

RESEARCH ARTICLE

Bacterial Pathogens and Community Composition in Advanced Sewage Treatment Systems Revealed by Metagenomics Analysis Based on High-Throughput Sequencing

Xin Lu¹, Xu-Xiang Zhang^{1*}, Zhu Wang¹, Kailong Huang¹, Yuan Wang², Weigang Liang³, Yunfei Tan³, Bo Liu^{1,4*}, Junying Tang¹

1 State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing, China, **2** Environmental Science Research Institute of Jiangsu, Nanjing, China, **3** Zhengzhou Sewage Purification Company, Zhengzhou, China, **4** Research Institute of Nanjing University at Lianyungang, Lianyungang, China

* zhangxx@nju.edu.cn (XXZ); yongboliu@163.com (BL)



OPEN ACCESS

Citation: Lu X, Zhang X-X, Wang Z, Huang K, Wang Y, Liang W, et al. (2015) Bacterial Pathogens and Community Composition in Advanced Sewage Treatment Systems Revealed by Metagenomics Analysis Based on High-Throughput Sequencing. PLoS ONE 10(5): e0125549. doi:10.1371/journal.pone.0125549

Academic Editor: Horacio Bach, University of British Columbia, CANADA

Received: September 17, 2014

Accepted: March 25, 2015

Published: May 4, 2015

Copyright: © 2015 Lu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The authors would like to confirm that all data underlying the findings described in this manuscript are fully available without restriction. The pyrosequencing data have been deposited into the NCBI Short Reads Archive Database (URL: <http://www.ncbi.nlm.nih.gov/sra>; accession numbers: SRR1103268- SRR1103272 and SRR1060401, see [S2 Table](#) in the Supporting Information). The Illumina shotgun sequencing data have been deposited into MG-RAST server (URL: <http://metagenomics.nmpdr.org>; accession numbers:

Abstract

This study used 454 pyrosequencing, Illumina high-throughput sequencing and metagenomic analysis to investigate bacterial pathogens and their potential virulence in a sewage treatment plant (STP) applying both conventional and advanced treatment processes. Pyrosequencing and Illumina sequencing consistently demonstrated that *Arcobacter* genus occupied over 43.42% of total abundance of potential pathogens in the STP. At species level, potential pathogens *Arcobacter butzleri*, *Aeromonas hydrophila* and *Klebsiella pneumonia* dominated in raw sewage, which was also confirmed by quantitative real time PCR. Illumina sequencing also revealed prevalence of various types of pathogenicity islands and virulence proteins in the STP. Most of the potential pathogens and virulence factors were eliminated in the STP, and the removal efficiency mainly depended on oxidation ditch. Compared with sand filtration, magnetic resin seemed to have higher removals in most of the potential pathogens and virulence factors. However, presence of the residual *A. butzleri* in the final effluent still deserves more concerns. The findings indicate that sewage acts as an important source of environmental pathogens, but STPs can effectively control their spread in the environment. Joint use of the high-throughput sequencing technologies is considered a reliable method for deep and comprehensive overview of environmental bacterial virulence.

Introduction

Sewage is considered as a habitat for various human bacterial pathogens, and discharge of sewage into environmental water may threaten public health [1–3]. Bacterial pathogenesis often depends on specific virulence factors (VFs) [3], which enable microorganisms to proliferate and cause tissue damage or systemic inflammation in a host niche [4]. In China, sewage

4545873.3-4545879.3, 4545897.3-4545900.3, 4545924.3-4545927.3, 4546809.3-4546811.3, 4545974.3, 4545779.3, 4547054.3, 4547057.3, 4547080.3 and 4547081.3, see [S2 Table](#) in the Supporting Information). The DNA marker sequences of the four pathogens have been deposited in GenBank (URL: <http://www.ncbi.nlm.nih.gov/genbank/>; accession numbers: KP284853-KP284856, see [S4 Table](#) in the Supporting Information).

Funding: This study was financially supported by National Science and Technology Major Project of China (No. 2012ZX07204-001, PI: BL; No. 2012ZX07101-002, PI: XXZ; URL: <http://www.nmp.gov.cn/>) and Environmental Protection Research Foundation of Jiangsu Province (China) (No. 2012044 and 2012045; PI: XXZ; URL: <http://www.jshb.gov.cn/jshbw/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Two co-authors (Weigang Liang and Yunfei Tan) of this manuscript are working for the Production and Technology Department of a commercial company Zhengzhou Sewage Purification Company. The authors would like to state that the two authors have no competing or financial interests to this manuscript. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

treatment plants (STPs) normally consist of primary sedimentation and secondary biological treatment [5, 6]. Because of the specific effects of light, temperature and oxygen availability on shaping bacterial community structure, the primary sedimentation and conventional secondary biological treatment could remove about 30% and 80% pathogens from influent, respectively [7, 8]. Despite of high removal of pathogens in STPs [9, 10], the residual pathogens in effluent still deserve public concerns. It has been indicated that STPs are important sources for environmental pathogens [11], such as *Legionella pneumophila* breaking out from Norway STPs [12]. In order to reduce the potential health risk induced by pathogens, disinfection methods (e.g. chlorination) are usually used in sewage treatment [13], although the disinfection techniques are known capable of inducing adverse effects, such as disinfection by-products generation [14] and antibiotic resistance enhancement [15]. However, few studies have been conducted to investigate the fates of bacterial pathogens in advanced sewage treatment systems.

Human pathogens can survive and grow in STPs [2], but they are minor components of bacterial community in sewage and have to be enriched for detection [3]. Traditionally, total coliforms and fecal coliforms are used to indicate the contamination of water or wastewater by pathogens [16]. However, most of the environmental bacteria are unculturable [17], and the bacteria culture may cause bias in detection of most pathogens. Recently, molecular technologies, such as polymerase chain reaction (PCR) [18], quantitative real time PCR (q-PCR) [19, 20] and microarray [20], have been widely used to detect pathogens in sewage and natural waters. Nevertheless, these technologies can only specifically detect certain pathogens and cannot provide a comprehensive insight of potential pathogens in the environment. High-throughput sequencing, a promising technology, can provide enough sequencing depth to cover the complex microbial community of the environment [21]. Recent studies have investigated the diversity of pathogens in activated sludge of STPs by using 454 pyrosequencing of 16S rRNA gene amplicons [22], or Illumina high-throughput sequencing of environmental genomes [23, 24].

This study aimed to investigate the fates of pathogens and VFs in both conventional and advanced treatment processes of a full-scale STP by joint use of 454 pyrosequencing, Illumina high-throughput sequencing and q-PCR. This study provided a comprehensive insight into bacterial pathogens and community composition in the STPs and demonstrated the applicability of the high-throughput sequencing technologies in environmental pathogens detection.

