This is particularly prevalent in the global south region, such as Thailand, where a significant proportion (>75%) of generated wastewater is inadequately treated and discharged directly to the environment, owing to lacking or ineffective or lacking WWT. (Wongburi and Park

2018). Furthermore, 80-90% of retained faecal during WWT undergoes no additional treatment to reduce pathogen and microbial load and are discharged directly into the environment (Koottatep et al. 2021). In addition to ineffective or lacking WWT, poor regulations on antibiotic usage have fueled over-the-count-purchase and self-medication with antibiotics (Siltrakool et al. 2021) leading to high antibiotic consumption in these regions, which further exacerbates the global AMR burden.

In an effort to improve the treatment quality of wastewater from septic tanks (Connelly et al. 2019), the solar septic tank (SST) (Figure 1.5) was developed in Thailand and it's currently implemented in some areas of Thailand (Polprasert et al. 2018; Connelly et al. 2019). The SST is an emerging technology that was modified from CST. This technology differs from the CST in that it incorporates a central disinfection chamber which is heated through an internal copper coil connected to a passive solar heat collection system installed on the roof of the toilet block served by the SST (Figure 1.5) (Polprasert et al. 2018; Connelly et al. 2019)

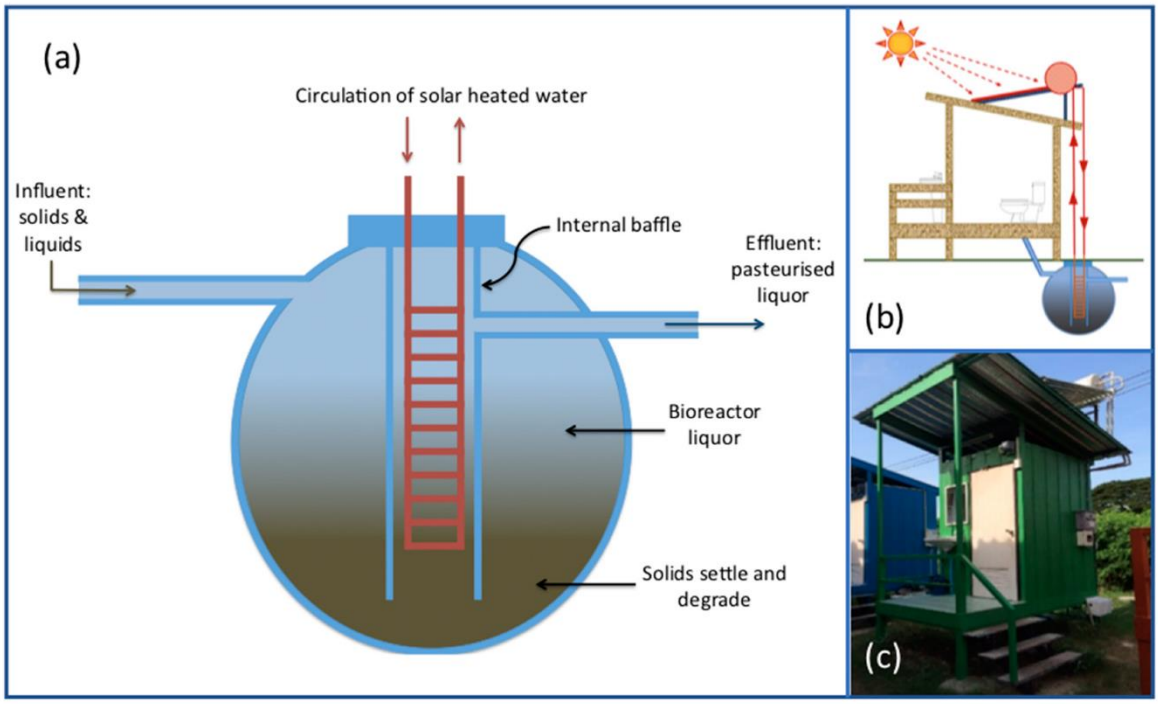

Figure 1.5: A schematic and actual image of the solar septic tank. A) schematic drawing, B) Schematic illustration showing the buried tank at field site, C) actual photograph of the solar tank unit implemented in one of the field sites. Figure copied from Connelly et al., (2019).

By design, the central chamber generates heat (50 - 60°C), which promotes partial pasteurisation of the effluent as it flows through the chamber prior to discharge. This, as a result, improves effluent water quality by reducing microbial biomass including potential

host pathogens (Polprasert et al. 2018; Connelly et al. 2019). Moreover, the heat generated by the central chamber is passively transferred to the rest of the tank, leading to an increase in the in-tank temperature, thus promoting enhanced microbial degradation of both retained solids (sludge) and soluble compounds (Polprasert et al. 2018; Connelly et al. 2019). In a preliminary study investigating CST and SST tanks over four months, Connelly et al., (2019) reported higher log removal of pathogens (total coliforms and *E.coli*) from the effluent of the SST unit as compared to the CST unit. However, the fate of AMR genes or removal efficiency of AMR genes from the SST unit is unknown.

1.9 Environmental AMR monitoring and current tools utilised in monitoring

With the plethora of AMR gene subtype and their notable abundance within WWT (centralised and decentralised WWT alike) and the discharged sludge and effluent, effective broad-spectrum monitoring of AMR genes from sources such as WWT to the wider environment is crucial but remains challenging (Bürgmann et al. 2018). Effective AMR monitoring allows for the comprehensive assessment of WWT removal efficiency, establishment of a baseline level of environmental resistance and identification of hotspot environments that pose the highest risk of AMR dissemination to humans and animals (Bürgmann et al. 2018; Liguori et al. 2022; Davis et al. 2023).

This, in turn, will facilitate an in-depth understanding of the role of WWT (including decentralised WWT) in disseminating AMR to the environment and provide informed knowledge upon which environmental engineers or policymakers can act on and implement strategies/ policies for reducing the global contributions of AMR from WWT.

Current research has employed a suite of molecular tools including real-time Q-PCR, the use of proxy genes (e.g., *intI*1 gene), high-throughput QPCR (HT-QPCR) and shotgun metagenomics to monitor AMR in polluted environments including WWT (Liu et al. 2019). However, each of these methods presents its unique challenges.

1.9.1 Real-time Quantitative PCR (Q-PCR)

Real-time Quantitative PCR (Q-PCR) is a targeted molecular technique, and it is by far the most commonly used molecular tool in AMR monitoring from WWT. Q-PCR is known to be a highly reproducible and sensitive technique for the quantification of genes (Smith and Osborn 2009; Liguori et al. 2022) to evaluate and monitor the anthropogenic influence of WWT discharge to the environment and assess ARGs removal from WWT (Liguori et al. 2022).

For example, Chen and Zhang, (2013b) showed that rural domestic sewage effluent contributed to a higher abundance of tetracycline ARGs (*tetM, tetO, tetQ and tetW*), sulphonamide ARGs (*sul*1, *sul*2) and CL1-integron as compared to municipal WWT effluent. Similarly, Chen and Zhang (2013a) found that the removal efficiency of six ARGs (*tetM, tetO, tetQ, tetW, sul1, sul2*) and CL1-integron (*intI*1), targeted by Q-PCR, did not show any statistical difference (p-value> 0.05) between three municipal WWT that utilised different advanced treatment process (biological aerated filter, constructed wetland, and ultraviolet disinfection). Although higher removal was observed for constructed wetlands among the three advanced treatment processes.

Quintela-Baluja et al.,(2021) found hospital effluents contained higher (10 times higher) anthropogenic impacted CL1-integron (CL1-integron carrying AMR gene) per bacterial cell compared to other WWT compartments in a WWT network. Their finding indicates that the anthropogenic impacted CL1-integrons are acquiring AMR genes, possibly due to the stronger selection pressure exerted by the hospital source.

Shamsizadeh et al.,(2021) found that the abundance (copies/ml) of targeted ARGs (*sul1, erm-B, bla*CTX-m-32, *tetW, cml-A*) and *intI*1 gene in the irrigation water source (wastewater, surface water, freshwater) decreased from wastewater to surface water to freshwater, suggesting that the irrigation water source can influence the abundance on ARG and *intI*1 in the soil or crops. In addition, crop and soil samples irrigated with wastewater showed higher abundance (copies/ g) of ARGs and *intI*1 gene than those irrigated with freshwater and surface water. However, this difference was not statistically significant (p-value > 0.05) for all targeted ARGs and *intI*1 except for the *cml-A* resistance gene, which showed a significant difference in abundance in the soil and crop samples across the three irrigation water sources.

In spite of its usefulness in AMR monitoring, this approach is associated with inherent shortcomings. These include 1) difficulties in selecting appropriate gene targets or subsets of genes (e.g., AMR genes and MGEs) to target for quantification, from the thousands of AMR gene subtypes and variants currently characterised (The Comprehensive Antibiotic Resistance Database (CARD) has over 5000 reference sequence) (Liguori et al. 2022). 2) Only a limited subset of ARGs or MGEs can be targeted simultaneously. 3) Only known genes can be targeted, therefore the discovery of novel genes with this approach is not feasible. 4) The presence of inhibitors in a sample or poor primer designs might yield no quantification leading to biased conclusions. (Miłobedzka et al. 2022).

1.9.2 Monitoring AMR pollution using proxy genes such as *intI*1

Utilising techniques such as Q-PCR to target a specific or subset of AMR genes is not ideal as the selected target gene(s) may be absent from the sample due to spatial and temporal differences in AMR composition within WWTs (Gillings et al. 2015). This is particularly true in cases where the researcher has no prior knowledge about the presence of that specific gene(s) in the sample.

Therefore, one proposed solution to circumvent challenges with multiple AMR monitoring, especially within anthropogenic polluted environments such as WWT, was the use of a proxy gene for inferring AMR pollution. Specifically, the use of the highly conserved CL1-integron-integrase (*intI*1), commonly found within clinical and polluted environments (Gillings et al. 2015; Zheng et al. 2020). This proposal was primarily because: 1) the *intI*1 is linked to genes conferring antibiotics, disinfectants and heavy metals resistance (Figure 1.1); 2) *intI*1 is commonly found in diverse taxonomic groups of both pathogenic and non-pathogenic bacteria and can move across taxa via HGT due to its physical linkage to plasmids and transposons; 3) its abundance can rapidly change in response to external pressures like antibiotic because its host cells can have rapid generation times and; 4) selection pressures imposed by recent human activities resulted in the emergence of the clinical *intI*1 variant (Gillings et al., 2015).

Despite this proposal, the suitability of the *intI*1 as an adequate proxy for inferring AMR pollution remains elusive. Some studies (Thakali et al. 2020; Zheng et al. 2020) have found a significant positive correlation between the *intI*1 abundance and the abundance of targeted ARGs and reported the gene *intI*1 as a good marker for AMR monitoring, while other studies

have only found a significant positive correlation between the *intI*1 abundance and the abundance of a subset of genes targeted (Chen and Zhang 2013b; Leng et al. 2020).

For example, Zheng et al., (2020) found a significant (p-value <0.05) positive correlation between *intI*1 abundance and the abundance of genes conferring resistance to aminoglycoside, beta-lactams, tetracyclines, and Macrolide-lincosamide-streptogramin B (MLSB) in both activated sludge and permeate WWT samples. As a result, they concluded that *intI*1 abundance could be used to infer overall ARG abundance in WWTP. Thakali et al., (2020) also recommend the use of the *intI*1 gene to monitor ARG abundance in WWT effluent after observing a significant positive correlation (p-value <0.05) of *intI*1 abundance and the abundance of ARGs (*blaTem, tetA, ermF*) quantified in the WWT samples (influent, secondary effluent and final effluent) via Q-PCR.

In contrast, Leng et al., (2020) only observed a significant positive correlation between *intI*1 abundance and the abundance of two (out of four) ARGs targeted via Q-PCR. Similarly, Chen and Zhang, (2013) only reported significant positive correlations between *intI*1 abundance and the abundance of one (of the six) ARGs targeted via Q-PCR.

1.9.3 High-throughput Q-PCR (HT-QPCR)

The development of the HT-QPCR array technology attempts to address the former limitation in AMR monitoring by targeting hundreds of AMR genes and selected MGEs, including *intI*1, simultaneously in nanoscale and on a single run (Waseem et al. 2019). This reduces the need to select the right suite of target AMR genes as faced when using conventional Q-PCR while offering the same benefits of sensitivity and specificity (Waseem et al. 2019; Liguori et al. 2022). Some pathogen-specific genes can also be included on the array to monitor pathogens in WWT (Liguori et al. 2022). Additionally, compared to other next-generation sequencing (NGS) approaches (i.e., amplicon sequencing and shotgun metagenomics), HT-QPCR has higher detection limits of genes, faster turnaround time in terms of quantification and analysis of data and there is no need for complex bioinformatic pipeline or steep-learning to analyse HT-QPCR dataset (Liu et al. 2019; Waseem et al. 2019).

In the last decade, HT-QPCR use in AMR monitoring has increased and been used to monitor AMR in environments such as WWT (An et al. 2018; Quintela-Baluja et al. 2019; Lin et al. 2021) and hospitals (Majlander et al. 2021), as well as to assess the long-term impact of sewage sludge on soil ARG abundance (Chen et al. 2016).

Utilising HT-QPCR to monitor AMR dynamics through a WWT network and receiving water body, Quintela-Baluja et al.,(2019) found that hospital wastewater source contributed significantly to a higher ARG richness (p-value <0.05) and higher ARG abundance (p-value >0.05) entering the WWT as compared to the of the receiving WWT influent. Furthermore, Quintela-Baluja et al.,(2019) observed that WWT was effective at significantly (p-value <0.05) reducing the abundance of total ARGs in the effluent from influent but not the ARG subtypes (p-value >0.05). Finally, the authors found that while the abundance of ARGs in the sludge was higher (activated sludge) than in the effluent, significantly lower ARG richness was observed in the sludge than in effluent and the composition of ARGs and bacteria in the sludge did not resemble that of the influent, whereas the composition of the effluent did. Therefore, the effluent is a major contributor of ARGs to the receiving environment than sludge.

In a recent study, Majlander et al.,(2021) utilised the HT-QPCR to monitor AMR genes in two hospitals over nine weeks and reported high AMR gene richness in both hospitals over the nine weeks but significantly higher richness for four of the weeks (at week 27-30) in both hospitals as compared to the other weeks. Furthermore, Majlander et al.,(2021) found that the hospital with the higher consumption of antibiotics had higher AMR gene richness and abundance (relative to the *16S rRNA* gene abundance). In addition, the gene profile between the two hospitals was found to be significantly different (p-value <0.05) owing to the different quantities of antibiotics used in both hospitals.

Whilst high-throughput in nature, the use of the HT-Q-PCR in AMR monitoring comes with some associated drawbacks such as high per-array run cost, a trade-off between sample number and gene targets and limited accessibility at present (Waseem et al. 2019; Liguori et al. 2022). Variants of ARG or MGE sequence are missed. More importantly, due to the high number of target assays on the array, it has been implied that conditions for some assays may not be optimal (Waseem et al. 2019).

1.9.4 Amplicon sequencing, a targeted approach

NGS has proven to be a powerful and useful monitoring tool (Davis et al. 2023), particularly for AMR. Compared to non-targeted approaches such as shotgun metagenomics, this approach is more cost-effective, provides higher detection limits and enables faster data processing and analysis (Gibson et al. 2023). For example, Gibson et al., (2023) recently surveyed ARGs from WWT influents across 16 WWTs using multiplex amplicon

sequencing, a novel technique, and found high ARG richness (total of 60 out of 114 ARGs targeted) in the WWT influents. In addition, they observed that 16 (of the 60 ARGs identified) exhibited varying sequence diversity, thus demonstrating that within a single sample, variants of a single gene can be present, which QPCR or HT-QPCR methods do not inform.

Although a promising tool, the reliance on PCR primers, however, makes this approach susceptible to the inherent bias from primers and PCR (Liu et al. 2019). Additionally, the researcher is still required to select ARGs or MGE targets from thousands of choices available.

1.9.5 Shotgun metagenomics; a non-targeted approach

Shotgun metagenomics is another NGS approach. However, unlike amplicon sequencing, this approach is non-targeted; therefore, it provides an overview of the total AMR genes, including known and unknown AMR genes, present in a given environmental samples (Zaheer et al., 2018), thus, circumventing the need for primer selection, which is an associated limitation with QPCR, HT-QPCR and amplicon-sequencing approaches. Coupled to a declining sequencing cost and advances in bioinformatic pipelines for analysis large and complex data (Garrido-Cardenas et al. 2017), the number of studies (Bengtsson-Palme et al. 2019; Petrovich et al. 2020; Manoharan et al. 2021; Rodríguez et al. 2021) employing shotgun metagenomics approach to characterise and monitor AMR genes from WWT has exponential increase in recent years. For example, utilising shotgun metagenomics, Petrovich et al., (2020) applied shotgun metagenomic to monitor AMR removal from hospital wastewater and found that ARG richness between influent, sludge and effluent was high, and a low decrease (16% reduction) in overall ARG abundance from the influent.

In another study, Bengtsson-Palme et al., (2019) studied the impact of pharmaceutical wastewater (from a macrolide antibiotic production company) on the abundance and profile of ARGs in sludge of the receiving WWT plant. The study found that, although the ARGs richness was lower in the WWT plant receiving pharmaceutical wastewater, the total abundance of ARGs was three times higher in its sludge as compared to a municipal WWT that did not receive pharmaceutical wastewater.

Furthermore, Bengtsson-Palme et al., (2019) found that, whilst the higher concentration of macrolide antibiotics was received in the WWT, compared to the municipal WWT,

enrichment of ARGs conferring macrolide resistance was not observed while significant (p-value <0.05) enrichment of MGEs including integrons was seen.

Despite decreasing NGS costs, there are some associated limitations in using shotgun metagenomics in AMR monitoring. These include: 1) high cost (in terms of regent and sequencing cost) compared to targeted approaches (i.e., conventional QPCR and HT-QPCR and amplicon sequencing methods; 2) preparation of metagenomic library is labour-intensive and time-consuming; 3) requires complex analysis and interpretation of obtained data, therefore, the expertise of a highly trained individual to accurately analyse and interpret data is required (Manaia et al. 2018); 4) Processing and analysing obtained data are time-consuming; 5) shotgun metagenomic is semi-quantitative; therefore, the detected AMR genes must be normalised to sequence library size, the *16S rRNA* housekeeping gene or single copy gene, such as the *ropB,* to estimate their relative abundance to permit subsequent cross-study comparisons (Liguori et al. 2022); 6) detection of novel AMR gene(s) is impeded when reference databases are used to map detected sequence (Bengtsson-Palme et al. 2017). Finally, the vast majority of studies employing shotgun metagenomics to monitor AMR are on centralised WWT (i.e., municipal WWT). Studies on decentralised WWT, in particular septic tanks, are scarce.

1.10 Research aims and objectives

In light of ongoing challenges in AMR monitoring and the knowledge gap of the contributions of decentralise WWT, specifically septic tanks, in the disseminating AMR to the environment, this thesis set out to:

1) Evaluate and validate the suitability of the *int*I1 gene as a proxy for inferring potential AMR pollution using the conventional QPCR and HT-QPCR on decentralised wastewater treatment plants from Thailand.
2) Evaluate the contributions of conventional septic tanks associated with household and healthcare usage, and the newly developed SST tank associated with household usage in the dissemination of AMR genes and MGE using both HT-QPCR and Shotgun metagenomics.

1.10.1 Thesis Outline

In Chapter 1 of this thesis, a general overview of AMR including some global challenges and subsequent impacts of AMR has been provided. The role of wastewater in disseminating AMR and current approaches used in AMR monitoring have been discussed. Finally, current knowledge gaps in both AMR monitoring and contributors of decentralised WWT in AMR dissemination have been highlighted.

In Chapter 2 of this thesis, the first experiment work will be presented. This work aimed to select highly suitable primer sets for optimal detection and quantification of the class 1 integron-integrase (*intI*1) from environmental samples. Subsequently, the selected primers were used to quantify the *intI*1 gene from septic tank wastewater from Thailand. First, current *intI*1 primers were evaluated via *in-silico* methods following a systematic review of the literature. Second, laboratory validation of selected *intI*1 primers, quantification of *intI*1 genes from septic tanks wastewater samples using selected primers and confirmation specificity of generated amplicon by selected *intI*1 primers via amplicon sequencing methods ensued. Finally, selected *intI*1 primers were empirically validated to confirm they are suitable for the quantification of *intI*1 gene transcripts from environmental samples. We demonstrated that selected intI1 primers were highly suitable for *intI*1 gene detection and quantification as well as quantification of *intI*1 gene transcript for gene expression analysis.

A version of this work has been published (DOI: https://doi.org/10.1128/aem.01071-23).

***intI*1 gene abundance from septic tanks in Thailand using validated *intI*1 primers**
Valentine Okonkwo, Fabien Cholet, Umer Z. Ijaz, Thammarat Koottatep, Tatchai Pussayanavin, Chongrak Polpraset, William T. Sloan, Stephanie Connelly, Cindy J. Smith

***Accepted for publication*: Applied Environmental Microbiology (AEM) Journal**

In Chapter 3 of this thesis, the second experimental work will be presented. This work employed the high-throughput QPCR tool to quantify AMR genes and the integrase genes associated with the spread of AMR genes (*intI*1, *intI*2, *intI*3) from the Thai wastewater samples. Next, the link between the integrase abundance, in particular *intI*1 abundance, to overall AMR abundance was assessed. First, the HT-QPCR array was validated, in light of

the implied sub-optimal condition for some assays on the array, using two array assays targeting the *16S rRNA* and *intI*1 gene.

In the final experimental chapter (Chapter 4), shotgun metagenomic PCR-Free approach was utilised to comprehensively characterise AMR genes including those conferring resistance to antibiotics (ARGs), heavy metal, biocides, acid and heat resistance (stress genes). A schematic illustration of the work packages for this thesis can be found in Figure 1.6.

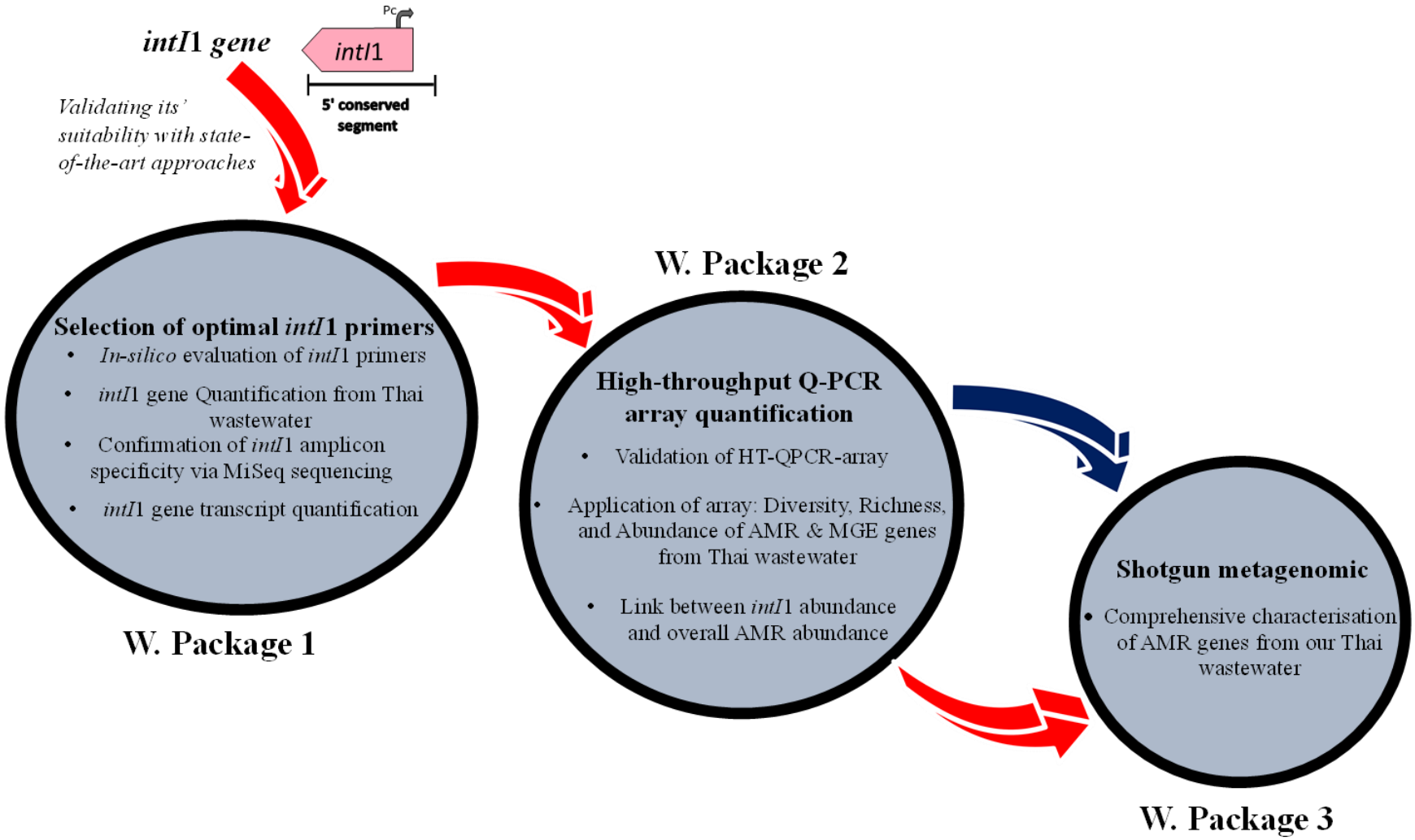

Figure 1.6: Schematic illustration of work packages carried out for this thesis.

Chapter 2

*intI*1 primer selection for class1 integron integrase gene and transcript quantification – validation and application for monitoring *intI*1 gene abundance within septic tanks in Thailand

A version of this Chapter has been published in Applied Environmental Microbiology (AEM) Journal (DOI: https://doi.org/10.1128/aem.01071-23).

Titled: *intI*1 gene abundance from septic tanks in Thailand using validated *intI*1 primers.

Valentine Okonkwo, Fabien Cholet, Umer Z. Ijaz, Thammarat Koottatep, Tatchai Pussayanavin, Chongrak Polpraset, William T. Sloan, Stephanie Connelly, Cindy J. Smith

2.1 Introduction

AMR, the ability of microbes to grow and thrive in the presence of compounds capable of limiting their cellular growth or killing cells, is a serious growing public health concern globally; and has recently been classified as "*one of the top ten global threats facing humanity*" by the World Health Organisation (WHO, 2021).

The occurrence of AMR via mutation and subsequent vertical gene transfer or acquisition of AMR genes via HGT is an inevitable natural phenomenon in the evolution of microbes (Holmes et al. 2016; Hayward et al. 2019). Nonetheless, recent global challenges including extensive consumption and misuse of antimicrobials, particularly antibiotics, in clinical settings, agri-and aqua-culture and their subsequent release to the environment, have given rise to the emergence and rapid dissemination of AMR genes amongst bacteria, including microbes of clinical importance, and the environment (Holmes et al. 2016). Consequently, high global mortality, as a result of patient treatment failure, has been associated with AMR-related infections (1.27 million global deaths in 2019 directly attributed to bacterial infections from AMR (Murray et al. 2022)). Moreover, the global death toll from AMR-

related infections has been projected to increase to 10 million deaths per year by 2050 surpassing death from cancer, assuming no change to the current trends/policies, coupled with an economic burden of 100 trillion US Dollars (O' Neill 2014).

WWT, including decentralised treatment systems such as septic tanks, receive significant amounts of antibiotics from human and animal waste (30 to 90% of antibiotics are excreted in urine and faeces (Sarmah et al. 2006)) and are now recognised as a unique interface where both human pathogenic and non-pathogenic microorganisms from humans, animals and the environment interact and exchange genes via HGT (Che et al. 2022). The selective pressure introduced by the often multiple, low-level, sub-inhibitory concentrations of antimicrobials found in wastewater, promotes AMR gene acquisition amongst microbes and selection for AMR bacteria. WWT and septic tanks are unable to effectively remove these (Gillings et al. 2015; Gillings 2017a; Hayward et al. 2019), completely from treated waste (sludge and effluent) resulting in increased AMR genes and bacteria discharged directly to the environment, contributing significantly to the global burden (Amos et al. 2018). The global AMR burden from wastewater is further exacerbated in the Global South due to the high prevalence of extensive antibiotic usage-propelled by poor regulations on usage, ineffective or lacking WWT, coupled with increasing populations and rapidly expanding megacities.

The necessity to tackle AMR discharge from WWT to the environment requires a comprehensive understanding of the role of WWT in the dissemination of AMR to the environment. This understanding will create unique opportunities to implement key strategies to mitigate AMR spread, and in turn, allow for the safeguarding of global public health. Accurate and sensitive detection, quantification, and tracking of AMR genes from source (e.g., WWT) to the environment are crucial for this purpose.

However, multiple AMR genes exist within WWT. Monitoring numerous AMR genes simultaneously is a major challenge, particularly if a rapid assessment is needed (Gillings et al. 2015). Similarly, monitoring one or a subset of AMR genes is neither ideal, as selected AMR gene(s) may be absent (Gillings et al. 2015). Previously, the clinical CL1-integron integrase (*intI*1*)* gene was proposed as a proxy for inferring potential AMR, which circumvents multiple monitoring limitations, by acting as a proxy for potential AMR pollution (Gillings et al. 2015). *intI*1 gene was proposed as a proxy as it is linked to genes that confer resistance to antibiotics, disinfectants and heavy metals; it is found in diverse taxonomic groups of pathogenic and non-pathogenic bacteria and can move across taxa via HGT due to its physical linkage to mobile genetic elements (MGEs) such as plasmid and

transposons; its abundance can rapidly change in response to external pressures such as the presence of antibiotics; selection pressures imposed by recent human activities have resulted in the emergence of the highly conserved clinical *intI*1 variant (Gillings et al. 2015), the elevated presence of which in the environment indicates pollution and potential hotspot for AMR transfer (Gillings et al. 2015; Pruden et al. 2021).

Currently, molecular approaches, specifically Q-PCR, have emerged as the methods of choice for AMR and CL1-integron detection and quantification in the environment. By far the most prevalent approach for detecting or quantifying the CL1-integron is the amplification of the *intI*1 gene at the 5' conserved segment (CS) across diverse ecological niches including engineered systems e.g. WWTs (Chen and Zhang 2013b; Berglund et al. 2015; Li et al. 2016) and natural ecosystems such as sediments (Lapara et al. 2011; Dong et al. 2019). Whilst targeting the *intI*1 gene provides no information about the structure beyond the 5' CS, quantification of the *intI*1 gene as an initial screening to infer potential AMR contamination within complex environments is invaluable and a useful initial screening approach. However, within the literature numerous primers targeting the *intI*1 gene are available (see Appendix Table A.1) and different sets are used across different studies. The current lack of standardisation prevents cross-study comparisons and limits the current understanding of AMR in the environment. As such, selecting optimal *intI*1 primers with both high coverage and specificity suitable for environmental monitoring is a challenge. Moreover, several primers have been designed based on the highly conserved clinical *intI*1 gene sequences (≥98% protein similarity to e each other), and the extent to which these primers target the less conserved *intI*1 gene variants (<98% protein similarity) found also in environmental samples (Gillings et al. 2008b; Gillings et al. 2008a; Hardwick et al. 2008; Gillings et al. 2015) and on the chromosome non-pathogenic *Betaproteobacteria* which carries gene cassettes not associated with AMR genes, has yet to be determined. As such a comprehensive and comparative evaluation of published *intI*1 primers to determine their coverage and specificity against clinical and environmental *intI*1 sequences to identify a consensus optimal *intI*1 primers for monitoring AMR within environmental samples is urgently needed (Zhang et al. 2018a).

With this need identified, we undertook to review, evaluate, and then apply *intI*1 primers to quantify the gene across a suite of wastewater samples from septic tanks in Thailand. Specifically, we compare the recent solar septic tank (SST) technology currently implemented in some areas of Thailand (Polprasert et al. 2018; Connelly et al. 2019) to that of conventional septic tanks (CST) treating household and healthcare wastewater. The SST

technology differs from CST primarily by the incorporation of a central disinfection chamber containing a heated copper coil connected to a passive solar heat collection system installed on the roof of the toilet block served by the SST (Polprasert et al. 2018; Connelly et al. 2019). The heat from the central chamber (50 - 60°C by design) promotes partial pasteurisation as the effluent passes through the chamber prior to discharge. Effluent water quality is improved by reducing microbial biomass including potential pathogens, and by extension reduction of the microbial load should reduce the AMR burden to receiving water bodies. Moreover, the in-tank temperature is raised by the passive transfer of heat from the chamber to the rest of the tank; thus, promoting enhanced microbial degradation of both retained solids (sludge) and soluble compounds (Connelly et al. 2019). As such, we hypothesise that *intI*1 gene abundance would be lower in the SST than in the CST sludge and effluent owing to the enhanced treatment caused by the increased temperature.

To address this methodological knowledge gap and our hypothesis a systematic review of the literature was undertaken to obtain published *intI*1 primers followed by a comprehensive *in-silico* analysis of primer coverage and specificity against a curated database of clinical and environmental *intI*1 sequences was completed to select the best-performing primers. A subset of the best-performing primer sets was used to quantify *intI*1 gene abundance from 30 septic tank wastewater samples comparing conventional (healthcare and household wastewater) and solar septic tank (household wastewater), with *intI*1 specificity validated by Illumina MiSeq. We further confirmed the suitability of the primers to quantify *intI*1 gene transcripts. Thus, we propose a validated *intI*1 primer set for the quantification of genes and transcripts from environmental samples towards the goal of achieving standardisation across *intI*1 studies.

2.2 Materials and Methods

2.2.1 *intI*1 primer evaluation

2.2.1.1 Systematic review of the literature and primer alignment to *intI*1 reference sequence

A systematic review of >3000 published papers was conducted to retrieve *intI*1 primers and probe sequences across a range of settings including clinical and environmental e.g., agricultural, and human-impacted settings including WWTPs. For this, the "Web of Knowledge" database (https://www.webofscience.com/; last assessed 04/10/2022) was

searched using the term “Class 1 integron”. Only published articles in English language were considered. 3266 published articles were subsequently recovered. The *intI*1 primer sequences from the respective literature were either retrieved in the main text or from the accompanying supplementary material.

Obtained *intI*1 primer and probe sequences were aligned to a *Pseudomonas aeruginosa* plasmid pVS1 nucleotide sequence (M73819.1) using the ClustalX2 algorithm (Version 2.1.0.0), with default settings (Larkin et al. 2007) and visualised with BioEdit (version 7.0.5.3) (Hall 1999). The alignment position of each primer and probe sequence was renamed according to position along the *Pseudomonas aeruginosa* reference *intI*1 gene nucleotide sequence (Figure 2.1, Appendix Table A.1).

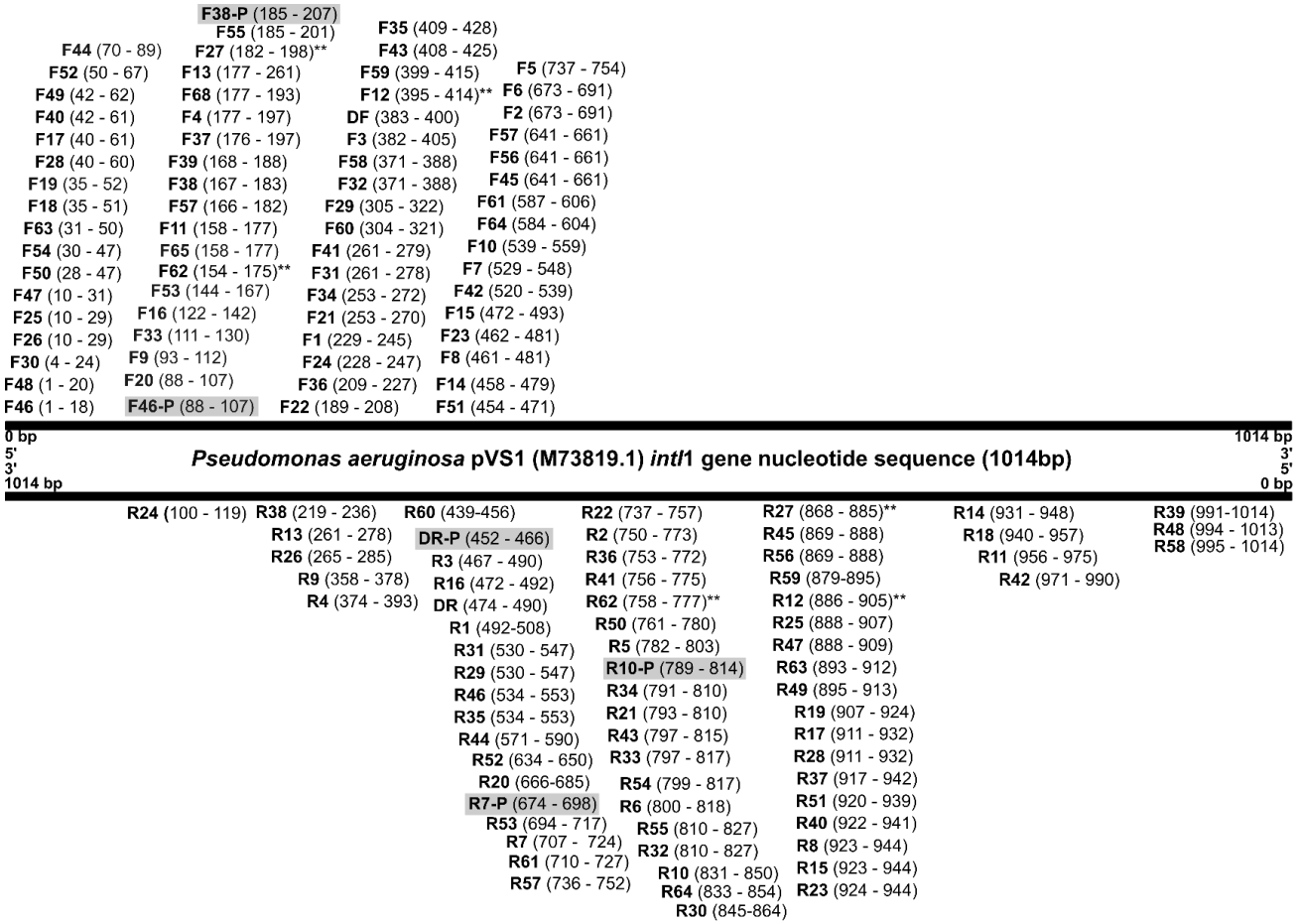

Figure 2.1: Alignment of published and newly designed *intI*1 primers and probe sequence hit position to a *Pseudomonas aeruginosa* plasmid pVS1 nucleotide sequence (M73819.1). F refers to forward primer. R refers to reverse primer. The number after F or R (i.e., F1 or R5) refers to the assigned primer ID- See Appendix Table A.1 for more detail. The number in parenthesis () denotes the position of the primer sequence on the reference *Escherichia coli intI*1 gene. ** denotes primer hit position based on Primer Prospector alignment to CP003684.1 *intI*1 nucleotide sequence. Highlighted in grey are the probe binding positions for the primer-probe primer set.

2.2.1.2 Databases construction and curation

The integron-integrase database by Zhang et al., (2018a) consisting of 922 and 2462 *intI*1 gene and integron-integrase (*intI*) of other class protein sequences respectively (herein referred to as non-*intI*1 database) was employed for the analysis of primers (Figure 2.2). Whilst the *intI* of other class database was mostly populated with protein sequences from other integron-integrase classes, it also contained a number of XerCs integrases (*n= 78*) and transposases protein sequences (*n= 66*) as recently reported by Roy et al., (2021). In this study, however, the inclusion of these protein sequences within this non-*intI*1 database is not of significance, as the goal was to confirm that analysed *intI*1 primer sets were unable to amplify sequences within this database via *in-silico* testing, thus confirming their specificity.

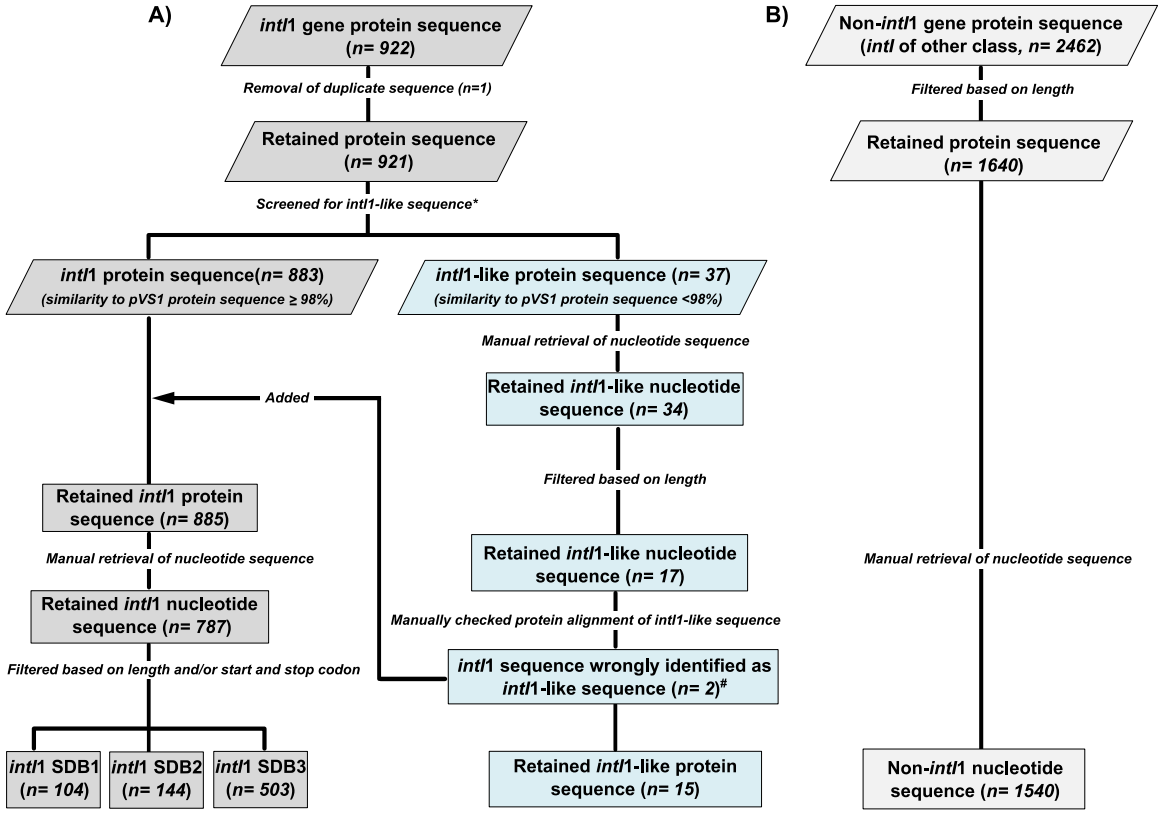

Figure 2.2: Workflow of constructed integrase sub-databases for primer evaluation using 922 *IntI*1 (A) and 2462 non-*IntI*1 protein sequences (B). Duplicate (*n=1*) *IntI*1 sequence was discarded. Retained protein sequences were compared to the reference *IntI*1 protein of pVS1 plasmid (AAA25857.1) using NCBI BlastP. Sequences with ≥ 98% identity to the pVS1 protein sequence were classified as *IntI*1 protein sequences, while those with <98% identity were classed as *IntI*1-like. Three *intI*1 gene nucleotide sub-databases (SDB1, SDB2, SDB3) were constructed, based on criteria specified in Table 2.1, for primer coverage assessment. *intI*1-like (*n=15*) and non-*intI*1 (*n=1540*) sub-databases were used to evaluate the primers specificity. * Denotes removal of one protein sequence (CP006631.1) from the 921 non-duplicate *IntI*1 protein sequence, due to low sequence similarity. # Denotes the two *IntI*1 protein sequences(WP_058137959.1 and WP_058135314.1) misidentified as *IntI*1-like protein sequences likely due to their partial length.

The *IntI*1 protein database was curated by discarding duplicate protein sequences (*n=1*) from the 922 *intI*1 protein sequences (Figure 1). Retained *IntI*1 protein sequences were then compared to the reference *IntI*1 protein sequence of pVS1 plasmid (AAA25857.1) using NCBI BlastP, to ensure *intI*1 nucleotide sequences used for in-silico assessment of primer and probe sequence coverage were indeed *intI*1 sequences as suggested by Roy et al., (2021). Further, protein sequences whose percentage identity to the reference pVS1 *IntI*1 plasmid protein sequence was ≥ 98%, were characterised as *IntI*1 sequence, while sequences whose protein similarity to the reference pVS1 *IntI*1 plasmid protein sequence were <98% were categorised as *IntI*1-like protein sequences (Roy et al. 2021) (Figure 2.2). Additionally, protein sequences identified as *IntI*1-like were manually checked to ensure the percentage similarity score to the pVS1 protein sequence reported by NCBI was not due to missing sequence caused by the alignment of a partial sequence to a complete length sequence. As such, protein sequences (*n=2*; WP_058137959.1 and WP_058135314.1) incorrectly identified as *IntI*1-like were added to the *IntI*1 protein database (Figure 2.2). Retained protein IDs for the *intI*1 and *intI*1-like, were then used to manually obtain the nucleotide sequences from NCBI in Fasta format.

In parallel, the non-*intI*1 sequence sub-database was constructed from the 2462 *intI* of other class protein sequences by applying a ≥ 300 amino acid length thresholds (900bp nucleotide length) to filter out shorter-length protein sequence (Table 2.1; Figure 2.2).

Table 2.1: Criterion for Construction of Integrase Sub-databases

Sub_databases ID	Criteria			Number of sequences within sub_database
	Nucleotide sequence length (bp)	Beginning start codons	Ending stop codons	
***intI*1 SDB1**	≥ 1000	ATG, TTG, GTG	TAA, TAG, TGA	104
***intI*1 SDB2**	≥ 900	ATG, TTG, GTG	TAA, TAG, TGA	144
***intI*1 SDB3**	≥ 600	N/A	N/A	502
***intI*1_like**	≥ 600	N/A	N/A	16
Non-*intI*1 SDB	>900	N/A	N/A	1540

SDB= Sub-database

Finally, three *intI*1 gene nucleotide sub-databases (SDB1, SDB2, and SDB3) an *intI*1-like and non-*intI*1 sequences sub-databases were created for robust primer analysis based on the

criterion specified in Table 2.1. SDB1 (*n=104*) contained full-length *intI*1 sequences >1000 bp, confirmed by the presence of a start and stop codon; SDB2 (*n=144*), contained full-length *intI*1 sequences > 900bp confirmed by the presence of a start and stop codon. Sequences within SDB1 are all present in SDB2. The final *intI*1 sub-database (SDB3, *n=503*) contained both complete and partial sequences (Table 2.1 and Figure 2.2). All sequences within SDB1 and SDB2 were also present within SDB3. The *intI*1-like (<98% similarity to pVS1) sub-database contained both complete and partial-length sequences (SDB *intI*1-like, *n=15*; Table 2.1; Figure 2.2). The non-*intI*1 database contained 1540 integrase sequences of other classes (Figure 2.2).

To summarise, *intI*1 sequences from this study were defined as *intI*1 protein sequences whose percentage identity shared a ≥ 98% similarity to pVS1 *intI*1 plasmid protein sequence (AAA25857.1), while *intI*1-like sequences were defined as *intI*1 protein sequences sharing a <98% similarity to pVS1 *intI*1 plasmid protein sequence (Figure 2.2).

2.2.1.3 Primer evaluation

Published *intI*1 primers were analysed as primer pair (Appendix Table A.1), using Primer Prospector (Walters et al. 2011), to evaluate coverage and specificity against constructed integrase sub-databases (Figure 2.2). The analyze_primers.py function with the default settings on Primer Prospector was used to generate an alignment profile file for each primer against unaligned individual nucleotide sequences in each test sub-database. For each primer alignment to a nucleotide sequence, a weighted score (WS) was given.

Overall WS was calculated as:

> Non-3' mismatches * 0.4 per mismatch + 3' mismatches * 1.0 per mismatch + Final 3' base mismatch * 3.0 per mismatch + Non-3' gaps * 1.0 per gap + 3' gap * 3.0 per gap.

The first 5 bases of the primer and the target sequence were defined as the 3'end and thus, mismatches within these bases were termed 3' mismatches. The remaining bases of the primer and the target sequence were defined as the non-3' end. Therefore, mismatches within these non-3' end bases were regarded as non-3'mismatches. Gaps in the alignment of the primer and the target sequence in the first 5 bases were termed 3' gaps while gaps in the alignment for the remaining primer and template sequence were known as non-3'gaps.

A higher WS was given for mismatches or gaps at the 3'end compared to mismatches or gaps at the non-3' end. This is simply because mismatches or gaps at the 3'end region have a significant chance of affecting PCR amplification, whereas mismatches or gaps at the non-3' end can be tolerated.

As such, the lower the WS, the better the compatibility between the primer and target DNA sequence which suggests higher primer coverage potential. A WS of 0 indicates perfect alignment. As such, the lower the WS, the better the compatibility between the primer and target DNA sequence. Primer Prospector, however, forces a primer sequence to bind anywhere within the target sequence if the primer binding site is unavailable to generate a WS for the primer.

Therefore, to evaluate primer coverage and specificity of each primer pair, the primer-binding orientation (i.e., reverse primer alignment was before the forward primer alignment position) against each unaligned nucleotide sequence was first verified from the primer hit position using R (R Core team 2022). the primer binding orientation of each analysed primer pair was checked for each sequence. Only sequences that had the correct primer-binding orientation (5' to 3' directionality) were analysed since DNA synthesis is from 5' to 3'. Sequences with incorrect primer-binding orientation as a result of missing primer-binding sites were discarded from each sub-database. Following this, the number of amplicons estimated to be amplified by each primer pair was calculated using the sum of the WS of the forward and reverse primer for each primer pair. If the sum was ≤ a defined threshold (0 - perfect match to 10 – incompatible match), then the primer set was considered to amplify the target sequence within the test database. Furthermore, the mean overall WS for the forward and reverse primer of each primer set with the correct binding orientation was calculated. Lastly, the R package "ggplot2" (Wickham 2009) was then used to generate a WS plot of each primer set based on the defined WS threshold.

In the case of primer pairs that incorporated a TaqMan hydrolysis probe, the primer-probe-binding-orientation (forward, probe and reverse) against each unaligned sequence was first verified, for each unaligned sequence by checking the hit positions of the forward, probe, and reverse primer sequence in R. Unaligned sequences with correct primer-probe orientation were subsequently retained and analysed in the manner same described above.

2.2.1.4 Design of a new *intI*1 primer set and TaqMan-minor-groove binder (TaqMan-MGB) probe

To improve *intI*1 sequence coverage and specificity for Q-PCR the *intI*1 primer set F3-R3, (Rosewarne et al., 2010, Appendix Table A.1) was modified to generate a new *intI*1 primer incorporating an MGB TaqMan probe set (*intI*1 DF-DR, Appendix Table A.1) following guidelines for primer-probe design outlined by McKew and Smith (McKew and Smith 2015). An MGB probe of 15bp was designed using Primer Express software (Version 3.0.1; Applied Biosystems)^{TM}. Briefly, *intI*1 sequences within SDB3 (*n=503*) were aligned using the MAFFT algorithm (Katoh et al. 2002). Aligned sequences were imported into EMBOSS Cons website (Last accessed 04/08/2021 https://www.ebi.ac.uk/Tools/msa/emboss_cons/) to generate a consensus sequence. The consensus sequence was exported into Primer Express software (Version 3.0.1; Applied Biosystems)^{TM} with the selection of the TaqMan MGB Quantification option. MGB probe design parameter was set to a minimum length of 13 bp and a maximum of 15 bp. Probe sequence with minimal secondary structures, and closer to the position of the forward or reverse primer was selected. Modified *intI*1 primer set (DF-DR) and designed MGB TaqMan-probe sequence (Appendix Table A.1) were BLAST searched (BLASTN) to validate the sequence specificity before proceeding to the alignment of probe sequence to reference pSV1 *intI*1 gene sequence (Figure 2.1) and subsequently, *in-silico* validation of the newly designed primer and probe set across constructed integrase sub-databases as specified above **(see section 2.2.1.3).**

2.2.2 Validation of selected *intI*1 primers from *in-silico* analysis on wastewater samples

2.2.2.1 Optimisation of selected primer sets for Q-PCR

The amplicon produced from selected primers for laboratory validation was assessed *in-silico* first using sequences within SDB1 and then in the laboratory by end-point PCR. Selected *intI*1 primer sets (Table 2.7). that resulted in the correct size amplicon were further optimised for Q-PCR assays.

Q-PCR standard curves were constructed by amplifying synthetic *intI*1 gene fragments containing the primer binding site for all selected primers inserted in a circularised, next-generation sequencing (NGS) verified, ampicillin-resistant vector from Integrated DNA Technologies (Figure 2.3). Briefly, *Escherichia coli* (NC_011964.1) *intI*1 gene fragment containing the primer-binding site for the three selected primers, with additional 10 bases at the ends of the total primer site, was flanked with a T7-forward (5'-TAATACGACTCACTATAGGG-3') and M13 reverse primers (5'-

CAGGAAACAGCTATGAC-3'), resulting in a 660bp gene fragment (Figure 2.3). The 660bp gene fragment (sequence can be found in below Figure 2.3) was then inserted in a circularised, NGS-verified, ampicillin-resistance vector (pUCIDT-AMP). pUCIDT-AMP vector DNA was resuspended in a 20µl IDTE (10mM Tris, 0.1mM EDTA) buffer at pH 7.8, with a final concentration of 200ng/µl.

Endpoint-PCR was performed using T7 (5'-TAATACGACTCACTATAGGG-3') and M13 (5'-CAGGAAACAGCTATGAC-3') flanking primers, with the HotStartTaq PCR kit (Qiagen) in a 25µl volume, which consisted of 15.875µl nuclease-free water, 2.5µl 10x PCR Buffer, 0.125µl HotStartTaq, 0.5µl dNTPs (10µM), 0.5µl of each primer (10 µM each), and 5µl (10ng) template DNA (pUCIDT- vector). The reaction condition was as follows: 95°C-15 min, (94°C-30s, 57°C-30s, 72°C-60s) ×29 cycles and a final extension at 72°C for 10 mins. The resultant PCR product was cleaned, and size selected with the Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA) per manufacturer's recommendation, using a 1:1 ratio of beads volume to PCR product volume, and eluted in a 25µl volume nuclease-free water. Cleaned PCR products were quantified fluorometrically using Qubit (Invitrogen, according to the manufacturer recommendations) and the gene copy number was determined using EndMemo DNA copy number calculator (http://endmemo.com/bio/dnacopynum.php). The purified concentrated stock was subsequently diluted to 10^9 copies/µl, followed by a five, 10-fold serial dilution (10^7-10^3 copies/µl) for amplification by Q-PCR. A standard curve was obtained by plotting the average of each triplicate threshold cycle (Cq) against the natural log of standard concentration (copies/µl). Standard curve descriptors including efficiency, slope, y-intercept and R^2 are reported (Smith and Osborn 2009).

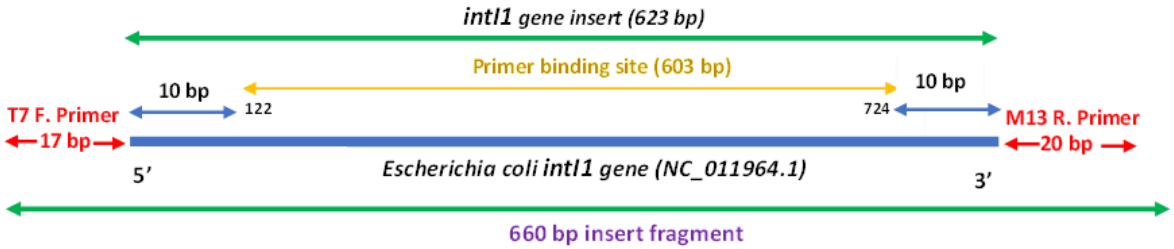

Figure 2.3: Description of synthetic gene *intI*1 fragment inserted into a circularised, double-stranded NGS verified plasmid vector used for constructing QPCR standard curve. A 603bp region on *Escherichia coli intI*1 sequence (NC_011964.1) containing the binding site for the selected *intI*1 primers for laboratory validation was flanked with an extra 10 bases (total *E. coli intI*1 gene sequence= 623 bases). The 623 bp *intI*1 gene fragments were subsequently flanked with the T7-Foward and M13-Reverse primer. Total insert fragment= 660bp. The 660bp gene fragment was then inserted into an NGS-verified circular, ampicillin-resistant plasmid vector. F denotes forward primer and R reverse primer. The number following the forward primer indicates the hit start position of the first base of the forward primer while the number following the reverse primer indicates the hit position of the last base of the reverse primer. See Figure 2.1 for detailed information of the binding position of the primers. The 660bp insert fragment sequence was: >NC_011964.1 *Escherichia coli* plasmid insert fragment
TAATACGACTCACTATAGGGGTCCACTGGGTTCGTGCCTTCATCCGTTTCCACG
GTGTGCGTCACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGCATTTCTGTCC
TGGCTGGCGAACGAGCGCAAGGTTTCGGTCTCCACGCATCGTCAGGCATTGGC
GGCCTTGCTGTTCTTCTACGGCAAGGTGCTGTGCACGGATCTGCCCTGGCTTCA
GGAGATCGGAAGACCTCGGCCGTCGCGGCGCTTGCCGGTGGTGCTGACCCCGG
ATGAAGTGGTTCGCATCCTCGGTTTTCTGGAAGGCGAGCATCGTTTGTTCGCCC
AGCTTCTGTATGGAACGGGCATGCGGATCAGTGAGGGTTTGCAACTGCGGGTC
AAGGATCTGGATTTCGATCACGGCACGATCATCGTGCGGGAGGGCAAGGGCTC
CAAGGATCGGGCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGC
AGCTGTCGCGTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGCAG
CGGCGTTGCGCTTCCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATT
CCTGGCCGTGGTTCTGGGTTTTTGCGCAGCACACGCATTCGACCGATCCACGG
AGCGGGTCATAGCTGTTTCCTG

2.2.3 Application of selected *intI*1 primers from SST and CST wastewater samples

2.2.3.1 Solar and Conventional tank sampling

Two household scale solar septic tank (SST; SST01 and SST07) units and three conventional septic tank (CST; two household tanks and one healthcare tank) units, operational within the Pathum Thani province and Samut Prakan province, Thailand, were sampled between April 2018 to September 2019 (Table 2.2).

Table 2.2: Selected Sample Timepoint for Each Septic Tank Investigated

Reactor type	Reactor ID	April 2018	May 2018	June 2018	Nov 2018	March 2019	June 2019	July 2019	August 2019	Sept 2019
Solar septic tank	SST-01	EFF/ SLG	EFF/ SLG	EFF/ SLG						
	SST-07	EFF/ SLG			EFF/ SLG	EFF/ SLG				
Conventional septic tank	CST-P3						INF/EFF/ SLG	INF/EFF/ SLG	INF/EFF/ SLG	
	CST-J6						EFF/SLG	EFF/SLG	EFF/SLG	
	CST-HC								EFF/SLG	EFF† /SLG

INF Influent; EFF Effluent; SLG Sludge. SST Solar septic tank; CST Conventional septic tank

† Excluded from *intI*1 Q-PCR quantification due to insufficient sample volume but was included in MiSeq amplicon sequencing.

The SST and the household CST units (CST-P3 and CST-J6) have a 1000L total working capacity, whilst the healthcare CST units (CST-HC2; herein referred to as CST-HC) has a 2000L total working capacity each. Each tank was buried to approximately 1.5 metres below ground level; with the tank surface (lid) at ground level, and so exposed to atmospheric temperatures (Connelly et al. 2019). Details and descriptions of the tanks can be found in Table 2.3

Table 2.3: Details and descriptions of the Thai septic tanks

Unit type	Unit details	
Solar septic tank (SST)	Tank ID (Size):	**SST-01** (1000 L)
	Heating Device:	Evacuated tube collector (36 tubes) and 200 L of storage tank
	Date of Installation:	March 2015
	Operation Period (Site):	March 2015 – Present (Samut Prakan, Thailand)
	Number of toilets:	1 toilet
	User:	Public toilet at a factory (Santavee Factory, Thailand)
	Toilet type:	Flush sitting toilet
	Tank ID (Size):	**SST-07** (1000 L)
	Heating Device:	Evacuated tube collector (36 tubes) and 200 L of storage tank
	Date of Installation:	August 2017
	Operation Period (Site):	September 2017 – Present (Pathum Thani province, Thailand)
	Number of toilets:	2 toilets
	User:	5 people
	Toilet type:	Pour squat toilet and flush sitting toilet
Conventional septic tank (CST)	Tank ID (Size):	**CST-P3** (1000 L)
	Date of Installation:	-
	Operation Period (Site):	- (Pathum Thani province, Thailand)
	Number of toilets:	2 toilets
	User:	-
	Latitude:	14°05'24.3"N
	Longitude:	100°35'29.0"E
	Tank ID (Size):	**CST-J6** (1000 L)
	Date of Installation:	August 2017
	Operation Period (Site):	September 2015 – Present (Pathum Thani province, Thailand)
	Number of toilets:	2 toilets
	User:	-
	Latitude:	14°04'13.4"N
	Longitude:	100°35'27.1"E
	Tank ID (Size):	**CST-HC2** (2000 L) **(referred to as CST-HC in this thesis)**
	Date of Installation:	-
	Operation Period (Site):	- (Public Health Service Centre 3, Pathum Thani province, Thailand)
	Number of toilets:	2-4 toilets
	User:	-
	Latitude:	13°58'31.3"N
	Longitude:	100°37'32.2"E

CST influent was sampled by disconnecting inflow to the CST septic tank via a sampling valve for 24 hours. Waste generated during the 24 hours was collected in a sealed bucket, followed by homogenisation of the buckets' content. Three 1 L homogenised samples were then collected in storage bottles and stored at -80˚c if not in use immediately for downstream processing. The physiochemical and tank operational parameters of all septic tank units were measured (Connelly et al. 2019). Due to the inaccessibility of influent samples, influent was only collected for one CST-household unit (CST-P3).

CST and SST effluent and sludge samples were collected prior to influent sampling to ensure samples were representative of the system under normal operating conditions. Sampling of effluent was done by flushing the toilet once to clear the outflow pipe of residual materials, followed by collection of effluent in a 10L bucket after a second flush. The effluent was homogenised by mixing, and three 1L sub-samples were collected in 1L bottles for later use.

Conversely, sampling of sludge was done by homogenising the tank contents by mixing using a submersible pump. Subsequently, 2L of homogenised sample was collected into a plastic beaker through tubing (2cm internal diameter) inserted into the tank and connected to an external vacuum pump (Sacco, Model SC-1A). Contents of the beaker were thoroughly mixed by stirring and then four sub-samples were taken in 50mL centrifuge tubes before storing on ice (approximately 2 hours) and transported to the laboratory for downstream processing (Connelly et al. 2019).

100 ml of effluent and 40 ml of sludge were sampled from the SST, CSTJ7 and CST-HC2, while 100 ml of influent was also collected from CST-P3. 40 ml of sludge was sampled from each reactor. All samples were pelleted for DNA extraction. The months for sampling the SST were selected based on the highest recorded internal temperature of the 12-month sampling campaign conducted (Table 2.2). All samples were centrifuged and pelleted, and DNA was extracted from 0.5g sludge. The months for sampling the SST were selected based on the highest recorded internal temperature of the 12-month sampling campaign conducted (Table 2.2).

2.2.3.2 DNA extraction

From each sample, DNA extraction was performed with the DNeasy PowerSoil Kit (Qiagen), following the manufacturer's instructions. The integrity of extracted genomic DNA was assessed via agarose gel electrophoresis and DNA concentration was quantified fluorometrically using the Qubit (Invitrogen) according to manufacturer instructions.

2.2.3.3 Q-PCR quantification of *intI*1 gene from wastewater

*intI*1 genes were quantified from septic tank wastewater samples from Thailand (Table 2) using optimised Q-PCR conditions for the three selected *intI*1 primer pairs (DF-DR, F3-R3 and F7-R7). For each primer set, Q-PCR amplification was carried out in a 20µl volume reaction using 2µl (1:50 diluted) template DNA. Reaction volume, conditions, primer sequences and probe type for the three selected optimal *intI*1 primer pairs are detailed in Table 2.7. Triplicate/ duplicate no template control (NTC) was included for each primer set. Reactions were performed on the Bio-Rad CFX96 Touch Real-Time PCR Detection System and analysed with the Bio-Rad CFX Manager 3.1 software. Melt curve analysis was performed, for the SYBR Green assay, from 65°C to 95°C with 0.5°C increments every 5 secs, and a single peak was confirmed to ensure assay specificity.

Statistical analyses were performed in R (R Development Core Team, 2008). The Shapiro-Wilk test was first used to assess the normality of data, with a p-value of 0.05 chosen as the significance threshold. Two-way analysis of variance (ANOVA) followed by a Turkey HSD post hoc test, was subsequently employed to compare *intI*1 and *16S rRNA* gene abundance for each of the sample types (influent, sludge, and effluent) and reactor type (CST and SST) for each primer set. Finally, a Spearman rank correlation analysis was applied, following the Shapiro-Wilk normality test, to calculate the relationship between the abundance of *intI*1 detected between each primer set.

2.2.4 MiSeq Amplicon sequencing to confirm the specificity of Q-PCR amplicon

The specificity of the selected *intI*1 primer sets (DF-DR, F3-R3 and F7-R7) used to quantify *intI*1 gene from septic tank sludge and wastewater, were confirmed by Illumina MiSeq amplicon sequencing of the *intI*1 gene from 31 wastewater samples (Table 2.2) using the optimised endpoint PCR conditions outlined in Table 2.7. A two-step PCR was performed to barcode samples as detailed previously (Bourlat et al. 2016; Cholet et al. 2019). To do so, a two-step PCR was performed using a similar method detailed previously (Bourlat et al. 2016; Cholet et al. 2019). The first PCR step amplified the target region using respective *intI*1 primers (primer sequences outlined in Table 2.7), attached with Illumina adaptors at the 5' end: 5′-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG- 3' (forward adaptor); 5′-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G- 3' (reverse adaptor). For each primer set, PCR amplification was carried out in a 25µl volume reaction using 5µl (5ng) template DNA and the HotStartTaq PCR kit (Qiagen). Each 25µl volume reaction consisted of: 15.75µl (DF-R7 and F3-R3)/ 14.75µl (F7-R7) nuclease-free water, 0.5µl (DF-R7 and F3-R3)/ 1µl (F7-R7) of each primer (10 µM each), 0.5µl dNTPs (10 µM each), 0.25µl HotStartTaq and 2.5µl of 10x PCR Buffer.

A no template control and positive control (plasmid vector containing target sequence) were included for each primer set. The PCR thermocycling conditions are specified in Table 2.7. Gel electrophoresis was performed on generated amplicons to confirm the expected amplicon size and quality. For each primer set, duplicate, or triplicate PCRs were carried out on the samples (depending on the intensity of the band seen on gel following gel image analysis), using the same conditions, and then pooled together for further processing. PCR amplicons were cleaned, and size selected using a 1.5X volume ratio Agencourt AMPure

XP beads (Beckman Coulter, Brea, CA, USA) according to the manufacturer's recommendation and eluted in 30µl of nuclease-free water.

The second PCR step (index PCR) was performed to incorporate Illumina dual index (i5 and i7) using the Nextera XT Index Kit in a 25µl volume reaction which consisted of: 6.75µl nuclease-free water, 2.5µl of 10x PCR Buffer (Qiagen), 0.25µl HotStartTaq, 0.5µl dNTPs (10 µM each), 5µl of each index primer (10 µM each), and 5µl template. The cycle conditions are detailed in Table 2.7. Amplicons generated were purified using a 1.5X volume ratio Agencourt AMPure XP beads and eluted in 25µl nuclease-free water. Following this, three samples were chosen at random from each primer set and ran on the Bioanalyser following the DNA 1000 Assay protocol (Agilent Technologies, UK) to determine the average length of the amplicons generated by each primer set, and to confirm the absence of unspecific products. The DNA concentration of each amplicon, from each primer set, was determined fluorometrically (Qubit) and molarity was calculated using the following equation:

(Concentration in ng/µl) × 106 = (660 g/mol × average library size)

For each primer set, prepared libraries were pooled at an equimolar amount into individual tubes. Subsequently, the three libraries were pooled at an equimolar amount to make the final pooled library. Lastly, the final pooled library was measured fluorometrically (Qubit) before being sent to GENEWIZ (GENEWIZ Sequencing, UK) for MiSeq amplicon sequencing on the Illumina platform (2 × 250 bp paired-end).

2.2.4.1 Bioinformatics

Primer sequences were used to extract the *intI*1 gene from the resulting reads, particularly for shorter primer pairs, using the Cutadapt algorithm (Martin 2011). Abundance tables were then generated by constructing amplicon sequencing variants (ASVs) using the Qiime2 pipeline and the DADA2 algorithm (Bolyen et al. 2019) with details given at https://github.com/umerijaz/tutorials/blob/master/qiime2_tutorial.md. Constructed ASVs blast searched on NCBI and closest hit sequences retrieved for each ASV. The phylogenetic distance between sequences was investigated. First, multiple sequence alignments of ASV sequences, retrieved NCBI sequences, complete length *intI*1 and *intI*1-like and an *intI*3 (class three integrase gene) sequence were done using MAFFT (Katoh et al. 2002) for each primer set. Aligned sequences were visualised in BioEdit (version 7.0.5.3) (Hall, 1999) and trimmed to retain only aligned regions without gaps. Phylogenetic trees were constructed using a

maximum likelihood approach with a generalised timer-reversible substitution model implemented in RAxML version 8 (Stamatakis 2014). Consensus trees were calculated after 1000 bootstrapping permutations.

Phylogenetic tree of the trimmed and aligned sequence, for each primer pair, was constructed with RAxML (Price et al. 2009). A heat tree of the constructed ASVs, after log2 transformation of ASVs abundance per sample for each primer set, was mapped to analysed samples, coloured, and visualised using the ggtree package (Yu et al. 2017). The tip of the tree was coloured based on the sequence isolation source.

Availability of Supporting Data

The sequence data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB65102 with the metadata provided in the supplementary material 2 (https://www.ncbi.nlm.nih.gov/bioproject/PRJEB65102/)."

2.2.5 Validation of selected primers to quantify *intI*1 mRNA transcripts from environmental samples

2.2.5.1 Sample collection, filtration, and DNA/RNA co-extraction

As the septic tank wastewater samples were previously collected and only DNA extracted and stored at -80°C for an extended period (over two years), they were not suitable for RNA analysis (Cholet et al. 2019). Therefore, we tested the suitability of the optimised primer sets to detect *intI*1 mRNA using freshly collected environmental samples of river water collected from the Kevin River, Glasgow (UK), to determine if *intI*1 mRNA transcripts could be quantified in receiving water bodies. 3L of surface water was collected in March and April 2022 and filtered through a sterile glass microfibre filter (FisherBrand MF200; retention 1.2µm) and onto a 0.22µm Sterivex filter. Filters were immediately extracted from or frozen at -80°c for later use.

DNA-RNA co-extraction was carried out according to the protocol previously described (Griffiths et al. 2000; Tatti et al. 2016; Cholet et al. 2019), First, all glassware was baked at 180°C overnight to inactivate RNases. Lids of glassware and stirrers were soaked overnight in RNase Zap (Ambion) and disposable plasticwares used, including tubes, were RNase free. All solutions used for nucleic acid extraction were prepared using diethylpyrocarbonate

(DEPC) treated Milli-Q water. Total nucleic acid using the phenol-chloroform method described by Griffiths et al., (2000), with a minor modification to the bead-beating time (45 sec) as outlined by Lim et al., (2016). Briefly, glass microfibre filters were split into halves using sterile forceps, and each half was placed in a matrix E-bead-beating tube (MP Biomedical) and immediately transferred on ice. 0.5ml 5% CTAB (hexadecyltrimethylammonium bromide)/phosphate buffer (120mM, pH8; consisting of 2.58g K2HPO4.3H2O; 0.10g KH2PO4; 5.0g CTAB; 2.05g NaCl; 100ml DEPC water) and 0.5ml Phenol: Chloroform: Isoamyl alcohol (25:24:1; v:v:v; pH8) were added to each bead-beating tube. Samples were lysed on the FastPrep system (MP Biomedical) at 6.0m s^{-1} for 45 sec and then centrifuged at 12,000g for 20 mins at 4°C. The top aqueous layer was transferred to a sterile 2ml tube and mixed with 0.5 ml chloroform: isoamyl alcohol (24:1 v:v) followed by a centrifugation at 16,000 g for 5 mins at 5^{O}C. The top aqueous layer was transferred to a new sterile 2 ml tube and total nucleic acids were precipitated by adding 2 volumes of 30% polyethyleneglycol 6000 (PEG6000)/ NaCl (1.6M) solution to the tube. The resulting mixture was incubated on ice for 2 hours and then centrifuged at 16,000g for 30 mins at 4°C. The pellet was carefully recovered, by discarding the supernatant, and then washed with 1ml ice-cold 70% ethanol, followed by centrifugation at 16,000g for 30 mins at 4˚C. The ethanol wash was carefully discarded, and the tube (containing the pellet) was spun briefly for 5 sec at 4°C to remove residual ethanol. The recovered total nucleic acids pellet was air dried and subsequently re-suspended in 50µl DEPC treated water. The concentration of DNA in samples was determined fluorometrically (Qubit) and gel electrophoresis was performed to confirm the success of DNA and RNA co-extraction. Co-extracts (DNA and RNA) were stored at -80˚c if not used immediately.

2.2.5.2 RNA preparation and cDNA synthesis

RNA was prepared from the raw DNA/RNA co-extract by DNase treating with Turbo DNase Kit (Ambion) following the manufacturer's recommendation, with modification to the incubation time and volume of DNase added as previously described (Cholet et al. 2019). 1µl DNase volume was added to the samples and incubated at 37˚C for 1 hour, followed by further addition of 1µl DNase volume to the sample and a re-incubation at 37˚C for another hour. Subsequently, the success of DNase treatment was confirmed by no PCR amplification of the V4 - V5 region bacterial *16S rRNA* gene using the 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 926R (5'-CCGYCAATTYMTTTRAGTTT-3') primers (Suzuki et al. 2000). The PCR amplification was carried out in a 25µl volume reaction with the HotStartTaq PCR kit (Qiagen) containing 18.8µl nuclease-free water, 2.5µl

10x PCR Buffer, 0.2µl HotStartTaq, 0.5µl dNTPs (10µM), and 0.5µl of each primer (10µM each) and 2µl neat template (DNase treated RNA). The reaction condition was as follow: 95°C -15 min, (94°C -45s, 50°C -30s, 72°C -40s) × 35 cycles and a final extension at 72°C for 10 mins. DNA-free total RNA concentration in the samples was quantified fluorometrically (Qubit) and by Bioanalyser following the RNA 6000 Nano Assay protocol (Agilent Technologies, UK). The RNA integrity number (RIN) was also determined by the Bioanalyser based on the 23S/16S rRNA ratio.

DNA-Free RNA was immediately reverse-transcribed using the superscript IV reverse transcription kit (Invitrogen). Both Random hexamer (RH) and gene-specific (GS) reverse transcription were performed for the DF-DR and F7-R7 primer sets and only GS reserve transcription for the F3-R3 primer set. The initial reverse transcription reaction mix which consisted of 8µl water, 1µl primer (10µM gene-specific/ 50µM random hexamer), 1µl dNTP's (10µM each) and 3µl RNA template was incubated at 65°C for 5 min and immediately transferred to ice for 1 min. A second reaction mix which contained 4µl 5X first-strand buffer, 1µl 0.1 mM dithiothreitol (DTT), 1µl RNAse inhibitor (40 unit/µl) and 1µl SuperScript IV (200 unit/µl) was subsequently added and then incubated at 55°C for 10 min and 80°C for 10 min for gene specific priming/ 23°C for 10 min, 55°C for 10 min and 80°C for 10 min for Random hexamer priming (Cholet et al. 2019).

2.2.5.3 RT-Q-PCR quantification of *intI*1 genes and transcripts from river water

Q-PCR DNA standard curves were constructed as above (see section 2.2.2.1). For each primer set, *intI*1 cDNA and DNA Q-PCR amplification was carried out in a 20µl volume reaction using 2µl (1:2 and/ 1:5 diluted) template DNA/ cDNA. In addition, two priming strategies, Gene-specific (GS) and/ or Random (RH) priming were used to reverse transcribe *intI*1 mRNA to cDNA as described above (see section 2.2.5.2). Q-PCR Reaction volume, conditions, primer sequences and probe type for the three selected optimal *intI*1 primer pairs are the same as specified in Table 2.7. Reactions were performed on the Bio-Rad CFX96 Touch Real-Time PCR Detection System and analysed with the Bio-Rad CFX Manager 3.1 software. Melt curve analysis was performed, for the SYBR Green assay, from 65°C to 95°C with 0.5°C increments every 5 secs, and a single peak was confirmed to ensure assay specificity. Standard curve descriptors including efficiency, slope, y-intercept and R^2 are reported.

2.3 Results

2.3.1 *intI*1 primer evaluation

2.3.1.1 Evaluation of primers for coverage

In total 64 different *intI*1 primer sets, including 4 TaqMan primer-probe sets were retrieved from the systematic review (Appendix Table A.1). In addition, the primer and probe set designed in this study were included in the analysis, resulting in 65 primers evaluated (Appendix Table A.1). Primers were initially aligned against the reference *Pseudomonas aeruginosa* plasmid pVS1 nucleotide sequence (M73819.1) (Figure 2.1) and renamed for ease of identification (Appendix Table A.1). Next, primers were aligned with SDB1 (full-length *intI*1 database) to ensure binding sites were present (in froward or reserve orientation) and that the expected amplicon size would be generated. From this, 10 primer sets were discarded, which included two sets (F61-R61 and F64-R64) that were not *intI*1 primers (Table 2.4). The F61-R61 primer set targeted the *aadA1a* aminoglycoside adenylyl transferases gene (Sandvang et al. 1997; Guerra et al. 2001), while the F64-R64 primer pair targeted the class two integron-integrase gene (*intI*2) (Gündoğdu et al. 2011). In addition, these primer sets (F61-R61 and F64-R64) aligned poorly to the reference *Pseudomonas aeruginosa* pVS1 *intI*1 nucleotide sequence (Figure 2.1) and had no hit (High WS) with complete length *intI*1 sequences within SDB1 (Table 2.4). This left 55 primer pairs.

Table 2.4: Listed of *intI*1 primer sets excluded from further analysis in this study

Primer Pair	Sequence (5'-3' Direction)	Position[a]	Position[b]	Mean W.S*	Expected amplicon size (bp)	Observed amplicon size (bp)	Comment(s)	Study reference
F11	CTTCAGCCTTTTCCAGCAAC	158	157	3.41	308	818	Large difference between observed and expected amplicon size coupled with a high mean weighted score for the primers	(Najafgholizadeh Pirzaman and Mojtahedi 2019)
R11	GAAACCTGCTCCAGCACTTC	956	42	6.42				
F12	TCATGGCTTGTTATGACTGT	973	972	7.22	600	-	Primer sequences same as those of primer set hep58 and hep59, commonly used to target cassette region of class 1 integron.	(Mobaraki et al. 2018)
R12	GTAGGGCTTATTATGCACGC	752	886	5.76				
F20	AGCTTACGAACCGAACAGGC	88	87	0.37	950	597	A large difference between the estimated and expected amplicon size	(Borruso et al. 2016)
R20	TCCGCCAGGATTGACTTGCG	666	665	3.69				
F24	CCCGAGGCATAGACTGTA	238	227	6.88	160	-	Wrong primer binding orientation even when the forward sequence is used as the reverse sequence and vice versa	(Koeleman et al. 2001)
R24	CAGTGGACATAAGCCTGTTC	100	99	0.15				
F27	CGAGGCATAGACTGTAC	876	182	2.81	925	703	A large difference between the observed and expected amplicon size	(Orman et al. 2002)
R27	TTCGAATGTCGTAACCGC	869	868	0.01				
F28	GTCAAGGTTCTGGACCAGTTG	40	39	**5.4**	550	892	A large difference between the estimated and expected amplicon size; the primer sequence same as F17-R17 (Rosser and Young 1999) with a 1bp difference in the forward sequence.	(Bashir et al. 2015)
R28	ATCATCGTCGTAGAGACGTCGG	911	910	**4.64**				
F30	AAAACCGCCACTGCGCCGTTA	4	3	0.19	1201	996	A large difference between the expected and observed amplicon size. Additionally, the expected amplicon size exceeds the size of a complete length *intI*1 gene (1014bp)	(Falbo et al. 1999; Fonseca et al. 2005)
R30	GAAGACGGCTGCACTGAACG	977	275	6.25				
F55	AAGCAGACTTGACCTGA	185	184	5.46	457	643	A large difference between the observed and expected amplicon size	(Kainuma et al. 2018)
R55	GGTGTGGCGGGCTTCGTG	810	809	0.09				
F61	GTGGATGGCGGCCTGAAGCC	587	586	4.41	-	141	Targets aminoglycoside adenylyl transferases (*aadA1a;* previously *ant(3")Ia*) gene (Sandvang et al. 1997; Guerra et al. 2001)	(Kennedy et al. 2018)
R61	ATTGCCCAGTCGGCAGCG	710	319	3.99				
F64	CACGGATATGCGACAAAAAG	584	935	5.02	160	271	Targets Class 2 integron-integrase (*intI*2) gene (Gündoğdu et al. 2011)	(Karami et al. 2020)
R64	GATGACAACGAGTGACGAAATG	833	832	4.99				

F, forward primer; R, reverse primer; W.S, weighted score. a, start position of primers on *intI*1 sequence based on alignment to the reference *Pseudomonas aeruginosa* plasmid pVS1 nucleotide sequence (Figure 2.1). b, start position of primers on *intI*1 sequence based on alignment with Primer Prospector to sequence within the complete length *intI*1 sub-database, SDB1 (*n=104*).

Of the remaining 55 primers, when aligned to SDB1 (complete length sub-database *n=104*) > 97%-100 of sequences had the correct primer binding orientation with the majority of these primers (49 sets; 89%) amplifying 69-100% of sequences in the correct primer binding orientation with 0 mismatches (Figures 2.4A, 2.4B, Appendix Table A.2). One primer (F40-R40) performed poorly, amplifying only 0.9% (*n=1* amplicon) sequences with a correct primer binding orientation at 0 mismatch (Figures 2.4A, 2.4B, Appendix Table A.2), and was removed from further consideration.

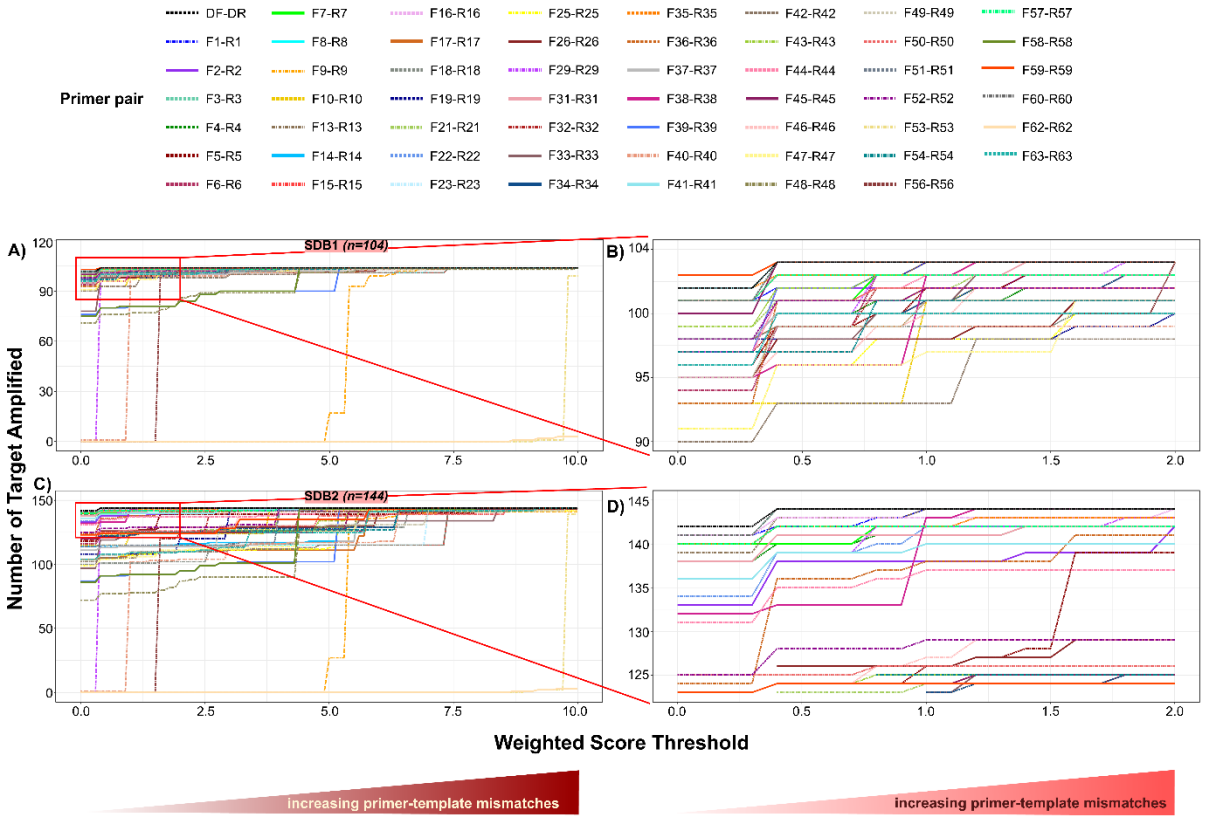

Figure 2.4: Performance of *intI*1 primer sets against *intI*1 nucleotide sequences of SDB1 (*n=104*) and SDB2 (*n=144*), to evaluate primer coverage**.** Primers were evaluated as pairs, for their ability to generate amplicon based on a defined weighted score (WS) threshold that varied from 0 (strict) to 10 (less stringent). A WS of 0 indicates a perfect match (0 mismatch) between the primer and template sequence. A WS >0 indicates mismatches between the primer and template sequence. The top-performing primers were defined as those primer sets that were able to generate the highest number of amplicons at 0 WS in the test sub-database. A) SDB1 and C) SBD2 WS plot for all evaluated primer sets based on WS threshold that varied from 0 to 10. The red rectangular box indicates the zoomed-in area of the WS plot for SDB1 B) and SBB2 D). Each line colour and line type represent a different set of primer.

Primer coverage was then tested against the other *intI*1 complete length and partial length sub-databases, SDB2 (*n=144*) (Figures 2.4C, 2.4D) and SDB3 (*n=503*) (Figure 2.5). Here, the number of sequences with the correct primer binding orientation declined (SDB2: 79-100%; SDB3: 41%-100%), as did the number of amplicons amplified with a 0 mismatch

(SDB2: 50-99%; SDB3: 16%- 99%) for the correct primer binding orientation sequences (Appendix Table A.2).

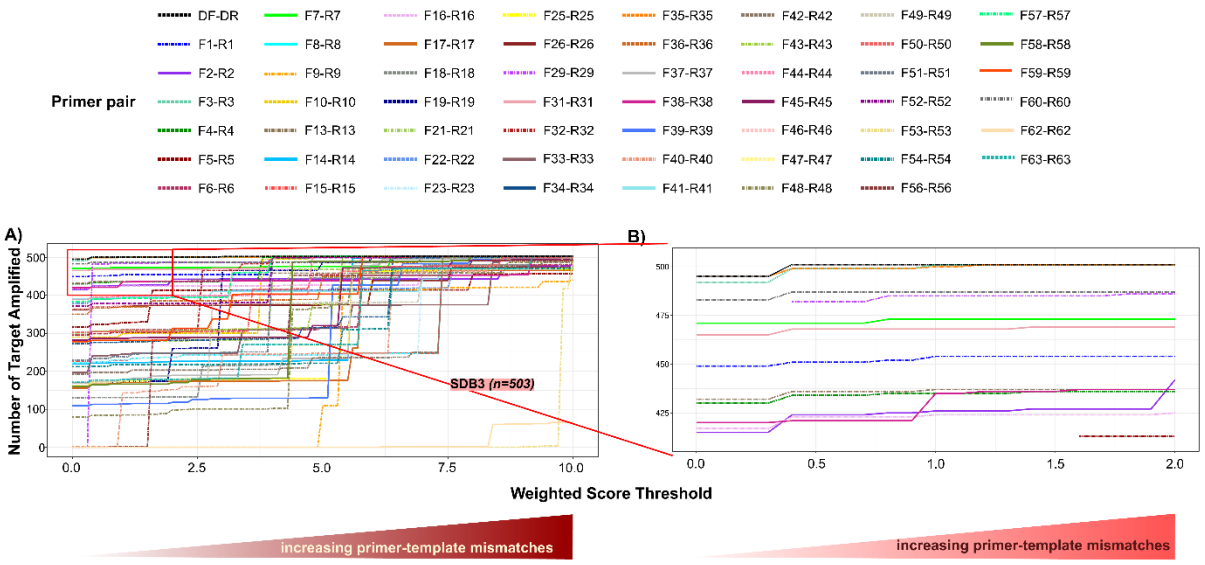

Figure 2.5: Performance of *intI*1 primer sets against *intI*1 nucleotide sequences of SDB3 (*n= 503*), to evaluate primer coverage. Primers were evaluated as pairs, for their ability to generate amplicon based on a defined weighted score (WS) threshold that varied from 0 (strict) to 10 (less stringent). A WS of 0 indicates a perfect match (0 mismatch) between the primer and template sequence. A WS >0 indicates mismatches between the primer and template sequence. The top-performing primers were defined as those primer sets that were able to generate the highest number of amplicons at 0 WS in the test sub-database. A) WS plot for all evaluated primer sets based on WS threshold that varied from 0 to 10. The red rectangular box indicates a zoomed-in area. B) WS plot representing a zoomed-in area of plot A. Each line colour and line type represent a different set of primer.

Five (9%) primer sets produced no amplicon at a WS of 0 across the three *intI*1 sub-databases (Figures 2.4, 2.5, Appendix Table A.2) and were removed from further consideration. This included one set (F29-R29) which performed optimally with the addition of a single mismatch (WS=0.4) (Table 2.5), but as there were several primers with better coverage at a WS of 0, this primer set was removed. In summary, a further 6 primer sets (F9-R9, F29-R29, F40-R40, F53-R53, F56-R56, F62-R62) were discarded from the primer coverage analysis.

