# STUDY ON THE PREVELANCE OF ANTIBIOTIC RESISTANCE GENES FOLLOWING HURRICANE HARVEY

An Undergraduate Research Scholars Thesis

by

KELCI CHAMBERS

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisor: Dr. Jessica M. Labonté

May 2020

Major: Marine Biology
Marine Fisheries

# TABLE OF CONTENTS

# ABSTRACT

Study on the Prevalence of Antibiotic Resistant Gene Following Hurricane Harvey

Kelci Chambers
Department of Marine Biology
Texas A&M University

Research Advisor: Dr. Jessica Labonté
Department of Marine Biology
Texas A&M University

On August 25, 2017, Hurricane Harvey made landfall in Port Aransas, Texas and introduced 130 cm of rainfall to the Houston area. Hurricane Harvey's record rainfall caused city-wide flooding and large quantities of stormwater runoff, bringing an incredible amount of contaminated waters into Galveston Bay. The contaminated flood waters contained high levels of soil and sediment bacteria, as well as fecal contamination, which are usually associated with a higher prevalence of antibiotic resistance genes. Antibiotic resistance genes such as *vanA* (resistance to vancomycin), *sul1* (resistance to sulfonamides) and *tetA* (resistance to tetracycline) are commonly found throughout the environment due to the heavy usage of vancomycin, sulfonamide chemical groups, and tetracyclines antibiotics. Here, metagenomic data sampled every week for five weeks after Hurricane Harvey from a transect performed in Galveston Bay were mined for the presence of antibiotic resistance genes. The hypothesis was that there would be an increase in antibiotic resistance genes in the Bay right after Hurricane Harvey due to the input of terrestrial and wastewater-related microbes introduced into the ecosystem. Antibiotic resistance genes *sul1*, *vanA*, and *tetA* were detected at each sample station from the San Jacinto River to the Gulf of Mexico, with *sul1* being the most abundant antibiotic resistance genes

observed throughout the samples. Unexpectedly, there was no increase in the prevalence of antibiotic resistance genes after Hurricane Harvey. However, the genes found in the environment had a great diversity and differed from the ones found in clinical isolates, suggesting that there is a high potential for bacteria to acquire novel resistance genes. The data from this experiment will contribute to the understanding of the dispersal of antibiotic resistant genes in the environment after heavy rain events.

# DEDICATION

For Glenda Boyett, one of the strongest and motivational women I knew. Thank you for encouraging me to be something great.

# ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Labonté, for her guidance and encouragement throughout the course of this research. Thanks also goes to the members of the Labonté Viral Ecology Lab for their assistance and contributions to this research.

I would like to thank Texas Sea Grant for their contribution and providing me with the tools I needed to conduct this study. I would like to extend my gratitude to my friends and family for supporting me and keeping me motivated.

Finally, thanks to Fisher for his encouragement, dedication, and love.

# NOMENCLATURE

| | |
|---|---|
| ARGs | Antibiotic Resistant Genes |

# CHAPTER I

# INTRODUCTION

Since the 1930s, antibiotics have been used to treat bacterial infections and delay the growth or kill bacteria (Amabile-Cuevas et al., 1995). Antibiotics are natural defense mechanisms manufactured by soil bacteria and fungi. They occur naturally within the environment and are utilized by organisms when competition increases in the environment. However, microorganisms have gained resistance to the effects of antibiotics by developing bacterial defenses against the antimicrobial compound (Figure 1). Once acquired, the resistance genes can be passed down through horizontal and vertical gene transfer (Amabile-Cuevas et al., 1995). This resistance has been maintained in microorganisms by selective pressure provided by the widespread use of antibiotics (Wright, 2010). Antibiotic resistance genes (ARGs) can be disruptive to an environment and its microbial community, it can complicate treatment of bacterial infections, and pose potential risks for human health (Amabile-Cuevas et al., 1995). For this study three different antibiotic resistance genes (ARGs) were chosen: *vanA*, *tetA*, and *sul1*. These ARGs are commonly found in the environment, and are indicators of anthropogenic sources (Pingfeng et al., 2018).