Materials and Methods

Sampling and DNA extraction

Wastewater and sludge samples were collected from Wulongkou STP (Zhengzhou City, Henan Province, China; 34.78 deg N, 113.61 deg E) with daily treatment of 100,000 m³ of sewage. [S1 Table](#) shows detailed information about the operational processes and wastewater quality. We would like to state that the plant have approved this study which did not involve endangered or protected species. In the STP, sewage underwent a series of conventional treatment processes including grating, grit filtration and activated sludge treatment and secondary settling. The effluent from secondary settling tank was separately subject to two advanced treatment processes of coagulation/precipitation and resin filtration ([S1 Fig](#)). Samples were simultaneously collected from six locations along wastewater flow, including sewage influent (SI), primary effluent (PE), activated sludge (AS), secondary effluent (SE), final sand filter effluent (FFE) and final resin effluent (FRE) at three time points of November 2012, February 2013 and May 2013 ([S1 Fig](#), [S2 Table](#)).

Activated sludge samples (about 50 ml each) were centrifuged at 4,000 rpm for 10 min to collect approximately 200 mg of the pellet for DNA extraction. The influent and effluent

samples were filtered through 0.45- μ m micropore membrane with diameter of 50 mm (Xinya Co., China) to collect cell pellets for DNA extraction. For each sample, the extraction was conducted in duplicate using the FastDNA Soil Kit (MP Biomedicals, CA, USA). The DNA concentration and purity were determined by microspectrophotometry (NanoDropND-2000, NanoDrop Technologies, Wilmington, DE). The volume of the raw and treated sewage filtered through membrane was recorded to calculate the amount of DNA in given volume of sewage water.

454 Pyrosequencing and Illumina high-throughput sequencing

To avoid temporal variation, equal mass of DNA extracted from the three samples collected from each location at different time points were mixed for pyrosequencing on Roche 454 FLX Titanium platform (Roche, USA) and high-throughput sequencing on Illumina HiSeq 2000 (Illumina, USA). The extracted DNA was sent to Beijing Genome Institute (BGI, China) for Illumina shotgun sequencing. The sequencing strategy was “Index 101 PE” (Paired End sequencing, 101-bp reads and 8-bp index sequence), which generated nearly equal amount of clean reads for each sample. The raw reads containing three or more “N” or contaminated by adapter (> 15 bp overlap) were removed, and the filtered clean reads were used for metagenomic analyses.

The bacterial 16S rRNA gene was amplified with a set of primers targeting the hypervariable V3-V4 regions (about 460 bp). The primers used were V3F (5'- ACTCCTACGGGAGGCAG CAG-3') and V4R (5'- TACNVGGGTATCTAATCC-3'). PCRs were conducted in 50 μ L reaction volume, containing 1 \times Pfx Amplification Buffer (Invitrogen, USA), 0.4 mM dNTP, 2 mM MgSO₄, 0.4 μ M each fusion primer, 1 μ L of template DNA and 2 U of Platinum Pfx DNA Polymerase (Invitrogen, USA). PCR protocol included initial denaturation at 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, annealing at 62°C for 30 s and extension at 70°C for 45 s, with a final elongation step at 70°C for 7 min. A 10-nucleotide “barcode” was added into forward primer to separate the corresponding reads from the data pool generated in a single pyrosequencing run. To minimize the impact of potential early round PCR errors, PCR amplicon libraries were prepared by combining three independent PCR products for each sample. The PCR products were purified by MiniBest DNA Fragment Purification Kit Ver.4.0 (TaKaRa, Japan). Purified PCR products were mixed and sent to Majorbio Co. Ltd. (Shanghai, China) for pyrosequencing. Average length of the pyrosequencing reads generated was about 494 bp, and the detailed sequencing information is shown in [S3 Table](#). The pyrosequencing and Illumina sequencing data have been deposited into the NCBI Short-reads Archive Database and MG-RAST server, respectively, under the accession numbers listed in [S2 Table](#).

Bioinformatics analysis on Illumina high-throughput sequencing

The bioinformatics strategy of the Illumina data included four steps: (i) quality filtering and normalization, (ii) alignment, (iii) annotation, and (iv) statistics analysis. For quality control, the sequences were denoised by FASTX toolkit tools in GALAXY using parameters (quality cut-off value 30; percent of bases cut-off value 75) [25]. In order to compare the six samples at the same sequencing depth, the number of the filtered reads from each sample was normalized to 9,000,000 by a self-written Python script.

To identify the 16S rRNA gene, sequences of each dataset were aligned to the Silva database [26] using the local BLASTN tool. The output files were then imported into the MEGAN tool [27] to conduct the taxonomic assignment with the lowest common ancestor (LCA) algorithm using default parameters (min score 50, top percent 10, win score 0.0, and min complexity 0.3) [28]. The synonyms file (silva2ncbi.map) downloaded from MEGAN website (<http://ab.inf>

unituebingen.de/software/megan/) was used to map Silva accession numbers to NCBI taxa. Each of the dataset was imported into the MetaPhlAn online platform in GALAXY (<http://huttenhower.sph.harvard.edu/galaxy>) to profile the microbial community using the default parameters [29]. The output files were calculated and collated into a table file to generate a heatmap by R.

To establish a local database of VFs, the protein sequences of pathogenicity islands and virulence proteins with specific pathogenic function were downloaded from MvirDB [1], a microbial database of protein toxins, VFs and antibiotic resistance genes. The Illumina datasets were aligned against the purified MvirDB by off-line BLASTX, and a sequencing read was annotated as one VF (pathogenicity islands or virulence proteins) if its best BLASTX hit had a similarity above 90% over an alignment of more than 50 bp [15]. To compile, confirm and validate the collection of data, each of the MvirDB protein sequences were manually aligned against the NCBI Protein Database (<http://blast.ncbi.nlm.nih.gov/>) to check whether one protein is specific for the pathogenic hosts, and the ones not directly involving pathogenesis were deleted from the database.

Bioinformatics analysis on 454 pyrosequencing

Pyrosequencing reads of each data set were denoised using the Mothur platform [30]. Sequences were trimmed using the “trim.seqs” command [31], (1) to trim off the adapters, barcodes and primers, (2) to remove the low quality reads and the reads containing ambiguous “N”, and (3) to remove the reads shorter than 300 bp. Subsequently, pyrosequencing errors were removed using the “pre.cluster” command [21], and PCR chimeras were removed using Chimera slayer [21]. For fair comparison, sequence number was normalized to 6,200 for each sample using a self-written Python script.

All the filtered sequences were compared with the Silva database [26] by the local BLASTN tool, and the output sequences were then assigned to NCBI taxonomies with MEGAN [22] to profile the bacterial communities. Sequences identified as archaea and eukaryon were filtered out. To further identify pathogens at species level, the retained sequences were aligned to a human pathogenic bacteria 16S rRNA gene database [23] downloaded from a homepage of the University of Hong Kong (<http://web.hku.hk/~zhangt/ZhangT.htm>), and pathogenic bacteria were determined if hits similarity was over 97% [32].