Table 2.5: *intI*1 primer sets with no amplicon produced at 0 WS and the WS at which an amplicon was produced

Primer pair	SDB1 *(n=104)*		SDB2 *(n=144)*		SDB3 *(n=503)*	
	WS at which an amplicon was produced	No of amplicon generated	WS at which an amplicon was produced	No of amplicon generated	WS at which an amplicon was produced	No of amplicon generated
F9-R9	5	17	5	27	5	109
F29-R29	0.4	100	0.4	139	0.4	482
F53-R53	9.2	1	9	1	9	1
F56-R56	1.6	100	1.6	120	0.8	1
F62-R62	8.7	1	8.7	1	6.2	1

F=Forward primer; R=Reverse primer; SDB, sub-database; WS, Weighted score

Of the 49 amplicon producing primers at 0 WS retained, 10 primer sets (DF-DR, F1-R1, F3-R3, F7-R7, F13-R13, F16-R16, F31-R31, F35-R35, F57-R57 and F60-R60) consistently had a high number of sequences with the correct primer-binding orientation and amplified ≥97% amplicons within the complete length sub-databases (SDB1: *n=104* and SDB2 *n=144*). Moreover, these primers consistently had a low mean WS for the forward and reverse primer within each pair (Appendix Table A.2). As such, these 10 primers were considered the best-performing *intI*1 primer sets.

Five primer sets analysed in this study incorporated a TaqMan probe (Table 2.6), two of which (DF-DR and F7-R7 sets) were among the best-performing primer sets. Of these, the DF-P-DR primer-probe set, designed in this study, consistently produced the highest number of amplicons at 0 WS across the three *intI*1 sub-databases with 102 (98%), 142 (99%) and 494 (99%) of sequences amplified within SDB1, SDB2 and SDB3 respectively. In addition, allowing for a single non-3' mismatch (WS=0.4) between primer and probe, resulted in all sequences with the correct primer-binding orientation to be amplified across the three *intI*1 sub-databases (Table 2.6).

Table 2.6: Coverage of published and newly modified *intI*1 primer sets that incorporate a reporter probe

Primer pair and probe combination	SDB1 *(n= 104)*				SDB2 *(n= 144)*				SDB3 *(n= 503)*				*intI*1-like *(n=15)*			
	Number of sequences with correct primer orientation (%)	Weighted Score Threshold			Number of sequences with correct primer orientation (%)	Weighted Score Threshold			Number of sequences with correct primer orientation (%)	Weighted Score Threshold			Number of sequences with correct primer orientation (%)	Weighted Score Threshold		
		0 (0 mismatch)	0.4 (1 mismatch at 5' end)	1 (mismatch at 3' end/non-3' gaps/ 2 mismatches at 5' end)		0 (0 mismatch)	0.4 (1 mismatch at 5' end)	1 (mismatch at 3' end/non-3' gaps/ 2 mismatches at 5' end)		0 (0 mismatch)	0.4 (1 mismatch at 5' end)	1 (mismatch at 3' end/non-3' gaps/ 2 mismatches at 5' end)		0 (0 mismatch)	0.4 (1 mismatch at 5' end)	1 (mismatch at 3' end/non-3' gaps/ 2 mismatches at 5' end)
DF-P-DR	104 (100%)	102 (98%)	104 (100%)	104 (100%)	144 (100%)	142 (99%)	144 (100%)	144 (100%)	501 (100%)	494 (99%)	501 (100%)	501 (100%)	14 (93%)	9 (64%)	11 (79%)	12 (86%)
F7-P-R7	104 (100%)	92 (88%)	96 (92%)	102 (98%)	144 (100%)	131 (91%)	135 (94%)	141 (98%)	475 (94%)	454 (96%)	465 (98%)	471 (99%)	10 (67%)	8 (80%)	8 (80%)	8 (80%)
F10-P-R10	104 (100%)	91 (88%)	92 (88%)	99 (95%)	126 (88%)	113 (90%)	114 (90%)	121 (96%)	302 (60%)	288 (95%)	289 (96%)	297 (98%)	9 (60%)	6 (67%)	6 (67%)	6 (67%)
F38-P-R38	104 (100%)	92 (88%)	96 (92%)	103 (99%)	144 (100%)	127 (88%)	131 (91%)	143 (99%)	438 (87%)	415 (95%)	419 (96%)	435 (99%)	13 (87%)	9 (69%)	9 (69%)	9 (69%)
F46-P-R46	102 (98%)	34 (33%)	54 (53%)	95 (93%)	130 (90%)	41 (32%)	67 (52%)	121 (93%)	380 (76%)	88 (23%)	185 (49%)	360 (95%)	13 (87%)	3 (23%)	3 (23%)	8 (23%)

F, forward primer; R, reverse primer; P, probe; SDB, sub-database

Conversely, primer-probe set F46-P-R46 performed the worst, of the five primer-probe combinations assessed, with only 34 (33%), 41 (32%) and 88 (23%) of sequences with the correct primer-binding-orientation amplified at 0 WS within SDB1, SDB2 and SDB3 respectively. Nonetheless, allowing for an increased WS of 1 (i.e., mismatch caused by either a single 3' mismatch or two non-3' mismatches) resulted in a significant increase in the number of amplicons amplified across *intI*1 sub-databases (SDB1: *n=95* (93%), SDB2: *n=121* (93%), SDB3: *n=360* (95%)) (Table 2.6)

The F7-P-R7 primer-probe set (commonly used TaqMan assay in *intI*1 gene study), F10-P-R10 and F38-P-R38 primer-probe sets showed similar coverage to each other, but lower than DF-P-DR set, with 92 (88%), 91 (88%) and 92 (88%) of amplicons amplified at 0 WS within SDB1 respectively (Table 2.6). However, the F7-P-R7 primer-probe set amplified the highest number of amplicons at 0 WS (or second highest after the DF-DR set) for the correct primer binding sequences across the other two *intI*1 sub-databases (SDB2: *n=131* (91%), SDB3: *n=454* (96%)) among the three primer sets (Table 2.6). As such, the DF-P-DR and F7-P-R7 primer-probe sets were put forward as the top-performing primer-probe set.

2.3.1.2 Evaluation of primers for specificity

The primer sets were tested for specificity against the *intI*1-like (*n=15*) and non-*intI*1 (*n=1540*) sub-databases respectively (Figure 2.6, Appendix Table A.2). Here, the aim was for the primers to amplify the least amount of non-target sequence reflected by a higher forward and reverse primer WS for sequences where primers bind in the correct orientation. The 10 best-performing primer sets identified above were focused on.

For the best-performing primer sets, the number of sequences with correct primer-binding orientation ranged from 67-100% and 41-65% for the *intI*1-like and non-*intI*1 sub-databases respectively with 57-80% and 0% of these correct primer-binding orientation sequences amplified at 0 mismatch in *intI*1-like and non-*intI*1 sub-databases (Figures 2.6, Appendix Table A.2).

Of these best-performing sets, the F16-R16 primer set amplified the highest number of *intI*1-like amplicons (*n=11*, 79%) at 0 mismatches and was removed, while the primer pairs F57-R57 and F31-R31 amplified the lowest number of *intI*1-like sequence (*n=7*) (Figure 2.6A, Appendix Table A.2). The incorporation of a TaqMan reporter probe generally improved primer specificity, however, two of the primer sets which incorporated a TaqMan probe (DF-

P-DR and F7-P-R7) both amplified *intI*1-like sequence. The number of *intI*1-like amplicons amplified by the DF-P-DR (*n=9*) and F7-P-R7 (*n=8*) primer-probe sets at a 0 WS were similar. Of note, whilst the 10 best performing were focused on, the other primer sets analysed (see section 2.3.1.1) also amplified *intI*1-like sequences (Figure 2.6A, Appendix Table A.2).

Next, the nine remaining primer sets were tested against the non-*intI*1 sequences (Figures 2.6B, 2.6C). None of the primers amplified the non-*intI*1 sequence at a 0 mismatch (Appendix Table A.2). In general, primers only produced amplicons from the non-*intI*1 database with very high weighted scores (sum of forward and reverse primer mean WS ranged from 8.39-11.6) (Appendix Table A.2). However, the primer pairs F1-R1 (WS: 2) and F13-R13 (WS: 3.2) performed worst, having the lowest WS required to amplify at least one non-*intI*1 target. As such, were removed from further analysis, leaving seven sets (DF-DR, F3-R3, F7-R7, F31-R31, F35-R35, F57-R57 and F60-R60) to be considered the best overall performing *intI*1 primer sets in terms of coverage and specificity.

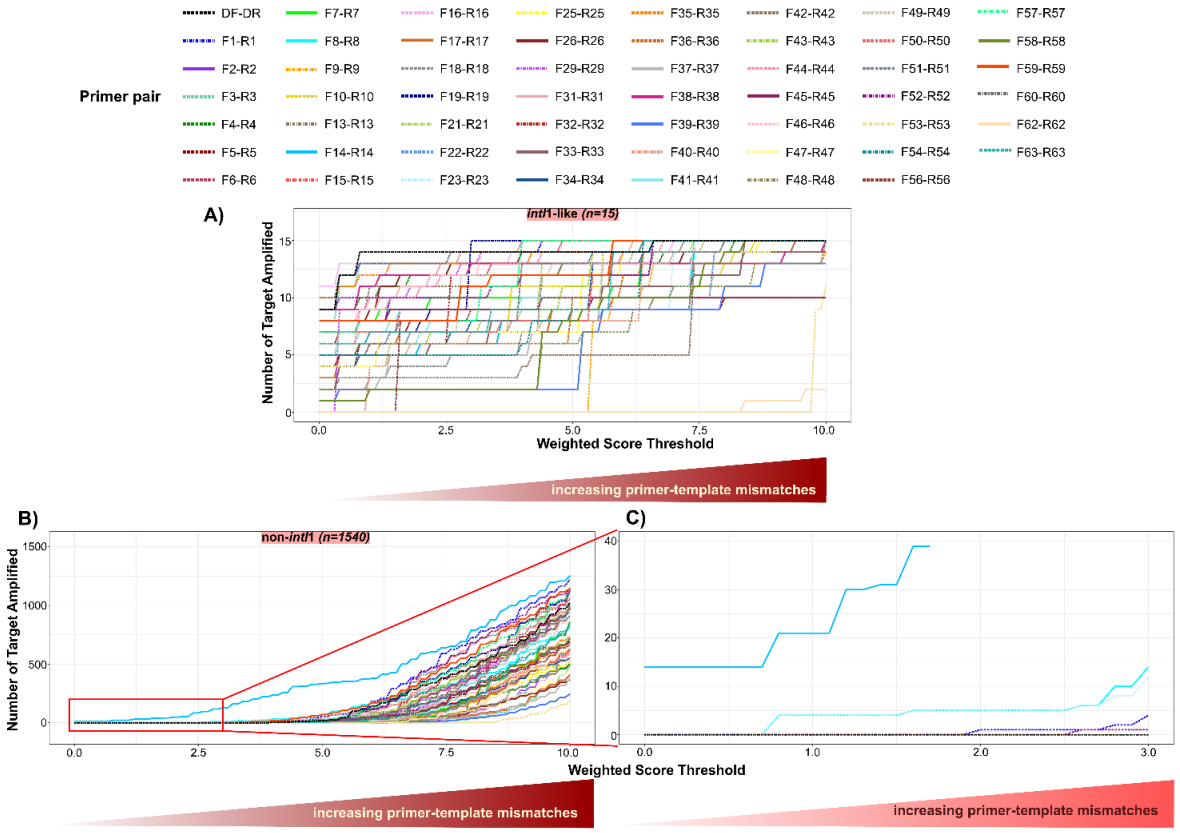

Figure 2.6: Performance of *intI*1 primer sets against *intI*1-like nucleotide sequences (*n= 15*) and non-*intI*1 nucleotide sequences (*n= 1540*), to evaluate primer specificity. Primers were evaluated as pairs, for their ability to generate amplicon based on a defined weighted score (WS) threshold that varied from 0 (strict) to 10 (less stringent). A WS of 0 indicates a perfect match (0 mismatch) between the primer and template sequence. A WS >0 indicates mismatches between the primer and template sequence. A) *intI*1-like and B) non-*intI*1-like WS plots for all evaluated primer sets based on WS threshold that varied from 0 to 10. The red rectangular box indicates a zoomed-in area of the WS plot for the non-*intI*1 sub-database C). Each line colour and line type represent a different set of primer.

2.3.1.3 Recommendation of optimal primer sets for *in situ* laboratory validation and *in-silico* amplicon size distribution

From the initial 65 primer sets, seven (DF-DR, F3-R3, F7-R7, F31-R31, F35-R35, F57-R57 and F60-R60) were identified that had high coverage in our *intI*1 database, but low-specificity to the non-*intI*1 database, indicating there are good primer sets targeting a broad range of *intI*1 targets, while discriminating against non-*intI*1 sequences. Two published sets (F3-R3 and F7-R7) were selected (Appendix Table A.2), as they not only had the highest WS required to amplify the non-*intI*1 target (i.e., needed the highest number of mismatches to target the sequence) but also had short amplicons (100-200bp), making them ideal for both qPCR and high-throughput amplicon sequencing. In addition, each of these selected primer sets targeted a different region of the *intI*1 gene and was commonly used within the

literature. F7-R7 incorporated a TaqMan probe. The primer and probe set, DF-P-DR, designed in this study, was also included resulting in three *intI*1 primer sets selected for laboratory validation.

2.3.2 Application of selected *intI*1 primers on septic wastewater samples

2.3.2.1 Q-PCR quantification of *intI*1 gene from Thai Septic Tanks wastewater

The three selected and optimised *intI*1 primer sets and probes (DF-DR, F3-R3 and F7-R7; Table 2.7) were used to quantify *intI*1 gene abundance across 30 septic tank wastewater samples (influent, sludge, effluent) from CST-household, CST-healthcare and SST-household reactors (Table 2.2). Each of the standard curves from all three primer sets had high efficiencies which ranged from 91.29 to 95.7%, y-intercepts of 35.71 to 39.6, slope of -3.43 to -3.55 and a No Template Control Ct from undetected to 36.9 (Table 2.7).

Table 2.7: Primers and probe sets selected and optimised for Q-PCR to quantify intI1 gene copies from wastewater

Primer ID	Sequence (5'- 3')	Orientation	Target (length)	Assay type	Q-PCR reaction	Experimental cycle condition	Q-PCR Standard Curv Descriptors					Reference
							Efficiency (%)	R^2	Slope	Intercept	NTC	
DF	TTCTGGAAGGCGAGCATC	Forward	*intI*1 (108bp)	MGB TaqMan probe	10µl 2x iTaq Universal Probes Supermix (Bio-Rad); 0.8µl of each primer (10µM); 0.4µl probe (10µM); 1.8µl MgCl2 (final concentration= 3µM); 4.2µl nuclease-free water	**PCR:** 95°-15min; [94°-30sec; 61°-30sec; 72°-30sec] x35; 72°-10min **Q-PCR:** 95°-10min; [95°-30sec; 60°-60sec, plate read]x45 **MiSeq Sequencing 1st step PCR:** 95°-15min; [94°-30sec; 61°-30sec; 72°-30sec]x25;72°-10min **2nd-step PCR**: 95°-15min; [95°-30sec, 55°-30sec, 72°-30sec) ×8; 72°-5min	95.7	0.999	-3.43	37.71	36.9	This study
DR	TGCCGTGATCGAAATCC	Reverse										
DF-P	Fam-TGACCCGCAGTTGCA-MGB Eclipse	Probe										
F3	TTTCTGGAAGGCGAGCATCGTTTG	Forward	*intI*1 (109bp)	SYBR green	10µl 2x QuantiTect SYBR Green PCR Master Mix (Qiagen); 1µl of each primer (10µM); 0.4µl MgCl2 (final concentration= 3µM); 5.6µl nuclease-free water	**PCR:** 95°-15min; [94°C-30sec; 60°-30sec; 72°-30sec] x35; 72°-10min **Q-PCR:** 95°-15min; [94°-15sec; 65°-30sec; 72°-30sec, plate read] x40; Melt curve: 65°-95° (0.5° increment/5sec) **MiSeq Sequencing 1st step PCR**: 95°-15min; [94°-30sec; 61°-30sec; 72°-30sec] x25; 72°-10min **2nd step PCR:** 95°-15min; [95°-30sec, 55°-30sec, 72°-30sec) ×8; 72°-5min	91.64	1	-3.54	35.71	36.2	(Rosewarne et al., 2010)
R3	TGCCGTGATCGAAATCCAGATCCT	Reverse										
F7	GCCTTGATGTTACCCGAGAG	Forward	*intI*1 (196bp)	TaqMan probe	10µl 2x iTaq Universal Probes Supermix (Bio-Rad); 0.8µl of each primer (10µM); 0.4µl probe (10µM); 1.8µl MgCl2 (final concentration= 3µM); 4.2µl nuclease-free water	**PCR:** 95°-15min; [94°-30sec; 60°-30sec; 72°-30sec] x35; 72°-10min **Q-PCR :** 95°-10min; [95°-30sec; 60°-60sec, plate read]x45 **MiSeq Sequencing 1st step PCR**: 95°-15min; [94°-30sec; 60°-30sec; 72°-30sec] x25; 72°C-10min **2nd-step PCR:** 95°-15min; [95°-30sec, 55°-30sec, 72°-30sec) ×8; 72°-5min	91.29	0.997	-3.55	39.6	0	(Barraud et al., 2010)
R7	GATCGGTCGAATGCGTGT	Reverse										
F7-P	6Fam-ATTCCTGGCCGTGGTTCTGGGTTTT-BHQ1	Probe										

Between primers sets, there was no significant difference in *intI*1 gene abundance for the same sample type (influent, sludge, effluent) (p-value >0.05, Figure 2.7, Appendix Table A.3), and Spearman rank correlation coefficient analysis indicated that the *intI*1 gene copy number per ml DNA (influent and effluent) or per g DNA (sludge) amplified by each of the primers were highly correlated (r=0.989 (p-value <0.001), r=0.993 (p-value <0.001), and r=0.994 (p-value <0.001) for DF-DR and F7-R7, DF-DR and F3-R3, and F3-R3 and F7-R7 primer sets respectively). As such, each primer set resulted in the same overall pattern of *intI*1 gene abundance, with higher *intI*1 gene copies observed in the sludge (DF-DR: $1.90x10^8$ ±SD$9.98x10^7$; F3-R3: $1.69x10^8$±SD$8.64x10^7$; F7-R7: $1.85x10^8$±SD$1.07x10^8$ copies/g DNA) than in the effluent (DF-DR: $4.94x10^6$±SD$4.87x10^6$; F3-R3: $4.34x10^6$±SD$4.20x10^6$; F7-R7: $5.35x10^6$±SD$5.39x10^6$ copies/ml DNA) and influent (DF-DR: $4.20x10^6$±SD$4.58x10^6$; F3-R3: $3.69x10^6$± SD $3.93x10^6$; F7-R7: $5.03x10^6$±SD $5.39x10^6$ copies/ml DNA) (Figure 2.7). The distribution of CL1-integron among the sample types (influent, sludge, effluent) could have been affected by factors including the septic tank volume, number of users and frequency of septic tank usage, and the duration of wastewater effluent or sludge retention prior to discharge.

For the *16S rRNA* gene, higher bacteria biomass was found in the sludge ($4.99x10^9$±SD$3.91x10^9$ copies/g DNA) rather than in the influent ($2.71x10^7$±SD$2.03x10^7$ copies/g DNA) or effluent ($2.48x10^7$±SD$1.93x10^7$ copies/g DNA) (Figure 2.7D).

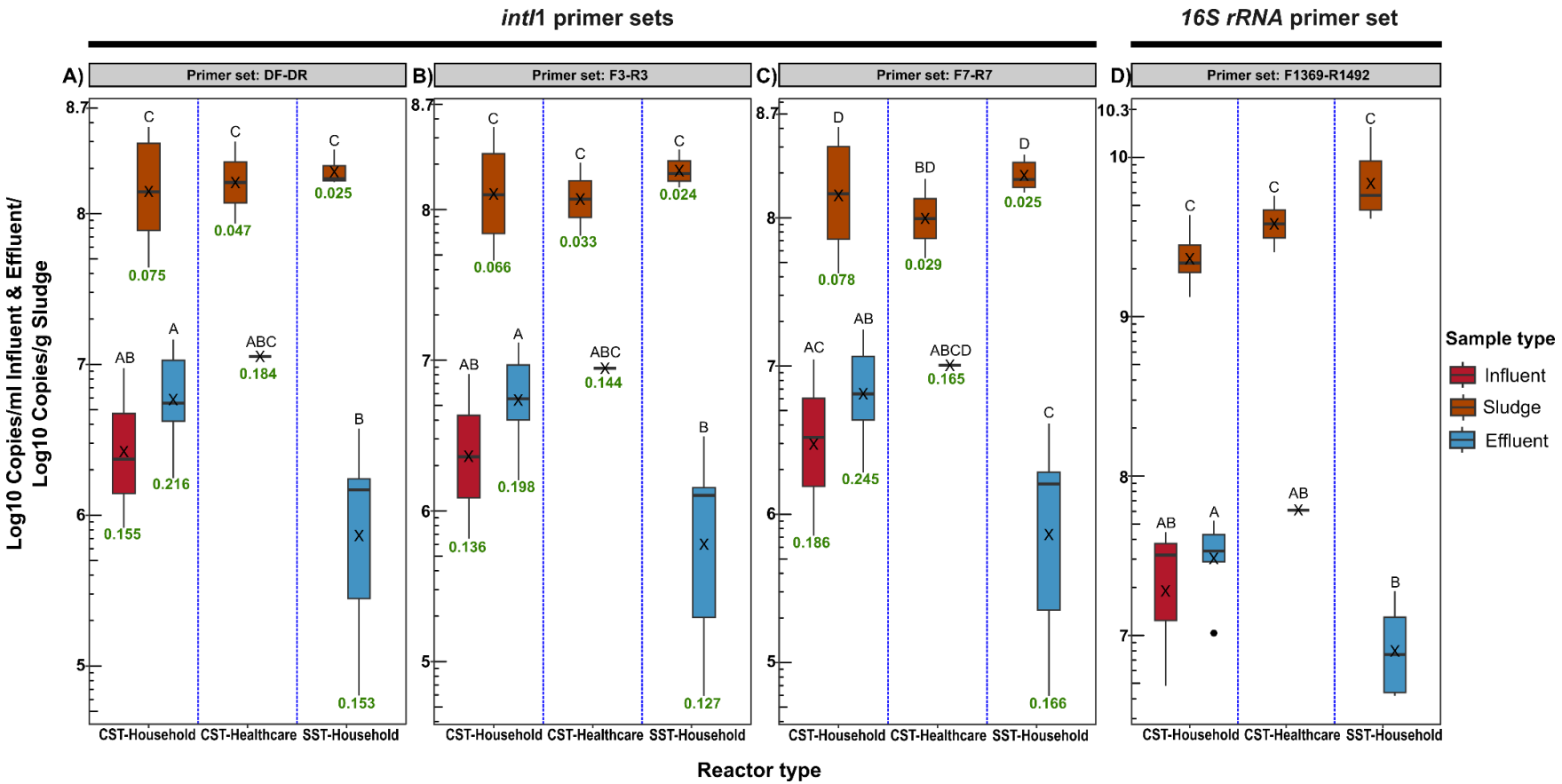

Figure 2.7: Impact of primer choice on *intI*1 gene copies quantified from CST-Household, CST-Healthcare and SST-Household septic tank wastewater reactors, and three wastewater sample types (influent, effluent, sludge). Results of the two-way ANOVA analysis showing statistically significant differences in *intI*1 gene copies quantified between reactor types and sample types for the DF-DR (A), F3-R3 (B) and F7-R7 (C) primer set. D) *16S rRNA* gene copies quantified between reactor types and sample types. For each primer set, boxplot sharing the same letter indicates no statistically significant difference at p-value >0.05, while boxplot with different letters indicates a statistically significant difference at p-value <0.05. A statistically significant difference in *intI*1 gene abundance between primer sets for the same sample was not observed (p-value >0.05; see Appendix Table A.3). X icon indicates mean *intI*1 copies. The black dot indicates data outlier. Numbers written in green below each boxplot indicate the ratio of *intI*1: *16S rRNA* copies.

Furthermore, although a similar overall pattern of *intI*1 gene abundance was observed, the F7-R7 primer set was the only primer set that reported no statistical difference (p-value > 0.05) in gene abundance between the CST-household effluent and CST-healthcare sludge samples (Figure 2.7C). Additionally, both DF-DR and F3-R3 primer sets showed very similar patterns of the *intI*1 gene copies across sample types and tank types. However, the DF-DR primer consistently reported higher *intI*1 to *16S rRNA* ratios (Figure 2.7A) than the F3-R3 set (Figure 2.7B).

SST-household units incorporated an internal pasteurisation effect and were therefore expected to have lower *intI*1 gene abundance in the effluent. *intI*1 gene abundance per ml of DNA was lower in effluent than in both the CST-household and CST-healthcare tanks for all three primer sets (Figure 2.7A-C). However, these differences were only statistically significant (p-value < 0.05) between the SST household and CST household effluent (Figure 2.7A-C). Nonetheless, the lower *intI*1 gene abundance quantified in the solar septic tank (SST-household) effluent suggests a potential role of temperature in reducing CL1-integron gene abundance from wastewater.

To assess the daily environmental impact of the reactors in releasing CL1-integrons to the environment, the number of *intI*1 gene copies per litre entering the environment through the effluent of the three tank types was calculated based on their respective flow rates (CST: 72.1±SD32.1, SST: 93.7±SD48.1 L/day):

- CST-household- DF-DR: $2.31x10^5$, F3-R3: $2.60x10^5$, F7-R7: $2.44x10^5$ copies/L per day
- CST-healthcare- DF-DR: $3.65x10^5$, F3-R3: $4.17x10^5$, F7-R7: $2.95x10^5$ copies/L per day
- SST-household- DF-DR: $4.52x10^4$, F3-R3: $5.85x10^4$, F7-R7: $5.03x10^4$ copies/L per day

Regardless of the selected primer set, the CST-healthcare reactor contributed to the highest daily release of CL1-integron copies/L into the environment, while the SST-household scale reactor contributed the lowest contribution.

Although *intI*1 gene copies per g of sludge in the reactors (CST household, CST healthcare, SST household) was higher than *intI*1 gene copies per ml of influent and effluent, irrespective of the primer set employed, the ratio of *intI*1 to *16S rRNA* gene in the sludge

remained consistently lower compared to that in the influent and effluent across reactors. The SST household sludge had the lowest *intI*1 to *16S rRNA* gene ratio compared to the CST household and CST healthcare sludge samples (Figure 2.7A-C). Moreover, even with the lower *intI*1 to *16S rRNA* gene ratio per g of sludge, the presence of CL1-integron per bacteria genome in the sludge remained high regardless of the primer set (Figure 2.7).

The primer sets F3-R3 (Figure 2.7B) and F7-R7 (Figure 2.7C) indicated the SST-household sludge (F3-R3: 1. $85x10^8$±SD$4.26x10^7$, F7-R7: $1.98x10^8$±SD$5.04x10^7$ copies/g DNA) as the higher contributor of *intI*1 gene copies to the environment, especially when directly released without additional treatment, and the CST-healthcare sludge (F3-R3: 1. $36x10^8$±SD$9.75x10^7$, F7-R7: $1.18x10^8$±SD$9.14x10^7$ copies/g DNA) to be the lowest contributor of the three reactors. Meanwhile, the DF-DR primer set (Figure 2.7A) showed the CST-healthcare sludge to be the higher contributor ($1.93x10^8$±SD$1.51x10^8$ copies/g DNA) of CL1-integron to the environment but reported the CST-household ($1.87x10^8$±SD$1.40x10^8$ copies/g DNA) as the least contributor. As such, the SST-household in general had the highest *intI*1 gene abundances in sludge when primer sets F3-R3 and F7-R7 were used, but not when the DF-DR primer set was used (Figure 2.7). This again showed how the reactor with the higher abundance and thus likely to contribute most to the environment changed depending on the primer set used.

Influent samples were only accessible from the CST units with *intI*1 gene abundance higher in the **effluent** (DF-DR: $4.25x10^6$±SD$2.56x10^6$; F3-R3: $4.37x10^6$±SD$2.98x10^6$; F7-R7: $5.14x10^6$±SD$3.40x10^6$; copies/ml DNA) than **influent** (DF-DR: $4.20x10^6$±SD$4.58x10^6$; F3-R3: $3.69x10^6$±SD$3.93x10^6$; F7-R7: $5.03x10^6$±SD$5.39x10^6$ copies/ml DNA), indicating an increase ([mean influent- mean effluent/ mean influent]*100) of 1.20%, 18.45%, 2.18%, for the DF-DR, F3-R3 and F7-R7 primer sets respectively (Figure 2.7).

In summary, primer sets used did not change the overall pattern of *intI*1 gene abundances nor did it result in statistical difference (p-value >0.05) in *intI*1 gene abundance for the same sample type (influent, sludge, effluent) quantified with the different primers. However, comparing samples within the same primer set did sometimes result in statistical differences between samples, which could alter the interpretation of the risk of *intI*1 gene abundances, and in turn AMR pollution to the environment.

2.3.3 MiSeq amplicon sequencing

MiSeq amplicon sequencing was undertaken on all septic tank samples (*n=31*; Table 2.2), to confirm the specificity of the selected *intI*1 primers and assess the diversity of the short *intI*1 amplicons retrieved from the septic tanks (Table 2.7). Overall, the number of unique ASVs generated by each primer set was low, with 4 ASVs for DF-DR; 4 ASVs for F3-R3 and 11 ASVs for F7-R7 primer set. One ASV for the DF-DR primer set was removed due to cross-contamination with the F3-R3 set, owing to the high similarity of both primer sets and similar target regions. This excluded ASV from the DF-DR primer set had a maximum abundance of 8 ASVs across 3 samples (total of 20 ASVs removed). In summary, 3, 4 and 11 ASVs were generated for DF-DR, F3-R3 and F7-R7 primer sets respectively. The summary statistics of the sequencing output are provided in Table 2.8.

Table 2.8: Summary statistics of the ASVs abundances per sample by MiSeq amplicon sequencing

Primer set	No of ASVs	ASV abundance summary statistics					
		Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
DF-DR	3	37801	47492	48890	51945	54056	113204
F3-R3	4	40103	43406	45630	46602	48959	57196
F7-R7	11	27723	34570	36653	36684	39880	45592

A phylogenetic tree was constructed with the highly conserved *intI*1 sequences from a range of environmental and clinical samples. In addition, *intI*1-like sequences were added to the tree to determine if the primer sets could distinguish between these and *intI*1 sequences. All *intI*1 and *intI*1-like sequences, including the ASVs from our samples, showed high sequence similarity to each other, likely owing to the short amplicon region designed over conserved regions (Figures 2.8, 2.9, 2.10). *intI*1-like sequences clustered among the *intI*1 and our ASVs for all three primer sets, indicating that the primer sets could not differentiate between both variants.

All three primer sets amplified an abundant ASV-1 phylotype as the dominant *intI*1 sequence present in all septic tank types and samples. It was highly similar to *intI*1 and *intI*1-like sequences found in a range of environmental samples including freshwater biofilm, tannery effluent, hospital sewage and activated sludge (Figures 2.8, 2.9, 2.10).

For primer set DF-DR, a single cluster of *intI*1 sequences was present, albeit not supported by a bootstrap value, which also contained a second ASV (ASV-3) only present in the CST-household effluent (CST-P3_08-19). A third ASV (ASV-2), again only detected in the CST-household effluent (CST-P3_08-19), clustered outside the main group, highly similar to the *intI*1 sequence from activated sludge, although again not supported by a bootstrap value (Figure 2.8).

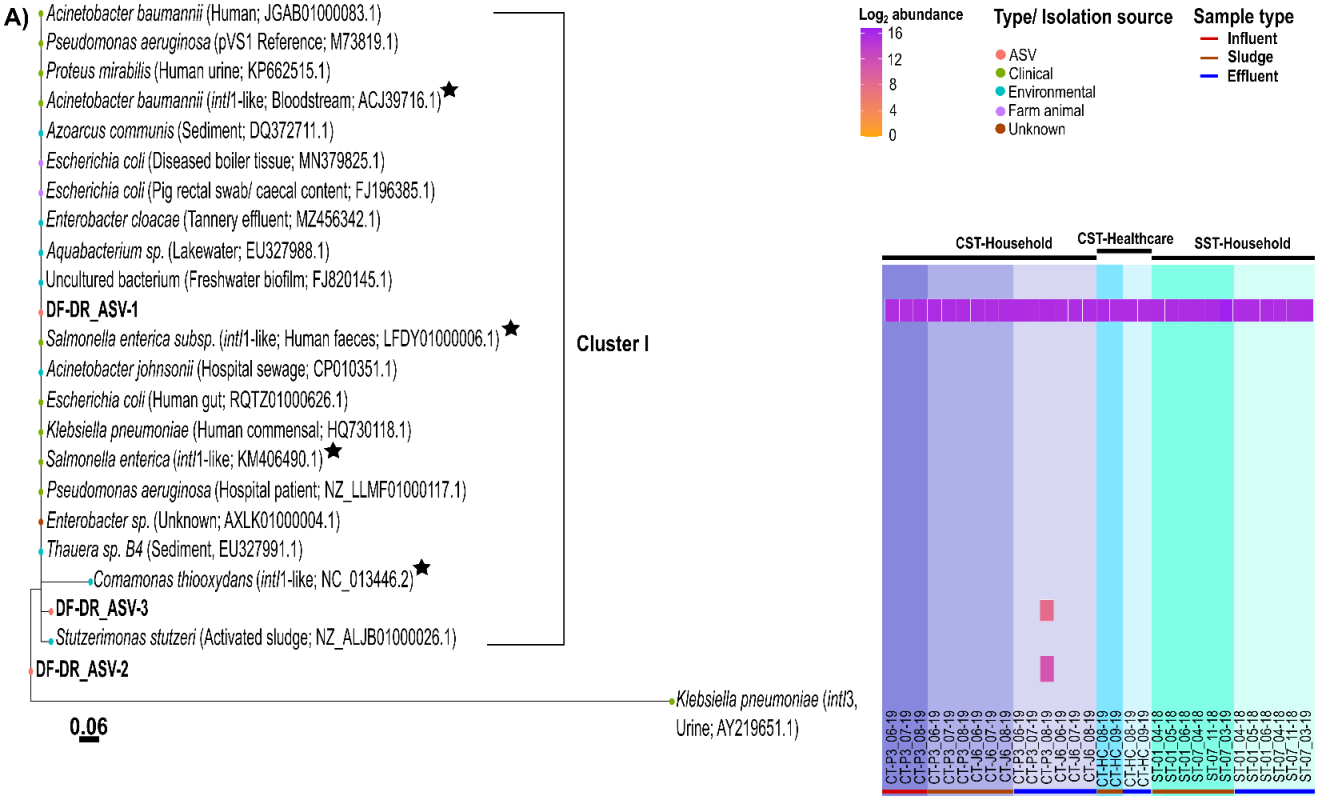

Figure 2.8: Detected ASVs abundance in Thai septic tank wastewaters (SST-Household, CST-Healthcare and SST-Household) by the DF-DR *intI*1 primer sets. Generated ASVs coupled with known and unknown *intI*1 within SDB1 (*n=104*), best hit NCBI sequences, and *intI*1-like sequences were aligned with Mafft, trimmed to only aligned region with no gaps, and phylogenetic tree constructed using the RAxML with 1000 bootstrap permutations. The number at the node represents a bootstrap value > 50% (from 1000 permutations). Bootstrap values at node <50 are not shown. The class 3 integron-integrase (*intI*3) gene (nucleotide ID: AY219651.1), which on protein level, shared a 60.74% similarity to the pVS1 protein sequence (AAA25857.1) was used as the outgroup. The colour of the tree tips indicates isolation source of sequence/ ASVs generated by the primer set. Heatmap shows log2 fold abundance (mean number of ASVs-DF-DR:5.1955x10^4; Table 2.8) of detected ASVs within each wastewater sample. CTP3 and CTJ6 samples originated from two independent CST-Household reactors. CT-HC sample was from a CST-Healthcare tank. ST01 and ST07 are two independent SST-Household units. The sampling month and year are indicated by the format month_year (i.e., 06_19= June 2019). CST, Conventional septic tank; SST, Solar septic tank.

For primer set F3-R3, a single cluster was observed, supported by a 95% bootstrap value containing ASV1, 3 and 4. These clustered with unknown and known *intI*1 from Tannery effluent and activated sludge as well as known *intI*1-like sequences from clinical origin (Figure 2.9). While ASV-1 was present in all samples, ASV3 and 4 were only detected in the CST-household effluent (CST-P3_08-19). Outside of this cluster was ASV-2, highly similar to *intI*1 from *Acinetobacter baumannii*, a clinical pathogenic bacterium. It was present in both the CST-household and SST-household tanks sludge and effluent but only in one CST-healthcare sludge sample (CT-HC_09-19) **(Figure 2.9)**.

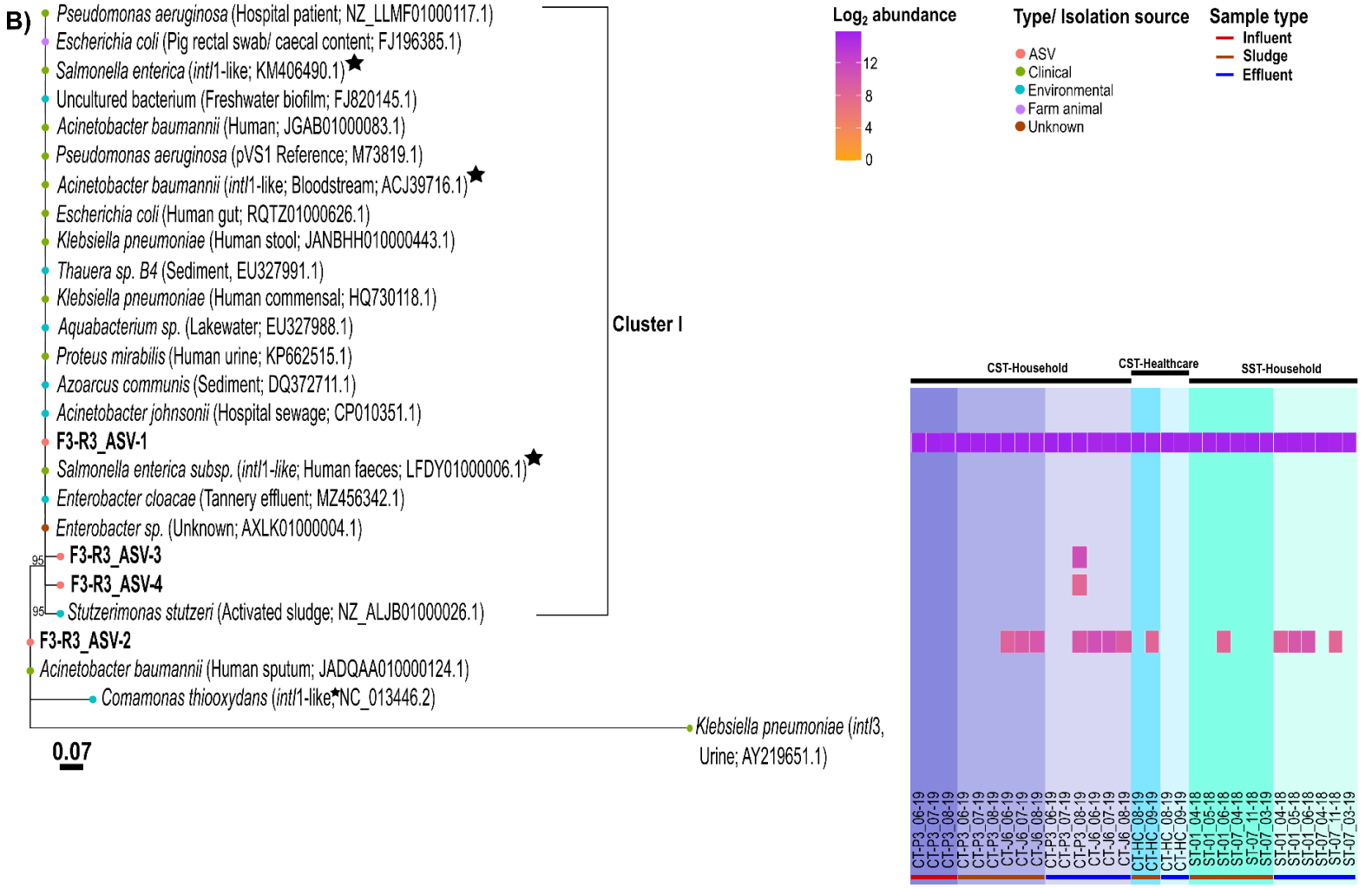

Figure 2.9: Detected ASVs abundance in Thai septic tank wastewaters (SST-Household, CST-Healthcare and SST-Household) by the F3-R3 *intI*1 primer sets. All other descriptions are the same as in Figure 2.8.

As primer sets DF-DR and F3-R3 targeted the same region of the *intI*1 (Figure 2.1), the ASVs generated by each primer set (DF-DR and F3-R3- ASV1; DF-DR-ASV2 and F3-R3-ASV3; DF-DR-ASV3 and F3-R3-ASV4) had 100% sequence similarity to each other but only ASV1 from each primer set showed a 100% sequence similarity when aligned against the full-length *intI*1 nucleotide sequences (pVS1, M73819.1). In addition, F3-R3-AV2 did not align to the full-length *intI*1 with a 100% similarity.

Finally, primer set F7-R7 which targeted the mid to downstream region (position 529-724 on pVS1 reference sequence, Figure 2.1), detected 11 ASVs within two clusters. Within cluster I, ASV-1 present in all samples was highly similar to ASV 8, 4, 7, 11, 10, 5 and 6 detected in CST-household (sludge), CST-healthcare (sludge and effluent) and SST-household (sludge and effluent) reactors. It clustered with known *intI*1 from sources such as hospital sewage and Tar-Pond, but also *intI*1-like sequences. Clustering was not supported by a high bootstrap value. Within cluster II, ASV-9 and 2 were detected in CST-household influent, CST-healthcare sludge and effluent and SST-household effluent samples and clustered with unknown and known *intI*1 sequence, as well as *intI*1-like sequence, found in environmental sources such as sediment and wastewater biofilm. However, clustering was not supported by a high bootstrap value (<50%) (Figure 3). A final ASV (ASV-3), again only detected in the CST-household effluent (CST-P3_08-19), clustered outside the main group, but was not supported by a high bootstrap value (<50%) (Figure 2.10).

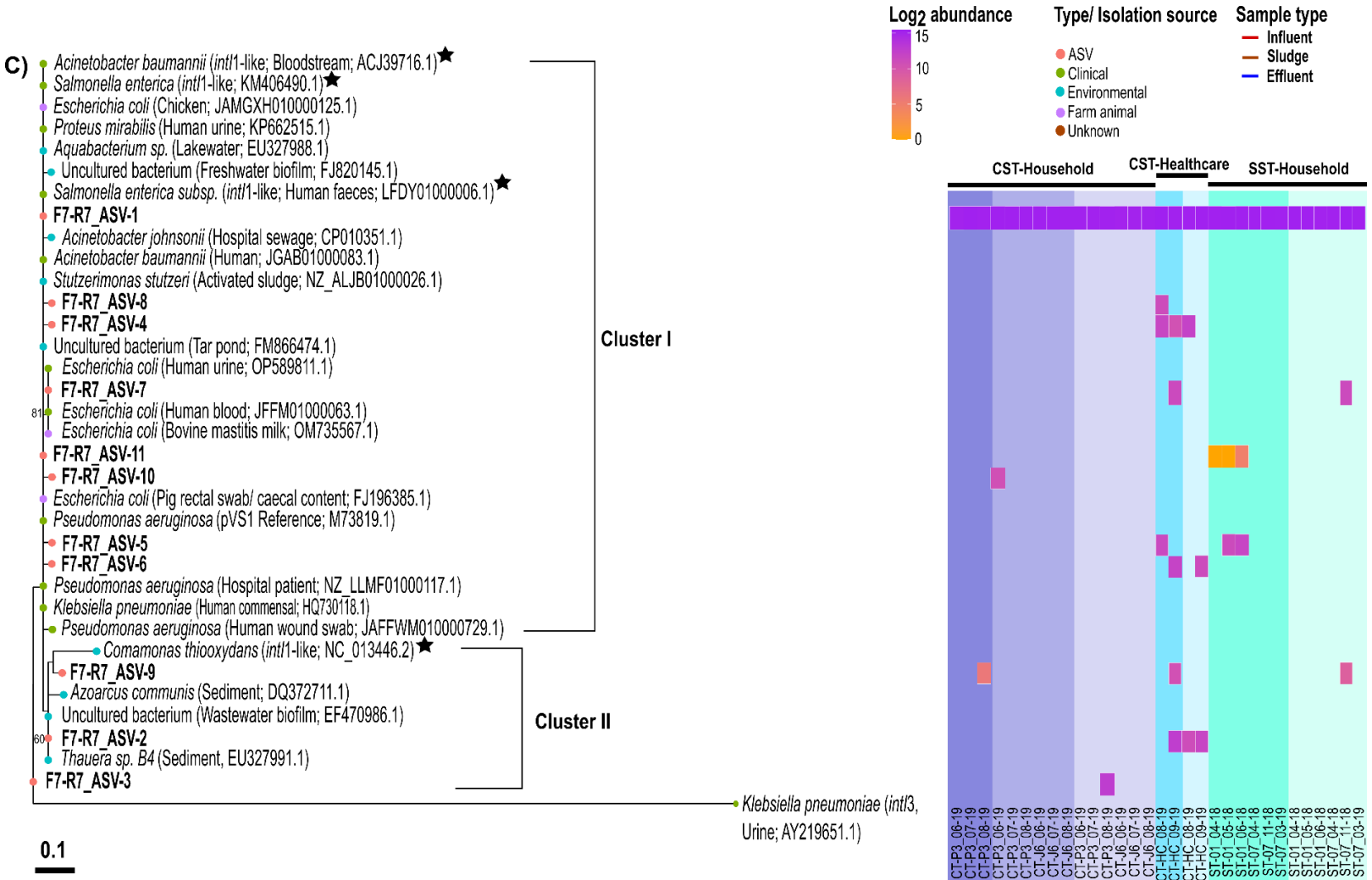

Figure 2.10: Detected ASVs abundance in Thai septic tank wastewaters (SST-Household, CST-Healthcare and SST-Household) by the F7-R7 *intI*1 primer sets. All other descriptions are the same as in Figure 2.8.

In summary, *intI*1 recovered diversity showed all samples to be dominated by a single ASV-1. It was highly similar to *intI*1 from clinical and environmental samples, however, *intI*1-like samples also clustered with it. CST-Household was the most diverse with primer set DF-DR and F3-R3, but a different picture arose with the F7-R7 primer set with the CST-Healthcare effluent having the highest *intI*1 diversity.

2.3.5 Laboratory Validation of selected *intI*1 primers to quantify *intI*1 mRNA transcript from environmental samples

As the detection of *intI*1 DNA does not infer integrase activity, each of the validated primer sets was tested for their ability to quantify *intI*1 mRNA transcripts. For this, fresh river water samples were used as the quality of RNA extracted wastewater nucleic acids may be of poor quality due to long-term storage (Cholet et al. 2019), although the quality was not measured. For each primer set, the reverse transcriptase reaction was carried out with random hexamers (RH) and gene-specific (GS) primers as previous work showed increased specificity with gene-specific priming (Cholet et al. 2020). In addition, TaqMan assays were carried out with (Figure 2.11C) and without the probes (i.e., SYBR Green) (Figure 2.11A, 2.11B, Table 2.9).

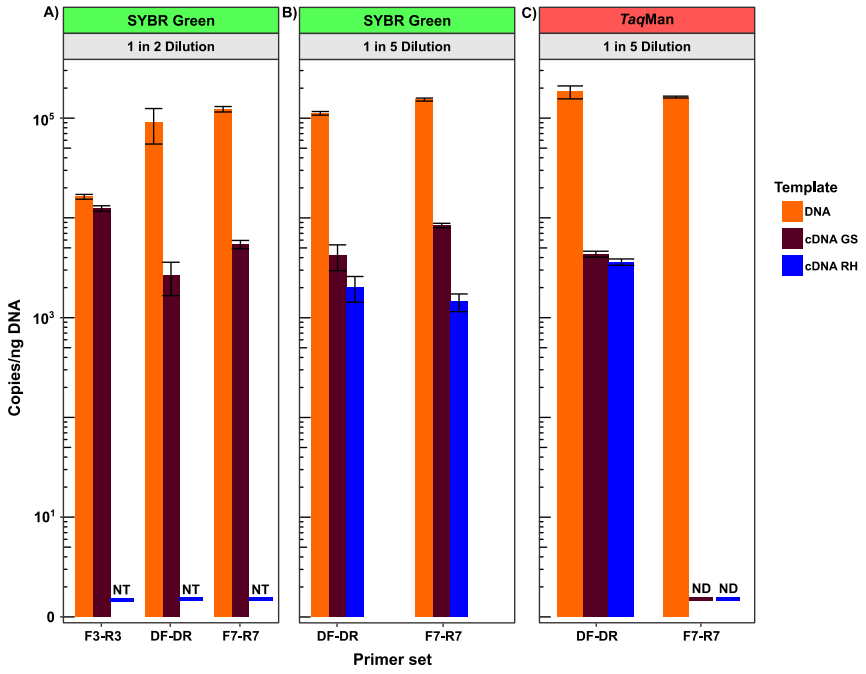

Figure 2.11: *intI*1 DNA and mRNA transcript quantified from river water sample by the DF-DR, F3-R3, F7-R7 *intI*1 primer sets. Reverse transcriptase reaction for each primer set was performed with random hexamers (RH) and gene-specific (GS) primers. Additionally, TaqMan assays were carried out with (C) and without the probes (i.e., SYBR Green) (A, B). NT denotes not-tested and ND denotes non-detected. Error bars show standard deviation.

Table 2.9: *intI*1 mRNA transcripts copies/ng DNA

Primer set	Assay	Priming strategy	Mean gene copy number (*n= 3)*				Q-PCR Standard Curve Descriptors				
			1 in 2 Dilution		1 in 5 Dilution						
			cDNA	DNA	cDNA	DNA	Efficiency (%)	R^2	Slope	Intercept	NTC
F3-R3	SYBR Green	GS	$1.24x10^4 \pm 7.83x10^2$	$1.63x10^4 \pm 9.15x10^3$	N.T	N.T	92.4	1	-3.52	35.84	0
DF-DR	SYBR Green	GS	$2.62x10^3 \pm 9.63x10^2$	$8.99x10^4 \pm 3.49x10^4$	$4.17x10^3 \pm 1.2x10^3$	$1.12x10^5 \pm 4.6x10^3$	94.3	0.997	-3.47	35.25	0
		RH	N.T		$2.01x10^3 \pm 5.8x10^2$						
	TaqMan	GS	N.T	N.T	$4.33x10^3 \pm 2.98x10^2$	$1.83x10^5 \pm 2.72x10^4$	94.9	0.999	-3.45	37.49	0
		RH	N.T		$3.61x10^3 \pm 2.67x10^2$						
F7-R7	SYBR Green	GS	$5.43x10^3 \pm 5.12x10^2$	$1.23x10^5 \pm 7.79x10^3$	$8.41x10^3 \pm 4.35x10^2$	$1.54x10^5 \pm 5.25x10^3$	90	0.999	-3.59	35.35	41.31
		RH	N.T		$1.44x10^3 \pm 2.89x10^2$						
	TaqMan	GS	N.T	N.T	ND	$1.63x10^5 \pm 3.66x10^3$	93	0.999	-3.5	37.41	37.64
		RH	N.T	N.T	ND						

All primer sets successfully quantified *intI*1 DNA and mRNA from river water, with *intI*1 gene abundances greater than *intI*1 transcripts (Figure 2.11). As previously shown (Cholet et al., 2020), gene-specific priming was more efficient than random hexamer priming. The F7-R7 primer set did not work as a TaqMan probe assay but worked in a SYBR green assay. It should be noted that, whilst higher *intI*1 transcript copies per ng DNA were quantified by the F3-R3 SYBR Green assay (1 in 2 dilutions), direct comparison to DF-DR and F7-R7 cannot be made as they were not done on the same sample. The aim here was simply to demonstrate that the primer sets were able to quantify *intI*1 mRNA transcripts. In summary, the primer sets tested are appropriate to quantify *intI*1 mRNA transcripts from environmental samples, using both gene-specific and random hexamer priming, albeit that the TaqMan probe chemistry must be swapped to SYBR green chemistry if the F7-R7 primer set is to be used.

2.4 Discussion

Accurate quantitative data is key to inform evidence-based management strategies and policies to reduce the global AMR burden. Quantitative approaches, alongside unified methodologies to enable comparison among data sets is a powerful tool to enable this. The clinical class 1 integron (CL1-integron) integrase gene (*intI*1*)* has been proposed as a proxy for inferring potential AMR The first step to investigating the potential for this is to select appropriate primers, however, our systematic literature review revealed over 65 *intI*1 primer sets with little consensus on the best primer to use. Through *in-silico* testing of the published primer sets, in addition to the design of an optimised primer set in this study, we selected three *intI*1 primer sets for laboratory validation and further testing of their specificity on septic tanks from Thailand associated with healthcare and household usage to investigate their contribution in disseminating CL1-integrons to the environment. This included a novel solar septic tank designed with internal heating ranging from 39 to 63.6°C in the disinfection chamber. From the 65 primers in the literature, three were selected- two published primer sets, F3-R3 (Rosewarne et al. 2010) and F7-R7 (Barraud et al. 2010) which have been extensively applied to survey CL1-integron abundance in a range of ecological settings including WWT (Stalder et al. 2014; Paiva et al. 2015) and agricultural settings (Johnson et al. 2016; McKinney et al. 2018) and a newly designed primer, DF-DR, modified from the F3-R3 primer set and an MGB probe added to increase specificity showed good coverage and specificity. All were successful PCR and RT-Q-PCR assays.

To confirm their specificity, MiSeq amplicon sequencing of the short amplicons was undertaken from the Thai septic tank samples. While the diversity of the amplicons was low, likely reflecting the short amplicon length, the *intI*1 gene was ubiquitous in our samples supporting previous findings where it was dominant in polluted environments such as WWT (Gillings et al. 2015; Zheng et al. 2020). The ASVs generated from each primer set were highly similar to *intI*1 and *intI*1-like sequences obtained from known and unknown bacteria, which were isolated from a range of clinical settings (e.g., human commensal) and environmental sources (e.g., wastewater-activated sludge) (Figures 2.8, 2.9, 2.10). Interestingly, a few of the *intI*1-like sequences characterised were from known bacterial species isolated within clinical context including human faeces and bloodstream **(**Figures 2.8, 2.9, 2.10**).** This observation challenges the well-established knowledge that *intI*1 sequence recovered within clinical settings have identical/nearly identical protein (≥98% protein identity) (Roy et al. 2021) and/nucleotide sequence (99-100% nucleotide identity) (Gillings et al. 2008b) to each other. As such, this implies that *intI*1-like sequences can also be present within clinical settings and not just restricted to environmental settings as originally thought.

Our sequence results also highlight that the primer sets show that it was not possible to distinguish between *intI*1 and the lesser conserved *intI*1 variants (*intI*1-like, <98% protein similarity) that have been shown to co-exist within these settings and similar environments (Gillings et al. 2008b; Gillings et al. 2015). These less conserved CL1-integron integrases (*intI*1-like) have been found, for example, on the chromosome of non-pathogenic *Betaproteobacteria* isolated from biofilms and soil and the entrained gene cassettes encoded currently unknown function rather than AMR (Gillings et al. 2008a). *intI*1-like may therefore not contribute to AMR but will contribute to *intI*1 Q-PCR signal. None of the primers, not even with the addition of a TaqMan or MGB probe, were able to distinguish between *intI*1 and *intI*1-like. As such, quantified *intI*1 gene abundance could potentially be overestimated. However, designing new primers over longer region but still suitable for QPCR, capable of distinguishing both variants can be a challenge. This is because the *IntI*1-like protein sequence identity between bacteria species can vary when compared to the reference *intI*1 *Pseudomonas aeruginosa* reference sequence (pVS1, AAA25857.1). For example, the *intI*1-like protein sequence from *Salmonella enterica subsp* (KMJ40944.1), a gamma-proteobacteria and *Comamonas thiooxydans* beta-proteobacteria (WP_012838479.1) shared 87.8% and 92.8% identity to the reference pVS1 *intI*1 protein sequence respectively. As such, the varying conserved region shared between the *intI*1, and *intI*1-like sequence variant makes it a challenge to design a primer that exclusively distinguish both variants.

The potential contributions of *intI*1-like abundance to the overall abundance of *intI*1 gene quantified via Q-PCR suggests that *intI*1 abundance may not be an adequate or reliable proxy for inferring overall AMR abundance. Therefore, other potential proxies such as the *qacEΔ*1 (confer antiseptic resistance) or *VanA* (confer vancomycin resistance) (Abramova et al. 2022) should be investigated for reliable estimation of overall AMR abundance in polluted environments.

This work has also shown the impact of using different primers on the interpretation of the findings and in turn our understanding of the risk of AMR. While across our septic tanks, the three best-performing primer sets revealed the same overall trends (Figure 2), they did on occasion change the statistical difference between samples. For example, there were statistically higher *intI*1 gene abundances in the CST-household effluent than CST-healthcare sludge when quantified with two of the primer sets (DF-DR and F3-R3, Figure 2.7A-B) but no statistically significant difference when using the third primer set (F7-R7) (Figure 2.7C).

Depending on the primer set used, our understanding of the role of wastewater in the dissemination of CL1-integron and entrained AMR gene to the wider environment differed, highlighting the need for primer standardisation if comparisons and environmental meaning are to be gained from the large body of literature and work currently being undertaken in this area. With this in mind, from the work carried out validating and comparing the primer sets, we arrived at three very good primer sets, albeit with the lack of specificity for *intI*1. Recently, a new Q-PCR primer-probe set (*aint*1) has been developed to specifically target CL1-integron carrying gene cassettes associated with human-impacted or anthropogenic sources (Quintela-Baluja et al. 2021). Unlike the *intI*1 primer sets investigated in this study which targets the *intI*1 at the 5'CS of the CL1-integron structure, the newly designed anthropogenic impacted CL1-integron primer-probe set targets the *attc*/*qacΔE1* region at the 3' CS of the CL1-integron structure. This newly designed primer-probe set effectively differentiated CL1-integrons associated with AMR from those not associated with AMR (*intI*1-like integrons). Future testing is needed to compare the performance of the newly designed anthropogenic primer (ant1) with the selected *intI*1 primers, especially the newly designed set.

Nonetheless, as the addition of the TaqMan and MGB probes did not offer increased specificity, we recommend the F3-R3 primer set and SYBR green assay (Rosewarne et al. 2010). This primer set has previously been extensively used in the literature to survey CL1-

integrons from a wide-ranging environment (Paiva et al. 2015) and here we have further demonstrated their suitability to quantify mRNA also. For this, a gene-specific RT-Q-PCR assay performed best as previously demonstrated (Cholet et al 2020).

The combination of Q-PCR and amplicon sequencing approach offered a rapid targeted and cost-effective alternative, in contrast to shotgun metagenomics. This approach permitted reliable and accurate profiling of functional genes from various environments; and is therefore, highly recommended in future studies.

2.4.1 Risk assessment of septic tanks in contributing to *intI*1 gene abundance to the environment

Comparing the abundance of *intI*1 gene copies among the different septic tanks, we showed that they were higher in the sludge (copies/g DNA) compared to effluent (copies/ml DNA), for all three reactors (CST-household, CST-healthcare, SST-household), irrespective of the *intI*1 primer set used, with the highest gene abundance quantified in the solar septic tank (SST-household) sludge and lowest conventional healthcare (CST-healthcare) sludge (Figure 2.7). Although, the ratio of *intI*1 to *16S rRNA* gene in the sludge remained consistently lower compared to that of the effluent across reactors indicating, that there were more *intI*1 genes per genome in the effluent than sludge. Nonetheless, the low *intI*1 gene copy (copies/g DNA) in the CST-healthcare reactor sludge compared to household reactors is surprising given that healthcare institutions are among the primary consumers of antimicrobials particularly antibiotics (Stalder et al. 2014).

Of the three reactors, lower *intI*1 gene abundance (copies/ ml) was quantified in the solar septic tank (SST) effluent compared to the conventional tanks (CST-healthcare and CST-household), irrespective of the primer set used. In addition, *intI*1 gene abundance (copies/ ml) was highest in the CST-healthcare effluent among the three tanks, again regardless of the primer set. These observations were further supported by the calculated daily release of CL1-integron into the environment, based on each tank's flow rate.

The higher *intI*1 gene abundance in the healthcare reactor effluent reported in this study was consistent with a previous study that reported higher *intI*1 relative gene abundance (normalised abundance to the *16S rRNA* copies) in hospital effluent compared to urban or municipal WWTP effluent (Stalder et al., 2014). The higher *intI*1 gene abundance in the CST-healthcare effluent could potentially be attributed to stronger selective pressures

imposed on the bacteria communities, given the higher and diverse antibiotics consumed within healthcare settings. Consequently, bacterial populations within these settings become more inclined to acquire resistance genes, including those carried within key vectors such as CL1-integron, to ensure their survival from the constant threat of antimicrobials within WWT system.

Indeed, the low *intI*1 gene copies quantified in the SST-household effluent imply that the increased temperature in the tank's disinfection chamber potentially plays a role in reducing CL1-integron from WWT and thus, the abundance entering the environment. This finding partly agrees with our proposed hypothesis of decreased *intI*1 gene abundance as a result of increased temperature driving enhanced wastewater treatment. Although the target internal temperature (50-60°C) within the solar tank was not consistently achieved, our finding is consistent with the recent study by Zhang and colleagues (Zhang et al. 2022), who investigated removal of CL1-integron and entrained AMR genes from anaerobic digestors operated at higher (thermophilic- 55°C) and lower (mesophilic-35°C and 25°C) temperatures and reported statistically lower *intI*1 gene abundance and removal at higher temperature. In addition, statistically lower *16S rRNA* gene abundance was reported at the higher temperature, coupled with a lower relative abundance of AMR gene cassettes, albeit slightly higher ARG subtypes were detected with the higher temperature.

Typified by poor treatment performance (Connelly et al. 2019), the conventional household tank with accessible influent was found to have increased *intI*1 gene copies (copies/ml) in the effluent compared to the influent. This increase ranged from 1.20 -18.45% depending on the primer set used. These findings contrast those of Thakali et al.,(2020) and Cuetero-Martínez et al.,(2023). Thakali et al.,(2020) reported decrease *intI*1 gene copies (copies/ml) in the final effluent compared to influent for two of the WWTPs investigated. Similarly, Cuetero-Martínez et al.,(2023) reported decrease *intI*1 gene copies (copies/ml) in the effluents (before and after disinfection) compared to influent for four of the WWTP investigated across two seasons (rainy and dry). Moreover, the decrease was statistically significant in some cases. Although this study and those of Thakali et al.,(2020) and Cuetero-Martínez et al.,(2023) have investigated different treatment systems (centralised vs decentralised system), the findings highlight the inadequacies of septic tanks in reducing CL1-integron abundance in the effluent.

WWT sludge represents an additional source of CL1-integron and entrained AMR genes to the environment, particularly if improperly managed (i.e. improperly disposed of without

further treatment), which further exacerbates the global AMR burden (Koottatep et al. 2021). In the Global south region such as Thailand and Vietnam, only 10-20% of the faecal sludge generated is estimated to be adequately disposed of, whilst the vast majority are discharged directly to the environment (Koottatep et al. 2021). With the high *intI*1 abundance quantified in the sludge for the three reactors, coupled with the already high abundance in the effluent, a significant amount of CL1-integron enters the environment via the CST-household, CST-healthcare and SST-household respectively. This is significant when taking into account the proportion of the global population (2.7 billion people) estimated to be served by onsite decentralised WWT including septic tanks (Harada et al. 2016). Thus, highlighting septic tanks as an important source of CL1 to the environment, and further supports the broader knowledge that WWT in general, are a major source of CL1-integrons and entrained resistance genes to the environment.

The increased abundance of CL1-integrons entering the natural environment from WWT coupled with a slow decay rate (*intI*1 halve-life estimated ≥ 1 month in soil (Burch et al. 2014)), increases the risk of acquisition and dissemination into broader bacteria taxa especially clinically relevant human pathogenic bacteria including *Acinetobacter baumannii* (Nikibakhsh et al. 2021), *Proteus mirabilis* (Chen et al. 2017; Lu et al. 2022) and *Pseudomonas aeruginosa* (Liu et al. 2020; Khademi et al. 2021).

2.5 Conclusions

This present study has provided insight into the importance of primer choice, especially in the context of validating the *intI*1 as a suitable proxy for AMR pollution, and the need for standardisation across studies to comprehensively understand the role in which wastewater plays in disseminating CL1-integrons and by extension AMR genes to the environment. Further work is needed to determine if the *intI*1 is indeed a suitable proxy for overall AMR gene abundances.

Moreover, we showed septic tank decentralised wastewater can be a significant source of CL1-integron to the environment via the effluent and sludge, especially if the sludge is directly applied to the environment without undergoing additional treatments. Thus, supports growing evidence that WWTs, in general, are a significant source of CL1-integrons and associated resistance genes to the wider environment which further exacerbates the global burden from AMR.

Chapter 3

Validation and quantification of AMR genes using high-throughput Q-PCR array technology

3.1 Introduction

With the rapid rise and spread of Antimicrobial resistance (AMR), exacerbated primarily by increased consumption and mismanagement of current antimicrobials, and compelling evidences identifying WWT as a critical source of AMR genes and MGEs to the environment (Karkman et al. 2016; Su et al. 2020; Majlander et al. 2021), effective routine monitoring of broad-spectrum AMR genes and MGEs from WWT, as well as other polluted source, to the environment remains a challenge (Smith et al. 2022). This challenge not only hinders a comprehensive insight into the role WWT plays in the mitigation or spread of AMR to the wider environment but also hinders the identification of environments that pose the highest risk of AMR acquisition to human and animal health (Abramova et al. 2022). One such approach proposed to overcome this challenge in AMR monitoring was the use of a proxy gene, the CL1-integron-integrase gene (*intI*1), to infer potential AMR pollution (Gillings et al. 2015). This is owing to CL1-integron linkage to gene conferring resistance to antibiotics, heavy metals and biocide and its elevated abundance in an anthropogenic polluted environment (Gillings et al. 2015; Pruden et al. 2021).

The suitability of this gene (*intI*1), however, as an adequate and reliable proxy to infer AMR overall AMR abundance in pollution environment remains to be fully investigated, with some studies reporting statistically positive correlations between quantified *intI*1 abundance and overall ARG abundance (including AMR genes associated and not-associated with CL1-integron) (Su et al. 2020) whilst others (Chen and Zhang 2013b; Dungan and Bjorneberg 2020) only reported statistically positive correlations between the abundance of a subset of genes investigated to the abundance of the *intI*1. For example, Su et al., (2020) reported strong statistical positive correlations (correlation: 0.71 to 0.96, p-value <0.01) between the abundance of *intI*1 and overall ARGs (*n=14,* belonging to six antibiotic classes) investigated which were both associated (*blasPSE-1,dfrA1*) and non-associated (*tetA, tetC, tetO, qnrA, qnrB, qnrS, ermA, ermB ereB, mphA, vatB, sul2*) with CL1-integrons. Chen et al., on the other hand, observed statistically positive correlation between the copy number of *intI*1 and *sul*1 ($R = 0.756$, $P < 0.05$) and not to the other ARGs genes *(sul*2, *tetM, tetO, tetQ and tetW*) investigated (Chen and Zhang 2013b). In fact, Paulus et al.,(2020) found that the observed

significant positive correlation (R^2 =0.72, p-value < 0.01) between *intI*1 relative abundance (normalised to 16S *rRNA* abundance) and overall ARGs relative abundance (13 ARGs) quantified from river water in their study was primarily driven by the abundance of the *Sul*1 gene (second highest gene abundance quantified in their study after *intI*1) abundance. Subsequent re-analysis with the exclusion of *sul*1 abundance revealed a weak and insignificant correlation (R^2 =0.05, p-value >0.05) between *intI*1 abundance and overall ARGs abundance. Therefore, they concluded that the *intI*1 gene abundance is not a good proxy for overall ARG abundance (Paulus et al. 2020).

An emerging approach rapidly being adopted as an alternative to tackle challenges in broad-spectrum AMR monitoring, is the use of the high-throughput Q-PCR (HT-Q-PCR) array. HT-Q-PCR targets hundreds of ARGs and selected MGEs simultaneously within a single reaction on a nanoscale level (Waseem et al. 2019). This allows for a high number of genes to be amplified simultaneously while maintaining the benefits (sensitivity and specificity) characteristic of conventional Q-PCR (Waseem et al. 2019). There is however a trade-off between the number of genes on the array and the number of samples you can screen. As such, the HT-QPCR array is gradually becoming widely adopted for used to profile ARGs and MGEs in various ecological niches including WWTs (Karkman et al. 2016; Wang et al. 2018; Majlander et al. 2021), urban park soils exposed to reclaimed water (Wang et al. 2014), drinking water treatment plants (Xu et al. 2016), Baltic Sea Fish farm sediments (Muziasari et al. 2016) and river water (Zheng et al. 2017). For example, Karkman et al., (2016) targeted 285 genes to investigate seasonal (summer, autumn, winter, spring) variations of AMR and transposase abundance in a year at an urban wastewater treatment plant in Finland. The authors detected and quantified 175 AMR genes and nine transposase genes and reported minor seasonal variation in the ARGs relative abundance between seasons for the samples (influent, effluent, sludge); thus, implied that the ARGs abundance remained relatively stable over the year. Additionally, a reduction in ARG relative abundance was reported in the sludge and effluent from the influent for the seasons (Karkman et al. 2016).

Environmental AMR monitoring with any/both approaches (use of *intI*1 and/ HT-Q-PCR) remains promising and invaluable as it can facilitate and enable the identification of environments with the highest risk to human and animal health. However, in light of the above-mentioned concerns raised for both methods, specifically the potential sub-optimal condition within the HT-QPCR array and the *intI*1 as an unreliable proxy for inferring potential AMR pollution, a comprehensive and reliable AMR monitoring to gain insights into WWTs in contributing to the overall global AMR burden. As such, we undertook to

investigate the link between *intI*1 gene abundance and overall AMR abundance quantified on the HT-QPCR array, to validate its suitability as an adequate and reliable proxy for inferring overall AMR pollution using a suite of wastewater samples from septic tanks in Thailand associated with household and healthcare usage.

In addition, we utilised the HT-QPCR array to characterise and quantify AMR and MGEs from our septic wastewater. Specifically, we compared the solar septic tank (SST), the recent technology currently implanted in some areas of Thailand and described in the previous chapter (Chapter 2) and elsewhere (Polprasert et al. 2018; Connelly et al. 2019) to the conventional tanks which treat household and healthcare wastewater, with the hypothesis that the increased incorporated temperature within the SST units will not only decrease the abundance of mobile integrons, particularly the CL1-integron (as shown previously in chapter 2) but also the AMR subtype and abundance quantified. In addition, we hypothesised that a poorer statistical correlation would be observed between *intI*1 abundance of and abundance of quantified AMR genes. This is because *intI*1 primer sets used on the array have lower *intI*1 sensitivity (i.e., coverage) at stringent threshold (i.e., no mismatch between primer and template sequence) compared to other *intI*1 primer currently available as indicated by our *in-silico* primer analysis conducted in previous study (see Chapter 2, Appendix Table A.2; HT-QPCR array *intI*1 primer sets AY289 and AY293, corresponded to the F4-R4 and F10-R10 respectively).

To address the outlined research aims and hypotheses, we first pre-screened our septic wastewater samples to investigate overall similarities and/ differences between reactor types and sample types, and to inform gene targets to select for subsequent individual samples by pooling the wastewater samples by reactor type (CST-healthcare, CST-household, and SST-household) and sample type (influent, sludge, effluent). Next, we selected 72 genes, informed by the initial pooled samples, and targeted 23 wastewater samples. Selected genes have known association and non-association to mobile resistance integrons (MRIs) and conferred resistance to the 11 major antibiotic classes. Two array assays targeting the *16S rRNA* and *intI*1 gene were selected, from the 72 genes, to validate conditions within the HT-QPCR by quantifying these genes in-house and comparing obtained absolute Ct to that obtained by the HT-QPCR array. Additionally, the HT-QPCR array *16S rRNA* primer set was compared to a well-described and validated *16S rRNA* primer-probe set to evaluate the impact of *16S rRNA* primer choice on quantified *16S rRNA* gene abundance on the same wastewater samples, owing to a low mean Ct and indistinguishable melt curve peak from samples observed for the no template control sample quantified by HT-QPCR *16S rRNA*

primer set. Furthermore, differences and similarities in AMR and integrase gene abundance between the sample type and rector type were also investigated. Finally, the link between integrase gene abundance, particularly *intI*1 gene abundance, to over AMR gene abundance quantified on the HT-QPCR array was assessed to ascertain the suitability of the *intI*1 gene as an adequate and reliable proxy for inferring overall AMR pollution.

3.2 Materials and Methods

3.2.1 Solar and Conventional septic tank sampling

Sampling of solar and conventional septic tanks was the same as described in the previous chapter (See Chapter 2, section 2.3.1)

3.2.2 DNA extraction

From each sample, DNA extraction was performed with the DNeasy PowerSoil Kit (Qiagen), following the manufacturer's instructions. The integrity of extracted genomic DNA was assessed via agarose gel electrophoresis and DNA concentration was quantified fluorometrically using the Qubit (Invitrogen) according to manufacturer instructions.

3.2.3 Sample pooling for AMR and MGEs pre-screen

3.2.3.1 Sample selection and pooling for HT-QPCR quantification

10ng of DNA in given volume (lower limit: 50µl, max volume: 100µl) is required for the HT-PCR array. Thai septic tank wastewater samples with sufficient DNA concentration (10ng) and volume (≥50µl) were selected and pooled by reactor type (CST-healthcare, CST-household, and SST-household) and sample type (influent, sludge, effluent) (Table 3.1)**.** Pooled samples were pre-screened for their AMR and MGE profile to investigate overall similarities/ differences between reactor types and sample types; and to inform gene targets for individual samples AMR and MGE quantification.

Briefly, influent (*n=2*), sludge (*n=4*) and effluent (*n=4*) samples from CST household unit; sludge (*n=2*) and effluent samples (*n=1*) from CST healthcare unit; and sludge (*n=5*) and effluent (*n=2*) samples from SST household units were pooled into individual tubes resulting in seven pooled sample tubes (Table 3.1). An eighth tube containing nuclease-free water was

included as the no-template control (NTC). Pooled samples were sent to Resistomap (Helsinki, Finland) for Q-PCR quantification on the HT-QPCR AMR array (2.1).

Table 3.1: Thai septic tank wastewater samples and time points selected for sample pooling for AMR and MGE profile pre-screening

Reactor type	Reactor ID	April 2018	May 2018	June 2018	Nov 2018	March 2019	June 2019	July 2019	Aug 2019	Sept 2019
SST	SST-01	SLG	-	EFF/ SLG						
	SST-07	SLG			SLG	EFF/ SLG				
CST	CST-P3						INF/EFF/ SLG	INF/EFF/ SLG	EFF/ SLG	
	CST-J6								EFF/ SLG	
	CST-HC								EFF/ SLG	SLG

SST: Solar septic tank; CST: Conventional septic tank; INF: Influent; SLG: Sludge; EFF: Effluent

3.2.3.2 High-throughput QPCR (HT-QPCR) quantification of ARGs and MGEs on wastewater samples

HT-QPCR was performed using the SmartChip Real-time PCR system (Takara Bio, Mountain View, CA, USA) at Resistomap (Helsinki, Finland). Pooled samples were pre-screened for AMR genes and MGEs profiles on two independent SmartChips using the 384-primer set 2.1 (Table 3.2). The 384 primers targeted 323 genes conferring resistance to the major classes of antibiotic (aminoglycoside, trimethoprim, β-lactam, phenicol, tetracycline, sulphonamide, multidrug resistance (MDR), macrolide-lincosamide-streptogramin B (MLSB), vancomycin, quinolone and other), 4 integrase genes from three mobile resistance integrons (class 1, 2 and 3), 48 MGEs (including Transposons, insertion sequence and plasmids), 8 taxonomic bacterial genes and the 16S *rRNA* gene (see Appendix Table B.1). The 16S *rRNA* gene served as the positive control for the QPCR reaction (Majlander et al. 2021).

Each SmartChip had a 5184-reaction capability with 100nl volume per reaction. Each 100nl reaction volume consisted of 1X SmartChip TB Green Gene Expression Master Mix (TakaraBio), nuclease-free PCR-grade water, 300nM of each primer and a DNA template of 2 ng/µl. The SmartChips were filled using the SmartChip Multisampling Nano-dispenser (Takara Bio) (Majlander et al. 2021). The cycling condition was as follows: 95°C-10 min, (95°C-15 sec, 60°C-15 sec, 72°C 30 sec)× 40 cycles and a final melting curve analysis with

temperature increased to 97°C with 0.4°C increments per step to observe assay specificity. Amplification was conducted in triplicate for each primer set (i.e., assay).

3.2.3.3 Data processing of raw pooled sample HT-QPCR results

Arbitrary cycle threshold (Ct) cut-off values as the limit of detection for the HT-QPCR primers continue to be widely employed in studies utilising the HT-QPCR array to profile AMR genes and MGEs. These Ct cut-offs, which include 31 (Chen et al. 2016; An et al. 2018), 30 (Chen et al. 2019b), 28 (Stedtfeld et al. 2018) and 27 (Muziasari et al. 2016; Majlander et al. 2021), would imply low abundant genes (low copy number genes) with Ct above the selected limit of detection are deemed false positive amplifications and thus, discarded. In addition, coupled with a often lack of comparison to the NTC, assays with Ct's similar to the NTC and/or less than one log difference between the sample and NTC are retained introducing bias to reported data. Here, in this study, a 3.32 Ct difference (1 log fold difference) between NTC and sample was adopted as an integral part of the data processing step (Smith and Osborn 2009). All data processing and analysis were performed in R (R Development Core Team, 2016). First, primer sets (i.e., assays) within each pooled wastewater sample with no amplification in any of the three technical replicates were removed. Next, primer sets with amplification in only one of the three technical replicates were deemed a false positive and were further removed from analysis for the specific sample (Karkman et al. 2016). Amplification in two or three of the technical replicates was regarded as true positive amplification and the mean cycle threshold (Ct) was calculated from these replicates.

NTC assays were performed for all primer sets. Amplification in at least one of the technical replicates for each primer set was considered real amplification and was used to compare the Ct difference with the corresponding assay in samples. A mean was calculated and used for comparison of Ct with the corresponding assay in samples if amplification was observed in two or three of the NTC technical replicates. Assays with mean Ct difference between the NTC primer and corresponding sample primer <3.32 Ct were discarded from further analysis for the specific sample. The total number of retained quantified genes for each pooled sample coupled to a presence and absence heatmap was visualised using the ggplot2 package (Wickham 2009). The resistance mechanism of retained AMR genes was retrieved from the CARD (Comprehensive Antibiotic Resistance Database) database (Alcock et al. 2023) (https://card.mcmaster.ca/ last accessed: November 2022) or published literature (Stedtfeld et al. 2018).

A Non-metric multidimensional scaling (NMDS) analysis based on the Jaccard distance matrix was used to investigate similarities and differences in the gene profile (presence and absence) of each pooled sample.

3.2.3.4 Selection of target genes (primer sets) for individual samples for HT-QPCR array quantification and data processing

From the pre-screened pooled samples, a subset of AMR and MGE genes were selected for quantification from individual samples (*n=23*, Table 3.2). Selected AMR genes included - known mobile integrons associated and non-associated genes that conferred resistance to the major antibiotic class (aminoglycoside, trimethoprim, β-lactam, phenicol, tetracycline, sulphonamide, multidrug resistance (MDR), macrolide-lincosamide-streptogramin B (MLSB), vancomycin, quinolone and other). In addition, a primer set targeting the *16S rRNA* gene, which also served as a positive control of the HT-QPCR reaction was included in the selected genes targeted on the individual samples that were quantified on the HT-QPCR array. A no-template control (PCR-free water) was also included on the HT-QPCR array resulting in a total of 24 individual samples targeted on the HT-QPCR array.

HT-QPCR array raw data for individual samples were processed in the same manner as specified above (see section 3.2.3.3). Retained genes within each sample were used for downstream analysis. In addition, a subset of genes quantified on the HT-QPCR array, specifically the *16S rRNA* (AY1) and *intI*1 gene (AY289), were used to validate the HT-QPCR array.

Table 3.2: Thai septic tank wastewater samples and time points selected for individual AMR and MGE gene quantification on the HT-QPCR array

Reactor type	Reactor ID	April 2018	May 2018	June 2018	Nov 2018	March 2019	June 2019	July 2019	August 2019	Sept 2019
SST	SST-01	SLG	SLG	EFF/ SLG						
	SST-07	SLG			EFF/ SLG	EFF/SLG				
CST	CST-P3						INF/EFF/ SLG	INF/EFF/ SLG	INF/EFF/ SLG	
	CST-J6								EFF/SLG	
	CST-HC								EFF/SLG	SLG

SST: CST: Solar septic tank; Conventional septic tank;I NF: Influent; SLG: Sludge; EFF: Effluent

3.2.4 HT-QPCR Array Validation and Best Practices: The Good, The Bad and The Ugly

3.2.4.1 *16S rRNA* and *intI*1 gene QPCR standard curve for absolute quantification

16S rRNA and *intI*1 gene standard curves were constructed similar to the method described in the previous chapter (see Chapter 2, section 2.2.1) with a slight addition to the method. For the *intI*1 gene, QPCR standard curves were constructed by amplifying a synthetic *intI*1 gene fragment containing the binding site for all *intI*1 primers used.

For the HT-QPCR array *16S rRNA* primers (Table 3.4) standard curves were constructed by amplifying the V3-V9 region of cloned *E. coli 16S rRNA* gene (Smith et al. 2006; Cholet et al. 2020) using the 515F and Prok 1492R *16S rRNA* primers. Similarly, standard curves for the TaqMan *16S rRNA* primers (Bact 1369F and Prok 1492R and TaqMan probe TM 1389F (Suzuki et al. 2000)) were constructed by amplifying cloned *E.coli 16S rRNA* amplicon (1369F and 1492R amplicon) using the T7-forward (5'-TAATACGACTCACTATAGGG-3') and M13-reverse (5'-CAGGAAACAGCTATGAC-3') primers (Smith et al. 2006; Cholet et al. 2020). The primer sequences, reaction volumes and cycling conditions, for all primers quantified in-house are listed in Table 3.4.

3.2.4.2 Comparison of quantified *16S rRNA* and *intI*1 gene on the HT-QPCR array and In-house quantification

As assays within the HT-QPCR array experience the same cycle condition (Waseem et al. 2019), the effectiveness of the HT-QPCR array in quantifying AMR genes and MGEs, especially within complex samples such as wastewater samples was validated in-house. First, a subset of HT-QPCR array primer sets, specifically the *16S rRNA* (AY1) and *intI*1 (AY289) primers, were selected and optimised in-house for QPCR. Optimised primers were subsequently used to quantify the same Thai wastewater samples (*n=23*) as quantified on the HT-QPCR array.

For each primer set validated in-house, QPCR amplification was carried out in a 20µl volume reaction using 2µl (1:50 diluted) template DNA. Reaction volume, conditions, primer sequences and probe type for the three selected optimal *intI1* primer pairs are detailed in

Table 3.4. Assays were performed in duplicates. Triplicate no template control (NTC) was included for each validated primer set.

Quantified in-house Ct of *16S rRNA* and *intI*1 primer (primer used on HT-QPCR array) were corrected to reflect expected Ct when the same input DNA concentration used on the HT-QPCR array (2ng/ul) was used (Equation 1).

Equation 1

Dilution factor= [DNA] of HT-QPCR sample A/ [DNA] of in-house QPCR sample A

Log_2 (Dilution factor sample A)

Expected CT= in-house QPCR Ct sample A - Log_2 (Dilution factor sample A)

A subsequent linear correlation was performed to compare the corrected in-house QPCR Ct to that of the HT-QPCR Ct for the quantified *16S rRNA* and *intI*1 genes. In addition, a two-way analysis of variance (ANOVA) followed by a Turkey HSD post hoc test, was used to compare *16S rRNA* and *intI*1 gene abundance quantified on the HT-QPCR array and in-house QPCR for each of the sample types (influent, sludge, and effluent) and reactor type (CST-Household, CST-Healthcare and SST-Household). Pearson correlation coefficient analysis was performed to assess the correlation between gene abundance quantified by each primer set.

3.2.4.3 In-house comparison of HT-QPCR array *16S rRNA* primer and TaqMan *16S rRNA* primer

The quantitative results of HT-QPCR array *16S rRNA* primers (1108F-1132R) were compared to that of a TaqMan *16S rRNA* primer set (1369F-1492R) by quantifying the *16S rRNA gene* using both primer sets on the same wastewater samples quantified on the HT-QPCR array.

QPCR quantification was carried out in the same manner as described above (see section 3.2.4.2). Reaction volume, conditions, primer sequences and probe type are detailed in Table 3.4.

Two-way analysis of variance (ANOVA) followed by a Turkey HSD post hoc test, was employed to compare *16S rRNA* gene abundance quantified by both *16S rRNA* primer sets for each of the sample types (influent, sludge, and effluent) and reactor type (CST-Household, CST-Healthcare and SST-Household). A linear regression analysis was performed to assess the linear relationship between observed mean sample Cts amplified by both primer sets. Pearson correlation coefficient analysis was performed to assess the correlation between gene abundance quantified by each primer set.

3.2.4.4 To report gene abundance as mean Ct/relative abundance to *16S rRNA* gene or not?

Whilst normalisation of target gene abundance to the abundance of quantified *16S rRNA* gene continues to be widely used to express relative gene abundance of quantified targets (Jiao et al. 2018; Majlander et al. 2021), this approach remains controversial as the *16S rRNA* gene copy number per bacterial genome can vary significantly between species (1-15 copies). Thus, variation in total bacteria load between samples masks real differences in quantified targets when cross-sample comparisons are made (Smith and Osborn 2009). As such, we investigated the impact of reporting obtained QPCR data as either mean Ct or relative abundance (normalised to the *16S rRNA* abundance). First, *intI*1 gene abundance was normalised to the *16S rRNA* gene abundance quantified on the HT-QPCR array. Next, the abundance of the *intI*1 gene was normalised to the *16S rRNA* gene abundance quantified in-house. Relative gene abundance of *intI*1 gene to the *16S rRNA* gene (*intI*1 gene copy per *16S rRNA* gene copy) in each sample was calculated according to the $2^{-\Delta Ct}$ method (where - ΔCt = Ct of detected gene – Ct of *16S rRNA* gene) (Schmittgen and Livak 2008; Majlander et al. 2021). Normalised *intI*1 gene abundance quantified on the HT-QPCR array and in-house were compared to each other and to the obtained *intI*1 absolute Ct.

3.2.5 Application of HT-QPCR array: Risk assessment of the individually targeted septic tanks wastewater samples in disseminating AMR genes and integrases (*intI*1, *intI*2, *intI*3) to the environment

Retained genes post data processing were analysed for their AMR gene and MGE profile within individual samples. All analysis was performed in R. Heatmap of gene abundance was visualised using the ggplot2 package (Wickham, 2009). NMDS analysis based on Bray-

Curtis dissimilarity matrix using the vegan package in R (Oksanen et al. 2022a) was used to visualise similarities and differences in gene profile and abundance between sample type and reactor type.

One-way analysis of variance (ANOVA) followed by a Turkey HSD post hoc test, was used to compare overall abundance of genes within each antibiotic class and overall abundance of the three integrases class (*intI*1, *intI*2, *intI*3). In addition, two-way ANOVA followed by a Turkey HSD post hoc test, was used to compare the effect of sample types (influent, sludge, and effluent) and reactor type (CST-Household, CST-Healthcare and SST-Household) on ARGs and integrase gene abundance.

3.2.6 Link between *intI*1 gene abundance and overall AMR abundance using HT-QPCR array

Pairwise correlation (Pearson correlation coefficient) analysis was performed to analyse the relationship between integrase gene abundance and ARG abundance using wastewater samples (*n=23*) quantified with the HT-QPCR array. Briefly, the abundance of ARGs associated and non-associated with mobile resistance integrons was correlated with the abundance of integrase genes (*intI*1, *intI*2 and *intI*3) and *sul*1 resistance gene. Where no amplification was quantified in a sample, a Ct of 40 (maximum HT-QPCR array cycle) was used to permit pairwise data comparison. A p-value <0.05 was used as the threshold for significance. The correlation heatmap of the pairwise comparison was visualised using the ggplot2 package (Wickham 2009).

3.2.6.1 In-house *intI*1 gene QPCR quantification from same wastewater samples quantified on HT-QPCR array using HT-QPCR array *intI*1 primer sets and previously optimised *intI*1 primers

The same Thai wastewater samples (*n=23*) quantified on the HT-QPCR array were subjected to in-house *intI*1 gene QPCR quantification using the HT-QPCR array *intI*1 primers (AY289 and AY293) and previously optimised *intI*1 primer sets (DF-DR, F3-R3 and F7-R7; see chapter 2). The HT-QPCR array *intI*1 primer sequences AY289 and AY293 corresponded to the F4-R4 and F10-R10 primer sequences respectively from the previous Chapter (herein referred to as F4-R4 and F10-R10 *intI*1 primer sets).

For each primer set, QPCR amplification was carried out in a 20µl volume reaction using 2µl (1:50 diluted) template DNA. Reaction volume, conditions, primer sequences and probe type for the *intI*1 primer sets are detailed in Table 3.4. Assays were performed in duplicates. Duplicate NTC was included for each primer set. Reactions were performed on the Bio-Rad CFX96 Touch Real-Time PCR Detection System and analysed with the Bio-Rad CFX Manager 3.1 software. Melt curve analysis was performed, for the SYBR Green assay, from 65°C to 95°C with 0.5°C increments every 5 secs, and a single peak was confirmed to ensure assay specificity.

Two-way statistical analysis of variance (ANOVA) followed by a Turkey HSD post hoc test, was employed to compare gene abundance for each of the sample types (influent, sludge, and effluent) and reactor type (CST-Household, CST-Healthcare and SST-Household) for each primer set, when Shapiro-Wilks test indicates normality of data. A p-value of 0.05 was chosen as the significance threshold. Kruskal-Wallis test was performed alternatively when the Shapiro-Wilks test indicated a non-normal distribution of the data (p-value <0.05). Dunn post-hoc test was subsequently employed to compare gene abundance for each of the sample types (influent, sludge, and effluent) and reactor type (CST-Household, CST-Healthcare and SST-Household) for each primer set. Pearson correlation coefficient/ Spearman ranks sum correlation analysis was used, following the Shapiro-Wilks normality test, to assess the correlation between gene abundance quantified by each primer set.

