https://doi.org/10.1016/j.scitotenv.2019.03.278

Greay, T.L., Gofton, A.W., Zahedi, A., Paparini, A., Linge, K.L., Joll, C.A. and Ryan, U.M. (2019) Evaluation of 16S next-generation sequencing of hypervariable region 4 in wastewater samples: An unsuitable approach for bacterial enteric pathogen identification. Science of The Total Environment

Evaluation of 16S next-generation sequencing of hypervariable region 4 in wastewater samples: An unsuitable approach for bacterial enteric pathogen identification

Telleasha L. Greay, Alexander W. Gofton, Alireza Zahedi, Andrea Paparini, Kathryn L. Linge, Cynthia A. Joll, Una M. Ryan

PII:	S0048-9697(19)31270-7
DOI:	https://doi.org/10.1016/j.scitotenv.2019.03.278
Reference:	STOTEN 31488
To appear in:	*Science of the Total Environment*
Received date:	5 November 2018
Revised date:	15 March 2019
Accepted date:	18 March 2019

Evaluation of 16S next-generation sequencing of hypervariable region 4 in wastewater samples: an unsuitable approach for bacterial enteric pathogen identification

Telleasha L Greay[1,2], Alexander W Gofton[1], Alireza Zahedi[1,2], Andrea Paparini[1], Kathryn L Linge[3,4], Cynthia A Joll[3] and Una M Ryan[1*]

[1]Vector and Waterborne Pathogens Research Group, School of Veterinary and Life Sciences, Murdoch University, Perth, Western Australia, Australia

[2]Western Australian State Agricultural Biotechnology Centre, Murdoch University, Perth, Western Australia, Australia

[3]Curtin Water Quality Research Centre, Chemistry, School of Molecular and Life Sciences, Curtin University, GPO Box U1987, Perth, Australia

[4]ChemCentre, PO Box 1250, Perth, Western Australia, Australia

* Correspondence: una.ryan@murdoch.edu.au

Emails:

TLG: telleasha.greay@outlook.com

AWG: alexander.gofton@csiro.au

AZ: a.zahediabdi@murdoch.edu.au

AP: a.paparini@murdoch.edu.au

KLL: klinge@chemcentre.wa.gov.au

CAJ: c.joll@curtin.edu.au

Abstract

Recycled wastewater can carry human-infectious microbial pathogens and therefore wastewater treatment strategies must effectively eliminate pathogens before recycled wastewater is used to supplement drinking and agricultural water supplies. This study characterised the bacterial composition of four wastewater treatment plants (WWTPs) (three waste stabilisation ponds and one oxidation ditch WWTP using activated sludge treatment) in Western Australia. The hypervariable region 4 (V4) of the bacterial 16S rRNA (16S) gene was sequenced using next-generation sequencing (NGS) on the Illumina MiSeq platform. Sequences were pre-processed in USEARCH v10.0 and denoised into zero-radius taxonomic units (ZOTUs) with UNOISE3. Taxonomy was assigned to the ZOTUs using QIIME 2 and the Greengenes database and cross-checked with the NCBI nr/nt database. Bacterial composition of all WWTPs and treatment stages (influent, intermediate and effluent) were dominated by Proteobacteria (29.0-87.4%), particularly Betaproteobacteria (9.0-53.5%) and Gammaproteobacteria (8.6-34.6%). Nitrifying bacteria (*Nitrospira* spp.) were found only in the intermediate and effluent of the oxidation ditch WWTP, and denitrifying and floc-forming bacteria were detected in all WWTPs, particularly from the families Comamonadaceae and Rhodocyclales. Twelve pathogens were assigned taxonomy by the Greengenes database, but comparison of sequences from genera and families known to contain pathogens to the NCBI nr/nt database showed that only three pathogens (*Arcobacter venerupis*, *Laribacter hongkongensis* and *Neisseria canis*) could be identified in the dataset at the V4 region. Importantly, Enterobacteriaceae genera could not be differentiated. Family level taxa assigned by Greengenes database agreed with NCBI nr/nt in most cases, however, BLAST analyses revealed erroneous taxa in Greengenes database. This study highlights the importance of validating taxonomy of NGS sequences with databases such as NCBI nr/nt, and recommends including the V3 region of 16S in future short amplicon NGS studies that aim to identify bacterial enteric pathogens, as this will improve taxonomic resolution of most, but not all, Enterobacteriaceae species.

Keywords: Wastewater, next-generation sequencing, 16S rRNA, V4, Greengenes, Enterobacteriaceae.

1. Introduction

Water is becoming an increasingly scarce global resource, and as the overall demand for water grows, the quantity of wastewater produced and its overall pollution load are continuously increasing worldwide (Connor et al., 2017). Recycled wastewater is an essential resource in addressing this problem, as properly treated water can be safely released back into the environment, and used to supplement limited drinking water supplies. However, unless effectively treated, recycled wastewater has the potential to carry microbial pathogens (viruses, bacteria, protozoa and helminths), toxic chemicals and heavy metals. Therefore, treatment strategies must effectively eliminate these major public health risks (Rodriguez-Manzano et al., 2012).