Quantitative real-time PCR

The genes encoding beta-glucuronidase (*uidA*) of *Escherichia coli* [19], extracellular lipase (*lip*) of *Aeromonas hydrophila* [33], gyrase A subunit (*gyrA*) outside the quinolone resistance determining region of *Arcobacter butzleri* [34] and phosphate-regulated porin (*phoE*) of *Klebsiella pneumoniae* [35] were used as specific indicators to quantify the four potential pathogens in the sewage treatment plant, respectively. Specificity of the primer sets for the pathogens was confirmed by online BLAST search (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Abundance of the four genes in the samples was determined by q-PCR on Corbett Real-Time PCR Machine with Rotor-Gene 6000 Series Software 1.7 (QIAGEN, the Netherlands). q-PCR of each gene was conducted in triplicate with the primer sets listed in S4 Table. The recombinant plasmids obtained by molecular cloning of the target genes were used to generate standard curves. In detail, PCRs of target genes were conducted in a 30 μ L volume containing 1 \times PCR buffer, 100 μ M dNTP, 2 pmol of each primer, 150 ng of DNA template and 0.8 U of ^{EX}Taq polymerase (TaKaRa, Japan). PCR products of target genes were purified using PCR Quick-Spin PCR Product Purification Kit Ver.4.0 (TaKaRa, Japan) and cloned using pMD18-T Vector (TaKaRa, Japan). The sequences of the amplicons were deposited into GenBank under the accession

numbers listed in [S4 Table](#). Plasmids carrying target genes were extracted and purified using MiniBest Plasmid Purification Kit (TaKaRa, Japan). Purified plasmid concentrations were determined by microspectrophotometry (NanoDropND-2000, NanoDrop Technologies, Wilmington, DE) [36]. For q-PCR, the 20- μ L reaction volume contained 10 μ L of SYBR Premix^{EX}Taq Super Mix (TaKaRa, Japan), 8 μ L of DNA template, 10 μ M each primer and ddH₂O [37], and the reaction conditions included: initial denaturation at 94° for 3 min, followed by 40 thermal cycles ([S4 Table](#)) and final extension at 72°C for 45 s. The specificity of q-PCR products was confirmed by melting curves observation and agarose (1%) gel electrophoresis.

Results

Detection of potential pathogens by 454 pyrosequencing

Pyrosequencing of 16S rRNA gene amplicons revealed that the bacterial community shifted greatly along the treatment processes ([S2 Fig](#)). At genus level, the relative number of the sequences assigned to known pathogens was 3,428 (55.33%) for SI, 3,883 (62.63%) for PE, 102 (1.65%) for AS, 76 (1.23%) for SE, 83 (1.34%) for FFE and 56 (0.90%) for FRE ([S5 Table](#)), and a total of 20 genera containing potentially pathogenic species were detected in the six samples ([Fig 1A](#)). Among the genera identified, *Acrobacter* had the highest abundance in each of SI, PE, AS, SE and FFE, accounting for 43.42%-97.37% of all pathogenic sequences, and the genus also occupied 14.29% in FRE ([Fig 1A](#)). At species level, pyrosequencing showed that 1,422 (6 species, 22.95%), 1,302 (7 species, 21.00%), 23 (7 species, 0.37%), 23 (4 species, 0.37%), 21 (7 species, 0.34%) and 6 (4 species, 0.10%) sequences were closely related (with similarity over 97%) to known pathogens in SI, PE, AS, SE, FFE and FRE, respectively ([Fig 1B](#)), demonstrating that the sewage treatment processes can effectively reduce the relative abundance of the pathogens. In absolute terms, over 99% of the potential pathogens in SI (2.68×10^4 sequences per milliliter) were eliminated by the treatment processes ([S6 Table](#)). Among the species identified, *Arcobacter butzleri* had the highest abundance in the STP, and the relative abundance ranged from 33.33% to 94.87% of total pathogenic bacteria in the six samples ([Fig 1C](#)). Additionally, *Aeromonas hydrophila* was also found prevalent in SE and FRE ([Fig 1C](#)).

Detection of potential pathogens by Illumina sequencing

For a further investigation of potential pathogens in the STP, the Illumina sequencing reads assigned to 16S rRNA gene were used to determine the diversity of potential pathogens. Results showed that the potential pathogens in the STP were relatively affiliated with 29 genera in each of SI (6,952 sequences) and PE (10,086 sequences), among which *Arcobacter* had the highest relative abundance ([Fig 2A](#)). Both the conventional and advanced sewage treatment processes applied in the STP demonstrated high pathogen removal efficiency since only 440, 418, 292 and 219 sequences were found closely related to known pathogens in AS, SE, FFE and FRE, respectively ([S7 Table](#)). Comparison in absolute terms showed the similar results ([S6 Table](#)). Further analysis of the Illumina sequencing datasets by MetaPhlan showed that a total of 32 species of the potential pathogens were detected in the six samples ([Fig 2B](#)), among which 30, 25, 18, 7, 5 and 4 species were present in SI, PE, AS, SE, FFE and FRE, respectively. Among the species, *A. butzleri* was most abundant in each sample. *A. hydrophila*, *E. coli* and *K. pneumoniae* were also found prevalent in SI and PE (over 1.85% each). *A. hydrophila*, *Bordetella pertussis* and *Pseudomonas aeruginosa* also had high abundance in SE after oxidation ditch treatment. Comparatively, *A. hydrophila* had a decreased abundance and *Brucella suis* had an increased abundance in both FFE and FRE. *Bordetella pertussis* in FFE and *Pseudomonas aeruginosa* in FRE were also found to be the dominant potential pathogens.

