Electronic Supplementary Material (ESI) for Environmental Science: Water Research & Technology.

Removal of bacterial cells, antibiotic resistance genes and integrase genes by on-site hospital wastewater treatment plants: Surveillance of treated hospital effluent quality

Kenda Timraz [1], Yanghui Xiong [1], Hamed Al Qarni [2], Pei-Ying Hong * [1]

1. King Abdullah University of Science and Technology (KAUST), Water Desalination and Reuse Center (WDRC), Biological and Environmental Sciences & Engineering Division (BESE), Thuwal, 23955-6900, Saudi Arabia
2. Urban Water Technology Centre, School of Science, Engineering and Technology, University of Abertay Dundee, Bell Street, Dundee DD1 1HG, Scotland, UK

Supplemental information

1. **Preparation procedures for amplicon-based sequencing** Barcoded PCR for 16S rRNA gene-based Ion PGM sequencing was carried out with primer pairs 515F: (5′-Barcode-GTGYCAGCMGCCGCGGTA-3′) and 909R: 5′-CCCCGYCAATTCMTTTRAGT-3′). Each PCR reaction contained 0.3 μL of 0.025 U/μL Ex Taq polymerase (Takara Bio, Dalian, China), 25 μL of 2X Epicentre Biotechnologies FailSafe™ Premix F (Illumina, Madison, WI, USA), 1 μL each of 10 μM each of forward and reverse primer, 22 μL H_2O and 1 μL of DNA template. The thermal cycling program included an initial denaturation stage at 95 °C for 3 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 45 s and extension at 72 °C for 60 s, and then a final extension stage at 72 °C for 10 min. The presence of PCR product was verified using Invitrogen SYBR® green nucleic acid gel stain (Thermo Fisher Scientific, Carlsbad, CA, USA) and gel electrophoresis. Controls for PCR reactions were negative for amplification. Amplicons were of the correct anticipated size of ~450 bp, and were gel-purified with the Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA). The concentrations were then measured by Invitrogen Qubit® 2.0 fluorometer (Thermo Fisher Scientific, Carlsbad, CA, USA). Purified amplicons were submitted to KAUST Genomics Core lab for sequencing on an Ion Torrent™ 314 chip.
2. **Amplicon-based high-throughput sequencing data analysis** Sequences were initially sorted by the Bioinformatics Team at KAUST based on a Phred score of > 20. The sorted sequences were then trimmed off for the primers, barcodes and adaptor sequences, and removed of any sequences > 250 nt in length. Chimeras were identified and removed on UCHIME [1] by referencing against a core reference fasta file that was

downloaded from Greengenes (i.e., gold strains gg16 – aligned.fasta, last updated on 19 March 2011). The first approach is to assign the chimera-free sequences to bacterial/archaeal taxonomic hierarchy at a 95% confidence level using the Ribosomal Database Project (RDP) Classifier [2]. Chimera-free sequences were also aligned using the RDP Infernal Aligner, and aligned files were submitted to their individual cluster [3]. The cluster files were used in rarefaction analysis. Microbial richness for each sample was obtained from the rarefaction curves based on a defined sequencing depth of 3250 sequences. The relative abundances of the bacterial and archaeal genera were calculated, collated and then square-root transformed. Square-root transformation is performed so as to down-weight the dominant taxa, and to achieve a better balance of the abundant and rarer species that are present in the samples. This allows the rarer species, which are common in the data generated from high-throughput sequencing to exert some influence on the calculation of similarity, and is done based on recommendations suggested by statistician developers of Primer-E. The transformed data set were then computed for their Bray–Curtis similarities and represented graphically for spatial distribution in a metric multidimensional scaling (mMDS) plot using Primer-E version 7 [4]. Bio-Env (BEST) correlates the microbial data with the measured chemical data, and was conducted using Primer-E version 7 [5].

3. **Phylogenetic identification of bacterial isolates** The 16S rRNA gene sequences were then PCR-amplified using forward primer 11F (5'- GTTYGATYCTGGCTCAG -3') and reverse primer 1492R (5'-GGYTACCTTGTTACGACTT-3'). Each PCR reaction contained 0.3 μL of 0.025 U/μL Ex Taq polymerase (Takara Bio, Dalian, China), 25 μL of 2X Epicentre Biotechnologies FailSafe™ Premix F (Illumina, Madison, WI, USA), 1

μL each of 10 μM each of forward and reverse primer, 22 μL H_2O and 1 μL of DNA template. Thermal cycling was conducted on C1000 Touch Thermocycler (BioRad Laboratories Inc., Hercules, CA, USA) with the following conditions: an initial denaturation stage at 95 °C for 3 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 50 °C for 45 s and extension at 72 °C for 60 s, and a final extension stage at 72 °C for 10 min. The presence of PCR product was verified using Invitrogen SYBR® green nucleic acid gel stain (Thermo Fisher Scientific, Carlsbad, CA, USA) and gel electrophoresis. Controls for PCR reactions were negative for amplification. 16S rRNA gene amplicons obtained from the bacterial isolates were of the correct size of ~1500 bp, and were purified using the Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA). The purified products were then sent in to the KAUST Genomics core lab for Sanger sequencing on ABI 3730 XL using the primer 338F (5'-ACTCCTACGGGAGGCAGC-3'). Gene sequences were matched against the National Center for Biotechnology Information (NCBI) 16S rRNA sequences database using Blastn to determine the gene identities of isolates.