Wastewater recycling in urban areas typically employs reverse osmosis membranes or advanced oxidation treatment after activated sludge wastewater treatment. This results in high purity recycled water, fit for potable reuse, but is technically challenging and expensive (Rajasulochana and Preethy, 2016; Garrido-Cardenas et al., 2017). In contrast, rural WWTPs typically use simple, non-mechanical waste stabilisation ponds (WSPs) or lagoons consisting of open basins that rely on natural microorganisms and algae to assist in the breakdown and settlement of degradable organic matter. Wastewater "influent" enters on one side of the WSP and exits on the other side as "effluent", after spending days or even months undergoing treatment processes in the pond, depending on plant capacity and flow rate. The treated effluent is discharged generally for non-potable purposes, such as irrigation of public open spaces or agricultural/horticultural uses (Von Sperling, 2007; Anon, 2009). These WSPs are widely used across the world as passive wastewater treatment for domestic wastewaters as they can offer low cost, low maintenance and effective pathogen removal (Von Sperling, 2007; Ho et al., 2017; Eland et al., 2018).

Removal and inactivation of pathogens from WSPs is achieved via long retention times, increased temperature and pH, the presence of algal antibacterial compounds and sunlight penetration. Therefore shallow (<1 m) WSPs with low turbidity, high pH and maximal sunlight exposure will achieve the most efficient pathogen removal (Sharafi et al., 2012). While WSP systems can achieve high removal efficiencies (4-6 $\log_{10}$), the efficiency of pathogen removal in full-scale systems is highly variable, and many WSP systems achieve only 2-3 $\log_{10}$ removal (Verbyla et al., 2017).

In contrast to WSPs, many conventional WWTPs use an activated sludge process in which a biological sludge containing living microorganisms is mixed with wastewater and aerated in a reactor, forming a mixed liquor. Microbial populations within the activated sludge include a range of bacteria, yeast, fungi, protozoa and higher organisms such as rotifers that can digest organic matter in wastewater, and clump together (by flocculation), producing a treated wastewater that is relatively free from suspended solids and organic material. The removal mechanisms of pathogens in an activated sludge system are inactivation, hunting by ciliate protozoa, adsorption to solids and capsulation inside the sludge flocs (Sharafi et al., 2012).

Understanding the diversity of bacterial microorganisms in wastewater is essential for understanding the performance for biological wastewater treatment systems (Inaba et al., 2017). DNA-based approaches for identification of bacteria, such as polymerase chain reaction (PCR) and Sanger sequencing, can overcome the limitations of conventional bacterial identification techniques (e.g. microscopy, culture-dependent assays and biochemical techniques) that are laborious and time-consuming, by allowing for the identification of microbes that are morphologically indistinguishable, uncultivable, fastidious, and obligate intracellular. Molecular bacterial identification approaches often target the 16S rRNA (16S) gene, which enables species differentiation based on genetic dissimilarity. However, the throughput of species identification with PCR and Sanger sequencing is limited by individual clone library preparation, and species-specific PCR approaches require *a priori* hypotheses regarding the taxa to be targeted. Wastewater can be comprised of hundreds of bacterial species (Berlec 2012; Kim et al. 2015), therefore assessments of bacterial diversity on this scale using PCR and/or Sanger sequencing is impractical. The rapid advances of next-generation sequencing (NGS) technologies have revolutionised the ability to identify large numbers of bacteria from various types of environmental and biological samples (Garrido-Cardenas et al., 2017). Primers targeting one or more of the nine hypervariable (V) regions of 16S can be used with NGS to identify bacteria. Other studies that have used 16S NGS to identify bacteria in WWTPs have targeted V3-4 (Lu et al., 2015), V4 (Zhang et al., 2012) and V5-6 (McLellan et al., 2010), and there is no consensus on the most suitable region to target for bacterial assessments in WWTPs. The V4 region of 16S is commonly targeted in microbiome studies with the widely used 515F/806R primers (Caporaso et al., 2011). These primers are recommended in the Earth Microbiome Project’s Illumina NGS protocol (http://www.earthmicrobiome.org/protocols-and-standards/16s/) and have been modified by other studies to include additional degeneracies to allow amplification of additional taxa (Apprill et al., 2015; Parada et al., 2015). Therefore, the present study evaluated the ability of the V4 16S NGS to identify bacteria, particularly enteric pathogens, in WWTPs, and used this NGS approach to characterise bacterial compositions in different treatment stages (influent, intermediate and effluent) of three WSPs and one oxidation ditch WWTP, which is a modified activated sludge WWTP, that utilises prolonged aeration to remove biodegradable organic compounds (Baars, 1962), in Western Australia (WA).