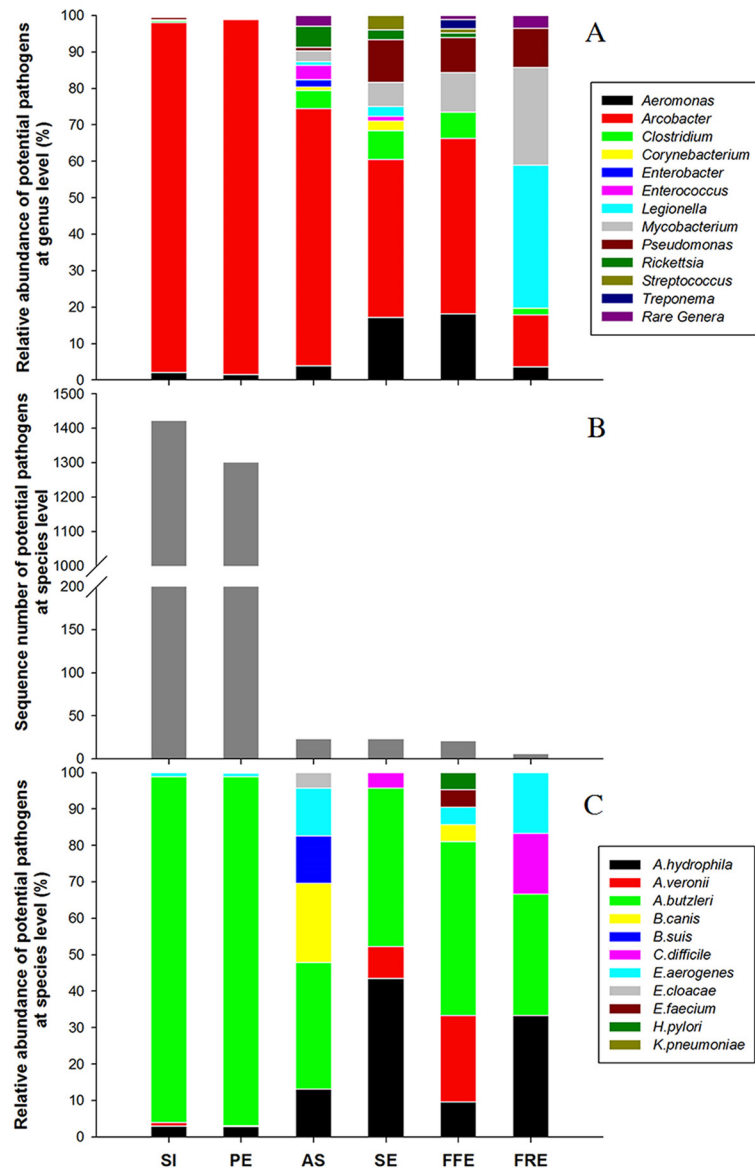


Fig 1. Occurrence and abundance of potential pathogens at genus and species levels revealed by 454 pyrosequencing of 16S rRNA gene amplicons. (A) Relative abundance of potential pathogens at genus level. The effective pyrosequencing reads were classified using MEGAN. Rare genera refer to the taxa with maximum abundance <1% in any sample, including *Bacillus*, *Campylobacter*, *Helicobacter*, *Klebsiella*, *Leptospira*, *Neisseria* and *Serratia*. (B) Sequence number of potential pathogens at species level under the same sequencing depth (6200 reads). (C) Relative abundance of potential pathogens at species level. All the effective sequences identified as potentially pathogenic species were aligned using the local BLASTN tool. The relative abundance was obtained by normalizing the sequence number of each pathogenic taxon to the total number of all pathogenic taxa in one sample.

doi:10.1371/journal.pone.0125549.g001

Removal of four pathogens quantified by q-PCR

To confirm the metagenomic results, q-PCR was used to investigate the removals of four potential pathogens in relative and absolute terms by different sewage treatment processes (Fig 3 and S4 Fig). Results consistently showed that *A. butzleri*, among the four pathogens, had the highest abundance in each sampling location and time point, followed by *A. hydrophila*, *E. coli*

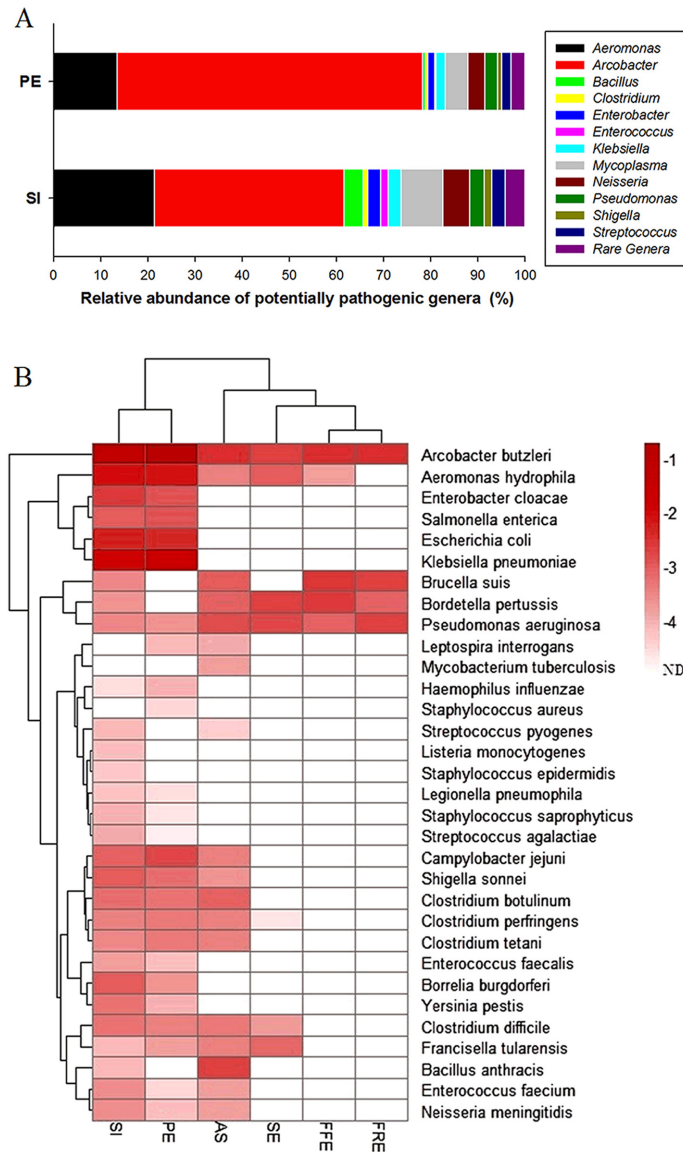


Fig 2. Diversity and relative abundance of potential pathogens at genus (A) and species (B) levels revealed by Illumina shotgun sequencing of the metagenomes. (A) Relative abundance of potentially pathogenic genera. The denoised reads were aligned using MEGAN against 16S rRNA gene Silva database. Rare genera, indicating the taxa with maximum abundance lower than 1% in any sample, include *Bordetella*, *Campylobacter*, *Corynebacterium*, *Escherichia*, *Francisella*, *Haemophilus*, *Helicobacter*, *Legionella*, *Leptospira*, *Listeria*, *Mycobacterium*, *Salmonella*, *Serratia*, *Staphylococcus*, *Treponema*, *Vibrio* and *Yersinia*. (B) Heat map illustrating relative abundance (log) of potential pathogens at species level generated by MetaPhlAn. The relative abundance was obtained by normalizing the sequence number of each pathogenic taxon to the total number of all pathogenic taxa in one sample.

doi:10.1371/journal.pone.0125549.g002

and *K. pneumoniae*, which agrees with the results of the high-throughput sequencing technologies. Overall, about 99% of pathogens in the raw sewage were finally removed in the STP, and comparison of the different treatment processes used in the STP showed that the removal effectiveness mainly depended on the oxidation ditch. Magnetic resin processes were found to remove 99% of *E. coli*, but sand filter tank seemed to accumulate the pathogen. Magnetic resin

