Metagenomics-based evaluation of groundwater microbial profiles in response to treated wastewater discharge

Noor Zaouri [1], Muhammad Raihan Jumat [2], Tariq Cheema [3] and Pei-Ying Hong [1*]

[1] Water Desalination and Reuse Center (WDRC), Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal-Jeddah 23955-6900, Saudi Arabia

[2] Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857

[3] Hydrology Department, Faculty of Earth Sciences, King Abdulaziz University (KAU), Jeddah, Saudi Arabia

* Corresponding author:

Pei-Ying Hong

Tel: +966-(12)-808-2218;

Email: peiying.hong@kaust.edu.sa

Highlights

- Metagenomics as a tool to assess groundwater quality
- Lower alpha diversity of bacteria and virus due to treated wastewater exposure
- Increase in relative abundance of Planctomycetes and Picornaviridae
- Antibiotic resistance genes may not be good markers to differentiate wastewater exposure

Abstract

This study aims to demonstrate the use of metagenomics to assess groundwater quality. Metagenomics revealed a lower alpha diversity for both bacteria and virus in wastewater-exposed groundwater compared to the upstream controls. However, an increase in the relative abundance of Planctomycetes and Picornaviridae, as well as increasing counts of unique antibiotic resistance genes, was observed in wastewater-exposed groundwater. Findings suggest that metagenomics can detect microbial signatures indicative of treated wastewater discharge.

Keywords: Shotgun metagenomics; aquifer recharge; groundwater quality; water reuse

1. Introduction

Water scarcity is projected to exacerbate by 2040, with 33 countries including Saudi Arabia predicted to face extreme water scarcity by that time (Luo et al., 2015). In recent years, recharging aquifer with treated wastewater has been proposed as a mitigation strategy to replenish the groundwater supplies. In Saudi Arabia, most wastewater treatment plants (WWTPs) are designed to achieve secondary or tertiary treatment with a final chlorination step prior to discharge. Such WWTPs typically achieve a cumulative 6-log reduction of microorganisms (Al-Jassim et al., 2015). Considering that untreated wastewater generally has a cell density of 10^{11} to 10^{12} cells/L (Manti et al., 2008), the final treated wastewater still contains a significant microbial load that may potentially perturb the indigenous groundwater microbial community. Aquifers are sensitive ecosystems, and therefore recharging aquifer with treated wastewater can only be adopted with an improved understanding on how these treated wastewater would affect the existing groundwater.

The current approaches to detecting changes in microbial community include qualitative approaches like terminal restriction fragment length polymorphism (T-RFLP) or quantitative approaches like real time PCR (qPCR). T-RFLP generates fingerprinting profiles for comparative analysis but does not provide information on the taxonomic identities of microbial contaminants. qPCR is a targeted approach that does not provide a comprehensive outlook on the wide variety of bacteria, virus and genes contaminants present in treated wastewater. The advent of high throughput untargeted sequencing approaches (e.g. metagenomics) have allowed a simultaneous evaluation on the bacteria, virus and genes present in an environmental sample (Quince et al., 2017). By comparing the diversity and relative abundance of these biological contaminants, one can denote if the groundwater has been perturbed by the treated wastewater.

This study aims to demonstrate the use of metagenomics as a tool to assess microbial water quality. Changes in the diversity and relative abundance of bacteria, virus and antibiotic resistance genes were monitored to denote the extent of perturbation imposed by treated wastewater.

2. Material and method

2.1. Description of sampling sites and sampling procedure

A subsurface reservoir, otherwise also known as Wadi Fatimah in Arabic, situated along the Yamaniyah-Az Zibarah-Fatimah drainage system in Saudi Arabia (Grainger, 1992) was identified for groundwater monitoring. Wadi Fatimah has been receiving treated wastewaters as part of a study commissioned in 2010-2016 by Makkah water utilities to recharge local aquifers. Groundwater at this site was collected over a period of 9 months, and include the following: wells 1, 15, 14 and 17 in March 2016; wells 1, 14 and 17 in August 2016; wells 12, 15 and 17 in December 2016 (Figure S1). Treated wastewaters injected to Wadi Fatimah were from a local WWTP upstream of well 1, and utilized schematics including clarifiers, activated sludge tanks, oxidation ponds and chlorine disinfection. Sampling wells varied throughout the sampling period because of various reasons including lack of sufficient water to facilitate representative sampling, groundwater well was closed for access or no permission was granted from the well owners on the day of sampling. Two wells situated approximately 60 km upstream of Wadi Fatimah were also identified, and groundwater was collected from these wells during the same time frame as mentioned above. Samples from these two wells serve as upstream controls since both the wells are not known to actively receive any treated wastewater discharge (Figure S1). The collected volume of all groundwater samples ranged from 20-40 L per well, and were

comprised of a single grab sample collected per well per sampling event. Before collection, the water was flushed for > 5 min to remove stagnant pools of water retained within the sampling pipes. Water was also flushed until no apparent turbidity was observed prior to collection in sterile polyethylene bottles. All samples were placed into a cooler and kept cool during transport. Upon arrival in laboratory, samples were stored in 4 °C prior to analyses.

2.2. Groundwater concentration and DNA extraction

Each 20-40 L of groundwater sample was passed through a filtration system fitted with a T-series cassette with a 100-kDa cut-off (approximately 0.005 µm) (Pall Corporation, Port Washington, NY, USA). This cassette has been shown previously to retain viruses at an efficiency of up to 87% and was particularly suited for large volume concentration (Jumat et al., 2017). Hence, this procedure was chosen to concentrate any microbial matter present in the low turbidity groundwater samples such as that collected in this study. Each groundwater sample was concentrated to 5 mL of retentate. To collect any microbial matter adhering to the 100 kDa cassette, the system was washed with 50 mL of wash solution (1 x Phosphate Buffered Saline (PBS) and 0.01% Tween 60). After circulating the wash solution through the filtration system for several minutes, the wash solution was itself concentrated to 5 mL. Both retentate and wash solutions derived from the same sample were extracted for DNA was by using the DNeasy Blood and Tissue Kit (Qiagen, Hilgen, Germany). MilliQ water was collected and subjected to the same filtration procedure at the same time when all groundwater samples were filtered. This acts as a negative control, and provide comparative analysis that would allow determination of microbial populations unique to the groundwater.