2. Methods

2.1 Study sites and sample collection

Wastewater samples (*n* = 26) were collected from three WSPs (WWTPs 1, 2 and 3) and an oxidation ditch (WWTP 4) in 2015 in WA (Table 1 and Figure 1). Samples were collected in February, July and September in 2015 and covered two seasons for each site. Samples were collected from WWTP 1, located in north-west WA and in a tropical climate, during the wet and dry seasons, while samples from WWTPs 2, 3 and 4, located in south-west WA and in a temperate climate, were collected during summer and winter (Table 1). Wastewater samples were also collected at different treatment stages (influent, intermediate and effluent) during summer and winter (or dry and wet seasons for WWTP 1 samples) (Table 1). The wastewater samples were collected in 1 L sterile containers that were treated with chlorine and rinsed with the sample before filling. Samples were kept cool in an ice box during transport back to the laboratory, and then stored at 4 °C and processed within 48 hours prior to DNA isolation.

2.2 DNA isolation

After 100 mL of each wastewater sample was filtered through sterile 0.2 µm Sterivex filters (Millipore, USA), genomic DNA (gDNA) was extracted from the filters using a PowerWater Sterivex DNA Isolation Kit (MO BIO Laboratories, California, USA). Extraction reagent blank controls (ExCs; *n* = 6) were included alongside each batch of gDNA extractions. Purified DNA was stored at -20 °C prior to molecular analysis.

2.3 Next-generation sequencing library preparation

The NGS library was prepared and sequenced following the 16S metagenomic sequencing library preparation protocol from Illumina (Part # 15044223 Rev. B; Illumina, USA), with minor modifications to the first stage PCR. V4 16S was amplified using modified 515F/806R primers [originally designed by Caporaso et al. (2011)]: 515FB 5′-GTGYCAGCMGCCGCGGTAA-3′ (Parada et al., 2015) and 806RB 5′-GGACTACNVGGGTWTCTAAT-3′ (Aprill et al., 2015). The 515FB/806RB primers were modified to include Illumina MiSeq adapter sequences (Part # 15044223 Rev. B; Illumina, USA), and conventional PCRs were carried out as described elsewhere (www.earthmicrobiome.org/protocols-and-standards/16s/; https://doi.org/10.17504/protocols.io.nuudeww). No-template controls (NTCs) were included alongside each PCR. The V4 16S library was sequenced on the Illumina Miseq platform (San Diego, CA, USA) with v2 sequencing chemistry.

2.4 16S Bioinformatic analysis

Paired-end 16S reads were merged (minimum 50 bp overlap), trimmed of primers and distal bases, quality filtered (maximum expected error threshold of 1.0) and singletons were removed with USEARCH v10.0 (Edgar, 2010), resulting in reads that were 247 bp in length on average. Reads were denoised into zero-radius operational taxonomic units (ZOTUs) and chimeras were filtered with UNOISE3 (Edgar, 2016). Taxonomic assignment of ZOTUs was performed in QIIME 2 v2018.2 (Caporaso et al., 2010, https://qiime2.org) using the QIIME 2 feature classifier plugin (Bokulich et al, 2018) and the August 2013 release of the Greengenes sequence database (McDonald et. al., 2012). The sequences were also BLAST searched against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide (nr/nt) database to cross-check Greengenes assigned taxonomy. ZOTUs that were in low abundance (<0.05% sequence composition) across all samples may represent PCR or sequencing error, therefore, they were bioinformatically removed from the samples. To assess sequencing depth, alpha rarefaction plots were generated with the R package vegan (Oksanen et al., 2018) using R software (R Core Team, 2013).

2.5 Phylogenetic analysis

Enterobacteriaceae ZOTUs were aligned using the MAFFT program (Katoh et al., 2002) with closely related sequences retrieved from the NCBI nr/nt database in Geneious v10.2.2 (http://www.geneious.com, Kearse et al., 2012). Sequences in the alignment were trimmed to the same length, then were imported into the program PhyML (Guindon et al., 2010) and assessed for the most appropriate nucleotide substitution model (GTR+G+I) based on Akaike Information Criterion (AIC). Maximum likelihood trees were constructed using RAxML (Stamatakis, 2014). Genetic distance estimates were calculated with Kimura distance matrices (Kimura, 1980) in Geneious v10.2.2.

ZOTU sequences generated from this study have been submitted to GenBank under the accession numbers MH892609 to MH892828. Raw sequence files were deposited in the NCBI Sequence Read Archive under the accession number PRJNA526519 (refer to Table 1 for sample names and metadata).

3. Results

3.1 Next-generation sequencing library summary

Approximately 1.4 million paired-end V4 16S sequences were obtained for all samples and controls (*n* = 34) (Table 2). After the reads were pre-processed (merged, quality filtered with singletons and chimeras removed), there was a total of ~800,000 sequences for all samples (~24,000 average). The processed 16S sequences (total of ~700,000) excluded sequences that were not classified as bacteria and low abundance (<0.05%) ZOTUs, and on average, there were ~27,000 processed bacterial 16S sequences for the WWTP samples (*n* = 26). Few sequences were detected in the ExCs and NTCs, which had an average of 8 sequences (Table 2).