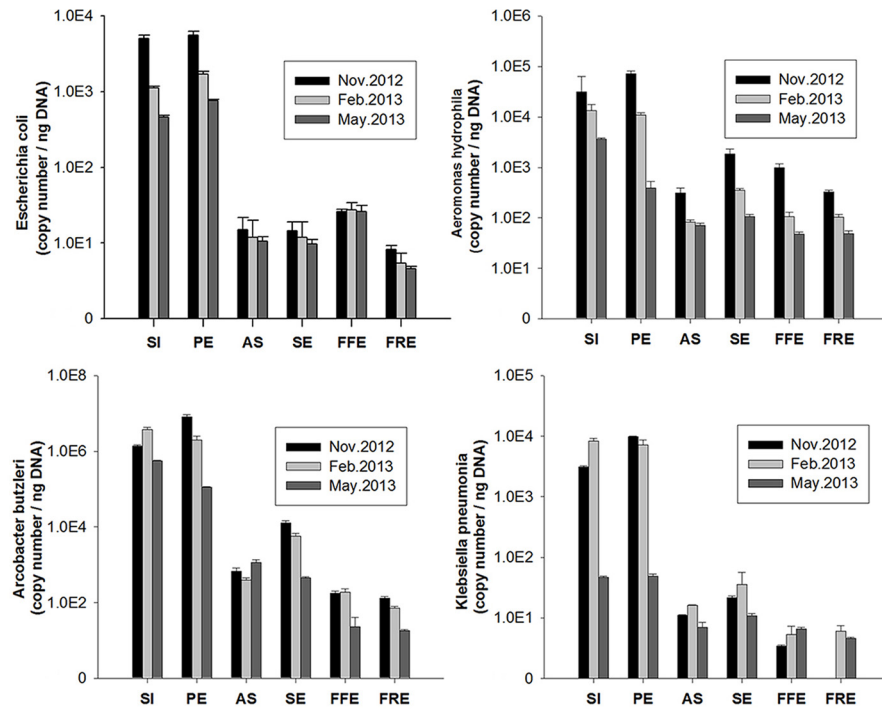


Fig 3. Abundance of four potentially pathogenic bacteria in the samples collected from the STP at different time points. The abundance was determined by quantitative real time PCR and normalized to one nanogram of extracted DNA.

doi:10.1371/journal.pone.0125549.g003

also showed higher removals in *A. hydrophila* and *K. pneumoniae* than sand filtration ($p < 0.05$ each), which is consistent with the results of metagenomic analyses.

Detection of VFs by Illumina high-throughput sequencing

BLAST against MvirDB protein database showed that 716 sequences from SI, 507 sequences from PE, 10 sequences from AS, 18 sequences from SE, 9 sequences from FFE and 6 sequences from FRE were annotated as the known pathogenicity islands, which were assigned to 94, 81, 9, 12, 3 and 2 types, respectively. A total of 208 sequences from SI, 165 sequences from PE, 13 sequences from AS, 15 sequences from SE, 9 sequences from FFE and 4 sequences from FRE were assigned to 31, 23, 5, 9, 5 and 4 types of known virulence proteins, respectively. Similar to pathogens, both pathogenicity islands and virulence proteins in the STP were mainly removed by oxidation ditch (Fig 4), which agrees with the results in absolute terms (S6 Table).

Discussion

In this study, potential pathogens in the STP were detected by joint use of 454 pyrosequencing and Illumina high-throughput sequencing, two promising technologies reliable for quantification of genetic diversity within environmental communities [38]. All sequences generated by the two technologies were blasted against Silva database and then imported into MEGAN using the same LCA algorithm parameters. LCA has been proved more accurate than Best Hit for annotating short sequences at genus or species level [23]. It has also been indicated that MEGAN can provide more reliable taxonomy results than RDP classifier in 454 pyrosequencing data process [22] and can assign sequences to species deposited at NCBI [27], so we used MEGAN for metagenomic analyses to detect the potential pathogens in the STP. Moreover,

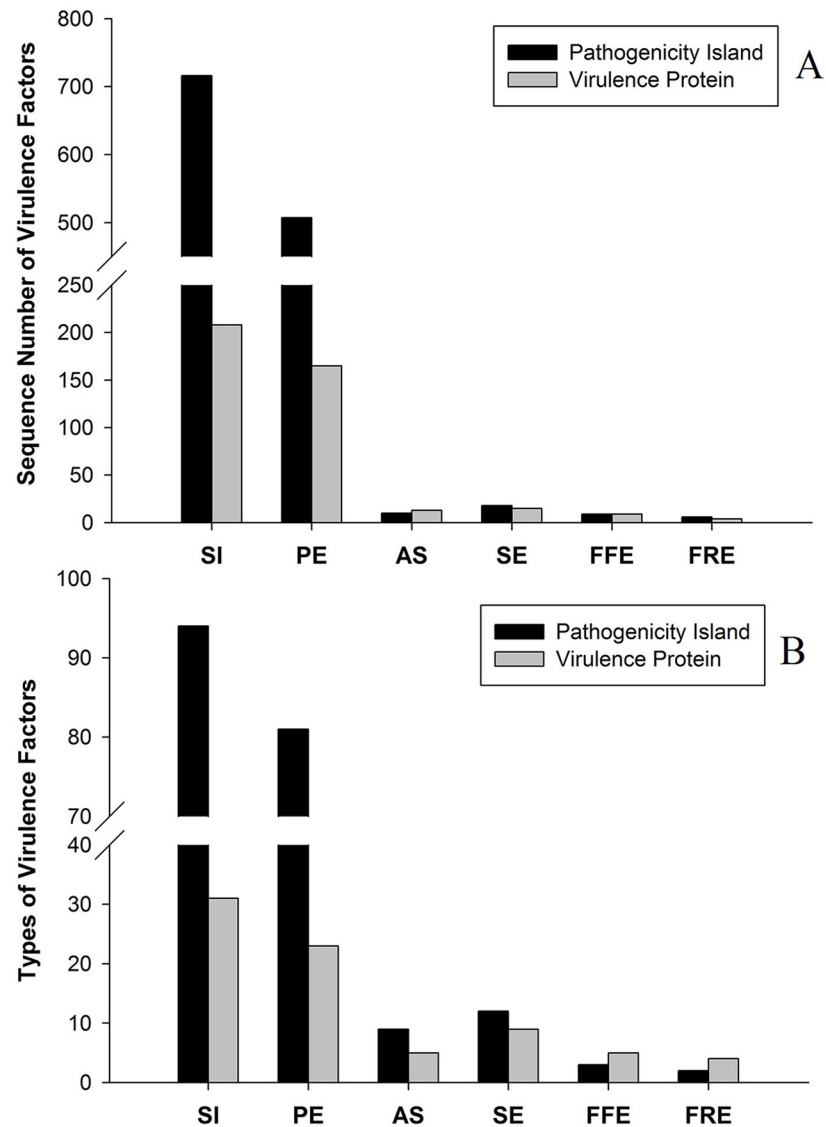


Fig 4. Metagenomic analyses of virulence factors (VFs) at six locations along sewage flow in the STP. Sequence number (A) and types (B) of virulence factors were calculated based on alignment of the Illumina shotgun sequences (normalized to 9,000,000 reads for each sample) against MvirDB protein database.

doi:10.1371/journal.pone.0125549.g004

MetaPhlAn, a more reliable tool to profile environmental communities using the database of clade-specific marker genes [5, 23], was applied to determine pathogens based on Illumina sequencing.

This study showed that Illumina sequencing portrayed a much more complex pathogenic community than 454 pyrosequencing. Due to bias from PCRs of 16S rRNA gene, pyrosequencing only identified 11 species of potentially pathogenic bacteria, overlooking some minor populations. The PCR bias can be avoided by direct shotgun sequencing of environmental genomes, so Illumina sequencing revealed more species of the potential pathogens in the STP [23].