3.3 Results

3.3.1 Sample pooling for AMR and MGEs pre-screen

3.3.1.1 Arbitrary Ct cut-offs retain assays with similar Ct to the no template control: Data processing of raw pooled sample HT-QPCR results

Step 1: Removal of assays with amplification in only one of the three replicates

From pooled samples, 35 CST-household influent, 35 sludge and 33 effluent assays were discarded as a result of no amplification in any of the three technical replicates or due to amplification in only one of the three replicates. Similarly, 22 and 41 assays from the CST-healthcare sludge effluent respectively and 31 and 41 assays from the SST-household unit sludge and effluent respectively were discarded owing to no amplification in any of the three

technical replicates or due to amplification in only one of the three replicates (Table 3.3, Figure 3.1).

Step 2: Next to ensure that gene abundances (inferred by Ct) are a log-fold greater than the NTC

239 assays' NTC (63%) had no amplification in the three technical replicates, while 145 assays' NTC (37%) had amplification in at least one of the three NTC technical replicates (Table 3.3) and this Ct value was used to assess the log-fold difference between sample and NTC. This is to ensure that Ct values were above the negative control, and therefore represented real amplification the Ct of a gene target had to be 3.32 Cts (a log value) greater than the equivalent NCT (blank) as outlined in the method (see section 3.2.3.3) (Smith and Osborn 2009).

Following this approach, assays from the CST-household influent (*n=37*), sludge (*n=35*) and effluent (*n=37*); CST-healthcare sludge (*n=40*) and effluent (*n=47*); and SST-household sludge (*n=27*) and effluent (*n=38*) pooled samples (Table 3.3, Figure 3.1) were further removed. Of note, the *intI*1 primer set (AY293) was discarded from all pooled samples, as it had a mean NTC Ct of 13.77±0.11SD that was similar/ lower than that of quantified pooled samples, leaving only one *intI*1 primer set (AY289).

After NTC cut-off processing 312, *314 and 314* genes from the CST-household influent, sludge and effluent respectively were retained; *322 and 296* from the CST-healthcare sludge and effluent; *326 and 305* from the SST-household sludge and effluent pooled sample were retained for downstream analysis (Table 3.3, Figure 3.1).

Table 3.3: Data decarded and retained following data processing of pooled samples

Reactor type	Sample type	Genes targeted (*n= 384*)	Assays (primer sets) with no amplification or amplification in one of the three technical replicates (n)	Amplification in two or three replicates	Discarded genes following adoption of 3.32 Ct difference (n)	Retained total genes
CST-Household	Influent	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48* Taxanomic: *n= 8* 16S_rRNA: *n= 1*	*35*	AMR: *n= 294* Integron: *n= 4* MGE: *n= 42* Taxanomic: *n= 8* *16S rRNA*: *n= 1* ***n= 349* genes**	*37*	AMR: *n= 263* Integron: *n= 2* MGE: *n= 38* Taxanomic: *n= 8* *16S rRNA*: *n= 1* ***n= 312* genes**
	Sludge	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48* Taxanomic: *n= 8* *16S rRNA*: *n= 1*	*35*	AMR: *n= 295* Integron: *n= 3* MGE: *n= 43* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 349* genes**	*35*	AMR: *n= 264* Integron: *n= 2* MGE: *n= 40* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 314* genes**
	Effluent	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48* Taxanomic: *n= 8* *16S rRNA*: *n= 1*	*33*	AMR: *n= 296* Integron: *n= 3* MGE: *n= 44* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 351* genes**	*37*	AMR: *n= 263* Integron: *n= 2* MGE: *n= 41* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 314* genes**
CST-Healthcare	Sludge	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48* Taxanomic: *n= 8* *16S rRNA*: *n= 1*	*22*	AMR: *n= 305* Integron: *n= 4* MGE: *n= 44* Taxanomic: *n= 8* *16S rRNA*: *n= 1* ***n= 362* genes**	*40*	AMR: *n= 272* Integron: *n= 2* MGE: *n= 39* Taxanomic: *n= 8* *16S rRNA*: *n= 1* ***n= 322* genes**
	Effluent	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48* Taxanomic: *n= 8* *16S rRNA*: *n= 1*	*41*	AMR: *n= 286* Integron: *n= 3* MGE: *n= 45* Taxanomic: *n= 8* *16S rRNA*: *n= 1* ***n= 343* genes**	*47*	AMR: *n= 245* Integron: *n= 2* MGE: *n= 40* Taxanomic: *n= 8* *16S rRNA*: *n= 1* ***n= 296* genes**
SST-Household	Sludge	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48* Taxanomic: *n= 8* *16S rRNA*: *n= 1*	*31*	AMR: *n= 297* Integron: *n= 4* MGE: *n= 44* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 353* genes**	*27*	AMR: *n= 274* Integron: *n= 3* MGE: *n= 41* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 326* genes**
	Effluent	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48* Taxanomic: *n= 8* *16S rRNA*: *n= 1*	*41*	AMR: *n= 288* Integron: *n= 4* MGE: *n= 43* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 343* genes**	*38*	AMR: *n= 256* Integron: *n= 2* MGE: *n= 39* Taxanomic: *n= 7* *16S rRNA*: *n= 1* ***n= 305* genes**
No template control (NTC)	NTC	AMR: *n= 323* Integron: *n= 4* MGE: *n= 48*	*239*	***n = 145***	-	-

3.3.1.1 High diversity and richness of AMR genes and mobile gene elements in pooled samples, with drug inactivation as the dominant resistance mechanism

Step 3: Comparison of AMR diversity among the pooled sample

AMR genes and mobile elements (MGEs and integrons) diversity within each pooled sample was high (Figure 3.1). The identified AMR genes conferred resistance to all major classes of antibiotics (and some heavy metals and biocides), including Tetracycline, Sulphonamide, Aminoglycoside, MLSB Vancomycin, MDR, Quinolone, Phenicol, Other, β-lactam, Trimethoprim (Figure 3.1).

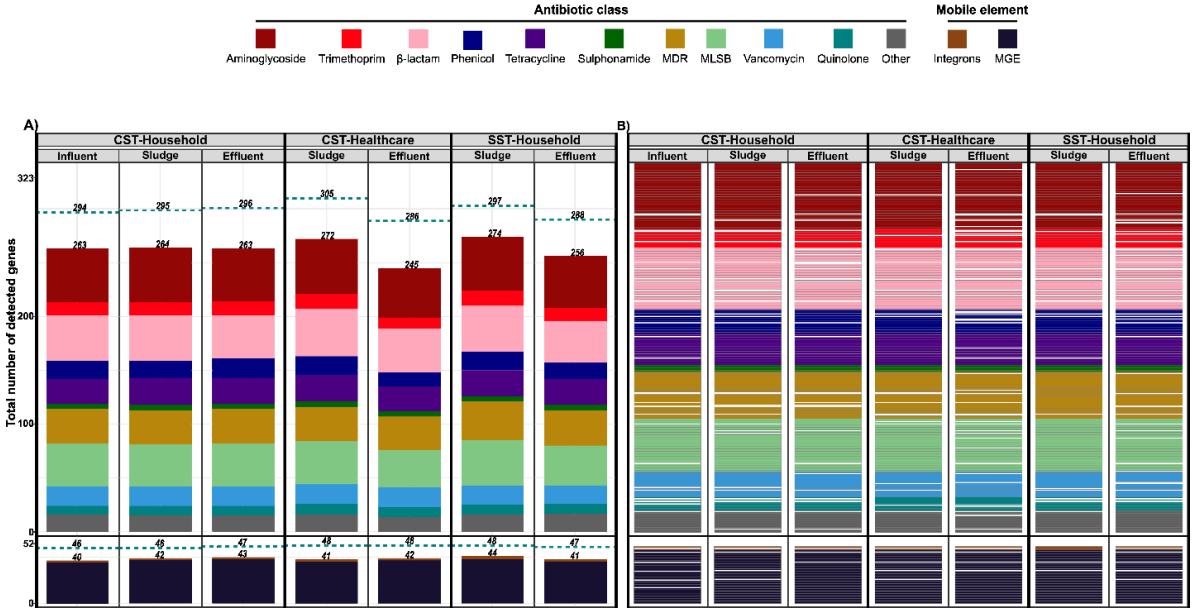

Figure 3.1: AMR genes and MGEs in pooled Thai wastewater samples on the high-throughput QPCR array SmartChip. Pooled wastewater samples were grouped by reactor type (CST-Household, CST-Healthcare and SST-Household) and sample type (influent, sludge, effluent). A) Total number of AMR genes and MGEs detected and quantified in pooled samples and B) presence and absence of quantified genes within each pooled sample. Each colour corresponds to a different antibiotic resistance class or mobile element. *16S rRNA* or Taxanomic genes are not included. Teal dash-line and value above the dashed line indicate the total number of genes quantified within each pooled sample from the 323 AMR genes and 52 MGEs targeted by the array. The number at the top of each stacked bar-plot indicates the total number of AMR genes and mobile element genes analysed following data processing (see section 3.2.3.3 for further details). CST denotes conventional septic tank; SST denotes solar septic tank.

Additionally, the identified AMR genes within each pooled sample encompassed five major resistance mechanisms, with the dominant mechanism reported as drug inactivation > drug efflux > target alteration > target protection > target replacement in all sample types

(influent, sludge, effluent) across the three reactors (CST-household, CST-healthcare, SST-household) (Figure 3.2). Conversely, for the mobile elements (Figure 3.1) identified genes included plasmids, transposons, insertional sequences and integrons (class 1, 2, 3).

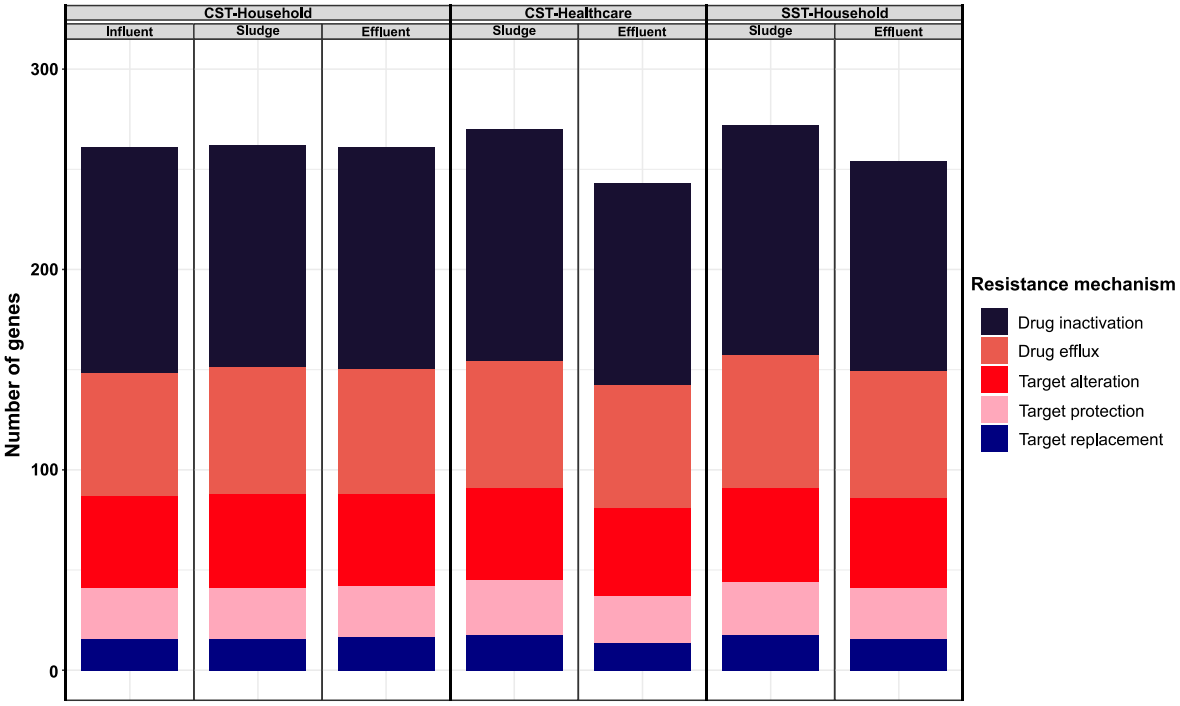

Figure 3.2: Resistance mechanisms of quantified AMR genes on the HT-QPCR array from pooled Thai wastewater samples. CST denotes conventional septic tank; SST denotes solar septic tank.

Richness of quantified AMR genes was higher in the sludge (CST-household: *n=264*, CST-healthcare: *n=272*, SST-household: *n=274*) > influent (CST-household: *n=263*) > effluent (CST-household: *n=263*, CST-healthcare: *n=245*, SST-household: *n=256*) (Table 3.3, Figure 3.1A). In contrast, the richness of mobile elements (Integrons and MGEs) (Table 3.3, Figure 3.1A) were generally higher in the pooled effluent (CST-household: *n=43*, CST-healthcare: *n=42*, SST-household: *n=41*) > sludge (CST-household: *n=42*, CST-healthcare: *n=41*, SST-household: *n=44)* > influent (CST-household: *n=40*), although the SST-household pooled sludge sample had the highest number of integrons and MGEs richness (Figure 3.1A). Interestingly, the *intI*2 gene was only quantified in the pooled SST-household sludge sample (Figure 3.1B).

The AMR (Figures 3.3A, 3.1B) and mobile elements (MGEs and integrons) (Figure 3.3B, 3.1B) gene profile between the three-reactor type (CST-household, CST-healthcare, SST-household) appear different from each other, and PERMANOVA indicated that reactor type explained 51.5% (p-value <0.01) (Figures 3.3A) and 19.2% (p-value = 0.76) (Figures 3.3B)

of gene profile variance observed for the AMR and mobile elements genes respectively. Within each reactor, the AMR (Figures 3.3A, 3.1B) and mobile element (Figures 3.3B, Figure 3.1B) gene profile of the samples (sludge and effluent) appeared to be different for CST-healthcare and SST-household unit, although the CST-healthcare sludge and SST-household effluent have the same mobile element gene profile (Figures 3.3B, 3.1). Meanwhile, for the CST-household samples (Influent, sludge, effluent), the AMR gene profile (Figures 3.3A, 3.1) appeared to be very similar to each other, with little variation between the sample types (Table 3.3, Figures 3.1, Figure 3.3A), while for the mobile elements (Figures 3.3B, 3.1), only the influent and sludge appeared to have similar gene profiles as compared to the effluent sample.

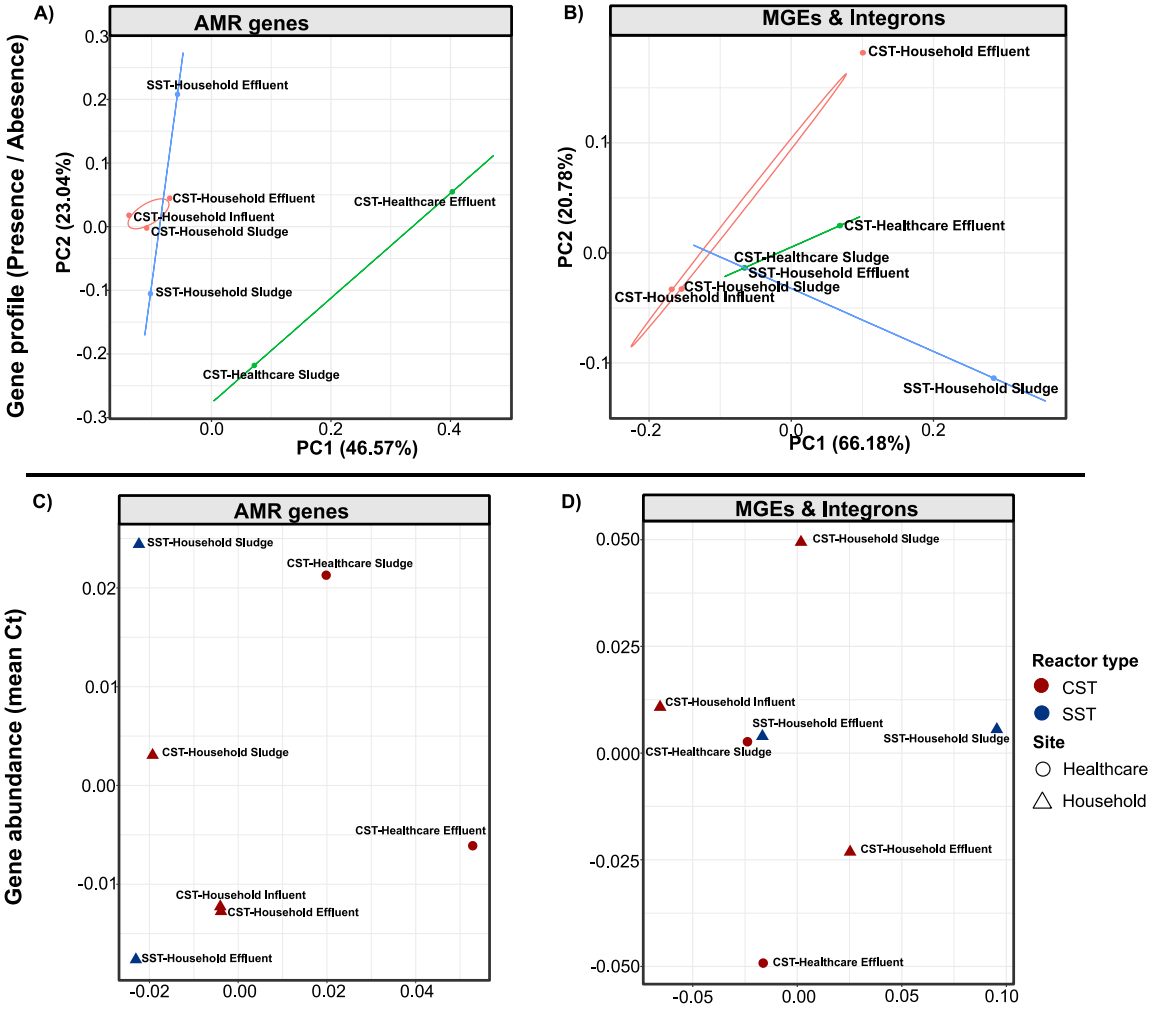

Figure 3.3: Non-metric dimensional scaling (NMDS) plot indicating similarities/ difference between AMR, MGEs and integron gene profile and abundance quantified from pooled wastewater samples on the HT-QPCR array. Pooled samples are grouped based on sample type (influent, sludge, effluent) and reactor type (CST-household, CST-healthcare, SST-household). A) AMR genes and D) mobile element (MGEs and Integrons) gene profile quantified from pooled samples. Quantified gene abundance from pooled samples reported as mean Ct (B- AMR genes, E- MGEs and Integron) and normalised gene abundance relative to 16S rRNA gene (C- AMR genes, F- MGEs and Integron). CST= Conventional septic tank; SST= Solar septic tank.

Despite observed similarities in the AMR gene profile between the CST-household samples (influent, sludge, effluent) (Figures 3.1A, 3.1B, 3.3A), only the influent and effluent sample had similar AMR gene abundance (mean Ct) as compared to the sludge (Figure 3.3C). While, for the mobile elements (Figure 3.3D), the gene abundance between the three sample types (influent, sludge, effluent) appeared to be very different from each other despite the influent and sludge having a similar gene profile. Additionally, the CST-household influent appeared to have similar mobile element gene abundance with the CST-healthcare sludge and SST-household effluent (Figure 3.3D).

The SST-household samples (sludge and effluent) had very dissimilar gene abundance, but the AMR gene abundance between the sludge and effluent showed more dissimilarities as compared to the mobile elements (Figure 3.3C). Similarly, the CST-healthcare samples (sludge and effluent) had very dissimilar AMR gene and mobile element gene abundance (Figure 3.3D).

In brief, high AMR gene and mobile element gene (including integrons) diversity and richness were quantified from pooled samples. Drug inactivation was the dominant resistance mechanism identified for all sample types (influent, sludge, effluent) across the three reactors (CST-household, CST-healthcare, SST-household). Additionally, the gene profile and gene abundance between sample types (sludge and effluent) for the CST-healthcare and SST-household tank appeared to be dissimilar, whilst the CST-household tank samples (influent, sludge, effluent) appeared to have similar AMR gene and mobile element gene profile, but very dissimilar gene abundance between the sample types (influent, sludge, effluent), although, the AMR gene abundance for the influent and effluent were the same (Figure 3.3C).

3.3.1.2 Selection of target genes (primer sets) for individual samples for HT-QPCR array quantification and data processing

Step 4: Gene selection for target on the HT-QPCR array

Owing to constraints of the HT-QPCR SmartChip configuration a trade-off between sample number and target genes on the array must be made. Here we selected 72 genes (67 AMR genes, four integrase genes (Class 1, 2 and 3 mobile resistance integron), and the *16S rRNA* gene; Appendix Table B.2) as informed from the initial pre-screen of pooled samples,

targeting 23 wastewater samples (Table 3.2). A no template control sample (nuclease-free water) was included, resulting in a total of 24 samples quantified on the HT-QPCR array.

Selected AMR genes had known association (*35/67* (52%) genes targeted) and non-association (*32/67* (48%) genes targeted) to mobile resistance integron (MRI), and conferred resistance to the major antibiotic classes (aminoglycoside, trimethoprim, β-lactam, phenicol, tetracycline, sulphonamide, multidrug resistance (MDR), macrolide-lincosamide-streptogramin B (MLSB), vancomycin, quinolone and other).

Post data processing steps, two genes (*DfrA8* (AY589) and *intI*1_1 (AY293)) were removed from all samples as the mean NTC Ct (*DfrA8*: 19.61±0SD; *intI*1_1: 16.47±0.18SD) that was similar/ lower than that of quantified samples. Of note, the *intI*1 primer removed was the same one discarded from the pooled samples. Subsequently, primers for 66 AMR genes, three integrase genes (*intI*1, *intI*2 *intI*3) and the *16S rRNA* gene were retained. The *16S rRNA* (AY1) and *intI*1 gene (AY289) assays from the HT-QPCR array were selected and used to validate the HT-QPCR array.

3.3.2 HT-QPCR Array Validation and Best Practices: The Good, The Bad and The Ugly

3.3.2.1 In-house Q-PCR validation of HT-QPCR array primers

All assays (primers) on the HT-QPCR array undergo the same Q-PCR reaction and cycling conditions, which has been suggested may be sub-optimal for some assays (Waseem et al. 2019). Therefore, two array assays targeting the *16S rRNA* (AY1) and *intI*1 gene (AY289) were used to validate the HT-QPCR. To do so, array primers were optimised in-house and used to quantify the same Thai wastewater samples (*n=23*, Table 3.2) as on the HT-QPCR array. In addition, *intI*1 primers from the previous chapter (see Chapter 2) and the standard highly cited *16S rRNA* Q-PCR TaqMan assay designed by Suzuki and colleagues (Suzuki et al. 2000) were used to cross-validate the array *16S rRNA* assay for the same gene (Table 3.4).

3.3.2.1.1 HT-QPCR array quantified higher *16S rRNA* and *intI*1 gene abundance as compared to in-house quantification for the same wastewater samples except for the influent

We compare the absolute Ct values of the samples quantified on the array to the absolute Ct values of the same samples quantified in-house for the *16S rRNA* and *intI*1 gene. The *16S rRNA* (AY1) and *intI*1 gene (AY289) assays from the array were optimised in-house and used to quantify the same Thai wastewater samples (*n=23*, Table 3.2) as on the HT-QPCR array. To permit reliable comparison between obtained Ct for the *16S rRNA* and *intI*1 gene quantified in-house Q-PCR and on the HT-QPCR array, the in-house Q-PCR Cts were corrected by adjusting the DNA concentration used in the assay to reflect the amount in the HT-QPCR array using Equation 1.

A similar Ct in the NTC was observed for the *16S rRNA* gene quantified on the HT-QPCR array (mean Ct= 26.11 ±0.4) and in-house Q-PCR (mean Ct= 26.35±0.03SD (Table 3.4)), and in-house Q-PCR melt-curve analysis showed an inability to distinguish sample melt-curve peak from that of NTC, although resulting in-house Q-PCR mean Ct for all wastewater samples were at least 2.12 log-fold higher than the Ct of the NTC. Nonetheless, the low NTC Ct quantified on the HT-QPCR array and in-house Q-PCR signifies an inability to reliably quantify low *16S rRNA* copy number using HT-QPCR *16S rRNA* primer set (AY1).

Table 3.4: Primer and probe sets selected and optimised for Q-PCR to quantify the *16S rRNA* and *intI*1 gene copies from Thai wastewater

Primer ID	Sequence (5'- 3')	Orientation	Target (length)	Assay type	Experimental Condition	QPCR Standard Curve Descriptors					Reference
						Efficiency (%)	R^2	Slope	y-Intercept	NTC (Mean Ct)	
1108F** 1132R (AY1)	ATGGYTGTCGTCAGCTCGTG GGGTTGCGCTCGTTGC	Forward Reverse	V7 Bacterial 16S *rRNA* (59 bp)	SYBR Green	**PCR:** 95°C-15min; [94°C-30sec; 60°C-30sec; 72°C-30sec] x35; 72°C-10min **Q-PCR:** 95°C-15min; [94°C-15sec; 60°-30sec; 72°-30sec, plate read] x40; Melt curve: 65°-95° (0.5° increment/5sec)	95.68	0.999	-3.43	35.95	26.35	(Lee et al., 1993; Wilmotte et al., 1993)
1369F 1389P 1492R	CGGTGAATACGTTCYCGG CTTGTACACACCGCCCGTC GGWTACCTTGTTACGACTT	Forward Probe Reverse	V9 Bacterial 16S *rRNA* (123 bp)	*Taq*Man	**Q-PCR :** 95°C-10min; [95°C-10sec; 60°C-30sec, plate read]x40; 40°C-10min	95.68	1	-3.43	38	34.46	(Suzuki et al., 2000)
*intI1*_1** (AY293)	CGAACGAGTGGCGGAGGGTG TACCCGAGAGCTTGGCACCCA	Forward Reverse	*intI*1 gene (312 bp)	SYBR Green	**PCR:** 95°C-15min; [94°C-30sec; 60°C-30sec; 72°C-30sec] x35; 72°C-10min **Q-PCR:** 95°C-15min; [94°C-15sec; 60°-30sec; 72°-30sec, plate read] x40; Melt curve: 65°-95° (0.5° increment/5sec)	92.71	1	-3.51	35	38.01	(Gillings et al., 2015)
*intI1*_2** (AY289)	CGAAGTCGAGGCATTTCTGTC GCCTTCCAGAAAACCGAGGA	Forward Reverse	*intI*1 gene (217 bp)	SYBR Green	**PCR:** 95°C-15min; [94°C-30sec; 60°C-30sec; 72°C-30sec] x35; 72°C-10min **Q-PCR:** 95°C-15min; [94°C-15sec; 60°-30sec; 72°-30sec, plate read] x40; Melt curve: 65°-95° (0.5° increment/5sec)	95.3	0.999	-3.44	34.5	40.21	(Muziasari et al., 2014)
intI1-DF intI1-DR intI1-MGB	TTCTGGAAGGCGAGCATC TGCCGTGATCGAAATCC Fam-TGACCCGCAGTTGCA-MGB Eclipse	Forward Reverse Probe	*intI*1 (108bp)	MGB *Taq*Man probe	**Q-PCR:** 95°-10min; [95°-30sec; 60°-60sec, plate read]x45	96.06	1	-3.42	37.23	39.53	This study
F3 R3	TTTCTGGAAGGCGAGCATCGTTTG TGCCGTGATCGAAATCCAGATCCT	Forward Reverse	*intI*1 (109bp)	SYBR Green	**Q-PCR:** 95°-15min; [94°-15sec; 65°-30sec; 72°-30sec, plate read] x40; Melt curve: 65°-95° (0.5° increment/5sec)	92.35	0.999	-3.52	36.06	0	(Rosewarne et al., 2010)
F7 R7 F7-probe	GCCTTGATGTTACCCGAGAG GATCGGTCGAATGCGTGT 6Fam-ATTCCTGGCCGTGGTTCTGGGTTTT-BHQ1	Forward Reverse Probe	*intI*1 (196bp)	*Taq*Man probe	**Q-PCR :** 95°-10min; [95°-30sec; 60°-60sec, plate read]x45	92.71	1	-3.51	38.27	40.45	(Barraud et al., 2010)

** denotes HT-QPCR array primer set of which ***intI*1_1(AY293)** corresponds to **F10-R10** and ***intI1*_2 (AY289)** correspond to **F4-R4**

There was no statistical difference (p-value >0.05) between the *16S rRNA* gene (AY1) absolute Cts quantified in-house and on the array for the same samples (influent, sludge, effluent) across reactors (CST-household, CST-healthcare, SST-household) (Table 3.5A). Similarly, no statistical difference (p-value >0.05) in *intI*1 gene abundance was observed between absolute Cts for the same sample (influent, sludge effluent) quantified in-house and on the array across reactors (Table 3.5B).

Table 3.5: One-way ANOVA analysis of in-house and HT-QPCR *16S rRNA* (A) and Kruskal Wallis analysis of in-house and HT-QPCR *intI*1 (B) quantification of the same sample type (influent, sludge, effluent) across the three reactors (CST-household, CST-healthcare, SST-household)

A	***16S rRNA* (AY1) gene abundance (mean Ct) quantified in-house and on HT-QPCR array**	***p-value***
	CST-household-Influent	0.533
	CST-household-Sludge	0.31
	CST-household-Effluent	0.491
	CST-healthcare-Sludge	0.957
	CST-healthcare-Effluent	NA
	SST-household-Sludge	0.431
	SST-household-Effluent	0.264
B	***intI*1 (AY289) gene abundance (mean Ct) quantified in-house and on the HT-QPCR array**	***p-value***
	CST-household-Influent	0.827
	CST-household-Sludge	0.387
	CST-household-Effluent	0.564
	CST-healthcare-Sludge	0.439
	CST-healthcare-Effluent	NA
	SST-household-Sludge	0.2
	SST-household-Effluent	0.827

For *16S rRNA*, *both* array and in-house quantification indicated no statistical difference in 16S rRNA Ct values between sample types (influent, effluent, sludge) within each of the tanks (CST-household, CST-healthcare, SST-household). Nonetheless, comparing Ct values between sample types and reactors, both array and in-house Ct values for 16S rRNA were found to be statistically different (p-value >0.05) between CST-household sludge and SST-household sludge (Figure 3.6A).

In the case of the *intI*1 gene, neither the absolute Ct values quantified on the array nor in-house were significantly different (p-value >0.05) between the sample types for each of the three reactor types. Additionally, no significant difference (p-value >0.05) in Ct values quantified on the array and in-house was observed when comparing between sample types and reactors (Figure 3.6B).

Pearson correlation coefficient analysis indicated that both *16S rRNA* and *intI*1 gene abundance quantified in-house and on the HT-QPCR array were highly correlated (*16S*

rRNA r=0.81 (p-value <0.001); *intI*1 r=0.909 (p-value <0.001)). As such, each primer set produced the same overall pattern in gene abundance in-house QPCR and HT-QPCR array quantification.

However, generally higher gene abundance (inferred by absolute Ct values) was reported by the HT-QPCR array, except for CST-household influent, which was higher in-house than on the array (Figure 3.4). The difference between the in-house and the array for the other samples (sludge and effluent) between the tanks (CST-household, CST-healthcare, SST-household) was always less than 0.15 log for the *16S rRNA* gene and less than 0.25 log for the *intI*1 gene. Furthermore, a fitted linear regression relationship model of gene abundance (absolute Ct) quantified in-house and on the array estimated a 0.29 log difference for the *16S rRNA* gene (adjusted R2= 0.6, y-intercept= 0.97 (p-value > 0.05), slope = 0.94 (p-value <0.001)) and 1.06 log-fold difference for the *intI*1 gene (adjusted R2= 0.82, y-intercept= -3.54 (p-value > 0.05), slope =1.21 (p-value <0.001)).

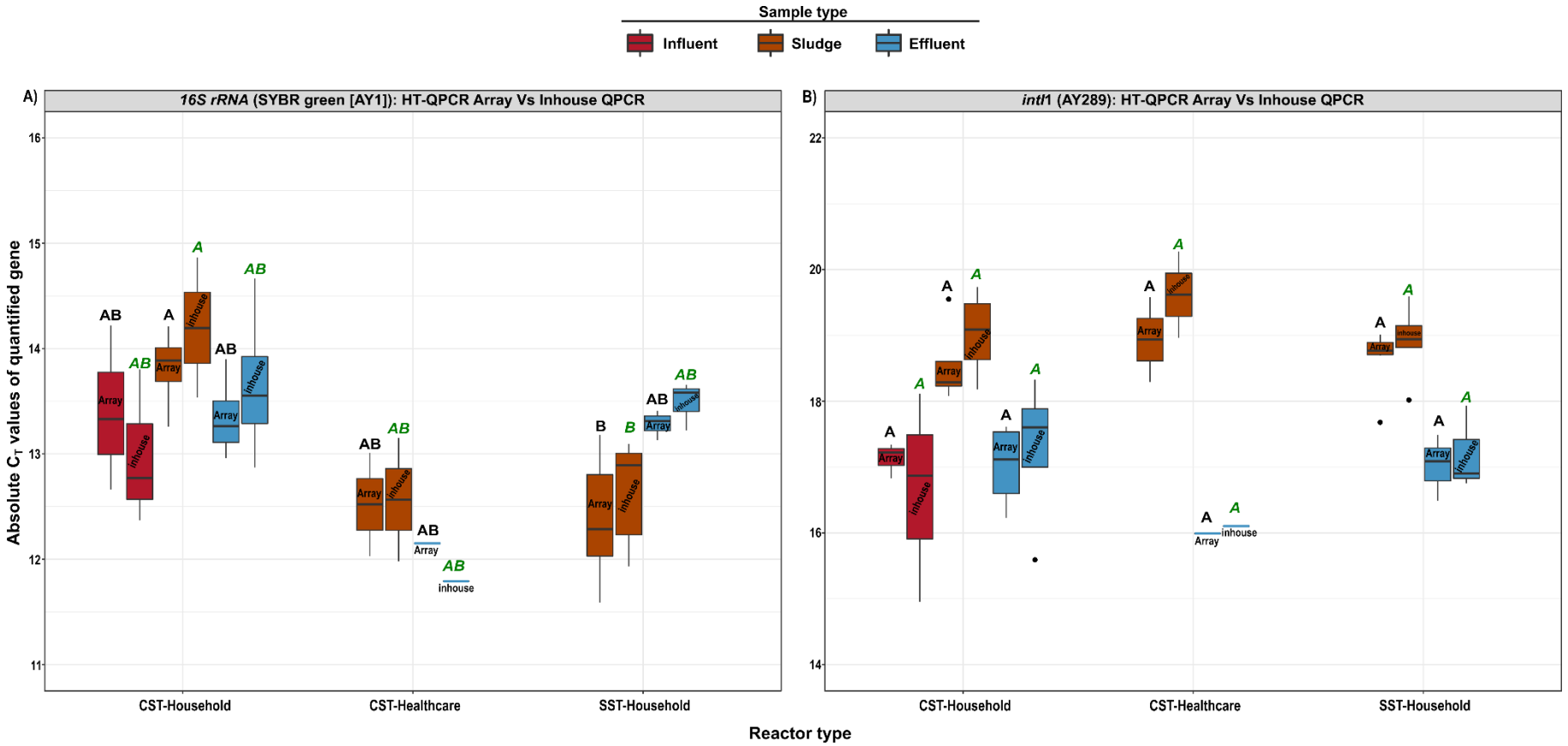

Figure 3.4: Comparison of absolute Ct values quantified on the HT-QPCR array and in-house for the *16S rRNA* A) and *intI*1 B) gene target from the same Thai wastewater samples (*n=23*, Table 3.2). Results of a 2-way ANOVA or Kruskal Wallis analysis (array *intI*1) showing the effect of sample type (influent, sludge, effluent) and tank type (CST-household, CST-healthcare, SST-household) on quantified Ct values using HT-QPCR array *16S rRNA* and *intI*1 primer sets to validate the HT-QPCR array. Black dot represents outliers. Boxplot sharing the same letter indicates no statistically significant difference at p-value >0.05, while boxplot with different letters indicates statistically significant difference at p-value <0.05. Boxplot letters in *green italics* show the statistical difference for the in-house quantification. CST denotes conventional septic tank; SST denotes solar septic tank.

To summarise, HT-QPCR generally quantified higher gene abundance (absolute Ct) for both the *16S rRNA* and *intI*1 gene as compared to in-house-QPCR, though the higher quantification observed was <0.15-log higher for the 16S rRNA but up to 1.06-log higher for the *intI*1 (according to fitted linear regression model between Ct obtained in-house and on the HT-QPCR array). Thus, it implies that conditions within the array may be appropriate for *16S rRNA* assay but not *intI*1.

3.3.2.1.2 HT-QPCR array *16S rRNA* primer (AY1) quantified up-to a log higher *16S rRNA* abundance compared to the TaqMan *16S rRNA* primer-probe set for the same samples quantified in-house

In this section, we used a well-described and validated *16S rRNA* TaqMan primer-probe set from the literature (Suzuki et al. 2000) and the HT-QPCR array *16S rRNA* SYBR green primer set (AY1) (Lee et al. 1993; Wilmotte et al. 1993) to quantify the same wastewater samples quantified on the array in-house and compared resulting mean Ct and copy numbers to assess how choice of primer impact quantification of the *16S rRNA* gene.

Standard curves for each gene had similar slopes and efficiencies -3.43, and 95.68%; y-intercepts of 35.95 to 38.00, and No Template Control Ct of 26.35 (AY1) and 34.46 (TaqMan) (Table 3.4). A similar overall trend in Ct values (Figure 3.5 A) and absolute gene abundance (copies/ml or g, Figure 3.5B) was observed for each primer set (SYBR (AY1) and TaqMan), used to quantify *16S rRNA* gene in-house for the same sample type and reactor. Subsequently, Pearson correlation (absolute Ct: r=0.991, p-value <0.001) and Spearman rank correlation (absolute copies/ml or g: r=0.993, p-value <0.001) analysis indicated that the gene abundance quantified in-house by the *16S rRNA* primers (TaqMan and SYBR) were highly correlated (Appendix Table B.5).

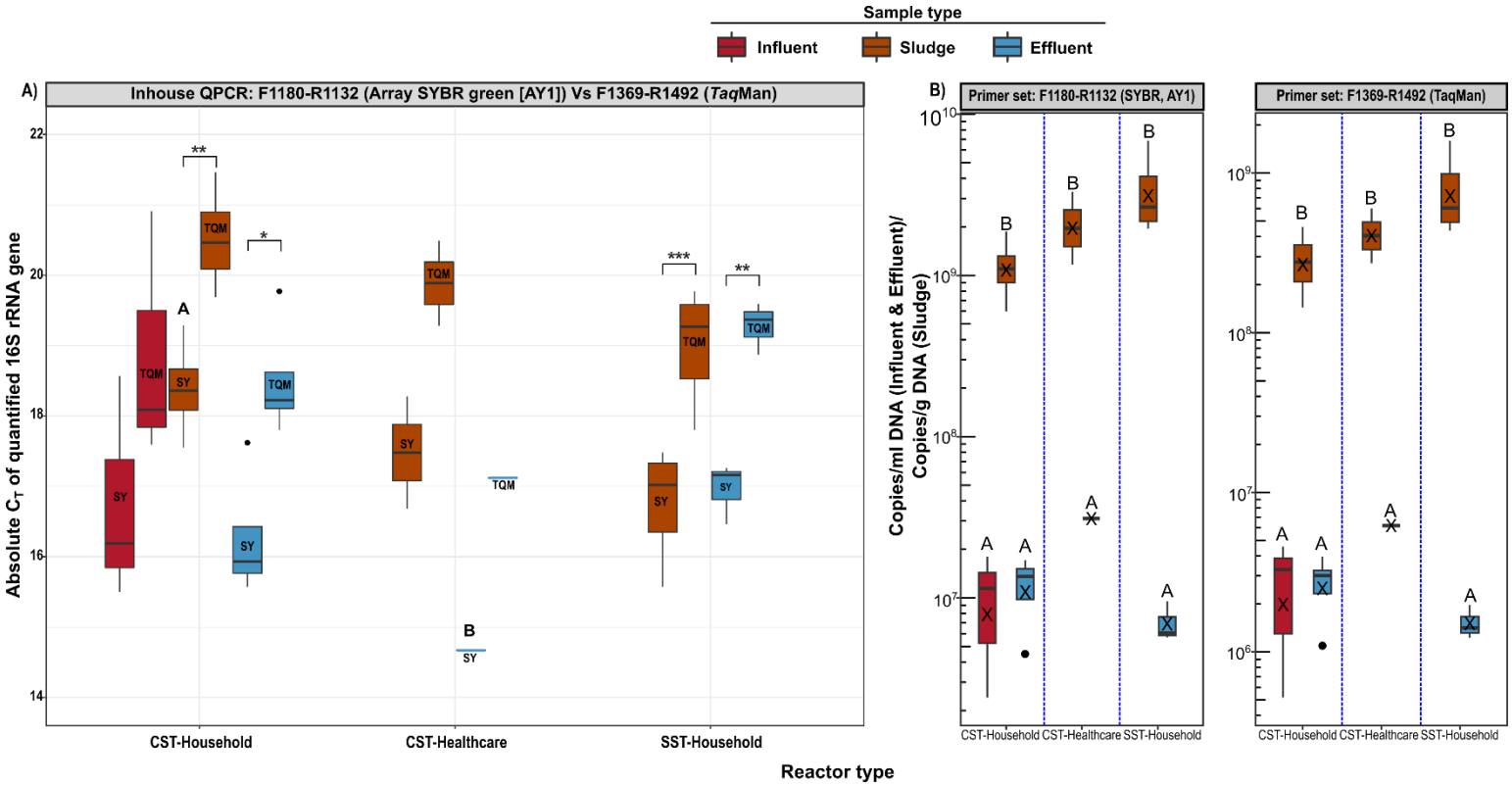

Figure 3.5: Comparison of *16S rRNA* gene abundance quantified in-house using HT-QPCR array *16S rRNA* primer (SYBR green- AY1) and validated TaqMan *16S rRNA* primer. A) Comparison of gene abundance inferred by absolute Ct quantified with both *16S rRNA* primer sets (SYBR and TaqMan). B) Absolute gene abundance (copies/ml or copies/g) quantified by the two *16S rRNA* primer sets. Boxplot with no letter indicates no statistically significant difference at p-value>0.05, while boxplot with different letters indicates a statistically significant difference at p-value <0.05. A statistically significant difference in *16S rRNA* gene abundance between primer sets for the same sample was observed and indicated by significant stars at the top (* p-value <0.05, ** p-value <0.01, *** p-value <0.001). CST denotes conventional septic tank; SST denotes solar septic tank; TQM denotes TaqMan; SY denotes SYBR green.

Nonetheless, irrespective of the reporting method (absolute Ct/ copies/ml or g), statistical differences (p-value <0.05) within the same sample type were observed when the different primer sets (SYBR green vs. TaqMan). Specifically, statistically significant differences were observed in the CST-household and SST-household unit sludge and effluent samples for both reporting methods, while no statistically significant differences were observed for the CST-healthcare tank samples (Figure 3.5, Table 3.6).

Table 3.6: One-way ANOVA analysis of in-house *16S rRNA* quantification using the well-validated TaqMan assay and HT-QPCR array SYBR green assay (AY1) for the same sample type (influent, sludge, effluent) across the three reactors (CST-household, CST-healthcare, SST-household)

TaqMan (F1369-R1492) and SYBR green (AY1- F1180-R1132) *16s rRNA* gene abundance (inferred by mean Ct) quantified in-house	***p-value***
CST-household-Influent	0.204
CST-household- Sludge	0.00635**
CST-household - Effluent	0.0122*
CST-healthcare-Sludge	0.139
CST-healthcare- Effluent	NA
SST-household- Sludge	0.000535***
SST-household-Effluent	0.00216**
TaqMan (F1369-R1492) and SYBR green (AY1- F1180-R1132) *16s rRNA* gene abundance (copies/ml or copies/g) quantified in-house	***p-value***
CST-household-Influent	0.17
CST-household- Sludge	0.0181*
CST-household - Effluent	0.015*
CST-healthcare-Sludge	0.239
CST-healthcare- Effluent	NA
SST-household- Sludge	0.0081**
SST-household-Effluent	0.0155*

*p-value * <0.05, ** <0.01, *** <0.001*

Higher gene abundance (inferred by mean Ct (Figure 3.5A) and absolute gene copies/ml or g (Figure 3.5 B)) were generally quantified by array primer set (SYBR green assay (AY1)). For gene abundance inferred by the absolute Ct, almost a log-fold higher *16S rRNA* abundance was quantified by array primer set than (SYBR green assay (AY1)) by the TaqMan *16S rRNA* probe set [e.g., CST-household (influent: 0.7, sludge: 0.64, effluent: 0.67- log-fold higher); CST-healthcare (sludge: 0.73 effluent: 0.74-log-fold higher); SST-household (sludge: 0.67, effluent: 0.7- log-fold higher)] (Figure 3.5A). This log-fold difference between the same sample quantified with the different assays (SYBR green verses TaqMan) was confirmed with statistically significant fitted linear regression (Adjusted R2=0.982, y-intercept= 2.644 (p-value <0.001), slope= 0.976 (p-value <0.001)). Additionally, only array primer set (SYBR green (AY1)) reported statistical differences in Ct values between sample type and reactors, specifically between the CST-household sludge

and CST-healthcare effluent (Figure 3.5A). In contrast, no statistically significant difference in Ct was observed between sample type and tank type for the TaqMan primer set (Figure 3.5A). Furthermore, both primer sets reported significant differences in gene *16S rRNA* gene abundance (copies per ml or g) between influent, sludge and effluent across the reactors.

3.3.2.1.3 Normalising AMR gene Cts with *16S rRNA* changes the gene abundance reported between samples (influent, sludge, effluent)

HT-Q-PCR array data findings are often reported in the literature as relative Ct normalised to the Ct of the *16S rRNA* gene from the same sample. However, with the varying *16S rRNA* copy number (1-15 copies) per bacterial genome, normalisation of target gene abundance to the abundance of *16S rRNA* gene can bias the estimation of total microbial abundance based on Q-PCR methods and can mask real differences in *16S rRNA* gene abundance and thus, affect the abundance of normalised target genes (Smith and Osborn 2009; Větrovský and Baldrian 2013; Angly et al. 2014).

Therefore, in this section, we investigated the effect of gene abundance normalisation (normalised AMR gene Ct to *16S rRNA* Ct) on interpretations of observed results, by comparing the normalised *intI*1 relative abundance quantified on the array and in-house to the *intI*1 abundance reported as mean Ct, as discussed in the above section (see section 3.3.2.1.1). There was no statistical difference in the absolute Ct of *intI*1 (Table 3.5B) and the normalised *intI*1 Ct (Table 3.7) between array and in-house quantification for the same sample (influent, sludge effluent) (p-value >0.05).

Table 3.7: One-way ANOVA analysis of normalised *intI*1 abundance (normalised to array *16S rRNA* (AY1)) quantified on the HT-QPCR and in-house for the same sample type (influent, sludge, effluent) across the three reactors (CST-household, CST-healthcare, SST-household)

Normalised intI1 abundance (normalised to array 16S rRNA) quantified on the HT-QPCR and in-house	*p-value*
CST-household-Influent	0.831
CST-household-Sludge	0.8
CST-household-Effluent	0.812
CST-healthcare-Sludge	0.0844
CST-healthcare-Effluent	NA
SST-household-Sludge	0.884
SST-household-Effluent	0.965

Neither the absolute Ct (Figure 3.6A) nor normalised *intI*1 gene relative abundance (Figure 3.6B), quantified using the array and in-house, showed any statistical significance difference (p-value >0.05) between sample types (influent, effluent, sludge) within each reactor (CST-household, CST-healthcare, SST-household), nor between sample type and reactors.

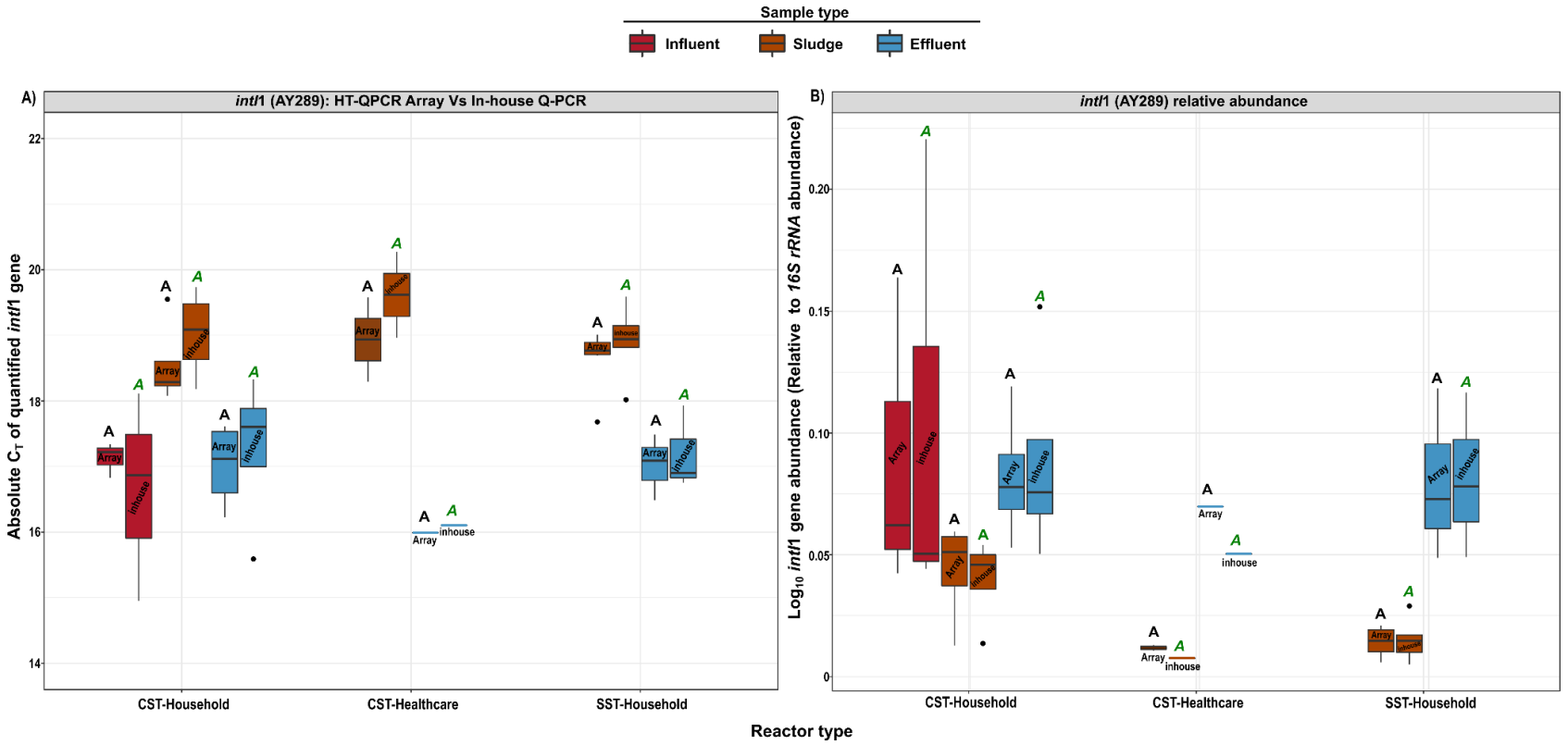

Figure 3.6: Comparison of quantified *intI*1 gene abundance on the HT-QPCR array and in-house for the same Thai wastewater samples (*n=23*, see Table 3.2). Result of a 2-way ANOVA comparing *intI*1 absolute Ct A) and normalised *intI*1 (normalised to *16S rRNA* gene Ct) B) between sample type (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household). *intI*1 gene absolute Ct quantified on the HT-QPCR array was normalised to the *16S rRNA* gene Ct on the HT-QPCR array. Similarly, *intI*1 gene absolute Ct quantified in-house was normalised to the *16S rRNA* gene Ct quantified in-house. Black dots indicate outliers. Boxplots sharing the same letter indicate no statistical difference. Boxplot letters in *green italic* show the statistical difference for the in-house quantification. No Statistically significant difference (p-value> 0.05) in *intI*1 absolute Ct between sample types (influent, sludge, effluent) and reactors (CST-Household, CST-Healthcare, SST-Household) was observed for the array and in-house quantification A). No statistically significant difference (p-value> 0.05) was observed for normalised *intI*1abundance between samples (influent, sludge, effluent) and reactors quantified on the array and in-house. B). CST denotes conventional septic tank; SST denotes solar septic tank.

For the most part, similar trends were observed between absolute Ct and relative abundance (normalised Ct to the *16S rRNA* Ct) (Figure 3.6). However, on the array, the absolute Ct indicated that *intI*1 in the SST-household effluent (17.02±0.50SD) was higher than the CST-household influent (17.13±0.27SD) (Figure 3.6B). Conversely, the normalised relative abundance (to the *16S rRNA*) showed *intI*1 in the CST-household influent ($8.94x10^{-2}$±$6.52x10^{-2}$SD) to be greater than the SST-household effluent ($7.99x10^{-2}$±$3.53x10^{-2}$SD) (Figure 3.6A).

Furthermore, for both in-house and the array data, absolute Ct abundance indicated higher *intI*1 gene abundance in the CST-healthcare effluent (array: 15.99±0SD, in-house: 16.1±0SD) than in the CST-household influent (array: 17.13±0.27SD, in-house: 16.64±1.59SD) (Figure 3.6B). However, normalised *intI*1 gene relative abundance was higher in the CST-household influent (array: 8.94x10-2±6.52x10-2SD, in-house: 1.05x10-1±1.00x10-1SD) than CST-healthcare effluent (array: 6.98x10-2±0SD, in-house: 5.04x10-2±0SD) Figure 3.6A).

Additionally, absolute Ct indicated higher *intI*1 in the CST-healthcare effluent (array: 15.99±0SD, in-house: 16.1±0SD) than in the CST-household effluent (array: 17.02±0.66SD, in-house: 17.28±1.18SD) and SST-household effluent (array: 17.02±0.50SD, in-house: 17.19±0.64SD) (Figure 3.6B). However, normalised *intI*1 gene relative abundance was higher in the CST-household effluent (array: 8.19x10-2±2.76x10-2SD, in-house: 8.84x10-2±4.40x10-2SD) and SST-household effluent (array: 7.99x10-2±3.53x10-2SD, in-house: 8.12x10-2±3.39x10-2SD) than in the CST-healthcare effluent (array: 6.98x10-2±0SD, in-house: 5.04x10-2±0SD) (Figure 3.6A).

In summary, although a similar trend in *intI*1 gene abundance was observed between samples (influent, effluent, sludge) and reactors (CST-household, CST-healthcare, SST-household) when gene abundance was reported an absolute Ct and normalised to the *16S rRNA* gene Ct, the presence of varying *16S rRNA* gene copies amongst bacteria taxa resulted in changes to gene abundance reported between samples. For example, mean Ct reported higher *intI*1 abundance in effluent, while normalised abundance indicated higher abundance for the influent.

Moreover, coupled with the strong signal observed in the NTC of the SYBR green assay, and the double melt curve peak on the HT-QPCR array for the *16S rRNA* gene, the results from this section clearly highlight how normalising Ct values of target genes to the *16S rRNA* gene can alter interpretation of result. In this context, highlights how normalised Ct values

(to *16S rRNA* Ct) alter the reactor type and sample type posing the most risk of potential CL1-intregon spread to the environment.

That said, we processed forward with normalised values rather than reporting absolute Ct's. Note absolute results are also included in the appendix section (Appendix B). This decision stemmed from practical constraints, which include time limitation, cost consideration and limited available sample volumes, preventing the re-quantification of the samples on the array using a different well-validated *16S rRNA* primer set or a single copy gene.

Nonetheless, it is critical to be aware of the potential limitations of normalising target gene Ct to that of the *16S rRNA* Ct, especially when normalising to *16S rRNA* Ct quantified by array *16S rRNA* primer set (AY1).

3.3.3 Application of HT-QPCR array: Risk assessment of septic tanks in disseminating AMR genes and integrases (*intI*1, *intI*2, *intI*3) to the environment

3.3.3.1 Quantification of AMR genes and MGEs within Thai Septic Tanks

Temporal (sampling months) of conventional and solar septic tanks showed that the richness and diversity of AMR genes were similar in all wastewater samples based on the selected target genes, except for SST-household sludge (ST-07_11-18) and CST-healthcare effluent (CT-HC2_08-19) (Figure 3.7A, Appendix B.1). These had a slightly lower richness and diversity owing to the absence of *aacC2* (Aminoglycoside N-acetyltransferase resistance gene encoding antibiotic inactivation resistance mechanism) and *dfrB* (Trimethoprim dihydrofolate reductase resistance gene which encodes antibiotic target replacement resistance mechanism) resistance genes.

Indeed, a higher gene richness across all wastewaters could have been obtained by targeting more genes on the array, as evidenced by the pre-screen run. However, associated drawbacks of the array including high per-array run cost and limited accessibility of the array at present (Waseem et al. 2019; Liguori et al. 2022), means a trade-off between sample number and gene targets were carefully considered resulting in reduced number of genes targeted for the number of wastewater samples used.

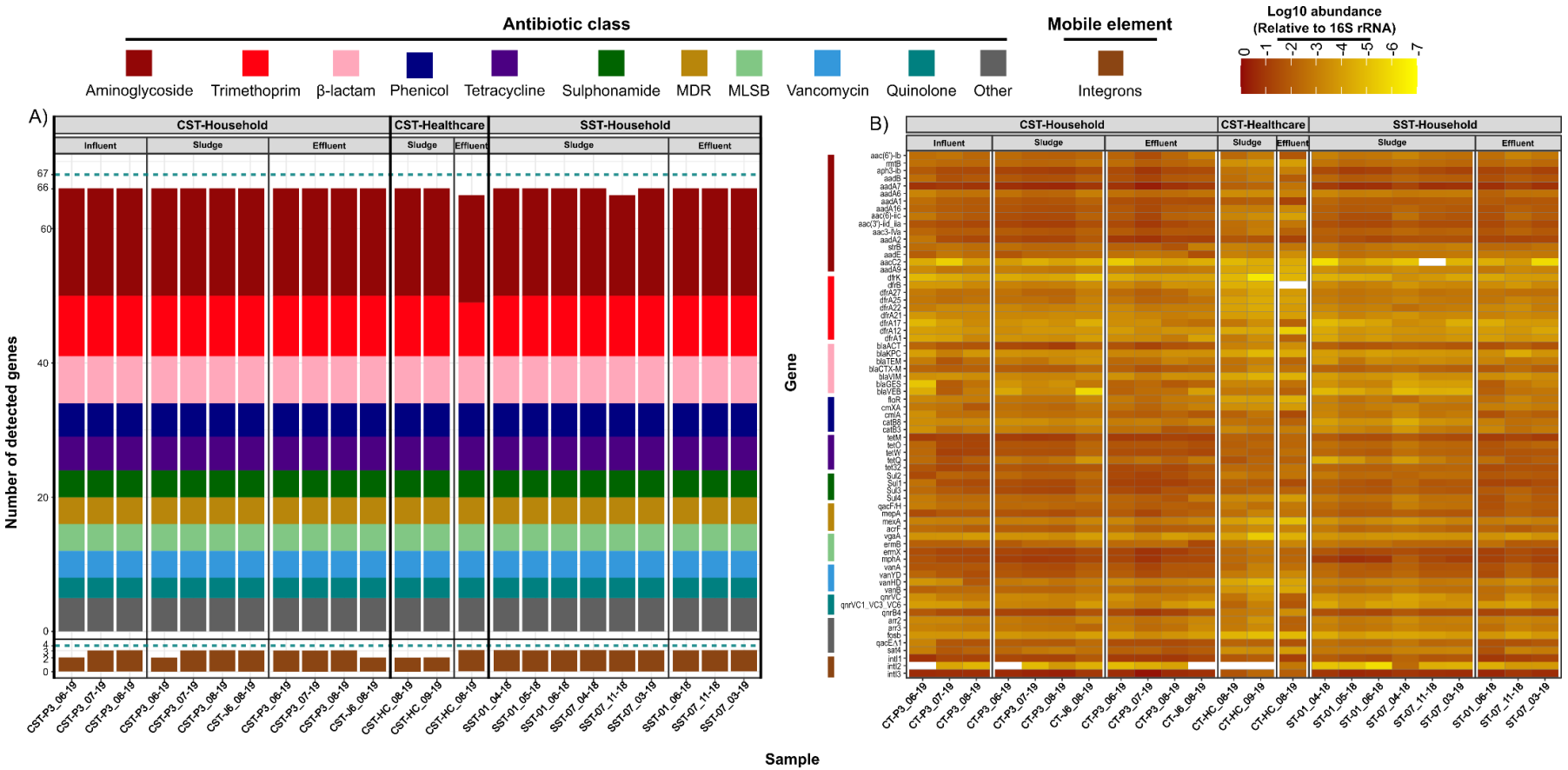

Figure 3.7: Quantified AMR genes and MGEs on the HT-QPCR array for the individual wastewater samples (*n=23*) grouped by reactor type (CST-Household, CST-Healthcare and SST-Household) and sample type (influent, sludge, effluent). A) Total number of detected genes for each sample and B) gene abundance (relative to *16S rRNA* gene abundance) for each sample, with each colour representing a different antibiotic class/ mobile element. Teal dash-line shows total gene count targeted per sample. CST= Conventional Septic Tank; SST= Solar Septic Tank. CTP3 and CTJ6 samples originated from two independent CST-Household reactors. CT-HC sample was from a CST-Healthcare tank. ST01 and ST07 are two independent SST-Household units. Sampling month and year are represented as month_year (e.g., 06_19 for June 2019). CST, Conventional septic tank; SST, Solar septic tank.

The relative gene abundance values (relative to the *16S rRNA* gene) indicated higher overall AMR genes in the effluent (mean relative abundance: $1.59x10^{-2}$±$3.31x10^{-2}$SD) > influent (mean relative abundance: $1.20x10^{-2}$±$2.15x10^{-2}$SD) > sludge (mean relative abundance: $1.07x10^{-2}$±$2.67x10^{-2}$SD) (Appendix Tables B.4). In addition, quantified AMR genes conferred resistance to all major class of antibiotics and covered the five major resistance mechanisms with the dominant resistance mechanism indicated as drug inactivation > target replacement > target protection > drug efflux > target alteration. (Figure 3.8).

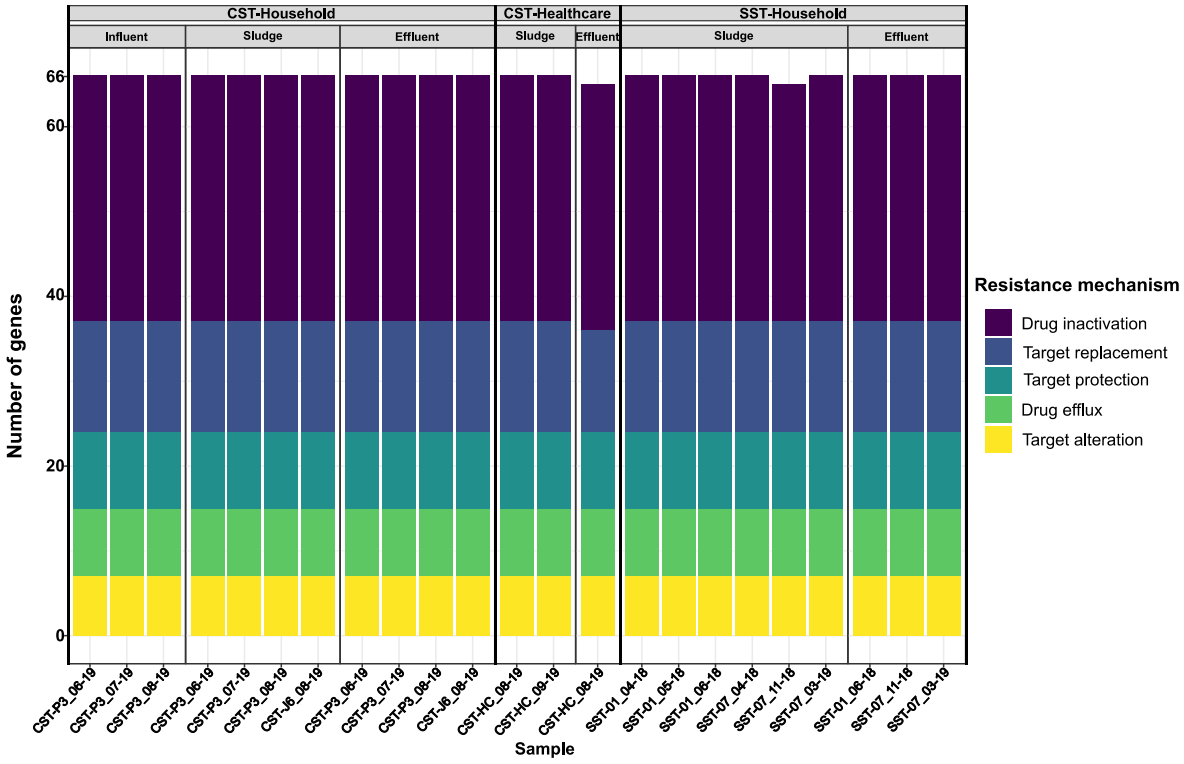

Figure 3.8: Resistance mechanisms of quantified AMR genes on the HT-QCR array from individual Thai wastewater samples (*n=23*). CST denotes conventional septic tank; SST denotes solar septic tank.

Overall, the most abundant (relative to the *16S rRNA* gene) resistance genes was MLSB ($2.91x10^{-2}$±$4.0x10^{-2}$SD) > Aminoglycoside ($2.28x10^{-2}$±$4.62x10^{-2}$SD) > Tetracycline ($2.18x10^{-2}$±$2.43x10^{-2}$SD) > Sulphonamide ($1.68x10^{-2}$±$1.76x10^{-2}$SD) > Quinolone ($1.30x10^{-2}$±$2.30x10^{-2}$SD) > Vancomycin ($8.78x10^{-3}$±$1.02x10^{-2}$SD) > MDR ($6.73x10^{-3}$±$7.74x10^{-3}$SD) > β-lactam ($5.05x10^{-3}$±$1.01x10^{-2}$SD) > Other ($4.23x10^{-3}$±$7.03x10^{-3}$SD) > Phenicol ($2.73x10^{-3}$±$4.63x10^{-3}$SD) > Trimethoprim ($8.69x10^{-4}$±$1.37x10^{-3}$SD) (Figure 3.7B, Appendix B.2A).

Similarly, for the integrases, the most abundant gene was *intI*3 ($2.20x10^{-1}$±$1.52x10^{-1}$SD) > *intI*1 ($5.34x10^{-2}$±$4.25x10^{-2}$SD) > *intI*2 ($3.54x10^{-4}$±$9.83x10^{-4}$SD).

3.3.3.1.1 Risk assessment between the three tanks (CST-household, CST-healthcare, SST-household)

Of the three reactors, the CST-healthcare reactor consistently had the lowest AMR gene abundance (relative to the *16S rRNA* gene abundance) in the sludge (mean AMR gene: $2.33x10^{-3}\pm3.52x10^{-3}$SD) and effluent (mean AMR gene: $5.78x10^{-3}\pm1.14x10^{-2}$SD) sample as compared to the CST-household (sludge: $1.63x10^{-2}\pm3.38x10^{-2}$SD, Effluent: $2.02x10^{-2}\pm1.14x10^{-2}$SD) and SST-household (sludge: $9.71x10^{-3}\pm2.47x10^{-2}$SD, Effluent: $1.36x10^{-2}\pm2.38x10^{-2}$SD) reactors. This suggests that the CST-healthcare reactor was the least contributor of AMR genes to the environment via sludge and effluent. On the other hand, the CST-household samples (sludge and effluent) consistently had the highest relative AMR gene abundance among the three reactors; suggesting that it was the highest contributor of AMR gene abundance via its sludge and effluent into the environment.

Similarly, for integrases, the CST-healthcare samples (sludge and effluent) had the lowest *intI*1 (sludge:$1.21x10^{-2}\pm3.47x10^{-2}$SD; effluent: $7.04x10^{-2}\pm0$SD) and *intI*3 (sludge:$1.22x10^{-2}\pm1.05x10^{-2}$SD; effluent: $8.87x10^{-3}\pm0$SD) abundance (relative to the *16S rRNA* gene) among the three reactors. This suggests the CST-healthcare unit as the lowest contributor of CL1-and-CL3 integrons to the environment.

In contrast, the CST-household unit samples (sludge and effluent) had the highest *intI*1 (sludge:$4.31x10^{-2}\pm2.14x10^{-2}$SD, Effluent: $9.02x10^{-2}\pm2.73x10^{-2}$SD) and *intI*3 (sludge:$2.91x10^{-1}\pm1.41x10^{-1}$SD, Effluent: $3.42x10^{-1}\pm2.53x10^{-1}$SD) gene abundance among the three reactors; suggesting that the CST-household unit played a more significant role as the highest contributor of CL1-and-CL3 integrons to the environment.

Finally, with no *intI*2 gene detected in the CST-healthcare sludge, the SST-household sludge ($6.80x10^{-4}\pm3.19x10^{-5}$SD) had the highest *intI*2 gene abundance (relative to the *16S rRNA* gene) among the three reactors. This indicated that the SST-household unit was a higher contributor of CL2-integron via sludge of the three reactors.

In the case of the effluent, *intI*2 gene abundance was highest in the CST-healthcare effluent ($1.60x10^{-3}\pm0$SD) and lowest in the CST-household effluent ($3.86x10^{-5}\pm3.19x10^{-5}$SD), indicating that the CST-healthcare effluent was the most contributor of CL2-intergon to the environment and the CST-household effluent the least contributor, among the three reactors.

3.3.3.1.2 Risk assessment within each septic tank (CST-household, CST-healthcare, SST-household) samples (Influent, Sludge and Effluent)

Within each septic tank unit (CST-household, CST-healthcare, SST-household), no statistical difference in gene abundance (relative to the *16S rRNA* gene) was observed between sludge and effluent for all targeted antibiotic classes expected for tetracycline class in the SST-household sludge and effluent (Figure 3.9).

In addition, overall AMR gene abundance (relative to the *16S rRNA* gene) was higher in the effluent (CST-household: $2.02x10^{-2}\pm4.10x10^{-2}$SD, CST-healthcare: $5.78x10^{-3}\pm1.14x10^{-2}$SD, SST-household: $1.36x10^{-2}\pm2.38x10^{-2}$SD) than sludge (CST-household: $1.63x10^{-2}\pm3.38x10^{-2}$SD, CST-healthcare: $2.33x10^{-3}\pm3.52x10^{-3}$SD, SST-household: $9.71x10^{-3}\pm2.47x10^{-2}$SD) for each tank.

Of the different antibiotic classes, resistance genes from the MLSB class dominated as the most abundant in the household tanks (CST-household and SST-household) sludge (CST-household: $3.57x10^{-2}\pm3.75x10^{-2}$SD and SST-household: $3.38x10^{-2}\pm5.02x10^{-2}$SD) and effluent (CST-household: $4.24x10^{-2}\pm5.18x10^{-2}$SD and SST-household: $2.80x10^{-2}\pm2.77x10^{-2}$SD) samples. Meanwhile, for the CST-healthcare units, tetracycline and sulphonamide resistance genes were most abundant in the sludge ($7.91x10^{-3}\pm4.23x10^{-3}$SD) and effluent ($9.79x10^{-3}\pm1.42x10^{-2}$SD) respectively. Additionally, tetracycline resistance was the most abundant in the CST-household influent ($3.89x10^{-2}\pm2.66x10^{-2}$SD).

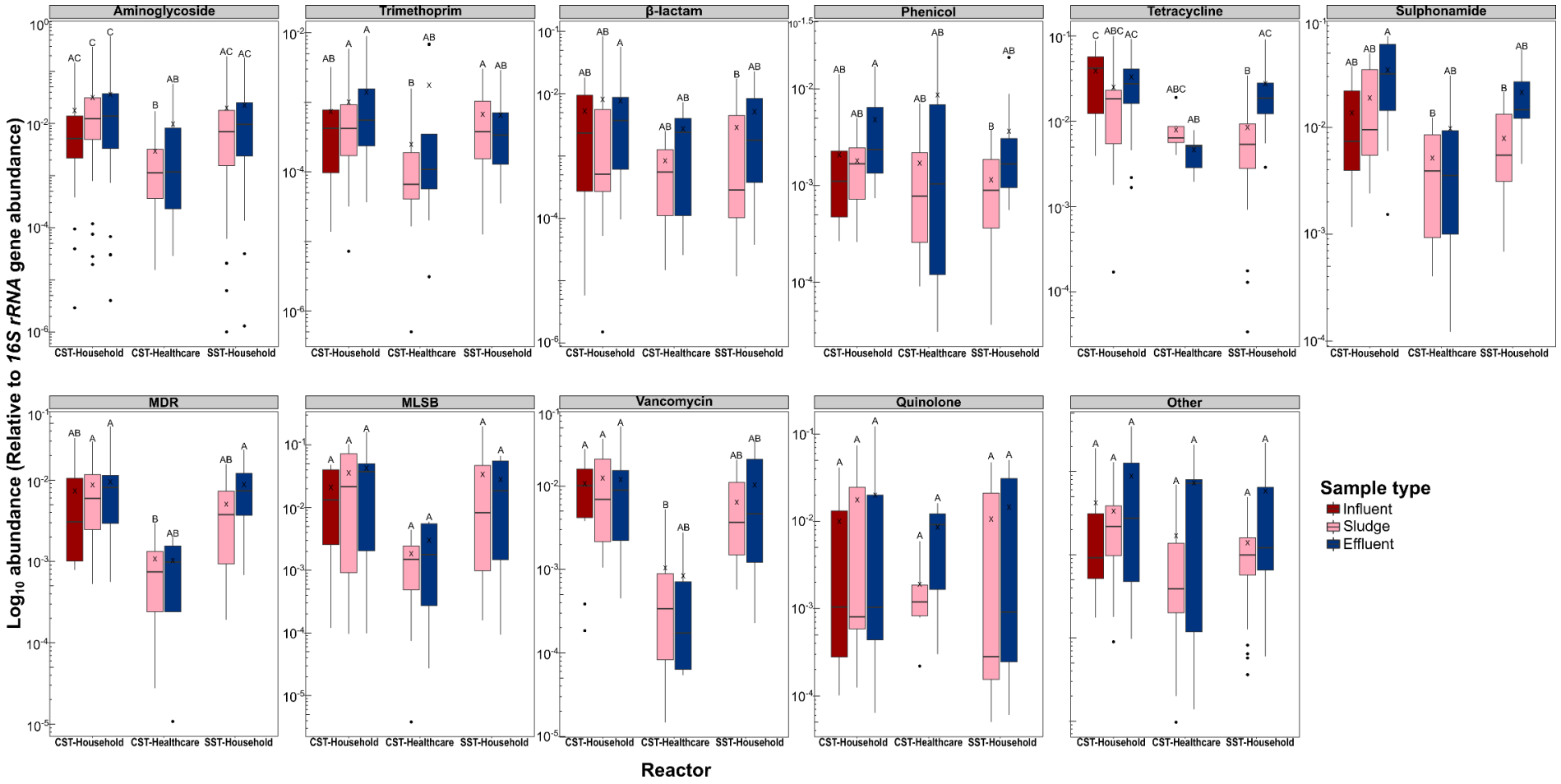

Figure 3.9: Comparison of relative AMR gene abundance between samples (influent, sludge, effluent) and reactor type (CST-Household, CST-Healthcare, SST-Household) quantified on the HT-QPCR. X icon within each boxplot indicates the mean. Black dots represent outliers. Boxplot sharing the same letter indicates no statistically significant difference at p-value >0.05. A statistically significant difference (p-value> 0.05) between samples and reactors was only observed for the antibiotic classes: Aminoglycoside, Trimethoprim and Vancomycin. CST= Conventional septic tank; SST= Solar septic tank.

However, gene abundance (relative to the *16S rRNA* gene), on a per antibiotic class basis, indicated that AMR gene was higher in the effluent than sludge for all antibiotic class except vancomycin (CST-household, CST-healthcare), β-lactam (CST-household), tetracycline and MDR (CST-healthcare) and MLSB and trimethoprim (SST-household) (Figure 3.9). In addition to this, the AMR gene abundance of the CST-healthcare unit samples (sludge and effluent) was different from that of the household (CST-household and SST-household) tank samples (influent, sludge, effluent) (Figure 3.10A). Meanwhile, the household tanks appeared to have similar gene abundance between sludge and effluent but were separate from the CST-household influent (P3_06-19_INF and P3_07-19_INF) (Figures 3.10A).

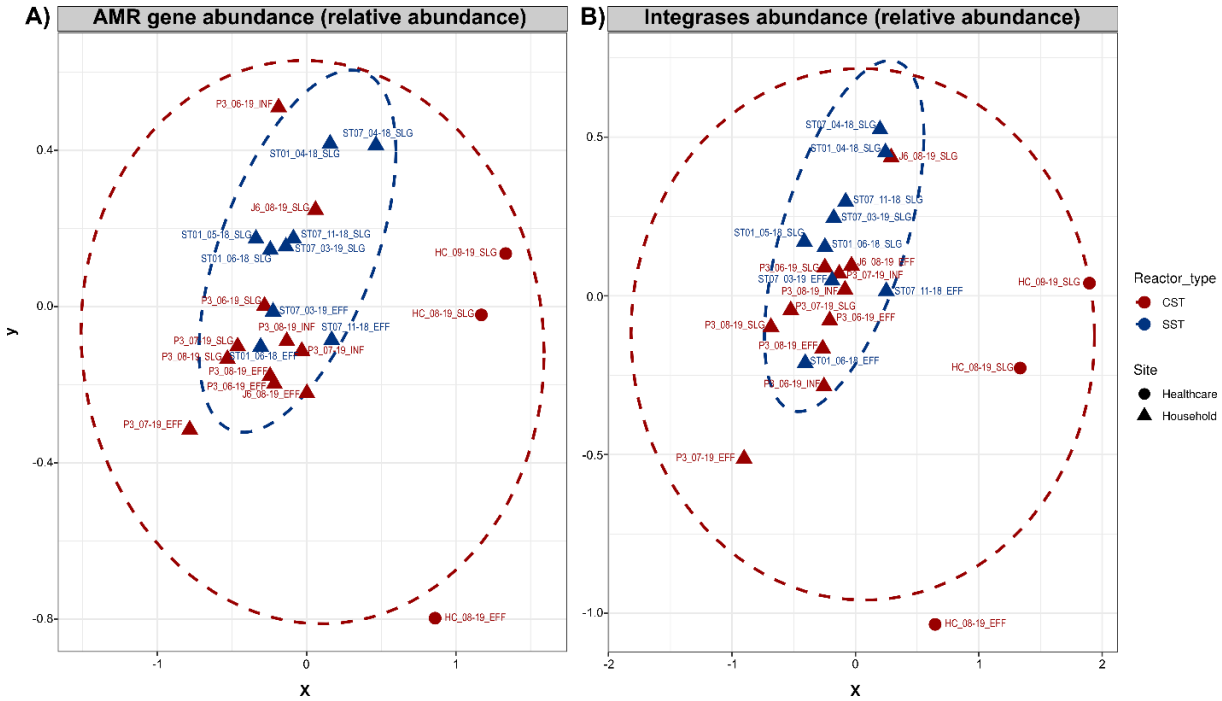

Figure 3.10: Non-metric distance scaling (NMDS) indicating similarities/ differences in gene abundance between samples (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household) based on their relative gene abundance (normalised to *16S_rRNA* gene abundance). A) Relative abundance of the AMR genes from all targeted antibiotic classes quantified on the HT-QPCR array. B) Relative abundance of the integrases (*intI*1, *intI*2, *intI*3) quantified on the HT-QPCR array. Ellipses represent a 95% confidence interval of standard error for a given group (CST vs SST). CST= Conventional septic tank, SST= Solar septic tank.

For the integrases (*intI*1-3, Figure 3.11), higher *intI*1 relative gene abundance was found in the effluent (CST-household: $9.02x10^{-2}$±$2.73x10^{-2}$SD, CST-healthcare: $7.04x10^{-2}$±0SD, SST-household: $7.71x10^{-2}$±$3.42x10^{-2}$SD) than sludge (CST-household: $4.31x10^{-2}$±$2.14x10^{-2}$SD, CST-healthcare: $1.21x10^{-2}$±$3.47x10^{-3}$SD, SST-household: $1.57x10^{-2}$±$7.36x10^{-3}$SD),

for the three reactors although statistical difference (p-value <0.05) between sludge and effluent for each of the tanks was not apparent. However, a comparison between sample types and reactor types indicated statistical differences, particularly between the CST-household effluent and the SST-household sludge (Figure 3.11A).

In the case of the *intI*3 gene, it was higher in the effluent (CST-household: $3.42x10^{-1}\pm2.53x10^{-1}$SD, SST-household: $2.20x10^{-1}\pm9.43x10^{-2}$SD) than sludge (CST-household: $2.91x10^{-1}\pm1.41^{-1}$SD, SST-household: $1.98x10^{-1}\pm6.59x10^{-2}$SD) for the household reactors. Nonetheless, for the healthcare units (CST-healthcare), *intI*3 showed higher relative abundance in the sludge ($1.22x10^{-2}\pm1.05x10^{-2}$SD) than effluent ($8.87x10^{-3}\pm0$SD) (Figure 3.11C). Moreover, there was no statistically significant difference in *intI*3 gene abundance between sample types for each reactor. Additionally, no statistical difference in *intI*3 gene abundance was found between sample type and rector type (Figure 3.11C).

*intI*2 gene was not detected in the CST-healthcare sludge but was quantified in the effluent ($1.60x10^{-3}\pm0$SD). However, for both household units (CST-household and SST-household), *intI*2 gene was higher in the sludge (CST-household: $6.41x10^{-5}\pm7.72x10^{-5}$SD, SST-household: $6.80x10^{-4}\pm1.63x10^{-3}$SD) than effluent (CST-household: $3.86x10^{-5}\pm3.19x10^{-5}$SD, SST-household: $1.05x10^{-4}\pm6.49x10^{-5}$SD) (Figure 3.11B). In addition, no statistical difference (p-value >0.05) was found between sludge and effluent for the two tanks (CST-household and CST-household) with quantifiable *intI*2 gene. Furthermore, no difference in relative intI2 gene abundance was reported between the sample and tank types (Figure 3.11B).

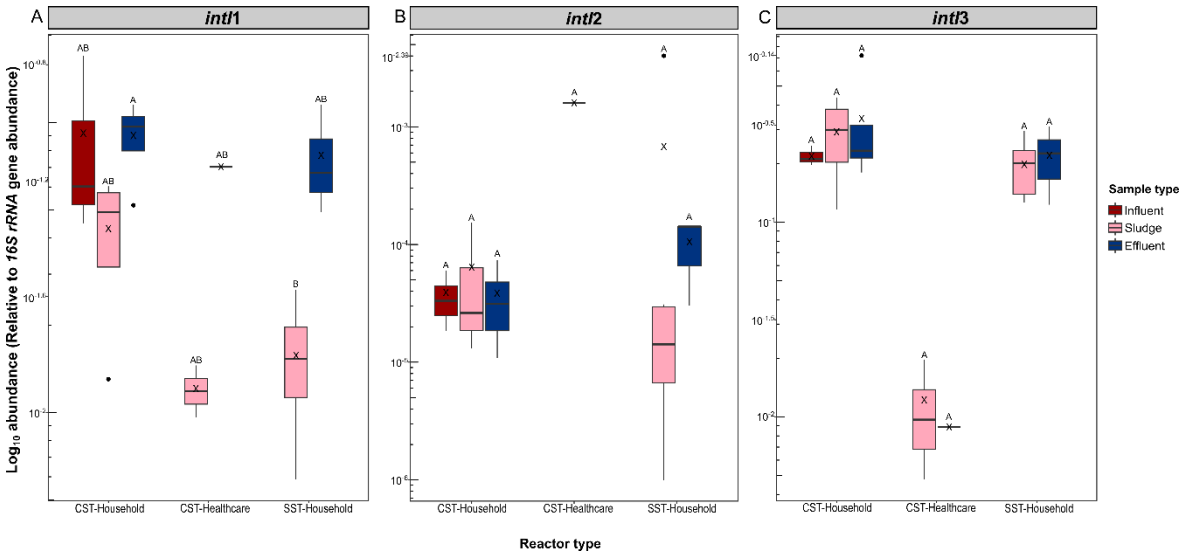

Figure 3.11: Comparison of relative integrases (*intI*1, *intI*2, *intI*3) gene abundance between samples (influent, sludge, effluent) and reactor type (CST-Household, CST-Healthcare, SST-Household) quantified on the HT-QPCR array. Result of a two-way ANOVA for the *intI*1 A) and Kruskal Wallis for the *intI*2 B), and *intI*3 C) relative gene abundance between samples and reactors. intI2 was undetected in the CST-Healthcare sludge. Black dots represent outliers. Boxplot sharing the same letter indicates no statistically significant difference at p-value>0.05. CST= Conventional septic tank; SST= Solar septic tank.

The total integrase (*intI*1, *intI*2, *intI*3) relative gene abundance in the CST-healthcare samples (sludge and effluent) appeared different to the SST-household and CST-household samples (influent, sludge, effluent) (Figure 3.10B). Both household tanks (SST-household and CST-household) generally clustered together, though total integrase abundance in SST-household sludge appears to be different from that of the SST-household effluent and CST-household samples. Finally, the total integrase relative abundance of the SST-household effluent samples was more similar to the CST-household samples, particularly the CST-household influent (Figure 3.10B).

3.3.3.1.3 Septic tanks increase AMR gene loading entering the environment.

Influent samples were accessible for collection for the CST-household tank (CT-P3), enabling the evaluation of septic tanks ability to evaluate their ability to mitigate AMR removal. It was observed that AMR genes were highest in the effluent (mean relative abundance of the 11 antibiotic classes: $2.26x10^{-2}\pm4.59x10^{-2}$SD) than in the influent (mean relative abundance of the 11 antibiotic classes: $1.20x10^{-2}\pm2.15x10^{-2}$SD).

The higher AMR relative gene abundance reported in the effluent, was as a result of an increase in gene abundance from the influent, for the three sampling months (June, July,

August) (Appendix Tables B.4 or B.3 for the absolute Ct values). Of the three months, the highest number of total AMR genes increased in the effluent from the influent was in the June and July (n=53/66 (80%)) > August (*n= 42/66* [64%]) sampling months (Appendix Tables B.4). Of these AMR gene that showed increase in relative abundance between three sampling months (June, July, August), 31 AMR genes were commonly increased across the three months, whilst four, five, and none were exclusively increased in the June, July and August sampling month respectively. Additionally, 12 AMR genes were increased between the June and July months, five genes increased between the July and August months and six genes commonly increased between the June and August sampling months (Figure 3.12).

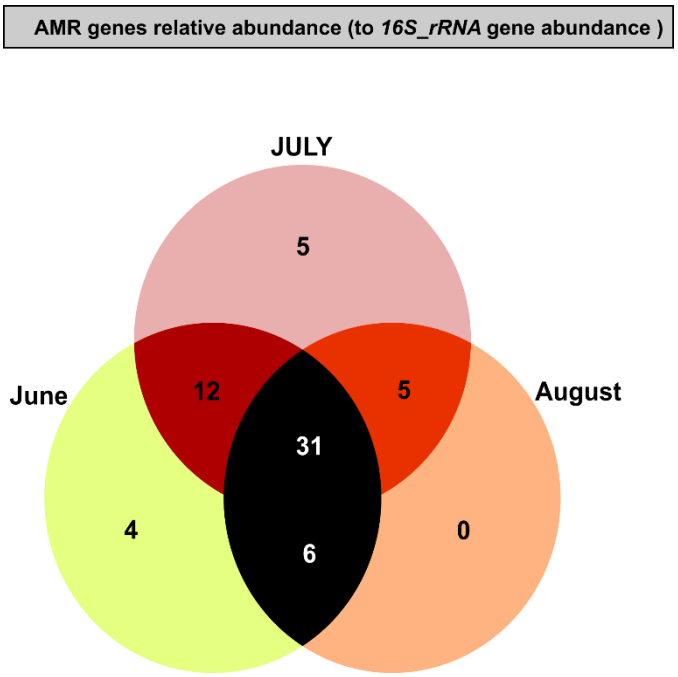

Figure 3.12: Number of gene with increased abundance (relative to the *16S rRNA* gene) in the effluent from the influent of the CST-Household unit (CT-P3 only) across the three sampling months (June, July, August). A) Number of AMR genes increased in the effluent from influent for all targeted antibiotic genes in the CST-household unit (CT-P3 only) across the three sampling months (June, July, August). B) Number of increased integrase genes (*intI*1, *intI*2, *intI*3) in the effluent from influent for the CST-household unit (CT-P3 only) across the three sampling months (June, July, August).

For the integrases (*intI*1, *intI*2, *intI*3), no integrase gene was increased between the three sampling months (June, July, August). Furthermore, no integrase gene relative abundance increased between the June and July months, while two integrase genes (*intI*1 and *intI*3) increased in relative abundance in the effluent from influent between the July and August sampling months. Finally, one integrase gene (*intI*2) increased in relative abundance in the effluent from influent between the June and August sampling months.

General Summary

MLSB genes were on average the most abundant and trimethoprim resistance genes the least abundant across the samples (influent, sludge, effluent) and tanks (CST-household, CST-healthcare, SST-household).

Surprisingly, the CST-healthcare tank contributed less AMR and integrase (*intI*1 and *intI*3) gene abundance (relative to the *16S rRNA*) to the environment via sludge and effluent, but higher *intI*2 abundance via effluent. The CST-household unit on the other hand, contributed to a higher AMR and integrase (*intI*1 and *intI*3) via sludge and effluent entering the environment among the three reactors but was the least contributor of *intI*2 via effluent to the environment.

Lastly, even though lower AMR and integrase (*intI*1 and *intI*3) gene abundance were quantified in the SST-household samples (sludge and effluent) than in the CST-household samples (the highest contributor), the AMR and integrases (*intI*1 and *intI*3) genes were still relatively high in abundance. Thus, the SST-household tank could potentially be another source of AMR genes and mobile elements to the environment. Moreover, this also indicated a limited role in incorporated temperature in greatly reducing AMR and integrase genes from SST-household samples.

With the high number of AMR and integrase genes quantified (on the array) between the samples (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household), we investigated, whether or not the abundance of integrase gene (*intI*1, *intI*2, *intI*3) or the *sul*1 gene (typically linked to CL1-integrons from clinical/ polluted environment) quantified on the array can serve as proxy to overall AMR gene abundance.