3.2 Bacterial sequence composition in WWTPs

A total of 3,598 ZOTUs (Supplementary File B.1) were obtained for the pre-processed sequences, and a total of 1,644 ZOTUs remained for the processed sequences. For the processed sequences, sequencing depth was adequate for all samples at ~5,000 sequences (Supplementary Figures A.3 and A.4), but the alpha rarefaction plots did not reach a plateau for the pre-processed sequences (Supplementary Figures A.1 and A.2). The archaeal sequence compositions were low and two archaeal phyla were detected: Euryarchaeota was found in the influent of WWTP 4 (<0.1%) and effluent of WWTP 2 (0.1%), and Parvarchaeota was found in the effluent of WWTP 4 (0.1%). Two different types of Euryarchaeota were detected, *Methanobrevibacter* sp. from the class Methanobacteria in WWTP 4 influent and *Methanosaeta* sp. from the class Methanomicrobia in WWTP 2 effluent. The taxonomy for Parvarchaeota was assigned as Parvarchaea for class, and WCHD3-30 and YLA114 for Parvarchaea orders, with no further taxonomic classifications assigned by Greengenes (Supplementary File B.2).

Bacteria were classified into 28 phyla (Supplementary File B.2); the most dominant phylum was Proteobacteria across all WWTPs and treatment stages (influent, intermediate and effluent), with sequence compositions ranging from 29.0% in the effluent of WWTP 2 to 87.4% in the intermediate stage of WWTP 3 (Figure 2). Other abundant phyla (>10% composition in WWTP samples) were Bacteroidetes (ranging from 4.1% in WWTP 1 influent to 31.5% WWTP 3 effluent), Cyanobacteria (0% (not detected) in WWTP 1 and 3 influent to

47.2% WWTP 2 effluent), Firmicutes (0.1% in WWTP 3 effluent to 22.1% in WWTP 1 influent) and Actinobacteria (1.1% in WWTP 4 influent to 10.3% in WWTP 2 influent) (Figure 2).

Six classes of Proteobacteria were identified: Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria, Gammaproteobacteria and "TA18". Betaproteobacteria and Gammaproteobacteria sequences were abundant (≥8.6%) and prevalent across all WWTPs and treatment stages. There were also six classes for Bacteroidetes: WWTP 1 and 4 exhibited a similar pattern in sequence composition of Bacteroidetes, with classes Bacteroidia and Flavobacteriia detected in the influent, and in addition to these two classes, three other classes (Saprospirae and Sphingobacteriia) were also detected in the intermediate and effluent stages (Figure 2). Like WWTP 1 and 4, Bacteroidia and Flavobacteriia were detected in all stages of WWTP 3, and the same classes that were found in WWTP 1 and 4 were also found in WWTP 3, but in the intermediate stage of WWTP 3, only Bacteroidia, Flavobacteriia and Saprospirae sequences were obtained. WWTP 2 had similar Bacteroidetes in the influent and effluent; all aforementioned Bacteroidetes classes were found in the influent and effluent, and an additional class, Rhodothermi, was also found in the effluent (Figure 2 and Supplementary File B.2).

Cyanobacteria were not found in the influent of WWTP 1 and 3, but sequences were detected in the intermediate and effluent stages of these plants, and were detected in all stages of WWTP 2 and 4. Oscillatoriophycideae was dominant in the intermediate and effluent of WWTP 1 and 2 (11.2% and 14.3%, respectively) and was also detected in the effluent of WWTP 2 and 3. Other classes of Cyanobacteria included Synechococcophycideae in the intermediate of WWTP 1 and effluent of WWTP 1 and 4, Nostocophycideae in WWTP 2 effluent, and a class designated as 4C0d-2 by the Greengenes database was found in WWTP 3 intermediate and WWTP 4 intermediate and effluent. Among the Firmicutes, three classes were detected: Bacilli (Bacillales, Lactobacillales and Turicibacterales), Clostridia (Clostridiales) and Erysipelotrichi (Erysipelotrichales). Bacilli and Clostridia sequences were the most abundant classes of Firmicutes and were found in the influent of WWTPs 1, 3 and 4, ranging from 0.8-22.1%, and sequences were in low abundance (≤2.0%) or not detected in the intermediate and effluent stages (Figure 2 and Supplementary File B.2).