2.3. Metagenomic sequencing and analysis

Genomic DNA was sent to CosmosID (Rockville, MD, USA) for sequencing on HiSeq, as described (Hasan et al., 2014). Shotgun metagenomic datasets were directly analyzed by CosmosID metagenomic software (CosmosID Inc., Rockville, MD) as described elsewhere (Connelly et al., 2018; Hasan et al., 2014; Kaleko et al., 2016; Ponnusamy et al., 2016) to reveal associated microbial community composition. Briefly, the system utilizes a high performance data-mining k-mer algorithm and highly curated dynamic comparator databases (GenBook®) that rapidly disambiguate millions of short reads into the discrete genomes or genes engendering the particular sequences. The GeneBook databases were composed of over 150,000 microbial genomes and gene sequences representing over 1000 bacterial and 5000 viral species, as well as over 5500 antibiotic resistance genes. For each sample, the reads from a bacteria or viral species are assigned to the strain with the highest aggregation statistics, and subsequently determined for the corresponding alpha diversity. Community resistome was also identified using the CosmosID Inc. bioinformatics software package by querying the unassembled sequence reads against the CosmosID curated antibiotic resistance gene databases. Paired end Fastq files are deposited onto European Nucleotide Archive under accession number PRJEB32173.

2.4. Determination of organic water quality

Organic constituents of the DOC were fractionated according to molecular weight (MW) with size exclusion chromatography. The fraction distribution of the DOC in the groundwater was analyzed using Liquid Chromatography equipped with a macro-porous silica separation column, an Organic Carbon Detector (OCD), a UV 254 nm Detector (UVD) and an Organic Nitrogen Detector (OND) (LC-OCD Model 8, DOC LABOR, Germany). ChromCaluni software was used to determine the concentration of each fraction by integrating the area underneath the curve according to manufacture instructions.

3. Results and discussion

3.1. Taxonomic profile of bacterial communities

Phyla Proteobacteria, Actinobacteria and Bacteroidetes constitute the three dominant bacterial groups in the groundwater samples of both Wadi Fatimah and upstream controls. However, Proteobacteria and Actinobacteria were also present in high relative abundance in the negative MilliQ controls (Figure 1a). In contrast, Bacteroidetes was detected in higher relative abundance and in more consistent frequency in only the groundwater samples. Planctomycetes was detected in higher relative abundance among the Wadi Fatimah samples than that in upstream and negative controls (Figure 1a). Planctomycetes has been shown to be closely associated with organic detritus in the marine environment (Halliday et al., 2014). Similarly, a higher relative abundance of Planctomycetes was observed in groundwater samples that had a higher dissolved organic carbon (Table 1). Among these five detected DOC fractions, building block (BB) and humic substances (HS) were significantly higher in groundwater sampled from Wadi Fatimah compared to upstream controls ($p < 0.10$, Table 1). Specifically, HS which is predominated by aromatic moieties was only detected in Wadi Fatimah at concentrations ranging from 0.54 to 0.70 mg/L. Despite the higher DOC concentrations in Wadi Fatimah, the alpha diversity of bacterial community was lower in all of its samples compared to the controls (Figure 1b), suggesting the presence of an unknown factor that bias against certain families (e.g. Bradyrhizobiaceae, Methylobacteriaceae) in Wadi Fatimah. These families were only detected in the upstream controls during the sampling period. There was however, a positive trend of increasing alpha diversity with time, although the reason behind this trend remains unknown.

3.2. Community resistome

Ninety six unique ARGs were detected in Wadi Fatimah while only 56 unique ARGs were detected in the upstream controls (Figure 1c). The 5 most abundant unique ARGs detected in Wadi Fatimah were related to aminoglycoside, beta-lactam, macrolide and quinolone resistance, while the 5 most abundant ARGs detected in upstream controls also included resistance against beta-lactam. Shared ARGs between Wadi Fatimah and control samples revealed additional genes conferring resistance to tetracycline and trimethoprim (Table 2). A positive detection of ARGs in both Wadi Fatimah and control samples suggests a ubiquitous presence of ARGs in most environments, including pristine environment (Van Goethem et al., 2018). This may be related to long term global consumption of antibiotics for both hospital-associated and community-acquired infections (Versporten et al., 2018). Hence, the choice of ARGs as metagenomics markers to denote anthropogenic contamination events may not be provide definitive results.

3.3. Taxonomic profile of viral communities

Podoviridae and Siphoviridae account as the two most dominant viral families in the groundwater (Figure 2a). It was further observed that upstream controls had a significantly higher relative abundance of Siphoviridae, along with Myoviridae, Inoviridae and Tectiviridae ($p < 0.10$). All four families are mainly comprised of bacteriophages that rely on bacteria and archaea as their natural hosts. Along with one of the negative controls which denoted an exceedingly high relative abundance of Picornaviridae, Picornaviridae was consistently detected in all groundwater samples and especially at a higher relative abundance in Wadi Fatimah samples (average 45%) than in the upstream controls (average 11%). Picornaviridae comprises of several notable human enteric viruses including Enteroviruses and Salivirus which are proposed as potential human virus indicator (Reuter et al., 2017). The high relative abundance of Picornaviridae may be related to the type of treatment process used by the WWTP situated in

Wadi Fatimah. Secondary and tertiary treatment processes typically do not achieve high removal rates of viruses (Zhang and Farahbakhsh, 2007). In contrast, a WWTP retrofitted with membrane bioreactor reported only sporadic occurrences of Enterovirus (Jumat et al., 2017). The alpha diversity of viruses are higher in the controls than in Wadi Fatimah (Figure 2b). Coupled with the higher abundance of families associated with bacteriophages, the higher alpha diversity may be due to a wide variety of bacteriophages present in relatively unperturbed environments compared to those exposed to treated wastewater. Unlike the bacterial alpha diversity, the viruses did not exhibit an increasing trend of alpha diversity with time (Figure 2b).

4. Conclusion

Findings from this study demonstrates the use of metagenomics to provide a simultaneous evaluation of bacteria, virus and antibiotic resistome in groundwater exposed to treated wastewater discharge. It was observed that alpha diversity of both bacteria and viruses were lower in exposed groundwater than the upstream controls. In addition, Planctomycetes and Picornaviridae, both of which were detected by metagenomics simultaneously, may be appropriate markers to indicate increased organic content and presence of human fecal contamination. As more countries explore the use of treated wastewater to recharge diminishing groundwater levels, metagenomics could suffice as an appropriate monitoring tool to ensure that a wide diversity of biological contaminants are assessed and unintentional aberrations to the groundwater minimized. However, metagenomics alone is unable to provide insights to the potential risks that may arise from recharging aquifers with treated wastewater. This is because metagenomics analysis does not provide absolute quantification of biological contaminants and are derived from DNA template sequencing. Hence it is unable to differentiate the viable bacteria

populations against the dead and non-infectious ones. Instead, metagenomics can possibly narrow down the suite of contaminants to a selected few (e.g. Picornaviridae) for more in-depth evaluation by other quantitative methods like quantitative PCR or digital PCR. By coupling metagenomics with other quantitative methods, potential risks arising from recharging aquifer with treated wastewater can be better assessed.