3.3.4 *intI*1 gene abundance as a proxy for AMR abundance

CL1-integrons are ubiquitous in the environment, particularly within anthropogenic polluted environments and have an elevated presence in polluted environments and as such have been suggested as a proxy for inferring AMR pollution (Gillings et al. 2015; Pruden et al. 2021). Here we investigated the link between integrases (*intI*1, *intI*2, *intI*3) and *sul*1 abundance (inferred by the absolute Ct), particularly *intI*1 gene abundance, to overall AMR abundance by analysing correlations between AMR gene abundance and integrase gene abundance quantified on the HT-QPCR array.

3.3.4.1 *intI*3 and not *intI*1 could serve as a proxy for overall AMR abundance

The abundance of the *intI*1 gene (relative to the *16S rRNA* gene) positively correlated to the abundance (relative to the *16S rRNA* gene) of most MRI associated and non-associated AMR genes, although only 21% (*n=7*) of MRI-associated AMR genes (*sul*1, *qacEΔ*1, *aac*3-*IVa*, *strB, blaGes, aadA*6, *dfrA*25) (Figure 3.13A) and 31% (*n=10*) of non-associated MRI genes (*tetM, tet32, mepA, sul*2, *blaCTX-M, sul*4, *sat*4, *vanB, mexA, fosB*) (Figure 3.13B) quantified were statistically correlated (p-value <0.05).

The relative abundance of *sul*1 (sulphonamide resistance gene encoding antibiotic target replacement resistance mechanism) and *qacEΔ1* (biocide/ antiseptic resistance gene encoding drug efflux resistance mechanism), which are genetically linked to the CL1-integron structure typically found within clinical and anthropogenic settings like WWT (Gillings et al. 2015) showed statistically significant correlation (p-value <0.05) to the relative abundance of *intI*1 (Figures 3.13A, Appendix Figure B.8A).

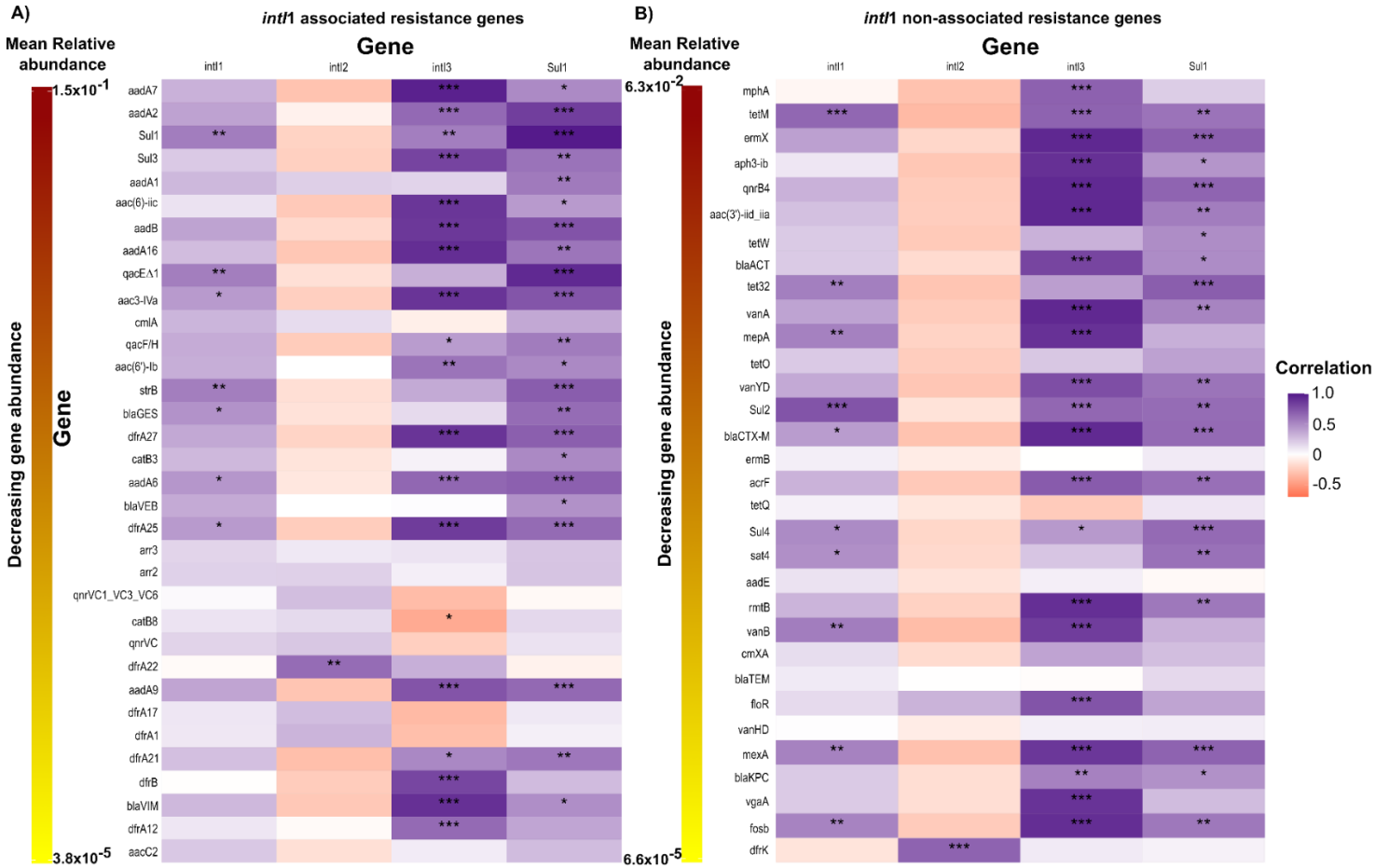

Figure 3.13: Correlation analysis based on gene abundance (relative to the *16S rRNA* gene), investigating link between integrases (*intI*1, *intI*2, *intI*3) and *Sul*1 gene abundance to the abundance of overall AMR genes quantified on the HT-QPCR array. Quantified AMR genes are separated based on their association A) and non-association with genes mobile resistance integron known disseminating AMR genes between and within bacterial taxa. Quantified genes are ranked from most abundant to least abundant. Statistically significant correlation are indicated with star(s) (* p-value <0.05, ** p-value <0.01, *** p-value <0.001).

The relative abundance of the *intI*2 gene correlated both positively and negatively to the relative abundance of MRI-associated and non-associated AMR, although only one MRI-associated AMR gene (*dfrA*22- encode Trimethoprim resistance, Figures 3.13A) and one non-associated MRI AMR (*dfrK*- encode Trimethoprim resistance, Figures 3.13B) statistically correlated (p-value <0.05) to the abundance of the *intI*2 gene.

*IntI*3 gene abundance (relative to the *16S rRNA*) correlated both positively and negatively to the abundance of MRI AMR genes, albeit 56% *(n=19)* of MRI-associated genes abundance were statistically correlated (95% *(n=18)* positively and 5% *(n=1)* negatively) (Figures 3.13A). Additionally, *intI*3 abundance predominately (*n=21* out of 32 genes (66%)) correlated positively and statistically (p-value <0.05) to the relative abundance of non-associated MRI AMR genes (Figures 3.13B).

Taken together, the relative abundance of *intI*3, emerged as a potential candidate for inferring overall AMR pollution from this dataset, owing to the high number of AMR-associated and non-associated MRI genes correlating positively and statistically with its relative abundance.

3.3.4.2 *sul*1 as an alternate proxy for mobile resistance integron associated AMR genes than *intI*1

*sul*1, which is typically found at CL1-integron 3' end isolated from clinical or other polluted settings like WWT, gene abundance correlated positively and statistically (p-value <0.05) to the abundance of a higher number (*n= 22* (65%)) of MRI-associated AMR genes subtypes than the *intI*1 gene (Figure 3.13A). In addition, like the *intI*1 gene, *sul*1 relative abundance positively correlated to the relative abundance of most non-associated MRI AMR genes, although only a subset (*n=19* (59%)) of these genes were statistically correlated (p-value <0.05) (Figure 3.13B). Thus, *sul*1 relative abundance appears to serve as a better proxy than the *intI*1 gene relative abundance for both MRI AMR-associated and non-associated genes.

3.3.4.3 Lower *intI*1 gene copies quantified by array primers (in-house) as compared to three previously selected *intI*1 primers (DF-DR, F3-R3, F7-R7)

Informed by our previous *in-silico* analysis of published and designed primers and probes (see Chapter 2), we hypothesised that the fewer statistically positive correlation observed between the *int*I1 gene abundance and the AMR gene abundance, particularly those

associated with MRI-AMR genes, could be due to the lower the sensitivity (at stringent threshold- i.e. no mismatch between primer and template sequence) of the HT-QPCR primer sets (AY289, AY293)) as compared to other published *intI*1 primer sets (see Chapter 2, Appendix Table A.2). As such, in this section, we validated the hypothesis that the array *intI*1 primers (AY289, AY293) generally quantified lower *intI*1 abundance by using array *intI*1 primers and the three selected *intI*1 primer sets from our previous chapter (Chapter 2), resulting in a total of five *intI*1 primer sets used.

Array *intI*1 primers (AY289, AY293) and our previously selected *intI*1 primer from *in-silico* testing (see Chapter 2, Table 2.7) were used to quantify the same wastewater samples (*n=23*, Table 3.2) as quantified on the array in-house.

Each of the standard curves from the five primer sets had high efficiencies which ranged from 92.35 to 96.06%, y-intercepts of 34.50 to 38.27, slope of -3.42 to -3.52, and a No-Template Control Ct of 0 (a number as convention for no amplificatiom) and 40.45 (Table 3.4). Between all five primers sets, no statistical difference in *intI*1 gene abundance was observed for the same sample types (influent, sludge, effluent) and reactors (CST-household, CST-healthcare and SST-household), except for SST-household sludge sample (Appendix Table B.6). This exception indicated that quantified gene abundance statistically differed when the F3-R3 and F10-R10 primer sets (p-value= 0.0019) are compared as well as the F4-R4 and F3-R3 (p-value= 0.011).

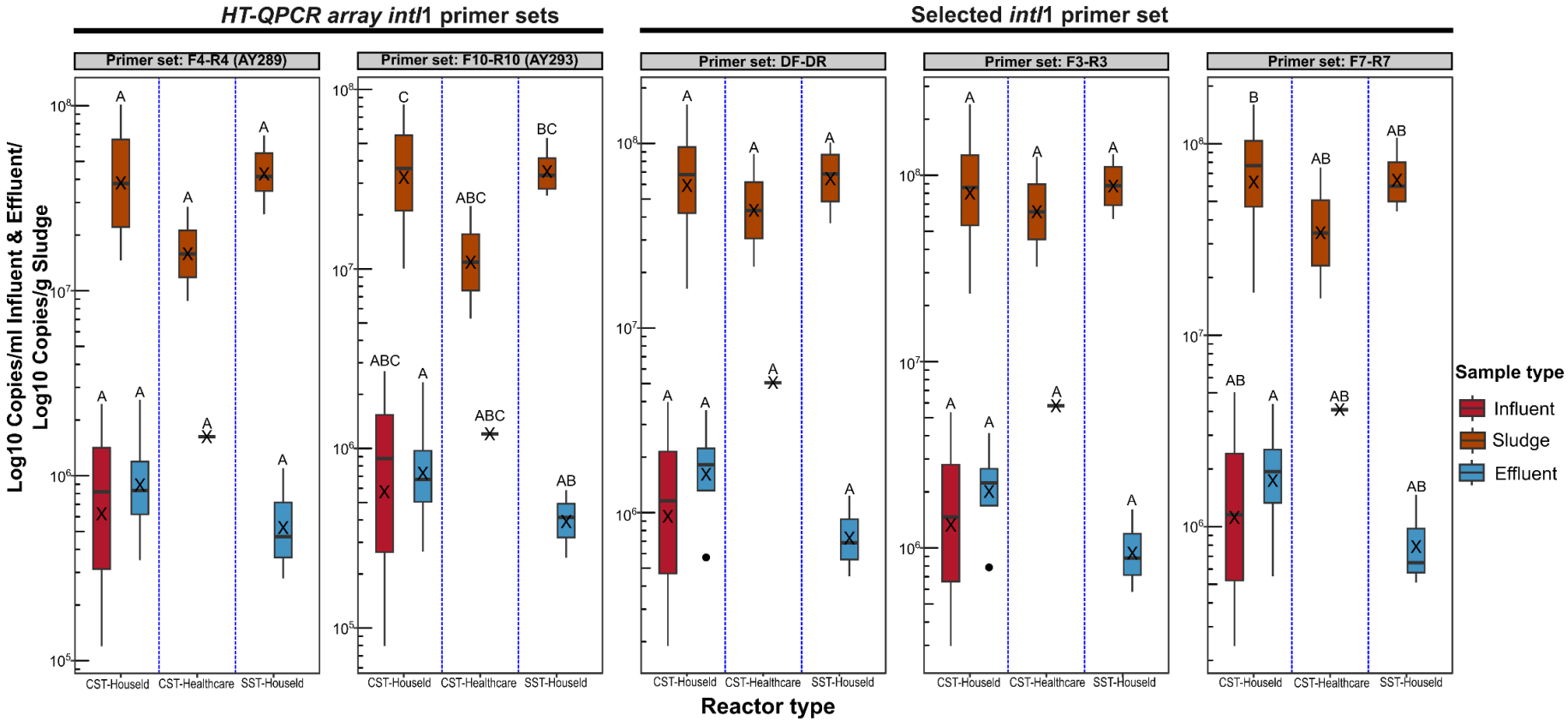

Figure 3.14: Impact of primer choice on quantified *intI*1 gene copies (per ml/g DNA) from CST-Household, CST-Healthcare and SST-Household wastewater reactors, and three wastewater sample types (influent, effluent and sludge). Results of the two-way ANOVA analysis indicated a statistically significant difference (p-value <0.05) in *intI*1 gene copies quantified between reactor types and sample types. Primer sets F4-R4 and F10-R10 correspond to HT-QPCR *intI*1 primer sets: AY289 and AY293 respectively. Primer sets: DF-DR, F3-R3 and F7-R7 are selected *intI*1 primers from *in-silico* analysis undertaken in the previous chapter (see Chapter 2). Outliers are represented by black dots. For each primer set, boxplot sharing the same letter indicates no statistically significant difference at p-value >0.05, while boxplot with different letters indicates a statistically significant difference at p-value <0.05. A statistically significant difference (p-value < 0.05) in *intI*1 gene abundance between primer sets for the same sample was only observed for the SST-household sludge samples (see Appendix Table B.6).

Spearman's rank correlation analysis, following Shapiro-Wilks normality test (p-value >0.05), indicated that the *intI*1 gene copies quantified by all five primer sets were highly correlated (ranged from 0.955 to 0.998 (p-value < 0.001), Appendix Table B.5) and thus, a similar overall pattern in gene abundance was observed for each primer set. Additionally, the statistical difference between samples and reactors observed was dependent on the primer set used. Only the F10-R10 (AY293) and F7-R3 primer set reported statistical differences between sample types (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household). Meanwhile, array primer sets- F4-R4 (AY289), DF-DR and the F3-R3 primer set reported no statistical difference for *intI*1 copies (per ml/g DNA) between the sample types and reactors (Figure 3.14).

In spite of the overall observed pattern in quantified intI1 gene abundance by the five primer sets, lower *intI*1 gene copies (per ml/g DNA) were quantified by HT-QPCR array *intI*1 primer sets (AY289, AY293) as compared to the previously selected primer sets (DF-DR, F3-R3, F7, R7).

To summarise, whilst a similar overall trend in *intI*1 gene abundance was reported by the five *intI*1 primer sets used to quantify *intI*1 abundance from the same Thai wastewater sample, array *intI*1 primers (AY289 and AY293) quantified lower *intI*1 gene abundance as compared the previously selected *intI*1 primer sets (DF-DR, F3-R3, F7-R7).

3.4 Discussion

The emergence and rapid adoption of the HT-QPCR array technology, and/ the use of the proxy gene, *intI*1, to infer AMR pollution is a promising step towards alleviating associated challenges in environmental AMR monitoring. However, with the potentially sub-optimal condition within the HT-QPCR array for some assays and the questionable reliability of the *intI*1 as a suitable proxy for inferring AMR pollution, we underwent to ascertain the suitability of the *intI*1 as a proxy for overall AMR abundance quantified on the HT-QPCR. First, by validating the array using the *16S rRNA* and *intI*1 array assays to quantify the same septic tank samples as on the array in-house and then compare the array and in-house results for both assays.

Following this, we utilised the array to characterise and quantify AMR and integrase (*intI*1, *intI*2, *intI*3) genes from our septic wastewater and compared the solar septic tank (SST-household) associated with household usage to conventional tanks associated with both

household (CST-household) and healthcare (CST-healthcare) usage to address our hypothesis that the increased temperature incorporated with the solar tanks will not only decrease the abundance of mobile integrons, particularly the CL1-integron (as shown previously in Chapter 2) but also the AMR subtype and abundance quantified.

3.4.1 HT-QPCR Array Validation and Best Practices: The Good, The Bad and The Ugly

The HT-QPCR, as observed from this study, can inform trends in ARGs and MGEs abundance between samples (in this specific case, trends in AMR gene and MGEs abundance between sample types (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household)) **(the good)**. However, these trends are biologically meaningful only when the absolute Cts abundance is compared, as the accuracy and reliability of observed trends within the data (or samples) become biased when quantified gene abundances (inferred by absolute Ct values) on the array are normalised to the *16S rRNA* Cts, owing to the multiple *16S rRNA* per bacterial genome (1-15 copies) (Smith and Osborn, 2009; Větrovský and Baldrian, 2013; Angly et al., 2014) **(the bad)**.

In HT-QPCR studies (Zhu et al. 2017; Chen et al. 2019a; Huang et al. 2019; Majlander et al. 2021) targeting hundreds of AMR gene and MGEs simultaneously, normalisation to the *16S rRNA* abundance gene remains the norm, owing to constraints and challenges in constructing standard curves for individual assays on the array for absolute quantification. Nonetheless, normalisation of quantified abundance (to the *16S rRNA* gene abundance) not only changes the antibiotic class or sample type (influent, sludge, effluent) with the higher gene abundance but can manifest statistical difference (p-values <0.05) in gene abundance between sample types or antibiotic class, where none originally exist (when compared to the absolute Ct abundance).

A fitting example from this study that clearly illustrates and supports this statement is shown in Appendix Table B.3 (reported AMR mean Ct abundance) and Appendix B.4 (reported AMR normalised abundance). Tetracycline resistance genes were, on average, the most abundant (of the 11 antibiotic resistance gene classes targeted) when absolute Ct abundance was reported. However, when abundance was normalised (to the *16S rRNA*), MLSB emerged as the most abundant.

In addition, tetracycline resistance gene was most abundant in the CST-healthcare sludge, as indicated by the absolute Ct values, compared to CST-household and SST-healthcare sludge. While in the effluent, tetracycline resistance gene was higher in the SST-household reactor than in the CST-household and CST-healthcare reactors when absolute Ct values were reported. Finally, no statistically significant difference between sample types (sludge and effluent) was reported by the absolute Cts for each of the reactors. However, a statistical difference (2-way-ANOVA p-value <0.05) between reactors and sample type, specifically between the CST-household influent and SST-household sludge was observed for the tetracycline class (Figure 3.9).

In contrast, reported relative abundance (normalised to the *16S rRNA* Ct) highlighted tetracycline resistance gene to be more abundant in the CST-household sludge and effluent than in the CST-healthcare and SST-household samples (sludge and effluent). Additionally, differences between sample types (sludge and effluent) were reported for the SST-household reactor. There was also a statistical difference between sample type and reactor type, specifically, between the CST-household influent, CST-healthcare effluent, and SST-household sludge for the tetracycline class. In summary, the normalisation of gene abundance to *16S rRNA* not only influences the results but alters how results are interpreted.

Bias in reported normalised abundance (to the *16S rRNA* gene abundance) is further worsened by the short amplicon (59 bp) *16S rRNA* primers (AY1) on the array **(the ugly)**. Array *16S rRNA* primer sets quantified not only a lower absolute Ct in the NTC, which was 2.44 (in-house) or 2.51 (array) log-fold higher when compared to the well-validated TaqMan primer-probe set NTC but quantified almost a log higher *16S rRNA* gene abundance in the samples when compared to the TaqMan primer-probe set quantified in-house. Thus, this indicates that the array *16S rRNA* primer set (AY1) potentially overestimates quantified *16S rRNA* abundance. A recommendation for future use would be to change the *16S rRNA* gene primer sets used on the array.

Considering the above, it becomes imperative and good practice to 1) ensure that absolute Cts compared are real, and not produced from amplification in the negative, by adopting the ≥3.3 Ct difference approach between sample and NTC mean Ct used in this study and described elsewhere (Smith and Osborn 2009) as an integral part of the data processing step. This avoids the need to employ arbitrary Ct-offs, which quite simply isn't the best approach, as it implies low abundance genes above the selected Ct cut-off cannot be quantified on/ by the array; 2) be wary of gene abundance normalisation to the *16S rRNA*, in particular,

normalisation of genes to *16S rRNA* quantified by array *16S rRNA* primer. The varying *16S rRNA* copies number per bacteria genome (1-15), choice of selected *16S rRNA* primers, coupled with any number of sample-specific factors (i.e. inhibitors) can disproportionately affect quantified *16S rRNA* abundance; thus, biasing the obtained results (Manor and Borenstein 2015).

Ideally, gene abundance quantified on the array should be reported as mean Ct or normalised to a single copy gene such as *rop*B (also quantified on the array) (Dai et al. 2020) to preserve original trends within data/sample. This makes cross-study comparisons far more reliable and accurate. Thus, permitting an in-depth knowledge of AMR dissemination from various environmental sources and identification of environments with the highest risk of AMR and MGE dissemination to humans and animals.

Despite these considerations and limitations of the array *16S rRNA* primers (AY1) highlighted above, quantified AMR and integrase gene abundance were normalised to the *16S rRNA* gene abundance quantified with the array *16S rRNA* primers. This decision stemmed from practical constraints, which include time limitation, cost consideration and limited available sample volumes, preventing the re-quantification of the samples on the array using a different well-validated *16S rRNA* primer set or a single copy gene.

3.4.2 HT-QPCR array applications: Risk assessment of quantified AMR and integrase (*intI*1, *intI*2, *intI*3) genes

As discussed above we moved forward with gene abundance normalised to the *16S rRNA* gene abundance. While there were overall similarities in AMR gene diversity, AMR gene abundance (relative to the *16S rRNA* gene abundance) within each antibiotic class differed between sample types (influent, sludge, effluent) and reactors.

Of the different antibiotic classes, resistance genes from the MLSB class dominated as the most abundant in the household tanks (CST-household and SST-household) sludge and effluent samples. Meanwhile, for the CST-healthcare units, tetracycline and sulphonamide resistance genes were most abundant in the sludge and effluent respectively. In Thailand, both macrolide and tetracycline are among the most commonly prescribed antibiotics consumed by humans (Siltrakool et al. 2021). Tetracycline's usage is owing to their broad-spectrum activity against gram-positive and negative bacteria and protozoan parasites (Grossman 2016). Tian et al.,(2019) also reported tetracycline resistance genes (normalised

abundance to the *16S rRNA* gene) as the most abundant AMR genes in wastewater sludge quantified on the array.

Genes conferring trimethoprim resistance were the least abundant across all sample types and reactors except for the CST-healthcare effluent, where genes conferring resistance to vancomycin were the least abundant (relative to the *16S rRNA* abundance). In Thailand, trimethoprim is commonly used in combination with sulfamethoxazole for the treatment of Melioidosis, an infectious disease endemic throughout Thailand and associated with a high fatality (fatality rate estimated at 37% based on a modelling of the 7572 cases diagnosed yearly in Thailand) (Hinjoy et al. 2018). Melioidosis is caused by the gram-negative and pathogenic bacteria *Burkholderia pseudomallei*, frequently isolated in contaminated water and soil (Saiprom et al. 2015; Limmathurotsakul et al. 2016; Hinjoy et al. 2018). Treatment of Melioidosis in Thailand involves an intravenous trimethoprim-sulfamethoxazole antibiotic combination for 10 days, coupled with an oral trimethoprim-sulfamethoxazole antibiotic combination for a further 20 weeks (Anunnatsiri et al. 2021). As 30-90% of consumed antibiotic is excreted unchanged via urine or faeces (Sarmah et al. 2006) and enters WWT works, driving selection pressure perhaps then, the trimethoprim genes resistance observed within the three septic tanks (CST-household, CST-healthcare, SST-household) could be/ in part ascribed to trimethoprim excreted as a result of Melioidosis treatment. Although challenging to confirm as data was not collected.

Of the three tanks, consistently lower AMR (relative abundance of the 11-antibiotic classes) and integrase gene (*intI*1 and *intI*3) abundance was reported in the CST-healthcare sludge and effluent samples than in the SST-household and CST-household samples (sludge and effluent). However, higher *intI*2 gene relative abundance was observed in the CST-healthcare effluent, when compared to the other two reactors. This suggests that the CST-healthcare unit contributed less to overall AMR and integron (CL1-and-CL3) abundance in the environment via sludge and effluent, but more to CL2-integron abundance via effluent.

The consistently lower overall AMR and integrase gene (*intI*1 and *intI*3) abundance reported in the CST-healthcare samples (sludge and effluent) was indeed surprising considering that healthcare WWT is widely acknowledged as an important source of AMR and integrase genes to the wider environment (Rodriguez-Mozaz et al. 2015) owing to broad-spectrum antibiotics that are generally consumed. The presence of broad-spectrum antibiotics, often higher concentration, within the WWT, particularly in septic tanks, exerts stronger selection pressure. This pressure can provide a competitive advantage to bacteria already possessing

resistance traits (Thongsamer et al. 2021) and/or drive the acquisition of AMR genes. In addition, it can promote integrase excision and integration of AMR gene cassettes onto the integron platform.

In contrast, the CST-household samples (sludge and effluent) consistently showed higher overall AMR and integrase (*intI*1 and *intI*3) gene relative abundance, among the three septic tank units. Thus, likely implies that they are the higher contributor of AMR and CL1-and-3-inetgron gene abundance to the environment via effluent and sludge discharge, particularly if the sludge fails to undergo additional treatment before discharge to the wider environment. Moreover, this higher overall AMR and integrase gene abundance in sludge and effluent is indicative of poor treatment performance within the tank, usually characteristic of septic tanks (Connelly et al. 2019).

In fact, assessing the conventional tank with accessible influent samples showed a high number of targeted AMR genes had a higher relative abundance in the effluent compared to the influent for the three sampling months. The number of AMR genes with higher relative abundance in the effluent was particularly higher in the June and July sampling months, with 80% (53/66) of the targeted AMR gene displaying higher relative gene abundance in the effluent than influent (Figure 3.12).

Whilst the SST-household reactor was not identified as the highest contributor of AMR and CL1-and-CLL3 integron to the environment via its sludge and effluent, among the three tanks, the abundance of AMR gene (relative abundance of 11 antibiotic class) and integrase (*intI*1 and *intI*3) gene were still high and slightly lower than that of the CST-household unit (most AMR contributor). As such, suggests a limited impact of the increased internal tank temperature in reducing integrase genes compared to the conventional tanks (CST-household and CST-healthcare).

Although, the lack of an accessible influent sample for the solar tank hinders a better understanding of the impact of incorporated internal temperature on the fate of AMR gene and integrases within the solar septic tank. The role of temperature on the fate of AMR genes and integrases gene remain unclear with some studies reporting reduced AMR gene abundance at increased temperature (Thermophilic temperature i.e., 55˚C) (Ghosh et al. 2009; Sun et al. 2016; Tian et al. 2019), whilst others (Huang et al. 2019) reported a better reduction of AMR abundance at mesophilic temperatures (i.e., 25˚C and 37˚C) than thermophilic temperature (55˚C) but highest increase of *intI*1 abundance at thermophilic temperature (55˚C) than mesophilic temperatures (i.e., 25˚C and 37˚C). For example,

utilising the HT-QPCR array, Huang et al., (2019), reported higher reduction of AMR gene abundance (normalised abundance to the *16S rRNA*) at mesophilic temperatures (25˚C and 37˚C) than at thermophilic temperatures (55˚C) after 30 days of anaerobic digestion swine manure treatment, although, the AMR abundance of thermophilic temperatures (55˚C) at day 30 was lower than day 0. In contrast, increased *intI*1 abundance (normalised abundance to the *16S rRNA*) was reported for all temperatures (mesophilic and thermophilic temperature) after 30 days, again, higher *intI*1 abundance was noted at the thermophilic than mesophilic temperature.

On the other hand Tian et al., (2019), also utilising the array, reported statistically (p-value <0.01) lower AMR and *intI*1 gene abundance (normalised to the *16S rRNA*) from excess WWT sludge subjected to increasing antibiotic concentration (0 to 1000mg/L oxytetracycline) and treated at thermophilic temperature (55˚C), as compared to sludge sample without thermophilic treatment and antibiotic stress. Although, as the concentration of antibiotic stress increases the abundance of AMR genes abundance also reported to increase (Tian et al. 2019).

In Thailand, and many other global south countries, discharged faecal sludge from WWT is rarely subjected to further treatment (only 10-20% are estimated to undergo additional treatment) and as such, discharged directly to the environment (Koottatep et al. 2021). Therefore, WWT sludge samples represent a major source of additional AMR genes and mobile resistance integron (Class 1, 2, 3) to the environment when discharged directly to the environment, which further exacerbates the global AMR burden.

Our findings do not strongly support our proposed hypothesis that the solar septic tanks reduce AMR and integrase gene abundance substantially from the effluent and sludge, of the three tanks. Further studies are needed to obtain a clearer understanding of the impact of temperature on AMR gene removal, especially for this innovative septic tank technology. Moreover, the generally diverse and high AMR and integrase genes abundance between the three tanks showcase the potential role of septic tanks (decentralised WWT) as a major contributor of AMR genes to the environment, which echoes the current understanding that WWT, in general, is a significant source of AMR to the environment. Thus, emphasises the need for optimisation of WWT, particularly decentralised WWT, to enhance AMR removal and thereby, reduce the global AMR burden. AMR genes between the three tanks were dominated by genes conferring antibiotic inactivation resistance mechanisms.

3.4.3 Link between integrases (*intI*1, *intI*2, *intI*3) and *sul*1 gene abundance and overall AMR abundance using HT-QPCR array

*intI*1, whilst proposed as a proxy for inferring AMR pollution, was shown in this study to correlate positively and statistically to only a few AMR genes associated and non-associated mobile resistance integron as compared to the *intI*3 gene (Figure 3.13). This was indeed surprising given that *intI*1 genes are generally ubiquitous in clinical settings and the environment, particularly within polluted environments such as WWT than *intI*3 (Gillings et al. 2015; Quintela-Baluja et al. 2021). *intI*3 are considered to be less frequent in the polluted environment than *intI*1 (Cambray et al. 2010; Quintela-Baluja et al. 2021) but was quantified in higher overall abundance from the septic tanks than *intI*1 and *intI*2. In fact, the higher number of AMR genes (both mobile integron-associated and non-associated AMR genes) correlating positively and significantly to the *intI*3 abundance prompted the suggestion that perhaps *intI*3 abundance could serve as a proxy for overall AMR abundance. Although further studies are required to support this observation.

However, the poor statistical positive correlation observed for the *intI*1 gene could also be a result of the *intI*1 primer sets used on the array. We showed, in the previous *in-silico* study (see Chapter 2) that the *intI*1 primer sets used on the array amplified lower *intI*1 abundance as compared to our three previously selected *intI*1 primer sets or other published primer sets analysed. Further, we confirmed this observation in the laboratory and showed that the array *intI*1 primer sets (AY289, AY293), when compared to our three previously selected *intI*1 primer sets (DF-DR, F3-R3, F7-R7), generally quantified lower *intI*1 gene copies per ml (influent and effluent) or copies per g (sludge) DNA for the same wastewater samples (Figure 3.14).

*Sul*1, which is typically linked to the CL1-integron 3' conserved region, was found to correlate positively and statistically to the *intI*1, which has also been reported in numerous studies (Su et al. 2012; Chen and Zhang 2013b; Paulus et al. 2020). In addition, *sul*1 abundance correlates positively and significantly to a higher number of genes, both MRI-associated and non-associated AMR genes, than *intI*1 indicating that *sul*1 abundance could serve as an alternate proxy for AMR-associated mobile resistance integrons and potentially for overall AMR abundance. This finding supports the study (Berendonk et al. 2015) suggesting the use of *sul*1 as an alternate proxy for monitoring AMR pollution. Although Gillings et al., (2015) previously argued that the use of specific resistance genes such as *sul*1

is generally not a good idea as their presence and abundance in the investigated polluted environment is dependent on the presence of the specific antibiotic they confer resistance to.

3.5 Conclusion

Though sub-optimal condition within the HT-QPCR array has only been implied, this current study provided a glimpse into HT-QPCR array's performance and highlights the good, bad and ugly aspects of using the array, particularly in the context of interpretation of reported results, which are negatively influenced by normalisation to the *16S rRNA* abundance, especially when the *16S rRNA* gene was quantified by the short amplicon array *16S rRNA* primers (AY1). Nonetheless, for reasons outlined in the discussion, the normalised AMR relative (normalised to the *16S rRNA*) gene abundance was reported. Furthermore, the suitability and reliability of the *intI*1 gene as a proxy for inferring overall AMR pollution remains inconclusive, but findings indicate that it may not be the best target. Future studies are in no doubt needed, especially a PCR-bias-free approach (i.e., shotgun metagenomic), to clearly discern whether or not the *intI*1 abundance can serve as a suitable and reliable proxy for inferring overall AMR pollution or just mobile integron-associated AMR genes. Finally, the results from this study have highlighted septic tank effluent, in particular the CST-household unit effluent, as a source of higher AMR gene abundance and *intI*3 abundance to the environment.

Chapter 4

Shotgun metagenomic characterisation of AMR genes from Thai septic tanks

4.1 Introduction

Despite the development of the HT-QPCR array to monitor AMR and/or the use of proxies such as *intI*1 to infer AMR pollution, comprehensive broad-spectrum AMR monitoring from polluted environments, such as WWT, remains a challenge. This is owing to reliance on Q-PCR primers and associated bias from these primers. Nonetheless, recent advances in next-generation DNA sequencing technologies, reduced sequencing costs, and fast turnaround, coupled with advanced bioinformatic pipelines (Krawczyk et al. 2018), facilitated the rapid and comprehensive profiling and characterisation of AMR genes through shotgun metagenomic sequencing (Ma et al. 2021). This non-targeted approach provides a holistic understanding of the dynamic fate of AMR genes through WWT works, as well as contributions from WWT to the overall global AMR burden. Thus, enabling the identification of high-risk environments to human and animal health and the implementation of intervening strategies to combat the global AMR burden.

Many studies have employed shotgun metagenomics to character AMR genes from WWT (Ekwanzala et al. 2020; Karaolia et al. 2021; Manoharan et al. 2021; Rodríguez et al. 2021) and/or their immediate discharge environment (Chu et al. 2018). For example, Karaolia et al.,(2021) used shotgun metagenomics, to investigate the fate of antibiotic resistance (ARGs) and biocides resistance genes (BRGs) from two full-sale urban wastewater treatment plants, one of which uses carbon-activated sludge (CAS) treatment and the other membrane bioreactor (MBR) treatment. Both WWTs significantly reduced ARGs (in terms of richness, evenness, relative abundance) and BRGs (in terms of richness and evenness) in the effluent from influent, although the MBR WWT exhibited higher removal efficiency than the CAS WWT. Furthermore, the CAS WWT enriched clinically relevant ARGs in the effluent, while the MBR WWT enriched triclosan in the effluent. Whilst shotgun metagenomics approach continues to gain popularity and is increasingly used, the vast majority of current studies employing this approach mainly focus on centralised WWT resulting in a better understanding of centralised WWT performance in reducing AMR from treated waste, as

well as their contribution of AMR genes, especially clinically relevant ARGs, to the environment. In contrast, studies on decentralised WWT are scarce as decentralised WWT are often overlooked (Bunce and Graham 2019), despite serving a significant portion of the global population (approximately 2.7 billion people (Harada et al. 2016)) and characterised by poor treatment performance. This knowledge gap significantly impedes an in-depth understanding and evaluation of decentralised WWT performance and their contributions to AMR to the environment. Such understanding is urgently needed if we are going to improve decentralised WWT to effectively reduce its contribution to the global AMR burden.

With this knowledge gap identified, this study utilised shotgun metagenomics sequencing to comprehensively characterise AMR genes (ARGs and stress genes) from decentralised WWT from Thailand, a country where antibiotics are readily accessible and poorly regulated. Specifically, we compared the solar septic tanks (SSTs) associated with household usage, to conventional tanks (CSTs) associated with household and healthcare usage, with the hypothesis that the SSTs would be more effective in reducing overall AMR gene burden from the sludge and effluent than CST, due to their higher internal temperature. Additionally, this study investigated removal efficiency of ARGs and stress genes from septic tanks by examining influent and sludge samples from septic tanks (CST-household) with accessible influent. Finally, this study employed a random forest model to identify genes (ARGs and/ stress genes) that could potentially serve as a useful marker for distinguishing the different tank types and sample types.

4.2 Materials and Methods

4.2.1 Solar and Conventional septic tank sampling

The sampling of solar and conventional septic tanks was the same as described in the previous chapter (See Chapter 2, section 2.3.1). Table 4.1 highlights the samples selected for this study.

Table 4.1: Thai septic tank wastewater samples and time points selected for metagenomics

Reactor type	Reactor ID	April 2018	May 2018	June 2018	Nov 2018	March 2019	June 2019	July 2019	Aug 2019	Sept 2019
SST	SST-01	EFF/ SLG	EFF/ SLG	EFF/ SLG						
	SST-07	EFF/ SLG			EFF‡/ SLG	EFF/ SLG				
CST	CST-P3						INF/ EFF/ SLG	INF/ EFF/ SLG	INF/ EFF/ SLG	
	CST-J6						EFF/ SLG	EFF/ SLG	EFF/ SLG	
	CST-HC								EFF/ SLG	EFF/ SLG

SST: Solar septic tank; CST: Conventional septic tank; INF: Influent; SLG: Sludge; EFF: Effluent. ‡ failed sequencing run

4.2.2 DNA extraction

Extraction of DNA from septic tank samples was as described in the previous chapter (See Chapter 2, section 2.3.2)

4.2.3 Construction of DNA metagenome libraries constructions and sequencing

Metagenomic DNA libraries were prepared by Dr. Anastasiia Kostrytsia at the University of Glasgow. This was because the COVID-19 global pandemic lockdown restricted access to the laboratory, which prevented me from making the libraries with Dr Kostrytsia. Briefly, libraries were prepared with the KAPA HyperPlus Kit (PCR-free) according to the manufacturer's instructions. KAPA Pure Beads (3X) was used to purify genomic DNA from Ethylenediaminetetraacetic acid (EDTA) to prevent enzymatic inhibition of genomic DNA fragmentation during the KAPA HyperPlus protocol. Input genomic DNA into library construction was 200ng, except for three effluent samples which were below (CST-healthcare-CT-HC_09-19: 87.15ng; SST-household-ST-01_04-18: 183.75ng) or above (SST-household-ST-07_04-18: 297.5ng) 200ng. Genomic DNA was fragmented enzymatically at 37°C for 8 min. KAPA unique-dual indexed (UDI) adapter oligos (Roche) were used for adapter ligation. Post-ligation clean-up (0.7X) and size selection (0.3 – 0.5X) were performed with KAPA Pure Beads (Roche).

Prior to normalisation and pooling (31 libraries), all libraries were quantified with the Q-PCR-based KAPA Library Quantification Kit for Illumina platforms (Roche). Library size

distributions were confirmed with a 2100 Bioanalyzer instrument and Agilent® DNA 12000 Kit (Agilent Technologies; Cheadle, UK). The pooled sample was sequenced on four lanes (31 libraries/lane) on the Illumina NovaSeq 6000 Sp (V.15 flow cell) platform at Earlham Institute (Norwich, UK) with 2 x 250 bp paired-end reads to yield a total of 1,600,000,000 (1.6 Billion) reads per 31 samples.

4.2.4 Bioinformatics

A total of 30 metagenomics samples were processed. Adapter trimmed reads were provided by the Earlham Institute which was then subjected to quality trimming using Sickle v1.200 (Joshi and Fass 2011) Reads were trimmed when the average Phred quality dropped below 20, and then paired-end reads were only retained when greater than 50bp. This gave a total of 1,501,034,710 reads from all samples. Forward and reverse reads were collated together and samples were co-assembled using megahit with the parameters --k-list 27,47,67,87 --kmin-1pass -m 0.95 --min-contig-len 1000 (Li et al. 2015). This gave a total of 22,743,024 contigs, a total of 22,341,627,403 base pairs (bp), a maximum of 737,789 bp and an N50 score of 1,454 bp. MetaWRAP pipeline was then used (Uritskiy et al. 2018) and the contigs were binned using metawrap binning --metabat2 command (Kang et al. 2019), which gave a total of 4,103 bins. On these bins, checkm (Parks et al. 2015) was used to assess their completion as well as contamination. For the bins, a mean genome completion of 52.65% was obtained with a mean contamination of 3.705%. For the bins, sample-wise coverages using CoverM were done (Robbins et al. 2017) using the --methods mean parameter. AMRFinderPlus (Feldgarden et al. 2021) was subsequently used to recover AMR genes for the above-detected bins. Three thresholds for matching amino acids in the reference database were used: 25 amino acids coverage with 40% identity (Sydenham et al. 2019); 50 amino acids coverage with 75% identity (Antelo et al. 2021); and 75 amino acids coverage with 90% identity (Wang et al. 2021). The obtained sample read coverages per bin $C_{i,j}$ was then multiplied with feature coverages (returned from above) per bin $F_{j,k}$ to obtain feature coverages per sample $n_{i,k}$ as a matrix product $n_{i,k} = \sum_j C_{i,j}F_{j,k}$. These tables were then used subsequently in the statistical analyses.

4.2.5 Statistical analysis

R's Vegan package (Oksanen et al. 2022b) was used for the analysis of alpha diversity of all tables. For alpha diversity, the indices used were (i) rarefied richness – the number of

expected features in a rarefied sample (to the minimum library size), which is often the exponential of Shannon entropy, (ii) Pielou's evenness – an index that compares measured diversity values to the maximum theoretical diversity, and that is constrained between 1 (complete evenness) and 0 (no evenness), and (iii) Shannon entropy – an index that takes into account both richness and diversity to provide a measurement of community balance.

We performed Local Contribution to Beta Diversity (LCBD) analysis (Legendre and De Cáceres 2013) by using LCBD.comp() from R's adespatial package (Dray et al. 2012). We have used the Hellinger transform on the obtained abundance tables (microbes/functional annotations). LCBD gives the sample-wise local contributions to beta diversity that could be derived as a proportion of the total beta diversity. To find sets of features (ARGs or stress genes) that were differentially abundant between tank types (SST-household, CST-healthcare, SST-household) or sample types (influent, sludge, effluent) Kruskal-Wallis test, a non-parametric test, was performed. As the test was individually performed on all the acquired parameters, the Benjamini and Hochberg procedure (Benjamini and Hochberg 1995) was subsequently applied to adjust the p-values for multiple comparisons (Rashid et al. 2022). All figures in this study were generated using R's ggplot2 package (Wickham 2016). For alpha diversity and LCBD, we have used ANOVA, and where two categories are significantly different, following annotations are used to denote significance: '***' ($p \leq 0.001$), '**' ($p \leq 0.01$), '*' ($p \leq 0.05$), and '.' ($p \leq 0.1$).

4.3 Results

4.3.1 ARGs and Stress genes stringency mapping parameter greatly impact observed richness and diversity

Three mapping parameters of increasing stringency (i.e., coverage length and percentage identity) for the identification of genes were applied.

1) Lowest stringency- 25 amino acids coverage with 40% identity (Sydenham et al. 2019);
2) Medium stringency- 50 amino acids coverage with 70% identity (Antelo et al. 2021);
3) Highest stringency- 75 amino acids coverage with 90% identity (Wang et al. 2021).

These different parameters constitute a trade-off between coverage and specificity and thus, significantly affect the richness and diversity of ARGs/ stress genes characterised (Lal Gupta et al. 2020; Sevillano et al. 2020).

At the lowest stringent parameter (25 amino acids coverage with 40% identity) (Sydenham et al., 2019), richness (rarefied count richness) of detected ARGs and stress genes were high in all sample types across the three septic tanks (Figures. 4.1A, 4.1B). Of the three reactors, ARGs richness was higher in the CST-household tanks (influent: 370.78 ± 80.62 SD, sludge: 408.39 ± 34.06SD, effluent: 373.65 ± 56.51SD rarefied count) as compared to the SST-household (sludge: 401.76 ± 11.5SD, effluent: 315.91 ± 59.78SD rarefied count) and the CST-healthcare (lowest richness; sludge: 358.1 ± 33.33SD, effluent: 329.94 ± 36.39SD rarefied count) unit, with significant difference (p-value <0.05) only reported for sample types between the reactors (Figure 4.1A).

Similarly, for the stress genes (Figure 4.1B), richness was generally higher in the CST-household reactor (influent: 83.10 ± 6.38SD, sludge: 87.18 ± 2.13SD, effluent: 85.56 ± 7.05SD rarefied count) than in both the CST-healthcare (sludge: 83.25±4.31SD, effluent: 81.96 ± 5.89SD rarefied count) and SST-household (sludge: 86.58 ± 4.99SD, effluent: 74.36 ± 9.43SD rarefied count), albeit that statistical difference (p-value <0.05) was only reported between CST-household sludge and SST-household effluent and between SST-household sludge and effluent (Figure 4.1B).

For both ARGs and stress genes, lower richness was consistently detected in the effluent than in sludge for the three reactors and only the SST-household effluent was statistically lower than the sludge for both ARGs and stress genes (Figures 4.1A, 4.1B).

Both Pileous and Shannon entropy indicated that ARGs evenness was statistically different (p-value <0.05) between some sample types and reactors, with higher ARGs evenness generally observed for the CST-household tanks than the SST-household and CST-healthcare tanks (lowest ARG evenness) (Figure 4.1A, Appendix Table C.1).

Conversely, for stress genes, no significant difference (p-value >0.05) in evenness (Pileous and Shannon) was noted between the sample types (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household). In addition, both Pielou's and Shannon indicated that the stress genes evenness between the sample types and reactors was generally similar (Appendix Table C.1).

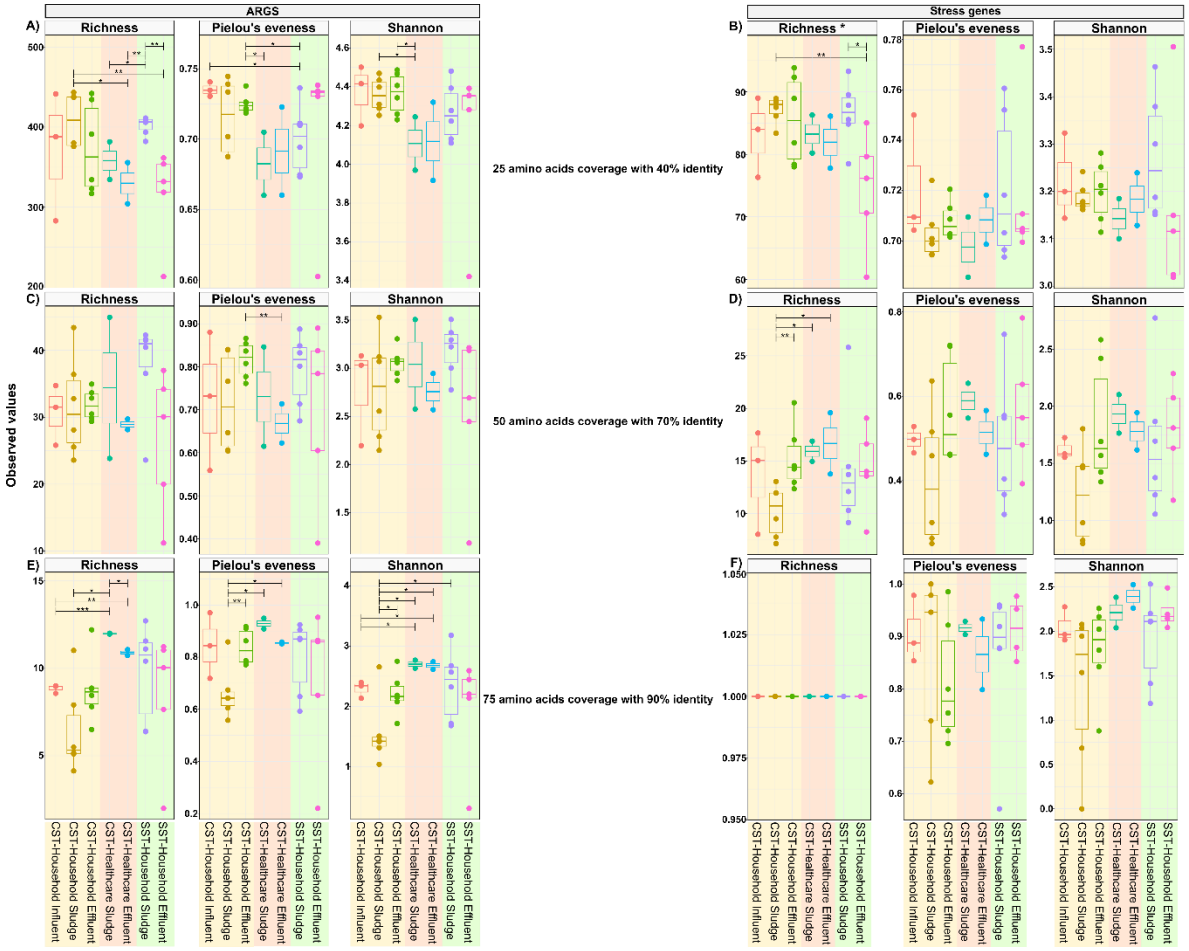

Figure 4.1: Impact of selected mapping stringency (i.e., coverage length and percentage identity) on characterised ARGs and stress genes Alpha diversity. Three stringent parameters: 25 amino acids coverage with 40% identity (A,B), 50 amino acids coverage with 70% identity (CD) and 75 amino acids coverage with 90% identity (EF) were employed to characterise ARGs and stress genes. Richness (rarefied count), Pileous evenness and Shannon entropy were used to compare the ARGs and stress genes within the septic tank wastewater which were grouped by sample type (influent, sludge, effluent) and reactor type (CST-household, CST-healthcare, SST-household). CST: Conventional septic tank; SST: Solar septic tank. * p-value <0.05; ** p-value <0.01; *** p-value <0.001.

Increasing the stringent parameters for classifying ARGs and stress genes substantially decreased the richness (rarefied richness) of detected ARGs and stress genes within each sample type across the three reactors (Figures 4.1C-F, Appendix Tables C.2-C.4). At medium (50 amino acids coverage with 70% identity) and higher (75 amino acids coverage with 90% identity) stringent parameter, ARG richness (rarefied count) decreased substantially by 91% (Figure 4.1C) and 98% (Figure 4.1E) when compared to the richness observed at the lower stringent parameter (25 amino acids coverage with 40% identity) (Table 4.2). Similarly, the stress gene richness decreased considerably by 83% (Figure 4.1D) and 100% (Figure 4.1F) for the medium and higher stringent parameter respectively, again when compared to the stress gene richness at the lower stringent parameter (Table 4.2).

Table 4.2: Trends in ARGs and stress genes richness (rarefied count) characterised with the three different parameters used in published literature

Mapping parameters	Richness (rarefied count)			
	ARGs	Trend (% decrease)	Stress genes	Trend (% decrease)
Lowest stringent parameter: 25 amino acids coverage with 40 % identity	372.36 ± 54.17SD	-	83.58 ± 7.13SD	-
Medium stringent parameter: 50 amino acids coverage with 70 % identity	31.94 ± 7 .62SD	↓ (91%)	13.87 ± 4.11SD	↓ (83%)
Highest stringent parameter: 75 amino acids coverage with 90 % identity	8.81 ± 2.73SD	↓ (98%)	1.0 ± 0SD	↓ (100%)

In addition, a statistical difference (p-value <0.05) in richness between sample types and reactors was observed only for the ARGs using the higher stringent parameter (75 amino acids coverage with 90% identity) (Figure 4.1E) and for the stress genes using the medium stringent parameter (50 amino acids coverage with 70% identity) (Figure 4.1D).

Similar to the richness, increasing the stringency parameter decreased the ARGs (Figures 4.1C, 4.1E) and stress genes evenness (Shannon) (Figures 4.1D, 4.1F) between the sample types across all three tanks, although, it is worth mentioning that for the stress genes, evenness was generally higher at the stricter parameter (75 amino acids coverage with 90% identity) (Figure 4.1F) than at the medium stringent parameter (50 amino acids coverage with 70% identity) parameter (Figures 4.1D) but not statistically significant (p-value >0.05). Moreover, a statistical difference (p-value <0.05) between ARG sample types and reactors was only reported for the higher parameter (Figure 4.1E) and not the medium stringent parameter (Figure 4.1C).

Pielou's evenness on the other hand indicated the ARGs and stress genes between the sample types and tanks become more even when increasing stringency parameters (Figures 4.1), except for the stress genes at the medium stringent parameter (75 amino acids coverage with 90%identity; Figure 4.1D), which showed decreased evenness in all sample types between the reactors as compared to the lower (Figure 4.1B) and higher stringent parameter (Figure 4.1F; Appendix Table C1:C3). In addition, no significant difference (p-value >0.05) was observed between sample types and reactors for the stress genes at the medium (Figure 4.1D) and higher stringent parameters (Figure 4.1F), whilst statistical difference (p-value <0.05) was reported only between the CST-household and healthcare samples at the higher stringent parameter (Figure 4.1E) and only between the CST-household and CST-healthcare effluent at the medium stringent parameter (Figure 4.1C).

As shown in Figure 4.1, lowering the stringent parameter for characterising ARGs or stress genes allows for the recovery of a higher number of ARGs and stress genes, however, the risk of false positive characterisation of genes is substantially increased (Sydenham et al., 2019). Nonetheless, we opted to move forward with the lower stringent parameter (25 amino acids coverage with 40% identity) for all downstream data analysis. This decision was supported and informed from our previous HT-QPCR array study (see chapter 3) where both pooled sample data (Chapter 3 Figure 3.1) and individually targeted wastewater samples (Chapter 3 Figure 3.7) detected and quantified higher number of AMR genes (ARGs and stress genes) than those estimated by medium and higher stringent parameter used to characterise ARGs and stress genes. Additionally, 158 of the AMR genes detected and quantified on the HT-QPCR array pooled samples were detected and characterised in metagenomics samples using the lowest stringent parameter (Figure 4.2).

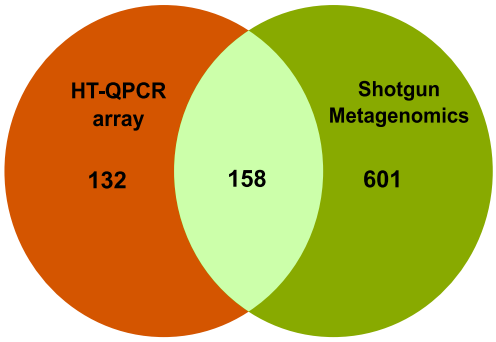

Figure 4.2: Shared and unique number of AMR genes (ARGs and stress genes) detected on the HT-QPCR array and shotgun metagenomics. Number of genes for the HT-QPCR array are the total number of genes retained (Out of 367 AMR genes targeted on the array) from pooled septic tank samples post data processing step (see Chapter 3 section 3.3.1 for further details).

Therefore, focusing solely on the selected lower stringent parameter (25 amino acids coverage with 40% identity), we observed that ARGs and stress genes richness and evenness were higher in the CST-household tanks as compared to the SST-household and CST-healthcare (lowest richness and evenness) units, and statistical difference between sample types and reactors was reported (Figure 4.1). In the next section that follows, we explored the dynamics of ARGs, and stress genes detected within the various sample types (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household) by categorising the detected genes (ARGs and stress genes) based on the antibiotic class

(ARGs) or antimicrobial class (stress genes) they confer resistance to and their mechanism of resistance.

4.3.2 Dynamics of detected genes within the septic tanks

To obtain a better understanding of the AMR genes (ARGs and stress genes) identified within the septic tanks (CST-household, CST-healthcare, SST-household), detected ARGs and stress genes were grouped by the antibiotic class (ARGs) or antimicrobial class (stress genes) they confer resistance to and their mechanism of resistance (Figures 4.2, 4.3).

A total of 646 ARGs and 113 stress gene subtypes (absolute count and not rarefied count) were detected. The 646 identified ARG subtypes conferred resistance to 29 antibiotic classes including aminoglycoside, antibacterial free fatty acid, avilamycin, bacitracin, β-lactam, bleomycin, fosfomycin, fusidic acid, macrolide, lincosamide, streptogramin, lincosamide/streptogramin (LS), macrolide/lincosamide/streptogramin (MLS), multidrug resistance (MDR), mupirocin, nitroimidazole, phenicol, pleuromutilin, quinolone, rifamycin, streptothricin, sulphonamide, tetracenomycin, tetracycline, trimethoprim, thiostrepton, colistin, tuberactinomycin and vancomycin (Figure 4.3A). Furthermore, the 646 ARG subtypes were associated with the five major antibiotic mechanisms including 1) antibiotic inactivation, 2) antibiotic target alteration, 3) antibiotic efflux, 4) antibiotic target protection and 5) antibiotic target replacement (Figure 4.3C) categorised based on resistance mechanisms listed on the CARD databases (Alcock et al. 2023). Among these assigned resistance mechanisms, the dominant resistance mechanism was for ARGs conferring antibiotic inactivation (45.7%(*n=295)*) > antibiotic target alteration (25.7%(*n=166)*) > antibiotic efflux (15.5%*(n=100)*) > antibiotic target alteration (7.4%*(n=48)*) > antibiotic target replacement (5.7%*(n=37)*).

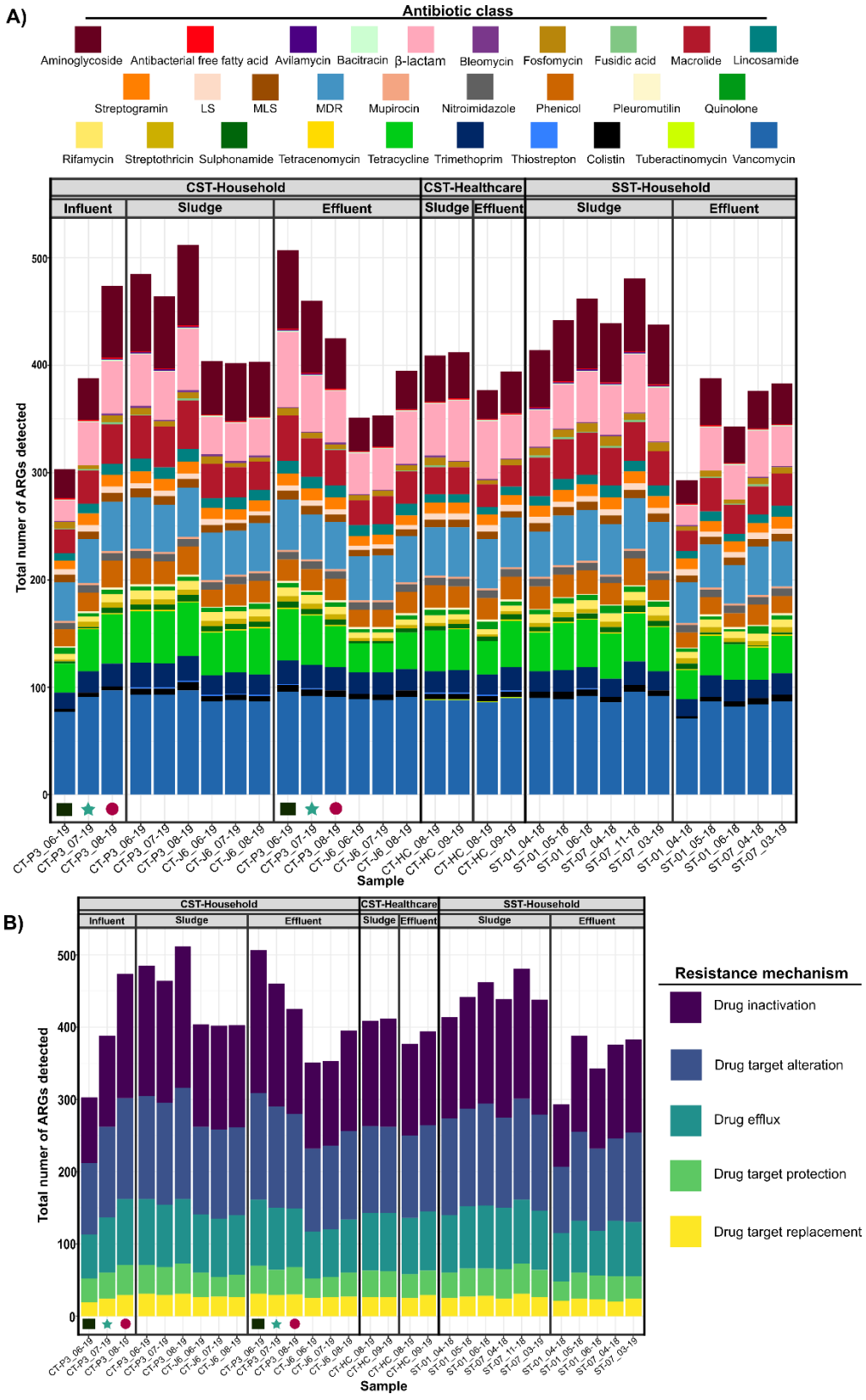

Figure 4.3: Detected ARGs in Thai wastewater samples (*n=30*), grouped by reactor type (CST-Household, CST-Healthcare, SST-Household) and sample type (influent, sludge, effluent). A) Total number of detected ARGs per sample. B) Detected gene resistance mechanisms. Each colour represents a different antibiotic class A)/ resistance mechanism B). CTP3 and CTJ6 samples are from two independent CST-Household reactors, CT-HC is from a CST-Healthcare tank, ST01 and ST07 are from two independent SST-Household units. Sampling month and year are shown as month_year (e.g., 06_19 for June 2019). LS: Lincosamide/Streptogramin; MLS: Macrolide/ Lincosamide/Streptogramin. □= influent and effluent for June. *=influent and effluent for July. O= influent and effluent for August.

The 113 stress genes, conferred resistance to 18 stress gene classes including Acid, heat, metals (copper, gold, silver, mercury, nickel, chromate, arsenic, tellurium, cadmium, fluoride, multi-metal resistance), biocide and metal resistance, biocides (quaternary ammonium, bacitracin and multi-biocide resistance) and drug and biocide resistance (Figure 4.4A). Three resistance mechanism including resistance protein, efflux and resistance regulator was categorised for the 113 stress genes based on the MEGARES database (Bonin et al., 2023) and NCBI reference gene catalogue databases (https://www.ncbi.nlm.nih.gov/pathogens/refgene/#: Last accessed 24th January 2023). Of these assigned stress genes mechanisms, the genes conferring resistance protein (49.6%(*n=56)*) > efflux (31%(*n=35*)) > resistance regulator (19.4%(*n= 22*)).

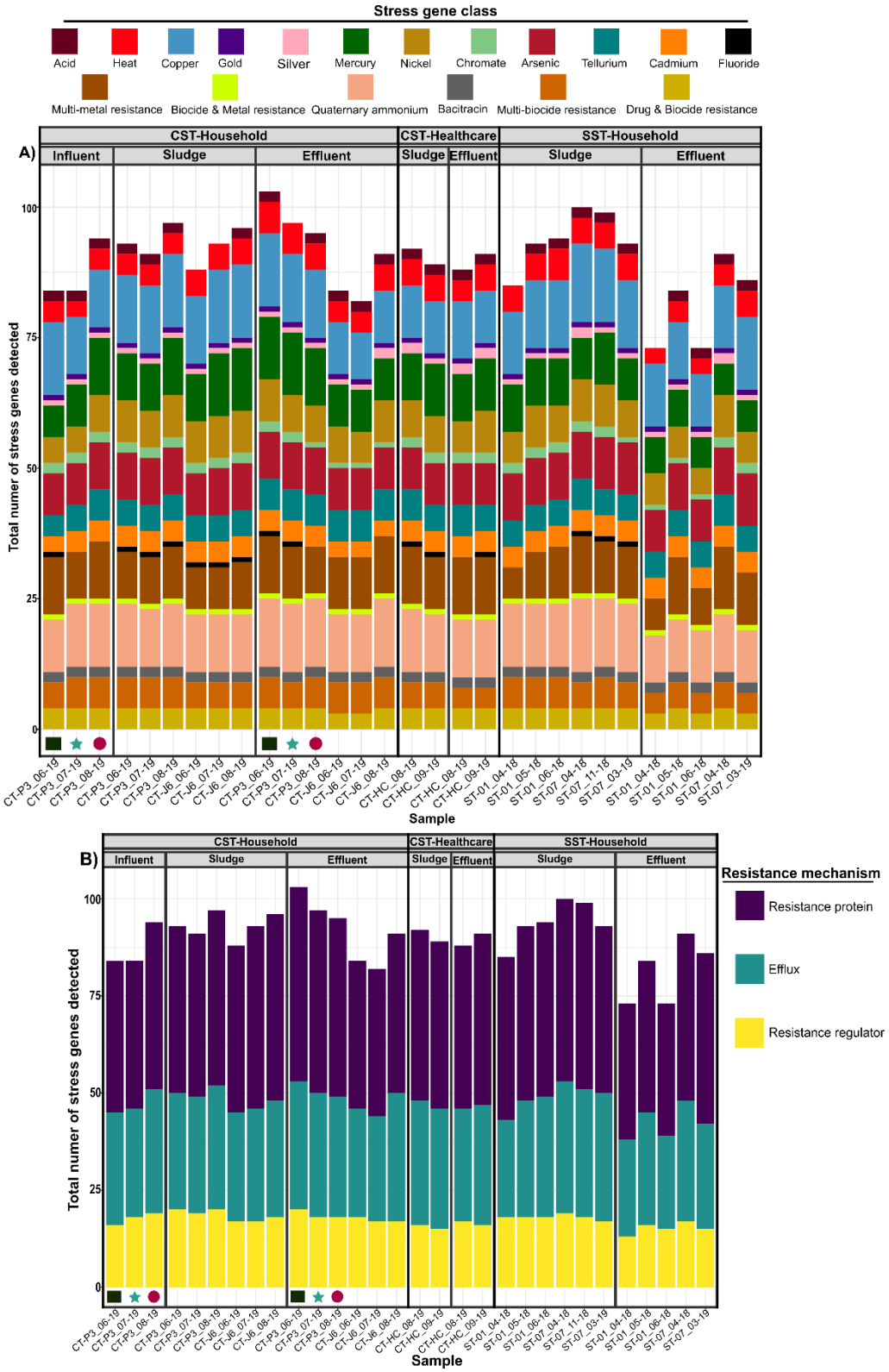

Figure 4.4: Detected stress genes in Thai wastewater samples (*n=30*) grouped by reactor type (CST-Household, CST-Healthcare, SST-Household) and sample type (influent, sludge, effluent). A) Total number of stress genes detected per sample. B) Detected gene resistance mechanisms. Each colour represents a different antibiotic class A)/ resistance mechanism B). CTP3 and CTJ6 samples are from two independent CST-Household reactors. CT-HC sample is from a CST-Healthcare tank. ST01 and ST07 are two independent SST-Household units. Sampling month and year are shown as month_year (e.g., 06_19 for June 2019). Heavy metals= Copper, Gold, Silver, Mercury, Nickel, Chromate, Arsenic, Tellurium, Cadmium and Fluoride. Biocides= Quaternary ammonium and Bacitracin. □ icon= influent and effluent for June. * =influent and effluent for July. O= influent and effluent for the August.

4.3.2.1 Risk assessment of ARGs and stress genes between the three tanks: CST-household unit higher contributor of ARGs and stress genes than the CST-healthcare and SST-household tanks

Regarding the detected ARG subtypes, 606, 490 and 575 ARGs were detected in the CST-household (influent, sludge, effluent), CST-healthcare (sludge, effluent) and SST-household (sludge, effluent) unit samples, respectively. Among the three tanks, 446 genes were shared (Figure 4.5A). Furthermore, only 37, 12 and 18 ARGs were exclusively unique to the CST-household, CST-healthcare, SST-household unit, respectively (Figure 4.5A). Additionally, 101 unique ARGs were shared between CST-household and SST-household unit, 10 unique ARGs were shared between SST-household and CST-healthcare unit, and 22 unique ARGs were common between CST-healthcare and CST-household tank (Figure 4.5A).

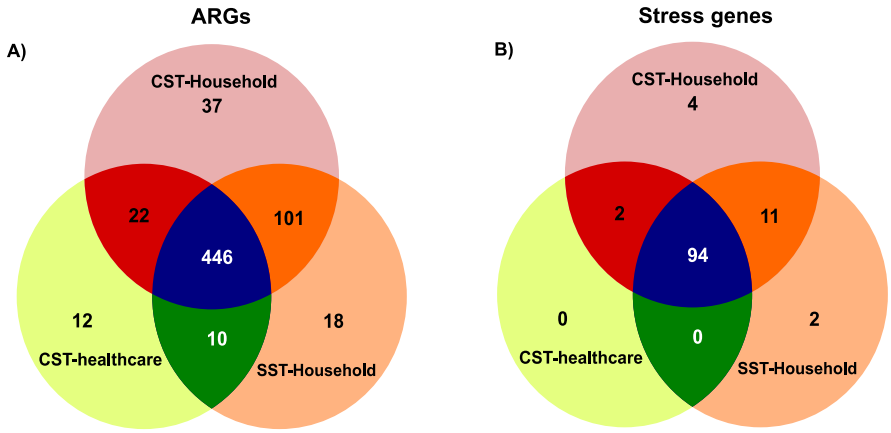

Figure 4.5: Shared and unique genes between the three septic tanks (CST-household, CST-healthcare, SST-household) and sample type (influent, sludge, effluent). A) shared and unique ARGs and B) shared and unique stress genes.

Similarly, for the 113 stress genes detected, 111, 96 and 107 stress genes were detected in the CST-household, CST-healthcare and SST-household tank samples respectively. Between the three septic tanks, 94 stress genes were shared (Figure 4.5B). Moreover, only four, none and two genes were exclusive to the CST-household, CST-healthcare, SST-healthcare reactors, respectively (Figure 4.5B). Furthermore, 11 unique genes were shared between the CST-household and SST-household unit, and two unique genes were common between CST-healthcare and CST-household tank, but none were shared between CST-healthcare and SST-household unit (Figure 4.5B).

4.3.2.2 Risk assessment of ARGs and stress genes between sample types (sludge and effluent) within each of the three septic tanks: ARGs and stress genes subtypes generally higher in the effluent than sludge

Consistently high numbers of shared ARGs and stress gene subtypes were observed between the sludge and effluent samples of each tank (Figure 4.6), indicating the sludge and effluent as important sources of rich and diverse AMR gene (ARGS and stress genes) to the environment following discharge. Of the three tanks, the CST-household tank was found to be the higher contributor of ARG subtypes via the sludge (*n=549*) and effluent (*n=552*) to the environment (Figures 4.5A), and a higher contributor of stress gene subtypes to the environment via its effluent (*n=107*) (Figures 4.5D). The CST-healthcare tank, on the other hand, was observed to be a lower contributor of ARGs and stress gene subtypes via its sludge (ARGs: *n=441*, stress genes: *n=95*) and effluent (*n=443*, stress genes: *n=94*) samples (Figures 4.5B, 4.5E).

Finally, the SST-household tank was found to be the higher contributor of stress gene subtypes to the environment via its sludge sample (*n=106*) (Figure 4.6F). The SST-household unit ARGs richness in the sludge (*n=541*) (Figure 4.6C) is marginally lower than that of the CST-household tank sludge (higher environmental contributor of ARGs and stress gene richness) (Figure 4.6A).

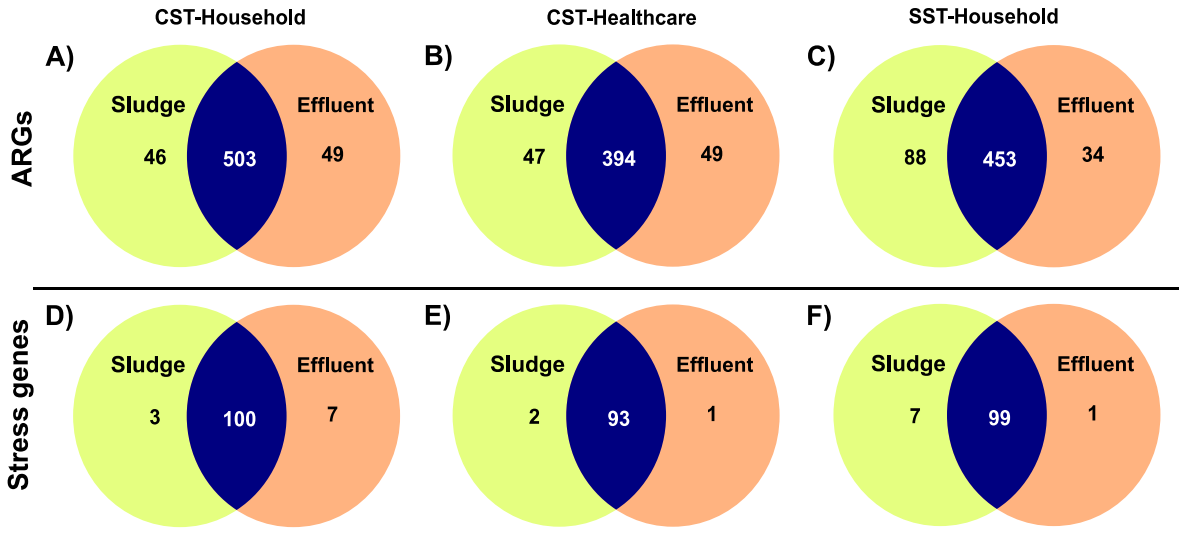

Figure 4.6: Venn diagram of shared and unique ARGs A-C) and stress genes D-F) between sludge and effluent for the three septic tank reactors (CST-Household, CST-Healthcare and SST-Household). CST denotes conventional septic tank; SST denotes solar septic tank.

ARGs conferring vancomycin, aminoglycoside, β-lactam, MDR and tetracycline resistance were the top five most frequently detected resistance gene subtypes between the sample types (sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household) (Figure 4.3A). Similarly, stress genes conferring arsenic, copper, mercury, multi-metal resistance and quaternary ammonium resistance were the top-most frequently detected stress genes between the sample types and reactors (Figure 4.4A). Vancomycin is listed by the World Health Organisation (WHO) (WHO 2021a) as an antibiotic class of last resort.

Of note, the *vph* gene, which confers resistance to tuberactinomycin, a second-line antibiotic used for the treatment of drug-resistance tuberculosis caused by *Mycobacterium tuberculosis* (Zane and Graeme 2022) was only detected in the CST-healthcare samples (sludge and effluent) (Figure 4.3A). Furthermore, genes conferring resistance to antibiotics deemed critical for human medicine by WHO (WHO 2021a) including cephalosporin and carbapenem (β-lactams), kanamycin (aminoglycoside), erythromycin (macrolide) acid and Rifamycin, were detected in all sample types (influent, sludge, effluent) between the reactors (Figure 4.3A). Moreover, genes conferring colistin resistance, another antibiotic class listed by WHO (WHO 2021a) as a last resort reserved only for the treatment of MDR bacteria, were detected in all sample types (influent, sludge, effluent) across the three reactors (Figure 4.3A).

Among the resistance mechanisms identified, antibiotic inactivation was the dominant resistance mechanism conferred by the ARG subtypes in the sludge and effluent for the three tanks (Figure 4.3B). Moreover, β-lactam and aminoglycosides antibiotics accounted for the majority of antibiotic inactivation resistance genes observed between sample type (sludge and effluent) and tank types (CST-household, CST-healthcare, SST-household). β-lactam antibiotic inactivation was observed to be predominately conferred through Class A beta-lactamase and subclass B3 metallo-β-lactamase, while aminoglycoside antibiotic inactivation was found to be conferred through N-acetyltransferase and O-phosphotransferase mostly.

Similarly, resistance protein was the dominant resistance mechanism conferred by the stress gene subtypes in the sludge and effluent for the three tanks (Figure 4.4B). Furthermore, genes conferring copper (*n=13,* out of 56) and mercury resistance (*n=9,* out of 56) contributed to the majority of resistance. Copper resistance protein mechanism was predominantly conferred via copper translocating P-type ATPase whereas, mercury resistance protein mechanism was primarily conferred through mercury transporter.

4.3.2.3 Risk assessment of ARGs and stress genes between influent and effluent for the septic tank unit (CST-household tank) with accessible influent sample: Highest enrichment of genes observed in June

Removal of ARG and stress gene subtypes was assessed for the conventional household septic tanks (CST-household- CT-P3 tank) with accessible influent. The diversity of the ARGs and stress genes subtypes conferring resistance to antibiotic and stress genes class remained similar for the influent and effluent samples between the three sampling months (June, July, August) (Figure 4.3A, 4.4A). The richness, however, was higher in the effluent than the influent for the three sampling months, indicating enrichment of ARGs and stress gene subtypes in the effluent (Figure 4.7). Thus, indicating the ineffectiveness of the CST-household tank in removing ARGs and stress genes from treated wastewater. Of the three sampling months, the highest enrichment of resistance gene subtypes in the effluent was observed for the June sampling month (ARGs: *n= 226*, stress genes: *n= 21*) (Figures 4.6A, 4.6E), while the number of enriched resistance genes steadily declined in the July (ARGs: *n= 105*, stress genes: *n= 17*) (Figures 4.6B, 4.6F).and August sampling months (ARGs: *n= 24*, stress genes: *n= 5*) (Figures 4.6C, 4.6G).

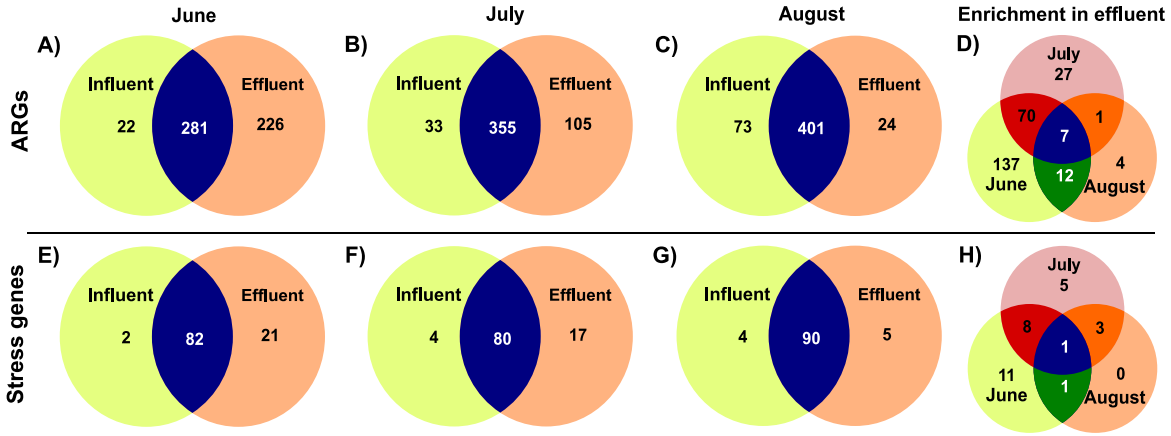

Figure 4.7: Share and unique ARGs (A-C) and stress genes (E-G) between the influent and effluent for the CST-household tank (CT-P3) with accessible influent sample for the three sampling months (June, July, August); and number of shared and unique ARGs (D) and stress genes (H) enriched in the effluent of the CST-household tank (CT-P3) between the three sampling months (June, July, August).

Of the enriched genes, only seven ARGs (*BlaPCD, blaPAM, Pen-I, blaPME, blaCPS, mcr-1, mcr-9*) (Figures 4.6D) and one stress gene (*merF*) (Figures 4.6H) was commonly enriched between the three-sampling months (June, July, August). *BlaPCD* and *blaPME confer resistance to* cephalosporin beta-lactam antibiotics. Cephalosporin is listed as an antibiotic of critical importance to human medicine (WHO 2021a), used for the treatment of gram-negative and positive bacterial infections. *blaPAM and blaCPS encode resistance to* carbapenem (β-lactam), an antibiotic class listed as critically importance to human health/last resort (WHO 2021a) and used for the treatment of complicated bacterial infection due to its broad-spectrum activity (Papp-Wallace et al. 2011). *Pen-I* is a class A beta-lactamase resistance gene while *mcr-1* and *mcr-9* genes are plasmid-borne colistin resistance genes antibiotics. *merF* stress gene confers resistance to mercury.

In summary, the number of detected ARGs and stress genes was higher in the CST-household> SST-household > CST-healthcare tank. AMR gene (ARGs and stress genes) subtypes were noted in the effluent of the CST-household tank compared to the influent, with the highest number of AMR subtypes observed in the June sampling month. Moreover, the number of enriched genes steadily decreased for the other two sampling months (July and August). The similar number of ARG and stress gene subtypes detected in the SST-household tanks (*n=682* combined ARG and stress gene total) compared to the CST-household tank (*n=717*, combined ARG and stress gene total) implies a limiting role of incorporated temperature within the SST-household unit in reducing AMR genes from the

treated wastewater. Finally, the detection of gene conferring resistance to critically important antibiotic for human use as well as antibiotics of last resort in all three tanks emphasise the need to improve the design of decentralised WWT to better remove AMR genes.

4.3.3 ARGs abundance are higher in CST-healthcare tank samples (sludge and effluent), while stress gene abundance are higher in SST-household samples (sludge and effluent)

Overall ARGs and stress genes abundance (count abundance per ng DNA) between the three tanks were high (Tables 4.3, 4.4)**.** Although, the overall stress gene abundance in all three tanks, which ranged from 347.1±893.8SD (CST-household) to 439.6±1167.7SD (SST-household) (Table 4.4), were higher than the overall ARGs, which ranged from 80.1±339.6SD (CST-household) to 103.5±485.1SD (CST-healthcare) (Table 4.3).

Table 4.3: ARGs abundance (in terms of count abundance) between reactors and sample types

Groups	Overall mean abundance (count abundance) between the three tanks						Overall mean abundance (count abundance) within each tank											
	CST-Household		CST-Healthcare		SST-Household		CST-Household						CST-Healthcare				SST-Household			
	Mean abundance (±SD)	Trend	Mean abundance (±SD)	Trend	Mean abundance (±SD)	Trend	Influent	Trend	Sludge	Trend	Effluent	Trend	Sludge	Trend	Effluent	Trend	Sludge	Trend	Effluent	Trend
Total ARGs	80.1±339.6	↓	103.5±485.1	↑	92.5±454.4	–	**62.1**±242.6	↓	**78.8**±351.5	–	**90.2±367.6**	↑	**95.6±458.9**	↓	**111.3±510.0**	↑	**87.9±406.5**	↓	**98.1±506.1**	↑
Aminoglycoside	21.6±106.8	↓	25.2±236.1	–	26.3±167.2	↑	24.1±123.1	↑	21.4±81.9	–	20.4±119.5	↓	24.7±200.3	↓	25.7±267.7	↑	21.4±82.4	↓	32.1±231.1	↑
Antibacterial free fatty acids	57.1±45.9	↓	184.3±224.6	↑	66.5±89.7	–	29.0±21.9	↓	39.5±23.2	–	88.8±55.7	↑	71.0±52.3	↓	297.5±311.8	↑	63.7±62.9	↓	70.0±123.0	↑
Avilamycin	0.4±0.5	–	0.0±0.0	↓	1.6±2.7	↑	0.3±0.6	–	0.5±0.5	–	0.3±0.5	–	0.0±0.0	–	0.0±0.0	–	3.0±3.0	↑	0.0±0.0	↓
Bacitracin	2134.7±919.5	–	1495.3±380.6	↓	2190.5±1137	↑	1232.0±560.3	↓	2877.8±789	↑	1842.8±604.3	–	1510.5±82.7	↑	1480.0±653.4	↓	2858±1161.1	↑	1390.0±274.8	↓
β-lactam	5.1±26.1	↓	6.9±28.8	↑	6.9±36.8	↑	3.8±14.4	–	3.4±13.5	↓	7.6±37.6	↑	6.0±21.9	↓	7.7±34.4	↑	4.4±20.3	↓	9.8±49.7	↑
Bleomycin	11.8±35.8	↑	2.0±5.1	↓	3.2±9.1	–	3.2±5.6	–	26.7±53.9	↑	1.3±2.6	↓	0.8±1.0	↓	3.2±7.3	↑	3.3±7.5	↑	3.0±11.1	↓
Fosfomycin	34.3±83.2	–	20.5±47.7	↓	61.2±232.5	↑	15.7±28.9	↓	53.0±115.0	↑	24.8±56.8	–	14.5±25.5	↓	26.5±62.7	↑	44.9±109.9	↓	80.9±323.8	↑
Fusidic acid	1.0±2.5	↓	1.5±3.5	–	6.3±9.8	↑	3.2±4.7	↑	0.8±1.6	–	0.3±0.6	↓	3.0±4.8	↑	0.0±0.0	↓	10.7±11.6	↑	1.1±2.2	↓
Macrolide	24.5±66.4	–	10.9±33.2	↓	31.8±244.1	↑	18.9±58.0	↓	28.3±67.0	↑	23.6±69.6	–	10.1±20.9	↓	11.8±42.1	↑	23.3±63.5	↓	42.1±355.4	↑
Lincosamide	12.4±25.9	–	12.2±26.4	↓	13.6±40.3	↑	14.5±47.4	↑	10.2±14.4	↓	13.6±19.2	–	12.2±20.8	–	12.2±31.4	–	13.1±20.4	↓	14.3±55.7	↑
Streptogramin	124.5±172.6	↓	255.9±471.1	↑	184.6±439.3	–	118.9±246.3	↓	121.6±117.2	–	130.2±177.4	↑	231.1±418.9	↓	280.6±526.9	↑	134.5±111.3	↓	244.8±638.2	↑
LS	76.9±217.0	–	78.6±166.2	↑	64.7±191.5	↓	59.9±97.1	–	36.1±53.9	↓	126.3±328.5	↑	61.2±133.3	↓	96.0±198.2	↑	27.1±32.2	↓	109.8±277.6	↑
MLS	75.9±141.3	–	29.9±59.7	↓	82.2±136.5	↑	27.0±60.9	↓	101.6±177.9	↑	74.7±123.2	–	20.9±31.4	↓	38.9±78.8	↑	109.6±143.3	↑	49.3±121.7	↓
MDR	104.4±242.2	–	155.8±621.4	↑	103.1±253.1	↓	100.0±276.3	–	72.8±144.6	↓	138.3±292.9	↑	110.9±476.7	↓	200.6±738.1	↑	87.7±179.1	↓	121.5±319.6	↑
Mupirocin	867.4±318.8	↓	1141.9±996.9	↑	1085.9±551.9	–	623.8±73.4	↓	851.6±366.8	–	1004.9±278.8	↑	1168.5±893.6	↑	1115.3±1232.2	↓	1123±549.7	↑	1041.0±580.8	↓
Nitroimidazole	87.8±144.9	–	457.1±1137.7	↑	48.6±65.4	↓	56.6±89.2	–	41.4±41.1	↓	149.7±202.6	↑	394.0±997.4	↓	520.3±1298.0	↑	50.6±62.7	↑	46.3±69.4	↓
Phenicol	55.3±125.2	–	48.9±126.5	↓	91.8±327.8	↑	42.9±108.4	↓	51.0±121.9	–	65.7±135.7	↑	39.0±74.2	↓	58.9±163.0	↑	77.0±207.3	↓	109.6±430.3	↑
Pleuromutilin	791.3±919.7	↓	1484.5±1953.9	↑	939.6±1225.1	–	823.5±926.2	–	566.5±657.0	↓	999.9±1142.4	↑	1544.3±2035.2	↑	1424.8±2181.0	↓	763.8±958.8	↓	1150.6±1512.5	↑
Quinolone	26.1±67.6	–	19.4±44.3	↓	69.2±384.1	↑	21.8±64.7	–	41.0±85.4	↑	13.5±42.9	↓	18.8±48.0	↓	19.9±41.8	↑	40.4±92.4	↓	103.8±562.7	↑
Rifamycin	118.4±190.5	–	52.2±82.7	↓	144.6±403.6	↑	54.0±98.3	↓	197.4±245.6	↑	71.5±126.5	–	60.3±103.2	↑	44.2±57.4	↓	127.5±167.4	↓	165.0±573.0	↑
Streptothricin	23.3±35.8	–	1.6±2.4	↓	121.0±433.4	↑	18.0±21.3	↓	26.9±39.7	↑	22.3±38.0	–	1.2±1.1	↓	1.9±3.2	↑	76.5±129.3	↓	174.3±630.0	↑
Sulphonamide	288.5±694.6	↓	397.6±988.2	–	405.0±980.4	↑	135.6±377.3	↓	406.8±873.4	↑	246.7±605.5	–	354.3±968.2	↓	440.8±1048.9	↑	413.9±995.4	↑	394.3±978.9	↓
Tetracenomycin	20.9±25.3	–	3.3±5.3	↓	41.4±45.0	↑	5.7±7.4	↓	41.7±27.6	↑	7.8±11.9	–	1.0±1.4	↓	5.5±7.8	↑	73.5±36.3	↑	2.8±2.2	↓
Tetracycline	137.7±606.6	–	117.0±469.9	↓	149.3±719.2	↑	73.8±284.0	↓	152.5±707.6	–	154.7±614.9	↑	88.8±369.9	↓	145.2±552.4	↑	146.2±741.6	↓	153.0±692.7	↑
Trimethoprim	62.4±211.1	–	46.7±142.6	↓	64.3±256.0	↑	72.1±272.3	–	39.6±116.4	↓	80.2±246.0	↑	42.7±92.4	↓	50.7±180.3	↑	39.6±102.9	↓	94.0±361.2	↑
Thiostrepton	0.6±1.3	–	1.4±2.4	↑	0.0±0.2	↓	0.0±0.0	↓	1.3±1.8		0.3±0.6	–	0.8±1.0	↓	2.0±3.4	↑	0.1±0.3	↑	0.0±0.0	↓
Colistin	7.3±19.2	–	20.2±61.7	↑	5.2±11.9	↓	5.8±14.7	↓	6.3±12.0	–	9.2±25.9	↑	14.3±35.1	↓	26.2±80.7	↑	4.3±9.7	↓	6.2±14.1	↑
Tuberactinomycin	0.0±0.0	↓	2.0±1.4	↑	0.0±0.0	↓	0.0±0.0	–	0.0±0.0	–	0.0±0.0	–	1.5±0.7	↓	2.5±2.1	↑	0.0±0.0	–	0.0±0.0	–
Vancomycin	186.9±557.1	↓	283.1±848.5	↑	207.1±778.8	–	152.4±414.9	↓	178.0±532.6	–	213.1±636.9	↑	289.0±870.0	↑	277.2±828.5	↓	219.1±679.9	↑	192.6±883.6	↓

LS= Lincosamide/Streptogramin; MLS= Macrolide/Lincosamide/Streptogramin; MDR= Multi-drug resistance

Table 4.4: Stress abundance (in terms of count abundance) between reactors and sample types

Groups	Overall mean abundance (count abundance) between the three tanks						Overall mean abundance (count abundance) within each tank											
	CST-Household		CST-Healthcare		SST-Household		CST-Household						CST-Healthcare				SST-Household			
	Mean abundance (±SD)	Trend	Mean abundance (±SD)	Trend	Mean abundance (±SD)	Trend	Influent	Trend	Sludge	Trend	Effluent	Trend	Sludge	Trend	Effluent	Trend	Sludge	Trend	Effluent	Trend
Total stress gene	**347.1±893.8**	↓	**381.2±988.9**	–	**439.6±1167.7**	↑	**325.6±868.5**	↓	**345.2±894.1**	–	**359.9±906.9**	↑	**336.8±897.0**	↓	**425.6±1073.1**	↑	**376.5±936.7**	↓	**515.3±1392.4**	↑
Acid	6.4±11.5	↓	11.3±4.6	↑	11.1±20.9	–	8.0±5.2	–	1.7±1.8	↓	10.4±17.0	↑	8.0±2.2	↓	14.5±4.1	↑	9.3±17.2	↓	13.4±25.6	↑
Heat	784.0±1430.7	↑	668.4±1548.4	↓	749.1±1382.7	–	579.9±1138.9	↓	667.3±1102.3	–	1002.8±1809.8	↑	558.8±1226.6	↓	778.0±1866.1	↑	795.7±1395.1	↑	693.1±1389.3	↓
Copper	649.2±1438.8	–	423.9±803.5	↓	675.9±1340.9	↑	703.9±1698.5	–	738.0±1602.7	↑	533.0±1090.4	↓	390.4±718.7	↓	457.4±890.6	↑	692.8±1524.4	↑	655.7±1089.9	↓
Gold	250.5±290.4	↓	258.3±155.2	–	495.3±1008.6	↑	438.0±659.3	↑	187.8±112.2	↓	219.5±150.8	–	153.5±120.9	↓	363.0±117.4	↑	300.5±445.4	↓	729.0±1473.1	↑
Silver	281.5±643.4	↑	166.6±206.4	↓	237.5±384.0	–	633.2±1406.5	↑	224.8±247.6	–	162.5±211.0	↓	97.0±111.4	↓	236.3±272.1	↑	279.9±453.0	↑	186.7±296.5	↓
Mercury	72.9±184.8	↑	36.3±82.4	↓	52.6±100.4	–	117.6±329.0	↑	48.9±66.1	↓	74.7±163.3	–	29.3±60.9	↓	43.3±100.1	↑	49.8±91.9	↓	55.8±110.4	↑
Nickel	69.7±148.7	–	67.9±146.9	↓	97.3±305.3	↑	76.7±238.8	↑	76.2±114.6	–	59.7±120.9	↓	55.1±99.3	↓	80.7±184.9	↑	76.7±177.1	↓	122.1±410.6	↑
Chromate	443.4±523.1	↑	135.8±255.9	↓	409.9±487.6	–	357.7±442.0	↓	361.8±394.2	–	567.8±670.2	↑	59.3±68.0	↓	212.3±364.1	↑	373.0±446.9	↓	454.1±553.7	↑
Arsenic	897.8±1198.7	↓	925.1±1310.8	–	1227.6±1939.6	↑	700.6±885.5	↓	952.5±1244.4	↑	941.8±1292.2	–	835.7±1282.4	↓	1014.5±1365.8	↑	952.7±1261.7	↓	1557.6±2499.6	↑
Tellurium	43.6±44.7	↓	579.5±1263.3	↑	185.0±457.7	–	44.7±48.2	–	53.5±51.6	↑	33.2±33.0	↓	534.8±1127.8	↓	624.3±1435.4	↑	242.9±488.5	↑	115.5±415.2	↓
Cadmium	1100.6±1236.9	↓	2177.6±2435.7	↑	1460.3±2064.3	–	879.2±806.6	↓	917.1±892.1	–	1394.9±1632.9	↑	2010.8±2341.3	↓	2344.4±2677.2	↑	998.7±1009.5	↓	2014.2±2796.2	↑
Fluoride	4.3±7.1	↑	3.5±5.1	–	2.8±6.4	↓	0.3±0.6	↓	10.3±8.4	↑	0.3±0.5	↓	1.5±0.7	↓	5.5±7.8	↑	5.2±8.2	↑	0.0±0.0	↓
Multi-metal resistance	357.2±847.8	↓	447.4±1048.3	↑	428.2±1024.5	–	304.1±683.1	↓	367.6±893.9	–	373.3±883.1	↑	415.0±1005.6	↓	479.9±1108.3	↑	463.8±1077.0	↑	385.4±964.4	↓
Biocide & Metal resistance	283.5±337.9	–	312.5±265.7	↑	232.1±393.0	↓	455.7±662.1	↑	58.8±17.0	↓	422.0±206.6	–	133.5±105.4	↓	491.5±269.4	↑	172.8±132.9	↓	303.2±593.6	↑
Quaternary ammonium	79.0±198.4	–	72.5±191.2	↓	139.4±576.1	↑	54.8±193.5	↓	83.6±158.7	–	86.6±234.4	↑	59.9±171.7	↓	85.1±211.1	↑	93.8±235.2	↓	194.2±814.5	↑
Bacitracin	31.4±54.4	↓	44.6±38.4	–	349.4±976.7	↑	18.3±11.3	↓	24.8±35.0	–	44.6±78.7	↑	58.0±48.5	↑	31.3±24.8	↓	49.8±74.1	↓	708.9±1398.5	↑
Multi-biocide resistance	92.9±199.2	↓	120.8±251.4	–	245.6±1319.2	↑	172.7±366.3	↑	73.4±108.6	–	72.5±140.7	↓	46.3±55.0	↓	195.3±342.1	↑	67.0±129.1	↓	460.0±1947.6	↑
Drug & Biocide resistance	117.1±160.0	↓	210.6±441.2	–	424.9±1430.0	↑	74.2±124.9	↓	100.7±134.7	–	154.8±193.2	↑	167.8±277.8	↓	253.4±579.4	↑	157.2±230.7	↓	746.2±2089.2	↑

The results of the beta diversity analysis based on a Bray-Curtis distance dissimilarity matrix indicated that overall abundance of ARGs and stress genes clustered based on the tank type they originate from (i.e., CST-household, CST-healthcare, and SST-household) (Appendices Figure C.1). Specifically, ARGs abundance in the household (CST-household and SST-household) tanks were more similar to each other, while the ARGs abundance in the healthcare (CST-healthcare) tank appeared different from the household tank (Table 4.3, Appendices Figure C.1A). Conversely, overall stress gene abundance in the CST-healthcare and SST-household tanks appeared more similar than in the CST-household tank (Table 4.4, Appendix Figure C.1B).