Despite of the difference mentioned above, pyrosequencing and Illumina sequencing provided consistent results for detection of the major genera (e.g. *Arcobacter*) and species of

bacterial pathogens (e.g. *A. butzleri*, *A. hydrophila* and *K. pneumonia*) in the STP, which was further supported by q-PCR. Moreover, the three technologies consistently indicated that *A. butzleri* was the most abundant pathogenic species in the STP. *A. butzleri*, a free-living and waterborne pathogen, can cause gastrointestinal diseases such as diarrhea [39]. Previous studies have revealed that *Arcobacter* genus dominates in influent sewage of different STPs based on pyrosequencing of 16S rRNA gene [21]. Within the genus, *A. butzleri* usually has highest abundance in sewage, and presence of *Arcobacter* sp. in environmental waters indicates high levels of fecal pollution [40]. The pathogens *A. hydrophila*, *K. pneumoniae* and *E. coli* were also frequently detected in sewage by using molecular methods [22] or culture-dependent methods [41, 42].

VFs responsible for pathogenesis were further analyzed by Illumina sequencing and metagenomic analyses. This study revealed occurrence of various types of pathogenicity islands and virulence proteins in the STP. Pathogenicity islands, a distinct class of genomic islands acquired by horizontal gene transfer, are often incorporated in the genome of pathogenic organisms, but absent in nonpathogenic ones [43]. Some mobile genetic elements (e.g. transposon, integron, plasmids etc.), which are normally capable of gene transfer, can directly encode pathogenicity factors [44]. Virulence proteins, often acquired by horizontal genetic transfer [45], can facilitate bacterial pathogenesis by interfering with host cell signal transduction and other cellular processes [46]. Although types and abundances of virulence factors decreased dramatically after the sewage treatment, several ones with serious pathogenicity and virulence were still present in the final effluent (S8 and S9 Tables). For example, STM1008, a gisyl-2 prophage, can induce *Salmonella*'s ability to establish a systemic infection in mice [47]. Heat shock proteins (Hsp70), which are normally generated in response to adverse environmental conditions, were reported to contribute to virulence as adhesins for invading the host cell or in signaling the immune system [48].

Pyrosequencing and Illumina sequencing consistently showed that the abundance and diversity of the potential pathogens and VFs obviously declined along water stream of the sewage treatment systems. This changing trend was subsequently confirmed by the results of q-PCR. Oxidation ditch was found highly effective in removing the pathogens and VFs, which may result from the microbial structure shift from raw sewage to sludge and treated sewage. The removal of suspended solids by oxidation ditch may also contribute to the decrease in the quantity of the potential pathogens and VFs in per milliliter of sewage water. Previous studies have also revealed that aerobic treatment can contribute to the removal of pathogens and VFs in sewage based on molecular detection or bacterial culture. By microarray and q-PCR, Lee et al. [20] revealed that activated sludge process had removal efficiency of over 99% for each of 12 tested pathogens including *A. hydrophila*, *K. pneumoniae* and *E. coli*. Membrane filtration culture of pathogens also demonstrated that activated sludge process can remove nearly all cells of *A. hydrophila* [49]. By q-PCR, our previous studies have also revealed that integrases and transposases genes can be effectively eliminated from sewage by aerobic treatment [50, 51]. The possible explanation is that growth of the pathogenic bacteria dominating in anaerobic or anoxic environments of raw sewage was inhibited when the sewage entered in the aerobic tank.

Comparison between the two advanced sewage treatment processes indicated that magnetic resin treatment had higher removals in most of the potential pathogens and VFs than sand filtration. Little information is available about the effectiveness of the two processes in removing pathogens or VFs. It is known that magnetic resin process can effectively remove hydrophilic organic compounds [52] and heavy metals [53] from biologically treated secondary effluent within a short contact time. Additionally, magnetic resin can increase disinfection capability of ozone by reducing the ozone decay rates [54], but no studies have been conducted to investigate

the effect of magnetic resin on pathogen removal. Its removal effect on pathogen may mainly result from the reinforced capability of magnetic resin in adsorbing and accumulating residuum of organisms from treated secondary effluent [52]. Although most of the pathogens and virulence factors were eliminated in the STP, presence of the opportunistic pathogens (e.g. *A. butzleri*) in the effluent still deserves research concerns, as little information is available on their environmental fates in receiving water bodies.

To our knowledge, this is the first study jointly using the two different high-throughput sequencing technologies to investigate bacterial virulence in STPs. Technical complementation of the two sequencing methods and q-PCR can provide a more comprehensive and accurate insight into environmental bacterial virulence. However, compared with culture-based methods, the molecular technologies still have some limitations on specificity, since pathogenicity may vary among the different strains within one species [22]. For example, adhesiveness, invasiveness, cytotoxicity and diarrhea infectivity of *A. butzleri* strains varied among different animal hosts [39].

In conclusion, joint use of Illumina high-throughput sequencing, 454 pyrosequencing and q-PCR reveals that the STPs are important environmental reservoirs of various potential pathogens, among which *Arcobacter* are prevalent. *A. butzleri*, *A. hydrophila* and *K. pneumonia* are the major species of potential pathogens dominating in raw sewage. Illumina sequencing also reveals occurrence of various types of pathogenicity islands and virulence proteins in the STPs. Most of the potential pathogens and VFs can be eliminated in the STPs applying conventional and advanced treatment processes, and the removal efficiency mainly depends on aerobic treatment. As a novel advanced treatment process, magnetic resin shows higher removals in most of the potential pathogens and VFs than sand filtration. However, *A. butzleri* can still persist in the final effluent, which deserves more concerns.

Supporting Information

S1 Table. Information of operational processes and wastewater quality of the sewage treatment plant.

(DOCX)

S2 Table. Information of the water/sludge sampling from the sewage treatment plant and accession numbers of the sequencing datasets deposited into publicly available databases.

(DOCX)

S3 Table. Sequencing information of Illumina high-throughput sequencing and 454 pyrosequencing data analyzed.

(DOCX)

S4 Table. Primer sequences and conditions for q-PCR of the target genes indicating human bacterial pathogens.

(DOCX)

S5 Table. Numbers of the sequences assigned to genera containing potentially pathogenic species obtained from annotation of 454 pyrosequencing reads by using MEGAN.

(DOCX)

S6 Table. Absolute abundance of genera containing potentially pathogenic species and virulence factors detected in sewage samples.

(DOCX)

S7 Table. Numbers of the sequences assigned to genera containing potentially pathogenic species obtained from annotation of Illumina high-throughput sequencing reads by using MEGAN.

(DOCX)

S8 Table. Pathogenicity islands detected at six locations along sewage flow in the STP by Illumina high-throughput sequencing.

(DOCX)

S9 Table. Virulence proteins detected at six locations along sewage treatment in the STP by Illumina high-throughput sequencing.

(DOCX)

S1 Fig. Sampling locations along the sewage treatment processes in Wulongkou STP located in Zhengzhou City, China.

(TIF)

S2 Fig. Relative abundance of different phyla and classes in *Proteobacteria* at six stages of sewage treatment.

(TIF)

S3 Fig. Percentages of unclassified pyrosequencing reads at seven taxonomic levels for the six samples.