Acknowledgements

The authors would like to thank Dr Moustapha Harb and Dhafer Alsaleh for assistance during the early sampling years. The authors would also like to thank Professor Tom Missimer and Mr. Mohammed AlRayaan for providing background information related to the hydrogeological characteristics of Wadi Fatimah. This study is funded by KAUST Baseline Grant BAS/1/1033-01-01 awarded to P.-Y. Hong.

Table 1. Summary of the organic matter components in Wadi Fatimah and upstream controls. DOC denotes dissolved organic carbon, BP denotes biopolymers, HS denotes humic substances, BB denotes building blocks, LMW denotes low molecular weight. Statistical comparison was made using t-test between Wadi Fatimah samples and upstream controls. ** denotes significant difference at 95% confidence level. * denotes significant difference at 90% confidence level. Value after ± denotes standard deviation.

		Wadi Fatimah			Upstream controls	P-value
		March 2016	August 2016	December 2016		
DOC	mg/L	8.59 ± 3.23	4.70 ± 0.89	6.12 ± 4.43	1.71 ± 2.1	0.08*
Carbon distribution						
BP	mg/L	0.08 ± 0.001	0.04 ± 0.01	0.07 ± 0.01	0.02 ± 0.05	0.17
HS	mg/L	0.70 ± 0.07	0.62 ± 0.02	0.54 ± 0.04	0.04 ± 0.05	0.01**
BB	mg/L	3.20 ± 0.15	2.58 ± 0.34	2.11 ± 0.09	0.68 ± 0.88	0.08*
LMW neutrals	mg/L	3.36 ± 0.33	0.58 ± 0.06	1.28 ± 0.11	0.14 ± 0.06	0.33
LMW acids	mg/L	1.13 ± 0.107	0.64 ± 0.04	1.56 ± 0.23	1.16 ± 1.64	0.95

Table 2. Antibiotic resistance genes detected in groundwater samples. * denotes that the average percentage matched reads were already corrected for the average percentage detected in MilliQ control. N.D. denotes not detected or below detection limits.

Top 5 identified ARGs unique to Wadi Fatimah	**Average % total matches ***	**Average % in MilliQ control**	**Denoted function**
Aminoglycoside *aph3' Ia*	0.03	N.D.	Aminoglycoside phosphotransferase gene, confers resistance by catalyzing the addition of phosphate from ATP to the 3'-hydroxyl group of a 4,6-disubstituted aminoglycoside
Aminoglycoside a*adA13*	0.04	N.D.	Aminoglycoside nucleotidyltransferase gene, confers resistance by adenylating the 2"-hydroxyl group of aminoglycoside
Beta-lactam-resistance bla_{CAU-1}	0.04	N.D.	Metallo-β-lactamase gene, confers resistance by catalyzing the hydrolysis of beta-lactam
Macrolide *ereA*	0.05	0.01	Erythromycin esterase gene, confers resistance through inactivation by hydrolyzing the lactone ring of the antibiotic
Plasmid-mediated-quinolone-resistance *qnrD1*	0.04		Quinolone resistance gene, confers resistance by preventing antibiotic binding
Top 5 identified ARGs unique to upstream controls	**Average % total matches ***	**Average % in MilliQ control**	**Denoted function**
Multidrug efflux pump *mdtE*	0.14	0.15	Part of the tripartite efflux system MdtEF-TolC, which confers resistance to compounds such as erythromycin, doxorubicin, benzalkonium and so on
Repressor for multidrug efflux pump *emrR*	0.12	0.13	Negatively regulates the transcription of the multidrug resistance pump-encoding operon, *emrRAB*, by binding to its regulatory region
Sensor kinase *cpxA*	0.11	0.13	Histidine kinase member of the two-component regulatory system CpxA/CpxR, responds to envelope stress response
Beta-lactam-resistance $bla_{OXA\ 281}$	0.10	N.D.	Oxacillinase (enzyme that confers resistance to oxacillin, a type of beta-lactam), presumably intrinsic to *Acinetobacter johnsonii* SH046
Beta-lactam-resistance $bla_{OXA\ 211}$	0.09	N.D.	Oxacillinase (enzyme that confers resistance to oxacillin, a type of beta-lactam), presumably intrinsic to *Acinetobacter johnsonii* SH046

ARGs identified in Wadi Fatimah and upstream controls	Average % total matches in Wadi Fatimah*	Average % total matches in upstream control*	Average % in MilliQ control	Denoted function
Beta-lactam-resistance $bla_{OXA-212}$	0.04	0.10	N.D.	Carbapenem-hydrolyzing class D beta-lactamase, confers resistance by catalyzing the hydrolysis of beta-lactam
Beta-lactam-resistance $bla_{OXA-373}$	0.03	0.14	N.D.	Functions as a transpeptidase, beta-lactam became suicide substrates of transpeptidases
Tetracycline *tet39*	0.04	0.16	0.05	Tetracycline resistance gene, confers resistance by active efflux mechanisms
Tetracycline *tetRG*	0.05	0.09	0.05	TetR is the repressor of the tetracycline resistance element, TetG is tetracycline efflux protein
Trimethoprim *dfr32*	0.05	0.03	0.12	Dihydrofolate reductase and trimethoprim resistance gene, confers resistance by replacing or substituting antibiotic action target

Figure legends

Figure 1. Bacterial communities and bacteria-associated antibiotic resistome in groundwater. **(a)** Heat map illustrates relative abundance of individual bacteria phyla in the negative controls (MilliQ water), upstream controls and wells in Wadi Fatimah across different sampling months. **(b)** Bacteria alpha diversity in groundwater sampled from Wadi Fatimah across different sampling months, upstream controls and negative controls. Linear line depicts an increasing trend in alpha diversity with time. **(c)** Venn diagram illustrating the number of unique antibiotic resistance genes detected in Wadi Fatimah (depicted as green) and upstream control samples (depicted as red). Overlapping region shows the number of shared antibiotic resistance genes detected in both Wadi Fatimah and upstream controls.