4. **Solid phase extraction and LC-MS/MS detection** All effluent samples were first filtered with 0.45 μm nylon membrane. The filtrate was adjusted to pH 3-4 with 40% sulfuric acid and 1N NaOH, and after that 4.726 g/L citric acid, 4.37 g/L Na_2HPO_4 and 14.892 g/L $Na_2EDTA{\cdot}2H_2O$ were added along with 20 μg/L each of tetracycline-d6 isotope and simeton (internal surrogate controls). The sample was thoroughly shaken and subject to solid phase extraction (SPE) (Dionex, AutoTrace 280, Thermo Scientific, USA). The SPE cartridge (Oasis HLB, 500 mg) was conditioned with methanol, 10 mM trifluoroacetic acid and deionized water, consecutively, before the sample was loaded.

The flowrate for the conditioning and loading was 15 mL/min and 3 mL/min, respectively. Three consecutive rinsing steps with 10 mL methanol/deionized water (5:95 v/v), 10 mL deionized water and 10 mL deionized water were performed after loading was completed. The cartridge was dried under gas stream for 60 min. The dried cartridge was then eluted twice with 5 mL of methanol with 0.1% formic acid at 1 mL/min. The collected eluate was evaporated under a nitrogen gas stream till the volume is 0.5 mL. The concentrated sample was stored in a 1.5 mL amber glass vial and then determined for tetracycline (TC) on LC-MS/MS (Transcend, Thermo Scientific, Massachusetts, US). All samples were separated using a Luna®-C18(2) in combination with a guard column (Phenomenex, California, USA). 0.1% formic acid in water (Phase A) and 0.1% formic acid in methanol (Phase B) was combined to form a gradient mobile phase to separate the TC as well as the two internal surrogates. The isolated sample by LC was detected by Tandem Mass Spectrometry (TSQ Vantage, Thermo Scientific, US). A series of tetracycline (TC) and isotope-TC standards ranging from 2 to 50 μg/L were conducted with the SPE before LC-MS/MS analysis. All extracted samples were detected in duplicate, and the SPE recovery for tetracycline is 52.7 ± 18.9 %.

Table S1. Water quality at the different stages of SH and IH WWTP.

Parameters	**SH WWTP**		**IH WWTP**	
	Influent	**Effluent**	**Influent**	**Effluent**
Chemical oxygen demand, COD (mg/L)	376	64	336	27
Total nitrogen, TN (mg/L)	23.8	1.6	19.7	2.7
Ammonium, NH_4^+ (mg/L)	22.3	0.3	19.0	Not detected
Nitrite, NO_2^- (mg/L)	0.3	0.6	0.3	0.3
Nitrate, NO_3^- (mg/L)	0.2	0.7	0.4	2.4

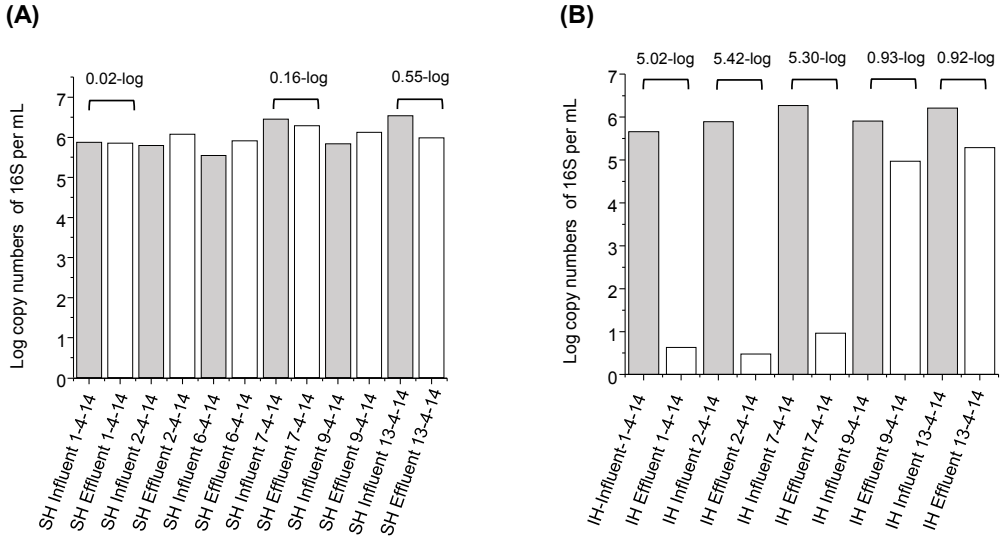

Figure S1. Total 16S rRNA gene copies (log) per mL, quantified by qPCR in influent and effluent of SH WWTP **(A)** and IH WWTP **(B).** The corresponding log removal values are denoted. Numbers accompanying each sample name in the x-axis correspond to the date at which the sample was collected.

References

1. R. C. Edgar, B. J. Haas, J. C. Clemente, C. Quince and R. Knight, *Bioinformatics (Oxford, England)*, 2011, **27**, 2194-2200.
2. J. R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity and J. M. Tiedje, *Nucleic acids research*, 2009, **37**, D141-145.
3. Q. Wang, G. M. Garrity, J. M. Tiedje and J. R. Cole, *Applied and environmental microbiology*, 2007, **73**, 5261-5267.
4. K. Clarke and R. Gorley, *PRIMER-E, Plymouth, UK*, 2015, **296**.
5. K. R. Clarke, *Australian journal of ecology*, 1993, **18**, 117-143.