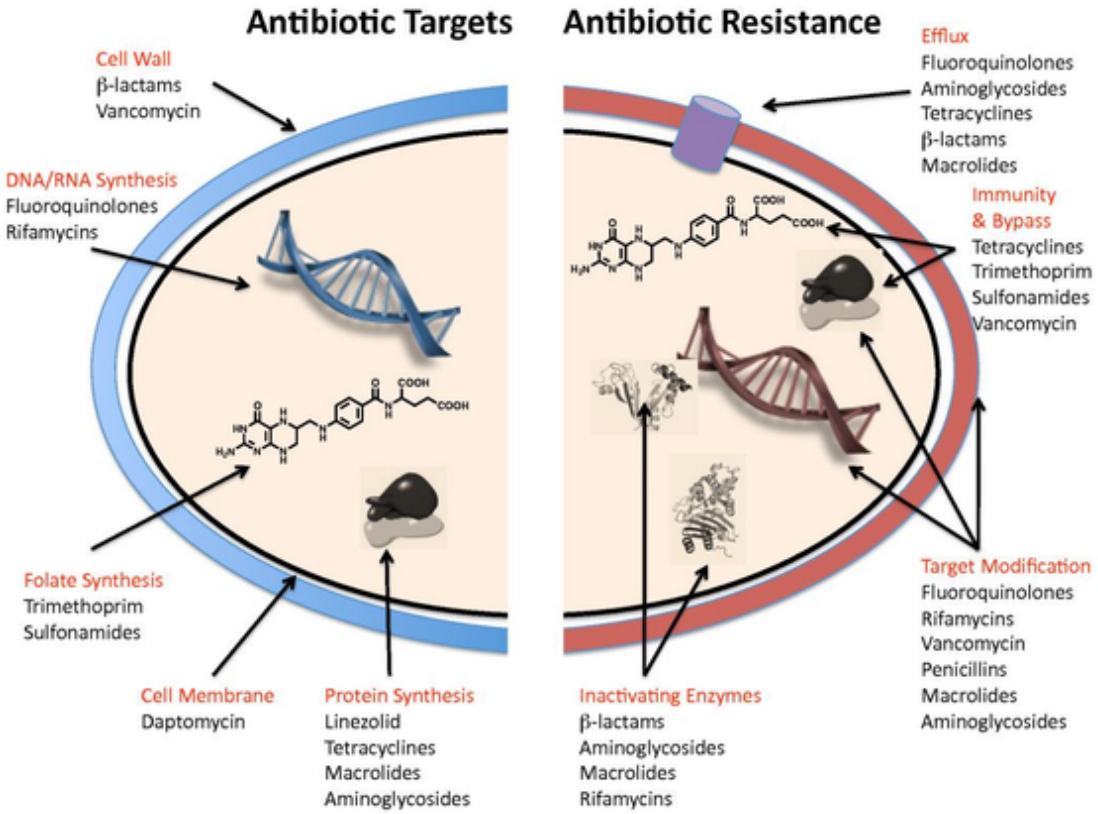



Figure 1. Antibiotic targets in the cell (left) and mechanisms of resistance against antibiotics (right) (Wright, 2010).

**Resistance to glycopeptides (*vanA*)**

The first glycopeptide antibiotic discovered was vancomycin in the late 1950s and is one of the only therapeutics currently used against Gram-positive microorganisms. Glycopeptide antibiotics such as vancomycin inhibits the bacterial cell wall formation by preventing peptidoglycan synthesis. Glycopeptides bind to the D-alanyl–D-alanine terminus of the cell wall peptidoglycan precursors and inhibit the transglycosylation reaction (Figure 1) (Van Hoek et al., 2011). However, when a microorganism possesses the gene *vanA*, modified peptidoglycan precursors are produced, making them resistant to glycopeptides like vancomycin (Figure 1) (Van Hoek et al., 2011). These modified precursors possess endings of D-Ala–D-Lac which

create a lower binding affinity for glycopeptides. *VanA* can be found on conjugative plasmids and is widespread in the environment (Berglund, 2015). Currently, resistance to vancomycin has been observed in six different Gram-positive bacterial genera: *Enterococcus, Erysipelothrix, Lactobacillus, Leuconostoc, Pediococcus, and Staphylococcus* (Van Hoek et al., 2011).

**Resistance to sulfonamides (*sul1*)**

Sulfonamides are synthetic antibiotics that contain sulfonamide chemical groups. Sulfonamides are the oldest introduced synthetic medicines and were first used in 1932 (Van Hoek et al., 2011). Sulfonamide antibiotics are considered bacteriostatic therefore these antibiotics can stop the reproduction of a bacteria without killing it (MacLeod, 1940). Also, sulfonamides competitively inhibit the enzyme dihydropteroate synthase which is required for thymine production and bacterial growth (Figure 1) (Van Hoek et al., 2011). Although sulfonamides are synthetic antibiotics, naturally occurring enzymes can degrade or modify the drug, making it ineffective. *Sul1* is a sulfonamide resistance gene that encodes forms of dihydropteroate synthase, which is not inhibited by antibiotics (Figure 1) (Van Hoek et al., 2011). *Sul1* is an example of an ARG that is widespread in the environment and, inside the cell, is often encoded by a conserved class 1 integrons, mobile genetic element (Berglund, 2015).

**Resistance to tetracyclines (*tetA*)**

Tetracycline antibiotics were first characterized in 1948 and are classified as either naturally occurring molecules or products of semi-synthetic approaches. Tetracyclines are considered to have broad spectrum activity against many gram-positive bacteria and are used widely across the world (Van Hoek et al., 2011). They inhibit protein synthesis by entering the bacterial cell and interacting with its ribosomes (Figure 1). Some tetracycline derivatives are poor inhibitors of protein synthesis and bind the ribosomes of the bacterial cell inefficiently or

not at all. Tetracycline resistance genes are chromosomally encoded but can be found on transposons and plasmids (Berglund, 2015). Tetracycline resistance genes have three different categories of resistance: energy-dependent efflux pumps, ribosomal protection proteins, and enzymatic inactivation (Figure 1). Ribosomal protection proteins are soluble cytoplasmic proteins that mediate tetracycline resistance (Connell et al., 2003). Enzymatic inactivation is the process referring to the point when an enzyme stops working. Enzymatic inactivation by tetracyclines resistance genes permanently eliminates tetracycline antibiotics (Markley & Wencewicz, 2018). Efflux pumps, which are transporters located in the membrane of cell, allow microorganisms to regulate their internal environment by pumping out toxic substances (Soto, 2013). The tetracycline resistance gene *tetA* encodes for an energy-dependent efflux pumps that decreases the accumulation of the antibiotic in whole cells (Van Hoek et al., 2011), and is commonly found among the bacterial genera: *Acinetobacter, Aeromonas, Bordetella, Chryseobacterium, Citrobacter, Edwardsiella, Enterobacter, Escherichia, Flavobacterium, Klebsiella, Laribacter, Plesiomonas, Proteus, Pseudomonas, Salmonella, Serratia, Shigella, Variovorax, Veillonella,* and *Vibrio* (Van Hoek et al., 2011).

**Prevalence of ARGs in the environment**

Antibiotic resistance genes are a naturally occurring phenomenon, but they can also be caused by anthropogenic sources. Berglund (2015) found that ARGs such as *tetA*, *vanA*, and *sul1* are naturally found within the environment, especially in surface waters and wastewater (Berglund, 2015). Their study also analyzed the prevalence of antibiotic contamination as an anthropogenic source and discovered high concentrations of *sul1* in a river downstream from a pharmaceutical formulation facility. Pingfeng et al. (2018) assessed the microbial water quality after Hurricane Harvey in flooded residential homes and bayous and demonstrated that the

microbial contamination within stagnant waters and sediments created a health risk for humans. They also revealed that anthropogenic indicator bacteria (*Escherichia coli*) and antibiotic resistance genes (*sul1*) increased post flooding (Pingfeng et al., 2018).