Figure 2. Viral communities in groundwater. **(a)** Heat map illustrates the relative abundance of individual viral families in the control (MilliQ water), upstream controls and wells in Wadi Fatimah across the different sampling times. **(b)** Virus alpha diversity in groundwater sampled from Wadi Fatimah across different sampling months, upstream controls and negative controls. Linear line depicts no apparent trend of increment with time.

References

Al-Jassim, N., et al., 2015. Removal of bacterial contaminants and antibiotic resistance genes by conventional wastewater treatment processes in Saudi Arabia: Is the treated wastewater safe to reuse for agricultural irrigation? Water Res. 73**,** 277-90.

Connelly, S., et al., 2018. Distinct consequences of amoxicillin and ertapenem exposure in the porcine gut microbiome. Anaerobe. 53**,** 82-93.

Grainger, D. J., 1992. A geological excursion in the Wadi Fatima area, near Jiddah, Saudi Arabia. Geology Today. 8**,** 215-219.

Halliday, E., et al., 2014. Comparison of bacterial communities in sands and water at beaches with bacterial water quality violations. PLoS One. 9**,** e90815.

Hasan, N. A., et al., 2014. Microbial Community Profiling of Human Saliva Using Shotgun Metagenomic Sequencing. Plos One. 9.

Jumat, M. R., et al., 2017. Membrane Bioreactor-Based Wastewater Treatment Plant in Saudi Arabia: Reduction of Viral Diversity, Load, and Infectious Capacity. Water. 9.

Kaleko, M., et al., 2016. Development of SYN-004, an oral beta-lactamase treatment to protect the gut microbiome from antibiotic-mediated damage and prevent Clostridium difficile infection. Anaerobe. 41**,** 58-67.

Luo, T., et al., Aqueduct Projected Water Stress Country Rankings. 2015. http://reliefweb.int/sites/reliefweb.int/files/resources/AQUEDUCT%20PROJECTED%20WATER%20STRESS%20Country%20Rankings.pdf. 31 August 2019

Manti, A., et al., 2008. Bacterial cell monitoring in wastewater treatment plants by flow cytometry. Water Environ Res. 80**,** 346-54.

Ponnusamy, D., et al., 2016. Cross-talk among flesh-eating *Aeromonas hydrophila* strains in mixed infection leading to necrotizing fasciitis. Proceedings of the National Academy of Sciences. 113**,** 722-727.

Quince, C., et al., 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology. 35**,** 833.

Reuter, G., et al., 2017. Saliviruses-the first knowledge about a newly discovered human picornavirus. Rev Med Virol. 27.

Van Goethem, M. W., et al., 2018. A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. Microbiome. 6**,** 40.

Versporten, A., et al., 2018. Antimicrobial consumption and resistance in adult hospital inpatients in 53 countries: results of an internet-based global point prevalence survey. Lancet Glob Health. 6**,** e619-e629.

Zhang, K., Farahbakhsh, K., 2007. Removal of native coliphages and coliform bacteria from municipal wastewater by various wastewater treatment processes: implications to water reuse. Water Res. 41**,** 2816-24.

Metagenomics-based evaluation of groundwater microbial profiles in response to treated wastewater discharge

Noor Zaouri [1], Muhammad Raihan Jumat [2], Tariq Cheema [3] and Pei-Ying Hong [1*]

[1] Water Desalination and Reuse Center (WDRC), Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal-Jeddah 23955-6900, Saudi Arabia

[2] Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857

[3] Hydrology Department, Faculty of Earth Sciences, King Abdulaziz University (KAU), Jeddah, Saudi Arabia

* Corresponding author:

Pei-Ying Hong

Tel: +966-(12)-808-2218;

Email: peiying.hong@kaust.edu.sa

Highlights

- Metagenomics as a tool to assess groundwater quality
- Lower alpha diversity of bacteria and virus due to treated wastewater exposure
- Increase in relative abundance of Planctomycetes and Picornaviridae
- Antibiotic resistance genes may not be good markers to differentiate wastewater exposure

Abstract

This study aims to demonstrate the use of metagenomics to assess groundwater quality. Metagenomics revealed a lower alpha diversity for both bacteria and virus in wastewater-exposed groundwater compared to the upstream controls. However, an increase in the relative abundance of Planctomycetes and Picornaviridae, as well as increasing counts of unique antibiotic resistance genes, was observed in wastewater-exposed groundwater. Findings suggest that metagenomics can detect microbial signatures indicative of treated wastewater discharge.

Keywords: Shotgun metagenomics; aquifer recharge; groundwater quality; water reuse

1. Introduction

Water scarcity is projected to exacerbate by 2040, with 33 countries including Saudi Arabia predicted to face extreme water scarcity by that time (Luo et al., 2015). In recent years, recharging aquifer with treated wastewater has been proposed as a mitigation strategy to replenish the groundwater supplies. In Saudi Arabia, most wastewater treatment plants (WWTPs) are designed to achieve secondary or tertiary treatment with a final chlorination step prior to discharge. Such WWTPs typically achieve a cumulative 6-log reduction of microorganisms (Al-Jassim et al., 2015). Considering that untreated wastewater generally has a cell density of 10^{11} to 10^{12} cells/L (Manti et al., 2008), the final treated wastewater still contains a significant microbial load that may potentially perturb the indigenous groundwater microbial community. Aquifers are sensitive ecosystems, and therefore recharging aquifer with treated wastewater can only be adopted with an improved understanding on how these treated wastewater would affect the existing groundwater.

The current approaches to detecting changes in microbial community include qualitative approaches like terminal restriction fragment length polymorphism (T-RFLP) or quantitative approaches like real time PCR (qPCR). T-RFLP generates fingerprinting profiles for comparative analysis but does not provide information on the taxonomic identities of microbial contaminants. qPCR is a targeted approach that does not provide a comprehensive outlook on the wide variety of bacteria, virus and genes contaminants present in treated wastewater. The advent of high throughput untargeted sequencing approaches (e.g. metagenomics) have allowed a simultaneous evaluation on the bacteria, virus and genes present in an environmental sample (Quince et al., 2017). By comparing the diversity and relative abundance of these biological contaminants, one can denote if the groundwater has been perturbed by the treated wastewater.

This study aims to demonstrate the use of metagenomics as a tool to assess microbial water quality. Changes in the diversity and relative abundance of bacteria, virus and antibiotic resistance genes were monitored to denote the extent of perturbation imposed by treated wastewater.