Permutational multivariate analysis of variance (PERMANOVA) showed that only 14.9% (p-value <0.001) and 12.1% (p-value =0.06) of overall ARGs and stress genes variance, respectively, could be explained by the different tank type (CST-household, CST-healthcare, SST-household) groups (Appendix Figure C.1).

ARGs conferring resistance to bacitracin, mupirocin and pleuromutilin were the top three most abundant (in terms of count abundance) between the three tanks, despite not being among the most frequently detected ARGs subtypes (see section 4.3.2.2). Among these top three most abundance ARGs, genes conferring bacitracin resistance had the most count abundance (CST-household: 2134.7±919.5SD, CST-healthcare: 1495.3±380.6SD, SST-household: 2190.5±1136.6SD count abundance) (Table 4.3). ARGs conferring Mupirocin resistance genes were the second most ARGs in the two of the tanks (CST-household: 867.4±318.8SD, SST-household:1085.9±551.9SD) and the third most abundance in the healthcare tank (CST-healthcare: 1141.9±996.9SD count abundance) (Table 4.3).

Lastly, ARGs conferring pleuromutilin resistance genes were the third most abundant count in the two household units (CST-household: 791.3±919.7SD, SST-household: 939.6±1225.1SD count abundance) but the second most abundant in the healthcare unit (1484.5±1953.9SD count abundance) (Table 4.3).

Similarly, stress genes conferring resistance to arsenic and cadmium (metal resistance) and heat were the top three most abundant genes between the three tanks (Table 4.4). Only genes conferring arsenic resistance were among the most frequently detected in between the tanks (see section 4.3.2.2). Among the three most abundance stress genes, cadmium resistance was the most abundant in the three tanks (CST-household: 1100.6±1236.9SD, CST-healthcare: 2177.6±2435.7SD, SST-household: 1227.6±1939.6) (Table 4.4). Stress gene conferring arsenic resistance was the second most abundant (CST-household: 897.8±1198.7SD, CST-

healthcare: 925.1±1310.8SD, SST-household: 1460.3±2064.3SD), while heat resistance gene was the third most abundant amongst the tanks (CST-household: 784.0±1430.7SD, CST-healthcare: 668.4±1548.4SD, SST-household: 749.1±1382.7SD) (Table 4.4).

Regarding sources (sludge and effluent) of ARGs and stress genes to the environment, the three tanks had somewhat similar ARGs (Table 4.3) and stress genes (Table 4.4) abundance in their sludge and effluent.

Of the three tanks, the CST-healthcare tank had the highest overall ARG count abundance in the sludge (95.6±458.9SD) and effluent (111.3±510.0SD) (Tables 4.3), but a lower stress genes count abundance in the sludge (336.8±897.0SD) (Tables 4.4). This likely suggests that the CST-healthcare likely contributes to a higher ARG abundance to the environment through sludge and effluent, among the three tanks, but not stress genes via the sludge. In contrast, the SST-household tank unit had the highest overall stress gene count abundance in the sludge (78.8±351.5SD) and effluent (90.2±367.6SD) (Table 4.4) among the three reactors, implying that it likely contributes to a higher stress gene abundance to the environment compared to the conventional tanks (CST-household and CST-healthcare).

The conventional household tank (CST-household), among the three tanks, was the lower contributor of ARGs abundance to the environment via its sludge (345.2±894.1SD) and effluent (359.9±906.9SD) sample (Table 4.3) and a lower contributor of stress gene abundance to the environment via its effluent sample (359.9±906.9SD) (Table 4.3).

Beta-diversity analysis showed that the septic tank wastewater samples clustered by sample type (influent, sludge, effluent) and tank type (CST-household, CST-healthcare, SST-household) for both the ARGs (Figures 4.5A) and stress genes (Figures 4.5B). Although, PERMANOVA analysis indicated that the sample groups (based on tank type and sample type) only explain 36.9% (p-value <0.001) and 28.1% (p-value = 0.068) of variance in ARGs and stress gene abundance, respectively (Figure 4.8). The remaining 62.1% (ARGs) and 71.9% (stress genes) that were not explained by the sample grouping could potentially be explained by other factors such as environmental (i.e., temperature, pH), and chemical (such as Dissolved organic carbon) parameters measured but not focused on in this study.

The abundance of ARGs within the effluent of the three tanks (CST-household, CST-healthcare, SST-household) appears similar to each other and similar to the CST-household influent (Figure 4.8A). One SST-household effluent sample (ST-01_04-18) appeared to have very dissimilar ARGs abundance to all other samples (Figure 4.8A).

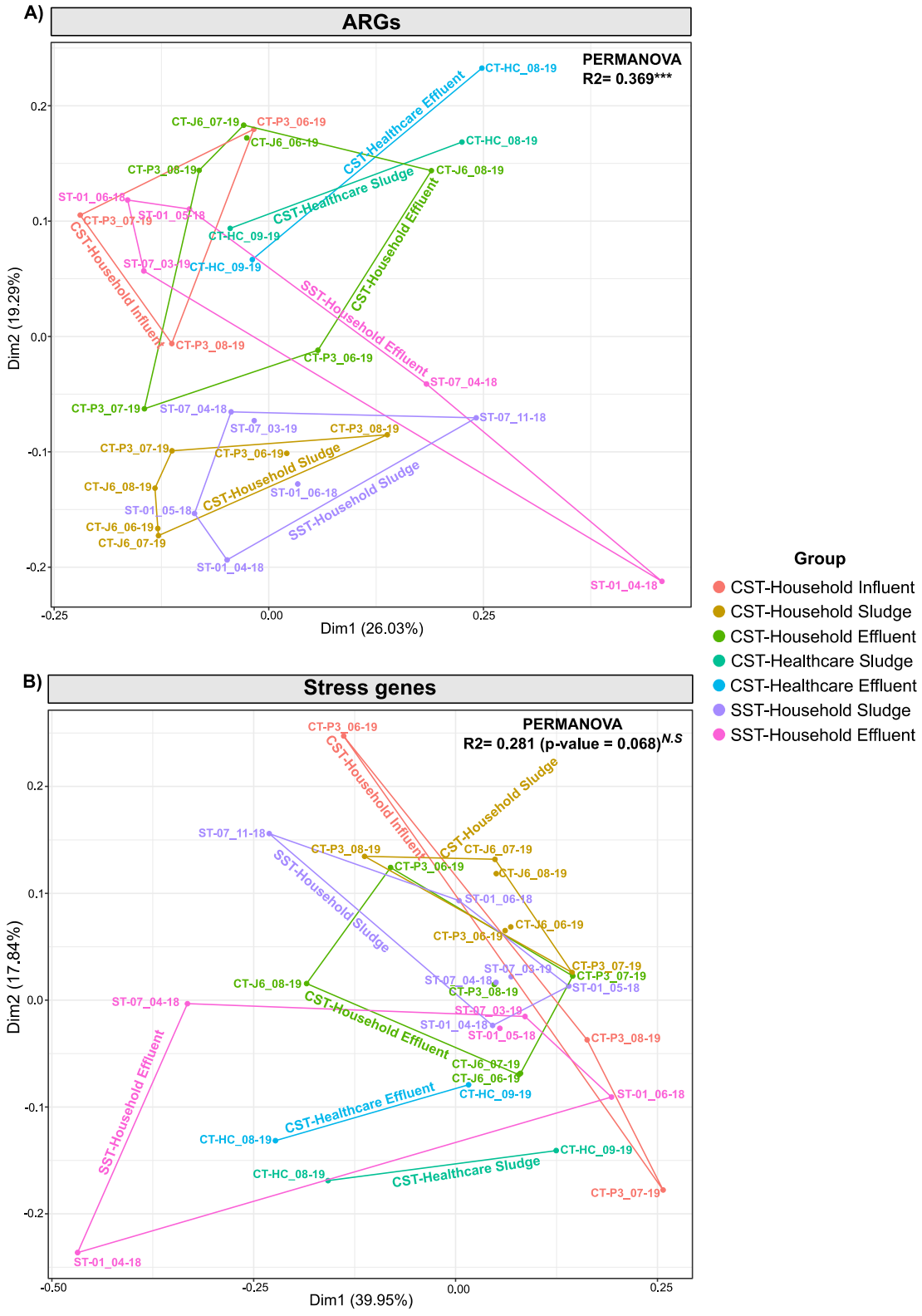

Figure 4.8: A non-metric dimensional scaling (NMDS) based on Bray-Curtis dissimilarity index of ARGS A) and stress gene B) abundance between sample types (influent, sludge, effluent) and reactor type (CST-household, CST-healthcare, SST-household). *** indicates p-value <0.001.N.S= Not statistically significant.

Similarly, the stress gene abundance in the effluent of the three tanks appeared similar to each other (Figure 4.8B). Although, some of the effluent samples from the different tanks clustered closer to the SST-household sludge and CST-household influent indicating some similarities in stress gene abundance (Figure 4.8B).

The household tanks (CST-household and SST-household) sludge samples appeared to have similar ARGs (Figure 4.8A) and stress gene (Figure 4.8B) abundance and were distinct from the other sample type (influent and effluent) and CST-healthcare tank sample (sludge and effluent). In addition, the stress genes abundance within the CST-healthcare sludge (CT-HC_09_19) appears to be more similar to the CST-household influent and effluent (Figure 4.8B).

Overall, the abundance of ARGs and stress genes in the CST-healthcare sludge and effluent appear different from that of the household (CST-household and SST-household) tanks sludge and effluent samples (Figure 4.8).

To further explore and understand how ARGs/ stress gene abundances within each sample group (samples grouped based on sample type and reactor type) differed from the average beta diversity, the local contribution to beta diversity (LCBD), which gives sample-wise contributions to beta diversity was calculated (McKenna et al. 2020). In addition, LCBD showed how each sampling month within each sample type group differs from the average. A higher LCBD value indicates that the diversity of a sample type is different from the others.

Between the three reactors, higher LCBD values, which were generally above the mean beta-diversity value (thick black line on Figure 4.9), were observed for the healthcare unit (CST-healthcare) samples (sludge: ARGs- 0.032 to 0.04, Stress gene: 0.03 to 0.051; effluent: ARGs- 0.037 to 0.048, Stress gene: 0.04 to 0.054) as compared to the two household tanks (Figures 4.8A, 4.8B). In addition, statistical difference (p-value <0.05) was observed between the CST-healthcare samples (sludge and effluent) and the other tank samples, specifically the CST-household sludge and effluent and SST-household effluent sample. Thus, indicating that the CST-healthcare samples, particularly the effluent, contributed most to beta diversity. In addition, these higher LCBD values (Figure 4.9) confirmed that the CST-healthcare unit sample (sludge and effluent) abundance was different from the two household tanks (CST-household and SST-household) as observed by the beta diversity (Figure 4.9).

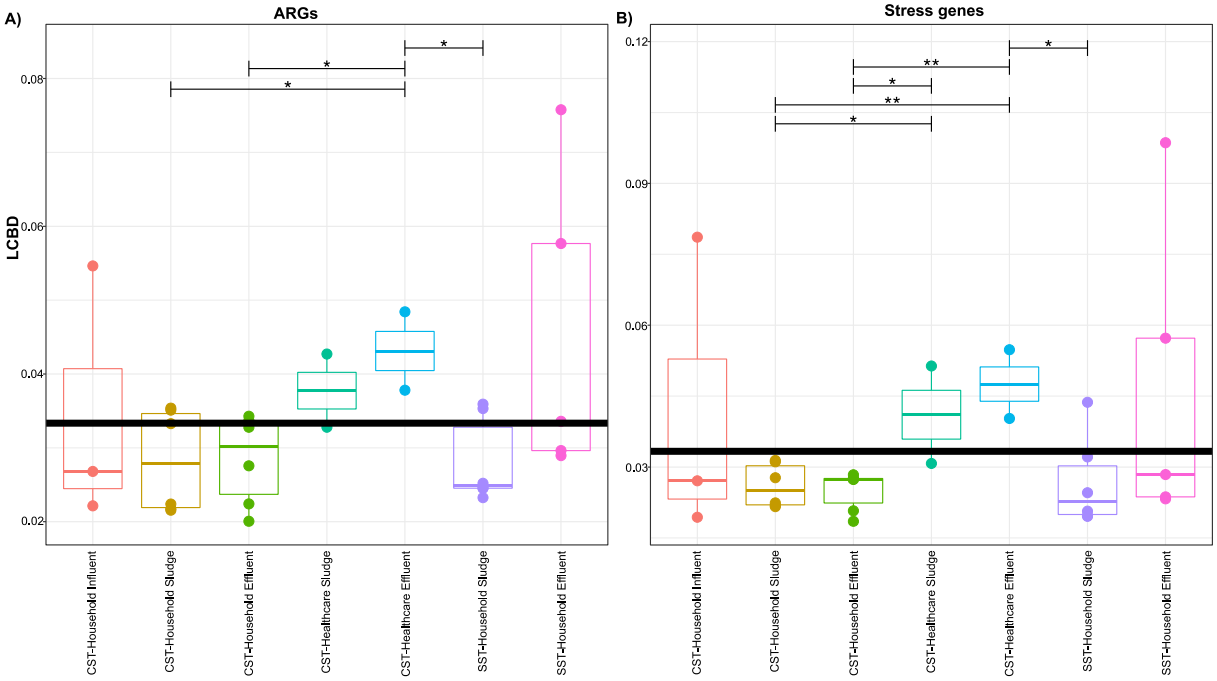

Figure 4.9: Local contribution to beta diversity for ARGs A) and stress genes B). Samples were group based on sample type (influent, sludge, effluent) and reactor type (CST-household, CST-healthcare, SST-household). * p-value <0.05; ** p-value <0.01. Thick black line indicates mean beta diversity.

Both household tanks (CST-household and SST-household) samples (influent, sludge, effluent) tend to have LCBD values that were slightly below or close to the mean beta diversity (Figure 4.9), indicating that the ARG and stress genes abundance were somewhat similar and would explain why the samples from these two tanks tend to cluster closer together (Figure 4.8).

Despite the lower ARGs and Stress genes richness detected in the SST-household unit (compared to the CST-household tanks) the close clustering between the tanks (CST-household and SST-household) sludge, suggests very limiting role of increased temperature within the SST unit in reducing the abundance of ARGs and stress gene.

4.3.4 Risk assessment of ARGs and stress genes between influent and effluent for septic tank unit (CST-household tank) with accessible influent sample: Highest abundance of enriched ARGs observed in June

Overall, ARG abundance (in terms of count abundance) was higher in the effluent than influent for the three sample months (June, July, August), with the highest abundance found in June (107±416SD) > August (74±266SD) > July (58±239SD) month (Table 4.5). On a class-by-class basis, a higher number of antibiotic classes were enriched in July (*n=22*) > July (*n=19*) > August (*n=11*) month (Table 4.5)

For the stress genes, higher overall abundance was found in the effluent than influent for only two of the sampling months (July and August) with the abundance in August (329±786SD) higher than the July month (259±633SD) (Table 4.5). In addition, the highest number of enriched stress gene classes in the effluent was also found in August (*n=14*) > July (*n=13*) > June (*n=6*) (Table 4.5). Taken together, this once again, showed that the septic tanks are ineffective for reducing ARGs and stress abundance from wastewater and are enriching AMRs in the effluent.

Table 4.5: Enrichment of ARG and stress gene abundance for the CST-household tank (CT-P3) with accessible influent sample

Group	ARGs								Stress genes								
	June			July			August			June			July			August	
	Influent	Effluent	Trend	Influent	Effluent	Trend	Influent	Effluent	Trend		Influent	Effluent	Trend	Influent	Effluent	Trend	Influent	Effluent	Trend
Total mean ARGs	**78±299**	**107±416**	↑	**42±157**	**58±239**	↑	**67±249**	**74±266**	↑	**Total mean stress gene**	**583±1297**	**463±1103**	↓	**149±369**	259±633	↑	**244±594**	**329±786**	↑
Aminoglycoside	45±186	29.3±134.1	↓	12.6±78.0	14.1±58.0	↑	14.9±67.5	21.3±122.2	↑	Acid	6.0±0.0	1.0±0.0	↓	3.5±0.7	0.0±0.0	↑	14.5±0.7	4.0±0.0	↓
Antibacterial free fatty acids	20±0.0	125.0±0.0	↑	13.0±0.0	70.0±0.0	↑	54.0±0.0	49.0±0.0	↓	Heat	625.5±1489.1	1348±2294	↑	454.8±870.8	739.5±1218.4	↑	659.3±1185.5	968.3±1983.4	↑
Avilamycin	0.0±0.0	1.0±0.0	↑	0.0±0.0	1.0±0.0	↑	1.0±0.0	0.0±0.0	↓	Copper	1468.6±2673.8	842.0±1656	↓	199.3±372.9	511.4±1033.7	↑	443.8±885.1	478.8±1043.2	↑
Bacitracin	909±0.0	2559.0±0.0	↑	908.0±0.0	1522.0±0.0	↑	1879.0±0.0	1359.0±0.0	↓	Gold	1199.0±0.0	165±0.0	↓	39.0±0.0	82.0±0.0	↑	76.0±0.0	450.0±0.0	↑
β-lactam	2.7±11.3	10.7±51.3	↑	4.2±15.6	5.5±33.1	↑	4.5±15.8	8.3±40.2	↑	Silver	1749.5±2474.2	316.0±446.9	↓	42.5±60.1	178.5±252.4	↑	107.5±152.0	107.0±151.3	↓
Bleomycin	4.3±7.5	3.3±5.8	↓	0.0±0.0	1.7±2.9	↑	5.3±6.8	1.7±2.1	↓	Mercury	302.1±532.9	91±115	↓	18.0±27.6	47.8±62.1	↑	32.7±50.7	66.9±74.2	↑
Fosfomycin	13.7±21.1	33.4±62.7	↑	9.0±15.7	16.7±31.4	↑	24.3±42.9	19.6±34.7	↓	Nickel	153.7±407.6	115±188	↓	32.4±48.5	55.6±94.0	↑	43.9±82.3	78.0±150.4	↑
Fusidic acid	0.5±0.7	0.0±0.0	↓	0.5±0.7	0.0±0.0	↓	8.5±4.9	0.5±0.7	↓	Chromate	567.0±763.7	537±734	↓	221.5±311.8	259.5±350.0	↑	284.5±400.9	464.0±656.2	↑
Macrolide	19.5±72.2	39.4±83.9	↑	12.4±27.8	21.8±50.8	↑	24.7±64.4	19.3±45.6	↓	Arsenic	1020±1145	1243±1606	↑	421.8±549.2	678±880	↑	660.2±844	848±1033	↑
Lincosamide	30.3±81.1	18.3±29.9	↓	4.1±5.0	10.1±18.7	↑	9.0±10.4	9.5±11.2	↑	Tellurium	28.7±28.5	48.5±22.8	↑	43.5±52.6	26.2±15.1	↓	62.0±60.4	31.7±12.5	↓
Streptogramin	214±411	135.7±82.2	↓	50.4±45.9	70.4±43.5	↑	92.7±84.5	75.1±68.2	↓	Cadmium	1184±919	1078±1145	↓	684.0±788.9	501.0±530.8	↓	769.3±853.0	1225±1247	↑
LS	68.8±124	66.3±100.0	↓	56.3±88.8	26.3±38.5	↓	54.7±92.7	85.3±150.1	↑	Fluoride	1.0±0.0	1.0±0.0	–	0.0±0.0	1.0±0.0	↑	0.0±0.0	0.0±0.0	–
MLS	9.0±8.8	123.6±199.0	↑	15.4±23.9	83.1±146.1	↑	56.5±100.1	20.4±22.7	↓	MMR	523±950	527±1176	↑	130.2±304.3	318.2±702.3	↑	259.5±628.9	297.5±622.4	↑
MDR	169.5±436	146.5±241.2	↓	62.8±137.3	81.5±134.3	↑	67.6±125.6	150±312	↑	BMR	1220.0±0.0	292±0.0	↓	58.0±0.0	135.0±0.0	↑	89.0±0.0	619.0±0.0	↑
Mupirocin	650.5±46	1428.0±22.6	↑	550.5±0.7	702.0±7.1	↑	670.5±91.2	970.0±189.5	↑	Quaternary ammonium	92.7±320.7	95.0±185.9	↑	19.7±49.3	56.9±108.4	↑	52.1±97.6	74.1±153.5	↑
Nitroimidazole	81.7±149	75.6±72.1	↓	39.4±39.7	37.6±37.0	↓	48.6±41.6	75.9±81.1	↑	Bacitracin	28.0±15.6	14.0±2.8	↓	12.0±9.9	6.5±0.7	↓	15.0±2.8	30.5±33.2	↑
Phenicol	38.6±97.3	79.5±161.5	↑	27.9±48.2	38.9±80.2	↑	62.4±153.7	60.8±110.5	↓	MBR	422±569	90±173	↓	43.2±96.0	40.0±71.2	↓	53.3±106.0	124.3±263.3	↑
Pleuromutilin	1011±14308	1031±1456	↑	682.5±962.4	396.0±560.0	↓	777.0±1097.4	839.5±1181.6	↑	DBR	106.0±182.8	116.8±184.0	↑	70.5±132.4	49.3±73.4	↓	46.0±61.7	131.0±158.8	↑
Quinolone	10.0±9.9	36.5±91.3	↑	4.1±7.1	17.8±44.3	↑	51.1±110.0	7.5±15.0	↓	MMR= Multi-metal resistance BMR= Biocide & Metal resistance MBR= Multi-biocide resistance DBR= Drug & Biocide resistance									
Rifamycin	84.8±151.2	141.9±205.9	↑	22.7±35.3	90.6±144.7	↑	54.6±71.4	45.8±59.7	↓										
Streptothricin	11.6±20.4	57.6±69.9	↑	13.4±17.6	35.0±44.2	↑	29.0±25.0	18.8±21.7	↓										
Sulphonamide	50.7±120.2	419.0±1002	↑	102.3±250.7	262.7±632.1	↑	253.7±616.5	197.3±478.5	↓										
Tetracenomycin	0.0±0.0	28.0±0.0	↑	3.0±.0.0	17.0±0.0	↑	14.0±0.0	2.0±0.0	↓										
Tetracycline	61±245	169.2±757.1	↑	54.8±195.6	93.9±437.8	↑	105.5±380.9	99.8±326.6	↓										
Trimethoprim	127.8±443.6	78.3±240.3	↓	44.0±125.1	34.4±113.4	↓	44.7±103.1	77.9±252.3	↑										
Thiostrepton	0.0±0.0	1.0±1.4	↑	0.0±0.0	0.5±0.7	↑	0.0±0.0	0.0±0.0	–										
Colistin	7.2±15.8	11.8±20.7	↑	7.2±20.0	5.5±11.0	↓	3.0±5.8	24.0±56.3	↑										
Tuberactinomycin	0.0±0.0	0.0±0.0	–	0.0±0.0	0.0±0.0	–	0.0±0.0	0.0±0.0	–										
Vancomycin	179.7±503.4	247.5±693.1	↑	106.3±274.4	140.8±405.0	↑	171.2±433.6	172.0±460.3	↑										

4.3.5 *tetA*(58) (ARG) and *copR* (stress gene) were the most abundant genes in CST-household and SST-household tanks, from the top 25 most abundant ARGs and stress genes, while *vanR-I* and *cadA* were the most abundant in CST-healthcare unit

The most abundant (in terms of count abundance) ARGs and stress genes between the sample types (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household) were somewhat similar (Figure 4.10). The top 25 most abundant ARGs (Figure 4.10A) belong to ten antibiotic classes namely: tetracycline (*tetA*(58), *tet*(37), *tetB*(46)), vancomycin (*vanR-I, vanR-F, vanRM, vanR-A, vanR-E, vanR-G, vanR-Cd, vanR-C, vanU-G, vanR-B, vanR-D*), bacitracin (*bcrA*), sulphonamide (*sul4*), pleuromutilin (*taeA*), MDR (*bepE, bepG, emhC*), mupirocin (*mupB, mupA*), aminoglycoside (*RanA*), trimethoprim (*dfrA3*) and phenicol (*estDL136*), six of which were not amongst the top five most frequently identified antibiotic class.

Similarly, the top 25 most abundant stress genes (Figure 4.10B) belong to ten stress gene classes including copper (*copR, pcoR, cueA, copB, copA*), arsenic (*arsR, arsC, acr3, arsB, arsA*), heat(*trxLHR, shsP, hsp20*), multi-metal resistance (*goLT, silR, silA*), cadmium (*cadA, cadC*) quaternary ammonium (*smr, chrR*), drug and biocide (*mtrF*), silver (*silP*), multi-biocide resistance (*smdA*) and tellurium (*terZ, terD*), again, six of which were not amongst the top five most frequently identified stress gene class.

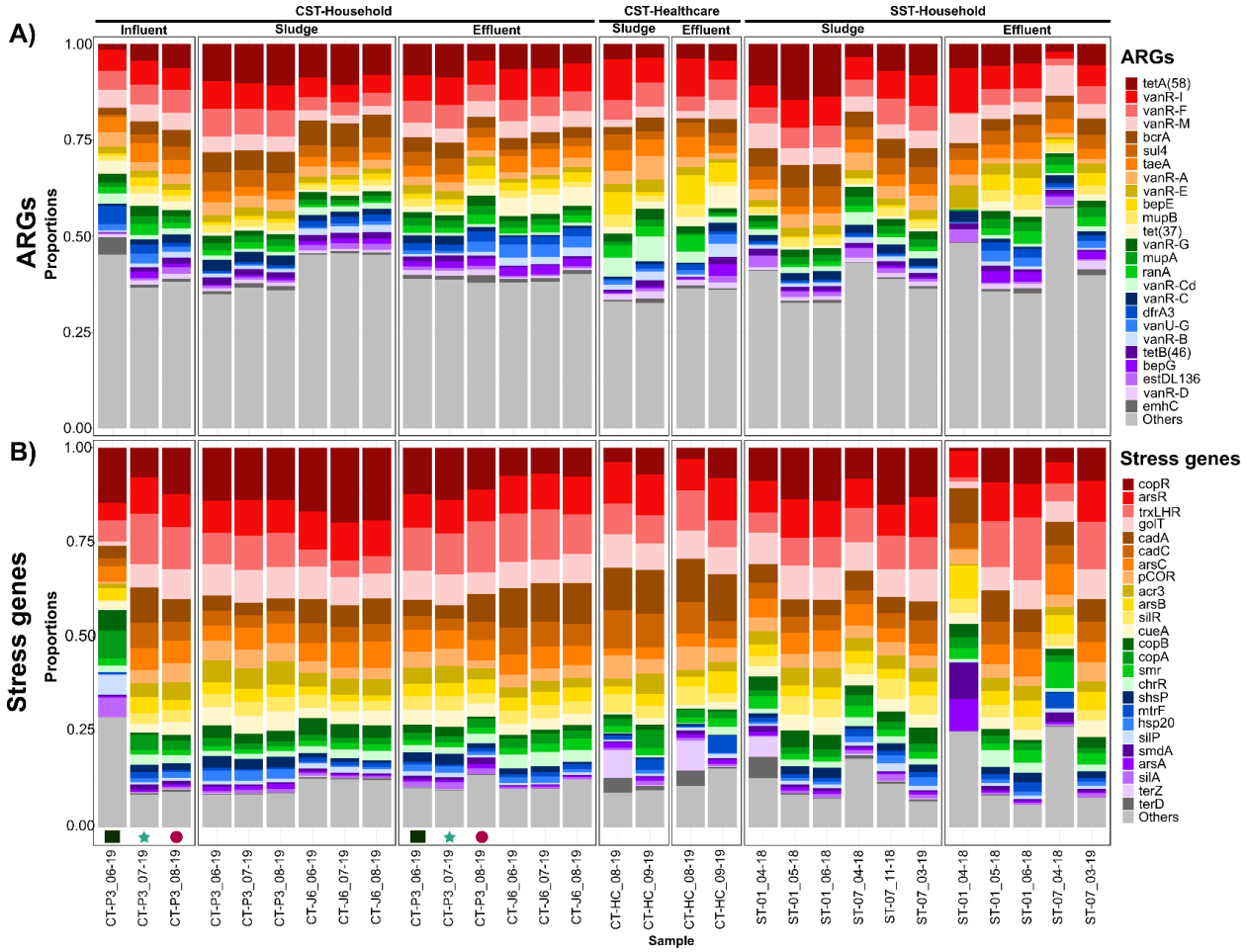

Figure 4.10: Top 25 most abundant (count abundance) ARGs and stress genes detected between the sample types (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household). A) Top 25 most abundant ARGs, B) Top 25 most abundant stress gene. Colour filled icon at the bottom of figure indicates septic tank where influent sample was collected. □ denote influent and effluent for the June sampling month. *- denotes influent and effluent for the July sampling month and, O- denotes influent and effluent for the August sampling month.

Overall, *tetA*(58) (which encodes tetracycline resistance; Figure 4.10A) and *copR* (copper stress gene; Figure 4.10B) were the most abundant resistance gene in the two household tanks (CST-household- *tetA*(58):7.3%±2.7%SD, *copR*:12.7%±4.1%SD; SST-household-*tetA*(58):7.4%±4%SD, *copR*- 9.5%±4.3%SD). While, for the healthcare reactor (CST-healthcare), *VanR-I* (confers vancomycin resistance, 7.9%±2.7%SD) and *cadA* (confers cadmium resistsance-11.6%±0.5%SD) were on overall the most abundant ARG and stress gene respectively (Figure 4.10).

Within each reactor sludge and effluent sample, *VanR-I* (Vancomycin ARG) and *trxLHR* (stress gene that confers heat resistance) were the most abundant in the effluent for both household tanks (CST-household- *VanR-I*:7.1%±0.6%SD, *trxLHR:* 12.0%±1.3%SD; SST-household: *VanR-I*-6.2%±3.5%SD, *trxLHR*:9.3%±6.3%SD) (Figure 4.10). However, in sludge *tetA(58)* (ARG) and *copR* (Stress gene) were the most abundant resistance gene for the two household units (CST-household-*tetA*(58):9.7%±1.1%SD, *copR*-16.2%±2.8%SD; SST-household- *tetA*(58):9.6%±4.2%SD, *copR*: 12.1%±2.9%SD). In contrast, *VanR*-I (ARG) and *cadA* were the most abundant genes in the CST-healthcare sludge and effluent samples (Figure 4.10)

Within the CST-household tank with accessible influent samples, enrichment of ARGs and stress genes within the top 25 most abundant genes were noted in the effluent for the three sampling months (June, July, August) (Figure 4.10). For the June sampling month, 16 of the 25 (64%) most abundant ARGs (*tetA*(58), *vanR-I, vanR-F, bcrA, sul4, vanR-E, bepE, mupB, mupA, ranA, vanR-C, vanR-B, tetB(46), bepG, estDL136, vanR-D*) and 20 of the 25 (80%) most abundant stress genes (*copR, arsR, trxLHR, golT, cadA, cadC, arsC, pcoR, acr3, arsB, silR, cueA, chrR, shsP, mtrF, hsp20, smdA, arsA, terZ, terD*) were enriched, with *tetA*(58) and *goLT* being the most enriched ARG and stress gene respectively.

For the July and August sampling months, the number of enriched genes from the top 25 most abundant ARGs and stress genes decreased. 12 ARGs (*tetA*(58), *vanR-I, vanR-F, bcrA, sul4, vanR-A, vanR-E, vanR-C, tetB(46), bepG, estDL136, emhC)* (Figure 4.8 top) and 13 stress genes (*copR, golT, pcoR, acr3, silR, cueA, copB, smr, shsP, hsp20, silP, arsA, silA*) (Figure 4.8 bottom) were enriched in July with *tetA*(58) and *trxLHR* being the most enriched ARG and stress gene respectively. Meanwhile, 13 ARGs (*vanR-I, vanR-E, bepE, mupB, tet*(37), *vanR-G, mupA, ranA, dfrA3, vanU-G, bepG, vanR-D, emhC*) and ten stress genes (*trxLHR, cadA, arsB, cueA, smr, chrR, mtrF, smdA, arsA, silA*) were enriched in the August, again with *bepE* and *trxLHR* the most enriched of the ARG and stress gene respectively.

Diving beyond the top-most abundant resistance genes, we then investigated the differential abundance of ARGs and stress genes between the three reactors (CST-household, CST-healthcare, SST-household) and then between sample type for each reactor by performing a differential abundance analysis. This was to help ARGs and/ stress genes that can potentially serve as a useful marker for discriminating the different reactors.

4.3.6 Identification of useful biomarkers (i.e., gene) to distinguish sample types or reactor types: *mexX* (ARG) and *klaB* (stress gene) identified important ARG and stress gene markers respectively

From this analysis, 15 ARGs *(aacA10, aph(9)-la, apmA, bepD, blaLRA5, catA1, dfrA22, mexA, mtrR, rmtG, tet(33), tet(O), ttgB , vanS-F, vanS-Pt2*) (Figure 4.11A) and two stress genes (*klaB, srpR*) (Figure 4.12A) whose abundance (abundance in terms of count) were significantly different between the three reactors were identified (p-adjusted value <0.05). The 15 ARGs identified confer aminoglycoside (*aacA10, aph(9)-la, apmA, rmtG*), MDR (*bepD, mexA, mtrR, ttgB*), β-lactamase(*blaLRA5*), phenicol (*catA1*), trimethoprim (*dfrA22*), tetracycline (*tet(33), tet(O)*) and vancomycin resistance (*vanS-F, vanS-Pt2*) while the two stress genes confer tellurium resistance (*klaB*) multi-biocide resistance (*srpR*).

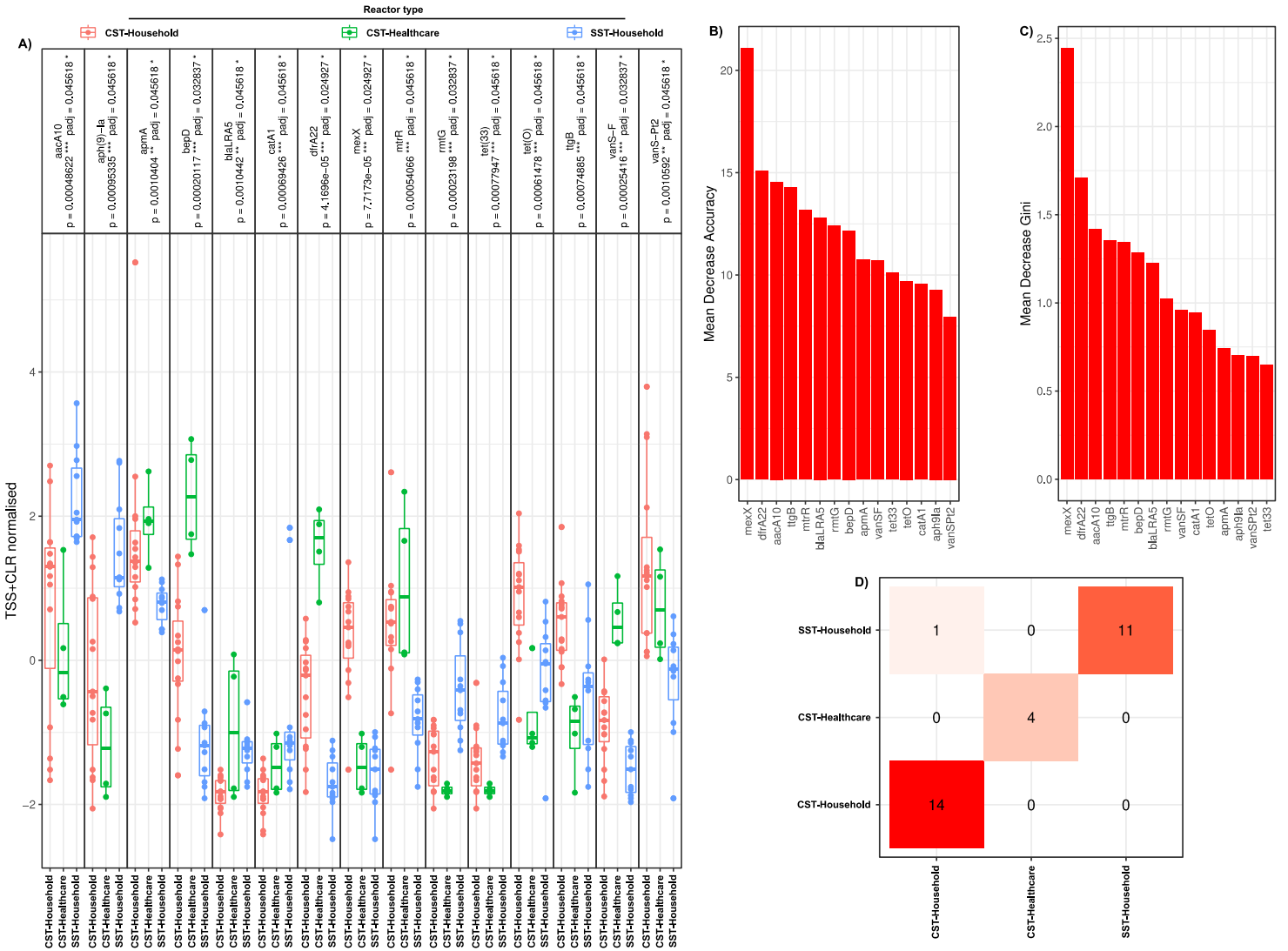

Figure 4.11: Differentially abundant ARGs between the three septic tank reactors (CST-household, CST-healthcare, SST-household). A) Differentially abundant ARGs and Stress gene (Combined) between the three septic tank reactors. B) Mean Decrease Accuracy and C) Mean Decrease Gini importance measures ranking the differentially abundant genes from most importance (highest value) to the least important for differentiating the groups (septic tank types). D) confusion matrix analysis for classification of each sample into their respective reactor type (CST-household, CST-healthcare, SST-household). * p-adj-value < 0.05.

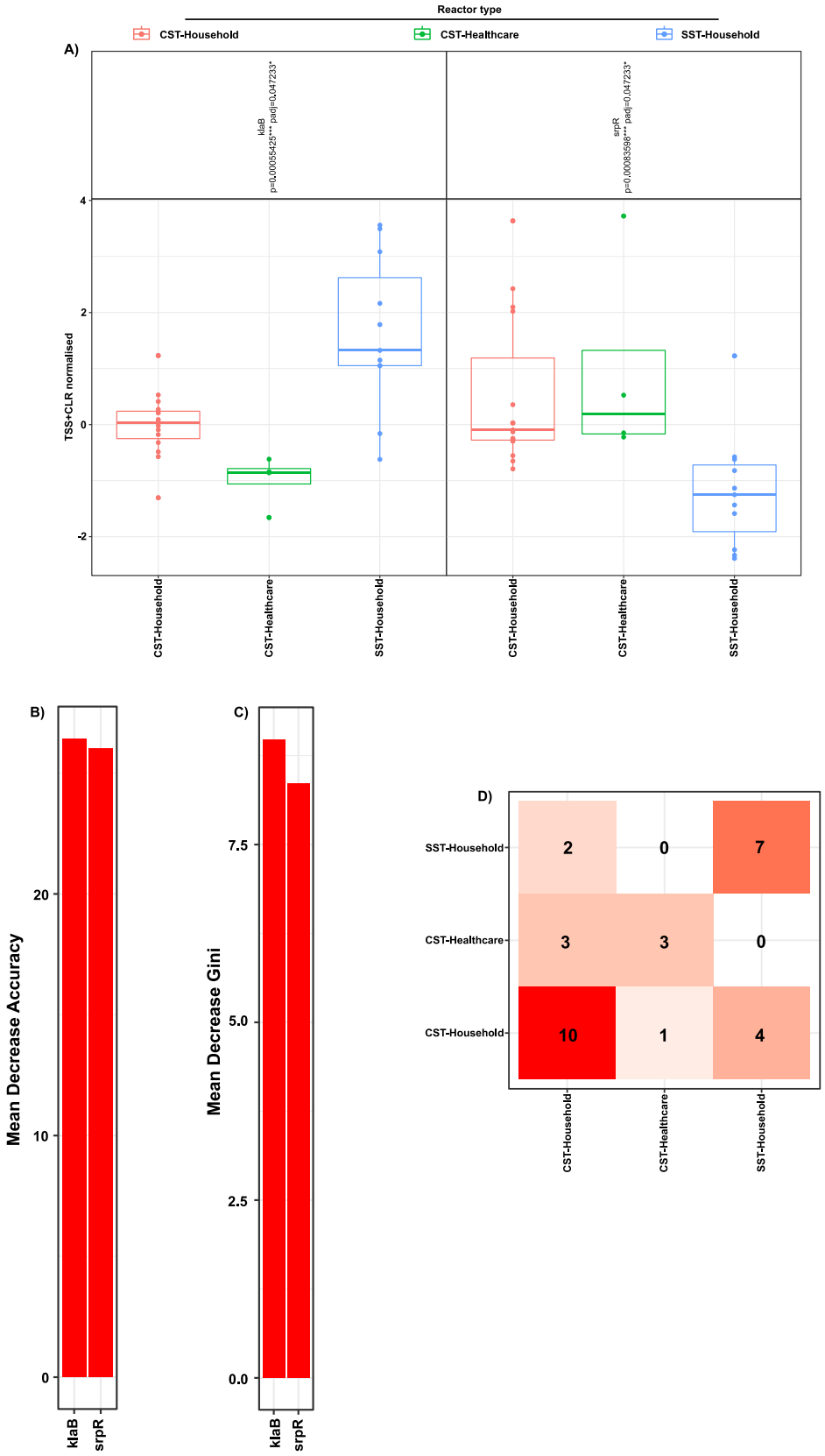

Figure 4.11: Differentially abundant Stress gene between the three septic tank reactors (CST-household, CST-healthcare, SST-household). A) Differentially abundant ARGs and Stress gene (Combined) between the three septic tank reactors. B) Mean Decrease Accuracy and C) Mean Decrease Gini importance measures ranking the differentially abundant genes from most important (highest value) to the least important for differentiating the groups (septic tank types). D) confusion matrix analysis for classification of each sample into their

respective reactor type (CST-household, CST-healthcare, SST-household). * p-adj-value < 0.05.

Random Forest (RF) classifier was subsequently performed to rank the importance of the differentially abundance resistance genes (ARGs and stress genes) identified. In addition, a confusion matrix analysis, which aided in evaluating the accuracy of the RF classifier was also performed. From the RF analysis, both variable importance measures (mean decrease in accuracy (MDA) and mean decrease in Gini (MDG)) identified the *mexX* ARG (confers antibiotic efflux resistance mechanism through the resistance nodulation cell division (RND) transporter; MDA: 20.97, MDG: 2.47) (Figures 4.10B, 4.10C) as the most important gene for predicting the different reactor types (CST-household, CST-healthcare, SST-household). Again, this was owing to *mexX* abundance being statistically different significant between the tanks (p-adj-value <0.05), with the higher abundance reported in the CST-household (10.73±7.76SD count abundance) > CST-healthcare (0.5±0.58SD count abundance) > SST-household (0.36±0.50SD count abundance) tank.

Meanwhile, the *klaB* (encoding resistance protein) was identified by both MDA (value: 28.99) (Figure 4.12B) and MDG (value: 9.24) (Figure 4.12C) importance measures as the most important stress gene for differentiating the different reactors, as its abundance was significantly differentially abundant (p-adj-value <0.05) between the three reactors (Figures 4.11A), with the higher abundance noted in the SST-household (418.91±689.07SD count abundance) > CST-household (28.33±20.307SD count abundance) > CST-healthcare (8.75±4.27SD count abundance) tank.

This means that the removal of either gene (*mexX* or *klaB*) from the model would result in a decrease in accuracy of 20.97% (ARG)/ 28.99% (stress gene) at which the model can correctly predict the different reactor types.

Results of the confusion matrix analysis indicated an almost perfect classification of each sample into their respective tank type for the ARGs (Figure 4.11D). All of the SST-household samples (sludge and effluent (*n=11);* Table 4.1) were correctly identified as SST-household tank samples. Similarly, each of the CST-healthcare samples (sludge and effluent, *n=4*) was correctly identified. For the CST-household samples (influent, sludge, effluent, *n=15*), all except 1 CST-household sample was incorrectly identified as a solar tank sample (SST-household) (Figure 4.11D).

Conversely, for the stress genes (Figure 4.12D), the number of samples incorrectly categorised between the tanks was higher. For the CST-household samples (*n=15*), 10 samples were correctly identified as belonging to this unit, while 3 and 2 samples were

incorrectly identified as originating from the CST-healthcare and SST-household tanks respectively (Figure 4.12D). Similarly, for the CST-healthcare tanks (*n= 4*), all except 1 sample were incorrectly identified as a CST-household sample. Furthermore, 7 of the 11 SST-household samples were correctly identified whilst 4 were incorrectly classified as a CST-household sample (Figure 4.12D).

A combination of both the ARGs and stress genes and subsequent subjection to the above analysis not only reduced the number (*n=11*) of genes identified as differentially abundant between the three tanks but also did not improve the results of the confusion matrix, instead the same output as ARGs alone, where 1 CST-household sample was misidentified as an SST-household sample was generated (Appendix Figure C.2). In addition, the *mexX* gene was again identified by both importance variables (MDA and MDG) as the most influential gene for distinguishing the rector types (CST-household, CST-healthcare, SST-household) (Appendix Figures C.2B, C.2C).

Within each of the three tanks, only the CST-household tank indicated that the ARGs and stress genes (Appendix Figure C.3) were significantly difference in abundance between the sludge and effluent. In fact, 70 ARGs and 16 stress genes were identified to be differentially abundant (p-adjusted-value <0.05) between the sludge and effluent. Of these differentially abundant genes, *erm32* (ARG-conferring resistance to Lincosamide/Streptogramin (LS) antibiotic; MDA: 4.89, MDG: 0.14) and *crcB* (stress gene-confer fluoride biocide resistance; MDA: 9.13, MDG: 0.46) (Appendix Figures C.3B, C.3C) were identified as the most influential gene for distinguishing the sample types (sludge and influent) within the CST-household tank. Additionally, confusion matrix analysis correctly identified all samples (*n=12*) to their respective sample type (sludge and effluent) within the CST-household unit for both ARGs and stress genes (Appendix Figure C.3D).

Finally, for the CST-household tank with an influent sample (CT-P3 tank), no ARGs or stress genes were identified to be statically different in abundance between the influent and effluent (p-adjusted-value >0.05). Again, demonstrates the inability of conventional wastewater to remove ARGs or reduce their abundance from the influent.

4.4 Discussion

Environmental monitoring of broad-spectrum AMR genes (ARGs and stress genes) using targeted approaches, such as the HT-QPCR and/ or proxy genes such as *intI*1, remains challenging due to the reliance on Q-PCR primers coupled with the inherent bias of the

primers. As such, this study employed shotgun metagenomic sequencing, a non-targeted approach, to comprehensively character AMR genes (ARGs and stress genes) from the Thai conventional (CST-household and CST-healthcare) and solar septic tanks (SST-household) associated with household and healthcare usage. To the best of our knowledge, this is the first that has applied shotgun metagenomic to characterise AMR genes from septic tank wastewater in Thailand, a country with poor regulations on antibiotic use and an ineffective WWT system that acts as a hotspot for AMR acquisition and dissemination to the environment. Here we show that septic tanks are major sources of AMR and stress genes.

4.4.1 ARGs and Stress genes stringency mapping parameter greatly impact observed richness and diversity

The results from this study found that using a lower stringent mapping parameter (25 amino acids coverage with 40% identity) resulted in higher richness and evenness of detected AMR genes from the three Thai septic tanks compared to medium and more stringent mapping parameters. However, using the lower stringent threshold increases the risk of characterising false-positive or non-functional AMR genes (Lal Gupta et al. 2020). This, therefore, highlights the trade-off that must be considered when characterising AMR genes (ARGs and stress genes) from metagenome-assembled genome (MAG) contigs against known and curated AMR databases such as CARD (Alcock et al. 2023). Despite the associated risks, this study used the AMR data detected at the lower stringent threshold as it provided a more realistic representation of AMR genes within the tanks, as informed by our HT-QPCR study of pooled and individually targeted samples (see Chapter 3).

4.4.2 Risk assessment of AMR gene dissemination from the sludge and effluent between the three Thai septic tanks

The widespread knowledge that WWT, including decentralised WWT such as septic tanks, contributes significantly to the subtype and abundance of AMR genes (ARGs and stress genes) in the environment has been supported by countless studies (Gupta et al. 2018; Xu et al. 2020; Raza et al. 2021; Rodríguez et al. 2021). In this study, we found that the septic tanks (CST-household, CST-healthcare, SST-household) were not only a rich source of diverse ARGs and stress genes that confer resistance to 29 broad-spectrum antibiotic classes and 18 stress gene classes (such as heavy metals and biocides) respectively, to the

environment but also play a major role in the abundance of both ARGs and stress genes entering the environment.

In shotgun metagenomics studies, due to a lack of standardised methodology for characterising detected genes, particularly with regard to a consensus mapping stringency parameter, different studies employ different stringency parameters. Consequently, differing richness, evenness, diversity and gene abundance (i.e., relative abundance or count abundance) are reported across studies (Ekwanzala et al. 2020), which limits in-depth understanding. Nonetheless, this study found similarities in the richness and diversity of ARGs and stress genes (such as heavy metals) reported in published studies (Gupta et al. 2018; Xu et al. 2020; Rodríguez et al. 2021; Zhang et al. 2021), with genes conferring resistance to different antibiotic class found in both this and others studies. Our findings support the already established knowledge that WWT is an important source of rich and diverse AMR genes to the environment.

ARGs conferring resistance to bacitracin, mupirocin and pleuromutilin were found to be the most abundant (count abundance) in the sludge and effluent for the three tanks (Table 4.3). This finding is consistent with that of Karaolia et al., (2021) who also reported ARGs conferring resistance to bacitracin, mupirocin and pleuromutilin among the top ten most abundant genes in two urban WWTs, albeit in the influents. This implies that these genes may be highly abundant within WWT samples.

Interestingly, the top 25 most abundant ARGs are dominated by genes conferring resistance to seven other classes of antibiotics, including tetracycline, vancomycin, sulphonamide, MDR, aminoglycoside, trimethoprim and phenicol, rather than just the three most abundant classes (bacitracin, mupirocin and pleuromutilin).

Beta-lactam, particularly Amoxicillin, Clavulanic acid, Ampicillin and Penicillinase-resistance penicillin, quinolone and tetracycline are among the most prescribed and consumed antibiotics in Thailand (Siltrakool et al., 2021). Although genes conferring both beta-lactam and quinolone resistance were not among the top 25 most abundant ARGs, they were characterised across the septic tanks. With tetracycline being among the most prescribed antibiotics in Thailand (Siltrakool et al., 2021), it is unsurprising to find genes conferring tetracycline resistance (*tetA*(58), *tet*(37), *tetB*(46)) among the top 25 most abundant ARGs. In Thailand, tetracycline is frequently used as an anti-inflammatory drug for the treatment of a range of inflammatory illnesses including endometritis (Siltrakool et al., 2021).

Detection of resistance genes does not provide insights into the microbial strain(s) harbouring these genes nor link it to the presence of pathogens, which is a major challenge in wider AMR studies and this study. Therefore, in future septic tank studies, one approach that could be employed to link resistance genes to specific microbial strain(s) could involve isolating individual strains from the septic tank samples and screening for their resistance phenotypes. Suspected genes from positive isolates can be targeted using PCR and subsequent sequencing can be performed to confirm the identity of each isolate and harboured resistance gene(s). An alternative approach to link resistance genes to specific microbial strain(s) could be the use of advanced single-cell molecular techniques such as the EPIC-PCR (Emulsion, Paired Isolation and Concatenation PCR) which permits linkage of functional genes (i.e., ARGs) to phylogenetic markers such as *16S rRNA* gene (Spencer et al. 2016).

Amongst the three tanks, the CST-healthcare unit was found to contribute to a lower richness of ARGs (Figures 4.1, 4.2) and stress genes (Figures 4.1, 4.3) through its sludge and effluent. Despite contributing to a lower ARG richness entering the environment, the CST-healthcare unit, of the three tanks, was the higher contributor of ARG abundance to the environment through its sludge and effluent.

The higher AMR abundance (count abundance) in the CST-healthcare tank sample (sludge and effluent) compared to the household tanks (CST-household and SST-household) sludge and effluent samples is not surprising. This is because hospital WWT usually receives broad-spectrum antibiotics and in higher concentrations than municipal WWT (Petrovich et al. 2020). In decentralised healthcare WWT, this concentration can be even higher and exposure of the microbial to broad-spectrum antibiotics and at higher concentration can impose stronger selective pressure on the microbes, increasing acquisition of AMR genes from the surrounding pool and thus, resulting to an increased abundance of ARGs harbouring microbes entering the environment (Hocquet et al. 2016; Rozman et al. 2020).

Of the three tanks, the SST-household unit did not have the highest ARG subtypes in the sludge and effluent samples or stress gene subtypes in the effluent. However, stress gene abundance (count abundance) was highest in the SST-household samples (sludge and effluent), while ARG abundance was second highest in the samples (sludge and effluent). These observations suggest that, while the SST-household unit may not be the highest contributor of ARG subtypes via its sludge and effluent to the environment or stress gene subtypes via its effluent, it does potentially contribute to the highest overall stress gene

abundance (Table 4.4) and second highest overall ARG abundance entering the environment (Table 4.3).

Conversely, the CST-household reactor had the highest ARG subtypes in the sludge and effluent, among the three reactors, but the lowest overall ARG abundance (count abundance) in the sludge and effluent. This implies that the CST-household reactor may contribute to the most ARG subtypes entering the environment but is a lower contributor to overall ARG abundance, among the three tanks (Table 4.3).

The solar septic tank (SST-household), compared to the conventional septic tanks (CST-household and CST-healthcare), was found to have a lower abundance of ARGs conferring Lincosamide/Streptogramin (LS) and colistin resistance in the sludge. In addition, lower abundance of eight ARGs class (antibacterial free fatty acids, bacitracin, MDR, nitroimidazole, tetracenomycin, thiostrepton, colistin, vancomycin) three stress gene class (heat, cadmium, biocide and metal resistance (BMR)) class was found to be in the SST-household effluent than the CST-household and CST-healthcare effluent (Tables 4.3, 4.4). This not only suggests that the effect of temperature may be resistance-specific (Nguyen et al. 2021) but also sample-type-specific.

The overall results suggest that the higher internal temperature in the solar tank unit was somewhat effective at reducing certain types of ARGs and stress gene classes. However, the tank's internal temperature may not be high enough to substantially reduce ARG abundance from the sludge and effluent.

The effect of temperature on the fate of ARGs remains unclear. For example, Xu et al., (2020), utilising shotgun metagenomic approach, reported a 30% removal of ARGs subtypes (from the 1360 subtypes characterised) from sludge treated at thermophilic temperature (55°C) than at mesophilic temperature (35°C). Additionally, the abundance of ARGs in 20 out of 41 antibiotic classes categorised was observed to be lower in sludge treated at thermophilic temperature. Furthermore, Xu and co-workers found significantly lower ARG richness and evenness (Shannon) as well as a lower relative abundance of most bacteria phyla (such as Proteobacteria and Chloroflexi among others) at thermophilic temperature. They then suggested that the enhanced reduction of ARG subtypes and abundance at higher temperatures (Thermophilic temperature) may be attributed to the increased rate of biomass degradation which includes microbial biomass. This, in turn, impedes horizontal gene transfer pathways, leading to a decrease in ARG acquisition by bacteria, and ultimately a decrease in ARGs abundance (Xu et al. 2020).

On the other hand, Zhang et al., (2015) who also employed a metagenomic approach, reported little difference in the diversity and total ARGs relative abundance (relative ARGs abundance as part per million) in sludge treated at mesophilic (35°C) and thermophilic temperature (55°C) temperature. Furthermore, the diversity and total ARGs abundance in the thermophilic and mesophilic treated sludge was similar to the untreated feed sludge.

ARGs in the effluent and sludge between the three tanks predominantly conferred antibiotic inactivation resistance mechanism. Similarly, the stress genes entering the environment via sludge and effluent for the three tanks conferred resistance proteins (i.e., small heat shock protein and Tellurium resistance membrane protein) as the dominant resistance mechanism.

Examination of ARGs and stress genes was observed in this study when the influent and effluent samples were compared for the tank with accessible influent sample revealed that total ARGs abundance (count abundance) for the three sampling months (June, July, August) was higher in the effluent than influent (Table 4.5). Similarly, the total stress gene abundance (count abundance) was also found to be higher in the effluent than influent for two (July and August) of the three sampling months (Table 4.5). Altogether, the results highlight the septic tanks as potentially a significant source of AMR gene (ARGs and stress genes) abundance to the environment. Zhang et al., (2021) also reported higher richness and evenness of ARGs and stress genes (notably biocide resistance genes) in the effluent than influent in a WWTP exclusively treating hospital wastewater, although the overall ARGs and stress genes abundance was found to be lower in the effluent.

WWT not only serve as a significant contributor to the richness, diversity and abundance of ARGs and stress subtypes to the environment but also plays a crucial role in the spread of genes that confer resistance to antibiotic classes that have been identified by WHO (WHO 2021a) as critically important and/or the highest priority for human health, including antibiotics of last resort.

Among the ARGs characterised from this study, ARGs conferring resistance to Rifamycin, a critically important antibiotics class (in particular Rifampicin) used in first-line treatment of Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (Brucoli and McHugh 2021; Arbiv et al. 2022), was detected in all of the sample types (influent, sludge, effluent) across the three tanks. The SST-household effluent, of the three tanks, was found to be the higher contributor of its abundance to the environment, while the same is true for the CST-household sludge (Table 4.4). In fact, the abundance of rifamycin resistance genes was observed to be enriched in the effluent of the tank (CST-household) with accessible influent samples for two (out of three) of the sampling months (June and July) (Table 4.5). Rodríguez

et al.,(2021) also reported an increase in the relative abundance of rifamycin in WWT effluent from the influent following characterisation of ARGs from wastewater compartments (i.e., influent, aeration tank, return activated sludge, effluent) using shotgun metagenomics.

Additionally, gene (*vph*) conferring resistance to antibiotic class (tuberactinomycin) used in second-line treatment of TB (Zane and Graeme 2022), was also detected but only in the conventional healthcare (CST-healthcare) tank (Figure 4.3), with its abundance higher in the effluent (2.5±2.1) than sludge (1.5±0.7) (Table 4.3). TB is a significant global health public concern (Chesdachai et al. 2016) and its prevalence is exacerbated by the recent rise in antibiotic resistance, resulting in an increase in multi-drug resistant TB and pan-drug resistant TB. Thailand, in particular, has a high TB burden (WHO 2022), which is primarily attributed to the increased epidemic of human immunodeficiency virus (HIV) infection (Boonsarngsuk et al. 2009).

Similarly, ARGs conferring resistance to highest priority antibiotics reserved as last resort (WHO 2021a), were detected in between the sample types between the reactors. Colistin resistance genes, in particular, were found in higher abundance in the CST-healthcare sludge and effluent samples, out of the three reactors. Moreover, the abundance of colistin was observed to be enriched in the effluent of the tank (CST-household) with accessible influent sample for two (out of three) of the sampling months (June and August) (Table 4.5). Since the first plasmid-borne mobile colistin resistance (*mcr*) gene discovery in China in 2015, *mcr* has spread globally across diverse microbial taxa, with 10 *mcr* variants (*mcr*-1 to *mcr*-10) described so far in human and bacterial isolates (Luk-in et al., 2021). These different *mcr* variants have been characterised in WWT (Rodríguez et al. 2021; Markkanen et al. 2023). Markkanen et al., (2023), for example, characterised all except one *mcr* variant (*mcr*-2) from hospital and non-hospital wastewater sources (i.e., tap water, river water) from three different countries (Benin, Burkina Faso, and Finland) with *mcr-5* reported as the most frequently observed variant in the majority of the hospital wastewater investigated among the three countries. In this study, nine (*mcr*-1 *to mcr*-9) of the ten variants were characterised from the septic tanks providing evidence that septic tanks can be an important source of *mcr* variants entering the environment.

The dissemination of *mcr* genes into clinically relevant bacterial pathogens, particularly those classified as “priority pathogens” by WHO (WHO 2017), is of great concern. These pathogens exhibit high resistance to broad-spectrum antibiotics, in particular, carbapenems and extended-spectrum cephalosporins, which are commonly used to treat multi-drug-

resistant bacteria (Luk-in et al. 2021). As a result, they have the potential to cause severe to life-threatening infections in patients (Grenni, 2022). The acquisition of *mcr* genes by these pathogens virtually renders them untreatable with current antibiotics (Walsh, 2018), leading to poor patient outcomes. In Thailand, mcr variants have been reported within some of the priority pathogens including Enterobacteriaceae (*Escherichia coli, Klebsiella pneumoniae, Salmonella spp.*) (Eiamphungporn et al. 2018; Luk-in et al. 2021).

4.4.3 Identification of useful biomarker (i.e., gene) to distinguish sample types or reactor types

Based on the differentially abundant ARGs identified, the random forest classifier model correctly classified all but one sample into their respective tank type and sample type. However, the accuracy of the model did drop when the two differentially abundant stress genes identified were used.

From the identified differentially abundant ARGs and stress genes, the classifier found *mexX* (confer MDR) and *klaB* (confer tellurium resistance) respectively, to be the most important for predicting the different tank type (CST-household, CST-healthcare, SST-household), which minimises the requirement (i.e., cost) to effectively distinguish the three tanks. Although further work is needed to support this.

With regards to distinguishing sample types within each reactor, only the CST-household tank showed that ARGs and stress gene abundance within the sample types (influent, sludge, effluent) were statistically different (p-value <0.05). of which, *erm32* (ARG conferring lincosamide/streptogramin (LS) resistance) and *crcB* (stress gene conferring fluoride resistance) were identified and the most useful markers for distinguishing the sample types. Again, further studies are needed to validate this.

4.5 Conclusion

This study has utilised shotgun metagenomic to gain a comprehensive understanding of decentralised WWT, specifically septic tanks, contributions to ARGs and stress genes burden to the environment. Thus, emphasising the urgency to improve current decentralised WWT design to effectively reduce ARGs and stress genes subtypes and abundance from treated wastewater prior to its discharge to the surrounding environment.

Further studies are needed to characterise mobile genetic elements, in particular mobile integrons associated with dissemination of resistance, from these septic tanks. This would help understand the potential for horizontal gene transfer within the tanks and their role as hotspot for ARGs or stress gene acquisition.

Chapter 5

Final discussion and future works

In light of ongoing challenges in AMR monitoring and the knowledge gap of the contributions of decentralised WWT, specifically septic tanks, in disseminating AMR to the environment, this thesis set out to:

1) Evaluate and validate the suitability of the *intI*1 gene as a proxy for inferring potential AMR pollution using the conventional QPCR and HT-QPCR on decentralised wastewater treatment plants from Thailand.

2) Evaluate the contributions of conventional septic tanks associated with household and healthcare usage, and the newly developed SST tank associated with household usage in the dissemination of AMR genes and MGE using both HT-QPCR and Shotgun metagenomics.

Main Findings

1) Suitability of *intI*1 as a proxy for AMR

The first key finding from this thesis was that the *intI*1 gene could not serve as a suitable or reliable proxy for inferring overall AMR gene -or at least a suitable proxy for mobile integrons associated AMR gene- pollution. This is owing to several key findings from this thesis. Firstly, current *intI*1 primers, as shown in Chapter 2, do not only have specificity to *intI*1 sequences (highly conserved *intI*1 variant; >98% protein identity to pVS1) but also to *intI*1-like sequence (the lesser conserved *intI*1 variants; >98% protein identity to pVS1). This means that *intI*1-like variants potentially contribute to the Q-PCR signal; therefore, leading to an overestimation of quantified *intI*1 gene abundance, which in turn influences its correlation to the abundance of AMR genes. In addition, ASVs obtained from the three selected *intI*1 primer sets were highly similar to both *intI*1 and *intI*1-like sequences from known and unknown bacteria, isolated from a range of clinical (e.g., human stool) and environmental (e.g., wastewater biofilm) settings. Secondly, *intI*1 gene relative abundance correlated positively and significantly to the abundance of a reduced number of AMR genes (both associated and non-associated with mobile resistance genes) quantified on the HT-

QPCR compared to the *intI*3 and *sul*1 gene (Figure 3.13). Therefore, suggesting that the abundance of the *intI*1 gene is not suitable as a proxy for overall AMR abundance.

In a previous study, Zheng et al.,(2020) reported a significant positive correlation between *intI*1 genes (termed clinical *intI*1 and *intI*1 in their study) absolute abundance and overall AMR absolute abundance in both activated sludge and permeate sample from WWT following utilisation of the HT-QPCR array. In the activated sludge the clinical *intI*1 absolute abundance was reported to have a higher correlation (R=0.794, p-value <0.001) to overall ARG absolute abundance than *intI*1 (R=0.787, p-value <0.001). Conversely, in the permeate sample, *intI*1 absolute abundance had a higher correlation (R=0.906, p-value <0.001) to overall ARG absolute abundance than clinical *intI*1 (R=0.641, p-value =0.013). Nonetheless, they recommended the use of the clinical *intI*1 abundance as marker of ARGs abundance. Whilst their findings (Zheng et al. 2020) contrast those of ours, a direct comparison between the two studies is a challenge, owing to the different methods that have been used in both studies to establish a correlation between *intI*1 gene abundance and AMR gene abundance.

In their study, Zheng et al., (Zheng et al. 2020) used absolute abundance (copies per g sludge or copies/L influent), calculated by multiplying the relative gene copy number (determined with an arbitrary Ct cutoff of 31) by the absolute *16S rRNA* copies quantified using a different *16S rRNA* primer set from the array, and on a different platform, to establish a correlation between *intI*1 and AMR genes. In this thesis, however, relative gene abundance (normalised to the *16S rRNA* Ct) following the ≥3.3 Ct difference approach between sample and NTC Ct, as described by Smith and Osborn, (2009) during the data processing step, without applying an arbitrary Ct cut off, was used to establish the correlation.

Factors such as inter-assay variations (Smith et al., 2006) from the use of different quantification platforms and the use of the different *16S rRNA* primers to obtain the absolute gene abundance could have potentially biased and influenced the results of their correlation analysis. In addition, the so-called "clinical *intI*1 primer set", which corresponded to F10-R10 in our *in-silico* study (Appendix Table A.1), did not exclusively target *intI*1 (the highly conserved *intI*1 variant) as originally suggested (Gillings et al. 2015) but also targets *intI*1-like variant (the lesser conserved *intI*1 variant) as shown from our detailed *in-silico* testing (Appendix Table A. 2). This means that *intI*1-like variant potentially contributed to the Q-PCR signal, leading to an overestimated of quantified *intI*1 abundance. Again, likely biasing

and influencing the significant positive correlation reported in their study (Zheng et al. 2020).

Furthermore, in our array study, we observed that this primer set repeatedly quantified similar Ct in the NTC and wastewater samples for both the pooled samples (used for pre-screening of AMR genes and MGEs) and the individually targeted wastewater samples and was removed from all analysis (NTC Ct: individually targeted sample-*intI*1_1: 16.47±0.18SD; pooled sample-13.77±0.11SD). This was only discovered by adopting the ≥3.3 Ct difference approach between sample and NTC mean Ct as described by Smith and Osborn, (2009) as part of the data processing step. It is plausible that this assay had a similar low absolute Ct in their sample and NTC as is the case in our study, which potentially suggests that the *intI*1 quantification isn't real- or at least reliable using the clinical intI1 primer set (F10-R10 primer set). However, with the arbitrary Ct cut-off of 31 applied in their study (Zheng et al. 2020), likely higher than the absolute Ct in the samples and NTC, this particular assay may have been retained, causing significant bias in the reported correlation analysis. Taken together, these findings raise questions as to the validity of the *intI*1 as a proxy for overall ARG abundance as suggested in their study (Zheng et al., 2020).

This said it is also plausible that the inadequacy of the *intI*1 gene to serve as a suitable proxy for inferring overall AMR abundance in our study could have been influenced by the choice of *intI*1 primer sets on the array. Array *intI*1 primers were found to not only quantify lower *intI*1 abundance, compared to the three selected *intI*1 primers (Chapter 3 Figure 3.14) but also impact the interpretation of results, such as significant differences between quantified samples.

With this knowledge, it becomes challenging to draw definitive conclusions regarding the suitability of the *intI*1 as a proxy for inferring overall AMR abundance within the scope of this thesis. Therefore, to address this issue and establish a more definitive conclusion, a way forward would be to characterise the *intI*1 gene from the shotgun metagenomic dataset and correlate its abundance to that of the overall AMR gene abundance This would aid in discerning whether or not the low number of AMR genes, whose abundance correlated statistically to the *intI*1 abundance was indeed due to primer selection on the array. In addition, this would help confirm that the overall significant positive correlation observed with transformed absolute abundance is indeed real. In doing so, provide experimental evidence that strongly supports the use of the *intI*1 gene as an initial pre-screening tool for the identification and monitoring of environments with the highest risk of AMR

dissemination to both humans and animals. Time permitting, this would have been the next step of this thesis.

Interestingly, many studies (Gupta et al., 2018; Yoo et al., 2020; Rodríguez et al., 2021; Markkanen et al., 2023) have characterised ARGs and integron-integrase genes including the *intI*1 gene from WWT or other polluted settings using shotgun metagenomic approach. However, not all of these studies have conducted a statistical correlation between *intI*1 abundance and ARGs abundance. Nonetheless, Markkanen et al.,(2023) reported a significant positive correlation between the relative *intI*1 gene and overall ARGs abundance characterised from hospital WWT from three countries (Benin ($R=0.81$, p-value <0.001), Burkina Faso ($R=0.54$, p-value <0.001) Finland ($R=0.83$, p-value <0.05)).

The metagenomic dataset from this thesis's body of work should also be interrogated in future work for alternative proxies such as the *intI*3 and *sul*1 genes. Interestingly, both *intI*3 and *sul*1 abundances quantified on the array correlated positively and significantly to the abundance of a higher number of AMR genes, both associated and non-associated with MRI, than *intI*1 in our array study (Chapter 3). In fact, the higher number of AMR genes (both mobile integron-associated and non-associated AMR genes) correlating positively and significantly to the *intI*3 abundance suggests that *intI*3 could serve as a proxy for overall AMR abundance even though it is considered rarer in the environment compared to *intI*1 and *intI*2 (Cambray et al. 2010). While the *sul*1 gene relative abundance could serve as proxy for MRI-associated AMR genes than *intI*1 and potentially proxy for overall AMR abundance.

Other suggested potential proxies, such as *VanA* (confer vancomycin resistance), *sul*3 (confer sulphonamide resistance) q*acEΔ*1(confer antiseptic resistance) (Abramova et al. 2022) could permit effective AMR monitoring should also be investigated in the metagenomic dataset in future works. Interestingly, Markkanen et al.,(2023) reported a significant positive correlation between the relative *qacEΔ*1abundance and overall ARG abundance characterised with metagenomic from hospital wastewater in three different countries (Benin ($R=0.76$, p-value <0.001), Burkina Faso ($R=0.64$, p-value <0.001), and Finland ($R=0.83$, p-value <0.05)).

2. **Septic tanks contribute to AMR burden in Thailand**

Another key find from this thesis was that septic tank samples represent a significant source of AMR genes and MGEs (in terms of richness, diversity and abundance) to the environment

via the effluent. This includes clinically relevant AMR genes conferring resistance to antibiotics reserved as a last resort for human use when all other antibiotics fail. Additionally, the direct release of septic tank sludge into the environment, as commonly practised in the global south regions such as Thailand, where 80-90% of faecal sludge is directly discharged this way (Koottatep et al. 2021), presents an additional source of AMR genes and MGEs. The role of septic as a significant source of AMR genes and MGEs was evident from the high richness, diversity and abundance of AMR genes, including stress genes, and MGEs quantified and/or characterised from the sludge and effluent samples from the three septic tank types (CST-household, CST-healthcare, SST-household).

In fact, examination of the influent and effluent samples for the septic tank (CST-household) with accessible influent through targeted (QPCR and HT-QPCR) and non-targeted (shotgun metagenomics) molecular approaches, consistently showed higher gene abundance (or count abundance in shotgun metagenomics) or ARGs or stress genes subtypes in the effluent than influent for the three sampling months. These findings suggest that the septic tanks may be inadequate in reducing the abundance, richness and diversity of AMR (ARGs and stress genes) and integrase gene from the effluent, highlighting their role potential role as a significant source of AMR and integrase gene to the environment. These findings echo the findings from the larger body of published work on AMR, MGEs and WWT, which recognises WWT including septic tanks, as a major source of AMR and MGEs to the wider environment (Rizzo et al. 2013; Connelly et al. 2019; Hayward et al. 2019). Yet, emphasises the necessity and urgency to improve septic tanks in AMR removal, particularly in light of the significant proportion of the global population (2.7 billion people (Harada et al. 2016)) estimated to be served by onsite decentralised WWT including septic tanks.

Whilst both the targeted (HT-QPCR and QPCR) and non-targeted (shotgun metagenomics) approaches have shown septic tanks to be a hotspot and major source of AMR and MGEs, the reactor type (or sample type) posing the highest risk of AMR genes and MGEs to the environment, differs with respect to the molecular tool used. The HT-QPCR approach showed that AMR gene richness and diversity were comparable between the reactors (CST-household, CST-healthcare, SST-household) and sample types (influent, sludge, effluent) (Figure 3.7) based on the selected target genes. However, this approach reported the highest relative abundance of AMR and integrase (*intI*1 and *intI*3) gene in the CST-household samples (sludge and effluent), among the three tanks, while the lowest AMR and integrase (*intI*1 and *intI*3) relative gene abundance was observed in the healthcare (CST-healthcare) reactor. This indicates that the CST-household reactor is the highest contributor of AMR and

integron (CL1-and-CL3) genes abundance to the environment, while the CST-healthcare reactor contributes the least.

In contrast, shotgun metagenomics, showed that the richness (both rarefied and absolute count) of AMR genes (ARGs and stress genes) was highest in the CST-household unit and lowest in the CST-healthcare reactor but similar diversity of AMR classes across the three tanks. However, the ARGs abundance (count abundance) was highest in the CST-healthcare samples (sludge and effluent), among the three reactors, while the lowest ARGs abundance was found in the CST-household reactor samples (sludge and effluent) indicating the CST-healthcare as the most contributor of ARGs abundance to the environment and the CST-household as the least contributor.

Both HT-QPCR and shotgun metagenomics are powerful approaches currently used in AMR gene profiling and monitoring (Liu et al. 2019). However, the limited available primers on the array potentially influence the interpretation of risk associated with the septic tanks. Therefore, emphasising the trade-off that must be considered when selecting either tool for effective AMR monitoring and/or identification of environments posing the greatest risk of AMR dissemination.

Compared to shotgun metagenomics, the HT-QPCR approach has the advantage of higher detection limits of genes, faster turnaround time, in terms of quantification and data analysis, no steep learning curve, expertise in bioinformatics, or complex bioinformatic pipeline and lower per assay run cost (Liu et al. 2019; Waseem et al. 2019). Therefore, highly suitable for effective routine AMR monitoring (Liu et al. 2019). Yet, biologically meaningful interpretation of obtained results from the array can only be achieved when quantified AMR gene abundance is reported as the absolute mean Ct or normalised abundance through normalisation of quantified AMR and integrase genes to a single copy gene such as *ropB*, preferably on the same instrument. Alternatively, quantified AMR and integrase gene can still be normalised to the *16S rRNA* gene but only to 16S rRNA abundance quantified with well-validated *16S rRNA* primer set such as Bact1369F-Prok1492R (Suzuki et al. 2000) and not the *16S rRNA* primers (AY1) currently used on array.

Shotgun metagenomic, on the other hand, eliminates the need for primer selection (Pruden et al. 2021) and can be used to complement the HT-QPCR approach to comprehensively profile AMR genes, MGEs and microbial taxa within a given environmental samples for robust monitoring. This not only helps gain an in-depth understanding of the fate and

dynamics of AMR genes and MGEs but also identifies potential for HGT of AMR and identification of microbial taxa that potentially harbour these AMR genes and MGEs.

Finally, both HT-QPCR and shotgun metagenomics indicated that the increased internal temperature within the solar septic tank served a somewhat limited role in sufficiently reducing AMR genes and MGEs from the sludge and effluent samples when compared to the conventional tanks (CST-household and CST-healthcare) samples (sludge and effluent).

Although, it is plausible that the varying internal temperature for the solar tank, which often was below the target core temperature (50 to 60°C) could have contributed to poor treatment performance observed. In addition, the inaccessibility of influent samples for the solar tanks further impeded the ability to reliability estimate AMR and MGEs removal efficiency in the effluent to adequately analyse the benefits of increased internal temperature on the fate and dynamic of AMR genes within the system.

Challenges and future directions

The results of this thesis have provided comprehensive evidence that septic tanks are not ineffective at AMR removal or reduction of their abundance from treated wastewater by-products (sludge and effluent). Thus, they contribute to a significant amount of AMR genes and MGEs entering the environment. Yet, further understanding is needed, in particular, on the role of temperature on the fate and dynamic of AMR genes and MGEs in septic tanks over an extended period. This knowledge is crucial if we are ever going improve septic tanks to adequately treat wastewater and significantly remove or reduce AMR genes, ARBs and MGEs from treated waste. Thus, contributing to tackling global AMR and averting the looming catastrophic predictions (10 million global deaths per year by 2050, (O'Neill, 2014)) from AMR-related infections in the next three decades.

With this in mind, future experimental work should include the influent for the solar septic tank during the sampling campaign and the sampling time points should be extended to obtain a clearer understanding of increased temperature impact over a longer period of time. Furthermore, if possible, the sampling should be done on re-designed and optimised solar septic tanks that achieve and maintain their intended target temperature (50-60°C).

Finally, a caveat of this work is that detected or quantified AMR and integrase genes were on DNA level, meaning that we do not know if the genes are intact, within organisms or available for uptake. Furthermore, even if the genes are harboured within organisms, we do

not know if the organisms are viable and/or metabolically active (i.e., actively expressing AMR or integrase genes, especially in the presence of selection pressures such as antibiotics) (Pruden et al. 2021). As harbouring a gene doesn't mean the genes are expressed even in the presence of selection pressure like antibiotics but at best infers potential for the genes to be expressed. More importantly, establish a link between AMR genes and the CL1-integron using viable and metabolically active cells. Therefore, further experiments could look to explore the use of Raman activated cell sorting approach to sort viable and metabolically active (i.e., cells grown in the presence of antibiotics and deuterium label), eject and culture each single cell and subsequently perform a colony PCR/Q-PCR to target the *16S rRNA* gene (taxonomic identification) and the *intI*1 gene (functional identification) on isolated, sorted single cells. Thus, providing experimental evidence that further strengthens *intI*1 gene use as a proxy for overall AMR gene abundance if it was indeed found to be a suitable proxy through the metagenomic dataset.

In conclusion, the work presented here has highlighted how primer choice potentially affects the quantification of the *intI*1 gene and thus, a conclusive result as to the suitability of the *intI*1 gene to serve as an adequate proxy for inferring AMR pollution could not be reached within the scope of this work. Future studies are in no doubt needed to address this. Additionally, septic tanks, a decentralised WWT, were found to potentially contribute to a significant AMR (ARGs and stress gene) and MGE gene subtypes and abundance to the environment. Data on the contribution of septic tanks to AMR and MGEs dissemination remains scarce compared to centralised WWT. Therefore, more studies focusing on AMR and MGEs dissemination from decentralised WWT, in particular septic tanks, are needed to gain a comprehensive understanding of the major contributors of AMR and MGES to the environment. Thus, permitting the implementation of mitigation strategies/ policies that could help mediate the global AMR burden.