Five Actinobacteria classes were identified: Acidimicrobiia, Actinobacteria, Coriobacteriia, Nitriliruptoria and Thermoleophilia. The sequence composition of the class Actinobacteria was higher than other classes of the phylum Actinobacteria (which were all ≤2.0% in various treatment plants and stages), particularly in the intermediate and effluent of WWTPs 1, 3 and 4 (1.7-8.0%), and the influent and effluent of WWTP 2 (9.1% and 4.8%, respectively) (Figure 2). Acidimicrobiia and Thermoleophilia were detected in the intermediate and effluent of WWTP 1, influent and effluent of WWTP 2 and intermediate of WWTP 4. A low sequence composition of Actinobacteria and Coriobacteriia were found in the influent of WWTPs 1, 3 and 4. Nitriliruptoria was only found in the effluent of WWTP 2 (Supplementary File B.2).

3.3 Bacterial pathogen identification

Based on Greengenes taxonomic assignments, seven ZOTUs were assigned to the family Enterobacteriaceae (Gammaproteobacteria: Enterobacteriales). Four of these ZOTUs were not assigned further taxonomy by Greengenes, but the remaining ZOTUs were designated as *Citrobacter* sp., *Escherichia coli* and *Trabulsiella* sp. with high confidence (0.95-1). However, comparison of the Enterobacteriaceae sp. ZOTUs to the NCBI nr/nt database using BLAST revealed that all these ZOTUs were 100% similar to multiple Enterobacteriaceae sp. genera (Table 3). The phylogenetic tree constructed with Enterobacteriaceae sp. ZOTUs and Enterobacteriaceae sequences from the NCBI nr/nt database showed that different genera grouped closely with short branch lengths, and most bootstrap values were low (Figure 3; refer to Supplementary file B.3 for pairwise genetic distances, which range from 94.3-100%). This supports the BLAST results from Table 3 that suggest that the Enterobacteriaceae sp. ZOTUs can only be confidently assigned to the family level, and suggests that the V4 region of 16S cannot distinguish between many Enterobacteriaceae species and genera.

Other ZOTUs from the class Gammaproteobacteria that were assigned to pathogenic species, or to taxa that contain pathogens, based on Greengenes taxonomy included *Acinetobacter* (*Acinetobacter johnsonii*, *Acinetobacter lwoffii* and unassigned species), Aeromonadaceae (unassigned genera and *Tolumonas*), Coxiellaceae (genus unassigned), Legionellaceae (genus unassigned), *Enterococcus* spp. (Enterococcaceae), Pseudomonadaceae (*Pseudomonas alcaligenes, Pseudomonas fragi*, *Pseudomonas nitroreducens*, *Pseudomonas stutzeri*, *Pseudomonas veronii*, *Pseudomonas viridiflava* and unassigned species), Piscirickettsiaceae (genus unassigned) and Pseudoalteromonadaceae (genus unassigned). Greengenes taxonomy that conflicted with BLAST analysis was identified for *Tolumonas* (Aeromonadaceae; ZOTU 483), which was most similar to *Pseudaeromonas sharmana* (100%; GenBank® accession no. MF280154), *Aeromonas sharmana* (99.2%; JF496528) and *Tolumonas* sp. (98.8%; MG801837). *Acinetobacter* ZOTUs assigned to the species level (*Acinetobacter johnsonii* and *Acinetobacter lwoffii*) with high confidence naïve Bayes confidence scores (0.94-1) were also 100% similar to several other *Acinetobacter* species. Similarly, many *Pseudomonas* species had sequence similarities of 100%, therefore had incorrect species level taxonomy assigned with Greengenes. Greengenes taxonomy was more conservative for ZOTU 589 (Pseudoalteromonadaceae sp.) as BLAST results showed that this sequence could be assigned to the genus *Vibrio*, but like

Acinetobacter and *Pseudomonas*, many *Vibrio* species were also 100% similar at the V4 region (Supplementary File B.4).

The BLAST results agreed with Greengenes taxonomic assignments for most Betaproteobacteria (Alcaligenaceae spp., Neisseriaceae spp. and *Vitreoscilla* spp.), except for ZOTU 55 that was classed as *Microvirgula* sp., but was 100% similar to *Laribacter hongkongensis* sequences (NR025167). Five Campylobacteraceae (class Epsilonproteobacteria) ZOTUs that were assigned to the genus *Arcobacter* or *Arcobacter cryaerophilus* were also 100% similar to *Campylobacter* sequences and therefore could only be confidently assigned to the Campylobacteraceae family. *Corynebacterium* spp. and *Mycobacterium* spp. (phylum Actinobacteria) taxonomy agreed for both Greengenes and BLAST results, but there were discrepancies for Streptococcaceae spp. (phylum Firmicutes) that were assigned as *Streptococcus* spp., *Streptococcus luteciae* and *Streptococcus minor* by Greengenes, but could not be assigned to the species or genus level by BLAST in most cases. The Greengenes taxa *Candidatus* Rhabdochlamydia sp. (ZOTU 1597; phylum Chlamydiae), *Clostridium* spp., *Proteiniclasticum* sp. (phylum Firmicutes) or *Treponema* spp. (Spirochaetes) could also not be confidently assigned to the genus level based on BLAST results (Supplementary File B.4).