2. Material and method

2.1. Description of sampling sites and sampling procedure

A subsurface reservoir, otherwise also known as Wadi Fatimah in Arabic, situated along the Yamaniyah-Az Zibarah-Fatimah drainage system in Saudi Arabia (Grainger, 1992) was identified for groundwater monitoring. Wadi Fatimah has been receiving treated wastewaters as part of a study commissioned in 2010-2016 by Makkah water utilities to recharge local aquifers. Groundwater at this site was collected over a period of 9 months, and include the following: wells 1, 15, 14 and 17 in March 2016; wells 1, 14 and 17 in August 2016; wells 12, 15 and 17 in December 2016 (Figure S1). Treated wastewaters injected to Wadi Fatimah were from a local WWTP upstream of well 1, and utilized schematics including clarifiers, activated sludge tanks, oxidation ponds and chlorine disinfection. Sampling wells varied throughout the sampling period because of various reasons including lack of sufficient water to facilitate representative sampling, groundwater well was closed for access or no permission was granted from the well owners on the day of sampling. Two wells situated approximately 60 km upstream of Wadi Fatimah were also identified, and groundwater was collected from these wells during the same time frame as mentioned above. Samples from these two wells serve as upstream controls since both the wells are not known to actively receive any treated wastewater discharge (Figure S1). The collected volume of all groundwater samples ranged from 20-40 L per well, and were

comprised of a single grab sample collected per well per sampling event. Before collection, the water was flushed for > 5 min to remove stagnant pools of water retained within the sampling pipes. Water was also flushed until no apparent turbidity was observed prior to collection in sterile polyethylene bottles. All samples were placed into a cooler and kept cool during transport. Upon arrival in laboratory, samples were stored in 4 °C prior to analyses.

2.2. Groundwater concentration and DNA extraction

Each 20-40 L of groundwater sample was passed through a filtration system fitted with a T-series cassette with a 100-kDa cut-off (approximately 0.005 µm) (Pall Corporation, Port Washington, NY, USA). This cassette has been shown previously to retain viruses at an efficiency of up to 87% and was particularly suited for large volume concentration (Jumat et al., 2017). Hence, this procedure was chosen to concentrate any microbial matter present in the low turbidity groundwater samples such as that collected in this study. Each groundwater sample was concentrated to 5 mL of retentate. To collect any microbial matter adhering to the 100 kDa cassette, the system was washed with 50 mL of wash solution (1 x Phosphate Buffered Saline (PBS) and 0.01% Tween 60). After circulating the wash solution through the filtration system for several minutes, the wash solution was itself concentrated to 5 mL. Both retentate and wash solutions derived from the same sample were extracted for DNA was by using the DNeasy Blood and Tissue Kit (Qiagen, Hilgen, Germany). MilliQ water was collected and subjected to the same filtration procedure at the same time when all groundwater samples were filtered. This acts as a negative control, and provide comparative analysis that would allow determination of microbial populations unique to the groundwater.

2.3. Metagenomic sequencing and analysis

Genomic DNA was sent to CosmosID (Rockville, MD, USA) for sequencing on HiSeq, as described (Hasan et al., 2014). Shotgun metagenomic datasets were directly analyzed by CosmosID metagenomic software (CosmosID Inc., Rockville, MD) as described elsewhere (Connelly et al., 2018; Hasan et al., 2014; Kaleko et al., 2016; Ponnusamy et al., 2016) to reveal associated microbial community composition. Briefly, the system utilizes a high performance data-mining k-mer algorithm and highly curated dynamic comparator databases (GenBook®) that rapidly disambiguate millions of short reads into the discrete genomes or genes engendering the particular sequences. The GeneBook databases were composed of over 150,000 microbial genomes and gene sequences representing over 1000 bacterial and 5000 viral species, as well as over 5500 antibiotic resistance genes. For each sample, the reads from a bacteria or viral species are assigned to the strain with the highest aggregation statistics, and subsequently determined for the corresponding alpha diversity. Community resistome was also identified using the CosmosID Inc. bioinformatics software package by querying the unassembled sequence reads against the CosmosID curated antibiotic resistance gene databases. Paired end Fastq files are deposited onto European Nucleotide Archive under accession number PRJEB32173.

2.4. Determination of organic water quality

Organic constituents of the DOC were fractionated according to molecular weight (MW) with size exclusion chromatography. The fraction distribution of the DOC in the groundwater was analyzed using Liquid Chromatography equipped with a macro-porous silica separation column, an Organic Carbon Detector (OCD), a UV 254 nm Detector (UVD) and an Organic Nitrogen

Detector (OND) (LC-OCD Model 8, DOC LABOR, Germany). ChromCaluni software was used to determine the concentration of each fraction by integrating the area underneath the curve according to manufacture instructions.

3. Results and discussion

3.1. Taxonomic profile of bacterial communities

Phyla Proteobacteria, Actinobacteria and Bacteroidetes constitute the three dominant bacterial groups in the groundwater samples of both Wadi Fatimah and upstream controls. However, Proteobacteria and Actinobacteria were also present in high relative abundance in the negative MilliQ controls (Figure 1a). In contrast, Bacteroidetes was detected in higher relative abundance and in more consistent frequency in only the groundwater samples. Planctomycetes was detected in higher relative abundance among the Wadi Fatimah samples than that in upstream and negative controls (Figure 1a). Planctomycetes has been shown to be closely associated with organic detritus in the marine environment (Halliday et al., 2014). Similarly, a higher relative abundance of Planctomycetes was observed in groundwater samples that had a higher dissolved organic carbon (Table 1). Among these five detected DOC fractions, building block (BB) and humic substances (HS) were significantly higher in groundwater sampled from Wadi Fatimah compared to upstream controls ($p < 0.10$, Table 1). Specifically, HS which is predominated by aromatic moieties was only detected in Wadi Fatimah at concentrations ranging from 0.54 to 0.70 mg/L. Despite the higher DOC concentrations in Wadi Fatimah, the alpha diversity of bacterial community was lower in all of its samples compared to the controls (Figure 1b), suggesting the presence of an unknown factor that bias against certain families (e.g. Bradyrhizobiaceae, Methylobacteriaceae) in Wadi Fatimah. These families were only detected in

the upstream controls during the sampling period. There was however, a positive trend of increasing alpha diversity with time, although the reason behind this trend remains unknown.