Appendices

Appendix A- Chapter 2

Table A.1: List of *intI*1 gene primer sets and probes reviewed in this study

Primer source ID	Assigned ID -Position*	Sequence (5'-3' Direction)	Primer length (bp)	Amplicon size (bp)	QPCR Study (QPCR Chemistry)	PCR Cycle Conditions	Environment	Reference
intI1-DF	DF- 383	TTCTGGAAGGCGAGCATC	18	108	Yes (MGB-TaqMan)	95^{O} – 10min; (95^{O}–30sec; 60^{O}–60sec) x45	Wastewater samples (Influent, sludge, and effluent)	**This study, modified from** (Rosewarne et al. 2010)
intI1-DR	DR- 474	TGCCGTGATCGAAATCC	17					
intI1-MGB	*DF*_P- 452	Fam-TGACCCGCAGTTGCA-MGB	15					
intI1/R	F1- 229	TCCACGCATCGTCAGGC	17	280	Yes (SYBR green I)	(94^{O}–30sec; 55^{O}–60^{O}sec; 72^{0}-15sec) x30	Clinical isolates (Avian)	(Bass et al. 1999)
intI1/F	R1- 492	CCTCCCGCACGATGATC	17					
345 R	F2- 673	CATTCCTGGCCGTGGTTCT	19	101	Yes (SYBR green)	95^{O}C – 10min; (95^{O} – 15sec; 60^{O} – 60sec) x30	Clinical isolates (Healthy adults)	(Skurnik et al. 2005)
245 F	R2- 750	TGAAAGGTCTGGTCATACATGTGA	24					
qINT-4	F3- 382	TTTCTGGAAGGCGAGCATCGTTTG	24	109	-	95^{O}–10min; (95^{O}-15sec ; 60^{O}-15sec) x40	Environmental samples (Sediments)	(Rosewarne et al. 2010)
qINT-3	R3- 467	TGCCGTGATCGAAATCCAGATCCT	24					
intI1-a-Fw	F4- 177	CGAAGTCGAGGCATTTCTGTC	21	217	Yes (SYBR Green)	95^{O}–7min; (95^{O}-10sec; 60^{O}–30sec) x40; Melt curve analysis	Environmental samples (Sediment from fish farms)	(Muziasari et al. 2014)
intI1-a-RV	R4- 374	GCCTTCCAGAAAACCGAGGA	20					
Int1F2	F5- 737	TCGTGCGTCGCCATCACA	18	67	Yes (SYBR Green)	(50^{O}-120sec; 95^{O}-10min) x1; (95^{O}-20sec ; 60^{O}-60sec) x40; Dissociation curve step	Environmental samples (Industrial waste, sewage sludge and pig slurry)	(Gaze et al. 2011)
Int1R2	R5- 782	GCTTGTTCTACGGCACGTTTGA	22					
intI1-a-RV	F6- 673	CATTCCTGGCCGTGGTTCT	19	146	Yes (SYBR Green)	95^{O}–15min; (95^{O}–15sec; 55^{O}–30sec; 72^{O}– 30sec) x45; melt curve: 65°to95° (0.5° increment/ 5sec)	Environmental samples (Surface water and sediment)	(Luo et al. 2010)
intI1-a-Fw	R6- 800	GGCTTCGTGATGCCTGCTT	19					
intI1-LC1	F7- 529	GCCTTGATGTTACCCGAGAG	20	196	Yes (TaqMan probe)	95^{O}–10 min; (95^{O}–30sec; 60^{O}–60sec) x45	Clinical /Laboratory strains	(Barraud et al. 2010)
intI1-LC5	R7- 707	GATCGGTCGAATGCGTGT	18					
intI1-probe	F7-P- 674	6 Fam-ATTCCTGGCCGTGGTTCTGGGTTTT-BHQ1	25					

intI1.F	F8- 461	GGGTCAAGGATCTGGATTTCG	21	483	Yes (SYBR Green)	94^{O}–5min; (94^{O}–30sec; 60^{O}–30sec; 72^{O}–60sec) x30	Laboratory reference strains	(Mazel et al. 2000)
intI1.R	R8- 923	ACATGCGTGTAAATCATCGTCG	22					
q*IntI1*F	F9- 93	ACCAACCGAACAGGCTTATG	20	286	Yes (SYBR Green)	95^{O}–15min; (95^{O}–30sec; 62^{O}–30sec; 72^{O}–30sec) x40; (72^{O}–10min) x1; Melt curve analysis	Environmental samples (Riverine estuarine, and Freshwater)	(Wright et al. 2008)
q*IntI1*R	R9- 358	GAGGATGCGAACCACTTCCAT	21					
*intI1*R476	F10- 539	TACCCGAGAGCTTGGCACCCA	21	312	Yes (TaqMan Probe)	95^{O}–5min; (95^{O}–30sec; 62^{O}–30sec; 72^{O}–45sec; plate read) x40	Environmental (Hospital, communal and urban WWTPs)	(Gillings et al. 2015; Paulus et al. 2019)
*intI1*F165	R10- 831	CGAACGAGTGGCGGAGGGTG	20					
	F10-P- 674	Texas615-TCGTGATGCCTGCTTGTTCTACGGCA	26					
IntI1-R	F11- 158	CTTCAGCCTTTTCCAGCAAC	20	308	No	94^{O}–3min; (95^{O}–30sec; 58.3^{O}–30sec; 72^{O}–40sec) x35; 72^{O}–7min	Clinical isolates (Urine, blood, and wound)	(Najafgholizadeh Pirzaman and Mojtahedi 2019)
IntI1-F	R11- 956	GAAACCTGCTCCAGCACTTC	20	818***				
intI-1F	F12- 395**	TCATGGCTTGTTATGACTGT	20	600	No	94^{O}–10min; (94^{O}–40sec; 57^{O}–50sec; 72^{O}–55sec) x30-40; 72^{O}–10min	Clinical isolates (Blood, urine, and wound)	(Mobaraki et al. 2018)
intI-1R	R12- 886**	GTAGGGCTTATTATGCACGC	20	512***				
ZANi1F	F13- 177	CGAAGTCGAGGCATTTC	17	102	Yes (SYBR Green)	95^{O}–3min; (95^{O}–20sec; 57.8^{O}–45sec; 72^{O}–30sec) x40; 72^{O}–10min	Environmental (Activated sludge, Pig Faeces)	(Yang et al. 2021)
ZANi1	R13- 261	ACCTTGCCGTAGAAGAAC	18					
hep35	F14- 458	TGCGGGTYAARGATBTKGATTT	22	491	No	(94^{O}–30sec; 55^{O}–30sec; 72^{O}–45sec) x30	Clinical isolates (Urinary tract)	(White et al. 2000)
hep36	R14- 931	CARCACATGCGTRTARAT	18					
HS463a	F15- 472	CTGGATTTCGATCACGGCACG	21	473	Yes (SYBR Green I)	94^{O}–3min; (94^{O}–30sec ; 60^{O}–30sec; 72^{O}–60sec) x35; Fluorescence acquisition step at 80^{O}	Environmental samples (Sediment)	(Hardwick et al. 2008)
HS464	R15- 923	ACATGCGTGTAAATCATCGTCG	22					
HS916	F16- 122	TTCGTGCCTTCATCCGTTTCC	21	371	No	94^{O}–3min; (94^{O}–30sec; 60/65^{O}–30sec; 72^{O}–90sec) x35; 72^{O}–5min	Clinical isolates (UTIs from outpatients).	(Márquez et al. 2008)
HS915	R16- 472	CGTGCCGTGATCGAAATCCAG	21					
intB	F17- 40	GTCAAGGTTCTGGACCAGTTGC	22	892	No	(96^{O}–5min; 55^{O}–1min; 70^{O}–3min) x1; (96^{O}–15sec; 55^{O}–30sec; 70^{O}–3min) x24; 70^{O}–5min	Environmental isolates (Estuarine environment)	(Rosser and Young 1999)
intA	R17- 911	ATCATCGTCGTAGAGACGTCGG	22					
*intI1*U/F	F18- 35	GTTCGGTCAAGGTTCTG	17	923	No	94^{O}–5min; (94^{O}–30sec; 50^{O}–30sec ; 72^{O}–90sec) x30; 72^{O}–7min	Clinical isolates (Healthy human patients)	(Zhang et al. 2004)
*intI1*D/R	R18- 940	GCCAACTTTCAGCACATG	18					
*intI1*F	F19- 35	GTTCGGTCAAGGTTCTGG	18	890	No	95^{O}–2min; (95^{O}–20sec; 54^{O}–30sec; 70^{O}–30sec) x35	Environmental isolate (Soil, wastewater)	(Xu et al. 2007)
*intI1*R	R19- 907	CGTAGAGACGTCGGAATG	18					

*intI1*F	F20- 88	AGCTTACGAACCGAACAGGC	20	950	No	94^{O}–5min; (92^{O}–30sec; 55^{O}–30sec; 72^{O}–1min) x35; 72^{O}–10min	Environmental samples (Sediments)	(Borruso et al. 2016)
Intf2	R20- 666	TCCGCCAGGATTGACTTGCG	20					
intI-F	F21- 253	GCCTTGCTGTTCTTCTAC	18	558	No	94^{O}–5min; (94^{O}–30sec; 55^{O}–30sec; 72^{O}–2.5min) x35; 72^{O}–5min	Clinical/Reference isolates	(Guerra et al. 2001)
intI-B	R21- 793	GATGCCTGCTTGTTCTAC	18					
*intI1*R	F22- 189	ATTTCTGTCCTGGCTGGCGA	20	568	No	-	Clinical isolates	(Ploy et al. 2000)
*intI1*L	R22- 737	ACATGTGATGGCGACGCACGA	21					
IntI1-F	F23- 462	GGTCAAGGATCTGGATTTCG	20	483	No	94^{O}–12min; (94^{O}–30sec; 62^{O}–30sec ; 72^{O}–1min) x30; 72^{O}–8min	Clinical isolates	(Machado et al. 2005)
IntI1-R	R23- 924	ACATGCGTGTAAATCATCGTC	21					
*IntI1*R	**F24- 238**	CCCGAGGCATAGACTGTA	18	160	No	(94^{O}–30sec; 55^{O}–30sec; 72^{O}–30sec) x35	Clinical isolates (including respiratory tracts)	(Koeleman et al. 2001)
*IntI1*F	**R24- 100**	CAGTGGACATAAGCCTGTTC	20					
Int-F	F25- 10	GCCACTGCGCCGTTACCACC	20	898	No	94^{O}– 5min; (94^{O}–15sec; 69^{O}–30sec; 72^{O}–60sec) x30; 72^{O}–7min	Clinical isolates (UTIs strains)	(Kerrn et al. 2002)
Int-B	R25- 888	GGCCGAGCAGATCCTGCACG	20					
Int-F	F26- 10	GCCACTGCGCCGTTACCACC	20	276	No	-	Clinical/ Environmental isolates (human, animal faeces and urine)	(Ho et al. 2012)
Int1-285B	R26- 265	GCACAGCACCTTGCCGTAGAA	21					
Inti1F	F27-182**	CGAGGCATAGACTGTAC	18	925	No	-	Clinical isolates (Stool, blood, and urine)	(Orman et al. 2002)
Inti1R	R27-868**	TTCGAATGTCGTAACCGC	17	703***				
Int1 R	F28- 40	GTCAAGGTTCTGGACCAGTTG	22	550		94^{O}–5min; (94^{O}–1min; 50^{O}–1min; 72^{O}–1min) x30; 72^{O}–5min	Clinical isolates (Urine)	(Bashir et al. 2015)
Int1 F	R28- 911	ATCATCGTCGTAGAGACGTCGG	21	892***				
IntI1-R	F29- 305	AGGAGATCCGAAGACCTC	18	243	No	94^{O}–5min; (94^{O}–1min; 55^{O}–1min; 72^{O}–30sec) x35	Clinical isolates	(Leverstein-Van Hall et al. 2003)
IntI1-F	R29- 530	TCTCGGGTAACATCAAGG	18					
Int1A	F30- 4	AAAACCGCCACTGCGCCGTTA	21	1201	No	-	Clinical isolate	(Falbo et al. 1999; Fonseca et al. 2005)
Int1B	R30- 977	GAAGACGGCTGCACTGAACG	20	996***				
Int I.R	F31- 261	GTTCTTCTACGGCAAGGT	18	287	No	94^{O}–5min; (94^{O}–20sec; 60^{O}–30sec; 72^{O}–60sec) x30	Clinical and animal facility (Faecal samples)	(Kheiri and Akhtari 2016)
Int I.F	R31- 530	TCTCGGGTAACATCAAGG	18					
F6	F32- 371	GCATCCTCGGTTTTCTGG	18	457	No	94^{O}–2min; (94^{O}–1min; 57^{O}–1min; 72^{O}–90sec) x30	Clinical (laboratory strains)	(Shibata et al. 2003)
R6	R32- 810	GGTGTGGCGGGCTTCGTG	18					
*intI*1F	F33- 111	TGTCCACTGGGTTCGTGCCT	20	707	No	94^{O}–5min; (94^{O}–30sec; 56^{O}–30sec; 72^{O}–1min) x30; 72^{O}–10min	Food (Raw meat samples)	(Zhou et al. 2019)

*intI*1R	R33- 797	GCTTCGTGATGCCTGCTTGTT	21					
*intI*1L	F34- 253	GCCTTGCTGTTCTTCTACGG	20	558	No	-	Clinical isolates	(Ng et al. 1999)
*intI*1R	R34- 791	GATGCCTGCTTGTTCTACGG	20					
*int*R	F35- 409	GCCCAGCTTCTGTATGGAAC	20	145	Yes (SYBR Premix Ex-Taq)	95^{O}–10sec; (95^{O}–10sec; 55^{O}–10sec; 72^{O}–10sec) x40	Clinical/ Laboratory strains	(Wei et al. 2011)
*int*F	R35- 534	CCAAGCTCTCGGGTAACATC	20					
*int*M1-U	F36- 209	ACGAGCGCAAGGTTTCGGT	19	565	No	94^{O}–5min; (94^{O}–30sec; 52^{O}–30sec; 72^{O}–2min) x30; 72^{O}–7min	Clinical isolates (Faeces, blood, and urine)	(Su et al. 2006)
*int*M1-D	R36- 753	GAAAGGTCTGGTCATACATG	20					
*Int*1-1	F37- 176	GCGAAGTCGAGGCATTTCTGTC	22	766	No	-	Clinical strains	(Rodríguez-Martínez et al. 2007)
*Int*1-2	R37- 917	ATGCGTGTAAATCATCGTCGTAGAGA	26					
*intI1*F	F38- 167	TGGGCAGCAGCGAAGTC	17	70	Yes (TaqMan Probe)	-	Clinical isolates (Veterinary or food isolates)	(Bugarel et al. 2011)
*intI1*R	R38- 219	TGCGTGGAGACCGAAACC	18					
intI1-Probe	F38 Pb- 185	AGGCATTTCTGTCCTGGCTGGCG	23					
D, intI	F39- 168	GGGCAGCAGCGAAGTCGAGGC	21	845	No	96^{O}–1min; (96^{O}–30sec ; 58^{O}–30sec) x35; 70^{O}–60sec; 70^{O}–7min	Environmental (Commercial pet turtle eggs and water ponds)	(Díaz et al. 2006)
A, intI	R39- 991	CTACCTCTCACTAGTGAGGGGCGG	24					
INT1-R	F40- 42	CAAGGTTCTGGACAGTTGC	19	900	No	-	Clinical isolates	(Adabi et al. 2009)
INT1-F	R40- 922	TGCGTGTAAATCATCGTCGT	20					
intI1-R	F41- 261	GTTCTTCTACGGCAAGGTG	19	515	No	94^{O}–3min; (94^{O}–60sec; - ; 72^{O}–10min) x30	Clinical isolates (Hospitalised patients with nosocomial infections)	(Wang et al. 2017a)
intI1-F	R41- 756	GCTGAAAGGTCTGGTCATAC	20					
*IntI1*F	F42- 520	AAGGATCGGGCCTTGATGTT	20	471	No	94^{O}–5min; (94^{O}–1min; 55^{O}–1min; 72^{O}–1min) x30	Clinical isolates (Blood, pus and urine)	(Pongpech et al. 2008)
*IntI1*R	R42- 971	CAGCGCATCAAGCGGTGAGC	20					
*intI1*R	F43- 408	CGCCCAGCTTCTGTATGG	18	408***	No	94^{O}–5min; (94^{O}–30sec; 51^{O}–40sec; 72^{O}–40sec) x35; 72^{O}–5min	Clinical isolate	(Gu et al. 2008)
*intI1*F	R43- 797	TTCGTGATGCCTGCTTGTT	19					
IntI1-f	F44- 70	ATACGCTACTTGCATTACAG	20	521	No	94^{O}–5min; (94^{O}–30sec; 51^{O}–45sec; 72^{O}–1-4min) x35; 72^{O}–7min	Clinical isolate (ICU patients)	(Zong et al. 2008)
IntI1-r	R44- 571	GCCCGTGCACGCGACAGCTG	20					
*intI1*R	F45- 641	ACGCCCTTGAGCGGAAGTATC	21	248	No	94^{O}–4min; (94^{O}–1min; 65^{O}–30sec[-1^{O} per cycle]; 70^{O}–2min) x10 touchdown cycles; (94^{O}–1min; 55^{O}–30ec; 70^{O}–2min) x24; 70^{O}–5min	Clinical (Human faecal samples and pig rectal swabs)	(Phongpaichit et al. 2007)
*intI1*F	R45- 869	GGTTCGAATGTCGTAACCGC	20					

INT1-R	F46- 1	CATGAAAACCGCCACTGC	18	553	No	-	Clinical isolates (including blood, urine and respiratory specimens)	(Hong et al. 2016)
INT1-F	R46- 534	CCAAGCTCTCGGGTAACATC	20					
INT1-mF	R46 Pb- 88	GCCTGTTCGGTTCGTAAGCT	20					
Int1 lower	F47- 10	GCCACTGCGCCGTTACCACCGC	22	900	No	94^{O}–5min; (94^{O}–30sec; 60^{O}–40sec; 72^{O}–1min) x35; 72^{O}–7min	Clinical isolates	(Frank et al. 2007)
Int1 upper	R47- 888	ATGGCCGAGCAGATCCTGCACG	22					
Int1-1B	F48- 1	ATGAAAACCGCCACTGCGCC	20	1012	No	-	Clinical isolates	(Yan et al. 2006)
Int1-1A	R48- 994	TACCTCTCACTAGTGAGGGG	20					
class1-B	F49- 42	CAAGGTTCTGGACCAGTTGCG	21	871	No	94^{O}–45sec; 5^{O} below melting temperature–60sec; 72^{O}–2min	Environmental isolates (Pig farm)	(Sandvang et al. 2002; Spindler et al. 2012)
class1-F	R49- 895	CGGAATGGCCGAGCAGATC	19					
IntI-1R	F50- 28	CCGCTGCGTTCGGTCAAGGT	20	753	No	-	Clinical isolate (Sputum, pus, blood and urine)	(Piyakul et al. 2012)
IntI-1F	R50- 761	GGCGCGCTGAAAGGTCTGGT	20					
Int I-R	F51- 454	CAACTGCGGGTCAAGGAT	18	542	No	-	Clinical isolates	(Yu et al. 2012)
Int I-F	R51- 920	CGTGTAAATCATCGTCGTAG	20	486***				
Δ*intI1*-F	F52- 50	TGGACCAGTTGCGTGAGC	18	601	No	-	Poultry (Chicken intestinal content and faeces)	(Lai et al. 2013)
Δ*intI1*-R	R52- 634	TCAAGGGCGTCGGGAAG	17					
int_F	F53- 144	CGATGCGTGGAGACCGAAACCTT	23	573***	No	95^{O}–5min; (95^{O}–1min; 58^{O}– 1min; 72^{O}–5min[+5sec each extension time cycle]) x35; 72^{O}–7min	Clinical isolates	(Krauland et al. 2009)
int_R	R53- 694	GTAACGCGCTTGCTGCTTGGATGC	24					
intI1-R	F54- 30	GCTGCGTTCGGTCAAGGT	18	788	No	-	Clinical and Food isolates	(Zhao et al. 2018)
intI1-F	R54- 799	GCTTCGTGATGCCTGCTTG	19					
Int1-F6	F55- 185	AAGCAGACTTGACCTGA	17	457	No	-	Clinical isolates	(Kainuma et al. 2018)
Int1-R6	R55- 810	GGTGTGGCGGGCTTCGTG	18	643***				
i1219 (r)	F56- 641	ACGCCCTTGAGCGGAAGTATC	21	254	No	94^{O}–4min; (94^{O}–1min; 65^{O}–30sec[-1^{O} per cycle]; 70^{O}–2min) x10 touchdown cycles; (94^{O}–1min; 55^{O}–30ec; 70^{O}–2min) x24; 70^{O}–5min	Clinical/Animal isolates	(Ebner et al. 2004)
i965 (f)	R56- 869	CCTTCGAATGCTGTAACCGC	20	248***				
IntI-B	F57- 166	TTGGGCAGCAGCGAAGT	17	587	No	94^{O}–5min; (94^{O}–40sec; 58^{O}–40ec; 72^{O}–45sec) x35; 72^{O}–5min	Clinical isolates (Sputum)	(Wang et al. 2017b)
IntI-F	R57- 736	TGATGGCGACGCACGAC	17					
intl1-f	F58- 371	GCATCCTCGGTTTTCTGG	18	644	No	-	Clinical isolates	(Kobayashi et al. 2013)
intl1-1014r	R58- 995	CTACCTCTCACTAGTGAGGG	20					

intR	F59- 399	TCGTTTGTTCGCCCAGC	17	497	No	95^{O}–5min; (95^{O}–45sec; 50^{O}–45ec; 72^{O}–1min) x30	Poultry and Swine isolates	(Chuanchuen et al. 2007)
intF	R59- 879	CCTGCACGGTTCGAATG	17					
int1LF	F60- 304	CAGGAGATCGGAAGACCT	18	152***	Yes (SYBR Green I)	-	Swine isolates	(Ekkapobyotin et al. 2008)
int1LR	R60- 439	TTGCAAACCCTCACTGAT	18					
*intI1*F	**F61**- 587	GTGGATGGCGGCCTGAAGCC	20	140***	No	94^{O}–5min; (95^{O}–1min; 55^{O}–30sec; 72^{O}–1min) x30; 72^{O}–5min	Animal (Foals faecal sample)	(Kennedy et al. 2018)
*intI1*R	**R61**- 710	ATTGCCCAGTCGGCAGCG	18					
intA	**F62**- 154**	ACAGGGCAAGCTTAGTAAAGCC	22	623**	No	(95^{O}–1min c; 67^{O}–1min c; 72^{O}–1min) x30; 72^{O}–5min	Environmental (Pig slurry and manured agricultural soils)	(Byrne-Bailey et al. 2009)
intB	**R62**- 758**	CTCGCTAGAACTTTTGGAAA	20					
int1-L	**F63**- 31	CTGCGTTCGGTCAAGGTTCT	20	882	No	-	Food isolates (commercial fish and seafood)	(Ryu et al. 2012)
int1-F	**R63**- 893	GGAATGGCCGAGCAGATCCT	20					
*Int1*F	**F64**- 584	CACGGATATGCGACAAAAAG	20	160	No	94^{O}–5min; (94^{O}–1min; 51^{O}–1min; 72^{O}–45sec) x35; 72^{O}–5min	Clinical (Blood) and Environmental (Tap water)	(Karami et al. 2020)
*Int1*R	**R64**- 833	GATGACAACGAGTGACGAAATG	22	271***				

F, forward primer; R, reverse primer; P, probe

* denotes hit start position of primer based on alignment to a reference *intI*1 gene sequence shown in Figure 2.1. ** denotes hit position based on Primer prospector alignment of CP003684.1 *** denotes estimated amplicon size based on alignment of primer to CP003684.1 *intI*1 reference nucleotide sequence.

Table A.2: Coverage and specificity of currently published and newly modified *intI*1 primer pairs

PCR Primer Set	Coverage Test									Specificity Test					
	*intI*1 Sub-databases									*intI*1-*Like sequence* (n= 15)			*Non-intI*1 *sequence* (n= 1540)		
	SDB1 (n= 104)			SDB2 (n= 144)			SDB3 (n=503)								
	Number of sequences with correct primer orientation (%)	Mean Weighted Score	Target amplified (%) (0 Mismatches)	Number of sequences with correct primer orientation (%)	Mean Weighted Score	Target amplified (%) (0 mismatch)	Number of sequences with correct primer orientation (%)	Mean Weighted Score	Target amplified (%) (0 mismatch)	Number of sequences with correct primer orientation (%)	Mean Weighted Score	Target amplified (%) (0 mismatch)	Number of sequences with correct primer orientation (%)	Mean Weighted Score	Target amplified (%) (0 mismatch)
DF-DR	104 (100%)	**F** 0.008 / **R** 0	102 (98%)	144 (100%)	**F** 0.006 / **R** 0	142 (99%)	502 (100%)	**F** 0.016 / **R** 0.002	493 (99%)	15 (100%)	**F** 0.6 / **R** 0.07	9 (60%)	813 (53%)	**F** 4.991 / **R** 4.223	0 (0%)
F1-R1	104 (100%)	**F** 0.017 / **R** 0.004	101 (97%)	144 (100%)	**F** 0.013 / **R** 0.003	141 (98%)	458 (91%)	**F** 0.041 / **R** 0.002	448 (98%)	14 (93%)	**F** 0.4 / **R** 0.057	9 (64%)	736 (48%)	**F** 4.342 / **R** 4.493	0 (0%)
F2-R2	104 (100%)	**F** 0.019 / **R** 0.042	97 (93%)	143 (99%)	**F** 0.014 / **R** 0.073	133 (93%)	443 (88%)	**F** 0.009 / **R** 0.08	413 (94%)	11 (73 %)	**F** 0.491 / **R** 0.782	7 (64%)	923 (60%)	**F** 5.215 / **R** 5.992	0 (0%)
F3-R3	104 (100%)	**F** 0.012 / **R** 0	101 (97%)	144 (100%)	**F** 0.008 / **R** 0	141 (98%)	501 (100%)	**F** 0.003 / **R** 0.006	490 (98%)	14 (93%)	**F** 0.171 / **R** 0.314	8 (57%)	997 (65%)	**F** 6.024 / **R** 5.579	0 (0%)
F4-R4	104 (100%)	**F** 0.037 / **R** 0.05	98 (94%)	144 (100%)	**F** 0.026 / **R** 0.036	138 (96%)	441 (88%)	**F** 0.034 / **R** 0.044	428 (97%)	14 (93%)	**F** 0.671 / **R** 0.529	9 (64%)	879 (57%)	**F** 5.94 / **R** 5.534	0 (0%)
F5-R5	104 (100%)	**F** 0.079 / **R** 0.035	96 (92%)	140 (97%)	**F** 0.059 / **R** 0.169	118 (84%)	455 (90%)	**F** 0.596 / **R** 1.026	315 (70%)	14 (93%)	**F** 2.371 / **R** 2.929	6 (43%)	912 (59%)	**F** 4.389 / **R** 4.968	0 (0%)
F6-R6	104 (100%)	**F** 0.019 / **R** 0.06	94 (90%)	127 (88%)	**F** 0.052 / **R** 0.085	116 (91%)	318 (63%)	**F** 0.028 / **R** 0.145	294 (93%)	9 (60%)	**F** 0.089 / **R** 0.267	7 (78%)	972 (63%)	**F** 5.37 / **R** 4.246	0 (0%)
F7-R7	104 (100%)	**F** 0.046 / **R** 0.008	101 (97%)	144 (100%)	**F** 0.065 / **R** 0.031	140 (97%)	475 (94%)	**F** 0.02 / **R** 0.009	469 (99%)	10 (67%)	**F** 0.260 / **R** 0.080	8 (80%)	778 (51%)	**F** 5.882 / **R** 4.636	0 (0%)
F8-R8	**F** 0			**F** 0			**F** 0.005			8	**F** 0.05			**F** 4.617	0

	101 (97%)	**R** 0	101 (100%)	115 (80%)	**R** 0.012	114 (99%)	243 (48%)	**R** 0.113	221 (92%)	(53%)	**R** 0.525	5 (63%)	1194 (78%)	**R** 4.634	(0%)
F9-R9	104 (100%)	**F** 0.396 **R** 5.019	0 (0%)	141 (98%)	**F** 0.372 **R** 5.014	0 (0%)	420 (83%)	**F** 0.345 **R** 5.005	0 (0%)	13 (87%)	**F** 0.615 **R** 5.108	0 (0%)	848 (55%)	**F** 5.302 **R** 5.055	0 (0%)
F10-R10	104 (100%)	**F** 0.063 **R** 0.117	93 (89%)	126 (88%)	**F** 0.052 **R** 0.097	115 (91%)	302 (60%)	**F** 0.022 **R** 0.044	289 (96%)	9 (60%)	**F** 1.178 **R** 0.8	6 (67%)	908 (59%)	**F** 5.438 **R** 5.549	0 (0%)
F13-R13	104 (100%)	**F** 0.013 **R** 0.004	101 (97%)	144 (100%)	**F** 0.01 **R** 0.008	139 (97%)	473 (94%)	**F** 0.392 **R** 0.231	431 (92%)	13 (87%)	**F** 0.492 **R** 0.338	9 (69%)	870 (56%)	**F** 4.226 **R** 4.85	0 (0%)
F14-R14	102 (98%)	**F** 0 **R** 0.025	101 (99%)	117 (81%)	**F** 0 **R** 0.055	114 (97%)	227 (45%)	**F** 0.004 **R** 0.041	217 (96%)	6 (40%)	**F** 0.133 **R** 0.333	4 (67%)	1055 (69%)	**F** 4.052 **R** 2.956	14 (0.9%)
F15-R15	101 (97%)	**F** 0 **R** 0	101 (100%)	115 (80%)	**F** 0 **R** 0.012	114 (99%)	243 (48%)	**F** 0.002 **R** 0.113	223 (93%)	8 (53%)	**F** 0.05 **R** 0.525	5 (63%)	1051 (68%)	F 5.496 R 4.356	0 (0%)
F16-R16	104 (100%)	**F** 0.008 **R** 0	102 (98%)	144 (100%)	**F** 0.019 **R** 0	141 (98%)	427 (85%)	**F** 0.02 **R** 0.014	415 (98%)	14 (93%)	**F** 0.3 **R** 0.029	11 (79%)	961 (62%)	**F** 5.64 **R** 5.543	0 (0%)
F17-R17	104 (100%)	**F** 0.067 **R** 0.121	98 (94%)	142 (99%)	**F** 0.444 **R** 0.866	104 (73%)	466 (93%)	**F** 1.028 **R** 2.596	154 (33%)	15 (100%)	**F** 0.827 **R** 2.360	3 (20%)	787 (51%)	**F** 5.847 **R** 5.709	0 (0%)
F18-R18	104 (100%)	**F** 0.04 **R** 0.121	98 (94%)	144 (100%)	**F** 0.246 **R** 1.158	100 (69%)	484 (96%)	**F** 0.568 **R** 2.632	128 (27%)	14 (93%)	**F** 0.257 **R** 3.186	2 (14%)	814 (53%)	**F** 4.153 **R** 4.17	0 (0%)
F19-R19	104 (100%)	**F** 0.035 **R** 0.115	98 (94%)	142 (99%)	**F** 0.166 **R** 0.465	108 (76%)	482 (96%)	**F** 0.438 **R** 1.413	168 (35%)	15 (100%)	**F** 0.347 **R** 1.280	4 (27%)	711 (46%)	**F** 4.44 **R** 4.224	0 (0%)
F21-R21	104 (100%)	**F** 0.004 **R** 0.048	99 (95%)	144 (100%)	**F** 0.003 **R** 0.542	121 (84%)	501 (99%)	**F** 0.427 **R** 1.554	272 (55%)	14 (93%)	**F** 0.629 **R** 1.957	4 (29%)	800 (52%)	**F** 4.659 **R** 4.509	0 (0%)
F22-R22	104 (100%)	**F** 0.015 **R** 0.062	97 (93%)	144 (100%)	**F** 0.025 **R** 0.075	134 (93%)	462 (92%)	**F** 0.879 **R** 0.06	377 (82%)	10 (71%)	**F** 2.020 **R** 0.08	5 (50%)	872 (57%)	**F** 4.962 **R** 5.87	0 (0%)
F23-R23	101 (97%)	**F** 0 **R** 0	101 (100%)	115 (80%)	**F** 0 **R** 0.012	114 (99%)	245 (49%)	**F** 0.033 **R** 0.17	221 (91%)	8 (53%)	**F** 0.05 **R** 0.525	5 (63%)	1149 (75%)	**F** 4.559 **R** 5.013	0 (0%)
F25-R25	102 (98%)	**F** 0.096 **R** 0.041	91 (89%)	130 (90%)	**F** 0.075 **R** 0.755	99 (76%)	381 (76%)	**F** 0.13 **R** 2.696	162 (43%)	13 (87%)	**F** 0.446 **R** 2.492	4 (31%)	843 (55%)	**F** 5.735 **R** 5.161	0 (0%)
F26-R26	102 (98%)	**F** 0.096 **R** 0.008	95 (93%)	130 (90%)	**F** 0.075 **R** 0.012	121 (93%)	380 (76%)	**F** 0.111 **R** 0.084	361 (96%)	13 (87%)	**F** 0.446 **R** 0.138	8 (62%)	659 (43%)	**F** 5.926 **R** 5.661	0 (0%)
F29-R29	104 (100%)	**F** 0.408 **R** 0.017	0 (0%)	143 (99%)	**F** 0.406 **R** 0.013	0 (0%)	489 (97%)	**F** 0.442 **R** 0.004	0 (0%)	13 (87%)	**F** 0.462 **R** 0.431	0 (0%)	661 (43%)	**F** 4.557 **R** 4.826	0 (0%)

F31-R31	104 (100%)	**F** 0.004 **R** 0.017	101 (97%)	143 (99%)	**F** 0.027 **R** 0.013	138 (97%)	502 (100%)	**F** 0.263 **R** 0.004	464 (93%)	14 (93%)	**F** 0.514 **R** 0.4	8 (57%)	674 (44%)	**F** 4.769 **R** 4.572	0 (0%)
F32-R32	104 (100%)	**F** 0.012 **R** 0.075	97 (93%)	126 (88%)	**F** 0.01 **R** 0.062	119 (94%)	324 (64%)	**F** 0.085 **R** 0.223	294 (91%)	8 (53%)	**F** 0.650 **R** 0.45	4 (50%)	896 (58%)	**F** 4.929 **R** 4.569	0 (0%)
F33-R33	104 (100%)	**F** 0.098 **R** 0.06	78 (75%)	137 (95%)	**F** 0.088 **R** 0.693	97 (71%)	432 (86%)	**F** 0.606 **R** 2.308	196 (46%)	13 (87%)	**F** 0.754 **R** 2.923	2 (15%)	903 (59%)	**F** 5.245 **R** 5.266	0 (0%)
F34-R34	104 (100%)	**F** 0.033 **R** 0.042	98 (94%)	144 (100%)	**F** 0.031 **R** 0.771	120 (83%)	502 (100%)	**F** 0.324 **R** 2.363	272 (54%)	14 (93%)	**F** 0.471 **R** 2.371	4 (29%)	787 (51%)	**F** 5.511 **R** 5.18	0 (0%)
F35-R35	104 (100%)	**F** 0 **R** 0.015	102 (98%)	143 (99%)	**F** 0.003 **R** 0.011	140 (98%)	502 (100%)	**F** 0.012 **R** 0.003	493 (99%)	14 (93%)	**F** 0. 3 **R** 0.186	10 (71%)	808 (52%)	**F** 5.52 **R** 5.068	0 (0%)
F36-R36	104 (100%)	**F** 0.069 **R** 0.033	93 (89%)	143 (99%)	**F** 0.067 **R** 0.057	124 (87%)	441 (88%)	**F** 0.771 **R** 0.066	349 (79%)	10 (67%)	**F** 1.62 **R** 1.28	5 (50%)	703 (46%)	**F** 5.239 **R** 5.468	0 (0%)
F37-R37	104 (100%)	**F** 0.037 **R** 0.148	98 (94%)	144 (100%)	**F** 0.026 **R** 1.11	111 (77%)	502 (100%)	**F** 0.619 **R** 2.888	167 (33%)	15 (100%)	**F** 1.107 **R** 2.88	2 (13%)	1105 (72%)	**F** 5.73 **R** 6.227	0 (0%)
F38-R38	104 (100%)	**F** 0 **R** 0.083	95 (91%)	144 (100%)	**F** 0 **R** 0.081	132 (92%)	438 (87%)	**F** 0.017 **R** 0.042	419 (96%)	13 (87%)	**F** 0.262 **R** 0.185	9 (69%)	801 (52%)	**F** 4.829 **R** 4.502	0 (0%)
F39-R39	104 (100%)	**F** 0.01 **R** 0.919	76 (73%)	144 (100%)	**F** 0.007 **R** 1.714	87 (60%)	493 (98%)	**F** 0.875 **R** 3.318	107 (22%)	13 (87%)	**F** 1.138 **R** 4.985	1 (8%)	751 (49%)	**F** 5.503 **R** 6.96	0 (0%)
F40-R40	104 (100%)	**F** 1.033 **R** 0.165	1 (0.9%)	142 (99%)	**F** 1.165 **R** 0.987	1 (0.7%)	466 (93%)	**F** 1.389 **R** 2.593	1 (0.2%)	15 (100%)	**F** 1.293 **R** 2.653	0 (0%)	1146 (74%)	**F** 4.994 **R** 3.732	0 (0%)
F41-R41	104 (100%)	**F** 0.004 **R** 0.004	102 (98%)	143 (99%)	**F** 0.013 **R** 0.062	136 (95%)	442 (88%)	**F** 0.511 **R** 0.1	389 (88%)	11 (73 %)	**F** 0.873 **R** 0.509	5 (45%)	699 (45%)	**F** 4.971 **R** 4.945	0 (0%)
F42-R42	101 (97%)	**F** 0.004 **R** 0.17	90 (89%)	114 (79%)	**F** 0.04 **R** 0.195	102 (89%)	208 (41%)	**F** 0.024 **R** 0.138	190 (92%)	4 (27%)	**F** 0.2 **R** 1.750	2 (50%)	757 (49%)	**F** 4.577 **R** 4.857	0 (0%)
F43-R43	104 (100%)	**F** 0 **R** 0.044	99 (95%)	137 (95%)	**F** 0.003 **R** 0.472	121 (88%)	439 (87%)	**F** 0.005 **R** 1.604	300 (69%)	11 (73 %)	**F** 0 **R** 1.964	6 (55%)	748 (49%)	**F** 4.74 **R** 4.518	0 (0%)
F44-R44	104 (100%)	**F** 0.06 **R** 0.012	97 (93%)	139 (97%)	**F** 0.045 **R** 0.035	131 (94%)	403 (80%)	**F** 0.041 **R** 0.015	387 (97%)	13 (87%)	**F** 0.138 **R** 0.692	6 (46%)	835 (54%)	**F** 5.151 **R** 5.451	0 (0%)
F45-R45	104 (100%)	**F** 0.015 **R** 0	100 (96%)	126 (88%)	**F** 0.056 **R** 0.037	120 (95%)	316 (63%)	**F** 0.023 **R** 0.409	281 (89%)	9 (60%)	**F** 0.044 **R** 0.111	8 (89%)	947 (61%)	**F** 5.508 **R** 4.907	0 (0%)
F46-R46	102 (98%)	**F** 0.041 **R** 0.016	95 (93%)	131 (91%)	**F** 0.079 **R** 0.069	121 (92%)	386 (77%)	**F** 0.166 **R** 0.023	363 (95%)	13 (87%)	**F** 0.246 R 0.677	8 (62%)	767 (50%)	**F** 4.834 **R** 4.968	0 (0%)

F47-R47	102 (98%)	**F** 0.057 **R** 0.057	91 (89%)	130 (90%)	**F** 0.045 **R** 0.808	99 (76%)	382 (76%)	**F** 0.114 **R** 2.895	161 (42%)	13 (87%)	**F** 0.446 **R** 2.677	4 (31%)	707 (46%)	**F** 5.875 **R** 5.318	0 (0%)
F48-R48	103 (99%)	**F** 0.134 **R** 0.746	71 (69%)	143 (99%)	**F** 0.55 **R** 1.432	72 (50%)	486 (97%)	**F** 1.209 **R** 2.807	77 (16%)	13 (87%)	**F** 0.554 **R** 4.523	2 (15%)	880 (57%)	**F** 5.306 **R** 5.722	0 (0%)
F49-R49	104 (100%)	**F** 0.081 **R** 0.075	95 (91%)	142 (99%)	**F** 0.552 **R** 0.449	103 (73%)	468 (93%)	**F** 1.309 **R** 1.229	167 (36%)	15 (100%)	**F** 1.053 **R** 1.24	5 (33%)	839 (54%)	**F** 5.465 **R** 4.357	0 (0%)
F50-R50	104 (100%)	**F** 0.081 **R** 0.008	101 (97%)	144 (100%)	**F** 0.59 **R** 0.124	125 (87%)	493 (98%)	**F** 1.281 **R** 0.534	301 (61%)	10 (67%)	**F** 0.58 **R** 1.0	7 (70%)	782 (51%)	**F** 5.203 **R** 5.264	0 (0%)
F51-R51	102 (97%)	**F** 0 **R** 0.029	101 (99%)	118 (82%)	**F** 0 **R** 0.097	114 (97%)	256 (51%)	**F** 0.017 **R** 0.184	227 (89%)	8 (53%)	**F** 0.15 **R** 0.35	5 (63%)	807 (52%)	**F** 4.188 **R** 5.266	0 (0%)
F52-R52	104 (100%)	**F** 0.081 **R** 0.008	98 (94%)	144 (100%)	**F** 0.392 **R** 0.026	125 (87%)	503 (100%)	**F** 0.95 **R** 0.061	369 (74%)	15 (100%)	**F** 0.52 **R** 1.32	9 (60%)	823 (53%)	**F** 4.126 **R** 4.243	0 (0%)
F53-R53	104 (100%)	**F** 4.219 **R** 5.65	0 (0%)	144 (100%)	**F** 4.214 **R** 5.631	0 (0%)	426 (85%)	**F** 4.208 **R** 5.612	0 (0%)	13 (87%)	**F** 4.231 **R** 6.4	0 (0%)	715 (46%)	**F** 5.867 **R** 6.198	0 (0%)
F54-R54	104 (100%)	**F** 0.079 **R** 0.06	97 (93%)	142 (99%)	**F** 0 **R** 0.748	121 (85%)	475 (94%)	**F** 1.096 **R** 2.215	211 (45%)	14 (93%)	**F** 0.357 **R** 3.171	5 (36%)	893 (60%)	**F** 4.744 **R** 5.006	0 (0%)
F56-R56	104 (100%)	**F** 0.015 **R** 1.6	0 (0%)	141 (98%)	**F** 0.011 **R** 1.923	0 (0%)	453 (90%)	**F** 0.004 **R** 2.620	0 (0%)	10 (67%)	**F** 0.04 **R** 1.98	0 (0%)	674 (44%)	**F** 5.644 **R** 4.724	0 (0%)
F57-R57	103 (99%)	**F** 0 **R** 0.008	101 (98%)	142 (99%)	**F** 0 **R** 0.006	140 (99%)	459 (91%)	**F** 0.458 **R** 0.014	381 (83%)	12 (80%)	**F** 0.667 **R** 0.533	7 (58%)	636 (41%)	**F** 4.115 **R** 4.273	0 (0%)
F58-R58	104 (100%)	**F** 0.012 **R** 0.812	75 (72%)	144 (100%)	**F** 0.008 **R** 1.497	86 (60%)	503 (100%)	**F** 0.055 **R** 2.937	158 (32%)	13 (87%)	**F** 0.4 **R** 4.523	1 (8%)	943 (61%)	**F** 4.866 **R** 5.583	0 (0%)
F59-R59	104 (100%)	**F** 0 **R** 0.004	103 (99%)	124 (86%)	**F** 0 **R** 0.003	123 (99%)	314 (62%)	**F** 0.012 **R** 0.206	277 (88%)	9 (60%)	**F** 0.378 **R** 0	8 (89%)	1082 (70%)	**F** 4.613 **R** 3.979	0 (0%)
F60-R60	104 (100%)	F 0.008 R 0.04	101 (97%)	144 (100%)	**F** 0.006 **R** 0.003	141 (98%)	487 (97%)	**F** 0.002 **R** 0.001	482 (99%)	14 (93%)	**F** 0.086 **R** 0.214	10 (71%)	774 (50%)	**F** 5.16 **R** 4.302	0 (0%)
F62-R62	102 (98%)	**F** 8.233 **R** 4.418	0 (0%)	141 (98%)	**F** 8.241 **R** 4.472	0 (0%)	437 (87%)	**F** 8.174 **R** 4.717	0 (0%)	13 (87%)	**F** 7.831 **R** 5.354	0 (0%)	638 (41%)	**F** 5.617 **R** 4.934	0 (0%)
F63-R63	104 (100%)	**F** 0.056 **R** 0.069	96 (92%)	142 (99%)	**F** 0.28 **R** 0.794	104 (73%)	471 (94%)	**F** 0.663 **R** 2.479	170 (36%)	15 (100%)	**F** 0.52 **R** 2.36	6 (40%)	815 (53%)	**F** 5.542 **R** 4.22	0 (0%)

F, forward primer; R, reverse primer; SDB, sub-database. WS is the sum of Weighted score is the sum of score for the forward and reverse primer for each primer set

Table A.3: Two-way ANOVA test between primer sets for the same sample types

TukeyHSD	*P*adj value
CST-Household Influent-F3-R3 : CST-Household Influent-DF-DR	0.981
CST-Household Influent-F7-R7 : CST-Household Influent-DF-DR	0.973
CST-Household Influent-F7-R7 : CST-Household Influent-F3-R3	0.913
CST-Household Effluent-F3-R3 : CST-Household Effluent-DF-DR	0.974
CST-Household Effluent-F7-R7 : CST-Household Effluent-DF-DR	0.93
CST-Household Effluent-F7-R7 : CST-Household Effluent-F3-R3	0.83
CST-Household Sludge-F3-R3 : CST-Household Sludge-DF-DR	0.905
CST-Household Sludge-F7-R7 : CST-Household Sludge-DF-DR	0.995
CST-Household Sludge-F7-R7 : CST-Household Sludge-F3-R3	0.859
CST-Healthcare Effluent-F3-R3 : CST-Healthcare Effluent-DF-DR	NA
CST-Healthcare Effluent-F7-R7 : CST-Healthcare Effluent-DF-DR	NA
CST-Healthcare Effluent-F7-R7 : CST-Healthcare Effluent-F3-R3	NA
CST-Healthcare Sludge-F3-R3 : CST-Healthcare Sludge-DF-DR	0.827
CST-Healthcare Sludge-F7-R7 : CST-Healthcare Sludge-DF-DR	0.724
CST-Healthcare Sludge-F7-R7 : CST-Healthcare Sludge-F3-R3	0.978
SST-Household Effluent-F3-R3 : SST-Household Effluent-DF-DR	0.843
SST-Household Effluent-F7-R7 : SST-Household Effluent-DF-DR	0.986
SST-Household Effluent-F7-R7 : SST-Household Effluent-F3-R3	0.758
SST-Household Sludge-F3-R3 : SST-Household Sludge-DF-DR	0.908
SST-Household Sludge-F7-R7 : SST-Household Sludge-DF-DR	0.969
SST-Household Sludge-F7-R7 : SST-Household Sludge-F3-R3	0.791

CST, conventional septic tank; SST, solar septic tank; F, forward primer; R, reverse primer

Appendix B- Chapter 3

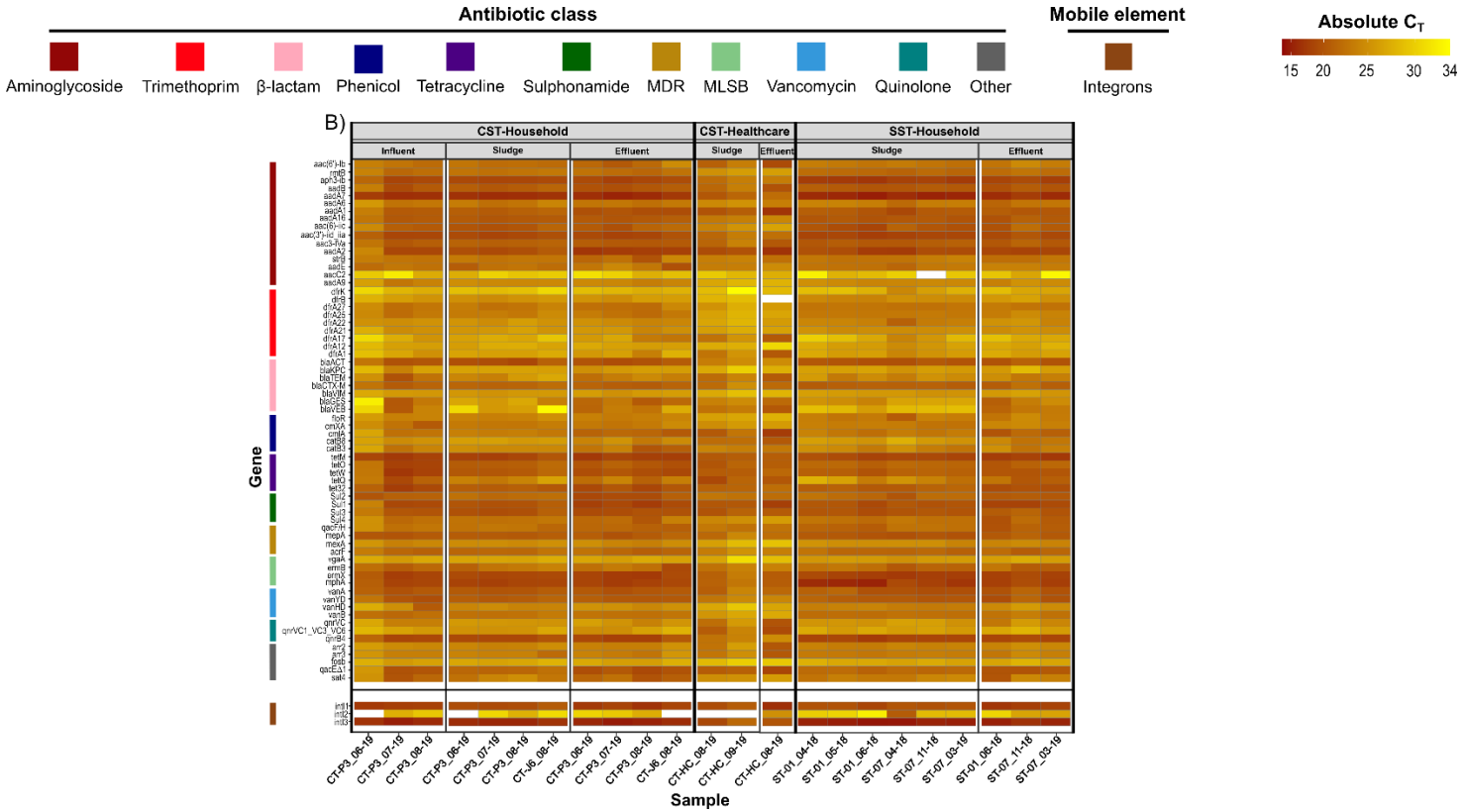

Figure B.1: AMR and integrase (*intI*1, *intI*2, *intI*3) abundance (inferred by the absolute Ct) heatmap of genes quantified on the HT-QPCR array for the individual wastewater samples (*n=23*) from three WWT reactors (CST-household, CST-healthcare, SST-household). CST denotes conventional septic tank; SST denotes solar septic tank. CTP3 and CTJ6 samples originated from two independent CST-Household reactor. CT-HC sample was from a CST-Healthcare tank. ST01 and ST07 are two independent SST-Household units. The sampling month and year is indicated by the format month_year (i.e., 06_19= June 2019). CST, Conventional septic tank; SST, Solar septic tank.

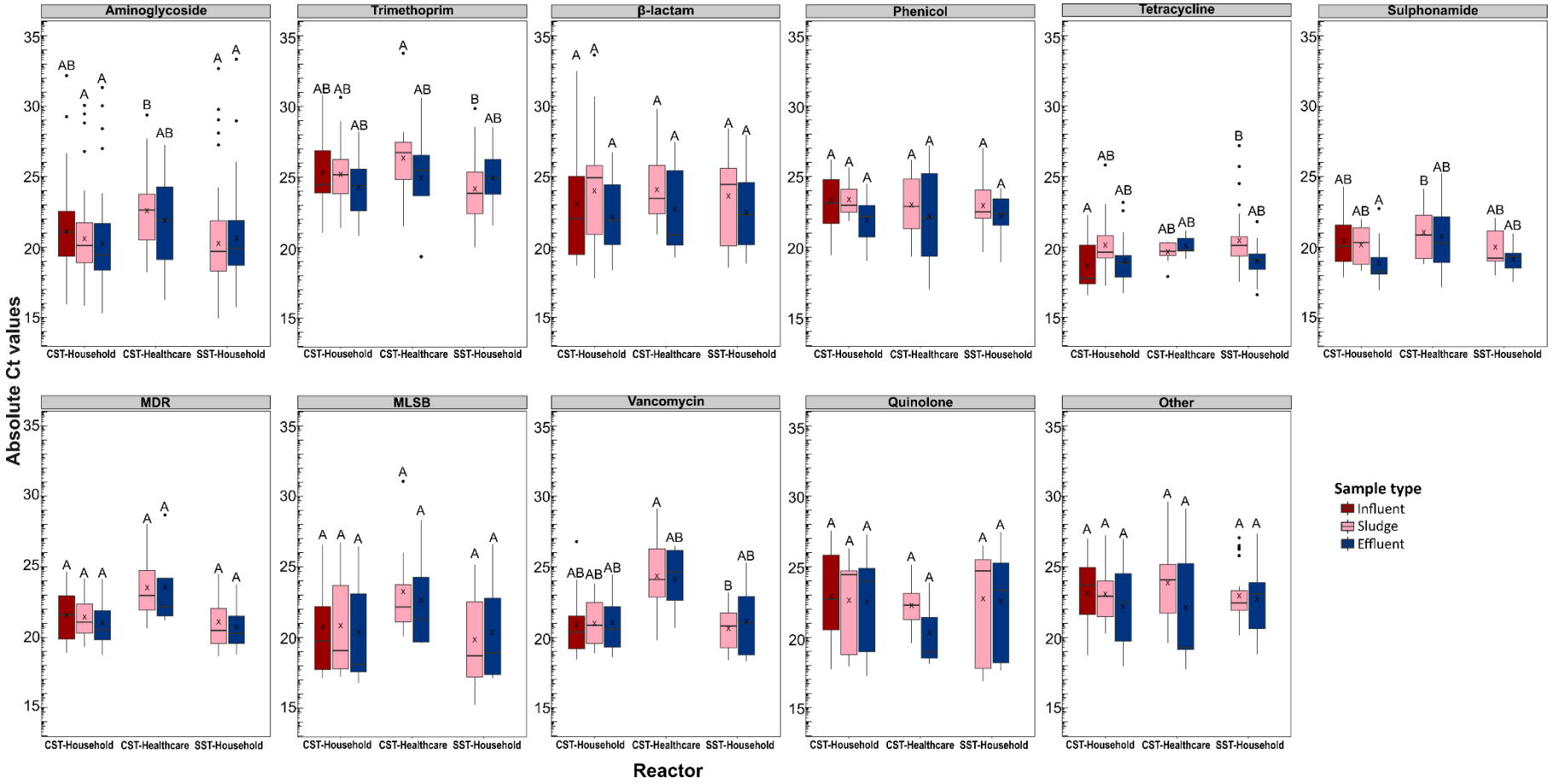

Figure B.2: AMR gene abundance (inferred by the absolute Ct) of each targeted antibiotic class quantified between samples (influent, sludge, effluent) and reactors (CST-Household, CST-Healthcare, SST-Household) on the HT-QPCR array. Black dots represent outliers. Boxplot sharing the same letter indicates no statistically significant difference at p-value >0.05. Statistically significant difference (p-value> 0.05) between samples and reactors was only observed for the antibiotic classes: Aminoglycoside, Trimethoprim and Vancomycin. CST denotes conventional septic tank; SST denotes solar septic tank.

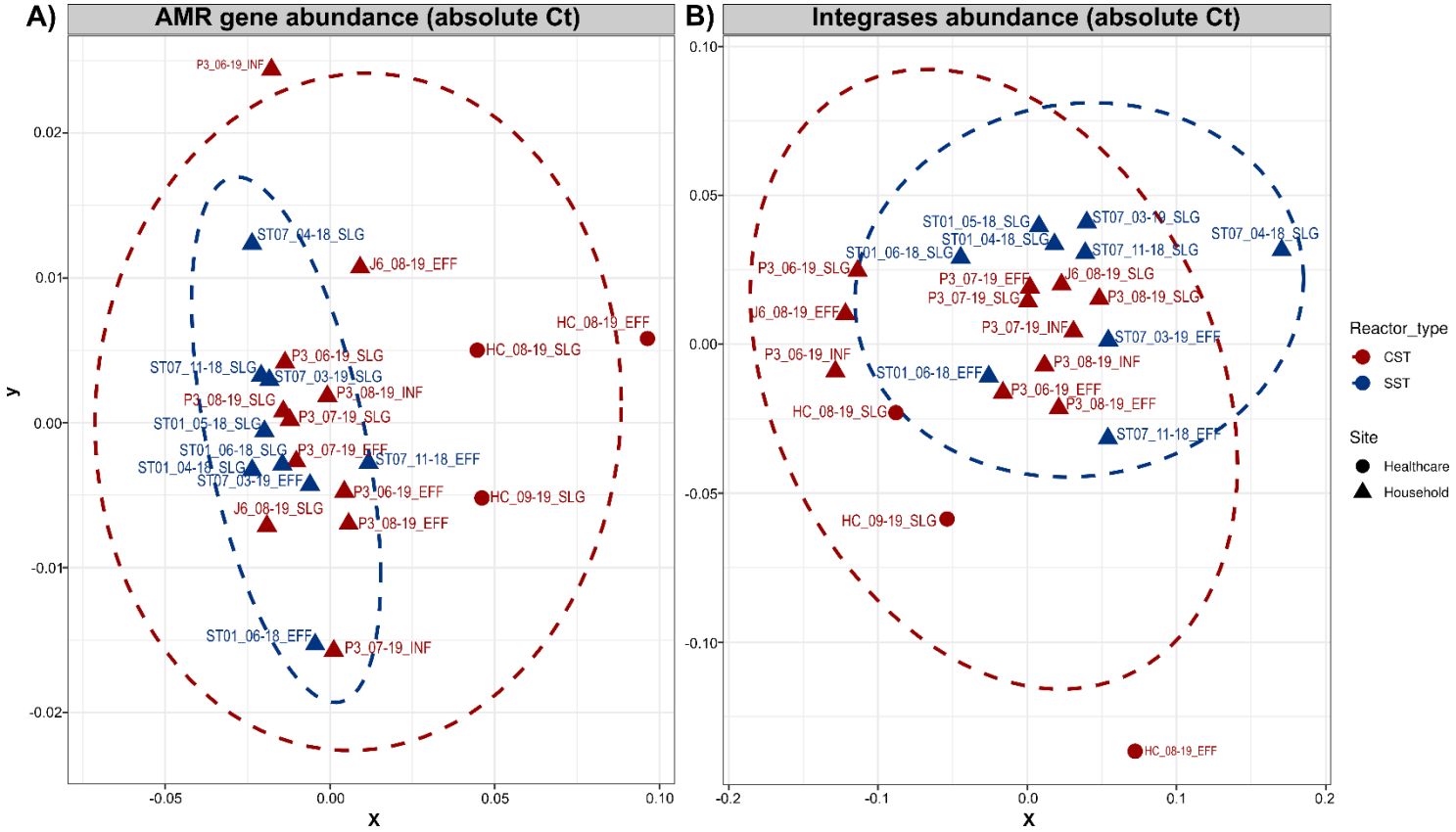

Figure B.3: NMDS plot indicating similarities/ differences in gene abundance (inferred by the absolute Ct) between samples (influent, sludge, effluent) and reactors (CST-household, CST-healthcare, SST-household). A) Gene abundance of the AMR genes from all targeted antibiotic class quantified on the HT-QPCR array. B) Gene abundance of the total integrases (*intI*1, *intI*2, *intI*3) quantified on the HT-QPCR array. Ellipses represent 95% confidence interval of standard error for a given group (CST vs SST). CST= Conventional septic tank, whilst SST= Solar septic tank.

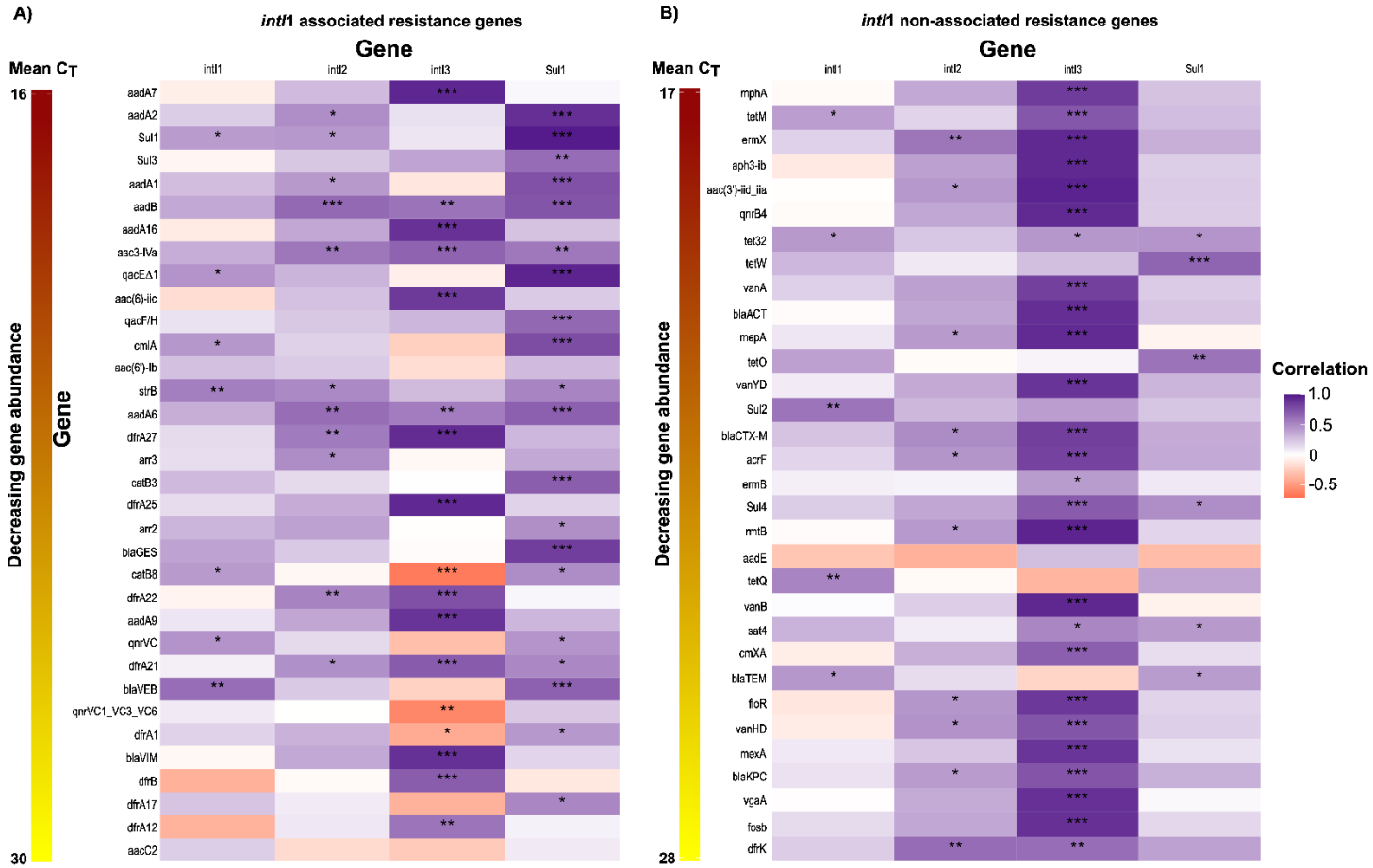

Figure B.4: Correlation analysis investigating link between AMR, integrases (*intI*1, *intI*2, *intI*3) and *Sul1* gene abundance (inferred by absolute Ct values) quantified on the HT-QPCR array. Quantified AMR genes were separated based on their association A) and non-association with genes mobile resistance integron known disseminating AMR genes between and within bacterial taxa. Quantified genes are ranked from most abundant (inferred by the mean Ct) to least abundance. Statistically significant correlation is indicated with star(s) (* p-value <0.05, ** p-value <0.01, *** p-value <0.001).

Table B.1: List of 384 primer sets and targeted genes quantified within pooled Thai wastewater samples on the HT-QPCR array

Antibiotic class	Assay	Gene	Forward primer (5'-3' direction)	Reverse primer (5'-3' direction)	Resistance mechanism
Housekeeping gene	AY1	*16S rRNA*	GGGTTGCGCTCGTTGC	ATGGYTGTCGTCAGCTCGTG	NA
Aminoglycoside	AY2	*aacC2*	ACGGCATTCTCGATTGCTTT	CCGAGCTTCACGTAAGCATTT	Drug inactivation
	AY4	*aacA_aphD*	AGAGCCTTGGGAAGATGAAGTTT	TTGATCCATACCATAGACTATCTCATCA	Drug inactivation
	AY6	*aac_6'_II*	CGACCCGACTCCGAACAA	GCACGAATCCTGCCTTCTCA	Drug inactivation
	AY7	*aphA3_1*	AAAAGCCCGAAGAGGAACTTG	CATCTTTCACAAAGATGTTGCTGTCT	Drug inactivation
	AY8	*aac_6'_Ib_1*	CGTCGCCGAGCAACTTG	CGGTACCTTGCCTCTCAAACC	Drug inactivation
	AY9	*aadA2_1*	ACGGCTCCGCAGTGGAT	GGCCACAGTAACCAACAAATCA	Drug inactivation
	AY10	*aadA_1*	GTTGTGCACGACGACATCATT	GGCTCGAAGATACCTGCAAGAA	Drug inactivation
	AY13	*aadD*	CCGACAACATTTCTACCATCCTT	ACCGAAGCGCTCGTCGTATA	Drug inactivation
	AY15	*aadA9_1*	CGCGGCAAGCCTATCTTG	CAAATCAGCGACCGCAGACT	Drug inactivation
	AY17	*aphA1_7*	TGAACAAGTCTGGAAAGAAATGCA	CCTATTAATTTCCCCTCGTCAAAAA	Drug inactivation
	AY21	*aadE*	TACCTTATTGCCCTTGGAAGAGTTA	GGAACTATGTCCCTTTTAATTCTACAATCT	Drug inactivation
	AY22	*str*	AATGAGTTTTGGAGTGTCTCAACGTA	AATCAAAACCCCTATTAAAGCCAAT	Drug inactivation
	AY23	*strA*	CCGGTGGCATTTGAGAAAAA	GTGGCTCAACCTGCGAAAAG	Drug inactivation
	AY24	*strB*	GCTCGGTCGTGAGAACAATCT	CAATTTCGGTCGCCTGGTAGT	Drug inactivation
	AY328	*aadA5_2*	ATCACGATCTTGCGATTTTGCT	CTGCGGATGGGCCTAGAAG	Drug inactivation
	AY330	*aph_2'_Ib*	TGAGCAGTATCATAAGTTGAGTGAAAAG	GACAGAACAATCAATCTCTATGGAATG	Drug inactivation
	AY331	*aadA2_3*	CAATGACATTCTTGCGGGTATC	GACCTACCAAGGCAACGCTATG	Drug inactivation
	AY385	*aac_6'_Iy*	GCCTCAATCCGCCACGATTA	ACGCGCTCTGTTTCCTCAAA	Drug inactivation
	AY386	*aac_6'_I1*	GGGAATTATCGGAATAGCTCTTGG	TTGGGCTGTTCTTCCTAGCTAA	Drug inactivation
	AY388	*aac3_IVa*	CCAACACGACGCTGCATC	GCTGTCGCCACAATGTCG	Drug inactivation
	AY389	*aph6_ia*	CGCTGGGAGCTGAAGAGG	AGCATCGTGCTGCTCTCC	Drug inactivation
	AY390	*spcN*	GCTATGTGCTGGTGGACTGG	GGAACCACTCGACGAACTCG	Drug inactivation
	AY391	*aac_3_ib*	CAGCGAGACGTTCATCGC	CACGCTTCAGGTGGCTAATC	Drug inactivation
	AY392	*aac_3_id_ie*	AGATAGTTATGCCCGCAACAAG	ACGCGCTGCGCCTATA	Drug inactivation
	AY393	*aac_3_iid_iia*	CGATGGTCGCGGTTGGTC	TCGGCGTAGTGCAATGCG	Drug inactivation
	AY394	*aac_3_xa_1*	GCAAGCGGTTCGTGACGTA	TCAGGTGCTCCTCGATCCAG	Drug inactivation
	AY396	*aac_6_ig*	GCGATGTTAGAAGCCTCAATTCG	CACACTTCGGCCTGTCGAA	Drug inactivation
	AY397	*aac_6_iic*	CAGTCTTTGGCTAATCCATCACAG	AACGAACCCGGCCTTCTC	Drug inactivation
	AY398	*aac_6_ij*	ATGCCTGTATCTGAATCCCTGATG	GGCAATCGCTTGTTGAGTATCTG	Drug inactivation
	AY399	*aac_6_im*	CGTGAGCATTATACAGAGCAATGG	CCATTTCCGTTCGTAGATATTGGC	Drug inactivation
	AY400	*aac_6_ir*	GCTATAACGATCAGCAGCAAGC	CGCGATGCATGGCATGAC	Drug inactivation
	AY401	*aac_6_is_iu_ix*	AAGCTTACTCTGGCCTGATCATG	TGCCTGAACGTCGATATTCAGG	Drug inactivation
	AY402	*aac_6_iv_ih*	TTGGCTTATACCGACACCCA	CCCGTTGCGATACCTGAAC	Drug inactivation
	AY403	*aac_6_iw*	TGCGTCAGTTACTTACACGAAC	CCTGATGCATTGCATGACTGA	Drug inactivation
	AY404	*aac_6_iz*	TGCGCCATGACTACGTGAAC	GACTGTCCGAAGCCAGTTCG	Drug inactivation
	AY406	*aac6_aph2*	CCAAGAGCAATAAGGGCATACCAA	GCCACACTATCATAACCACTACCG	Drug inactivation

	AY407	*aacA43*	CTTGGCCTACATTAGATTCAGCTC	GCTCTCAATCTTTGATAGGAGCAG	Drug inactivation
	AY408	*aadA10*	ACAGGCACTCAACGTCATCG	CGCGGAGAACTCTGCTTTGA	Drug inactivation
	AY409	*aadA16*	ACGGTGGCCTGAAGCC	GAATTGCAGTTCCCGTCTGG	Drug inactivation
	AY410	*aadA1_2*	TGTACGGCTCCGCAGTG	CACGGAATGATGTCGTCGTG	Drug inactivation
	AY411	*aadA6*	CCATCGAGCGTCATCTGGAA	CCCGTCTGGCCGGATAAC	Drug inactivation
	AY412	*aadA7*	CACTCCGCGCCTTGGA	TGTGGCGGGCTCGAAG	Drug inactivation
	AY413	*aadB*	CCTGCTTGGTGGGCAGAC	CGGCACGCAAGACCTCAA	Drug inactivation
	AY414	*ant4_ib*	GATGGCCGCTGACACATG	TCAACATTGCGCCATAGTGG	Drug inactivation
	AY415	*ant6_ia*	TCGCCATGAGCTGCTGA	CCTATCATACTCCGGATAGGCATA	Drug inactivation
	AY416	*ant6_ib*	AGAACATCCGACAGCACGTTC	CCAACCTTCCATGAAATCATTCGC	Drug inactivation
	AY417	*aph_viii*	TCGGTATCCCGGTTGTGAG	ACACGAGGTACGGGAATCC	Drug inactivation
	AY418	*aph_3''_ia*	TAACAGCGATCGCGTATTTCG	TCCGACTCGTCCAACATCAATA	Drug inactivation
	AY419	*aph3_ib*	AACAGGTTTGGGAGGCGATG	CGCAACAAGCCTCTCCTGAA	Drug inactivation
	AY420	*aph3_iii*	CAGAAGGCAATGTCATACCACTTG	GACAGCCGCTTAGCCGAA	Drug inactivation
	AY421	*aph3_viia*	CTCTCTCATGGAGATATGAGCGCTA	AATCCGGTTCAAGTCCCAACATG	Drug inactivation
	AY422	*aph3_via*	TCTCATGGCGATATCACGGATAG	TTTCCTCCGATGCATCCTCTC	Drug inactivation
	AY423	*aph4_ia*	CGCTCCCGATTCCGGAA	CACAGTTTGCCAGTGATACACA	Drug inactivation
	AY424	*aph4_ib*	GGGAACACCGTGCTCACC	GTTGGTCCCGTGCAGGTC	Drug inactivation
	AY426	*apmA*	GGCGCACATGCATTCATCA	CTATACTCCAGTCCCACCATTTGA	Drug inactivation
	AY427	*armA_1*	TCTTCGACGAATGAAAGAGTCG	GCTAATGGATTGAAGCCACAACC	Target alteration
	AY428	*aph9_ia*	GGTGCTGATATGAATGCCTTTGG	CATTGGGCGCATCAATAAATGG	Drug inactivation
	AY3	*aacC4*	CGGCGTGGGACACGAT	AGGGAACCTTTGCCATCAACT	Drug inactivation
	AY602	*armA_2*	TGCATCAAATATGGGGGTCT	TGAAGCCACAACCAAAATCT	Target alteration
	AY603	*rmtB*	GCTGTGATATCCACCAGGGA	AAGCTTAAAAATCAGCGCCA	Target alteration
	AY284	*dfrA1_1*	GGAATGGCCCTGATATTCCA	AGTCTTGCGTCCAACCAACAG	Target replacement
	AY285	*dfrA12*	CCTCTACCGAACCGTCACACA	GCGACAGCGTTGAAACAACTAC	Target replacement
	AY578	*dfrA10*	CTTCAACTATCACAGAGCACGAAG	TCTACCGGTACATACACATCAGC	Target replacement
	AY580	*dfrA15*	AGGCCGAAAGACTTTCGAGTC	TCACCTTCTGGCTCAATGTCG	Target replacement
	AY581	*dfra17*	CGGGAACGGCCCTGATATTCC	CGTGTTGCGACCGCATACTTTC	Target replacement
	AY582	*dfrA19*	GGAGCGAATCAAGGAGAAAGGAA	GCAATGCGTTGATCGGTATTCTC	Target replacement
	AY583	*dfra21*	TTGTTTCAACGCTGTCGCA	GGTTTCGGTTGAGACAAGCTC	Target replacement
	AY584	*dfrA22*	CAGCCGAACACGGCAAAG	CGGAGTGCGTGTACGTGA	Target replacement
Trimethoprim	AY585	*dfrA25*	TCAAACTGGACAGCGGCTA	GTCGATTGTCGACACATGCA	Target replacement
	AY586	*dfrA27*	GCCGCTCAGGATCGGTA	GTCGAGATATGTAGCGTGTCG	Target replacement
	AY588	*dfrA7*	GTAATCGGTAGTGGTCCTGA	ATCAGGACCACTACCGATTAC	Target replacement
	AY589	*dfrA8*	GGTCGCACCTGCATCGTTA	AGCGCCACCAATGACGTAG	Target replacement
	AY590	*dfrAB4*	CGGTTCGCATTCCCATCAAA	CGCAGTCATGGGATAAATCTGG	Target replacement
	AY591	*dfrB*	ACCAAGGCAGAAGTGAAGTCA	GGTGAGCCTCAGACTCGAC	Target replacement
	AY592	*dfrC*	GTCGCTCACGATAAACAAAGAGTC	CCCTTCATGGTGAAATGAAGCTTG	Target replacement
	AY593	*dfrG*	TCAATCGGAAGAGCCTTACCTGA	TGGGCAAATACCTCATTCCATTCC	Target replacement
	AY594	*dfrK*	TGCTGCGATGGATAAGAACAG	CTTCCAGGTAATGCTCTTCCG	Target replacement
β-lactam	AY97	*cfiA*	GCAGCGTTGCTGGACACA	GTTCGGGATAAACGTGGTGACT	Drug inactivation

AY101	*blaMOX_blaCMY*	CTATGTCAATGTGCCGAAGCA	GGCTTGTCCTCTTTCGAATAGC	Drug inactivation
AY102	*blaOCH*	GGCGACTTGCGCCGTAT	TTTTCTGCTCGGCCATGAG	Drug inactivation
AY103	*blaPAO*	CGCCGTACAACCGGTGAT	GAAGTAATGCGGTTCTCCTTTCA	Drug inactivation
AY105	*blaVEB*	CCCGATGCAAAGCGTTATG	GAAAGATTCCCTTTATCTATCTCAGACAA	Drug inactivation
AY107	*blaROB*	GCAAAGGCATGACGATTGC	CGCGCTGTTGTCGCTAAA	Drug inactivation
AY108	*blaOXY*	CGTTCAGGCGGCAGGTT	GCCGCGATATAAGATTTGAGAATT	Drug inactivation
AY109	*blaPSE*	TTGTGACCTATTCCCCTGTAATAGAA	TGCGAAGCACGCATCATC	Drug inactivation
AY111	*cphA_1*	GCGAGCTGCACAAGCTGAT	CGGCCCAGTCGCTCTTC	Drug inactivation
AY113	*bla_L1*	CACCGGGTTACCAGCTGAAG	GCGAAGCTGCGCTTGTAGTC	Drug inactivation
AY114	*cfxA*	TCATTCCTCGTTCAAGTTTTCAGA	TGCAGCACCAAGAGGAGATGT	Drug inactivation
AY115	*cepA*	AGTTGCGCAGAACAGTCCTCTT	TCGTATCTTGCCCGTCGATAAT	Drug inactivation
AY117	*ampC_blaDHA*	TGGCCGCAGCAGAAAGA	CCGTTTTATGCACCCAGGAA	Drug inactivation
AY125	*blaGES*	GCAATGTGCTCAACGTTCAAG	GTGCCTGAGTCAATTCTTTCAAAG	Drug inactivation
AY126	*blaSFO*	CCGCCGCCATCCAGTA	GGGCCGCCAAGATGCT	Drug inactivation
AY127	*blaTLA*	ACACTTTGCCATTGCTGTTTATGT	TGCAAATTTCGGCAATAATCTTT	Drug inactivation
AY128	*blaZ*	GGAGATAAAGTAACAAATCCAGTTAGATATGA	TGCTTAATTTTCCATTTGCGATAAG	Drug inactivation
AY129	*blaVIM*	GCACTTCTCGCGGAGATTG	CGACGGTGATGCGTACGTT	Drug inactivation
AY131	*pbp5*	GGCGAACTTCTAATTAATCCTATCCA	CGCCGATGACATTCTTCTTATCTT	Target alteration
AY132	*pbp*	CCGGTGCCATTGGTTTAGA	AAAATAGCCGCCCCAAGATT	Target alteration
AY133	*mecA*	GGTTACGGACAAGGTGAAATACTGAT	TGTCTTTTAATAAGTGAGGTGCGTTAATA	Target replacement
AY134	*blaCTX_M_5*	GCGATAACGTGGCGATGAAT	GTCGAGACGGAACGTTTCGT	Drug inactivation
AY138	*penA*	AGACGGTAACGTATAACTTTTTGAAAGA	GCGTGTAGCCGGCAATG	Target alteration
AY147	*blaCTX_M_8*	CGTCACGCTGTTGTTAGGAA	CGCTCATCAGCACGATAAAG	Drug inactivation
AY152	*blaNDM*	GGCCACACCAGTGACAATATCA	CAGGCAGCCACCAAAAGC	Drug inactivation
AY336	*blaACC*	CACACAGCTGATGGCTTATCTAAAA	AATAAACGCGATGGGTTCCA	Drug inactivation
AY338	*bla1*	GCAAGTTGAAGCGAAAGAAAAGA	TACCAGTATCAATCGCATATACACCTAA	Drug inactivation
AY339	*blaCMY_2*	AAAGCCTCAT GGGTGCATAAA	ATAGCTTTTGTTTGCCAGCATCA	Drug inactivation
AY430	*ampC_cefa*	CAGGATCTGATGTGGGAGAACTA	TCGGGAACCATTTGTTGGC	Drug inactivation
AY431	*blaSME*	GAGGAAGACTTTGATGGGAGGATTG	CGCTATATTGCAATGCAGCAGAAG	Drug inactivation
AY432	*blaCTX_M*	CGTACCGAGCCGACGTTAA	CAACCCAGGAAGCAGGCA	Drug inactivation
AY433	*blaFOX*	CCTACGGCTATTCGAAGGAAGATAAG	CCGGATTGGCCTGGAAGC	Drug inactivation
AY434	*imiR_2*	AGCCGGACTAGAGCTTCATG	GGCAGAACTCATCATCTGCAAA	Drug inactivation
AY435	*blaOXA51*	CGACCGAGTATGTACCTGCTTC	TCAAGTCCAATACGACGAGCTA	Drug inactivation
AY436	*blaOXY1*	AAAGGTGACCGCATTCGC	CCAGCGTCAGCTTGCG	Drug inactivation
AY437	*blaPER*	GCAAATGAAGCGCAGATGC	GACCACAGTACCAGCTGGTA	Drug inactivation
AY438	*blaSHV11*	TTGACCGCTGGGAAACGG	TCCGGTCTTATCGGCGATAAAC	Drug inactivation
AY439	*blaTEM*	CGCCGCATACACTATTCTCAG	GCTTCATTCAGCTCCGGTTC	Drug inactivation
AY440	*blaKPC*	GCCGCCAATTTGTTGCTGAA	GCCGGTCGTGTTTCCCTTT	Drug inactivation
AY441	*beta_ccra*	CACTGGCACGGCGATTGTA	CGGCAGCCAAACCACGATA	Drug inactivation
AY442	*bl1acc*	TGTTATCCGTGATTACCTGTCTGG	CTCAGCGAGCCAACTTCAAATA	Drug inactivation
AY443	*beta_B2*	GTAACGCCTACTGGAAGTCCA	CAGCTTCTCCTTGAGAATGCAG	Drug inactivation
AY444	*blaACT*	AAGCCGCTCAAGCTGGA	GCCATATCCTGCACGTTGG	Drug inactivation

	AY445	*blaB*	CGTGCCGGAGGTCTTGAATA	GGGATAGTAAACCTGAAACTCGGA	Drug inactivation
	AY446	*blaCARB*	TGATTTGAGGGATACGACAACTCC	CTGTAATACTCCGAGCACCAA	Drug inactivation
	AY447	*blaGOB*	CTTGGGCTTGAATGCTCAGGTA	TGTATGGTCGTAGTGAGCCTGA	Drug inactivation
	AY448	*blaHERA*	GGGCAACCGCATTCTGAC	GCATCTCCCACTTTATCGTCAC	Drug inactivation
	AY449	*blaIMI*	ACATCTACACCTGCAGCAGTAG	AATCGCTTGGTACGCTAGCA	Drug inactivation
	AY450	*blaIND*	CGCCTGTTAAACCCAACCTGTA	CGCTCTGTCATCATGAGAGTGG	Drug inactivation
	AY451	*blaLEN*	TGTTCGCCTGTGTGTTATCTCC	GCAGCACTTTAAAGGTGCTCAC	Drug inactivation
	AY452	*blaMIR*	CGGTCTGCCGTTACAGGTG	AAAGACCCGCGTCGTCATG	Drug inactivation
	AY453	*blaBEL_nonmobile*	ATGTCCATGGCACAGACTGTG	CCTGTCTTGTCACCCGTTACC	Drug inactivation
	AY454	*blaADC_nonmobile*	GGTATGGCTGTGGGTGTTATTCA	AGGCAAGGTTACCACTTGTATACG	Drug inactivation
	AY601	*blaOXA48*	TGTTTTTGGTGGCATCGAT	GTAAMRATGCTTGGTTCGC	Drug inactivation
Phenicol	AY29	*catB3*	GCACTCGATGCCTTCCAAAA	AGAGCCGATCCAAACGTCAT	Drug inactivation
	AY30	*catB8*	CACTCGACGCCTTCCAAAG	CCGAGCCTATCCAGACATCATT	Drug inactivation
	AY31	*ceoA*	ATCAACACGGACCAGGACAAG	GGAAAGTCCGCTCACGATGA	Drug efflux
	AY35	*cmlA_2*	TAGGAAGCATCGGAACGTTGAT	CAGACCGAGCACGACTGTTG	Drug efflux
	AY37	*cmxA*	GCGATCGCCATCCTCTGT	TCGACACGGAGCCTTGGT	Drug efflux
	AY38	*catA1*	GGGTGAGTTTCACCAGTTTTGATT	CACCTTGTCGCCTTGCGTATA	Drug inactivation
	AY41	*cmlA_4*	GCGCTCTTCGAGGATTCG	CCGCCCAAGCAGAAGTAGAC	Drug efflux
	AY555	*cat*	ATCGGCCAGACTGGATATCGA	CACAGCTCCAGTTGCAACAAC	Drug inactivation
	AY556	*cat_pC221*	AATGACCGTATGCTGCAAGAAG	TTTGCCTGCTATGGCATTCTG	Drug inactivation
	AY557	*catA2*	CCTGGAACCGCAGAGAACA	CGGAACTCCGGAAACTGATTAAC	Drug inactivation
	AY558	*catA3*	CTGATTGCTCAGGCCGTGAA	ATGAGTATGGGCAACTCAGTGC	Drug inactivation
	AY559	*catB2*	GCTACTATTCCGGCTATTACCATG	GGGCTCCTCGTTCATGTAGA	Drug inactivation
	AY560	*catB9*	CACCTTATGAAGTGGTCGGTTCA	GTCTGATGAACACAGAGACTGCA	Drug inactivation
	AY561	*catP*	CCTTTGGACTGAGTGTAAGTCTGA	TAAAGCCATCGAAGGTTGACCA	Drug inactivation
	AY561	*catP*	CCTTTGGACTGAGTGTAAGTCTGA	TAAAGCCATCGAAGGTTGACCA	Drug inactivation
	AY562	*catQ*	AGGTGCACTTACAGTATGACTGC	AACGTGGGAAGTTCTCGTCATAC	Drug inactivation
	AY563	*cmlV*	GCCCTCATCACCGTCTTCG	GGACGTTGGCGATGGAGAG	Drug inactivation
	AY564	*fexA*	TGGTGTGGCTGTTGCAATCTTA	CCAAGGTACAAAGCACCTTGGA	Drug efflux
	AY565	*optrA*	GGTGGATGAAGTCCGTACGG	AGGTTAGACCTCCAAGAGCCA	Target protection
	AY566	*floR*	AACCCGCCCTCTGGATCA	GCCGTCGAGAAGAAGACGAA	Drug efflux
	AY32	*floR_1*	ATTGTCTTCACGGTGTCCGTTA	CCGCGATGTCGTCGAACT	Drug efflux
	AY33	*yidY_mdtL*	GCAGTTGCATATCGCCTTCTC	CTTCCCGGCAAACAGCAT	Drug efflux
	AY34	*mdtL*	TGCTGATCGGGATTCTGATTG	CAGGCGCGACGAACATAAT	Drug efflux
Tetracycline	AY249	*tet36_1*	AGAATACTCAGCAGAGGTCAGTTCCT	TGGTAGGTCGATAACCCGAAAAT	Target protection
	AY250	*tet32*	CCATTACTTCGGACAACGGTAGA	CAATCTCTGTGAGGGCATTTAACA	Target protection
	AY254	*tetA_2*	CTCACCAGCCTGACCTCGAT	CACGTTGTTATAGAAGCCGCATAG	Drug efflux
	AY255	*tetA_B_1*	AGTGCGCTTTGGATGCTGTA	AGCCCCAGTAGCTCCTGTGA	Drug efflux
	AY258	*tetK*	CAGCAGTCATTGGAAAATTATCTGATTATA	CCTTGTACTAACCTACCAAAAATCAAAATA	Drug efflux
	AY259	*tetQ*	CGCCTCAGAAGTAAGTTCATACACTAAG	TCGTTCATGCGGATATTATCAGAAT	Target protection
	AY260	*tetH*	TTTGGGTCATCTTACCAGCATTAA	TTGCGCATTATCATCGACAGA	Drug efflux
	AY263	*tetW*	ATGAACATTCCCACCGTTATCTTT	ATATCGGCGGAGAGCTTATCC	Target protection

	AY264	*tetO_2*	CAACATTAACGGAAAGTTTATTGTATACCA	TTGACGCTCCAAATTCATTGTATC	Target protection
	AY267	*tetX*	AAATTTGTTACCGACACGGAAGTT	CATAGCTGAAAAAATCCAGGACAGTT	Drug inactivation
	AY268	*tetC_2*	ACTGGTAAGGTAAACGCCATTGTC	ATGCATAAACCAGCCATTGAGTAAG	Drug efflux
	AY269	*tetS*	TTAAGGACAAACTTTCTGACGACATC	TGTCTCCCATTGTTCTGGTTCA	Target protection
	AY273	*tetE*	TTGGCGCTGTATGCAATGAT	CGACGACCTATGCGATCTGA	Drug efflux
	AY274	*tetPB_1*	TGGGCGACAGTAGGCTTAGAA	TGACCCTACTGAAACATTAGAAATATACCT	Target protection
	AY276	*tetT*	CCATATAGAGGTTCCACCAAATCC	TGACCCTATTGGTAGTGGTTCTATTG	Target protection
	AY325	*tetR_1*	CAATCCATCGACAATCAC	GACAATCAGCTACTTCAC	Drug efflux
	AY367	*tetL_2*	ATGGTTGTAGTTGCGCGCTATAT	ATCGCTGGACCGACTCCTT	Drug efflux
	AY568	*tet39*	TATAGCGGGTCCGGTAATAGGTG	CCATAACGATCCTGCCCATAGATAAC	Drug efflux
	AY570	*tet38*	AAGCGACATTAGCCGGTTTAG	CTGCTCGTACTTAAGCCAAGG	Drug efflux
	AY571	*tetD*	AATTGCACTGCCTGCATTGC	GACAGATTGCCAGCAGCAGA	Drug efflux
	AY572	*tetG*	TCGCGTTCCTGCTTGCC	CCGCGAGCGACAAACCA	Drug efflux
	AY573	*tetJ*	CAGCGCCCATACGCCATTTA	CCTACTTCAGTAGTGTGCCAAGC	Drug efflux
	AY574	*tetM*	GGAGCGATTACAGAATTAGGAAGC	TCCATATGTCCTGGCGTGTC	Target protection
	AY575	*tetPA*	GGAAACCTTAGTTCAGTGACTTGG	CCCATTTAACCACGCACTGAA	Drug efflux
	AY576	*tet44*	CTCATGTAGATGCAGGAAAGACG	GTAACTGCTGCCTGAATTGTGA	Target protection
	AY577	*tetR*	CCGTCAATGCGCTGATGAC	GCCAATCCATCGACAATCACC	Drug efflux
Sulphonamide	AY241	*sul4*	TCAACGTCACTCCAGACAGC	TGGAAATAACGACGTCCACA	Target replacement
	AY245	*sul1_2*	GCCGATGAGATCAGACGTATTG	CGCATAGCGCTGGGTTTC	Target replacement
	AY247	*folA_1*	CGAGCAGTTCCTGCCAAAG	CCCAGTCATCCGGTTCATAATC	Drug inactivation
	AY361	*folP_2*	CAGGCTCGTAAATTGATAGCAGAAG	CTTTCCTTGCGAATCGCTTT	Target alteration
	AY365	*sul2_2*	TCATCTGCCAAACTCGTCGTTA	GTCAAAGAACGCCGCAATGT	Target replacement
	AY244	*sul3_1*	TCCGTTCAGCGAATTGGTGCAG	TTCGTTCACGCCTTACACCAGC	Target replacement
MDR	AY42	*pmrA*	TTTGCAGGTTTTGTTCCTAATGC	GCAGAGCCTGATTTCTCCTTTG	Drug efflux
	AY199	*acrB_1*	AGTCGGTGTTCGCCGTTAAC	CAAGGAAACGAACGCAATACC	Drug efflux
	AY201	*acrF*	GCGGCCAGGCACAAAA	TACGCTCTTCCCACGGTTTC	Drug efflux
	AY202	*adeA*	CAGTTCGAGCGCCTATTTCTG	CGCCCTGACCGACCAAT	Drug efflux
	AY206	*cmr*	CGGCATCGTCAGTGGAATT	CGGTTCCGAAAAAGATGGAA	Drug efflux
	AY207	*acrA_1*	GGTCTATCACCCTACGCGCTATC	GCGCGCACGAACATACC	Drug efflux
	AY208	*emrD_1*	CTCAGCAGTATGGTGGTAAGCATT	ACCAGGCGCCGAAGAAC	Drug efflux
	AY211	*mdtE*	CGTCGGCGCACTCGTT	TCCAGACGTTGTACGGTAACCA	Drug efflux
	AY215	*mexA*	AGGACAACGCTATGCAACGAA	CCGGAAAGGGCCGAAAT	Drug efflux
	AY219	*emrB_qacA_1*	CTTTTCTCTAACCGTACATTATCTACGATAAA	AGAACGTAGCGACTGATAAAATGCT	Drug efflux
	AY222	*mtrE*	CGATGTGTCGTTTTGGAAGGT	CCTGCACCATGATTCCTCAATA	Drug efflux
	AY224	*oprD*	ATGAAGTGGAGCGCCATTG	GGCCACGGCGAACTGA	Drug efflux
	AY226	*ttgA*	ACGCCAATGCCAAACGATT	GTCACGGCGCAGCTTGA	Drug efflux
	AY227	*mepA*	ATCGGTCGCTCTTCGTTCAC	ATAAATAGGATCGAGCTGCTGGAT	Drug efflux
	AY228	*mexE*	GGTCAGCACCGACAAGGTCTAC	AGCTCGACGTACTTGAGGAACAC	Drug efflux
	AY234	*cfr*	GCAAAATTCAGAGCAAGTTACGAA	AAAATGACTCCCAACCTGCTTTAT	Target alteration
	AY240	*mexB*	CTGGAGATCGACGACGAGAAG	GAAATCGTTGACGTAGCTGGAA	Drug efflux
	AY350	*mdsA*	CGGAGTCCATCGACCATTTG	ATCGTCGGCAAGGAGAATCA	Drug efflux