**Objectives and hypothesis**

On August 25, 2017, Hurricane Harvey made landfall in Port Aransas, Texas as a category 4 storm. Hurricane Harvey stalled over the greater Houston area from August 25 to August 29, and introduced 130 cm of rain fall, resulting in city-wide flooding. The flood waters contained high levels of soil and sediment bacteria, as well as, fecal contaminants (Pingfeng et al., 2018). These contaminants were most likely present due to the overflow of sewage from wastewater treatment plants. The introduction of foreign bacteria and contaminants creates the potential for an increase in antibiotic resistance genes (ARGs) (Kraemer et al., 2019). Indeed, antibiotic resistance genes can reduce diversity among microbial populations and alter the community structure leading to a reduction in biomass and microbial activity (Kraemer et al., 2019).

The objective of this study was to identify antibiotic resistance genes throughout Galveston Bay following Hurricane Harvey to assess their dispersal in order to obtain a better understanding of their prevalence in Galveston Bay. The central hypothesis was that there would be an increase in antibiotic resistance genes in Galveston Bay due to the Hurricane Harvey floodwater input. It was also expected that the *sul1* antibiotic resistance gene will be the most prevalent throughout the samples due to it being widespread in the environment.

# CHAPTER II

# METHODS

## Sample Collection

*Sample collection, sequencing, and metagenomic analysis were conducted prior to this thesis by members of the Labonté Viral Ecology Lab, especially Alaina Woods (MARB-IDP Master's student) and Jordan Walker, (UGRS Thesis entitled, Microbial Communities and Ecosystem Recovery of Galveston Bay Post-Hurricane Harvey)*. Samples were collected from four stations in Galveston Bay, after Hurricane Harvey's landfall, on September 6$^{th}$ (Sample 1 or H1), 9$^{th}$ (H3), 16$^{th}$ (H4), and 28$^{th}$ (H5) of 2017. A transect was sampled from the mouth of the San Jacinto River to the Gulf of Mexico (Figure 2). Control samples were collected from Texas A&M University at Galveston's boat basin on July 31$^{st}$ and August 22$^{nd}$, of 2017 prior to Hurricane Harvey's landfall. The volume of samples collected ranged from 4 to 20 L, depending on time constraints, available material, and personnel.

Samples were filtered in the field to remove small grazers and particles (nitex 30 mm filter). In the lab, the majority of the samples were filtered through glass fiber filters with a 0.7 mm pore-size followed by a polyvinylidene fluoride filter with a pore size of 0.22 mm. Due to the availability of supplies, the sample from September 9$^{th}$, was filtered with a 0.45 mm filter, and the virus concentrate was then filtered through 0.22 mm before long-term storage. After filtration, all filters were stored at -20°C. Viruses were concentrated using tangential flow filtration using a 30-kDa PrepScale cartridge (Suttle et al., 1991).

After processing, DNA was extracted from the filters using a standard phenol chloroform extraction protocol (Green and Sambrook, 2017). Samples were sent for sequencing using an

Illumina HiSeq 4000 instrument with 150 bp paired-end sequencing at the Texas A&M Genomics and Bioinformatics facility in College Station, TX.

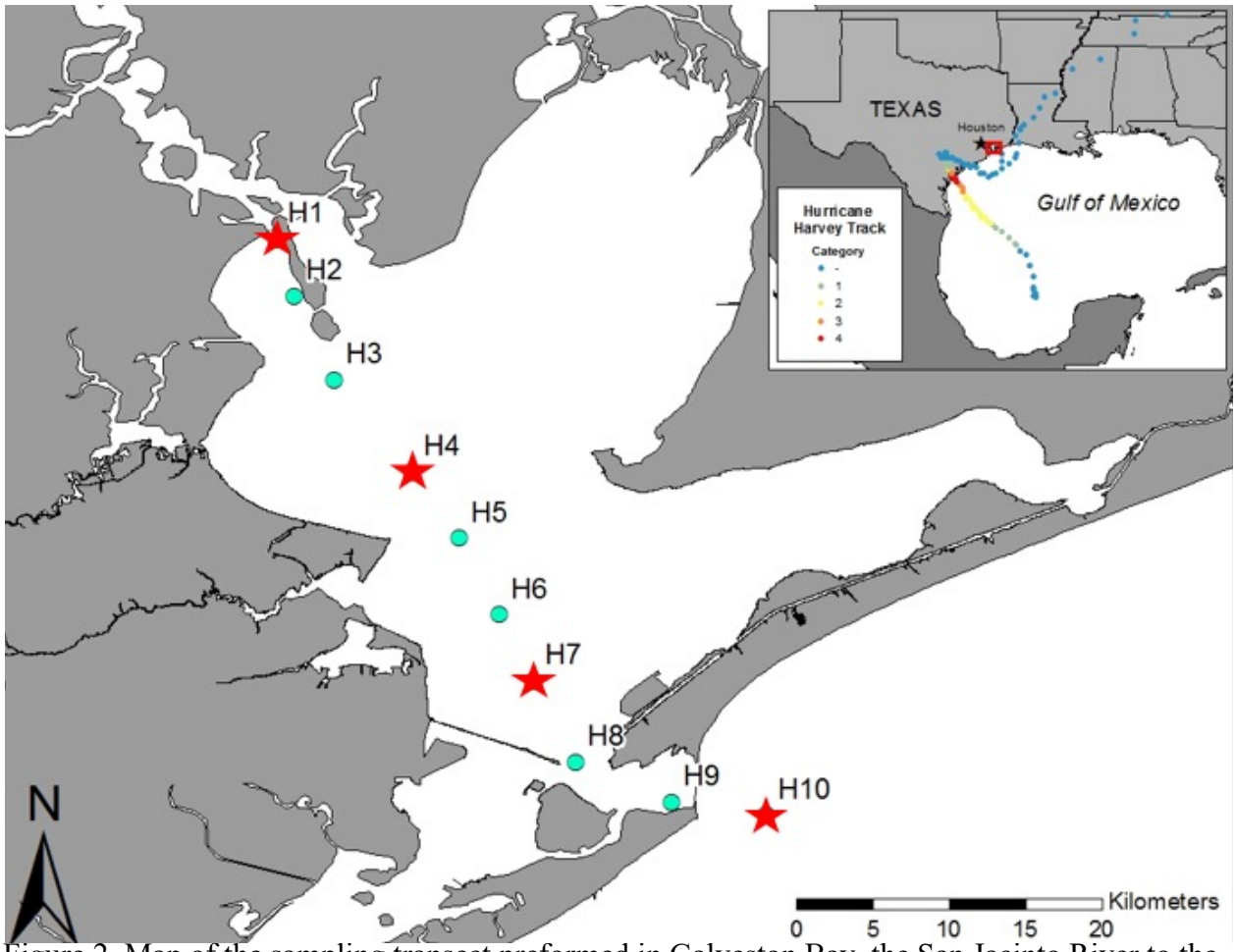



Figure 2. Map of the sampling transect preformed in Galveston Bay, the San Jacinto River to the mouth of the Gulf of Mexico. Samples from stations H1, H4, H7, and H10, marked with red stars, were used for this study.