3.2. Community resistome

Ninety six unique ARGs were detected in Wadi Fatimah while only 56 unique ARGs were detected in the upstream controls (Figure 1c). The 5 most abundant unique ARGs detected in Wadi Fatimah were related to aminoglycoside, beta-lactam, macrolide and quinolone resistance, while the 5 most abundant ARGs detected in upstream controls also included resistance against beta-lactam. Shared ARGs between Wadi Fatimah and control samples revealed additional genes conferring resistance to tetracycline and trimethoprim (Table 2). A positive detection of ARGs in both Wadi Fatimah and control samples suggests a ubiquitous presence of ARGs in most environments, including pristine environment (Van Goethem et al., 2018). This may be related to long term global consumption of antibiotics for both hospital-associated and community-acquired infections (Versporten et al., 2018). Hence, the choice of ARGs as metagenomics markers to denote anthropogenic contamination events may not be provide definitive results.

3.3. Taxonomic profile of viral communities

Podoviridae and Siphoviridae account as the two most dominant viral families in the groundwater (Figure 2a). It was further observed that upstream controls had a significantly higher relative abundance of Siphoviridae, along with Myoviridae, Inoviridae and Tectiviridae ($p < 0.10$). All four families are mainly comprised of bacteriophages that rely on bacteria and archaea as their natural hosts. Along with one of the negative controls which denoted an exceedingly high relative abundance of Picornaviridae, Picornaviridae was consistently detected in all groundwater samples and especially at a higher relative abundance in Wadi Fatimah

samples (average 45%) than in the upstream controls (average 11%). Picornaviridae comprises of several notable human enteric viruses including Enteroviruses and Salivirus which are proposed as potential human virus indicator (Reuter et al., 2017). The high relative abundance of Picornaviridae may be related to the type of treatment process used by the WWTP situated in Wadi Fatimah. Secondary and tertiary treatment processes typically do not achieve high removal rates of viruses (Zhang and Farahbakhsh, 2007). In contrast, a WWTP retrofitted with membrane bioreactor reported only sporadic occurrences of Enterovirus (Jumat et al., 2017). The alpha diversity of viruses are higher in the controls than in Wadi Fatimah (Figure 2b). Coupled with the higher abundance of families associated with bacteriophages, the higher alpha diversity may be due to a wide variety of bacteriophages present in relatively unperturbed environments compared to those exposed to treated wastewater. Unlike the bacterial alpha diversity, the viruses did not exhibit an increasing trend of alpha diversity with time (Figure 2b).

4. Conclusion

Findings from this study demonstrates the use of metagenomics to provide a simultaneous evaluation of bacteria, virus and antibiotic resistome in groundwater exposed to treated wastewater discharge. It was observed that alpha diversity of both bacteria and viruses were lower in exposed groundwater than the upstream controls. In addition, Planctomycetes and Picornaviridae, both of which were detected by metagenomics simultaneously, may be appropriate markers to indicate increased organic content and presence of human fecal contamination. As more countries explore the use of treated wastewater to recharge diminishing groundwater levels, metagenomics could suffice as an appropriate monitoring tool to ensure that a wide diversity of biological contaminants are assessed and unintentional aberrations to the

groundwater minimized. However, metagenomics alone is unable to provide insights to the potential risks that may arise from recharging aquifers with treated wastewater. This is because metagenomics analysis does not provide absolute quantification of biological contaminants and are derived from DNA template sequencing. Hence it is unable to differentiate the viable bacteria populations against the dead and non-infectious ones. Instead, metagenomics can possibly narrow down the suite of contaminants to a selected few (e.g. Picornaviridae) for more in-depth evaluation by other quantitative methods like quantitative PCR or digital PCR. By coupling metagenomics with other quantitative methods, potential risks arising from recharging aquifer with treated wastewater can be better assessed.

Acknowledgements

The authors would like to thank Dr Moustapha Harb and Dhafer Alsaleh for assistance during the early sampling years. The authors would also like to thank Professor Tom Missimer and Mr. Mohammed AlRayaan for providing background information related to the hydrogeological characteristics of Wadi Fatimah. This study is funded by KAUST Baseline Grant BAS/1/1033-01-01 awarded to P.-Y. Hong.

Table 1. Summary of the organic matter components in Wadi Fatimah and upstream controls. DOC denotes dissolved organic carbon, BP denotes biopolymers, HS denotes humic substances, BB denotes building blocks, LMW denotes low molecular weight. Statistical comparison was made using t-test between Wadi Fatimah samples and upstream controls. ** denotes significant difference at 95% confidence level. * denotes significant difference at 90% confidence level. Value after ± denotes standard deviation.

		Wadi Fatimah			Upstream controls	P-value
		March 2016	August 2016	December 2016		
DOC	mg/L	8.59 ± 3.23	4.70 ± 0.89	6.12 ± 4.43	1.71 ± 2.1	0.08*
Carbon distribution						
BP	mg/L	0.08 ± 0.001	0.04 ± 0.01	0.07 ± 0.01	0.02 ± 0.05	0.17
HS	mg/L	0.70 ± 0.07	0.62 ± 0.02	0.54 ± 0.04	0.04 ± 0.05	0.01**
BB	mg/L	3.20 ± 0.15	2.58 ± 0.34	2.11 ± 0.09	0.68 ± 0.88	0.08*
LMW neutrals	mg/L	3.36 ± 0.33	0.58 ± 0.06	1.28 ± 0.11	0.14 ± 0.06	0.33
LMW acids	mg/L	1.13 ± 0.107	0.64 ± 0.04	1.56 ± 0.23	1.16 ± 1.64	0.95

Table 2. Antibiotic resistance genes detected in groundwater samples. * denotes that the average percentage matched reads were already corrected for the average percentage detected in MilliQ control. N.D. denotes not detected or below detection limits.