	AY353	*tolC_2*	CAGGCAGAGAACCTGATGCA	CGCAATTCCGGGTTGCT	Drug efflux
	AY355	*acrR_1*	GCGCTGGAGACACGACAAC	GCCTTGCTGCGAGAACAAA	Drug efflux
	AY360	*marR_3*	GCTGTTGATGACATTGCTCACA	CGGCGTACTGGTGAAGCTAAC	Drug efflux
	AY482	*oqxA*	GAGTCAACCTACCTCCACTATCA	GCTGCGAGTTATCCAGCAG	Drug efflux
	AY483	*adeI*	CAGTCTGGTTTGCAGTAACCA	CACTCCTACAACAACAGGCAA	Drug efflux
	AY484	*bexA_norM*	TCGGGCATCCCGTTTATGATC	GTAGGCTGCGCATAATACCCA	Drug efflux
	AY485	*mdtA*	ACAAGCCCAGGGCCAAC	CCTTAATGGTGCCTTCGGTTTC	Drug efflux
	AY486	*mdtH*	ATGCTGGCTGTACAAGTGATG	CACTCCAGCGGGCGATA	Drug efflux
	AY487	*cefa_qacelta*	TAGTTGGCGAAGTAATCGCAAC	TGCGATGCCATAACCGATTATG	Drug efflux
	AY488	*qacA_B*	AAGGGCCACTGCATTAGCTG	CCAGTCCAATCATGCCTGCA	Drug efflux
	AY489	*qacF_H*	CTGAAGTCTAGCCATGGATTCACTAG	CAAGCAATAGCTGCCACAAGC	Drug efflux
	AY490	*arsA*	CAGGTCAGCCGCATCAACC	GCCTGAAACACGGCAATTTCTTC	Drug efflux
	AY491	*cadC*	CGCTCTGTGTCAGGATGAAGAG	CTTTCTTATGTGCTAGGGCGATCA	Drug efflux
	AY492	*copA*	TGCACCTGACVGGSCAYAT	GVACTTCRCGGAACATRCC	Drug efflux
	AY493	*czcA*	GCCTTGTTCATCGGCGAAC	GGCAATGTCGCCTTCGTTC	Drug efflux
	AY494	*pbrT*	GATGCGCACTGGGCTTG	TCGGAATATGCGGAAATGCG	Drug efflux
	AY495	*pcoA*	TGGCGTATGGAGTTTCAATGC	GAATAATGCCGTGCCAGTGAA	Drug efflux
	AY497	*sugE*	CTTAGTTATTGCTGGTCTGCTGGA	GCATCGGGTTAGCGGACTC	Drug efflux
	AY498	*tcrB*	GTGCCGGAACTCAAGTAGCA	GCACCGACTGCTGGACTTAA	Drug efflux
	AY499	*terW*	TCAAAGAGCTACGCGAGTCATA	CCTTCCCTGTGGACTCACC	Drug efflux
	AY212	*mdtG_1*	TGGCACAAAATATCTGGCAGTT	TTGTGTGGCGATAAGAGCATTAG	Drug efflux
	AY44	*ermD_K*	GAGCCGCAAGCCCCTTT	GTGTTTCATTTGACGCGGAGTAA	Target alteration
	AY46	*ermF_1*	CAGCTTTGGTTGAACATTTACGAA	AAATTCCTAAAATCACAACCGACAA	Target alteration
	AY53	*lmrA_1*	TTCAGATGCAATGGCGTTTG	ATAATCGGGAACATAATGAGCATAACTAC	Drug efflux
	AY54	*erm36*	GGCGGACCGACTTGCAT	TCTGCGTTGACGACGGTTAC	Target alteration
	AY57	*ermT_1*	GTTCACTAGCACTATTTTTAATGACAGAAGT	GAAGGGTGTCTTTTTAATACAATTAACGA	Target alteration
	AY58	*msrC_1*	TCAGACCGGATCGGTTGTC	CCTATTTTTTGGAGTCTTCTCTCTAATGTT	Target protection
	AY61	*mphB*	CGCAGCGCTTGATCTTGTAG	TTACTGCATCCATACGCTGCTT	Drug inactivation
	AY66	*msrA_1*	CTGCTAACACAAGTACGATTCCAAAT	TCAAGTAAAGTTGTCTTACCTACACCATT	Target protection
	AY68	*ermX_1*	GCTCAGTGGTCCCCATGGT	ATCCCCCCGTCAACGTTT	Target alteration
	AY72	*vgaB_1*	TAAAAGAGAATAAGGCGCAAGGA	TGTTTAGTAGCATGTTGCATTTTCC	Target protection
MLSB	AY73	*pncA*	GCAATCGAGGCGGTGTTC	TTGCCGCAGCCAATTCA	Target alteration
	AY75	*lnuA_1*	TGACGCTCAACACACTCAAAAA	TTCATGCTTAAGTTCCATACGTGAA	Drug inactivation
	AY77	*vatE_2*	GACCGTCCTACCAGGCGTAA	TTGGATTGCCACCGACAATT	Drug inactivation
	AY83	*ermY*	TTGTCTTTGAAAGTGAAGCAACAGT	TAACGCTAGAGAACGATTTGTATTGAG	Target alteration
	AY90	*ermA_ermTR*	ACATTTTACCAAGGAACTTGTGGAA	GTGGCATGACATAAACCTTCATCA	Target alteration
	AY91	*oleC*	CCCGGAGTCGATGTTCGA	GCCGAAGACGTACACGAACAG	Drug efflux
	AY92	*carB*	GGAGTGAGGCTGACCGTAGAAG	ATCGGCGAAACGCACAAA	Drug efflux
	AY94	*pikR2*	TCGTGGGCCAGGTGAAGA	TTCCCCTTGCCGGTGAA	Target alteration
	AY528	*ereA*	GATAATTCTGCTGGCGCACA	GCAGGCGTGGTCACAAC	Drug inactivation
	AY530	*erm34*	AAAGCGGTTTACAAGCGTTTCG	GGGTGCTCTAGGGTTGTTTAGTG	Target alteration
	AY531	*erm35*	CCTTCAGTCAGAACCGGCAA	GCTGATTTGACAGTTGGTGGTG	Target alteration

	AY532	*ermA*	TCGTTGAGAAGGGATTTGCGA	TTGCATGCTTCAAAGCCTGTC	Target alteration
	AY533	*ermB_2*	GAACACTAGGGTTGTTCTTGCA	CTGGAACATCTGTGGTATGGC	Target alteration
	AY534	*ermD*	TTTCCGGACAGCATTTGATGC	TCCACTGCCAATACCTTACCG	Target alteration
	AY535	*ermF*	TCTGATGCCCGAAATGTTCAAG	TGAAGGACAATTGAACCTCCCA	Target alteration
	AY536	*lnuB*	GGATCGTTTACCAAAGGAGAAGG	AGCATAGCCTTCGTATCAGGAA	Drug inactivation
	AY537	*lnuC*	GGGTGTAGATGCTCTTCTTGGA	CTTTACCCGAAAGAGTTTCTACCG	Drug inactivation
	AY538	*mefA*	TAATTATCGCAGCAGCTGGTTC	GTTCCCAAACGGAGTATAAGAGTG	Target protection
	AY539	*mphA*	TCAGCGGGATGATCGACTG	GAGGGCGTAGAGGGCGTA	Drug inactivation
	AY540	*vat_B*	GCAATTGTTGCTGCGAATTCAG	GTGCTGACCAATCCCACCA	Drug inactivation
	AY541	*vga_A_LC_1*	GTGAAGATGTCTCGGGTACAATTG	GAAATACCAGGATTCCCATGCAC	Target protection
	AY543	*erm42*	TGTTGAGATTGGGCCTGGA	CTAAGGGTGGGTTCTCACTATCTA	Target alteration
	AY544	*ermE*	GTCACGCAGCTGGAGTTCG	CGGTGAAGCACAGCTCGAC	Target alteration
	AY545	*ermC_2*	CCCTTGAATTAGTACAGAGGTG	GCAAACTCGTATTCCACGA	Target alteration
	AY546	*ermX_2*	TGATGACGGCTCAGTGG	GTGCACCAGCGCCTGA	Target alteration
	AY547	*ermB_3*	TGAAAGCCATGCGTCTGAC	TTCAGCTGGCAGCTTAAGC	Target alteration
	AY548	*ermO*	GAGTACGCCCGCAAACG	GCGTTCGATCCGGAGGA	Target alteration
	AY549	*lnuF*	ATACCGGTCATTTCCACTTGGC	GCATCAGGCTGATGAGGTTCAA	Drug inactivation
	AY550	*lsaC*	AAACGGCGTGAAAGTATCAGG	TTGTGGTGATGTAACGGATGC	Target protection
	AY551	*mefB*	CCGATAGGCTTACTTGTTGCAG	AGTCCACTTGCGGTTTCATTG	Drug efflux
	AY552	*msrD*	GGCAAGCTAGGTGTTGAGC	ATTGCTCAACACCTAGCTTGC	Target protection
	AY553	*msrE*	CGGCAGATGGTCTGAGCTTAAA	CGCACTCTTCCTGCATAAAGGA	Target protection
	AY554	*vat_A*	ATGAACGGAGCGAATCATCGG	CCATACCGATCCAAACGTCATTTC	Drug inactivation
	AY43	*ermD_1*	GGACTCGGCAATGGTCAGAA	CCCCGAAACGCAATATAATGTT	Target alteration
	AY56	*ermB_1*	TAAAGGGCATTTAACGACGAAACT	TTTATACCTCTGTTTGTTAGGGAATTGAA	Target alteration
	AY65	*mefA_1*	CCGTAGCATTGGAACAGCTTTT	AAACGGAGTATAAGAGTGCTGCAA	Target protection
	AY71	*vgaA_1*	CGAGTATTGTGGAAAGCAGCTAGTT	CCCGTACCGTTAGAGCCGATA	Target protection
Vancomycin	AY156	*vanC_2*	CCTGCCACAATCGATCGTT	CGGCTTCATTCGGCTTGATA	Target alteration
	AY159	*vanB_1*	TTGTCGGCGAAGTGGATCA	AGCCTTTTTCCGGCTCGTT	Target alteration
	AY160	*vanTE*	GTGGTGCCAAGGAAGTTGCT	CGTAGCCACCGCAAAAAAAT	Target alteration
	AY161	*vanD*	CAGAGGAACATAATGTTTCGATAAAATCT	GCCGGATTTTGTGATTCCAA	Target alteration
	AY162	*vanHD*	GTGGCCGATTATACCGTCATG	CGCAGGTCATTCAGGCAAT	Target alteration
	AY163	*vanHB*	GAGGTTTCCGAGGCGACAA	CTCTCGGCGGCAGTCGTAT	Target alteration
	AY164	*vanRA_1*	CCCTTACTCCCACCGAGTTTT	TTCGTCGCCCCATATCTCAT	Target alteration
	AY165	*vanSA*	CGCGTCATGCTTTCAAAATTC	TCCGCAGAAAGCTCAATTTGTT	Target alteration
	AY167	*vanWB*	CGGACAAAGATACCCCCTATAAAG	AAATAGTAAATTGCTCATCTGGCACAT	Target alteration
	AY170	*vanXB*	AGGCACAAAATCGAAGATGCTT	GGGTATGGCTCATCAATCAACTT	Target alteration
	AY174	*vanRB*	GCCCTGTCGGATGACGAA	TTACATAGTCGTCTGCCTCTGCAT	Target alteration
	AY175	*vanRC*	TGCGGGAAAAACTGAACGA	CCCCCCATACGGTTTTGATTA	Target alteration
	AY176	*vanRC4*	AGTGCTTTGGCTTATCTCGAAAA	TCCGGCAGCATCACATCTAA	Target alteration
	AY177	*vanRD*	TTATAATGGCAAGGATGCACTAAAGT	CGTCTACATCCGGAAGCATGA	Target alteration
	AY181	*vanTG*	CGTGTAGCCGTTCCGTTCTT	CGGCATTACAGGTATATCTGGAAA	Target alteration
	AY182	*vanYB*	GGCTAAAGCGGAAGCAGAAA	GATATCCACAGCAAGACCAAGCT	Target alteration

	AY183	*vanYD_1*	AAGGCGATACCCTGACTGTCA	ATTGCCGGACGGAAGCA	Target alteration
	AY380	*vanSC_2*	ATCAACTGCGGGAGAAAAGTCT	TCCGCTGTTCCGCTTCTT	Target alteration
	AY381	*vanTC_2*	ACAGTTGCCGCTGGTGAAG	CGTGGCTGGTCGATCAAAA	Target alteration
	AY595	*vanA*	GGGCTGTGAGGTCGGTTG	TTCAGTACAATGCGGCCGTTA	Target alteration
	AY596	*vanC2*	TGACTGTCGGTGCTTGTGA	GATAGAGCAGCTGAGCTTGTTC	Target alteration
	AY597	*vanG*	TGTTTCGCAGAACCGTGTCAA	CCCTGCACTGTTCCATCTTCTC	Target alteration
	AY598	*vanXA*	TCGTTGGGACGCTAAATATGC	GGACGGTAACCGTCCCATA	Target alteration
	AY599	*vanSB*	GAAGATAAAGAGGGAAGCGTACTC	CCGAATTGTCAGCCCTTGATAA	Target alteration
Quinolone	AY95	*qnrA*	AGGATTTCTCACGCCAGGATT	CCGCTTTCAATGAAACTGCAA	Target protection
	AY96	*qnrB*	GCGACGTTCAGTGGTTCAGA	GCTGCTCGCCAGTCGAA	Target protection
	AY455	*norA*	ATCGCCGTTTGGTGGTACG	TCCACCAATCCCTGGTCCTAAA	Drug efflux
	AY456	*qepA*	GGGCATCGCGCTGTTC	GCGCATCGGTGAAGCC	Drug efflux
	AY457	*qnrB4*	TCACCACCCGCACCTG	GGATATCTAAATCGCCCAGTTCC	Target protection
	AY458	*qnrB_2*	CGACGTTCAGTGGTTCAGATCTC	GCCAAGCCGCTCCATGAG	Target protection
	AY459	*qnrD*	CGCTGGAATGGCACTGTGA	GCTCTCCATCCAACTTCACTCC	Target protection
	AY460	*qnrS_1*	CCACTTTGATGTCGCAGATCTTC	CCCTCTCCATATTGGCATAGGAAA	Target protection
	AY461	*qnrS2*	TCCCGAGCAAACTTTGCCAA	GGTGAGTCCCTATCCAGCGA	Target protection
	AY462	*qnrVC1_VC3_VC6*	CTCACATCAGGACTTGCAAGAA	ATGAAGCATCTCGAAGATCAGC	Target protection
	AY463	*qnrVC_2*	TTCCTTTAAACGGGCAAACCTC	CGATACCTGATTCATGAAGCTAGC	Target protection
Other	AY142	*ttgB*	TCGCCCTGGATGTACACCTT	ACCATTGCCGACATCAACAAC	Drug efflux
	AY186	*nisB_1*	GGGAGAGTTGCCGATGTTGTA	AGCCACTCGTTAAAGGGCAAT	Drug inactivation
	AY188	*nimE*	TGCGCCAAGATAGGGCATA	GTCGTGAATTCGGCAGGTTTA	Drug inactivation
	AY191	*merA*	GTGCCGTCCAAGATCATG	GGTGGAAGTCCAGTAGGGTGA	Drug efflux
	AY197	*crAss56*	CAGAAGTACAAACTCCTAAAAAACGTAGAG	GATGACCAATAAACAAGCCATTAGC	Unknown
	AY198	*crAss64*	TGTATAGATGCTGCTGCAACTGTACTC	CGTTGTTTTCATCTTTATCTTGTCCAT	Unknown
	AY204	*sat4*	GAATGGGCAAAGCATAAAAACTTG	CCGATTTTGAAACCACAATTATGATA	Drug inactivation
	AY218	*qacEΔ1_1*	TCGCAACATCCGCATTAAAA	ATGGATTTCAGAACCAGAGAAAGAAA	Drug efflux
	AY236	*qacEΔ1_3*	GTCGGTGTTGCTTATGCAGTCT	CAACCAGGCAATGGCTGTAA	Drug efflux
	AY465	*bacA*	ATCCGCGGCACCCTGA	CCTGCTTGATGGACTTGATGAAGA	Target alteration
	AY466	*mcr1*	CACATCGACGGCGTATTCTG	CAACGAGCATACCGACATCG	Target alteration
	AY467	*mcr2*	CGGCGTACTTTAAGCGTTATGATG	GCATTTGGCATACCATGCAGATAG	Target alteration
	AY468	*fosb*	CTTGCAGGCCTATGGATTGC	TCTGTTCTCAAGTGTGCCAGTA	Drug inactivation
	AY469	*fosX*	AGCTGGTTTGTGGATTTGCA	CCACACCGAGAGCTTTAATCCG	Drug inactivation
	AY470	*arr3*	GATCGTCTTCGAACGGTCCTG	TTTGGCGATTGGTGACTTGCT	Drug inactivation
	AY471	*arr2*	TTGGCGATTGGTGACTTGCTAA	ATCGTCTTCGAACGGTCCTG	Drug inactivation
	AY472	*fabK*	CAGGAGCAGGAAATCCAAGC	CCAGCTTCCATTCCTTCTGC	Target alteration
Integrons	AY289	*intI1_2*	CGAAGTCGAGGCATTTCTGTC	GCCTTCCAGAAAACCGAGGA	Integrase
	AY293	*intI1_1*	CGAACGAGTGGCGGAGGGTG	TACCCGAGAGCTTGGCACCCA	Integrase
	AY294	*intI2_2*	TGCTTTTCCCACCCTTACC	GACGGCTACCCTCTGTTATCTC	Integrase
	AY500	*intI3*	CAGGTGCTGGGCATGGA	CCTGGGCAGCATCACCA	Integrase
MGE	AY297	*Tp614*	GGAAATCAACGGCATCCAGTT	CATCCATGCGCTTTTGTCTCT	Transposase
	AY298	*IS613*	AGGTTCGGACTCAATGCAACA	TTCAGCACATACCGCCTTGAT	Transposase

AY299	*tnpA_1*	GCCGCACTGTCGATTTTTATC	GCGGGATCTGCCACTTCTT	Transposase
AY300	*tnpA_2*	CCGATCACGGAAAGCTCAAG	GGCTCGCATGACTTCGAATC	Transposase
AY301	*tnpA_3*	GGGCGGGTCGATTGAAA	GTGGGCGGGATCTGCTT	Transposase
AY302	*tnpA_4*	CATCATCGGACGGACAGAATT	GTCGGAGATGTGGGTGTAGAAAGT	Transposase
AY303	*tnpA_5*	GAAACCGATGCTACAATATCCAATTT	CAGCACCGTTTGCAGTGTAAG	Transposase
AY304	*tnpA_6*	TGCAGATGGTTTAACCTTGGATATTT	TCGGTTCATCAAACTGCTTCAC	Transposase
AY305	*tnpA_7*	AATTGATGCGGACGGCTTAA	TCACCAAACTGTTTATGGAGTCGTT	Transposase
AY306	*trfA*	ACGAAGAAATGGTTGTCCTGTTC	CGTCAGCTTGCGGTACTTCTC	Transposase
AY307	*orf37_IS26*	GCCGGGTTGTGCAAATAGAC	TGGCAATCTGTCGCTGCTG	Insertional
AY309	*ISPps*	CACACTGCAAAAACGCATCCT	TGTCTTTGGCGTCACAGTTCTC	Insertional
AY310	*IS1247_2*	TGGATCGACCGGTTCCAT	GCTGACCGAGCTGTCCATGT	Insertional
AY311	*ISAba3*	TCAGAGGCAGCGGTATACGA	GGTTGATTCAGTTAAAGTACGTAAAACTTT	Insertional
AY312	*ISEfm1*	AGGTGTCCATGACGTGAAAGTG	TCCTTTGTCCCCTAGGATATTGG	Insertional
AY313	*IS1111*	GTCTTAAGGTGGGCTGCGTG	CCCCGAATCTCATTGATCAGC	Insertional
AY314	*IS1133*	GCAGCGTCGGGTTGGA	ACGCGTTCGAACAACTGTAATG	Insertional
AY315	*Tn5*	TCAGAGGCAGCGGTATACGA	GGTTGATTCAGTTAAAGTACGTAAAACTTT	Insertional
AY316	*IncN_rep*	AGTTCACCACCTACTCGCTCCG	CAAGTTCTTCTGTTGGGATTCCG	Plasmid-inc
AY317	*IncN_oriT*	TTGGGCTTCATAGTACCC	GTGTGATAGCGTGATTTATGC	Plasmid-inc
AY318	*IncP_oriT*	CAGCCTCGCAGAGCAGGAT	CAGCCGGGCAGGATAGGTGAAGT	Plasmid-inc
AY319	*IncQ_oriT*	TTCGCGCTCGTTGTTCTTCGAGC	GCCGTTAGGCCAGTTTCTCG	Plasmid-inc
AY320	*IncW_trwAB*	AGCGTATGAAGCCCGTGAAGGG	AAAGATAAGCGGCAGGACAATAACG	Plasmid-inc
AY321	*pAMBL*	CAGGCTCTTAATGTGATA	TTATGCTCAATACTCGTG	Plasmid- rep
AY324	*pAKD1*	GGTAAGATTACCGATAAACT	GTTCGTGAAGAAGATGTA	Plasmid- rep
AY501	*cro*	AGATGTTATCGACCACTTCGGA	CCGCTTGGCGATAAGCG	MGE
AY502	*EAE_05855*	CCCATCACCGCTGAACTGG	TGGGCGCTGCCATCTAAAC	MGE
AY503	*IncHI2_smr0018*	ATAATGATTCACCGGGGTAG	CTTCAGGCTATCGTTTCG	MGE
AY504	*IncI1_repI1*	CGAAAGCCGGACGGCAGAA	TCGTCGTTCCGCCAAGTTCGT	MGE
AY505	*IncN_korA*	GGAACGTTTGTAYCTTGTATTG	ACTCACTATCTTCTGTTGATTG	MGE
AY506	*IS1247_1*	CGGCCGTCACTGACCAA	TCGGCAGGTTGGTGACG	MGE
AY508	*IS200_1*	CCAAATACCGAAGACAAGCGTTC	CCAAACTGCTCGTAAAGCATCAG	MGE
AY509	*IS200_2*	GCACACCCGATGGAACTGTAAA	TCGGCGGGATCTCCAGAAG	MGE
AY510	*IS21_ISAs29*	GGTCCGTCAGGCACAAGTC	GGGATCGTATCGGCAAGCC	MGE
AY511	*IS256*	CTTGCGCATCATTGGATGATGG	AAGAACGGCTCCAATTAAGCGA	MGE
AY512	*IS26_1*	ATGGATGAAACCTACGTGAAGGTC	CGGTACTTAATCTGTCGGTGTTCA	MGE
AY513	*IS3*	CGGTCTGAGCTTCGGGAA	AGAACTGTCACTCCGGTCTG	MGE
AY514	*IS5_IS1182*	TTCTCGAAGAATCGCCATGGC	GCTTTGGATCGCTCCAATCGA	MGE
AY515	*IS6_257*	ATATCGTGCCATTGATGCAGAG	ACCATTGCTACCTTCGTTGAAG	MGE
AY516	*IS6100*	CGCACCGGCTTGATCAGTA	CTGCCACGCTCAATACCGA	MGE
AY517	*IS630*	CCGCCACCAGTGTGATGG	TTGGCGCTGACTGGATGC	MGE
AY519	*ISCR1*	ATGGTTTCATGCGGGTT	CTGAGGGTGTGAGCGAG	MGE
AY520	*ISEcp1*	CATGCTCTGCGGTCACTTC	GACGCACCTTCTTGATGACC	MGE
AY521	*lncF_FIC*	GTGAACTGGCAGATGAGGAAGG	TTCTCCTCGTCGCCAAACTAGAT	MGE

	AY523	*Tn3*	GCTGAGGTGTTCAGCTACATCC	GCTGAGGTAGTCACAGGCATTC	MGE
	AY524	*Tn5403*	AAGCGAATGGCGCGAAC	CGCGCAGGGTAAACTGC	MGE
	AY526	*traN*	GCTTGGCGGTCAGCAATT	TTAGGAATAACAATCGCTACACCTTTA	Plasmid
	AY527	*trbC*	CGGYATWCCGSCSACRCTGCG	GCCACCTGYSBGCAGTCMCC	Plasmid
Taxonomic	AY473	*A_baumannii*	TCTTGGTGGTCACTTGAAGC	ACTCTTGTGGTTGTGGAGCA	Taxonomic
	AY474	*Bacteroidetes*	GGARCATGTGGTTTAATTCGATGAT	AGCTGACGACAACCATGCAG	Taxonomic
	AY475	*Campylobacter*	CTGCTTAACACAAGTTGAGTAGG	TTCCTTAGGTACCGTCAGAA	Taxonomic
	AY476	*Enterococci*	AGAAATTCCAAACGAACTTG	CAGTGCTCTACCTCCATCATT	Taxonomic
	AY477	*Firmicutes*	GGAGYATGTGGTTTAATTCGAAGCA	AGCTGACGACAACCATGCAC	Taxonomic
	AY478	*K_pneumoniae*	ACGGCCGAATATGACGAATTC	AGAGTGATCTGCTCATGAA	Taxonomic
	AY479	*P_aeruginosa*	AGCGTTCGTCCTGCACAAGT	TCCACCATGCTCAGGGAGAT	Taxonomic
	AY480	*Staphylococci*	CGCAACGTTCAATTTAATTTTGTTAA	TGGTCTTTCTGCATTCCTGGA	Taxonomic

Table B.2: List of 72 primer sets and targeted genes quantified within individual Thai wastewater samples (*n=23*) on the HT-QPCR array

Antibiotic class	Assay	Gene	Forward Primer (5'-3' direction)	Reverse Primer (5'-3' direction)	Resistance mechanism
Housekeeping gene	AY1	*16S rRNA*	GGGTTGCGCTCGTTGC	ATGGYTGTCGTCAGCTCGTG	NA
Aminoglycoside	AY2	*aacC2*	ACGGCATTCTCGATTGCTTT	CCGAGCTTCACGTAAGCATTT	Drug inactivation
	AY8	*aac(6')-Ib_1*	CGTCGCCGAGCAACTTG	CGGTACCTTGCCTCTCAAACC	Drug inactivation
	AY15	*aadA9_1*	CGCGGCAAGCCTATCTTG	CAAATCAGCGACCGCAGACT	Drug inactivation
	AY21	*aadE*	TACCTTATTGCCCTTGGAAGAGTTA	GGAACTATGTCCCTTTTAATTCTACAATCT	Drug inactivation
	AY24	*strB*	GCTCGGTCGTGAGAACAATCT	CAATTTCGGTCGCCTGGTAGT	Drug inactivation
	AY331	*aadA2_3*	CAATGACATTCTTGCGGGTATC	GACCTACCAAGGCAACGCTATG	Drug inactivation
	AY388	*aac3-IVa*	CCAACACGACGCTGCATC	GCTGTCGCCACAATGTCG	Drug inactivation
	AY393	*aac(3)-iid_iia*	CGATGGTCGCGGTTGGTC	TCGGCGTAGTGCAATGCG	Drug inactivation
	AY397	*aac(6)-iic*	CAGTCTTTGGCTAATCCATCACAG	AACGAACCCGGCCTTCTC	Drug inactivation
	AY409	*aadA16*	ACGGTGGCCTGAAGCC	GAATTGCAGTTCCCGTCTGG	Drug inactivation
	AY410	*aadA1_2*	TGTACGGCTCCGCAGTG	CACGGAATGATGTCGTCGTG	Drug inactivation
	AY411	*aadA6*	CCATCGAGCGTCATCTGGAA	CCCGTCTGGCCGGATAAC	Drug inactivation
	AY412	*aadA7*	CACTCCGCGCCTTGGA	TGTGGCGGGCTCGAAG	Drug inactivation
	AY413	*aadB*	CCTGCTTGGTGGGCAGAC	CGGCACGCAAGACCTCAA	Drug inactivation
	AY419	*aph3-ib*	AACAGGTTTGGGAGGCGATG	CGCAACAAGCCTCTCCTGAA	Drug inactivation
	AY603	*rmtB*	GCTGTGATATCCACCAGGGA	AAGCTTAAAAATCAGCGCCA	Target alteration
Trimethoprim	AY284	*dfrA1_1*	GGAATGGCCCTGATATTCCA	AGTCTTGCGTCCAACCAACAG	Target replacement
	AY285	*dfrA12*	CCTCTACCGAACCGTCACACA	GCGACAGCGTTGAAACAACTAC	Target replacement
	AY581	*dfra17*	CGGGAACGGCCCTGATATTCC	CGTGTTGCGACCGCATACTTTC	Target replacement
	AY583	*dfra21*	TTGTTTCAACGCTGTCGCA	GGTTTCGGTTGAGACAAGCTC	Target replacement
	AY584	*dfrA22*	CAGCCGAACACGGCAAAG	CGGAGTGCGTGTACGTGA	Target replacement
	AY585	*dfrA25*	TCAAACTGGACAGCGGCTA	GTCGATTGTCGACACATGCA	Target replacement
	AY586	*dfrA27*	GCCGCTCAGGATCGGTA	GTCGAGATATGTAGCGTGTCG	Target replacement
	AY589	*dfrA8*	GGTCGCACCTGCATCGTTA	AGCGCCACCAATGACGTAG	Target replacement
	AY591	*dfrB*	ACCAAGGCAGAAGTGAAGTCA	GGTGAGCCTCAGACTCGAC	Target replacement
	AY594	*dfrK*	TGCTGCGATGGATAAGAACAG	CTTCCAGGTAATGCTCTTCCG	Target replacement
β-Lactam	AY105	*blaVEB*	CCCGATGCAAAGCGTTATG	GAAAGATTCCCTTTATCTATCTCAGACAA	Drug inactivation
	AY125	*blaGES*	GCAATGTGCTCAACGTTCAAG	GTGCCTGAGTCAATTCTTTCAAAG	Drug inactivation
	AY129	*blaVIM*	GCACTTCTCGCGGAGATTG	CGACGGTGATGCGTACGTT	Drug inactivation
	AY432	*blaCTX-M*	CGTACCGAGCCGACGTTAA	CAACCCAGGAAGCAGGCA	Drug inactivation
	AY439	*blaTEM*	CGCCGCATACACTATTCTCAG	GCTTCATTCAGCTCCGGTTC	Drug inactivation
	AY440	*blaKPC*	GCCGCCAATTTGTTGCTGAA	GCCGGTCGTGTTTCCCTTT	Drug inactivation
	AY444	*blaACT*	AAGCCGCTCAAGCTGGA	GCCATATCCTGCACGTTGG	Drug inactivation
Phenicol	AY29	*catB3*	GCACTCGATGCCTTCCAAAA	AGAGCCGATCCAAACGTCAT	Drug inactivation
	AY30	*catB8*	CACTCGACGCCTTCCAAAG	CCGAGCCTATCCAGACATCATT	Drug inactivation
	AY35	*cmlA_2*	TAGGAAGCATCGGAACGTTGAT	CAGACCGAGCACGACTGTTG	Drug efflux

	AY37	*cmxA*	GCGATCGCCATCCTCTGT	TCGACACGGAGCCTTGGT	Drug efflux
	AY566	*floR*	AACCCGCCCTCTGGATCA	GCCGTCGAGAAGAAGACGAA	Drug efflux
Tetracycline	AY250	*tet32*	CCATTACTTCGGACAACGGTAGA	CAATCTCTGTGAGGGCATTTAACA	Target protection
	AY259	*tetQ*	CGCCTCAGAAGTAAGTTCATACACTAAG	TCGTTCATGCGGATATTATCAGAAT	Target protection
	AY263	*tetW*	ATGAACATTCCCACCGTTATCTTT	ATATCGGCGGAGAGCTTATCC	Target protection
	AY264	*tetO_2*	CAACATTAACGGAAAGTTTATTGTATACCA	TTGACGCTCCAAATTCATTGTATC	Target protection
	AY574	*tetM*	GGAGCGATTACAGAATTAGGAAGC	TCCATATGTCCTGGCGTGTC	Target protection
Sulphonamide	AY241	*sul4*	TCAACGTCACTCCAGACAGC	TGGAAATAACGACGTCCACA	Target replacement
	AY245	*sul1_2*	GCCGATGAGATCAGACGTATTG	CGCATAGCGCTGGGTTTC	Target replacement
	AY365	*sul2_2*	TCATCTGCCAAACTCGTCGTTA	GTCAAAGAACGCCGCAATGT	Target replacement
	AY244	*sul3_1*	TCCGTTCAGCGAATTGGTGCAG	TTCGTTCACGCCTTACACCAGC	Target replacement
MDR	AY201	*acrF*	GCGGCCAGGCACAAAA	TACGCTCTTCCCACGGTTTC	Drug efflux
	AY215	*mexA*	AGGACAACGCTATGCAACGAA	CCGGAAAGGGCCGAAAT	Drug efflux
	AY227	*mepA*	ATCGGTCGCTCTTCGTTCAC	ATAAATAGGATCGAGCTGCTGGAT	Drug efflux
	AY489	*qacF/H*	CTGAAGTCTAGCCATGGATTCACTAG	CAAGCAATAGCTGCCACAAGC	Drug efflux
MLSB	AY539	*mphA*	TCAGCGGGATGATCGACTG	GAGGGCGTAGAGGGCGTA	Drug inactivation
	AY546	*ermX_2*	TGATGACGGCTCAGTGG	GTGCACCAGCGCCTGA	Target alteration
	AY547	*ermB_3*	TGAAAGCCATGCGTCTGAC	TTCAGCTGGCAGCTTAAGC	Target alteration
	AY71	*vgaA_1*	CGAGTATTGTGGAAAGCAGCTAGTT	CCCGTACCGTTAGAGCCGATA	Target protection
Vancomycin	AY159	*vanB_1*	TTGTCGGCGAAGTGGATCA	AGCCTTTTTCCGGCTCGTT	Target alteration
	AY162	*vanHD*	GTGGCCGATTATACCGTCATG	CGCAGGTCATTCAGGCAAT	Target alteration
	AY183	*vanYD_1*	AAGGCGATACCCTGACTGTCA	ATTGCCGGACGGAAGCA	Target alteration
	AY595	*vanA*	GGGCTGTGAGGTCGGTTG	TTCAGTACAATGCGGCCGTTA	Target alteration
Quinolone	AY457	*qnrB4*	TCACCACCCGCACCTG	GGATATCTAAATCGCCCAGTTCC	Target protection
	AY462	*qnrVC1_VC3_VC6*	CTCACATCAGGACTTGCAAGAA	ATGAAGCATCTCGAAGATCAGC	Target protection
	AY463	*qnrVC_2*	TTCCTTTAAACGGGCAAACCTC	CGATACCTGATTCATGAAGCTAGC	Target protection
Other	AY204	*sat4*	GAATGGGCAAAGCATAAAAACTTG	CCGATTTTGAAACCACAATTATGATA	Drug inactivation
	AY218	*qacEΔ1_1*	TCGCAACATCCGCATTAAAA	ATGGATTTCAGAACCAGAGAAAGAAA	Drug efflux
	AY468	*fosb*	CTTGCAGGCCTATGGATTGC	TCTGTTCTCAAGTGTGCCAGTA	Drug inactivation
	AY470	*arr3*	GATCGTCTTCGAACGGTCCTG	TTTGGCGATTGGTGACTTGCT	Drug inactivation
	AY471	*arr2*	TTGGCGATTGGTGACTTGCTAA	ATCGTCTTCGAACGGTCCTG	Drug inactivation
Integrons	AY289	*intI1_2*	CGAAGTCGAGGCATTTCTGTC	GCCTTCCAGAAAACCGAGGA	Integrase
	AY293	*intI1_1*	CGAACGAGTGGCGGAGGGTG	TACCCGAGAGCTTGGCACCCA	Integrase
	AY500	*intI3*	CAGGTGCTGGGCATGGA	CCTGGGCAGCATCACCA	Integrase
	AY294	*intI2_2*	TGCTTTTCCCACCCTTACC	GACGGCTACCCTCTGTTATCTC	Integrase

Table B.3: Gene abundance (absolute Ct) of the 72 genes (AMR and integrase) targeted for the 23 wastewater samples

Antibiotic class	Assay	Gene	CST-Household											CST-Healthcare			SST-Household								
			CTP3_06_19_INF	CTP3_06_19_SLG	CTP3_06_19_EFF	CTP3_07_19_ INF	CTP3_07_19_ SLG	CTP3_07_19_ EFF	CTP3_08_19_ INF	CTP3_08_19_ SLG	CTP3_08_19_ EFF	CTJ6_08_19_ SLG	CTJ6_08_19_ EFF	CTHC2_08_19_ SLG	CTHC2_08_19_ EFF	CTHC2_09_19_ SLG	SST01_04_18_ SLG	SST07_04_18_ SLG	SST01_05_18_ SLG	SST01_06_18_ SLG	SST01_06_18_ EFF	SST07_11_18_ SLG	SST07_11_18_ EFF	SST07_03_19_ SLG	SST07_03_19_ EFF
Housekeeping gene	**AY1**	***16S rRNA***	14.27	13.74	13.28	12.74	13.93	14.17	13.28	14.23	13.11	13.28	13.34	12.18	12.16	12.87	12.21	11.59	13.17	12.45	13.37	12.79	13.18	12.74	13.15
Aminoglycoside	**AY15**	***aadA9_1***	25.67	24.04	23.63	21.89	23.59	23.39	23	23.46	22.5	23.63	23.21	26.2	27.26	27.57	24.25	23.36	23.69	23.16	22.81	23.62	24.39	23.49	23.79
	AY2	***aacC2***	29.26	26.79	31.32	32.17	30.06	30.03	26.67	28.81	27	29.45	28.4	29.38	26.9	27.7	32.67	29.77	27.25	28.09	28.96	NA	26.06	29.03	33.33
	AY21	***aadE***	21.02	19.65	21.93	21.97	21.41	23.46	21.47	21.83	21.95	21.07	18.68	22.48	23.87	22.57	21.75	20.49	21.57	21.45	22.96	22.03	22.9	22.27	22.45
	AY24	***strB***	22.53	22.23	20.44	21.75	21.54	21.41	21.56	21.85	18.8	21.83	23.45	22.64	21.35	22.97	21.98	21.6	21.43	21.57	21.15	22.52	21.14	23	21.87
	AY331	***aadA2_3***	23.43	18.51	16.48	17.86	18.07	16.95	17.86	18.46	17.18	19.51	17.25	18.23	16.27	18.74	18.66	16.87	18.3	16.97	17.84	18.31	17.51	18.2	17.72
	AY388	***aac3-IVa***	21.8	20.2	19.68	19.15	19.9	19.36	19.79	20.02	19.24	20.26	19.42	20.37	19.24	23	19.16	19.29	19.69	19.24	19.17	19.73	20.38	20.04	19.29
	AY393	***aac_3_iid_iia***	20.04	18.13	18.06	17.74	17.66	17.49	17.85	17.98	17.73	18.55	18.82	20.59	21.39	22.48	17.63	17.49	17.84	17.11	18.05	17.8	19.17	17.82	17.99
	AY397	***aac_6_iic***	23.7	18.97	19.49	19.83	18.67	18.54	19.45	19.02	20.27	20.42	20.87	22.95	25.92	24.58	19.04	20.23	18.35	17.69	18.73	19.1	21.21	18.65	19.55
	AY409	***aadA16***	21.91	19.5	19.44	19.23	19.06	18.96	19.55	19.3	19.36	19.51	19.86	22.18	23.29	22.71	19.1	19.28	19.39	18.84	18.77	19.45	20.08	18.93	19.39
	AY410	***aadA1_2***	22.62	19.85	18.56	19.41	19.52	18.97	19.56	20.01	18.38	20.49	18.33	18.84	16.32	19.74	19.83	17.88	19.69	18.97	19.74	19.43	19.14	19.61	19.24
	AY411	***aadA6***	24.93	22.45	21.79	21.58	21.27	21.98	22.02	21.73	21.27	23.25	22.18	23.62	22.62	24.37	23.08	20.83	23.04	22.57	20.4	22.48	21.83	22.63	21.92
	AY412	***aadA7***	17.03	16.18	16.02	15.96	15.84	15.29	16.11	15.99	15.73	16.26	16.73	19.16	21.13	20.44	15.71	15.88	15.53	14.94	15.83	15.22	17.06	15.4	15.72
	AY413	***aadB***	22.6	19.77	18.76	18.12	19.2	18.45	19.44	19.61	19.19	20.34	19.73	20.52	18.74	22.67	19.88	19.12	19.42	18.98	18.6	19.56	19.7	19.32	18.9
	AY419	***aph3-ib***	20.56	17.98	18.14	17.59	17.74	17.25	18.02	17.83	17.79	18.07	18.28	21.04	22.4	22.62	17.05	17.4	16.79	16.47	17.65	17.52	18.23	16.87	17.15
	AY603	***rmtB***	22.94	21.69	21.6	20.56	21.35	20.96	21.15	21.73	20.49	22.11	21.63	24.18	25.44	25.4	21.09	20.23	20.83	20.66	21.02	20.91	21.78	20.37	20.96
	AY8	***aac_6'_Ib_1***	22.66	21.79	20.64	21.68	20.97	19.35	20.88	21.22	20.77	20.9	23.83	19.97	18.14	22.64	22.5	22.31	22.16	21.68	21.7	21.74	23.69	21.29	22.45
Trimethoprim	**AY284**	***dfrA1_1***	28.37	25.92	24.89	26.21	25.66	25.67	24.25	25.64	22.67	26.62	27.19	21.67	19.34	24.6	26.55	23.24	25.29	25.52	26.59	24.44	25.28	25.31	25.82
	AY285	***dfrA12***	27.84	25.4	26.48	25.66	26.71	25.71	25.46	26.83	25.34	25.9	25.5	26.67	30.6	27.02	26.67	23.79	25.38	25.25	26.4	24.72	25.65	25.86	27.29
	AY581	***dfra17***	30.79	25.78	25.03	26.84	26.71	26.23	24.49	26.11	22.28	28.95	21.79	21.5	19.38	24.6	29.84	23.02	28.37	26.39	28.52	24.78	26.38	27.88	26.25
	AY583	***dfra21***	26.87	24.69	24.86	23.75	24.84	24.87	24.23	25.71	23.11	24.5	24.09	25.46	25.78	26.77	24.38	23.88	23.57	23.91	23.79	23.89	24.38	24.1	23.84

	AY584	***dfrA22***	24.29	24.55	24.3	23.9	23.84	23.7	24.38	25.39	23.03	24.37	23.29	26.06	25.18	27.53	23.35	20.03	23.24	22.77	24.36	22.32	24.71	22.69	22.99
	AY585	***dfrA25***	23.56	22.33	22.4	21.03	22.52	21.55	22.12	22.02	20.84	23.74	22.11	25.76	26.08	27.41	22.06	22.06	22.11	21.97	22.46	22.24	23.78	21.96	23.04
	AY586	***dfrA27***	23.85	22.47	21.38	21.07	21.39	21.14	21.92	21.7	20.9	23.07	23.47	24.43	25.11	26.93	21.72	21.4	21.89	21.76	22.34	21.47	21.69	21.84	21.57
	AY591	***dfrB***	27.17	24.25	24.48	25.44	23.72	23.89	24.47	24.42	25.39	24.9	25.38	27.46	NA	28.19	23.47	24.18	23.09	22.96	24.93	23.96	25.83	23.8	24.15
	AY594	***dfrK***	30.51	28.39	28.2	27.56	28.05	27.65	26.84	28.53	26.82	30.63	27.26	28.14	27.91	33.75	28.24	23.4	28.55	27.17	28.25	25.54	26.2	27.15	26.1
β-Lactam	**AY105**	***blaVEB***	30.23	30.7	20.41	19.31	24.95	22.53	23.04	25.54	21.66	33.62	26.71	21.51	19.25	22.31	28.42	27.97	27.75	25.32	20.26	28.04	21.93	28.31	22.28
	AY125	***blaGES***	32.5	26.19	20.52	19.24	24.93	22.14	22.03	26.34	19.41	22.71	20.74	22.9	20.86	23.87	24.44	25.94	23.9	22.99	19.9	26.12	22.53	26.18	23.84
	AY129	***blaVIM***	26.44	24.83	25.35	25.02	24.41	24.23	24.81	24.9	25.13	25.03	25.39	26.44	27.47	28.44	25.55	24.96	25.22	24.9	24.98	24.3	25.8	24.51	24.56
	AY432	***blaCTX_M***	21.71	20.44	20.3	19.47	20.22	19.77	20.3	20.42	19.57	21.05	20.33	20.89	20.83	24.11	19.56	19.99	19.97	20.09	19.67	20.09	21.42	19.68	19.89
	AY439	***blaTEM***	24.18	23.34	21.18	18.7	22.73	22.41	21.55	25.14	23.67	26.32	22.64	20.91	19.44	23.09	25.19	22.38	22.18	23.84	23.99	24.4	22	23.74	22.7
	AY440	***blaKPC***	27.81	25.79	24.15	22.79	25.85	24.93	25.78	25.77	25.34	25.49	25.28	26.34	26.99	29.81	25.66	24.43	25.59	25.93	24.94	25.41	27.96	25.05	25.32
	AY444	***blaACT***	21.86	18.57	18.9	18.71	18.37	18.34	19.06	17.8	18.7	20.01	19.58	22.55	23.83	23.83	19.29	18.55	19	18.58	18.83	19.38	20.16	18.9	18.82
Phenicol	**AY29**	***catB3***	26.18	24.09	21.75	21.46	23.55	22.92	23.11	23.6	18.99	23.48	19.88	23.13	22.08	22.64	24.17	24.18	23.69	22.18	24.18	22.4	21.53	21.74	22.64
	AY30	***catB8***	25.84	25.49	23.09	23.77	25.65	24.49	24.41	24.93	22.08	25.28	23.77	20.97	19.34	21.14	25.59	27.04	25.06	24.44	23.5	25.05	21.55	24.43	21.98
	AY35	***cmlA_2***	26.11	22.77	20.2	21.31	22.4	20.77	21.74	22.65	20.51	22.05	19.43	19.32	16.98	21.83	23.01	22.47	21.99	21.14	18.93	21.79	20.86	22.45	19.95
	AY37	***cmxA***	23.92	22.26	22.14	21.59	21.97	21.71	19.41	21.87	22.25	22.51	23.73	23.97	25.22	25.1	22.77	22.46	22.46	22.56	23.34	21.64	23.64	21.55	22.82
	AY566	***floR***	25.17	22.98	22.79	22.8	22.73	22.21	23.02	22.95	22.58	24.2	23.16	25.67	27.19	26.18	22.65	19.63	21.71	22.2	21.72	23	23.8	22.54	22.37
Tetracycline	**AY250**	***tet32***	20.14	19.79	17.95	17.06	19.58	19.19	18.02	19.96	17.75	19.28	19.19	20.14	21.16	20.35	19.75	19.45	20.15	19.34	18.32	20.25	18.75	20.45	18.52
	AY259	***tetQ***	22.27	22.9	21.07	17.91	21.99	23.16	20.11	23.05	20	25.82	22.57	17.91	19.15	20.35	27.19	24.5	25.71	22.37	21.8	22.99	19.17	21.27	19.87
	AY263	***tetW***	21.99	19.22	18.41	16.59	19.2	19.15	17.42	19.45	18.37	20.77	17.9	19.63	20.61	19.68	19.93	20.1	19.88	18.79	19.12	20.27	19.03	20	19.09
	AY264	***tetO_2***	21.83	19.57	19.12	17.32	19.68	20.2	17.57	19.85	19.14	20.9	18.58	19.02	19.73	20.36	20.75	20.86	20.73	19.84	19.11	20.55	20.1	20.34	20.65
	AY574	***tetM***	17.78	17.42	16.73	16.71	17.26	17.71	17.43	17.81	17.24	19.07	17.16	19.32	19.79	19.73	18.76	17.93	18.27	17.51	17.01	18	17.48	17.63	16.61
Sulfonamide	**AY241**	***sul4***	24.29	21.58	20.67	20.69	21.37	20.97	21.35	21.52	19.06	21.99	22.74	23.3	25.26	24.16	21.78	22.09	21.43	20.82	18.65	21.34	20.97	21.39	19.48
	AY244	***sul3_1***	22.22	18.63	18.14	19.09	18.33	18.28	18.91	18.56	18.83	20	18.71	18.8	19.49	20.18	19.12	18.89	19.19	18.26	19.56	19.21	20.41	18.92	18.92
	AY245	***sul1_2***	22.98	18.96	17.23	17.88	18.63	18.01	18.1	18.86	16.97	19.88	18.18	19.15	17.19	19.21	19.23	19.21	19.14	18	17.59	18.74	17.52	19.03	18.18
	AY365	***sul2_2***	19.01	21.44	17.92	20.07	20.64	18.24	20.1	21.02	18.21	21.29	19.95	21.92	21.11	21.57	21.68	18.98	21.02	20.51	19.43	21.12	19.65	21.03	19.28
MDR	**AY201**	***acrF***	22.44	20.46	21.64	20.52	20.62	20.55	20.03	20.54	19.8	21.73	20.34	22.46	22.69	24.46	21.69	20.36	20.51	20.46	20.96	20.71	19.63	19.99	20.57
	AY215	***mexA***	24.66	23.78	24.14	22.81	23.95	23.16	23.39	23.63	22.65	24.2	23.96	25.58	28.67	28.02	24.49	24.1	24.06	24.01	23.14	24.08	23.75	23.06	23.09

	AY227	***mepA***	19.2	19.6	19.76	18.91	19.53	18.73	19.45	19.33	19.54	19.76	20.47	20.64	21.19	23.46	19.44	18.66	19.37	18.92	19.35	19.05	20.32	18.76	19.32
	AY489	***qacF_H***	24.22	21.55	20.56	21.31	21.94	20.69	21.88	21.77	20.02	20.54	19.83	22.07	21.63	21.46	20.36	21.58	20.16	19.6	18.77	21.58	20.24	21.36	19.72
MLSB	**AY539**	***mphA***	19.88	17.5	17.62	17.46	17.23	16.91	17.72	17.71	17.44	17.77	17.92	20.09	19.71	21.87	15.77	18.11	15.51	15.23	17.28	17.34	18.29	16.51	17.12
	AY546	***ermX_2***	19.47	18.14	17.63	17.12	17.77	16.93	17.68	17.91	16.77	18.11	18.18	20.05	19.6	22.45	18.14	16.76	17.57	17.22	17.3	17.2	18.23	17.21	17.4
	AY547	***ermB_3***	21.23	20	21.26	19.61	21.69	22.31	20.68	22.14	21.33	22.96	18	21.43	22.9	22.98	21.55	19.26	21.94	21.21	21.94	20.32	19.56	20.11	19.69
	AY71	***vgaA_1***	26.57	25.9	26.11	25.06	26.04	25.42	26.31	25.8	26.45	26.73	25.77	26.01	28.33	31.07	24.77	24.23	25.16	24.99	25.5	25.01	26.61	25.1	25.26
Vancomycin	**AY159**	***vanB_1***	21.38	21.41	21.71	20.63	21.36	21.41	21.34	21.54	21.34	22.29	21.77	24.63	26.48	26.31	21.1	21.57	21.22	21.07	21.6	21.36	22.64	21.23	21.47
	AY162	***vanHD***	26.79	23.26	24.44	24.09	23.83	24.27	19.42	23.05	23.39	23.12	24.33	26.26	26.05	29.13	22.64	22.38	22.99	23.17	23.71	22.21	25.28	22.39	23.83
	AY183	***vanYD_1***	21.84	19.78	19.86	18.9	19.28	19.21	18.44	19.61	19.2	20.35	19.38	21.8	23.28	23.57	20.37	19.4	19.7	19.32	18.29	19.01	20.51	19.1	18.73
	AY595	***vanA***	20.13	19.39	19.35	18.67	18.87	18.6	19.3	19	19.07	19.66	19.45	19.78	20.67	23.21	20.04	18.7	20.52	18.63	18.7	18.39	20.35	18.52	18.75
Quinolone	**AY457**	***qnrB4***	20.53	18.04	17.99	17.72	17.95	17.26	17.91	17.99	17.24	19	19.3	21.61	23.88	22.94	17.64	17.71	17.89	16.87	17.71	17.76	18.2	17.74	17.64
	AY462	***qnrVC1_VC3_VC6***	27.57	24.42	23.98	25.82	24.33	23.92	24.91	24.46	25.63	26.29	27.28	19.59	18.12	23.18	26.5	25.27	25.15	26.4	26.2	25.72	27.45	25.04	25.26
	AY463	***qnrVC_2***	26.26	24.99	25.62	22.71	24.87	24.45	22.67	24.66	22.87	24.45	24.63	21.14	18.93	25.11	24.64	25.73	24.79	24.42	21.97	25.58	25.2	24.33	23.33
Other	**AY204**	***sat4***	24.35	21.47	20.36	18.87	23.17	21.41	20.58	24.01	18.8	21.88	19.88	23.99	25.24	24.33	21.66	22.07	22.66	22.14	18.99	23.36	23.51	23.63	22.69
	AY218	***qacEΔ1_1***	24.87	20.52	18.4	18.73	20.29	19.32	19.01	20.51	17.96	21.42	18.82	19.58	17.73	20.05	20.89	20.88	20.97	20.14	18.87	20.46	18.81	20.76	19.34
	AY468	***fosb***	27	26.22	26.38	25.06	25.85	25.44	25.77	26.22	25.38	27.23	26.97	27.87	29.12	29.64	27.06	25.8	26.25	26.37	26.03	26.51	27.35	25.79	26.03
	AY470	***arr3***	24.62	23.07	22.97	22.68	22.78	22.36	22.69	22.88	21.75	20.95	24.75	21.84	19.29	24.15	22.28	22.4	22.19	21.86	21.87	22.6	23.67	22.46	22.69
	AY471	***arr2***	25.33	23.3	22.91	23.73	22.79	22.61	23.44	22.93	22.58	24.01	24.43	21.66	19.13	25.46	22.99	23.22	22.4	22.35	23.07	22.89	24.08	23.06	23.42
Integrons	**AY289**	***intI1_2***	16.83	18.29	16.72	17.22	18.08	17.51	17.34	18.28	16.23	19.55	17.61	18.29	15.99	19.58	18.78	19.01	18.76	17.68	16.49	18.69	17.09	18.93	17.49
	AY500	***intI3***	16.29	15.78	15.48	14.99	15.44	14.77	15.64	15.44	15.18	16.38	15.82	17.9	18.99	20.63	15.2	14.59	14.95	14.53	15.06	15.22	16.21	15.01	15.3
	AY294	***intI2_2***	NA	NA	30.26	27.04	30.48	29.22	29.01	27.13	26.85	30.9	NA	NA	24.17	NA	29.6	19.57	30.22	32.63	30.39	28.04	26.04	28.01	25.94

Table B.4: Relative gene abundance (normalised to the *16S rRNA* gene) of the 72 genes (AMR and integrase) targeted for the 23 wastewater samples.

Antibiotic class	Assay	Gene	CST-Household											CST-Healthcare			SST-Household								
			CTP3_06_19_I	CTP3_06_19_S	CTP3_06_19_E	CTP3_07_19_I	CTP3_07_19_S	CTP3_07_19_E	CTP3_08_19_I	CTP3_08_19_S	CTP3_08_19_E	CTJ6_08_19_S	CTJ6_08_19_E	CTHC2_08_19_S	CTHC2_08_19_E	CTHC2_09_19_S	SST01_04_18_S	SST07_04_18_S	SST01_05_18_S	SST01_06_18_S	SST01_06_18_E	SST07_11_18_S	SST07_11_18_E	SST07_03_19_S	SST07_03_19_E
Aminoglycoside	AY15	*aadA9_1*	0.00038	0.00082	0.00078	0.00177	0.00129	0.00177	0.00119	0.00168	0.00151	0.00078	0.00108	0.00006	0.00003	0.00004	0.00024	0.00029	0.00069	0.00062	0.00146	0.00058	0.00042	0.00061	0.00063
	AY2	*aacC2*	0.00004	0.00012	0.00000	0.00000	0.00003	0.00003	0.00009	0.00008	0.00007	0.00002	0.00003	0.00002	0.00004	0.00003	0.00000	0.00001	0.00006	0.00002	0.00003	NA	0.00014	0.00002	0.00000
	AY21	*aadE*	0.00928	0.01665	0.00249	0.00167	0.00566	0.00163	0.00344	0.00516	0.00219	0.00454	0.02469	0.00080	0.00030	0.00121	0.00134	0.00210	0.00299	0.00196	0.00131	0.00168	0.00121	0.00145	0.00158
	AY24	*strB*	0.00326	0.00280	0.00701	0.00195	0.00516	0.00676	0.00327	0.00510	0.01949	0.00267	0.00092	0.00071	0.00171	0.00092	0.00116	0.00097	0.00330	0.00182	0.00455	0.00118	0.00403	0.00084	0.00238
	AY331	*aadA2_3*	0.00177	0.03681	0.10984	0.02882	0.05700	0.15145	0.04185	0.05337	0.05958	0.01341	0.06670	0.01513	0.05792	0.01734	0.01143	0.02571	0.02866	0.04365	0.04528	0.02216	0.04964	0.02354	0.04200
	AY388	*aac3-IVa*	0.00558	0.01136	0.01197	0.01176	0.01629	0.02973	0.01105	0.01854	0.01449	0.00795	0.01491	0.00346	0.00745	0.00095	0.00820	0.00484	0.01095	0.00910	0.01812	0.00825	0.00686	0.00693	0.01422
	AY393	*aac_3_iid_iia*	0.01840	0.04770	0.03706	0.03143	0.07558	0.10351	0.04229	0.07478	0.04062	0.02607	0.02244	0.00297	0.00168	0.00134	0.02362	0.01698	0.03926	0.03974	0.03910	0.03122	0.01579	0.03057	0.03498
	AY397	*aac_6_iic*	0.00154	0.02658	0.01362	0.00736	0.03751	0.05028	0.01389	0.03629	0.00709	0.00710	0.00545	0.00058	0.00008	0.00031	0.00886	0.00251	0.02764	0.02666	0.02449	0.01275	0.00385	0.01772	0.01191
	AY409	*aadA16*	0.00506	0.01854	0.01404	0.01114	0.02868	0.03812	0.01327	0.02976	0.01315	0.01344	0.01098	0.00099	0.00045	0.00113	0.00851	0.00491	0.01349	0.01196	0.02374	0.00993	0.00843	0.01389	0.01347
	AY410	*aadA1_2*	0.00319	0.01451	0.02578	0.00982	0.02093	0.03749	0.01299	0.01846	0.02618	0.00679	0.03155	0.00993	0.05582	0.00857	0.00516	0.01283	0.01095	0.01090	0.01218	0.01008	0.01611	0.00890	0.01468
	AY411	*aadA6*	0.00064	0.00240	0.00274	0.00222	0.00617	0.00472	0.00236	0.00558	0.00352	0.00100	0.00222	0.00037	0.00071	0.00035	0.00054	0.00166	0.00107	0.00093	0.00768	0.00131	0.00252	0.00120	0.00235
	AY412	*aadA7*	0.14748	0.18492	0.15075	0.10776	0.26766	0.47578	0.14052	0.30025	0.16420	0.12659	0.09590	0.00800	0.00201	0.00528	0.08923	0.05140	0.19585	0.17937	0.18359	0.18727	0.06795	0.16296	0.16889
	AY413	*aadB*	0.00320	0.01526	0.02271	0.02401	0.02592	0.05382	0.01416	0.02426	0.01491	0.00755	0.01202	0.00310	0.01050	0.00113	0.00490	0.00542	0.01318	0.01094	0.02673	0.00943	0.01092	0.01096	0.01864
	AY419	*aph3-ib*	0.01291	0.05378	0.03594	0.03504	0.07280	0.12818	0.03762	0.08344	0.03947	0.03629	0.03267	0.00221	0.00083	0.00119	0.03501	0.01791	0.08236	0.06245	0.05229	0.03786	0.03087	0.05876	0.06259
	AY603	*rmtB*	0.00260	0.00426	0.00322	0.00450	0.00609	0.00997	0.00459	0.00563	0.00626	0.00231	0.00337	0.00026	0.00011	0.00016	0.00217	0.00271	0.00509	0.00347	0.00500	0.00386	0.00262	0.00561	0.00464
	AY8	*aac_6'_Ib_1*	0.00305	0.00379	0.00615	0.00204	0.00761	0.02866	0.00515	0.00806	0.00499	0.00512	0.00073	0.00455	0.01585	0.00115	0.00081	0.00060	0.00198	0.00170	0.00312	0.00213	0.00070	0.00282	0.00160
Trimet.	AY284	*dfrA1_1*	0.00006	0.00022	0.00033	0.00009	0.00030	0.00038	0.00053	0.00037	0.00133	0.00010	0.00007	0.00140	0.00688	0.00030	0.00005	0.00031	0.00023	0.00013	0.00010	0.00031	0.00025	0.00018	0.00016
	AY285	*dfrA12*	0.00010	0.00035	0.00012	0.00013	0.00017	0.00041	0.00022	0.00016	0.00022	0.00016	0.00022	0.00005	0.00000	0.00006	0.00005	0.00021	0.00022	0.00014	0.00012	0.00026	0.00018	0.00012	0.00006

	AY581	*dfra17*	0.00002	0.00025	0.00030	0.00009	0.00015	0.00028	0.00042	0.00027	0.00175	0.00003	0.00288	0.00157	0.00675	0.00030	0.00001	0.00037	0.00003	0.00007	0.00004	0.00025	0.00013	0.00003	0.00013
	AY583	*dfra21*	0.00016	0.00052	0.00036	0.00050	0.00055	0.00064	0.00056	0.00039	0.00103	0.00045	0.00062	0.00011	0.00010	0.00007	0.00022	0.00020	0.00076	0.00038	0.00077	0.00052	0.00043	0.00043	0.00066
	AY584	*dfrA22*	0.00109	0.00056	0.00049	0.00045	0.00107	0.00151	0.00048	0.00076	0.00105	0.00046	0.00105	0.00008	0.00013	0.00004	0.00046	0.00305	0.00098	0.00081	0.00051	0.00137	0.00034	0.00104	0.00110
	AY585	*dfrA25*	0.00164	0.00259	0.00182	0.00322	0.00271	0.00609	0.00218	0.00461	0.00472	0.00072	0.00230	0.00008	0.00008	0.00004	0.00109	0.00070	0.00204	0.00137	0.00184	0.00143	0.00065	0.00169	0.00110
	AY586	*dfrA27*	0.00134	0.00234	0.00373	0.00314	0.00588	0.00896	0.00260	0.00582	0.00453	0.00114	0.00093	0.00023	0.00013	0.00006	0.00137	0.00111	0.00238	0.00161	0.00199	0.00246	0.00278	0.00192	0.00291
	AY591	*dfrB*	0.00014	0.00071	0.00043	0.00015	0.00113	0.00121	0.00047	0.00089	0.00021	0.00032	0.00024	0.00003	NA	0.00003	0.00041	0.00018	0.00104	0.00071	0.00035	0.00044	0.00016	0.00053	0.00050
	AY594	*dfrK*	0.00001	0.00004	0.00004	0.00004	0.00007	0.00009	0.00009	0.00005	0.00008	0.00001	0.00007	0.00002	0.00002	0.00000	0.00002	0.00028	0.00003	0.00004	0.00005	0.00015	0.00012	0.00005	0.00013
β-Lactam	AY105	*blaVEB*	0.00004	0.00005	0.00714	0.01058	0.00049	0.00323	0.00116	0.00041	0.00267	0.00000	0.00010	0.00157	0.00735	0.00144	0.00002	0.00001	0.00004	0.00014	0.00842	0.00003	0.00233	0.00003	0.00179
	AY125	*blaGES*	0.00001	0.00019	0.00665	0.01108	0.00049	0.00405	0.00234	0.00024	0.01270	0.00146	0.00595	0.00060	0.00241	0.00050	0.00022	0.00005	0.00059	0.00069	0.01082	0.00010	0.00156	0.00010	0.00061
	AY129	*blaVIM*	0.00027	0.00047	0.00024	0.00020	0.00070	0.00100	0.00034	0.00062	0.00024	0.00030	0.00029	0.00005	0.00003	0.00002	0.00010	0.00009	0.00025	0.00018	0.00033	0.00035	0.00017	0.00033	0.00038
	AY432	*blaCTX_M*	0.00585	0.00961	0.00786	0.00946	0.01282	0.02154	0.00785	0.01380	0.01173	0.00464	0.00793	0.00253	0.00249	0.00042	0.00624	0.00298	0.00900	0.00515	0.01298	0.00671	0.00334	0.00883	0.00945
	AY439	*blaTEM*	0.00112	0.00129	0.00429	0.01614	0.00224	0.00342	0.00327	0.00053	0.00074	0.00021	0.00174	0.00236	0.00644	0.00087	0.00013	0.00057	0.00200	0.00038	0.00064	0.00035	0.00227	0.00050	0.00135
	AY440	*blaKPC*	0.00009	0.00024	0.00054	0.00095	0.00028	0.00064	0.00018	0.00034	0.00026	0.00025	0.00026	0.00007	0.00004	0.00001	0.00011	0.00014	0.00019	0.00009	0.00033	0.00016	0.00004	0.00021	0.00022
	AY444	*blaACT*	0.00558	0.03508	0.02049	0.01605	0.04596	0.05635	0.01826	0.08428	0.02098	0.00948	0.01333	0.00077	0.00031	0.00051	0.00746	0.00807	0.01754	0.01464	0.02288	0.01063	0.00801	0.01531	0.01981
Phenicol	AY29	*catB3*	0.00027	0.00077	0.00289	0.00238	0.00131	0.00248	0.00111	0.00154	0.01723	0.00085	0.01079	0.00051	0.00104	0.00118	0.00026	0.00016	0.00068	0.00118	0.00056	0.00129	0.00308	0.00198	0.00142
	AY30	*catB8*	0.00033	0.00029	0.00114	0.00049	0.00030	0.00085	0.00046	0.00061	0.00199	0.00026	0.00074	0.00226	0.00689	0.00326	0.00009	0.00004	0.00026	0.00025	0.00090	0.00021	0.00303	0.00032	0.00220
	AY35	*cmlA_2*	0.00029	0.00192	0.00828	0.00263	0.00282	0.01047	0.00285	0.00299	0.00599	0.00239	0.01478	0.00711	0.03540	0.00201	0.00057	0.00053	0.00223	0.00245	0.02127	0.00196	0.00513	0.00124	0.00896
	AY37	*cmxA*	0.00140	0.00273	0.00225	0.00220	0.00381	0.00565	0.01432	0.00503	0.00184	0.00167	0.00077	0.00029	0.00012	0.00025	0.00069	0.00054	0.00161	0.00095	0.00100	0.00219	0.00072	0.00238	0.00125
	AY566	*floR*	0.00053	0.00170	0.00142	0.00094	0.00226	0.00422	0.00118	0.00237	0.00144	0.00053	0.00111	0.00009	0.00003	0.00011	0.00075	0.00379	0.00272	0.00117	0.00308	0.00085	0.00064	0.00114	0.00168
Tetracycline	AY250	*tet32*	0.01722	0.01509	0.03933	0.05018	0.01993	0.03173	0.03746	0.01886	0.04019	0.01572	0.01746	0.00403	0.00196	0.00561	0.00539	0.00430	0.00795	0.00844	0.03249	0.00569	0.02120	0.00491	0.02416
	AY259	*tetQ*	0.00392	0.00180	0.00455	0.02775	0.00377	0.00219	0.00881	0.00222	0.00845	0.00017	0.00167	0.01899	0.00789	0.00567	0.00003	0.00013	0.00018	0.00103	0.00289	0.00092	0.01579	0.00287	0.00944
	AY263	*tetW*	0.00473	0.02230	0.02872	0.06963	0.02596	0.03280	0.05714	0.02702	0.02618	0.00556	0.04258	0.00575	0.00287	0.00893	0.00478	0.00277	0.00958	0.01243	0.01861	0.00570	0.01742	0.00682	0.01630
	AY264	*tetO_2*	0.00534	0.01756	0.01757	0.04174	0.01860	0.01638	0.05100	0.02037	0.01535	0.00511	0.02649	0.00873	0.00528	0.00558	0.00270	0.00162	0.00532	0.00598	0.01864	0.00464	0.00827	0.00530	0.00551
	AY574	*tetM*	0.08778	0.07804	0.09217	0.06403	0.09952	0.08895	0.05640	0.08369	0.05722	0.01813	0.07107	0.00711	0.00707	0.00869	0.01071	0.01232	0.02921	0.03001	0.07995	0.02723	0.05129	0.03441	0.09067

Sulphonamide	AY241	*sul4*	0.00117	0.00439	0.00600	0.00404	0.00576	0.00927	0.00376	0.00648	0.01622	0.00240	0.00153	0.00046	0.00012	0.00040	0.00132	0.00069	0.00326	0.00310	0.02577	0.00290	0.00457	0.00271	0.01250
	AY244	*sul3_1*	0.00404	0.03404	0.03483	0.01232	0.04729	0.06036	0.02035	0.04954	0.01898	0.00964	0.02441	0.01024	0.00626	0.00633	0.00837	0.00653	0.01540	0.01789	0.01383	0.01172	0.00667	0.01403	0.01836
	AY245	*sul1_2*	0.00242	0.02671	0.06490	0.02833	0.03838	0.07211	0.03549	0.04039	0.06874	0.01034	0.03510	0.00802	0.03076	0.01240	0.00771	0.00507	0.01601	0.02159	0.05349	0.01629	0.04956	0.01316	0.03056
	AY365	*sul2_2*	0.03748	0.00480	0.04012	0.00622	0.00957	0.06081	0.00884	0.00907	0.02925	0.00388	0.01028	0.00118	0.00203	0.00242	0.00142	0.00597	0.00435	0.00380	0.01504	0.00313	0.01133	0.00329	0.01422
MDR	AY201	*acrF*	0.00349	0.00972	0.00324	0.00467	0.00967	0.01300	0.00953	0.01266	0.00970	0.00289	0.00784	0.00083	0.00068	0.00033	0.00145	0.00236	0.00620	0.00401	0.00530	0.00422	0.01153	0.00665	0.00588
	AY215	*mexA*	0.00079	0.00103	0.00056	0.00096	0.00099	0.00216	0.00094	0.00151	0.00136	0.00053	0.00065	0.00009	0.00001	0.00003	0.00020	0.00019	0.00057	0.00038	0.00123	0.00041	0.00068	0.00078	0.00108
	AY227	*mepA*	0.03295	0.01799	0.01153	0.01404	0.02121	0.04572	0.01443	0.02966	0.01163	0.01140	0.00715	0.00296	0.00199	0.00067	0.00667	0.00769	0.01393	0.01180	0.01685	0.01321	0.00725	0.01578	0.01406
	AY489	*qacF_H*	0.00103	0.00450	0.00644	0.00268	0.00388	0.01139	0.00260	0.00538	0.00832	0.00654	0.01118	0.00106	0.00143	0.00262	0.00354	0.00099	0.00800	0.00714	0.02367	0.00228	0.00754	0.00270	0.01052
MLSB	AY539	*mphA*	0.02050	0.07447	0.05007	0.03840	0.10260	0.15470	0.04636	0.08978	0.05042	0.04462	0.04291	0.00421	0.00538	0.00202	0.08502	0.01104	0.19757	0.14745	0.06677	0.04373	0.02898	0.07871	0.06456
	AY546	*ermX_2*	0.02795	0.04819	0.05018	0.04834	0.07153	0.16195	0.04768	0.07977	0.08091	0.03560	0.03516	0.00443	0.00599	0.00136	0.01665	0.02786	0.04775	0.03709	0.06632	0.04717	0.03034	0.04693	0.05311
	AY547	*ermB_3*	0.00804	0.01305	0.00403	0.00853	0.00464	0.00360	0.00593	0.00418	0.00341	0.00123	0.03981	0.00166	0.00059	0.00091	0.00155	0.00490	0.00229	0.00230	0.00265	0.00545	0.00121	0.00626	0.01079
	AY71	*vgaA_1*	0.00021	0.00022	0.00014	0.00020	0.00024	0.00043	0.00012	0.00038	0.00010	0.00010	0.00019	0.00007	0.00003	0.00000	0.00017	0.00016	0.00025	0.00018	0.00025	0.00022	0.00009	0.00024	0.00025
Vancomycin	AY159	*vanB_1*	0.00728	0.00501	0.00297	0.00433	0.00583	0.00760	0.00379	0.00635	0.00338	0.00196	0.00290	0.00019	0.00005	0.00009	0.00212	0.00103	0.00383	0.00268	0.00335	0.00265	0.00142	0.00288	0.00312
	AY162	*vanHD*	0.00018	0.00147	0.00045	0.00039	0.00105	0.00102	0.00118	0.00223	0.00085	0.00111	0.00050	0.00006	0.00007	0.00001	0.00073	0.00057	0.00111	0.00060	0.00078	0.00157	0.00023	0.00129	0.00065
	AY183	*vanYD_1*	0.00558	0.01553	0.01047	0.01407	0.02448	0.03204	0.02791	0.02444	0.01489	0.00754	0.01543	0.00130	0.00045	0.00061	0.00351	0.00444	0.01095	0.00943	0.03350	0.01357	0.00649	0.01313	0.02140
	AY595	*vanA*	0.01766	0.01998	0.01542	0.01710	0.03341	0.05204	0.01572	0.03700	0.01623	0.01208	0.01449	0.00529	0.00279	0.00078	0.00501	0.00727	0.01129	0.01403	0.02509	0.02076	0.00704	0.01913	0.02081
Quinolone	AY457	*qnrB4*	0.01321	0.05148	0.03844	0.03178	0.06155	0.12241	0.04135	0.07422	0.05700	0.01912	0.01611	0.00148	0.00030	0.00095	0.02373	0.01475	0.03898	0.04736	0.05067	0.03210	0.03093	0.03166	0.04470
	AY462	*qnrVC1_VC3_VC6*	0.00010	0.00062	0.00061	0.00012	0.00079	0.00119	0.00033	0.00085	0.00017	0.00013	0.00006	0.00592	0.01612	0.00079	0.00005	0.00008	0.00027	0.00007	0.00014	0.00027	0.00006	0.00024	0.00027
	AY463	*qnrVC_2*	0.00028	0.00046	0.00025	0.00104	0.00051	0.00093	0.00153	0.00083	0.00116	0.00064	0.00053	0.00201	0.00915	0.00022	0.00018	0.00007	0.00034	0.00029	0.00282	0.00015	0.00025	0.00035	0.00091
Other	AY204	*sat4*	0.00099	0.00473	0.00813	0.01427	0.00167	0.00673	0.00639	0.00114	0.01952	0.00263	0.01081	0.00028	0.00012	0.00036	0.00145	0.00071	0.00147	0.00122	0.02053	0.00066	0.00078	0.00057	0.00134
	AY218	*qacEΔ1*	0.00069	0.00912	0.02891	0.01574	0.01218	0.02946	0.01889	0.01290	0.03480	0.00360	0.02251	0.00594	0.02101	0.00694	0.00243	0.00161	0.00449	0.00483	0.02210	0.00493	0.02024	0.00400	0.01367
	AY468	*fosb*	0.00019	0.00018	0.00012	0.00020	0.00028	0.00042	0.00018	0.00026	0.00022	0.00009	0.00010	0.00002	0.00001	0.00001	0.00004	0.00006	0.00013	0.00006	0.00018	0.00008	0.00006	0.00013	0.00017
	AY470	*arr3*	0.00079	0.00163	0.00123	0.00105	0.00218	0.00374	0.00152	0.00250	0.00255	0.00494	0.00041	0.00127	0.00715	0.00042	0.00094	0.00058	0.00194	0.00155	0.00303	0.00115	0.00071	0.00124	0.00139
	AY471	*arr2*	0.00052	0.00133	0.00127	0.00051	0.00216	0.00294	0.00092	0.00247	0.00143	0.00060	0.00050	0.00142	0.00796	0.00018	0.00057	0.00032	0.00167	0.00105	0.00120	0.00095	0.00060	0.00083	0.00085

Integrons	AY289	*intI1_2*	0.17042	0.04269	0.09223	0.04499	0.05627	0.10161	0.06026	0.06033	0.11527	0.01303	0.05188	0.01453	0.07035	0.00963	0.01052	0.00587	0.02072	0.02654	0.11519	0.01686	0.06704	0.01385	0.04916
	AY500	*intI3*	0.24733	0.24494	0.22646	0.21096	0.36362	0.71902	0.19694	0.43811	0.24031	0.11639	0.18036	0.01968	0.00887	0.00478	0.12652	0.12660	0.29480	0.23882	0.31151	0.18609	0.12310	0.21670	0.22615
	AY294	*intI2_2*	NA	NA	0.00001	0.00006	0.00001	0.00003	0.00002	0.00015	0.00007	0.00003	NA	NA	0.00160	NA	0.00001	0.00401	0.00001	0.00000	0.00003	0.00003	0.00014	0.00003	0.00014

Table B.5: Spearman's rank correlation and p-value of all primer sets including the five *intI*1 primer sets and 16S rRNA used to quantify Thai wastewater samples (*n=23*)

Primer set	rho	p-value
F3-R3 : F7-R7	0.989	1.71E-06
F3-R3 : DF-DR	0.998	1.45E-06
F3-R3 : F4-R4 (array 289)	0.965	2.40E-06
F3-R3 : F10-R10 (array 293)	0.955	2.69E-06
F7-R7 : DF-DR	0.986	1.79E-06
F7-R7 : F4-R4 (array 289)	0.966	2.37E-06
F7-R7 : F10-R10 (array 293)	0.971	2.23E-06
DF-DR : F4-R4 (array 289)	0.969	2.28E-06
DF-DR : F10-R10 (array 293)	0.958	2.59E-06
F4-R4 (array 289) : F10-R10 (array 293)	0.980	1.97E-06
F1369-R1492 (*16S rRNA* TaqMan) : F1180-R1132 (*16S rRNA* SYBR Green; AY1)	0.993	1.59E-06

Table B.6: Two-way ANOVA test between primer sets for the same sample types (influent, sludge, effluent) across the reactors (CST-Household, CST-Healthcare, SS-Household) for all five *intI*1 primer sets employed.

Comparison	P- value
SST-Household Effluent-F10-R10 : SST-Household Effluent-DF-DR	0.82
SST-Household Effluent-F3-R3 : SST-Household Effluent-DF-DR	0.96
SST-Household Effluent-F4-R4 : SST-Household Effluent-DF-DR	0.99
SST-Household Effluent-F7-R7 : SST-Household Effluent-DF-DR	1.00
SST-Household Effluent-F3-R3 : SST-Household Effluent-F10-R10	0.46
SST-Household Effluent-F4-R4 : SST-Household Effluent-F10-R10	0.98
SST-Household Effluent-F7-R7 : SST-Household Effluent-F10-R10	0.69
SST-Household Effluent-F4-R4 : SST-Household Effluent-F3-R3	0.77
SST-Household Effluent-F7-R7 : SST-Household Effluent-F3-R3	0.99
SST-Household Effluent-F7-R7 : SST-Household Effluent-F4-R4	0.94
SST-Household Sludge-F10-R10 : SST-Household Sludge-DF-DR	0.11
SST-Household SludgeF3-R3 : SST-Household Sludge-DF-DR	0.42
SST-Household Sludge-F4-R4 : SST-Household Sludge-DF-DR	0.38
SST-Household Sludge-F7-R7 : SST-Household Sludge-DF-DR	1.000
SST-Household Sludge-F3-R3 : SST-Household Sludge-F10-R10	**0.001****
SST-Household Sludge-F4-R4 : SST-Household Sludge-F10-R10	0.95
SST-Household Sludge-F7-R7 : SST-Household Sludge-F10-R10	0.13
SST-Household Sludge-F4-R4 : SST-Household Sludge-F3-R3	**0.01***
SST-Household Sludge-F7-R7 : SST-Household Sludge-F3-R3	0.38
SST-Household Sludge-F7-R7 : SST-Household Sludge-F4-R4	0.41
CST-Household Influent-F10-R10 : CST-Household Influent-DF-DR	1.00
CST-Household Influent-F3-R3 : CST-Household Influent-DF-DR	1.00
CST-Household Influent-F4-R4 : CST-Household Influent-DF-DR	0.99
CST-Household Influent-F7-R7 : CST-Household Influent-DF-DR	1.00
CST-Household Influent-F3-R3 : CST-Household Influent-F10-R10	0.95
CST-Household Influent-F4-R4 : CST-Household Influent-F10-R10	1.000
CST-Household Influent-F7-R7 : CST-Household Influent-F10-R10	0.98
CST-Household Influent-F4-R4 : CST-Household Influent-F3-R3	0.94
CST-Household Influent-F7-R7 : CST-Household Influent-F3-R3	1.000
CST-Household Influent-F7-R7 : CST-Household Influent-F4-R4	0.97
CST-Household-Sludge-F10-R10 : CST-Household-Sludge-DF-DR	0.91
CST-Household-Sludge-F3-R3 : CST-Household-Sludge-DF-DR	0.95
CST-Household-Sludge-F4-R4 : CST-Household-Sludge-DF-DR	0.96
CST-Household-Sludge-F7-R7 : CST-Household-Sludge-DF-DR	1.000
CST-Household-Sludge-F3-R3 : CST-Household-Sludge-F10-R10	0.54
CST-Household-Sludge-F4-R4 : CST-Household-Sludge-F10-R10	1.00
CST-Household-Sludge-F7-R7 : CST-Household-Sludge-F10-R10	0.87
CST-Household-Sludge-F4-R4 : CST-Household-Sludge-F3-R3	0.64
CST-Household-Sludge-F7-R7 : CST-Household-Sludge-F3-R3	0.97
CST-Household-Sludge-F7-R7 : CST-Household-Sludge-F4-R4	0.93
CST-Household Effluent-F10-R10 : CST-Household Effluent-DF-DR	0.81
CST-Household Effluent-F3-R3 : CST-Household Effluent-DF-DR	0.99
CST-Household Effluent-F4-R4 : CST-Household Effluent-DF-DR	0.89
CST-Household Effluent-F7-R7 : CST-Household Effluent-DF-DR	1.000
CST-Household Effluent-F3-R3 : CST-Household Effluent-F10-R10	0.56
CST-Household Effluent-F4-R4 : CST-Household Effluent-F10-R10	1.00
CST-Household Effluent-F7-R7 : CST-Household Effluent-F10-R10	0.65
CST-Household Effluent-F4-R4 : CST-Household Effluent-F3-R3	0.66
CST-Household Effluent-F7-R7 : CST-Household Effluent-F3-R3	1.00
CST-Household Effluent-F7-R7 : CST-Household Effluent-F4-R4	0.75
CST-Healthcare Sludge-F10-R10: CST-Healthcare Sludge-DF-DR	0.86
CST-Healthcare Sludge-F3-R3 : CST-Healthcare Sludge-DF-DR	0.97
CST-Healthcare Sludge-F4-R4 : CST-Healthcare Sludge-DF-DR	0.90
CST-Healthcare Sludge-F7-R7 : CST-Healthcare Sludge-DF-DR	1.00
CST-Healthcare Sludge-F3-R3 : CST-Healthcare Sludge-F10-R10	0.57
CST-Healthcare Sludge-F4-R4 : CST-Healthcare Sludge-F10-R10	1.000
CST-Healthcare Sludge-F7-R7 : CST-Healthcare Sludge-F10-R10	0.93
CST-Healthcare Sludge-F4-R4 : CST-Healthcare Sludge-F3-R3	0.63
CST-Healthcare Sludge-F7-R7 : CST-Healthcare Sludge-F3-R3	0.92
CST-Healthcare Sludge-F7-R7 : CST-Healthcare Sludge-F4-R4	0.96

CST-Healthcare Effluent-F10-R10 : CST-Healthcare Effluent-DF-DR	NA
CST-Healthcare Effluent-F3-R3 : CST-Healthcare Effluent-DF-DR	NA
CST-Healthcare Effluent-F4-R4 : CST-Healthcare Effluent-DF-DR	NA
CST-Healthcare Effluent-F7-R7 : CST-Healthcare Effluent-DF-DR	NA
CST-Healthcare Effluent-F3-R3 : CST-Healthcare Effluent-F10-R10	NA
CST-Healthcare Effluent-F4-R4 : CST-Healthcare Effluent-F10-R10	NA
CST-Healthcare Effluent-F7-R7 : CST-Healthcare Effluent-F10-R10	NA
CST-Healthcare Effluent-F4-R4 : CST-Healthcare Effluent-F3-R3	NA
CST-Healthcare Effluent-F7-R7 : CST-Healthcare Effluent-F3-R3	NA
CST-Healthcare Effluent-F7-R7 : CST-Healthcare Effluent-F4-R4	NA

Appendix C: Chapter 4

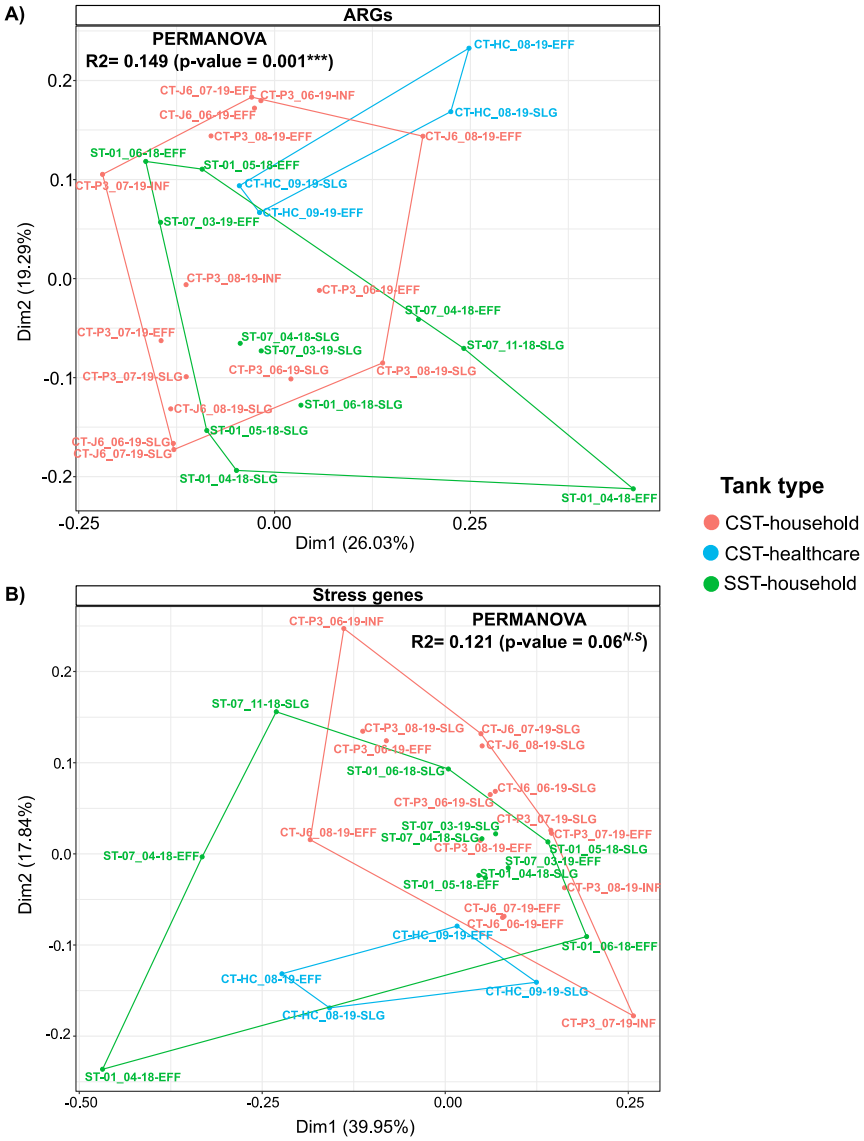

Figure C.1: NMDS plot based on Bray-Curtis dissimilarity index of ARGS A) and stress gene B) abundance between the three septic tank reactor (CST-household, CST-healthcare, SST-household). *** indicates p-value <0.001; N.S- not statistically significant.

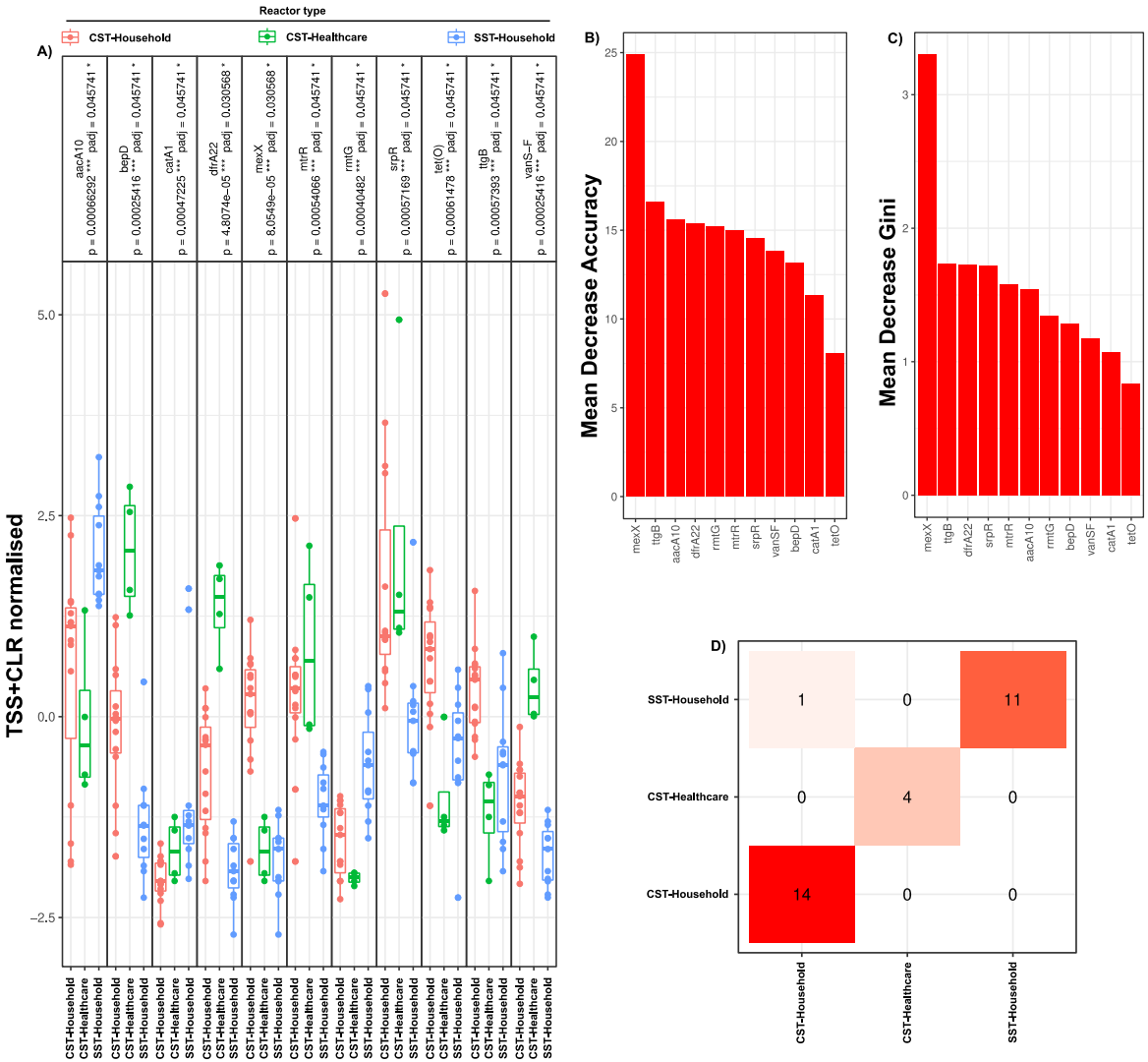

Figure C.2: Differentially abundant ARGs and Stress gene (Combined) between the three septic tank reactors (CST-household, CST-healthcare, SST-household). A) Differentially abundant ARGs and Stress gene (Combined) between the three septic tank reactors. B) Mean Decrease Accuracy and C) Mean Decrease Gini importance measures ranking the differentially abundant genes from most importance (highest value) to the least important for differentiating the groups (septic tank types). D) confusion matrix analysis for classification of each sample into their respective reactor type (CST-household, CST-healthcare, SST-household). * p-adj-value < 0.05.