The sequence compositions for pathogenic and potentially pathogenic taxa that were given final taxonomic assignments based on Greengenes and NCBI nr/nt sequence database comparisons are summarised in Table 4. Briefly, sequence compositions for these taxa were generally higher in the influent for WWTP 1, 3 and 4 and lower in the intermediate and effluent, with the exception of *Acinetobacter* spp. in the intermediate stage of WWTP 3 (17.1%) and *Aeromonas* sp. in the effluent of WWTP 1 (8.6%). Potentially pathogenic sequence compositions were relatively low in the influent and effluent of WWTP 2, with the highest composition of 2.0% in the effluent for Alcaligenaceae sp. (Table 4).

3.4 Nitrifying, denitrifying and floc-forming bacteria

Other bacteria of interest in WWTPs, such as nitrifying, denitrifying and floc-forming bacteria, also had Greengenes taxonomy validated with BLAST results from the NCBI nr/nt database. Compared to pathogenic bacteria, the nitrifying, denitrifying and floc-forming bacterial ZOTUs had more taxonomic assignments that agreed with both databases. All were assigned to the appropriate family, but some ZOTUs had conflicting genera. For example, ZOTU 387 was assigned as *Dechloromonas* sp. by Greengenes, but was also 100% similar to *Azonexus hydrophilus* (LN650477), and ZOTU 766 Greengenes taxonomy was *Comamonas* sp., but this ZOTU was 100% similar to *Comamonas* spp. (MH174324) and *Delftia* spp. (MF156914). Results of taxonomy database comparisons for nitrifying, denitrifying and floc-forming bacteria are provided in Supplementary File B.5, and the sequence compositions of validated taxa are presented in Table 5. Nitrifying bacteria, *Nitrospira* spp. (Nitrospirales: Nitrospiraceae), were only detected in the intermediate and effluent of WWTP 4, with sequence compositions of 1.2% and 1.5%, respectively. In WWTP 1, denitrifying bacteria with the highest compositions were found in the influent for *Comamonas* sp. (Comamonadaceae; 6.9%) and *Thauera* spp. (Rhodocyclaceae; 3.4%). The comamonads *Hydrogenophaga* spp. and *Aquabacterium* sp. had highest compositions in the effluent (2.4%) and influent (1.3%) of WWTP 2, respectively, and the floc-forming bacteria *Flavobacterium* spp. were higher in the effluent (4.7%) than in the influent (2.8%) of WWTP 2. WWTP 3 had a greater diversity of denitrifying and floc-forming bacteria in the influent and intermediate stages than the effluent; the highest sequence compositions in the influent was 4.4% for *Comamonas* sp., 6.0% for *Thauera* spp. in the intermediate stage and 7.6% for *Flavobacterium* spp. in the effluent. The abundance of *Comamonas* sp. sequences was also high in the influent of WWTP 4, and the denitrifying bacteria *Uliginosibacterium* spp. were highest in the intermediate stage, and *Flavobacterium* spp. was highest in the effluent of WWTP 4 (Table 5). Pseudomonadaceae (*Pseudomonas*) are also denitrifying bacteria, and are summarised in Table 4.

4. Discussion

Evaluation of BLAST results from the NCBI nr/nt database of V4 16S sequences that were assigned taxonomy by the 16S Greengenes taxonomy database to pathogenic species (or to bacterial groups that contain pathogenic species) showed that the V4 region of 16S resolves poorly at the species level, and genus level identification was also impeded in many instances. Comparison of the ZOTU sequences to the NCBI nr/nt database revealed that only three ZOTUs were 100% identical to the following pathogenic species: *Laribacter hongkongensis*, which causes gastroenteritis and diarrhoea (Beilfuss et al., 2015); *Neisseria canis*, which usually infects cats and dogs, but can also infect humans (Safton et al., 1999); and *Arcobacter venerupis*. There are 15 species of *Arcobacter*, and three (*Arcobacter butzleri*, *Arcobacter cryaerophilus* and *Arcobacter skirrowii*) have been associated with gastrointestinal infections (Kayman et al., 2012). *Arcobacter venerupis* has previously only been isolated from shellfish (Levican et al., 2012), and these sequences were in low abundance (0.7%) and only found in the influent of WWTP 1. The sequences from *L. hongkongensis* and *N. canis* were found in the influent of WWTPs 1, 3 and 4, and were in low abundance (≤0.1%) or not detected in the intermediate and effluent stages of these plants. Genera known to contain pathogenic species that were validated by BLAST analyses of the ZOTUs against the NCBI nr/nt database included *Aeromonas* sp., *Acinetobacter* spp., *Arcobacter* spp., *Candidatus* Rhabdochlamydia sp., *Corynebacterium* sp., *Enterococcus* spp., *Legionella* sp., *Mycobacterium* spp., *Neisseria* sp., *Pseudomonas* spp., *Streptococcus* spp., *Turneriella* sp. and *Vibrio* sp. A previous 16S NGS study on WWTPs in Australia that also used the Illumina MiSeq platform identified 25 potentially pathogenic genera (Ahmed et al., 2017), while another study of municipal activated sludge plants across four countries (China, USA, Canada and Singapore) identified 16 pathogenic genera using pyrosequencing (Ye and Zhang, 2011). The abundance of pathogenic genera may vary among studies due to DNA extraction kits, different sequencing technologies, inherent amplification biases during PCR and the 16S hypervariable region(s) targeted (Haft and Tovchigrechko 2012).