(TIF)

S4 Fig. Absolute abundance of four potentially pathogenic bacteria in sewage samples collected from the STP at different time points.

(TIF)

Author Contributions

Conceived and designed the experiments: XL XXZ YW BL. Performed the experiments: XL WL JT. Analyzed the data: XL XXZ ZW KH JT. Contributed reagents/materials/analysis tools: XXZ BL YW. Wrote the paper: XL XXZ BL YT.

References

1. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res.* 2007; 35(Database issue): D391–394. PMID: [17090593](#)
2. Okoh AI, Odjadjare EE, Igbinosa EO, Osode AN. Wastewater treatment plants as a source of microbial pathogens in receiving watersheds. *Afr J Biotechnol.* 2007; 6(25): 2932–2944.
3. Varela AR, Manaia CM. Human health implications of clinically relevant bacteria in wastewater habitats. *Environ Sci Pollut Res Int.* 2013; 20(6): 3550–3569. doi: [10.1007/s11356-013-1594-0](#) PMID: [23508533](#)
4. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 2005; 33(Database issue): D325–328. PMID: [15608208](#)
5. George T, Franklin L, Stensel H. *Wastewater engineering: treatment and reuse.* Edition 4th. Metcalf and Eddi Inc. 2003.
6. Sonune A, Ghate R. Developments in wastewater treatment methods. *Desalination.* 2004; 167: 55–63.
7. Godfree A, Farrell J. Processes for managing pathogens. *J Environ Qual.* 2005; 34(1): 105–113. PMID: [15647539](#)
8. Fu C, Xie X, Huang J, Zhang T, Wu Q, Chen J, et al. Monitoring and evaluation of removal of pathogens at municipal wastewater treatment plants. *Water Sci Technol.* 2010; 61(6): 1589–1599. doi: [10.2166/wst.2010.757](#) PMID: [20351439](#)

9. Lucena F, Duran A, Moron A, Calderon E, Campos C, Gantzer C, et al. Reduction of bacterial indicators and bacteriophages infecting faecal bacteria in primary and secondary wastewater treatments. *J Appl Microbiol*. 2004; 97(5): 1069–1076. PMID: [15479424](#)
10. Wen Q, Tutuka C, Keegan A, Jin B. Fate of pathogenic microorganisms and indicators in secondary activated sludge wastewater treatment plants. *J Environ Manage*. 2009; 90(3): 1442–1447. doi: [10.1016/j.jenvman.2008.09.002](#) PMID: [18977580](#)
11. Cheng HWA, Lucy FE, Broaders MA, Mastitsky SE, Chen CH, Murray A. Municipal wastewater treatment plants as pathogen removal systems and as a contamination source of noroviruses and *Enterococcus faecalis*. *J Water Health*. 2012; 10(3): 380–389. doi: [10.2166/wh.2012.138](#) PMID: [22960482](#)
12. Olsen JS, Aarskaug T, Thrane I, Pourcel C, Ask E, Johansen G, et al. Alternative routes for dissemination of *Legionella pneumophila* causing three outbreaks in Norway. *Environ Sci Technol*. 2010; 44(22): 8712–8717. doi: [10.1021/es1007774](#) PMID: [20949911](#)
13. Huang J-J, Hu H-Y, Tang F, Li Y, Lu S-Q, Lu Y. Inactivation and reactivation of antibiotic-resistant bacteria by chlorination in secondary effluents of a municipal wastewater treatment plant. *Water Res*. 2011; 45(9): 2775–2781. doi: [10.1016/j.watres.2011.02.026](#) PMID: [21440281](#)
14. Richardson SD, Postigo C. Drinking water disinfection by-products, in emerging organic contaminants and human health. 2012, Springer. P. 93–137.
15. Shi P, Jia S, Zhang XX, Zhang T, Cheng S, Li A. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water Res*. 2013; 47(1): 111–120. doi: [10.1016/j.watres.2012.09.046](#) PMID: [23084468](#)
16. Savichtcheva O, Okabe S. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. *Water Res*. 2006; 40(13): 2463–2476. PMID: [16808958](#)
17. Vartoukian SR, Palmer RM, Wade WG. Strategies for culture of ‘unculturable’ bacteria. *FEMS Microbiol Lett*. 2010; 309(1): 1–7. doi: [10.1111/j.1574-6968.2010.02000.x](#) PMID: [20487025](#)
18. Toze S. PCR and the detection of microbial pathogens in water and wastewater. *Water Res*. 1999; 33(17): 3545–3556.
19. Srinivasan S, Aslan A, Xagorarakis I, Alocilja E, Rose JB. *Escherichia coli*, *enterococci*, and *Bacteroides thetaiotaomicron* qPCR signals through wastewater and septage treatment. *Water Res*. 2011; 45(8): 2561–2572. doi: [10.1016/j.watres.2011.02.010](#) PMID: [21420709](#)
20. Lee DY, Shannon K, Beaudette LA. Detection of bacterial pathogens in municipal wastewater using an oligonucleotide microarray and real-time quantitative PCR. *J Microbiol Methods*. 2006; 65(3): 453–467. PMID: [16239042](#)
21. Zhang T, Shao MF, Ye L. 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *ISME J*. 2012; 6(6): 1137–1147. doi: [10.1038/ismej.2011.188](#) PMID: [22170428](#)
22. Ye L, Zhang T. Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. *Environ Sci Technol*. 2011; 45(17): 7173–7179. doi: [10.1021/es201045e](#) PMID: [21780772](#)
23. Cai L, Zhang T. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environ Sci Technol*. 2013; 47(10): 5433–5441. doi: [10.1021/es400275r](#) PMID: [23594284](#)
24. Bibby K, Peccia J. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ Sci Technol* 2013; 47(4): 1945–1951. doi: [10.1021/es305181x](#) PMID: [23346855](#)
25. Wang Z, Zhang XX, Huang K, Miao Y, Shi P, Liu B, et al. Metagenomic profiling of antibiotic resistance genes and mobile genetic elements in a tannery wastewater treatment plant. *PLOS ONE*. 2013; 8(10): e76079. doi: [10.1371/journal.pone.0076079](#) PMID: [24098424](#)
26. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007; 35(21): 7188–7196. PMID: [17947321](#)
27. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. Integrative analysis of environmental sequences using MEGAN4. *Genome Res*. 2011; 21(9): 1552–1560. doi: [10.1101/gr.120618.111](#) PMID: [21690186](#)
28. Ye L, Zhang T, Wang T, Fang Z. Microbial structures, functions, and metabolic pathways in wastewater treatment bioreactors revealed using high-throughput sequencing. *Environ Sci Technol*. 2012; 46(24): 13244–13252. doi: [10.1021/es303454k](#) PMID: [23151157](#)
29. Segata N, Waldron L, Ballarín A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods*. 2012; 9(8): 811–814. doi: [10.1038/nmeth.2066](#) PMID: [22688413](#)

30. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009; 75(23): 7537–7541. doi: [10.1128/AEM.01541-09](https://doi.org/10.1128/AEM.01541-09) PMID: [19801464](https://pubmed.ncbi.nlm.nih.gov/19801464/)
31. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLOS ONE.* 2011; 6(12): e27310. doi: [10.1371/journal.pone.0027310](https://doi.org/10.1371/journal.pone.0027310) PMID: [22194782](https://pubmed.ncbi.nlm.nih.gov/22194782/)
32. Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. Opportunistic pathogens enriched in showerhead biofilms. *Proc Natl Acad Sci USA.* 2009; 106(38): 16393–16399. doi: [10.1073/pnas.0908446106](https://doi.org/10.1073/pnas.0908446106) PMID: [19805310](https://pubmed.ncbi.nlm.nih.gov/19805310/)
33. Ahmed W, Huygens F, Goonetilleke A, Gardner T. Real-time PCR detection of pathogenic microorganisms in roof-harvested rainwater in Southeast Queensland, Australia. *Appl Environ Microbiol.* 2008; 74(17): 5490–5496. doi: [10.1128/AEM.00331-08](https://doi.org/10.1128/AEM.00331-08) PMID: [18621865](https://pubmed.ncbi.nlm.nih.gov/18621865/)
34. Abdelbaqi K, Buissonniere A, Prouzet-Mauleon V, Gresser J, Wesley I, Megraud F, et al. Development of a real-time fluorescence resonance energy transfer PCR to detect arcobacter species. *J Clin Microbiol.* 2007; 45(9): 3015–3021. PMID: [17652482](https://pubmed.ncbi.nlm.nih.gov/17652482/)
35. Sun F, Wu D, Qiu Z, Jin M, Wang X, Li J. Development of real-time PCR systems based on SYBR Green for the specific detection and quantification of *Klebsiella pneumoniae* in infant formula. *Food Control.* 2010; 21(4): 487–491.
36. Zhang XX, Zhang T. Occurrence, abundance, and diversity of tetracycline resistance genes in 15 sewage treatment plants across China and other global locations. *Environ Sci Technol.* 2011; 45(7): 2598–2604. doi: [10.1021/es103672x](https://doi.org/10.1021/es103672x) PMID: [21388174](https://pubmed.ncbi.nlm.nih.gov/21388174/)
37. Zhang T, Zhang M, Zhang XX, Fang HH. Tetracycline resistance genes and tetracycline resistant lactose-fermenting *Enterobacteriaceae* in activated sludge of sewage treatment plants. *Environ Sci Technol.* 2009; 43(10): 3455–3460. PMID: [19544839](https://pubmed.ncbi.nlm.nih.gov/19544839/)
38. Luo CW, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLOS ONE.* 2012; 7(2): e30087. doi: [10.1371/journal.pone.0030087](https://doi.org/10.1371/journal.pone.0030087) PMID: [22347999](https://pubmed.ncbi.nlm.nih.gov/22347999/)
39. Collado L, Figueras MJ. Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clin Microbiol Rev.* 2011; 24(1): 174–192. doi: [10.1128/CMR.00034-10](https://doi.org/10.1128/CMR.00034-10) PMID: [21233511](https://pubmed.ncbi.nlm.nih.gov/21233511/)
40. Collado L, Inza I, Guarro J, Figueras MJ. Presence of *Arcobacter* spp. in environmental waters correlates with high levels of fecal pollution. *Environ Microbiol.* 2008; 10(6): 1635–1640. doi: [10.1111/j.1462-2920.2007.01555.x](https://doi.org/10.1111/j.1462-2920.2007.01555.x) PMID: [18215159](https://pubmed.ncbi.nlm.nih.gov/18215159/)
41. Monfort P, Baleux B. Dynamics of *Aeromonas hydrophila*, *Aeromonas sobria*, and *Aeromonas caviae* in a sewage treatment pond. *Appl Environ Microbiol.* 1990; 56(7): 1999–2006. PMID: [2389929](https://pubmed.ncbi.nlm.nih.gov/2389929/)
42. Elmund GK, Allen MJ, Rice EW. Comparison of *Escherichia coli*, total coliform, and fecal coliform populations as indicators of wastewater treatment efficiency. *Water Environ Res.* 1999; 71(3): 332–339.
43. Pillar CM, Gilmore MS. Enterococcal virulence—Pathogenicity island of E-faecalis. *Front Biosci.* 2004; 9: 2335–2346. PMID: [15353291](https://pubmed.ncbi.nlm.nih.gov/15353291/)
44. Hacker J, Kaper JB. Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol.* 2000; 54: 641–679. PMID: [11018140](https://pubmed.ncbi.nlm.nih.gov/11018140/)
45. Finlay BB, McFadden G. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. *Cell.* 2006; 124(4): 767–782. PMID: [16497587](https://pubmed.ncbi.nlm.nih.gov/16497587/)
46. Hueck CJ. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev.* 1998; 62(2): 379–433. PMID: [9618447](https://pubmed.ncbi.nlm.nih.gov/9618447/)
47. Figueroa—Bossi N, Bossi L. Inducible prophages contribute to *Salmonella* virulence in mice. *Mol Microbiol.* 1999; 33(1): 167–176. PMID: [10411733](https://pubmed.ncbi.nlm.nih.gov/10411733/)
48. Neckers L, Tatu U. Molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell Host Microbe.* 2008; 4(6): 519–527. doi: [10.1016/j.chom.2008.10.011](https://doi.org/10.1016/j.chom.2008.10.011) PMID: [19064253](https://pubmed.ncbi.nlm.nih.gov/19064253/)
49. Poffe R, Beeck EO. Enumeration of *Aeromonas hydrophila* from domestic wastewater treatment plants and surface waters. *J Appl Microbiol.* 1991; 71(4): 366–370.
50. Ma LP, Zhang XX, Zhao FZ, Wu B, Cheng SP, Yang LY. Sewage treatment plant serves as a hot-spot reservoir of integrons and gene cassettes. *J Environ Biol.* 2013; 34(2): 391–399. PMID: [24620610](https://pubmed.ncbi.nlm.nih.gov/24620610/)
51. Ma L, Zhang X-X, Cheng S, Zhang Z, Shi P, Liu B, et al. Occurrence, abundance and elimination of class 1 integrons in one municipal sewage treatment plant. *Ecotoxicology.* 2011; 20(5): 968–973. doi: [10.1007/s10646-011-0652-y](https://doi.org/10.1007/s10646-011-0652-y) PMID: [21431316](https://pubmed.ncbi.nlm.nih.gov/21431316/)
52. Zhang R, Vigneswaran S, Ngo HH, Nguyen H. Magnetic ion exchange (MIEX) resin as a pre-treatment to a submerged membrane system in the treatment of biologically treated wastewater. *Desalination.* 2006; 192(1–3): 296–302.

53. Atia AA, Donia AM, Yousif AM. Removal of some hazardous heavy metals from aqueous solution using magnetic chelating resin with iminodiacetate functionality. *Sep Purif Technol.* 2008; 61(3): 348–357.
54. Wert EC, Edwards-Brandt JC, Singer PC, Budd GC. Evaluating magnetic ion exchange resin (MIEX) pretreatment to increase ozone disinfection and reduce bromate formation. *Ozone: Sci Eng.* 2005; 27(5): 371–379.