**Metagenomic analysis**

BBtools software suite (https://sourceforge.net/projects/bbtools/) was used for quality control of the metagenomic data. BBduk (version 38.31) was used twice to remove byproducts of Illumina sequencing. The first BBduk command was used to filter out contaminated reads including kmers of 31 bp in length and dropping the extra base that is occasionally added by Illumina sequencing using the force trim Modulo (ftm) argument. The second BBduk command

used the Phred algorithm for more accurate adapter trimming and trims kmers between 11 and 23 bp in length- a common length for adapter artifacts. BBmerge (version 38.31) (Bushnell et al 2017) was used to merge the forward and reverse reads using default settings. BBmask (version 38.31) was used to soft-mask human, cat, dog and mouse contamination, using default settings. Megahit (version 1.1.3) (Li et al 2015) was then used with for *de novo* assembly of the metagenomic data. Gene prediction was then done with PRODIGAL (version 2.6.3) (Hyatt et al 2010).

**Antibiotic Resistance Gene Search**

The antibiotic resistance genes *vanA*, *tetA*, and *sul1* were chosen based on their prevalence in the environment, and they are considered indicators of anthropogenic source (Berglund, 2015). Reference sequences were chosen from GenBank by selecting the protein category and searching "*tetA*", "*vanA*", and "*sul1*". Eukaryote sequences, partial sequences, and plasmid sequences were not considered. A database for each ARG was created using makeblastdb command in blast+ (version 2.6.0). Then, the protein sequences from the Hurricane Harvey metagenomes were compared to the selected ARGs databases with BLASTp to recover the sequences similar to the ARGs. An e-value of $1x10^{-10}$ was use to allow for the detection of similarity to be less likely by accidental coincidence and -outfmt -6 parameter to create a tabular text-based output. The files were exported to Geneious, an integrated and expendable software platform for the organization and analysis of sequence data (Kearse et al, 2012), where the sequences with a hit to the known ARGs were extracted into a single file. An alignment was conducted for each gene and included the extracted Harvey sequences and the reference ARG sequences using MAFFT, a multiple sequence alignment program for unix-like operating systems (Katoh et al., 2009). The algorithm FFT-NS-1 was used for the alignments as it is a fast,

progressive method. Any sequences that were lacking significant portions of amino acids were removed from the alignment before all sequences were re-aligned. MEGAX, a sophisticated and user-friendly software suite that has provided tools for exploring, discovering, and analyzing DNA and protein sequences from an evolutionary perspective (Kumar et al., 2018), was used to find the best fit DNA/Protein model suitable for each ARG alignment. A neighbor joining tree using the Poisson model was constructed for each ARG. Each tree was resampled using the bootstrap method with a 1000 replicates and sorted topologies. The phylogenetic tree analyses for *sul1* and *vanA* had replicates that failed. In the *sul1* phylogenetic tree, 232 replicates failed to resample, and in the *vanA* phylogenetic tree, 34 replicates failed to resample. This could be due to an issue with correlation or constraints within the data. Once the tree was constructed, the branch names were edited according to the sample site and type. Each phylogenetic tree was exported as a Newick file to be viewed and manipulated in FigTree, a graphical viewer of phylogenetic trees (Rambaut, 2007).

# CHAPTER III

# RESULTS

## Incidence of resistance genes following Hurricane Harvey

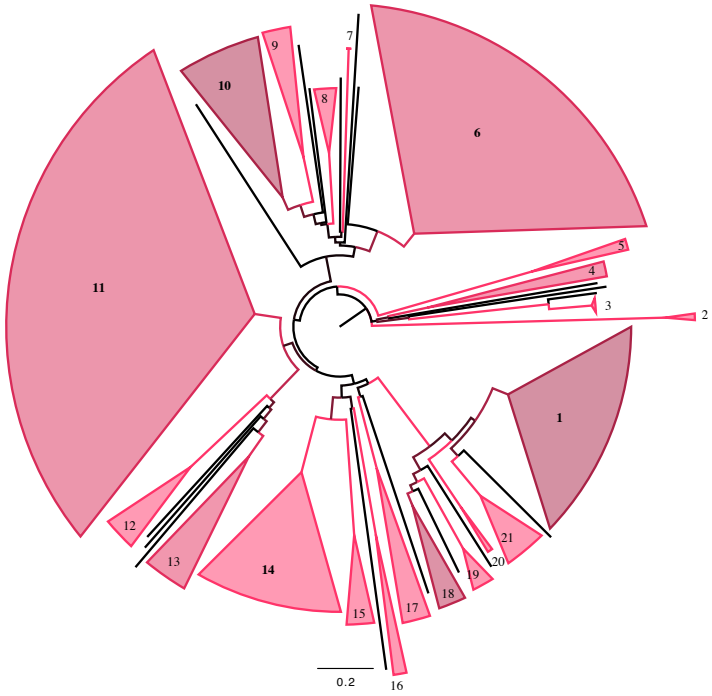
We identified the genes similar to known *vanA*, *tetA*, and *sul1* in marine metagenomes from samples following Hurricane Harvey along a transect from the San Jacinto River to the Gulf of Mexico (Table 1). A total of 322 sequences (frequency between 0.19 and 1.84 genes per 100,000 genes) were found to be similar to the 12 selected reference *tetA* protein sequences. A total of 440 sequences (0.99–1.95 genes per 100,000 genes) were found to be similar to the 10 selected reference *vanA* protein sequences. A total of 1,817 sequences (0.89–6.94 genes per 100,000 genes) were found to be similar to the 12 selected reference *sul1* protein sequences. The *sul1* resistance gene was to the most abundant throughout each sample stations (Table 1). The samples collected prior to Hurricane Harvey (labeled as VC for "virus concentrate") also shared a greater abundance of *sul1* antibiotic resistance genes (Table 1).