Other pathogenic genera that can infect people via contaminated drinking water, *Campylobacter* spp. and *Leptospira* spp., were not identified in the present study, but we cannot exclude the possibility of their presence, as several Campylobacteraceae spp. and Leptospiraceae spp. ZOTUs could not be resolved to the genus level (Table 4 and Supplementary File B.4). Most of the potentially pathogenic genera identified had higher sequence compositions in the influent, and had low compositions or were not detected in the effluent (Table 4). However, *Aeromonas* sp. had relatively high sequence compositions in WWTP 1 intermediate (2.2%) and effluent (8.6%) samples, and a similar trend was observed for *Aeromonas* sp. in WWTP 3, with compositions of 4.7% in the influent, 1.1% in intermediate samples and 4.8% in the effluent. *Acinetobacter* spp. also had a high sequence composition in the intermediate samples of WWTP 3 (17.1%), but was not detected in the effluent. Other studies have also found that *Acinetobacter* spp. sequence compositions were not significantly lower in treated wastewater samples compared to the influent (Ahmed et al., 2017). *Mycobacterium* spp. and *Pseudomonas* spp., which had lower compositions or were not detected in the influent, had higher compositions in the intermediate and effluent (Table 4). The absence or lower abundance of bacteria associated with human waste in the influent compared to the intermediate and effluent may be partly explained by the lack of sample replicates, as only 100 mL grab samples were collected per season at each site and treatment stage. However, the number of 16S sequences obtained by NGS does not represent the number of bacterial organisms present. A number of factors affect sequence composition, including PCR amplification bias (Hong et al., 2009), sequencing depth and copy number variation in the 16S gene (Kembel et al., 2012).

Many enteric bacteria (Enterobacteriaceae) can be transmitted to humans by faecal-oral transmission and can cause gastrointestinal illnesses with symptoms of abdominal pain, diarrhoea, fever, nausea and vomiting. Human enteric pathogens include *Citrobacter freundii*, *Escherichia coli*, *Klebsiella aerogene*, *Salmonella bongori*, *Salmonella enterica* and *Shigella* spp. (Cabral, 2010). Other pathogens from the family Enterobacteriaceae that can cause gastrointestinal illnesses are *Yersinia enterocolitica*, which is a food-borne pathogen associated with pork products (Bhaduri et al., 2005), and *Raoultella ornithinolytic*a (formerly *Klebsiella ornithinolytica*), which has been found in aquatic environments and hospitals, with one report of its isolation from human digestive organs (Seng et al., 2016). Enterobacteriaceae pathogens that cause urinary tract infections and other illnesses in humans include *Proteus mirabilis*, *Proteus penneri*, *Proteus vulgaris* and *Serratia marcescens* (Guentzel et al., 1996). Unfortunately, the V4 region of 16S lacked sufficient variability to distinguish between Enterobacteriaceae genera (Table 3 and Figure 3). The same issue was likely encountered in the V4 16S NGS study by Zhang et al. (2012), that reported the detection of sequences from the order Enterobacteriales. Similarly, a 16S NGS study on WWTPs that targeted the V6 region could not resolve one OTU that was 100% similar to several Enterobacteriaceae genera (primarily *Klebsiella* and *Shigella*) (McLellan et al., 2010). Other 16S wastewater studies that have targeted 16S regions that span two hypervariable regions appear to have been able to resolve Enterobacteriaceae genera. For example, Ahmed et al. (2017) sequenced regions V5-6 (300 bp), and reported the detection of *Escherichia*/*Shigella* (unclear if these could be differentiated), *Salmonella* and *Yersinia*, but species level assignments were not made. Lu et al. (2015) targeted the V3-4 region (460 bp) and reported the presence of *Klebsiella pneumoniae* and *Serratia* spp., but performed shotgun sequencing to identify pathogens to the species level, which included *E. coli*, *S. enterica*, *Shigella sonnei* and *Yersinia pestis*. According to a study by Chakravorty et al. (2007), V3 is a more suitable region for the differentiation of Enterobacteriaceae genera, and these authors recommended targeting V2, V3 and V6 to identify the bacterial genera assessed in their study, including *Acinetobacter*, *Bacillus*, *Clostridium*, *Corynebacterium*, *Chlamydia*, *Enterococcus*, *Listeria*, *Mycobacterium*, *Neisseria*, *Pseudomonas*, *Streptococcus*, *Staphylococcus*, *Treponema* and *Vibrio*. Using these three regions means that most of the 110 species examined in their study could be identified to the species level (Chakravorty et al., 2007). Using multiple regions does have some challenges, however. For example, the V2 region of *E. coli* starts at nucelotide (nt) position 137 and V6 ends at nt position 1,043 (Brosius et al., 1978), therefore V2-6 spans 906 bp of 16S. This amplicon is too long for current amplicon NGS sequencers; the maximum length is 600 bp on the Illumina MiSeq with v3 chemistry (http://www.illumina.com/). Regions V2-3 and V6 could be targeted separately, or full length 16S could be sequenced on long-read sequencing platforms such as PacBio for improved taxonomic resolution of a greater variety of taxa (Ibal et al. 2019). It is important for Enterobacteriaceae species such as *Escherichia coli* for serotypes to be differentiated at the strain level, as some strains are harmless gut bacteria whereas others are pathogenic, e.g. enterohemorrhagic *Escherichia coli* O157:H7. While some studies state that 16S sequencing is unsuitable for differentiating *E. coli* and *Shigella* spp. serotypes as the sequence similarity is high (97.9-99.9%) (Devanga Ragupathi et al. 2017), Ibal et al. (2019) were able to classify *E. coli* strains based on full length 16S sequences (Ibal et al. 2019). Other housekeeping genes that are conserved among bacteria, such as *gyrB*, *rpoB* and *mdh* have greater genetic variability for distinguishing *E. coli* and *Shigella* spp. strains than 16S (Devanga Ragupathi et al. 2017; Fukushima et al. 2002). These genes could also be targeted using amplicon NGS approaches for improved taxonomic resolution of bacterial strains, however the use of universal primers is more limited than 16S. Alternatively, shotgun sequencing could be performed, which can provide greater taxonomic and functional information (e.g. pathogenicity islands and toxin-producing genes) than amplicon NGS of several target genes (Sanapareddy et al., 2009; Lu et al., 2015). Shotgun sequencing of metagenomes has been considerably more expensive than amplicon NGS (Goodwin et al., 2016), however costs are reducing, particularly with new approaches such as "shallow shotgun sequencing", which can produce more accurate species level taxonomic and functional profiles of the human microbiome than 16S sequencing (Hillmann et al. 2018).