Top 5 identified ARGs unique to Wadi Fatimah	**Average % total matches ***	**Average % in MilliQ control**	**Denoted function**
Aminoglycoside *aph3' Ia*	0.03	N.D.	Aminoglycoside phosphotransferase gene, confers resistance by catalyzing the addition of phosphate from ATP to the 3'-hydroxyl group of a 4,6-disubstituted aminoglycoside
Aminoglycoside a*adA13*	0.04	N.D.	Aminoglycoside nucleotidyltransferase gene, confers resistance by adenylating the 2''-hydroxyl group of aminoglycoside
Beta-lactam-resistance bla_{CAU-1}	0.04	N.D.	Metallo-β-lactamase gene, confers resistance by catalyzing the hydrolysis of beta-lactam
Macrolide *ereA*	0.05	0.01	Erythromycin esterase gene, confers resistance through inactivation by hydrolyzing the lactone ring of the antibiotic
Plasmid-mediated-quinolone-resistance *qnrD1*	0.04		Quinolone resistance gene, confers resistance by preventing antibiotic binding
Top 5 identified ARGs unique to upstream controls	**Average % total matches ***	**Average % in MilliQ control**	**Denoted function**
Multidrug efflux pump *mdtE*	0.14	0.15	Part of the tripartite efflux system MdtEF-TolC, which confers resistance to compounds such as erythromycin, doxorubicin, benzalkonium and so on
Repressor for multidrug efflux pump *emrR*	0.12	0.13	Negatively regulates the transcription of the multidrug resistance pump-encoding operon, *emrRAB*, by binding to its regulatory region
Sensor kinase *cpxA*	0.11	0.13	Histidine kinase member of the two-component regulatory system CpxA/CpxR, responds to envelope stress response
Beta-lactam-resistance $bla_{OXA\ 281}$	0.10	N.D.	Oxacillinase (enzyme that confers resistance to oxacillin, a type of beta-lactam), presumably intrinsic to *Acinetobacter johnsonii* SH046
Beta-lactam-resistance $bla_{OXA\ 211}$	0.09	N.D.	Oxacillinase (enzyme that confers resistance to oxacillin, a type of beta-lactam), presumably intrinsic to *Acinetobacter johnsonii* SH046

ARGs identified in Wadi Fatimah and upstream controls	Average % total matches in Wadi Fatimah*	Average % total matches in upstream control*	Average % in MilliQ control	Denoted function
Beta-lactam-resistance $bla_{OXA-212}$	0.04	0.10	N.D.	Carbapenem-hydrolyzing class D beta-lactamase, confers resistance by catalyzing the hydrolysis of beta-lactam
Beta-lactam-resistance $bla_{OXA-373}$	0.03	0.14	N.D.	Functions as a transpeptidase, beta-lactam became suicide substrates of transpeptidases
Tetracycline *tet39*	0.04	0.16	0.05	Tetracycline resistance gene, confers resistance by active efflux mechanisms
Tetracycline *tetRG*	0.05	0.09	0.05	TetR is the repressor of the tetracycline resistance element, TetG is tetracycline efflux protein
Trimethoprim *dfr32*	0.05	0.03	0.12	Dihydrofolate reductase and trimethoprim resistance gene, confers resistance by replacing or substituting antibiotic action target

201

Figure legends

Figure 1. Bacterial communities and bacteria-associated antibiotic resistome in groundwater. **(a)** Heat map illustrates relative abundance of individual bacteria phyla in the negative controls (MilliQ water), upstream controls and wells in Wadi Fatimah across different sampling months. **(b)** Bacteria alpha diversity in groundwater sampled from Wadi Fatimah across different sampling months, upstream controls and negative controls. Linear line depicts an increasing trend in alpha diversity with time. **(c)** Venn diagram illustrating the number of unique antibiotic resistance genes detected in Wadi Fatimah (depicted as green) and upstream control samples (depicted as red). Overlapping region shows the number of shared antibiotic resistance genes detected in both Wadi Fatimah and upstream controls.

Figure 2. Viral communities in groundwater. **(a)** Heat map illustrates the relative abundance of individual viral families in the control (MilliQ water), upstream controls and wells in Wadi Fatimah across the different sampling times. **(b)** Virus alpha diversity in groundwater sampled from Wadi Fatimah across different sampling months, upstream controls and negative controls. Linear line depicts no apparent trend of increment with time.

References

Al-Jassim, N., et al., 2015. Removal of bacterial contaminants and antibiotic resistance genes by conventional wastewater treatment processes in Saudi Arabia: Is the treated wastewater safe to reuse for agricultural irrigation? Water Res. 73**,** 277-90.

Connelly, S., et al., 2018. Distinct consequences of amoxicillin and ertapenem exposure in the porcine gut microbiome. Anaerobe. 53**,** 82-93.

Grainger, D. J., 1992. A geological excursion in the Wadi Fatima area, near Jiddah, Saudi Arabia. Geology Today. 8**,** 215-219.

Halliday, E., et al., 2014. Comparison of bacterial communities in sands and water at beaches with bacterial water quality violations. PLoS One. 9**,** e90815.

Hasan, N. A., et al., 2014. Microbial Community Profiling of Human Saliva Using Shotgun Metagenomic Sequencing. Plos One. 9.

Jumat, M. R., et al., 2017. Membrane Bioreactor-Based Wastewater Treatment Plant in Saudi Arabia: Reduction of Viral Diversity, Load, and Infectious Capacity. Water. 9.

Kaleko, M., et al., 2016. Development of SYN-004, an oral beta-lactamase treatment to protect the gut microbiome from antibiotic-mediated damage and prevent Clostridium difficile infection. Anaerobe. 41**,** 58-67.

Luo, T., et al., Aqueduct Projected Water Stress Country Rankings. 2015. http://reliefweb.int/sites/reliefweb.int/files/resources/AQUEDUCT%20PROJECTED%20WATER%20STRESS%20Country%20Rankings.pdf. 31 August 2019

Manti, A., et al., 2008. Bacterial cell monitoring in wastewater treatment plants by flow cytometry. Water Environ Res. 80**,** 346-54.

Ponnusamy, D., et al., 2016. Cross-talk among flesh-eating *Aeromonas hydrophila* strains in mixed infection leading to necrotizing fasciitis. Proceedings of the National Academy of Sciences. 113**,** 722-727.

Quince, C., et al., 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology. 35**,** 833.

Reuter, G., et al., 2017. Saliviruses-the first knowledge about a newly discovered human picornavirus. Rev Med Virol. 27.

Van Goethem, M. W., et al., 2018. A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. Microbiome. 6**,** 40.

Versporten, A., et al., 2018. Antimicrobial consumption and resistance in adult hospital inpatients in 53 countries: results of an internet-based global point prevalence survey. Lancet Glob Health. 6**,** e619-e629.

Zhang, K., Farahbakhsh, K., 2007. Removal of native coliphages and coliform bacteria from municipal wastewater by various wastewater treatment processes: implications to water reuse. Water Res. 41**,** 2816-24.