Table 1. Number of identified genes (normalized per 100,000 genes per sample) similar to *vanA*, *tetA*, and *sul1* before (VC) and after Hurricane Harvey in Galveston Bay.

| **Sample Stations** | **VC** | **S1H1** | **S1H4** | **S1H7** | **S1H10** | **S3H1** | **S3H4** | **S3H7** | **S3H10** | **S4H1** | **S4H4** | **S4H7** | **S4H10** | **S5H1** | **S5H4** | **S5H7** | **S5H10** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***tetA*** | 0.97 | 1.29 | 1.19 | 0.80 | 1.01 | 0.88 | 1.21 | 0.19 | 1.03 | 1.04 | 1.84 | 0.91 | 0.43 | 1.25 | 0.99 | 1.16 | 0.55 |
| ***vanA*** | 1.34 | 1.77 | 1.95 | 1.02 | 1.75 | 1.18 | 1.32 | 0.21 | 0.90 | 1.86 | 1.61 | 1.48 | 0.99 | 1.36 | 1.48 | 1.53 | 1.15 |
| ***sul1*** | 6.94 | 4.34 | 5.24 | 4.26 | 4.45 | 6.05 | 4.78 | 0.89 | 5.86 | 5.58 | 4.48 | 7.72 | 5.70 | 6.05 | 6.79 | 6.50 | 6.84 |

**Diversity and distribution of *tetA* genes**

A total of 322 sequences with similarity to *tetA* were found in our dataset. Phylogenetic analysis revealed that the genes were distant from the clinical isolates, group 2 and 6 (Figure 3). A total of 21 highly supported groups were defined based on the bootstrap values >50%. No trend in distribution was observed among the groups containing 10 sequences or more. Groups 1, 6, 11, and 14 contained sequences from each sample station (Figure 3). Group 10 contained mainly VC sequences, suggesting that these sequences were present before Hurricane Harvey. Group 11 was the only group to contain sequences from each sampling time, and group 6 was the only group that contained *tetA* reference sequences. VC sequences were observed in groups 1, 6, 10, and 11. Group 14 did not contain any VC sequences, suggesting that those sequences could have been brought in by Hurricane Harvey.

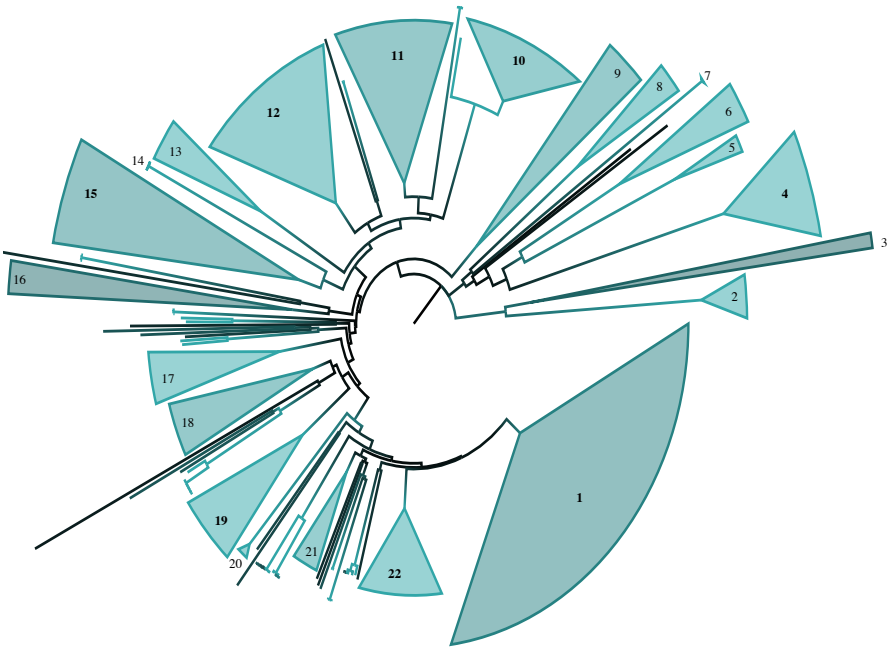



| | | | | |
|---|---|---|---|---|
| **Group 1:** | | | | |
| S1H1 – 4 | S3H1 – 2 | S4H4 – 6 | S5H4 - 2 | |
| S1H4 – 2 | S3H4 – 2 | S4H7 – 2 | S5H7 - 2 | |
| S1H7 – 1 | S3H7 – 2 | S4H10 - 2 | VC - 1 | |
| S1H10 - 1 | S4H1 – 7 | S5H1 - 5 | | |
| **Group 6:** | | | | |
| S1H1 – 1 | S3H4 – 6 | S4H4 – 5 | S5H4 – 2 | VC - 4 |
| S1H7 – 2 | S3H7 – 4 | S4H7 – 4 | S5H7 - 4 | |
| S1H10 – 3 | S3H10 – 4 | S4H10 – 1 | S5H10 - 5 | |
| S3H1 – 5 | S4H1 – 1 | S5H1 – 3 | tetA – 9 | |
| **Group 10:** | | | | |
| S1H1 – 1 | S3H7 – 1 | VC - 8 | | |
| S1H4 – 1 | S5H1 - 1 | | | |
| S1H7 -1 | S5H7 - 1 | | | |
| S3H4 – 1 | S5H10 - 1 | | | |
| **Group 11:** | | | | |
| S1H1 – 7 | S3H1 – 2 | S4H1 – 8 | S5H1 – 8 | VC - 5 |
| S1H4 – 3 | S3H4 – 3 | S4H4 – 12 | S5H4 - 7 | |
| S1H7 – 5 | S3H7 – 3 | S4H7 – 5 | S5H7 - 8 | |
| S1H10 – 3 | S3H10 – 3 | S4H10 – 3 | S5H10 - 3 | |
| **Group 14:** | | | | |
| S1H1 – 2 | S3H4 – 1 | S4H7 - 2 | S5H7 - 2 | |
| S1H4 – 3 | S3H7 – 1 | S4H10 - 2 | | |
| S1H7 – 2 | S4H1 - 1 | S5H1 - 2 | | |
| S1H10 – 4 | S4H4 - 5 | S5H4 – 2 | | |

Figure 3. Neighbor-joining phylogenetic with 1000 bootstrap replicates of 332 *tetA* sequences. The highly supported groups are denoted by the color pink.