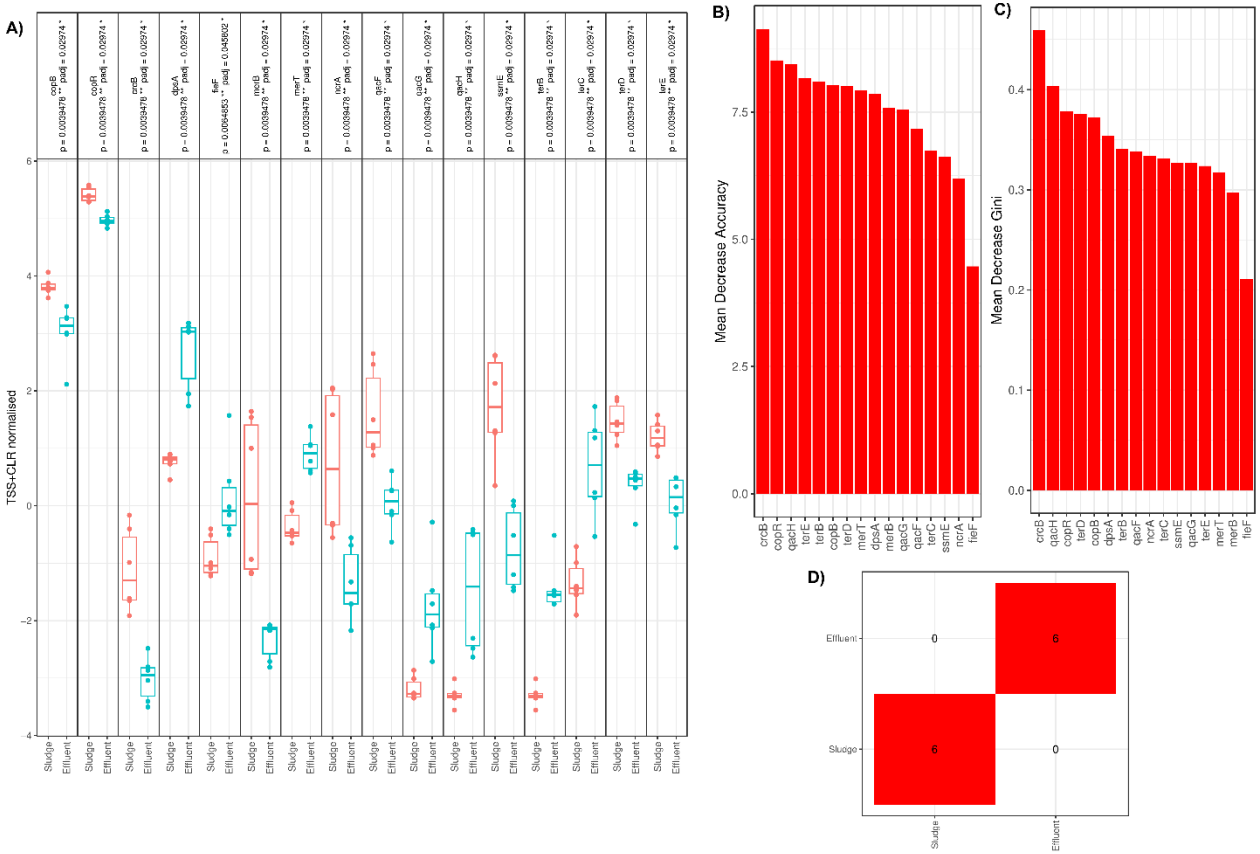

Figure C.3: Differentially abundant Stress gene between sludge and effluent for the CST-household reactor. A) Differentially abundant and Stress gene between the sample types (sludge and effluent). B) Mean Decrease Accuracy and C) Mean Decrease Gini importance measures ranking the differentially abundant genes from most importance (highest value) to the least important for differentiating the groups (sample types). D) confusion matrix analysis for classification of each sample into their respective sample type (sludge and effluent). * p-adj-value < 0.05.

Table C.1: Alpha diversity of detected ARGs and stress genes from septic tank wastewater samples characterised using the lowest stringent parameter: 25 amino acids coverage with 40% identity

Reactor	Sample type	Sample ID	ARGs			Stress genes		
			Richness (rarefied count)	Shannon entropy	Pielou's evenness	Richness (rarefied count)	Shannon entropy	Pielou's evenness
CST-Household	**Influent**	CT-P3_06-19	282.94	4.20	0.73	76.33	3.32	0.75
		CT-P3_07-19	388.00	4.42	0.74	84.00	3.14	0.71
		CT-P3_08-19	441.40	4.50	0.73	88.99	3.20	0.70
		Mean	**370.78**	**4.37**	**0.74**	**83.11**	**3.22**	**0.72**
		SD	**80.62**	**0.16**	**0.01**	**6.38**	**0.09**	**0.03**
	Sludge	CT-P3_06-19	437.74	4.25	0.69	88.75	3.20	0.71
		CT-P3_07-19	437.03	4.31	0.70	87.52	3.16	0.70
		CT-P3_08-19	443.40	4.29	0.69	88.92	3.18	0.69
		CT-J6_06-19	380.25	4.47	0.74	83.39	3.24	0.72
		CT-J6_07-19	375.94	4.40	0.73	86.14	3.17	0.70
		CT-J6_08-19	375.99	4.43	0.74	88.42	3.17	0.69
		Mean	**408.39**	**4.36**	**0.72**	**87.19**	**3.19**	**0.70**
		SD	**34.06**	**0.09**	**0.03**	**2.13**	**0.03**	**0.01**
	Effluent	CT-P3_06-19	441.96	4.49	0.72	93.85	3.25	0.70
		CT-P3_07-19	434.33	4.41	0.72	92.33	3.21	0.70
		CT-P3_08-19	391.23	4.47	0.74	88.94	3.28	0.72
		CT-J6_06-19	316.95	4.23	0.72	78.01	3.11	0.70
		CT-J6_07-19	323.19	4.26	0.73	78.37	3.14	0.71
		CT-J6_08-19	334.28	4.34	0.73	81.85	3.20	0.71
		Mean	**373.65**	**4.36**	**0.73**	**85.56**	**3.20**	**0.71**
		SD	**56.51**	**0.11**	**0.01**	**7.05**	**0.06**	**0.01**
CST-Healthcare	**Sludge**	CT-HC_08-19	334.53	3.97	0.66	80.20	3.10	0.69
		CT-HC_09-19	381.67	4.24	0.70	86.30	3.18	0.71
		Mean	**358.10**	**4.11**	**0.68**	**83.25**	**3.14**	**0.70**
		SD	**33.33**	**0.19**	**0.03**	**4.31**	**0.06**	**0.02**
	Effluent	CT-HC_08-19	304.21	3.92	0.66	77.78	3.13	0.70
		CT-HC_09-19	355.68	4.32	0.72	86.11	3.24	0.72
		Mean	**329.94**	**4.12**	**0.69**	**81.95**	**3.18**	**0.71**
		SD	**36.39**	**0.28**	**0.04**	**5.89**	**0.08**	**0.01**
SST-Household	**Sludge**	ST-01_04-18	382.08	4.28	0.71	78.48	3.38	0.76
		ST-01_05-18	411.29	4.11	0.67	87.87	3.19	0.70
		ST-01_06-18	410.40	4.13	0.67	84.85	3.15	0.69
		ST-07_04-18	405.96	4.48	0.74	93.26	3.46	0.75
		ST-07_11-18	407.06	4.39	0.71	89.49	3.30	0.72
		ST-07_03-19	393.76	4.22	0.69	85.55	3.16	0.70
		Mean	**401.76**	**4.27**	**0.70**	**86.58**	**3.27**	**0.72**
		SD	**11.51**	**0.15**	**0.02**	**4.99**	**0.13**	**0.03**
	Effluent	ST-01_04-18	213.44	3.42	0.60	60.35	3.02	0.70
		ST-01_05-18	353.93	4.35	0.73	76.17	3.15	0.71
		ST-01_06-18	331.97	4.28	0.73	70.60	3.02	0.70
		ST-07_04-18	318.68	4.35	0.73	85.04	3.51	0.78
		ST-07_03-19	361.55	4.39	0.74	79.65	3.12	0.70
		Mean	**315.91**	**4.16**	**0.71**	**74.36**	**3.16**	**0.72**

	SD	59.78	0.41	0.06	9.43	0.20	0.03

Table C.2: Alpha diversity of detected ARGs and stress genes from septic tank wastewater samples characterised using the medium stringent parameter: 50 amino acids coverage with 75% identity

Reactor	Sample type	Sample ID	ARGs			Stress genes		
			Richness	Shannon entropy	Pielou's evenness	Richness	Shannon entropy	Pielou's evenness
CST-Household	**Influent**	CT-P3_06-19	25.81	2.20	0.56	8.04	1.55	0.47
		CT-P3_07-19	34.73	3.13	0.88	15.05	1.58	0.53
		CT-P3_08-19	31.50	3.03	0.73	17.66	1.73	0.50
		Mean	**30.68**	**2.79**	**0.72**	**13.58**	**1.62**	**0.50**
		SD	**4.52**	**0.51**	**0.16**	**4.98**	**0.09**	**0.03**
	Sludge	CT-P3_06-19	32.78	3.07	0.77	9.51	0.98	0.30
		CT-P3_07-19	36.40	3.12	0.84	7.15	0.83	0.26
		CT-P3_08-19	43.41	3.53	0.84	7.78	0.80	0.25
		CT-J6_06-19	28.10	2.55	0.65	11.96	1.80	0.64
		CT-J6_07-19	23.59	2.15	0.60	11.94	1.46	0.52
		CT-J6_08-19	25.55	2.29	0.61	13.06	1.48	0.46
		Mean	**31.64**	**2.79**	**0.72**	**10.23**	**1.22**	**0.40**
		SD	**7.44**	**0.54**	**0.11**	**2.45**	**0.41**	**0.16**
	Effluent	CT-P3_06-19	34.95	3.30	0.84	14.25	1.57	0.46
		CT-P3_07-19	32.85	3.10	0.87	12.36	1.34	0.46
		CT-P3_08-19	33.66	2.95	0.76	17.01	2.42	0.72
		CT-J6_06-19	29.38	2.87	0.81	12.98	1.42	0.46
		CT-J6_07-19	30.50	3.06	0.85	14.56	1.69	0.56
		CT-J6_08-19	29.91	3.08	0.78	20.54	2.58	0.72
		Mean	**31.87**	**3.06**	**0.82**	**15.28**	**1.84**	**0.56**
		SD	**2.26**	**0.15**	**0.04**	**3.04**	**0.53**	**0.13**
CST-Healthcare	**Sludge**	CT-HC_08-19	23.84	2.58	0.61	14.96	2.10	0.63
		CT-HC_09-19	44.94	3.51	0.85	16.89	1.76	0.55
		Mean	**34.39**	**3.04**	**0.73**	**15.92**	**1.93**	**0.59**
		SD	**14.92**	**0.66**	**0.16**	**1.37**	**0.24**	**0.06**
	Effluent	CT-HC_08-19	28.15	2.57	0.62	13.76	1.62	0.46
		CT-HC_09-19	29.74	2.95	0.71	19.58	1.94	0.57
		Mean	**28.94**	**2.76**	**0.67**	**16.67**	**1.78**	**0.51**
		SD	**1.13**	**0.27**	**0.06**	**4.11**	**0.23**	**0.07**
SST-Household	**Sludge**	ST-01_04-18	23.60	2.78	0.71	10.31	1.69	0.55
		ST-01_05-18	41.52	3.51	0.89	14.48	1.37	0.40
		ST-01_06-18	40.46	3.37	0.85	9.15	1.06	0.32
		ST-07_04-18	36.54	3.00	0.67	25.80	2.77	0.75
		ST-07_11-18	41.67	3.29	0.80	13.71	1.87	0.55
		ST-07_03-19	42.27	3.22	0.83	12.10	1.22	0.37
		Mean	**37.68**	**3.20**	**0.79**	**14.26**	**1.66**	**0.49**
		SD	**7.20**	**0.26**	**0.08**	**6.00**	**0.62**	**0.16**
	Effluent	ST-01_04-18	11.20	1.19	0.39	14.00	2.07	0.79
		ST-01_05-18	34.19	3.19	0.84	13.58	1.63	0.49
		ST-01_06-18	30.08	2.69	0.78	8.26	1.18	0.39
		ST-07_04-18	19.99	2.45	0.60	19.07	2.29	0.63
		ST-07_03-19	37.00	3.21	0.89	16.62	1.81	0.55

	Mean	26.49	2.54	0.70	14.31	1.80	0.57
	SD	10.71	0.83	0.20	4.04	0.43	0.15

Table C.3: Alpha diversity of detected ARGs and stress genes from septic tank wastewater samples characterised using the high stringent parameter: coverage 75 amino acids with 90% identity

Reactor	Sample type	Sample ID	ARGs			Stress genes		
			Richness	Shannon entropy	Pielou's evenness	Richness	Shannon entropy	Pielou's evenness
CST-Household	**Influent**	CT-P3_06-19	8.97	2.34	0.84	1.00	2.28	0.89
		CT-P3_07-19	9.00	2.13	0.97	1.00	1.90	0.98
		CT-P3_08-19	8.57	2.39	0.72	1.00	1.97	0.85
		Mean	**8.85**	**2.29**	**0.84**	**1.00**	**2.05**	**0.91**
		SD	**0.24**	**0.14**	**0.13**	**0.00**	**0.20**	**0.06**
	Sludge	CT-P3_06-19	5.50	1.50	0.61	1.00	2.08	0.95
		CT-P3_07-19	7.91	1.41	0.64	1.00	1.95	1.00
		CT-P3_08-19	11.03	2.65	0.86	1.00	2.03	0.98
		CT-J6_06-19	5.16	1.43	0.56	1.00	0.00	NA
		CT-J6_07-19	4.13	1.03	0.64	1.00	0.68	0.62
		CT-J6_08-19	5.10	1.31	0.67	1.00	1.54	0.74
		Mean	**6.47**	**1.56**	**0.66**	**1.00**	**1.38**	**0.86**
		SD	**2.56**	**0.56**	**0.10**	**0.00**	**0.85**	**0.17**
	Effluent	CT-P3_06-19	7.81	2.08	0.79	1.00	1.79	0.72
		CT-P3_07-19	6.50	1.71	0.78	1.00	0.88	0.80
		CT-P3_08-19	12.21	2.75	0.92	1.00	1.60	0.70
		CT-J6_06-19	8.87	2.18	0.91	1.00	2.16	0.99
		CT-J6_07-19	8.46	2.14	0.86	1.00	2.02	0.92
		CT-J6_08-19	8.84	2.38	0.77	1.00	2.26	0.75
		Mean	**8.78**	**2.21**	**0.84**	**1.00**	**1.79**	**0.81**
		SD	**1.90**	**0.34**	**0.07**	**0.00**	**0.51**	**0.12**
CST-Healthcare	**Sludge**	CT-HC_08-19	11.96	2.77	0.91	1.00	2.38	0.90
		CT-HC_09-19	12.01	2.63	0.95	1.00	2.04	0.93
		Mean	**11.98**	**2.70**	**0.93**	**1.00**	**2.21**	**0.92**
		SD	**0.04**	**0.09**	**0.03**	**0.00**	**0.24**	**0.02**
	Effluent	CT-HC_08-19	10.74	2.61	0.86	1.00	2.53	0.93
		CT-HC_09-19	11.07	2.74	0.85	1.00	2.26	0.80
		Mean	**10.90**	**2.68**	**0.85**	**1.00**	**2.39**	**0.87**
		SD	**0.24**	**0.09**	**0.00**	**0.00**	**0.19**	**0.10**
SST-Household	**Sludge**	ST-01_04-18	6.40	1.68	0.59	1.00	1.41	0.88
		ST-01_05-18	10.41	2.57	0.87	1.00	2.12	0.92
		ST-01_06-18	6.38	1.71	0.65	1.00	1.19	0.57
		ST-07_04-18	12.72	3.18	0.87	1.00	2.53	0.88
		ST-07_11-18	11.57	2.67	0.92	1.00	2.20	0.96
		ST-07_03-19	11.11	2.32	0.90	1.00	2.11	0.96
		Mean	**9.76**	**2.35**	**0.80**	**1.00**	**1.93**	**0.86**
		SD	**2.72**	**0.58**	**0.14**	**0**	**0.51**	**0.15**
	Effluent	ST-01_04-18	2.00	0.30	0.22	0	0	0
		ST-01_05-18	11.04	2.59	0.86	1.00	2.04	0.85
		ST-01_06-18	10.04	2.13	0.86	1.00	2.15	0.98
		ST-07_04-18	7.65	2.20	0.65	1.00	2.49	0.88
		ST-07_03-19	11.24	2.44	0.95	1.00	2.19	0.95
		Mean	**8.39**	**1.94**	**0.71**	**1.00***	**2.22***	**0.91***

	SD	3.85	0.93	0.30	0*	0.19*	0.06*

*mean and standard deviation (SD) calculated without the ST-01_04-18 effluent samples which was 0.

References

Abbassi, B.E., Abuharb, R., Ammary, B., Almanaseer, N. and Kinsley, C. 2018. Modified septic tank: Innovative onsite wastewater treatment system. *Water (Switzerland)* 10(5). doi: 10.3390/w10050578.

Abramova, A., Berendonk, T.U. and Bengtsson-Palme, J. 2022. Meta-analysis reveals the global picture of antibiotic resistance gene prevalence across environments. *bioRxiv*, p. 2022.01.29.478248. Available at: http://biorxiv.org/content/early/2022/01/30/2022.01.29.478248.abstract.

Adabi, M., Bakhshi, B., Goudarzi, H., Zahraei, S.M. and Pourshafie, M.R. 2009. Distribution of class i integron and sulfamethoxazole trimethoprim constin in Vibrio cholerae isolated from patients in Iran. *Microbial Drug Resistance* 15(3). doi: 10.1089/mdr.2009.0885.

Adedeji, W.A. 2016. THE TREASURE CALLED ANTIBIOTICS. *Annals of Ibadan postgraduate medicine* 14(2).

Alcock, B.P. et al. 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research* 51(D1), pp. D690–D699. Available at: https://doi.org/10.1093/nar/gkac920.

Amos, G.C.A., Ploumakis, S., Zhang, L., Hawkey, P.M., Gaze, W.H. and Wellington, E.M.H. 2018. The widespread dissemination of integrons throughout bacterial communities in a riverine system. *ISME Journal* 12(3). doi: 10.1038/s41396-017-0030-8.

An, X.L. et al. 2018. Tracking antibiotic resistome during wastewater treatment using high throughput quantitative PCR. *Environment International* 117. doi: 10.1016/j.envint.2018.05.011.

Angly, F.E., Dennis, P.G., Skarshewski, A., Vanwonterghem, I., Hugenholtz, P. and Tyson, G.W. 2014. CopyRighter: A rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* 2(1). doi: 10.1186/2049-2618-2-11.

Antelo, V. et al. 2021. Metagenomic strategies identify diverse integron-integrase and antibiotic resistance genes in the Antarctic environment. *MicrobiologyOpen* 10(5). doi: 10.1002/mbo3.1219.

Anunnatsiri, S. et al. 2021. A Comparison Between 12 Versus 20 Weeks of Trimethoprim-

sulfamethoxazole as Oral Eradication Treatment for Melioidosis: An Open-label, Pragmatic, Multicenter, Non-inferiority, Randomized Controlled Trial. *Clinical Infectious Diseases* 73(11). doi: 10.1093/cid/ciaa1084.

Arbiv, O.A. et al. 2022. High-dose rifamycins in the treatment of TB: a systematic review and meta-analysis. *Thorax*. doi: 10.1136/thoraxjnl-2020-216497.

Arzanlou, M., Chai, W.C. and Venter, H. 2017. Intrinsic, adaptive and acquired antimicrobial resistance in Gram-negative bacteria. *Essays in Biochemistry* 61(1). doi: 10.1042/EBC20160063.

Babakhani, S. and Oloomi, M. 2018. Transposons: the agents of antibiotic resistance in bacteria. *Journal of Basic Microbiology* 58(11). doi: 10.1002/jobm.201800204.

Baker-Austin, C., Wright, M.S., Stepanauskas, R. and McArthur, J. V. 2006. Co-selection of antibiotic and metal resistance. *Trends in Microbiology* 14(4), pp. 176–182. doi: 10.1016/J.TIM.2006.02.006.

Barraud, O., Baclet, M.C., Denis, F. and Ploy, M.C. 2010. Quantitative multiplex real-time PCR for detecting class 1, 2 and 3 integrons. *Journal of Antimicrobial Chemotherapy* 65(8). doi: 10.1093/jac/dkq167.

Bashir, S., Haque, A. and Sarwar, Y. 2015. Prevalence of Integrons and Antibiotic Resistance among Uropathogenic Escherichia Coli from Faisalabad Region of Pakistan. *Archives of ClinalMicrobiology* 6(4).

Bass, L., Liebert, C.A., Lee, M.D., Summers, A.O., White, D.G., Thayer, S.G. and Maurer, J.J. 1999. Incidence and characterization of integrons, genetic elements mediating multiple-drug resistance, in avian Escherichia coli. *Antimicrobial Agents and Chemotherapy* 43(12). doi: 10.1128/aac.43.12.2925.

Bassetti, S., Tschudin-Sutter, S., Egli, A. and Osthoff, M. 2022. Optimizing antibiotic therapies to reduce the risk of bacterial resistance. *European Journal of Internal Medicine*. doi: 10.1016/j.ejim.2022.01.029.

Bengtsson-Palme, J., Larsson, D.G.J. and Kristiansson, E. 2017. Using metagenomics to investigate human and environmental resistomes. *Journal of Antimicrobial Chemotherapy* 72(10). doi: 10.1093/jac/dkx199.

Bengtsson-Palme, J., Milakovic, M., Švecová, H., Ganjto, M., Jonsson, V., Grabic, R. and Udikovic-Kolic, N. 2019. Industrial wastewater treatment plant enriches antibiotic resistance genes and alters the structure of microbial communities. *Water Research* 162. doi:

10.1016/j.watres.2019.06.073.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1). doi: 10.1111/j.2517-6161.1995.tb02031.x.

Bennett, J.W. and Chung, K.T. 2001. Alexander Fleming and the discovery of penicillin. *Advances in Applied Microbiology* 49. doi: 10.1016/S0065-2164(01)49013-7.

Bennett, P.M. 2008. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. In: *British Journal of Pharmacology*. doi: 10.1038/sj.bjp.0707607.

Berendonk, T.U. et al. 2015. Tackling antibiotic resistance: the environmental framework. *Nature Reviews Microbiology* 13(5), pp. 310–317. Available at: https://doi.org/10.1038/nrmicro3439.

Berglund, B., Fick, J. and Lindgren, P.E. 2015. Urban wastewater effluent increases antibiotic resistance gene concentrations in a receiving northern European river. *Environmental Toxicology and Chemistry* 34(1). doi: 10.1002/etc.2784.

Bijekar, S. et al. 2022. The State of the Art and Emerging Trends in the Wastewater Treatment in Developing Nations. *Water* 14(16). doi: 10.3390/w14162537.

Bolyen, E. et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37(8). doi: 10.1038/s41587-019-0209-9.

Boonsarngsuk, V., Tansirichaiya, K. and Kiatboonsri, S. 2009. Thai drug-resistant tuberculosis predictive scores. *Singapore Medical Journal* 50(4).

Borruso, L., Harms, K., Johnsen, P.J., Nielsen, K.M. and Brusetti, L. 2016. Distribution of class 1 integrons in a highly impacted catchment. *Science of the Total Environment* 566–567. doi: 10.1016/j.scitotenv.2016.06.054.

Bourlat, S.J., Haenel, Q., Finnman, J. and Leray, M. 2016. Preparation of amplicon libraries for metabarcoding of marine eukaryotes using illumina MiSeq: The dual-PCR method. In: *Methods in Molecular Biology*. doi: 10.1007/978-1-4939-3774-5_13.

Brucoli, F. and McHugh, T.D. 2021. Rifamycins: Do not throw the baby out with the bathwater. Is rifampicin still an effective anti-tuberculosis drug? *Future Medicinal Chemistry* 13(24). doi: 10.4155/fmc-2021-0249.

Bugarel, M., Granier, S.A., Weill, F.X., Fach, P. and Brisabois, A. 2011. A multiplex real-

time PCR assay targeting virulence and resistance genes in Salmonella enterica serotype Typhimurium. *BMC Microbiology* 11. doi: 10.1186/1471-2180-11-151.

Bunce, J.T. and Graham, D.W. 2019. A simple approach to predicting the reliability of smallwaste water treatment plants. *Water (Switzerland)* 11(11). doi: 10.3390/w11112397.

Burch, T.R., Sadowsky, M.J. and LaPara, T.M. 2014. Fate of antibiotic resistance genes and class 1 integrons in soil microcosms following the application of treated residual municipal wastewater solids. *Environmental Science and Technology* 48(10). doi: 10.1021/es501098g.

Bürgmann, H. et al. 2018. Water and sanitation: An essential battlefront in the war on antimicrobial resistance. *FEMS Microbiology Ecology* 94(9). doi: 10.1093/femsec/fiy101.

Byrne-Bailey, K.G., Gaze, W.H., Kay, P., Boxall, A.B.A., Hawkey, P.M. and Wellington, E.M.H. 2009. Prevalence of sulfonamide resistance genes in bacterial isolates from manured agricultural soils and pig slurry in the United Kingdom. *Antimicrobial Agents and Chemotherapy* 53(2). doi: 10.1128/AAC.00652-07.

Cai, Y., Liu, J., Li, G., Wong, P.K. and An, T. 2022. Formation mechanisms of viable but nonculturable bacteria through induction by light-based disinfection and their antibiotic resistance gene transfer risk: A review. *Critical Reviews in Environmental Science and Technology* 52(20). doi: 10.1080/10643389.2021.1932397.

Cambray, G., Guerout, A.M. and Mazel, D. 2010. Integrons. *Annual Review of Genetics* 44. doi: 10.1146/annurev-genet-102209-163504.

Carattoli, A. 2011. Plasmids in Gram negatives: Molecular typing of resistance plasmids. *International Journal of Medical Microbiology* 301(8). doi: 10.1016/j.ijmm.2011.09.003.

Di Cesare, A., Eckert, E.M. and Corno, G. 2016. Co-selection of antibiotic and heavy metal resistance in freshwater bacteria. *Journal of Limnology* 75(2S). doi: 10.4081/jlimnol.2016.1198.

Che, Y., Xu, X., Yang, Y., Břinda, K., Hanage, W., Yang, C. and Zhang, T. 2022. High-resolution genomic surveillance elucidates a multilayered hierarchical transfer of resistance between WWTP- and human/animal-associated bacteria. *Microbiome* 10(1). doi: 10.1186/s40168-021-01192-w.

Chen, C.M., Lai, C.H., Wu, H.J. and Wu, L.T. 2017. Genetic characteristic of class 1 integrons in proteus mirabilis isolates from urine samples. *BioMedicine (France)* 7(2). doi: 10.1051/bmdcn/2017070202.

Chen, H. and Zhang, M. 2013a. Effects of advanced treatment systems on the removal of antibiotic resistance genes in wastewater treatment plants from Hangzhou, China. *Environmental Science and Technology* 47(15). doi: 10.1021/es401091y.

Chen, H. and Zhang, M. 2013b. Occurrence and removal of antibiotic resistance genes in municipal wastewater and rural domestic sewage treatment systems in eastern China. *Environment International* 55. doi: 10.1016/j.envint.2013.01.019.

Chen, Q., An, X., Li, H., Su, J., Ma, Y. and Zhu, Y.G. 2016. Long-term field application of sewage sludge increases the abundance of antibiotic resistance genes in soil. *Environment International* 92–93. doi: 10.1016/j.envint.2016.03.026.

Chen, Y. et al. 2019a. High-throughput profiling of antibiotic resistance gene dynamic in a drinking water river-reservoir system. *Water Research* 149. doi: 10.1016/j.watres.2018.11.007.

Chen, Z. et al. 2019b. Antibiotic resistance genes and bacterial communities in cornfield and pasture soils receiving swine and dairy manures. *Environmental Pollution* 248. doi: 10.1016/j.envpol.2019.02.093.

Chesdachai, S., Zughaier, S.M., Hao, L., Kempker, R.R., Blumberg, H.M., Ziegler, T.R. and Tangpricha, V. 2016. The effects of first-line anti-tuberculosis drugs on the actions of vitamin D in human macrophages. *Journal of Clinical and Translational Endocrinology* 6. doi: 10.1016/j.jcte.2016.08.005.

Cholet, F., Ijaz, U.Z. and Smith, C.J. 2019. Differential ratio amplicons (R amp) for the evaluation of RNA integrity extracted from complex environmental samples. *Environmental Microbiology* 21(2). doi: 10.1111/1462-2920.14516.

Cholet, F., Ijaz, U.Z. and Smith, C.J. 2020. Reverse transcriptase enzyme and priming strategy affect quantification and diversity of environmental transcripts. *Environmental Microbiology* 22(6). doi: 10.1111/1462-2920.15017.

Chu, B.T.T., Petrovich, M.L., Chaudhary, A., Wright, D., Murphy, B., Wells, G. and Poretsky, R. 2018. Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. *Applied and Environmental Microbiology* 84(5). doi: 10.1128/AEM.02168-17.

Chuanchuen, R., Khemtong, S. and Padungtod, P. 2007. Occurrence of QACE/QACEΔ1 genes and their correlation with class 1 integrons in Salmonella enterica isolates from poultry and swine. *Southeast Asian Journal of Tropical Medicine and Public Health* 38(5).

Coates, A.R., Halls, G. and Hu, Y. 2011. Novel classes of antibiotics or more of the same? *British Journal of Pharmacology* 163(1). doi: 10.1111/j.1476-5381.2011.01250.x.

Connelly, S. et al. 2019. Solar septic tank: Next generation sequencing reveals effluent microbial community composition as a useful index of system performance. *Water (Switzerland)* 11(12). doi: 10.3390/W11122660.

Cook, M.A. and Wright, G.D. 2023. The past, present, and future of antibiotics. *Science Translational Medicine* 14(657), p. eabo7793. Available at: https://doi.org/10.1126/scitranslmed.abo7793.

Coque, T.M., Graham, D.W., Pruden, A., So, A.D. and Topp, E. 2023. Bracing for Superbugs :Strengthening environmental action in the One Health response to antimicrobial resistance. Available at: https://digitallibrary.un.org/record/4009999/files/antimicrobial_Report%25202023.pdf.

Cox, G. and Wright, G.D. 2013. Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *International Journal of Medical Microbiology* 303(6–7). doi: 10.1016/j.ijmm.2013.02.009.

Cuetero-Martínez, Y., Villamizar-Ojeda, K.N., Hernández-Santiago, M.J., De los Cobos-Vasconcelos, D., Aguirre-Garrido, J.F., López-Vidal, Y. and Noyola, A. 2023. Removal of intI1, ARGs, and SARS-CoV-2 and changes in bacterial communities in four sewage treatment facilities. *Science of the Total Environment* 903. doi: 10.1016/j.scitotenv.2023.165984.

Dai, Z. et al. 2020. Disinfection exhibits systematic impacts on the drinking water microbiome. *Microbiome* 8(1). doi: 10.1186/s40168-020-00813-0.

Darby, E.M., Trampari, E., Siasat, P., Gaya, M.S., Alav, I., Webber, M.A. and Blair, J.M.A. 2022. Molecular mechanisms of antibiotic resistance revisited. *Nature Reviews Microbiology*. Available at: https://doi.org/10.1038/s41579-022-00820-y.

Davis, B.C. et al. 2023. Recommendations for the use of metagenomics for routine monitoring of antibiotic resistance in wastewater and impacted aquatic environments. *Critical Reviews in Environmental Science and Technology*, pp. 1–26. Available at: https://doi.org/10.1080/10643389.2023.2181620.

Díaz, M.A., Cooper, R.K., Cloeckaert, A. and Siebeling, R.J. 2006. Plasmid-mediated high-level gentamicin resistance among enteric bacteria isolated from pet turtles in Louisiana. *Applied and Environmental Microbiology* 72(1). doi: 10.1128/AEM.72.1.306-312.2006.

Domingues, S., da Silva, G.J. and Nielsen, K.M. 2012. Integrons: Vehicles and pathways for horizontal dissemination in bacteria. *Mobile genetic elements* 2(5). doi: 10.4161/mge.22967.

Dong, P., Cui, Q., Fang, T., Huang, Y. and Wang, H. 2019. Occurrence of antibiotic resistance genes and bacterial pathogens in water and sediment in urban recreational water. *Journal of Environmental Sciences (China)* 77. doi: 10.1016/j.jes.2018.06.011.

Dray, S. et al. 2012. Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs* 82(3). doi: 10.1890/11-1183.1.

Dungan, R.S. and Bjorneberg, D.L. 2020. Antibiotic resistance genes, class 1 integrons, and IncP-1/IncQ-1 plasmids in irrigation return flows. *Environmental Pollution* 257. doi: 10.1016/j.envpol.2019.113568.

Ebner, P., Garner, K. and Mathew, A. 2004. Class 1 integrons in various Salmonella enterica serovars isolated from animals and identification of genomic island SGI1 in Salmonella enterica var. Meleagridis. *Journal of Antimicrobial Chemotherapy* 53(6). doi: 10.1093/jac/dkh192.

Edokpayi, J.N., Odiyo, J.O. and Durowoju, O.S. 2017. Impact of Wastewater on Surface Water Quality in Developing Countries: A Case Study of South Africa. In: *Water Quality*. doi: 10.5772/66561.

Eiamphungporn, W., Yainoy, S., Jumderm, C., Tan-arsuwongkul, R., Tiengrim, S. and Thamlikitkul, V. 2018. Prevalence of the colistin resistance gene mcr-1 in colistin-resistant Escherichia coli and Klebsiella pneumoniae isolated from humans in Thailand. *Journal of Global Antimicrobial Resistance* 15. doi: 10.1016/j.jgar.2018.06.007.

Ekkapobyotin, C., Padungtod, P. and Chuanchuen, R. 2008. Antimicrobial resistance of Campylobacter coli isolates from swine. *International Journal of Food Microbiology* 128(2). doi: 10.1016/j.ijfoodmicro.2008.09.005.

Ekwanzala, M.D., Dewar, J.B. and Momba, M.N.B. 2020. Environmental resistome risks of wastewaters and aquatic environments deciphered by shotgun metagenomic assembly. *Ecotoxicology and Environmental Safety* 197. doi: 10.1016/j.ecoenv.2020.110612.

Falbo, V., Carattoli, A., Tosini, F., Pezzella, C., Dionisi, A.M. and Luzzi, I. 1999. Antibiotic resistance conferred by a conjugative plasmid and a class I integron in Vibrio cholerae O1 El Tor strains isolated in Albania and Italy. *Antimicrobial Agents and Chemotherapy* 43(3). doi: 10.1128/aac.43.3.693.

Feldgarden, M. et al. 2021. AMRFinderPlus and the Reference Gene Catalog facilitate

examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports* 11(1). doi: 10.1038/s41598-021-91456-0.

Fernández, L. and Hancock, R.E.W. 2012. Adaptive and mutational resistance: Role of porins and efflux pumps in drug resistance. *Clinical Microbiology Reviews* 25(4). doi: 10.1128/CMR.00043-12.

Fonseca, É.L., Vieira, V. V., Cipriano, R. and Vicente, A.C.P. 2005. Class 1 integrons in Pseudomonas aeruginosa isolates from clinical settings in Amazon region, Brazil. *FEMS Immunology and Medical Microbiology* 44(3). doi: 10.1016/j.femsim.2005.01.004.

Frank, T., Gautier, V., Talarmin, A., Bercion, R. and Arlet, G. 2007. Characterization of sulphonamide resistance genes and class 1 integron gene cassettes in Enterobacteriaceae, Central African Republic (CAR). *Journal of Antimicrobial Chemotherapy* 59(4). doi: 10.1093/jac/dkl538.

Garrido-Cardenas, J.A., Polo-López, M.I. and Oller-Alberola, I. 2017. Advanced microbial analysis for wastewater quality monitoring: metagenomics trend. *Applied Microbiology and Biotechnology* 101(20). doi: 10.1007/s00253-017-8490-3.

Gaze, W.H. et al. 2011. Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. *The ISME Journal* 5(8), pp. 1253–1261. Available at: http://www.nature.com/articles/ismej201115.

Ghaly, T.M., Gillings, M.R., Penesyan, A., Qi, Q., Rajabal, V. and Tetu, S.G. 2021. The natural history of integrons. *Microorganisms* 9(11). doi: 10.3390/microorganisms9112212.

Ghosh, S., Ramsden, S.J. and Lapara, T.M. 2009. The role of anaerobic digestion in controlling the release of tetracycline resistance genes and class 1 integrons from municipal wastewater treatment plants. *Applied Microbiology and Biotechnology* 84(4). doi: 10.1007/s00253-009-2125-2.

Gibson, C. et al. 2023. Multiplexed amplicon sequencing reveals high sequence diversity of antibiotic resistance genes in Québec sewers. doi: 10.1101/2023.03.06.531290.

Gillings, M.R. 2014. Integrons: Past, Present, and Future. *Microbiology and Molecular Biology Reviews* 78(2). doi: 10.1128/mmbr.00056-13.

Gillings, M.R. 2017a. Class 1 integrons as invasive species. *Current Opinion in Microbiology* 38. doi: 10.1016/j.mib.2017.03.002.

Gillings, M.R. 2017b. Lateral gene transfer, bacterial genome evolution, and the

Anthropocene. *Annals of the New York Academy of Sciences* 1389(1). doi: 10.1111/nyas.13213.

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

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

Gillings, M.R., Krishnan, S., Worden, P.J. and Hardwick, S.A. 2008b. Recovery of diverse genes for class 1 integron-integrases from environmental DNA samples. *FEMS Microbiology Letters* 287(1). doi: 10.1111/j.1574-6968.2008.01291.x.

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G. and Bailey, M.J. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Applied and Environmental Microbiology* 66(12). doi: 10.1128/AEM.66.12.5488-5491.2000.

Grossman, H.. T. 2016. Tetracycline antibiotics and resistance. *Cold Spring Harbor Perspectives in Medicine* 6(4).

Gu, B., Pan, S., Wang, T., Zhao, W., Mei, Y., Huang, P. and Tong, M. 2008. Novel cassette arrays of integrons in clinical strains of Enterobacteriaceae in China. *International Journal of Antimicrobial Agents* 32(6). doi: 10.1016/j.ijantimicag.2008.06.019.

Guerra, B., Soto, S.M., Argüelles, J.M. and Mendoza, M.C. 2001. Multidrug resistance is mediated by large plasmids carrying a class 1 integron in the emergent Salmonella enterica serotype [4,5,12:i:-]. *Antimicrobial Agents and Chemotherapy* 45(4). doi: 10.1128/AAC.45.4.1305-1308.2001.

Gündoğdu, A., Long, Y.B., Vollmerhausen, T.L. and Katouli, M. 2011. Antimicrobial resistance and distribution of sul genes and integron-associated inti genes among uropathogenic Escherichia coli in Queensland, Australia. *Journal of Medical Microbiology* 60(11). doi: 10.1099/jmm.0.034140-0.

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

Gupta, P.D. and Birdi, T.J. 2017. Development of botanicals to combat antibiotic resistance.

Journal of Ayurveda and Integrative Medicine 8(4). doi: 10.1016/j.jaim.2017.05.004.

Gupta, S.K., Shin, H., Han, D., Hur, H.G. and Unno, T. 2018. Metagenomic analysis reveals the prevalence and persistence of antibiotic- and heavy metal-resistance genes in wastewater treatment plant. *Journal of Microbiology* 56(6). doi: 10.1007/s12275-018-8195-z.

Hall, R.M. 2012. Integrons and gene cassettes: Hotspots of diversity in bacterial genomes. *Annals of the New York Academy of Sciences* 1267(1). doi: 10.1111/j.1749-6632.2012.06588.x.

Hall, T.A. 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. *Nucleic Acids Symposium Series* 41.

Hansson, K., Sundström, L., Pelletier, A. and Roy, P.H. 2002. IntI2 integron integrase in Tn7. *Journal of Bacteriology* 184(6). doi: 10.1128/JB.184.6.1712-1721.2002.

Harada, H., Strande, L. and Fujii, S. 2016. Challenges and Opportunities of Faecal Sludge Management for Global Sanitation. In: *Towards Future Earth: Challenges and Progress of Global Environmental Studies*.

Hardwick, S.A., Stokes, H.W., Findlay, S., Taylor, M. and Gillings, M.R. 2008. Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. *FEMS Microbiology Letters* 278(2). doi: 10.1111/j.1574-6968.2007.00992.x.

Hayward, J.L., Huang, Y., Yost, C.K., Hansen, L.T., Lake, C., Tong, A. and Jamieson, R.C. 2019. Lateral flow sand filters are effective for removal of antibiotic resistance genes from domestic wastewater. *Water Research* 162. doi: 10.1016/j.watres.2019.07.004.

Hickman, A.B., Chandler, M. and Dyda, F. 2010. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Critical Reviews in Biochemistry and Molecular Biology* 45(1). doi: 10.3109/10409230903505596.

Hinjoy, S. et al. 2018. Melioidosis in Thailand: Present and Future. *Tropical Medicine and Infectious Disease* 3(2). doi: 10.3390/tropicalmed3020038.

Ho, P.L. et al. 2012. Dissemination of pHK01-like incompatibility group IncFII plasmids encoding CTX-M-14 in Escherichia coli from human and animal sources. *Veterinary Microbiology* 158(1–2). doi: 10.1016/j.vetmic.2012.02.004.

Hocquet, D., Muller, A. and Bertrand, X. 2016. What happens in hospitals does not stay in hospitals: antibiotic-resistant bacteria in hospital wastewater systems. *Journal of Hospital*

Infection 93(4). doi: 10.1016/j.jhin.2016.01.010.

Holmes, A.H. et al. 2016. Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet* 387(10014). doi: 10.1016/S0140-6736(15)00473-0.

Hong, J.S., Yoon, E.J., Lee, H., Jeong, S.H. and Lee, K. 2016. Clonal dissemination of Pseudomonas aeruginosa sequence type 235 isolates carrying blaIMP-6 and emergence of blaGES-24 and blaIMP-10 on novel genomic islands PAGI-15 and -16 in South Korea. *Antimicrobial Agents and Chemotherapy* 60(12). doi: 10.1128/AAC.01601-16.

Huang, X. et al. 2019. Higher temperatures do not always achieve better antibiotic resistance gene removal in anaerobic digestion of swine manure. *Applied and Environmental Microbiology* 85(7). doi: 10.1128/AEM.02878-18.

Hussein, M. et al. 2020. The Killing Mechanism of Teixobactin against Methicillin-Resistant Staphylococcus aureus: an Untargeted Metabolomics Study. *mSystems* 5(3). doi: 10.1128/msystems.00077-20.

Hutchings, M., Truman, A. and Wilkinson, B. 2019. Antibiotics: past, present and future. *Current Opinion in Microbiology* 51. doi: 10.1016/j.mib.2019.10.008.

Jackson, C., Hsia, Y., Bielicki, J.A., Ellis, S., Stephens, P., Wong, I.C.K. and Sharland, M. 2019. Estimating global trends in total and childhood antibiotic consumption, 2011-2015. *BMJ Global Health* 4(1). doi: 10.1136/bmjgh-2018-001241.

Jiao, Y.N., Zhou, Z.C., Chen, T., Wei, Y.Y., Zheng, J., Gao, R.X. and Chen, H. 2018. Biomarkers of antibiotic resistance genes during seasonal changes in wastewater treatment systems. *Environmental Pollution* 234. doi: 10.1016/j.envpol.2017.11.048.

Johnson, T.A. et al. 2016. Clusters of antibiotic resistance genes enriched together stay together in swine agriculture. *mBio* 7(2). doi: 10.1128/mBio.02214-15.

Joshi, N. and Fass, J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. *Available at https://github.com/najoshi/sickle*.

Kainuma, A. et al. 2018. An outbreak of fluoroquinolone-resistant Pseudomonas aeruginosa ST357 harboring the exoU gene. *Journal of Infection and Chemotherapy* 24(8). doi: 10.1016/j.jiac.2018.03.008.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. and Wang, Z. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019(7). doi: 10.7717/peerj.7359.

Kapoor, G., Saigal, S. and Elongavan, A. 2017. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology Clinical Pharmacology* 33(3). doi: 10.4103/joacp.JOACP_349_15.

Karami, P., Khaledi, A., Mashoof, R.Y., Yaghoobi, M.H., Karami, M., Dastan, D. and Alikhani, M.Y. 2020. The correlation between biofilm formation capability and antibiotic resistance pattern in Pseudomonas aeruginosa. *Gene Reports* 18. doi: 10.1016/j.genrep.2019.100561.

Karaolia, P., Vasileiadis, S., G. Michael, S., G. Karpouzas, D. and Fatta-Kassinos, D. 2021. Shotgun metagenomics assessment of the resistome, mobilome, pathogen dynamics and their ecological control modes in full-scale urban wastewater treatment plants. *Journal of Hazardous Materials* 418. doi: 10.1016/j.jhazmat.2021.126387.

Karkman, A., Johnson, T.A., Lyra, C., Stedtfeld, R.D., Tamminen, M., Tiedje, J.M. and Virta, M. 2016. High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS Microbiology Ecology* 92(3). doi: 10.1093/femsec/fiw014.

Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30(14). doi: 10.1093/nar/gkf436.

Kennedy, C.A. et al. 2018. Multi-drug resistant Escherichia coli in diarrhoeagenic foals: Pulsotyping, phylotyping, serotyping, antibiotic resistance and virulence profiling. *Veterinary Microbiology* 223. doi: 10.1016/j.vetmic.2018.08.009.

Kerrn, M.B., Klemmensen, T., Frimodt-Möller, N. and Espersen, F. 2002. Susceptibility of Danish Escherichia coli strains isolated from urinary tract infections and bacteraemia, and distribution of sul genes conferring sulphonamide resistance. *Journal of Antimicrobial Chemotherapy* 50(4). doi: 10.1093/jac/dkf164.

Khademi, F., Ashrafi, S.S., Neyestani, Z., Vaez, H. and Sahebkar, A. 2021. Prevalence of class I, II and III integrons in multidrug-resistant and carbapenem-resistant Pseudomonas aeruginosa clinical isolates. *Gene Reports* 25. doi: 10.1016/j.genrep.2021.101407.

Kheiri, R. and Akhtari, L. 2016. Antimicrobial resistance and integron gene cassette arrays in commensal Escherichia coli from human and animal sources in IRI. *Gut Pathogens* 8(1). doi: 10.1186/s13099-016-0123-3.

Klein, E.Y. et al. 2018. Global increase and geographic convergence in antibiotic

consumption between 2000 and 2015. *Proceedings of the National Academy of Sciences of the United States of America* 115(15). doi: 10.1073/pnas.1717295115.

Kobayashi, K. et al. 2013. Identification and Characterization of a Novel aac(6′)-Iag Associated with the blaIMP-1-Integron in a Multidrug-Resistant Pseudomonas aeruginosa. *PLoS ONE* 8(8). doi: 10.1371/journal.pone.0070557.

Koeleman, J.G.M., Stoof, J., Van der Bijl, M.W., Vandenbroucke-Grauls, C.M.J.E. and Savelkoul, P.H.M. 2001. Identification of epidemic strains of Acinetobacter baumannii by integrase gene PCR. *Journal of Clinical Microbiology* 39(1). doi: 10.1128/JCM.39.1.8-13.2001.

Koottatep, T., Taweesan, A., Kanabkaew, T. and Polprasert, C. 2021. Inconvenient truth: Unsafely managed fecal sludge after achieving MDG for decades in Thailand. *Journal of Water Sanitation and Hygiene for Development* 11(6). doi: 10.2166/washdev.2021.118.

Krauland, M.G., Marsh, J.W., Paterson, D.L. and Harrison, L.H. 2009. Integron-mediated multidrug resistance in a global collection of nontyphoidal Salmonella enterica isolates. *Emerging Infectious Diseases* 15(3). doi: 10.3201/eid1503.081131.

Krawczyk, P.S., Lipinski, L. and Dziembowski, A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Research* 46(6). doi: 10.1093/NAR/GKX1321.

Labbate, M., Case, R.J. and Stokes, H.W. 2009. The integron/gene cassette system: an active player in bacterial adaptation. *Methods in molecular biology (Clifton, N.J.)* 532. doi: 10.1007/978-1-60327-853-9_6.

Lai, J., Wang, Y., Shen, J., Li, R., Han, J., Foley, S.L. and Wu, C. 2013. Unique class 1 Integron and multiple resistance genes co-located on inchi2 plasmid is associated with the emerging Multidrug resistance of salmonella Indiana isolated from chicken in China. *Foodborne Pathogens and Disease* 10(7). doi: 10.1089/fpd.2012.1455.

Lal Gupta, C., Kumar Tiwari, R. and Cytryn, E. 2020. Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. *Environment International* 138. doi: 10.1016/j.envint.2020.105667.

Langbehn, R.K., Michels, C. and Soares, H.M. 2021. Antibiotics in wastewater: From its occurrence to the biological removal by environmentally conscious technologies. *Environmental Pollution* 275. doi: 10.1016/j.envpol.2021.116603.

Lapara, T.M., Burch, T.R., McNamara, P.J., Tan, D.T., Yan, M. and Eichmiller, J.J. 2011.

Tertiary-treated municipal wastewater is a significant point source of antibiotic resistance genes into Duluth-Superior Harbor. *Environmental Science and Technology* 45(22). doi: 10.1021/es202775r.

Larkin, M.A. et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21). doi: 10.1093/bioinformatics/btm404.

Lee, S.H., Malone, C. and Kemp, P.F. 1993. Use of multiple 16S rRNA-targeted fluorescent probes to increase signal strength and measure cellular RNA from natural planktonic bacteria. *Marine Ecology Progress Series* 101(1–2). doi: 10.3354/meps101193.

Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters* 16(8). doi: 10.1111/ele.12141.

Leng, Y., Xiao, H., Li, Z. and Wang, J. 2020. Tetracyclines, sulfonamides and quinolones and their corresponding resistance genes in coastal areas of Beibu Gulf, China. *Science of the Total Environment* 714. doi: 10.1016/j.scitotenv.2020.136899.

Lerminiaux, N.A. and Cameron, A.D.S. 2019. Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian Journal of Microbiology* 65(1). doi: 10.1139/cjm-2018-0275.

Leverstein-Van Hall, M.A., Blok, H.E.M., Donders, A.R.T., Paauw, A., Fluit, A.C. and Verhoe, J. 2003. Multidrug resistance among enterobacteriaceae is strongly associated with the presence of integrons and is independent of species or isolate origin. *Journal of Infectious Diseases* 187(2). doi: 10.1086/345880.

Li, B., Qiu, Y., Song, Y., Lin, H. and Yin, H. 2019. Dissecting horizontal and vertical gene transfer of antibiotic resistance plasmid in bacterial community using microfluidics. *Environment International* 131. doi: 10.1016/j.envint.2019.105007.

Li, D., Liu, C.M., Luo, R., Sadakane, K. and Lam, T.W. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10). doi: 10.1093/bioinformatics/btv033.

Li, J., Cheng, W., Xu, L., Jiao, Y., Baig, S.A. and Chen, H. 2016. Occurrence and removal of antibiotics and the corresponding resistance genes in wastewater treatment plants: effluents' influence to downstream water environment. *Environmental Science and Pollution Research* 23(7). doi: 10.1007/s11356-015-5916-2.

Liguori, K., Keenum, I., Davis, B.C., Calarco, J., Milligan, E., Harwood, V.J. and Pruden, A. 2022. Antimicrobial Resistance Monitoring of Water Environments: A Framework for

Standardized Methods and Quality Control. *Environmental Science & Technology* 56(13), pp. 9149–9160. Available at: https://doi.org/10.1021/acs.est.1c08918.

Lim, N.Y.N., Roco, C.A. and Frostegård, Å. 2016. Transparent DNA/RNA co-extraction workflow protocol suitable for inhibitor-rich environmental samples that focuses on complete DNA removal for transcriptomic analyses. *Frontiers in Microbiology* 7(OCT). doi: 10.3389/fmicb.2016.01588.

Limmathurotsakul, D. et al. 2016. Predicted global distribution of Burkholderia pseudomallei and burden of melioidosis. *Nature Microbiology* 1(1). doi: 10.1038/nmicrobiol.2015.8.

Lin, Z.J., Zhou, Z.C., Zhu, L., Meng, L.X., Shuai, X.Y., Sun, Y.J. and Chen, H. 2021. Behavior of antibiotic resistance genes in a wastewater treatment plant with different upgrading processes. *Science of the Total Environment* 771. doi: 10.1016/j.scitotenv.2020.144814.

Ling, L.L. et al. 2015. A new antibiotic kills pathogens without detectable resistance. *Nature* 517(7535). doi: 10.1038/nature14098.

Liu, M., Ma, J., Jia, W. and Li, W. 2020. Antimicrobial Resistance and Molecular Characterization of Gene Cassettes from Class 1 Integrons in Pseudomonas aeruginosa Strains. *Microbial Drug Resistance* 26(6). doi: 10.1089/mdr.2019.0406.

Liu, X., Xiao, P., Guo, Y., Liu, L. and Yang, J. 2019. The impacts of different high-throughput profiling approaches on the understanding of bacterial antibiotic resistance genes in a freshwater reservoir. *Science of the Total Environment* 693. doi: 10.1016/j.scitotenv.2019.133585.

Lopatkin, A.J., Sysoeva, T.A. and You, L. 2016. Dissecting the effects of antibiotics on horizontal gene transfer: Analysis suggests a critical role of selection dynamics. *BioEssays* 38(12). doi: 10.1002/bies.201600133.

Lu, W., Qiu, Q., Chen, K., Zhao, R., Li, Q. and Wu, Q. 2022. Distribution and Molecular Characterization of Functional Class 2 Integrons in Clinical Proteus mirabilis Isolates. *Infection and Drug Resistance* 15. doi: 10.2147/IDR.S347119.

Luk-in, S. et al. 2021. Occurrence of mcr-mediated colistin resistance in Salmonella clinical isolates in Thailand. *Scientific Reports* 11(1). doi: 10.1038/s41598-021-93529-6.

Luo, Y., Mao, D., Rysz, M., Zhou, Q., Zhang, H., Xu, L. and Alvarez, P.J.J. 2010. Trends in antibiotic resistance genes occurrence in the Haihe River, China. *Environmental Science*

and Technology 44(19). doi: 10.1021/es100233w.

Ma, L., Li, A.D., Yin, X. Le and Zhang, T. 2017. The Prevalence of Integrons as the Carrier of Antibiotic Resistance Genes in Natural and Man-Made Environments. *Environmental Science and Technology* 51(10). doi: 10.1021/acs.est.6b05887.

Ma, T., McAllister, T.A. and Guan, L.L. 2021. A review of the resistome within the digestive tract of livestock. *Journal of Animal Science and Biotechnology* 12(1). doi: 10.1186/s40104-021-00643-6.

Machado, E., Cantón, R., Baquero, F., Galán, J.C., Rollán, A., Peixe, L. and Coque, T.M. 2005. Integron content of extended-spectrum-β-lactamase-producing Escherichia coli strains over 12 years in a single hospital in Madrid, Spain. *Antimicrobial Agents and Chemotherapy* 49(5). doi: 10.1128/AAC.49.5.1823-1829.2005.

Majlander, J., Anttila, V.J., Nurmi, W., Seppälä, A., Tiedje, J. and Muziasari, W. 2021. Routine wastewater-based monitoring of antibiotic resistance in two Finnish hospitals: focus on carbapenem resistance genes and genes associated with bacteria causing hospital-acquired infections. *Journal of Hospital Infection* 117. doi: 10.1016/j.jhin.2021.09.008.

Manaia, C.M. et al. 2018. Antibiotic resistance in wastewater treatment plants: Tackling the black box. *Environment International* 115. doi: 10.1016/j.envint.2018.03.044.

Manoharan, R.K., Srinivasan, S., Shanmugam, G. and Ahn, Y.H. 2021. Shotgun metagenomic analysis reveals the prevalence of antibiotic resistance genes and mobile genetic elements in full scale hospital wastewater treatment plants. *Journal of Environmental Management* 296. doi: 10.1016/j.jenvman.2021.113270.

Manor, O. and Borenstein, E. 2015. MUSiCC: A marker genes based framework for metagenomic normalization and accurate profiling of gene abundances in the microbiome. *Genome Biology* 16(1). doi: 10.1186/s13059-015-0610-8.

Markkanen, M.A. et al. 2023. Metagenomic Analysis of the Abundance and Composition of Antibiotic Resistance Genes in Hospital Wastewater in Benin, Burkina Faso, and Finland. Castanheira, M. ed. *mSphere* 8(1). Available at: https://journals.asm.org/doi/10.1128/msphere.00538-22 [Accessed: 10 March 2023].

Márquez, C. et al. 2008. Urinary tract infections in a South American population: Dynamic spread of class 1 integrons and multidrug resistance by homologous and site-specific recombination. *Journal of Clinical Microbiology* 46(10). doi: 10.1128/JCM.00835-08.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing

reads. *EMBnet.journal* 17(1). doi: 10.14806/ej.17.1.200.

Mazel, D. 2006. Integrons: Agents of bacterial evolution. *Nature Reviews Microbiology* 4(8). doi: 10.1038/nrmicro1462.

Mazel, D., Dychinco, B., Webb, V.A. and Davies, J. 2000. Antibiotic resistance in the ECOR collection: Integrons and identification of a novel aad gene. *Antimicrobial Agents and Chemotherapy* 44(6). doi: 10.1128/AAC.44.6.1568-1574.2000.

McKenna, A. et al. 2020. Impact of industrial production system parameters on chicken microbiomes: Mechanisms to improve performance and reduce Campylobacter. *Microbiome* 8(1). doi: 10.1186/s40168-020-00908-8.

McKew, B.A. and Smith, C.J. 2015. Real-Time PCR Approaches for Analysis of Hydrocarbon-Degrading Bacterial Communities. doi: 10.1007/8623_2015_64.

McKinney, C.W., Dungan, R.S., Moore, A. and Leytem, A.B. 2018. Occurrence and abundance of antibiotic resistance genes in agricultural soil receiving dairy manure. *FEMS Microbiology Ecology* 94(3). doi: 10.1093/femsec/fiy010.

Messier, N. and Roy, P.H. 2001. Integron integrases possess a unique additional domain necessary for activity. *Journal of Bacteriology* 183(22). doi: 10.1128/JB.183.22.6699-6706.2001.

Miłobedzka, A. et al. 2022. Monitoring antibiotic resistance genes in wastewater environments: The challenges of filling a gap in the One-Health cycle. *Journal of Hazardous Materials* 424. doi: 10.1016/j.jhazmat.2021.127407.

Mobaraki, S., Aghazadeh, M., Soroush Barhaghi, M.H., Yousef Memar, M., Goli, H.R., Gholizadeh, P. and Samadi Kafil, H. 2018. Prevalence of integrons 1, 2, 3 associated with antibiotic resistance in Pseudomonas aeruginosa isolates from Northwest of Iran. *BioMedicine (France)* 8(1). doi: 10.1051/bmdcn/2018080102.

Mukherjee, S. 2017. Emerging infectious diseases: Epidemiological perspective. In: *Indian Journal of Dermatology*. doi: 10.4103/ijd.IJD_379_17.

Muralikrishna, I. V and Manickam, V. 2017. Chapter Twelve - Wastewater Treatment Technologies. In: Muralikrishna, I. V and Manickam, V. B. T.-E. M. eds. Butterworth-Heinemann, pp. 249–293. Available at: https://www.sciencedirect.com/science/article/pii/B9780128119891000129.

Murray, C.J. et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a

systematic analysis. *The Lancet* 399(10325). doi: 10.1016/S0140-6736(21)02724-0.

Muziasari, W.I. et al. 2014. Sulphonamide and trimethoprim resistance genes persist in sediments at Baltic Sea aquaculture farms but are not detected in the surrounding environment. *PLoS ONE* 9(3). doi: 10.1371/journal.pone.0092702.

Muziasari, W.I. et al. 2016. Aquaculture changes the profile of antibiotic resistance and mobile genetic element associated genes in Baltic Sea sediments. *FEMS Microbiology Ecology* 92(4). doi: 10.1093/femsec/fiw052.

Najafgholizadeh Pirzaman, A. and Mojtahedi, A. 2019. Investigation of antibiotic resistance and the presence of integron genes among ESBL producing Klebsiella isolates. *Meta Gene* 19. doi: 10.1016/j.mgene.2018.10.008.

Nepal, G. and Bhatta, S. 2018. Self-medication with Antibiotics in WHO Southeast Asian Region: A Systematic Review. *Cureus*. doi: 10.7759/cureus.2428.

Ng, L.K., Mulvey, M.R., Martin, I., Peters, G.A. and Johnson, W. 1999. Genetic characterization of antimicrobial resistance in Canadian isolates of Salmonella serovar Typhimurium DT104. *Antimicrobial Agents and Chemotherapy* 43(12). doi: 10.1128/aac.43.12.3018.

Nguyen, A.Q. et al. 2021. Monitoring antibiotic resistance genes in wastewater treatment: Current strategies and future challenges. *Science of the Total Environment* 783. doi: 10.1016/j.scitotenv.2021.146964.

Nikibakhsh, M., Firoozeh, F., Badmasti, F., Kabir, K. and Zibaei, M. 2021. Molecular study of metallo-β-lactamases and integrons in Acinetobacter baumannii isolates from burn patients. *BMC Infectious Diseases* 21(1). doi: 10.1186/s12879-021-06513-w.

O' Neill, J. 2014. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations The Review on Antimicrobial Resistance Chaired. (December).

Oksanen, J. et al. 2022a. vegan: Community Ecology Package. R package version 2.6-4. *Community ecology package* 2.6-4, pp. 1–297. Available at: https://cran.r-project.org/web/packages/vegan/vegan.pdf [Accessed: 30 November 2022].

Oksanen, J. et al. 2022b. vegan: Community Ecology Package. R package version 2.6-4. *Community ecology package* 2.6-4, pp. 1–297.

Oliphant, C.M. and Eroschenko, K. 2015. Antibiotic resistance, Part 1: Gram-positive pathogens. *Journal for Nurse Practitioners* 11(1). doi: 10.1016/j.nurpra.2014.09.018.

Orman, B.E. et al. 2002. Evolution of multiresistance in nontyphoid Salmonella serovars from 1984 to 1998 in Argentina. *Antimicrobial Agents and Chemotherapy* 46(12). doi: 10.1128/AAC.46.12.3963-3970.2002.

Paiva, M.C., Ávila, M.P., Reis, M.P., Costa, P.S., Nardi, R.M.D. and Nascimento, A.M.A. 2015. The microbiota and abundance of the class 1 integron-integrase gene in tropical sewage treatment plant influent and activated sludge. *PLoS ONE* 10(6). doi: 10.1371/journal.pone.0131532.

Pal, C., Asiani, K., Arya, S., Rensing, C., Stekel, D.J., Larsson, D.G.J. and Hobman, J.L. 2017. Metal Resistance and Its Association With Antibiotic Resistance. In: *Advances in Microbial Physiology*. doi: 10.1016/bs.ampbs.2017.02.001.

Papp-Wallace, K.M., Endimiani, A., Taracila, M.A. and Bonomo, R.A. 2011. Carbapenems: Past, present, and future. *Antimicrobial Agents and Chemotherapy* 55(11). doi: 10.1128/AAC.00296-11.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25(7). doi: 10.1101/gr.186072.114.

Partridge, S.R., Kwong, S.M., Firth, N. and Jensen, S.O. 2018. Mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews* 31(4). doi: 10.1128/CMR.00088-17.

Partridge, S.R., Tsafnat, G., Coiera, E. and Iredell, J.R. 2009. Gene cassettes and cassette arrays in mobile resistance integrons: Review article. *FEMS Microbiology Reviews* 33(4). doi: 10.1111/j.1574-6976.2009.00175.x.

Paulus, G.K., Hornstra, L.M., Alygizakis, N., Slobodnik, J., Thomaidis, N. and Medema, G. 2019. The impact of on-site hospital wastewater treatment on the downstream communal wastewater system in terms of antibiotics and antibiotic resistance genes. *International Journal of Hygiene and Environmental Health* 222(4). doi: 10.1016/j.ijheh.2019.01.004.

Paulus, G.K., Hornstra, L.M. and Medema, G. 2020. International tempo-spatial study of antibiotic resistance genes across the Rhine river using newly developed multiplex qPCR assays. *Science of the Total Environment* 706. doi: 10.1016/j.scitotenv.2019.135733.

Petrovich, M.L. et al. 2020. Microbial and Viral Communities and Their Antibiotic Resistance Genes Throughout a Hospital Wastewater Treatment System. *Frontiers in Microbiology* 11. doi: 10.3389/fmicb.2020.00153.

Phongpaichit, S., Liamthong, S., Mathew, A.G. and Chethanond, U. 2007. Prevalence of class 1 integrons in commensal Escherichia coli from pigs and pig farmers in Thailand. *Journal of Food Protection* 70(2). doi: 10.4315/0362-028X-70.2.292.

Piyakul, C., Tiyawisutsri, R. and Boonbumrung, K. 2012. Emergence of metallo-β-lactamase IMP-14 and VIM-2 in Pseudomonas aeruginosa clinical isolates from a tertiary-level hospital in Thailand. *Epidemiology and Infection* 140(3). doi: 10.1017/S0950268811001294.

Ploy, M.C., Denis, F., Courvalin, P. and Lambert, T. 2000. Molecular characterization of integrons in Acinetobacter baumannii: Description of a hybrid class 2 integron. *Antimicrobial Agents and Chemotherapy* 44(10). doi: 10.1128/AAC.44.10.2684-2688.2000.

Polprasert, C., Koottatep, T. and Pussayanavin, T. 2018. Solar septic tanks: A new sanitation paradigm for Thailand 4.0. *ScienceAsia* 44. doi: 10.2306/scienceasia1513-1874.2018.44S.039.

Pongpech, P., Naenna, P., Taipobsakul, Y., Tribuddharat, C. and Srifuengfung, S. 2008. Prevalence of extended-spectrum beta-lactamase and class 1 integron integrase gene intl1 in Escherichia coli from thai patients and healthy adults. *Southeast Asian Journal of Tropical Medicine and Public Health* 39(3).

Price, M.N., Dehal, P.S. and Arkin, A.P. 2009. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* 26(7). doi: 10.1093/molbev/msp077.

Pruden, A., Vikesland, P.J., Davis, B.C. and de Roda Husman, A.M. 2021. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. *Current Opinion in Microbiology* 64. doi: 10.1016/j.mib.2021.09.013.

Quintela-Baluja, M. et al. 2019. Spatial ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. *Water Research* 162. doi: 10.1016/j.watres.2019.06.075.

Quintela-Baluja, M., Frigon, D., Abouelnaga, M., Jobling, K., Romalde, J.L., Gomez Lopez, M. and Graham, D.W. 2021. Dynamics of integron structures across a wastewater network – Implications to resistance gene transfer. *Water Research* 206. doi: 10.1016/j.watres.2021.117720.

Ramage, S., Camacho-Muñoz, D. and Petrie, B. 2019. Enantioselective LC-MS/MS for

anthropogenic markers of septic tank discharge. *Chemosphere* 219. doi: 10.1016/j.chemosphere.2018.12.007.

Ramírez, M.S. et al. 2010. Novel insights about class 2 integrons from experimental and genomic epidemiology. *Antimicrobial Agents and Chemotherapy* 54(2). doi: 10.1128/AAC.01392-08.

Rashid, A., Mirza, S.A., Keating, C., Ijaz, U.Z., Ali, S. and Campos, L.C. 2022. Machine Learning Approach to Predict Quality Parameters for Bacterial Consortium-Treated Hospital Wastewater and Phytotoxicity Assessment on Radish, Cauliflower, Hot Pepper, Rice and Wheat Crops. *Water (Switzerland)* 14(1). doi: 10.3390/w14010116.

Raza, S., Jo, H., Kim, J., Shin, H., Hur, H.G. and Unno, T. 2021. Metagenomic exploration of antibiotic resistome in treated wastewater effluents and their receiving water. *Science of the Total Environment* 765. doi: 10.1016/j.scitotenv.2020.142755.

Rizzo, L. et al. 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. *Science of the Total Environment* 447. doi: 10.1016/j.scitotenv.2013.01.032.

Robbins, S.J. et al. 2017. *GitHub - wwood/CoverM: Read coverage calculator for metagenomics*. Available at: https://github.com/wwood/CoverM [Accessed: 22 March 2023].

Rodríguez-Martínez, J.M., Velasco, C., García, I., Cano, M.E., Martínez-Martínez, L. and Pascual, A. 2007. Characterisation of integrons containing the plasmid-mediated quinolone resistance gene qnrA1 in Klebsiella pneumoniae. *International Journal of Antimicrobial Agents* 29(6). doi: 10.1016/j.ijantimicag.2007.02.003.

Rodriguez-Mozaz, S. et al. 2015. Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. *Water Research* 69. doi: 10.1016/j.watres.2014.11.021.

Rodríguez, E.A., Ramirez, D., Balcázar, J.L. and Jiménez, J.N. 2021. Metagenomic analysis of urban wastewater resistome and mobilome: A support for antimicrobial resistance surveillance in an endemic country. *Environmental Pollution* 276. doi: 10.1016/j.envpol.2021.116736.

Rosewarne, C.P., Pettigrove, V., Stokes, H.W. and Parsons, Y.M. 2010. Class 1 integrons in benthic bacterial communities: Abundance, association with Tn402-like transposition modules and evidence for coselection with heavy-metal resistance. *FEMS Microbiology*

Ecology 72(1). doi: 10.1111/j.1574-6941.2009.00823.x.

Rosser, S.J. and Young, H.K. 1999. Identification and characterization of class 1 integrons in bacteria from an aquatic environment. *Journal of Antimicrobial Chemotherapy* 44(1). doi: 10.1093/jac/44.1.11.

Roy, P.H., Partridge, S.R. and Hall, R.M. 2021. Comment on "Conserved phylogenetic distribution and limited antibiotic resistance of class 1 integrons revealed by assessing the bacterial genome and plasmid collection" by A.N. Zhang et al. *Microbiome* 9(1). doi: 10.1186/s40168-020-00950-6.

Rozman, U., Duh, D., Cimerman, M. and Turk, S.Š. 2020. Hospital wastewater effluent: Hot spot for antibiotic resistant bacteria. *Journal of Water Sanitation and Hygiene for Development* 10(2). doi: 10.2166/washdev.2020.086.

Ryu, S.H. et al. 2012. Antimicrobial resistance and resistance genes in Escherichia coli strains isolated from commercial fish and seafood. *International Journal of Food Microbiology* 152(1–2). doi: 10.1016/j.ijfoodmicro.2011.10.003.

Saiprom, N., Amornchai, P., Wuthiekanun, V., Day, N.P.J., Limmathurotsakul, D., Peacock, S.J. and Chantratita, N. 2015. Trimethoprim/sulfamethoxazole resistance in clinical isolates of Burkholderia pseudomallei from Thailand. *International Journal of Antimicrobial Agents* 45(5). doi: 10.1016/j.ijantimicag.2015.01.006.

Sandvang, D., Aarestrup, F.M. and Jensen, L.B. 1997. Characterisation of integrons and antibiotic resistance genes in Danish multiresistant Salmonella enterica Typhimurium DT104. *FEMS Microbiology Letters* 157(1). doi: 10.1016/S0378-1097(97)00473-4.

Sandvang, D., Diggle, M. and Platt, D.J. 2002. Translocation of integron-associated resistance in a natural system: Acquisition of resistance determinants by Inc P and Inc W plasmids from Salmonella enterica typhimurium DT104. *Microbial Drug Resistance* 8(3). doi: 10.1089/107662902760326850.

Sarmah, A.K., Meyer, M.T. and Boxall, A.B.A. 2006. A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. *Chemosphere* 65(5). doi: 10.1016/j.chemosphere.2006.03.026.

Satria, Y.A.A., Utami, M.S. and Prasudi, A. 2022. Prevalence of antibiotics prescription amongst patients with and without COVID-19 in low- and middle-income countries: a systematic review and meta-analysis. *Pathogens and Global Health*, pp. 1–13. Available at: https://doi.org/10.1080/20477724.2022.2160892.

Schmittgen, T.D. and Livak, K.J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nature Protocols* 3(6). doi: 10.1038/nprot.2008.73.

Sevillano, M. et al. 2020. Differential prevalence and host-association of antimicrobial resistance traits in disinfected and non-disinfected drinking water systems. *Science of the Total Environment* 749. doi: 10.1016/j.scitotenv.2020.141451.

shamsizadeh, Z., Ehrampoush, M.H., Nikaeen, M., Farzaneh mohammadi, Mokhtari, M., Gwenzi, W. and Khanahmad, H. 2021. Antibiotic resistance and class 1 integron genes distribution in irrigation water-soil-crop continuum as a function of irrigation water sources. *Environmental Pollution* 289. doi: 10.1016/j.envpol.2021.117930.

Sharma, M.K., Tyagi, V.K., Singh, N.K., Singh, S.P. and Kazmi, A.A. 2022. Sustainable technologies for on-site domestic wastewater treatment: a review with technical approach. *Environment, Development and Sustainability* 24(3). doi: 10.1007/s10668-021-01599-3.

Shibata, N. et al. 2003. PCR Typing of Genetic Determinants for Metallo-β-Lactamases and Integrases Carried by Gram-Negative Bacteria Isolated in Japan, with Focus on the Class 3 Integron. *Journal of Clinical Microbiology* 41(12). doi: 10.1128/JCM.41.12.5407-5413.2003.

Shlaes, D.M. and Bradford, P.A. 2018. Antibiotics—from there to where?: How the antibiotic miracle is threatened by resistance and a broken market and what we can do about it. *Pathogens and Immunity* 3(1). doi: 10.20411/pai.v3i1.231.

Siltrakool, B., Berrou, I., Griffiths, D. and Alghamdi, S. 2021. Antibiotics' use in thailand: Community pharmacists' knowledge, attitudes and practices. *Antibiotics* 10(2). doi: 10.3390/antibiotics10020137.

Skurnik, D. et al. 2005. Integron-associated antibiotic resistance and phylogenetic grouping of Echerichia coli isolates from healthy subjects free of recent antibiotic exposure. *Antimicrobial Agents and Chemotherapy* 49(7). doi: 10.1128/AAC.49.7.3062-3065.2005.

Smith, C.J., Nedwell, D.B., Dong, L.F. and Osborn, A.M. 2006. Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. *Environmental Microbiology* 8(5). doi: 10.1111/j.1462-2920.2005.00963.x.

Smith, C.J. and Osborn, A.M. 2009. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiology Ecology* 67(1). doi: 10.1111/j.1574-6941.2008.00629.x.

Smith, S.D. et al. 2022. Diversity of Antibiotic Resistance genes and Transfer Elements-Quantitative Monitoring (DARTE-QM): a method for detection of antimicrobial resistance in environmental samples. *Communications Biology* 5(1). doi: 10.1038/s42003-022-03155-9.

Soucy, S.M., Huang, J. and Gogarten, J.P. 2015. Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics* 16(8). doi: 10.1038/nrg3962.

Spencer, S.J. et al. 2016. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME Journal* 10(2), pp. 427–436. doi: 10.1038/ISMEJ.2015.124.

Spindler, A., Otton, L.M., Fuentefria, D.B. and Corção, G. 2012. Beta-lactams resistance and presence of class 1 integron in Pseudomonas spp. isolated from untreated hospital effluents in Brazil. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* 102(1). doi: 10.1007/s10482-012-9714-2.

Stalder, T., Barraud, O., Casellas, M., Dagot, C. and Ploy, M.C. 2012. Integron involvement in environmental spread of antibiotic resistance. *Frontiers in Microbiology* 3(APR). doi: 10.3389/fmicb.2012.00119.

Stalder, T., Barraud, O., Jové, T., Casellas, M., Gaschet, M., Dagot, C. and Ploy, M.C. 2014. Quantitative and qualitative impact of hospital effluent on dissemination of the integron pool. *ISME Journal* 8(4). doi: 10.1038/ismej.2013.189.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9). doi: 10.1093/bioinformatics/btu033.

Stedtfeld, R.D. et al. 2018. Primer set 2.0 for highly parallel qpcr array targeting antibiotic resistance genes and mobile genetic elements. *FEMS Microbiology Ecology* 94(9). doi: 10.1093/femsec/fiy130.

Su, H.C., Ying, G.G., Tao, R., Zhang, R.Q., Zhao, J.L. and Liu, Y.S. 2012. Class 1 and 2 integrons, sul resistance genes and antibiotic resistance in Escherichia coli isolated from Dongjiang River, South China. *Environmental Pollution* 169. doi: 10.1016/j.envpol.2012.05.007.

Su, J., Shi, L., Yang, L., Xiao, Z., Li, X. and Yamasaki, S. 2006. Analysis of integrons in clinical isolates of Escherichia coli in China during the last six years. *FEMS Microbiology Letters* 254(1). doi: 10.1111/j.1574-6968.2005.00025.x.

Su, Z., Li, A., Chen, J., Huang, B., Mu, Q., Chen, L. and Wen, D. 2020. Wastewater

discharge drives ARGs spread in the coastal area: A case study in Hangzhou Bay, China. *Marine Pollution Bulletin* 151. doi: 10.1016/j.marpolbul.2019.110856.

Sulayyim, H.J., Ismail, R., Hamid, A.A. and Ghafar, N.A. 2022. Antibiotic Resistance during COVID-19: A Systematic Review. *International Journal of Environmental Research and Public Health* 19(19). doi: 10.3390/ijerph191911931.

Sultan, I., Rahman, S., Jan, A.T., Siddiqui, M.T., Mondal, A.H. and Haq, Q.M.R. 2018. Antibiotics, resistome and resistance mechanisms: A bacterial perspective. *Frontiers in Microbiology* 9(SEP). doi: 10.3389/fmicb.2018.02066.

Sun, W., Qian, X., Gu, J., Wang, X.J. and Duan, M.L. 2016. Mechanism and Effect of Temperature on Variations in Antibiotic Resistance Genes during Anaerobic Digestion of Dairy Manure. *Scientific Reports* 6. doi: 10.1038/srep30237.

Suzuki, M.T., Taylor, L.T. and DeLong, E.F. 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. *Applied and Environmental Microbiology* 66(11). doi: 10.1128/AEM.66.11.4605-4614.2000.

Sydenham, T. V., Overballe-Petersen, S., Hasman, H., Wexler, H., Kemp, M. and Justesen, U.S. 2019. Complete hybrid genome assembly of clinical multidrug-resistant bacteroides fragilis isolates enables comprehensive identification of antimicrobial-resistance genes and plasmids. *Microbial Genomics* 5(11). doi: 10.1099/mgen.0.000312.

Tagliabue, A. and Rappuoli, R. 2018. Changing priorities in vaccinology: Antibiotic resistance moving to the top. *Frontiers in Immunology* 9(MAY). doi: 10.3389/fimmu.2018.01068.

Tatti, E., McKew, B.A., Whitby, C. and Smith, C.J. 2016. Simultaneous dna-rna extraction from coastal sediments and quantification of 16S rRNA genes and transcripts by real-time PCR. *Journal of Visualized Experiments* 2016(112). doi: 10.3791/54067.

Thakali, O., Brooks, J.P., Shahin, S., Sherchan, S.P. and Haramoto, E. 2020. Removal of antibiotic resistance genes at two conventional wastewater treatment plants of louisiana, USA. *Water (Switzerland)* 12(6). doi: 10.3390/W12061729.

Thongsamer, T. et al. 2021. Environmental antimicrobial resistance is associated with faecal pollution in Central Thailand's coastal aquaculture region. *Journal of Hazardous Materials* 416. doi: 10.1016/j.jhazmat.2021.125718.

Tian, Z., Chi, Y., Yu, B., Yang, M. and Zhang, Y. 2019. Thermophilic anaerobic digestion reduces ARGs in excess sludge even under high oxytetracycline concentrations.

Chemosphere 222. doi: 10.1016/j.chemosphere.2019.01.139.

Uddin, T.M. et al. 2021. Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. *Journal of Infection and Public Health* 14(12). doi: 10.1016/j.jiph.2021.10.020.

Uritskiy, G. V., DiRuggiero, J. and Taylor, J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6(1). doi: 10.1186/s40168-018-0541-1.

Větrovský, T. and Baldrian, P. 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE* 8(2). doi: 10.1371/journal.pone.0057923.

Walters, W.A., Caporaso, J.G., Lauber, C.L., Berg-Lyons, D., Fierer, N. and Knight, R. 2011. PrimerProspector: De novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* 27(8). doi: 10.1093/bioinformatics/btr087.

Wang, D. et al. 2017a. Genetic characterization of novel class 1 Integrons In0, In1069 and In1287 to In1290, and the inference of In1069-associated integron evolution in Enterobacteriaceae. *Antimicrobial Resistance and Infection Control* 6(1). doi: 10.1186/s13756-017-0241-9.

Wang, F.H., Qiao, M., Su, J.Q., Chen, Z., Zhou, X. and Zhu, Y.G. 2014. High throughput profiling of antibiotic resistance genes in urban park soils with reclaimed water irrigation. *Environmental Science and Technology* 48(16). doi: 10.1021/es502615e.

Wang, Q., Wang, P. and Yang, Q. 2018. Occurrence and diversity of antibiotic resistance in untreated hospital wastewater. *Science of the Total Environment* 621. doi: 10.1016/j.scitotenv.2017.10.128.

Wang, W. et al. 2021. Whole-Genome Sequencing and Machine Learning Analysis of Staphylococcus aureus from Multiple Heterogeneous Sources in China Reveals Common Genetic Traits of Antimicrobial Resistance. *mSystems* 6(3). doi: 10.1128/msystems.01185-20.

Wang, Y., Kong, B., Yang, W. and Zhao, X. 2017b. Correlation between class 1 integron of Escherichia coli and multidrug resistance in lower respiratory tract infection. *Journal of Infection in Developing Countries* 11(8). doi: 10.3855/jidc.8247.

Waseem, H. et al. 2019. Contributions and challenges of high throughput qPCR for determining antimicrobial resistance in the environment: A critical review. *Molecules* 24(1).

doi: 10.3390/molecules24010163.

Wei, Q., Jiang, X., Li, M., Chen, X., Li, G., Li, R. and Lu, Y. 2011. Transcription of integron-harboured gene cassette impacts integration efficiency in class 1 integron. *Molecular Microbiology* 80(5). doi: 10.1111/j.1365-2958.2011.07648.x.

White, P.A., McIver, C.J., Deng, Y.-M. and Rawlinson, W.D. 2000. Characterisation of two new gene cassettes, aadA5 and dfrA17 . *FEMS Microbiology Letters* 182(2). doi: 10.1111/j.1574-6968.2000.tb08906.x.

WHO. 2017. *WHO publishes list of bacteria for which new antibiotics are urgently needed*. Available at: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed [Accessed: 20 March 2023].

WHO. 2021a. *2021 AWaRe classification*. Available at: https://www.who.int/publications/i/item/2021-aware-classification [Accessed: 10 March 2023].

WHO. 2021b. *Antimicrobial resistance*. Available at: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance [Accessed: 25 March 2023].

WHO. 2022. *Global Tuberculosis Report 2022*. Available at: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022 [Accessed: 9 March 2023].

Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York*.

Wickham, H. 2016. Package `ggplot2`: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

Wilmotte, A., Van der Auwera, G. and De Wachter, R. 1993. Structure of the 16 S ribosomal RNA of the thermophilic cyanobacterium chlorogloeopsis HTF ('mastigocladus laminosus HTF') strain PCC7518, and phylogenetic analysis. *FEBS Letters* 317(1–2). doi: 10.1016/0014-5793(93)81499-P.

Wilson, D.N., Hauryliuk, V., Atkinson, G.C. and O'Neill, A.J. 2020. Target protection as a key antibiotic resistance mechanism. *Nature Reviews Microbiology* 18(11). doi: 10.1038/s41579-020-0386-z.

Von Wintersdorff, C.J.H. et al. 2016. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Frontiers in Microbiology* 7(FEB). doi:

10.3389/fmicb.2016.00173.

Wongburi, P. and Park, J.K. 2018. Decision making tools for selecting sustainable wastewater treatment technologies in Thailand. In: *IOP Conference Series: Earth and Environmental Science*. doi: 10.1088/1755-1315/150/1/012013.

Wright, M.S., Baker-Austin, C., Lindell, A.H., Stepanauskas, R., Stokes, H.W. and McArthur, J.V. 2008. Influence of industrial contamination on mobile genetic elements: Class 1 integron abundance and gene cassette structure in aquatic bacterial communities. *ISME Journal* 2(4). doi: 10.1038/ismej.2008.8.

Xu, H., Davies, J. and Miao, V. 2007. Molecular characterization of class 3 integrons from Delftia spp. *Journal of Bacteriology* 189(17). doi: 10.1128/JB.00348-07.

Xu, L., Ouyang, W., Qian, Y., Su, C., Su, J. and Chen, H. 2016. High-throughput profiling of antibiotic resistance genes in drinking water treatment plants and distribution systems. *Environmental Pollution* 213. doi: 10.1016/j.envpol.2016.02.013.

Xu, R., Zhang, Y., Xiong, W., Sun, W., Fan, Q. and Zhaohui Yang. 2020. Metagenomic approach reveals the fate of antibiotic resistance genes in a temperature-raising anaerobic digester treating municipal sewage sludge. *Journal of Cleaner Production* 277. doi: 10.1016/j.jclepro.2020.123504.

Yan, J.J., Hsueh, P.R., Lu, J.J., Chang, F.Y., Ko, W.C. and Wu, J.J. 2006. Characterization of acquired β-lactamases and their genetic support in multidrug-resistant Pseudomonas aeruginosa isolates in Taiwan: The prevalence of unusual integrons. *Journal of Antimicrobial Chemotherapy* 58(3). doi: 10.1093/jac/dkl266.

Yang, Y., Zhang, A.N., Che, Y., Liu, L., Deng, Y. and Zhang, T. 2021. Underrepresented high diversity of class 1 integrons in the environment uncovered by PacBio sequencing using a new primer. *Science of the Total Environment* 787. doi: 10.1016/j.scitotenv.2021.147611.

Yu, G., Smith, D.K., Zhu, H., Guan, Y. and Lam, T.T.Y. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8(1). doi: 10.1111/2041-210X.12628.

Yu, L. et al. 2012. Multiple antibiotic resistance of Vibrio cholerae serogroup O139 in China from 1993 to 2009. *PLoS ONE* 7(6). doi: 10.1371/journal.pone.0038633.

Zane, L.T. and Graeme, C.L. 2022. Tuberactinomycin antibiotics: Biosynthesis, anti-mycobacterial action, and mechanisms of resistance. *Frontiers in Microbiology* 13. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.961921.

Zhang, A.N., Li, L.-G., Ma, L., Gillings, M.R., Tiedje, J.M. and Zhang, T. 2018a. Conserved phylogenetic distribution and limited antibiotic resistance of class 1 integrons revealed by assessing the bacterial genome and plasmid collection. *Microbiome* 6(1), p. 130. doi: 10.1186/s40168-018-0516-2.

Zhang, D. et al. 2021. Metagenomic Survey Reveals More Diverse and Abundant Antibiotic Resistance Genes in Municipal Wastewater Than Hospital Wastewater. *Frontiers in Microbiology* 12. doi: 10.3389/fmicb.2021.712843.

Zhang, H. et al. 2004. Identification and characterization of class 1 integron resistance gene cassettes among Salmonella strains isolated from healthy humans in China. *Microbiology and Immunology* 48(9). doi: 10.1111/j.1348-0421.2004.tb03473.x.

Zhang, Q.Q., Tian, G.M. and Jin, R.C. 2018b. The occurrence, maintenance, and proliferation of antibiotic resistance genes (ARGs) in the environment: influencing factors, mechanisms, and elimination strategies. *Applied Microbiology and Biotechnology* 102(19). doi: 10.1007/s00253-018-9235-7.

Zhang, T., Yang, Y. and Pruden, A. 2015. Effect of temperature on removal of antibiotic resistance genes by anaerobic digestion of activated sludge revealed by metagenomic approach. *Applied Microbiology and Biotechnology* 99(18). doi: 10.1007/s00253-015-6688-9.

Zhang, Y. et al. 2022. Temperature affects variations of class 1 integron during sludge anaerobic digestion. *Bioresource Technology* 364, p. 128005.

Zhao, H., Chen, W., Xu, X., Zhou, X. and Shi, C. 2018. Transmissible ST3-IncHI2 plasmids are predominant carriers of diverse complex IS26-class 1 integron arrangements in multidrug-resistant Salmonella. *Frontiers in Microbiology* 9(OCT). doi: 10.3389/fmicb.2018.02492.

Zheng, J. et al. 2017. High-throughput profiling and analysis of antibiotic resistance genes in East Tiaoxi River, China. *Environmental Pollution* 230. doi: 10.1016/j.envpol.2017.07.025.

Zheng, W., Huyan, J., Tian, Z., Zhang, Y. and Wen, X. 2020. Clinical class 1 integron-integrase gene – A promising indicator to monitor the abundance and elimination of antibiotic resistance genes in an urban wastewater treatment plant. *Environment International* 135. doi: 10.1016/j.envint.2019.105372.

Zhou, M., Li, X., Hou, W., Wang, H., Paoli, G.C. and Shi, X. 2019. Incidence and

Characterization of Salmonella Isolates From Raw Meat Products Sold at Small Markets in Hubei Province, China. *Frontiers in Microbiology* 10. doi: 10.3389/fmicb.2019.02265.

Zhu, Y.G. et al. 2017. Continental-scale pollution of estuaries with antibiotic resistance genes. *Nature Microbiology* 2. doi: 10.1038/nmicrobiol.2016.270.

Zong, Z., Lü, X., Valenzuela, J.K., Partridge, S.R. and Iredell, J. 2008. An outbreak of carbapenem-resistant Acinetobacter baumannii producing OXA-23 carbapenemase in western China. *International Journal of Antimicrobial Agents* 31(1). doi: 10.1016/j.ijantimicag.2007.08.019.