Figure 1

Click here to download high resolution image

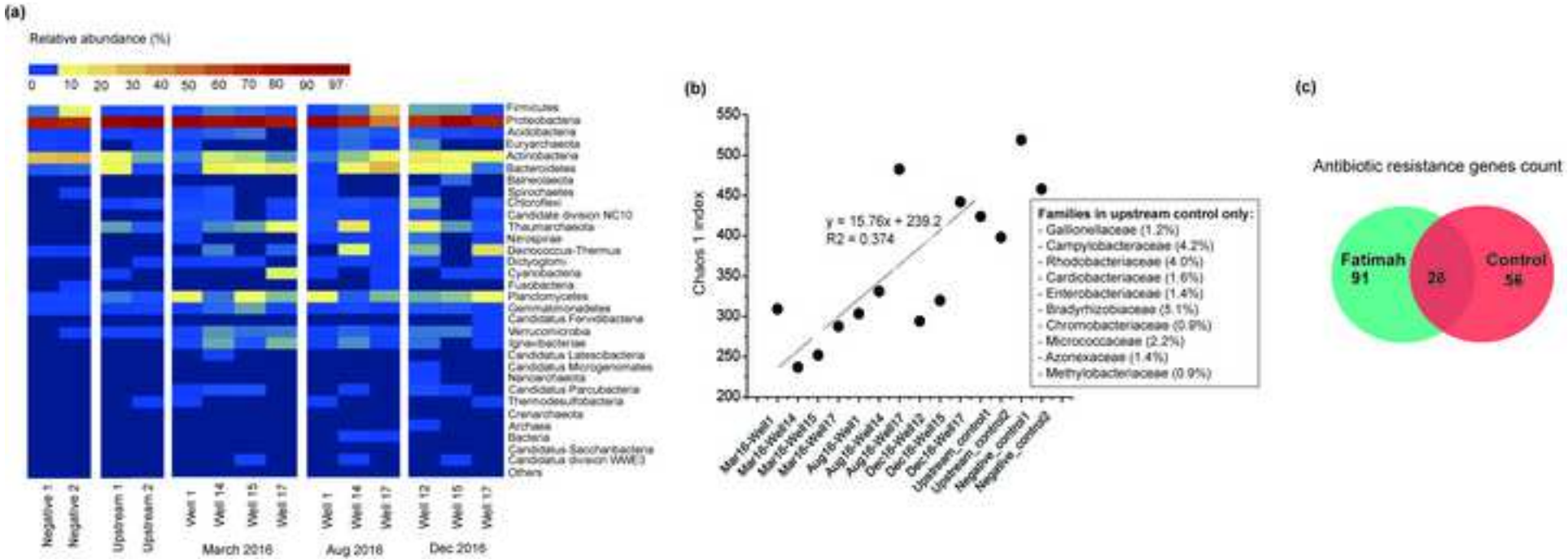
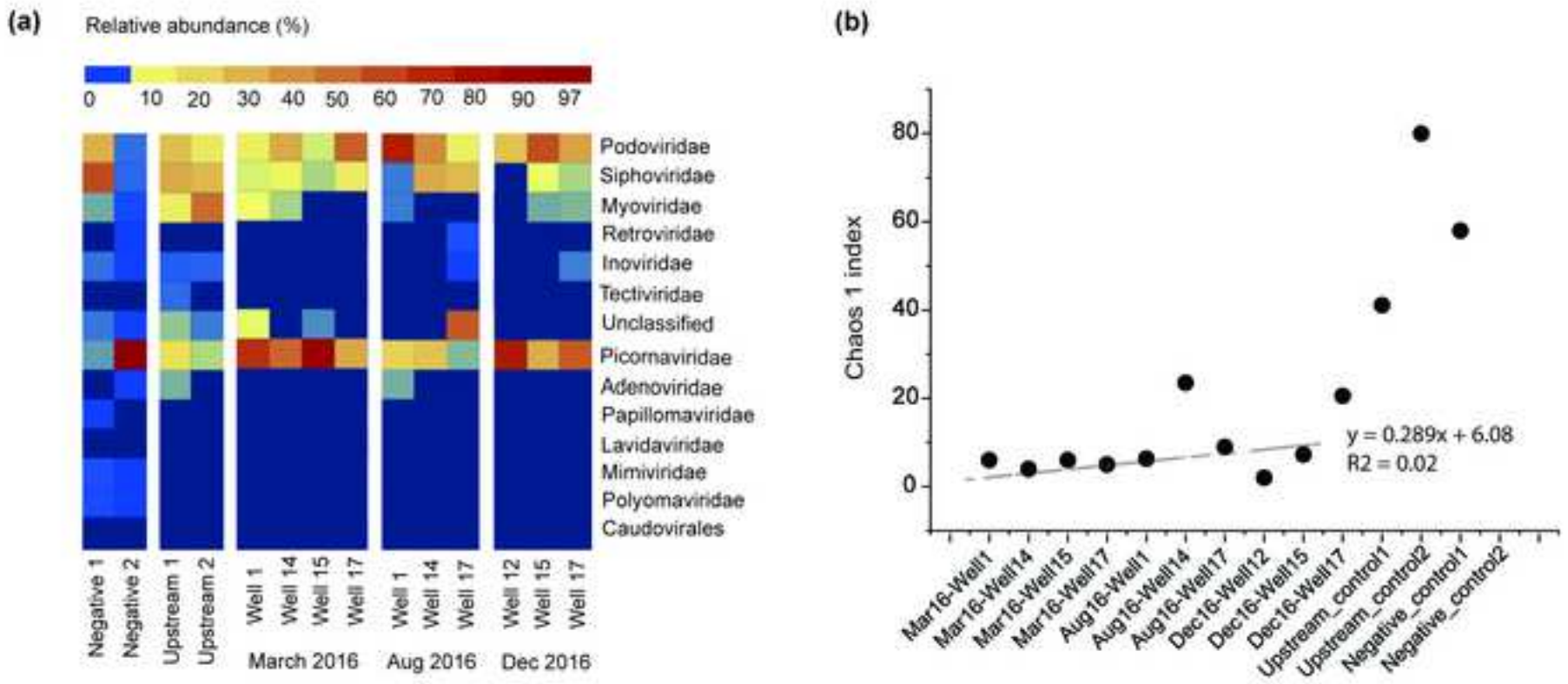


Figure 2

Click here to download high resolution image



Supplementary Material

Click here to download Supplementary Material: For review - Supplemental Information.docx

***conflict of Interest Statement**

Conflict of interest

The authors declare no conflict of interest.