**Diversity and distribution of *vanA* genes**

A total of 440 sequences with similarity to *vanA* were found in our dataset. Phylogenetic analysis revealed that the genes were distant from the clinical isolates, group 13 (Figure 4). A total of 22 highly supported groups were defined based on bootstrap values > 50%. No trend in distribution was observed among the groups containing 10 sequences or more. Sequences were found in the VC, or pre-Harvey, samples in each group except for groups 12, 13, 19, and 22. Suggesting that the sequences in these group may have been brought in by the storm. Groups 1, 11, 12, and 22 contained sequences from each sampling station with groups 1 containing sequences from every sampling time.




**Group 1:**

| | | | |
|---|---|---|---|
| S1H1 – 14 | S3H1 – 8 | S4H1 - 9 | S5H1 - 4 |
| S1H4 - 10 | S3H4 – 8 | S4H4 - 7 | S5H4 - 3 |
| S1H7 - 6 | S3H7 – 7 | S4H7 - 3 | S5H7 - 4 |
| S1H10 - 14 | S3H10 – 2 | S4H10 - 1 | S5H10-3 |
| VC - 1 | | | |

**Group 4:**

| | | | |
|---|---|---|---|
| S1H7 – 1 | S4H10 – 1 | S5H10 - 2 | |
| S4H1 – 2 | S5H1 – 2 | VC - 3 | |
| S4H4 – 1 | S5H4 – 1 | | |
| S4H7 – 1 | S5H7 – 3 | | |

**Group 10:**

| | | | |
|---|---|---|---|
| S3H1 – 1 | S4H10 – 2 | S5H10 - 10 | |
| S3H10 – 1 | S5H1- 2 | VC - 12 | |
| S4H4 – 2 | S5H4 - 5 | | |
| S4H7 – 3 | S5H7 – 4 | | |

**Group 11:**

| | | | |
|---|---|---|---|
| S1H1 – 1 | S3H1 – 1 | S4H4 - 2 | S5H4 - 1 |
| S1H4 – 3 | S3H4 – 21 | S4H7 – 2 | S5H7 - 1 |
| S1H7 – 2 | S3H7 – 2 | S4H10 – 4 | S5H10 - 1 |
| S1H10 – 3 | S4H1 – 2 | S5H1 – 3 | VC - 3 |

**Group 12:**

| | | | |
|---|---|---|---|
| S1H1 – 4 | S3H1 – 1 | S4H4 – 8 | S5H4 - 1 |
| S1H4 – 4 | S3H4 – 3 | S4H7 – 3 | S5H7 – 3 |
| S1H7 – 3 | S3H7 – 2 | S4H10 – 2 | S5H10 - 1 |
| S1H10 – 4 | S4H1 – 2 | S5H1 – 1 | |

**Group 13:**

vanA - 10

**Group 15:**

| | | | |
|---|---|---|---|
| S4H1 – 3 | S5H1 – 2 | S5H10 – 1 | |
| S4H7 – 1 | S5H4 – 1 | VC - 9 | |
| S4H10 – 1 | S5H7 – 1 | | |

**Group 19:**

| | | | |
|---|---|---|---|
| S3H1 – 1 | S4H1 – 3 | S5H4 - 1 | |
| S3H4 – 3 | S4H4 – 1 | S5H7 - 1 | |
| S3H7 – 2 | S4H10 - 1 | S5H10 - 2 | |
| S3H10 – 1 | S5H1 - 1 | | |

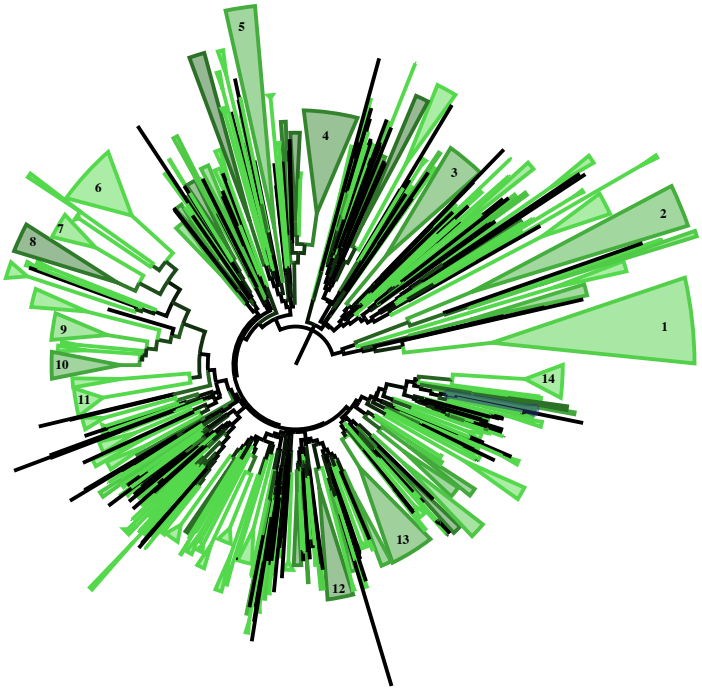
**Group 22:**

| | | | |
|---|---|---|---|
| S1H1 – 1 | S3H7 – 1 | S4H10 – 2 | S5H10 - 2 |
| S1H7 – 1 | S4H1 – 2 | S5H1 - 3 | |
| S3H1 – 2 | S4H4 – 4 | S5H4 - 3 | |
| S3H4 – 1 | S4H7 – 2 | S5H7 – 3 | |

Figure 4. Neighbor-joining phylogenetic with 1000 bootstrap replicates of 440 *vanA* sequences. The highly supported groups are denoted by the color blue.

**Diversity and distribution of *sul1* genes**

A total of 1,817 sequences with similarity to *sul1* were found in our dataset. Phylogenetic analysis revealed that the genes were distant from the clinical isolates, group 5 (Figure 5). A total of 183 highly supported groups were defined based on bootstrap values of >50%. No trend in distribution was observed among the groups containing 25 sequences or more. VC sequences were not observed in groups 5 and 10, suggesting that these sequences could have been brought in by the storm. Group 5 contained sequences from sample station S1 and S3, and all of the *sul1* reference sequences. Sequences for each sample station were observed in all groups except groups 5, 8, 10, and 12.




| | | | | |
|---|---|---|---|---|
| **Group 1:** | | | | |
| S1H7 – 3 | S4H10 – 8 | S3H10 – 1 | S5H7 - 7 | |
| S4H1 – 1 | S3H1 – 2 | S5H1 – 3 | S5H10 - 13 | |
| S4H7 – 11 | S3H7 – 1 | S5H4 – 1 | VC – 13 | |
| **Group 2:** | | | | |
| S1H1 – 2 | S1H10 – 1 | S3H7 – 2 | S4H7 – 2 | S5H4 - 3 |
| S1H4 – 3 | S3H1 – 1 | S4H1 – 1 | S4H10 – 1 | S5H7 - 2 |
| S1H7 – 1 | S3H4 – 2 | S4H4 – 2 | S5H1 – 2 | VC – 6 |
| **Group 3:** | | | | |
| S1H4 – 2 | S3H1 – 4 | S3H10 – 9 | S5H10 - 1 | |
| S1H7– 6 | S3H4 – 7 | S4H1 – 2 | VC - 2 | |
| S1H10 – 3 | S3H7 – 5 | S5H1 – 2 | | |
| **Group 4:** | | | | |
| S3H1 – 2 | S4H1 – 1 | S4H10 – 6 | S5H7 - 6 | |
| S3H7 – 1 | S4H4 – 3 | S5H1 – 7 | S5H10 - 9 | |
| S1H10 – 1 | S4H7 – 5 | S5H4 – 9 | VC – 14 | |
| **Group 5:** | | | | |
| S1H1 – 5 | S1H7 – 1 | S3H7 – 3 | sul1 - 12 | |
| S1H4 – 2 | S1H10 – 1 | S3H10 – 1 | | |
| **Group 6:** | | | | |
| S1H7 – 2 | S3H7 – 1 | S4H4 – 3 | S5H1 – 7 | S5H10 - 7 |
| S3H1 – 2 | S3H10 – 1 | S4H7 – 6 | S5H4 – 7 | VC - 5 |
| S3H4 – 1 | S4H1 – 7 | S4H10 – 9 | S5H7 – 7 | |
| **Group 7:** | | | | |
| S1H7 – 1 | S3H7 – 1 | S4H7 – 3 | S5H4 – 3 | VC - 3 |
| S3H1 – 2 | S3H10 – 1 | S4H10 – 2 | S5H7 - 3 | |
| S3H4 – 1 | S4H1 – 2 | S5H1 – 3 | S5H10 – 3 | |
| **Group 8:** | | | | |
| S4H7 – 5 | S4H10 – 9 | S5H7 – 2 | S5H10 – 4 | VC - 9 |
| **Group 9:** | | | | |
| S1H7 – 1 | S4H4 – 1 | S5H1 – 3 | S5H10 – 5 | |
| S3H1 – 1 | S4H7 – 2 | S5H4 - 5 | VC - 3 | |
| S4H1 – 3 | S4H10 – 5 | S5H7 – 4 | | |
| **Group 10:** | | | | |
| S3H1 – 2 | S4H4 – 2 | S5H1 – 6 | S5H10 - 3 | |
| S3H10 – 1 | S4H7 – 3 | S5H4 – 5 | | |
| S4H1 – 3 | S4H10 – 3 | S5H7 – 5 | | |
| **Group 11:** | | | | |
| S1H7 – 1 | S3H10 – 1 | S4H7 – 3 | S5H4 – 4 | VC - 2 |
| S3H1 – 2 | S4H1 – 4 | S4H10 – 2 | S5H7 - 5 | |
| S3H7 – 1 | S4H4 – 5 | S5H1 – 5 | S5H10 – 3 | |
| **Group 12:** | | | | |
| S3H1 – 3 | S3H10 – 1 | S4H7 – 1 | S5H4 – 1 | VC - 7 |
| S3H4 – 4 | S4H1 – 3 | S4H10 – 2 | S5H7 - 2 | |
| S3H7 – 3 | S4H4 – 2 | S5H1 – 2 | S5H10 – 2 | |
| **Group 13:** | | | | |
| S1H7 – 1 | S3H7 – 1 | S4H4 – 7 | S5H1 – 9 | S5H10 - 5 |
| S3H1 – 3 | S3H10 – 2 | S4H7 – 5 | S5H4 – 4 | VC - 2 |
| S3H4 – 3 | S4H1 – 8 | S4H10 – 5 | S5H7 - 5 | |
| **Group 14**: | | | | |
| S1H7 – 1 | S3H7 – 3 | S4H4 – 4 | S5H4 – 2 | |
| S3H1 – 4 | S3H10 – 2 | S4H7 – 4 | S5H7 - 1 | |
| S3H4 – 2 | S4H1 – 8 | S5H1 – 3 | S5H10 - 2 | |

Figure 5. Neighbor-joining phylogenetic with 1000 bootstrap replicates of 1,817 *sul1* sequences. The highly supported groups are denoted by the color green.

# CHAPTER IV

# DISCUSSION

The presence of antibiotic resistance genes was detected in Galveston Bay before and after Hurricane Harvey. The ARGs *vanA*, *tetA*, and *sul1* were all found at each sample station within the transect from the San Jacinto River to the Gulf of Mexico (Figure 1). The data revealed that the most abundant ARG was *sul1*, which is a known indicator of anthropogenic sources (Pingfeng et al., 2018). *Sul1*, encoded on a class 1 integron, is also widespread in the environment and can be found is in surface waters and wastewater (Berglund, 2015). The results also revealed that the ARGs were present in Galveston Bay before Hurricane Harvey, suggesting not all ARGs were brought by the storm but are always present. The prevalence of ARGs prior to Hurricane Harvey could be due anthropogenic sources, such as, wastewater, agricultural runoff, or could naturally be prevalent in Galveston Bay. The resistance genes *tetA* and *vanA* are found in surface waters and contaminated waters (wastewater) (Berglund, 2015). *VanA* is also widespread in the environment due to its widespread mobile genetic element (Tn*1546*) (Berglund, 2015).

**Abundance before and after Hurricane Harvey**

Over the course of the transect there was no change in the overall abundance of ARGs before and after the flooding event, but some groups of ARGs were most likely brought in by the storm. A similar study conducted by Pingfeng et al. (2018) observed the *sul1* and *E. coli* abundances in Buffalo Bayou and in closed indoor floodwaters before and after Hurricane Harvey. Their results revealed that *sul1* and *E. coli* abundance increased in both indoor floodwaters and Buffalo Bayou after flooding took place. Buffalo Bayou is located in the center

of Houston and has two wastewater treatment plant near it therefore the increase in *sul1* and *E. coli* may have resulted in direct discharge of untreated or partially treated wastewater. Galveston Bay has a greater surface area than Buffalo Bayou and the wastewater may have been diluted by the time it dispersed throughout the bay. The timing and location of collection may contribute to the increase they observed in their study, and the lack thereof in this study.

**Diversity and Distribution of ARGs in Galveston Bay**

For some groups of ARGs, there were no sequences recovered within the VC samples suggesting that the storm may have brought in those ARGs into Galveston Bay. Also, the *vanA* phylogenetic tree (Figure 4) contained a group with more sequences from sample station S1 than the other sample stations and groups. This could suggest that during collection week one *vanA* resistance genes were distributed into Galveston Bay via Hurricane Harvey. The genes found in the environment had a greater diversity that those found in clinical isolates, suggesting that there is a high potential for bacteria to acquire antibiotic resistance in the environment.

**Future work**

Due to the unforeseen events surrounding the COVID-19 virus in spring 2020, complete data analysis was unavailable at the time of publication for this URS thesis. This study could have consisted in the identification of the microbial genera that encode the ARGs. Currently, the most well studied resistant strains belong to the genera *Staphylococcus*, *Enterococcus*, *Escherichia*, *Klebsiella*, *Pseudomonas* (Van Hoek et al., 2011). We could have mined the metagenomic data to identify the species that carry each of the ARGs and determine if the groups follow a specific taxonomy or if they can be acquired by many different species. Due to the excess amount of wastewater from wastewater treatment plants, *Escherichia* and *Enterococcus* are often used as fecal contamination indicators. We could have searched in the metagenomic

data for the presence of *E. coli* and other fecal contaminants. The results of this data could also contribute to the prevalence of these ARGs in Galveston Bay and where they are most abundant. Moreover, we could have looked into more details into the dispersal of these genes to see if the same sequence was found in multiple samples over time.

# CHAPTER V

# CONCLUSION

Overall, the results identified *tetA*, *vanA* and *sul1* as antibiotic gene that are prevalent in Galveston Bay. These results indicate that these ARGs are either occurring natural or are a result of anthropogenic sources or both. The discovery of new groups of ARGs, and the further study on their dispersal and prevalence, will provide a better understanding the ecology and dissemination of these genes. Antibiotic resistance has made it difficult to treat bacterial infections, and new antibiotics are needed. This information regarding the distribution of ARGs could be essential to projects like the 10 x ’20 Initiative which are trying to develop 10 new antibacterial drugs by the year 2020 (Infectious Disease Society of America, 2010).

# REFERENCES


Amabile-Cuevas, C. F., Cardenas-Garcia, M., & Ludgar, M. (1995). Antibiotic resistance. *American Scientist*, *83*(4), 320-329.

Berglund, B. (2015). Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics. *Infection ecology & epidemiology*, *5*(1), 28564.

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge–accurate paired shotgun read merging via overlap. *PloS one*, *12*(10).

Connell, S. R., Tracz, D. M., Nierhaus, K. H., & Taylor, D. E. (2003). Ribosomal protection proteins and their mechanism of tetracycline resistance. *Antimicrobial agents and chemotherapy*, *47*(12), 3675-3681.

Green MR, Sambrook J. (2017). Isolation of high-molecular-weight DNA using organic solvents. *Cold Spring Harb Protoc* **2017**: 356–359.

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal prokaryotic dynamic programming genefinding algorithm. *Bmc Bioinformatics*, *11*(119), 10-1186.

Infectious Diseases Society of America. (2010). The 10×'20 Initiative: pursuing a global commitment to develop 10 new antibacterial drugs by 2020. *Clinical Infectious Diseases*, *50*(8), 1081-1083.

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, *30*(14), 3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

Kraemer, S. A., Ramachandran, A., & Perron, G. G. (2019). Antibiotic pollution in the environment: from microbial ecology to public policy. *Microorganisms*, *7*(6), 180.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6), 1547-1549.

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674-1676.

MacLeod, C. M. (1940). The inhibition of the bacteriostatic action of sulfonamide drugs by substances of animal and bacterial origin. *Journal of Experimental Medicine*, *72*(3), 217-232.

Markley, J. L., & Wencewicz, T. A. (2018). Tetracycline-inactivating enzymes. *Frontiers in microbiology*, *9*, 1058.

Suttle CA, Chan AM, Cottrell MT. (1991). Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Appl Environ Microbiol* **57**: 721–726.

Soto, S. M. (2013). Role of efflux pumps in the antibiotic resistance of bacteria embedded in a biofilm. *Virulence*, *4*(3), 223-229.

Van Hoek, A. H., Mevius, D., Guerra, B., Mullany, P., Roberts, A. P., & Aarts, H. J. (2011). Acquired antibiotic resistance genes: an overview. *Frontiers in microbiology*, *2*, 203.

Wright, G. D. (2010). Q&A: Antibiotic resistance: where does it come from and what can we do about it?. *BMC biology*, *8*(1), 123.

Yu, P., Zaleski, A., Li, Q., He, Y., Mapili, K., Pruden, A., ... & Stadler, L. B. (2018). Elevated levels of pathogenic indicator bacteria and antibiotic resistance genes after Hurricane Harvey's flooding in Houston. *Environmental Science & Technology Letters*, *5*(8), 481-486.