A large portion of the V4 16S sequences (68%) collected in this current study were not assigned to the genus level with the Greengenes database. Other 16S NGS studies on wastewater have used RDP Classifier (Zhang et al., 2012; Ahmed et al., 2017) and SILVA (McLellan et al., 2010; Lu et al., 2015) databases for taxonomic assignment. According to a recent study that compared the major taxonomy databases (Greengenes, RDP classifier, SILVA, NCBI and OTT), there were few conflicts when SILVA, RDP and Greengenes were mapped into NCBI and OTT (Balvočiūtė et al., 2017). However, we found many genus level conflicts, when potentially pathogenic and denitrifying bacteria were compared to the NCBI nr/nt database (Supplementary Files B.4 and B.5). Furthermore, we found erroneous taxonomy in the Greengenes database that causes 16S sequences deriving from chloroplasts in algae and plants to be classified to the bacterial phylum Cyanobacteria and the class “Chloroplast”, which is not a valid taxon. For 44 ZOTUs in our dataset that were classified to the class “Chloroplast”, the orders provided by the Greengenes database were Chlorophyta (phylum of green algae), Euglenozoa (phylum of flagellate excavates) and Stramenopiles (infrakingdom of algae and oomycetes). While chloroplast sequences in the Greengenes database can be useful to identify such sequences in an NGS dataset, researchers that aim to only analyse bacterial 16S sequences at higher levels of classification (kingdom and phylum) may be unaware that the chloroplast sequences are classified in the database at the class level. Classifying the chloroplast sequences as "Chloroplast" at the kingdom level, rather than as "Bacteria" may help researchers to detect these sequences at an earlier stage of the data analysis. We have provided a modified version of the Greengenes 99 OTU taxonomy file for all chloroplast sequences with the kingdom "Bacteria" renamed as "Chloroplast" in Supplementary File B.6. A custom curated sequence database for waterborne pathogens, with quality-checked sequences and taxonomy validated by phylogenetic analyses, may also reduce the errors in bacterial taxonomic assignment experienced with other 16S sequence databases.

Overall, of the 2 archaeal and 28 bacterial phyla detected, Proteobacteria, Bacteroidetes, Cyanobacteria, Firmicutes and Actinobacteria had high sequence compositions (>10%) in WWTP samples (Figure 2). The two most dominant phyla in all treatment stages for WWTPs 1-4 were Proteobacteria and Bacteroidetes, which has also been observed by a previous 16S NGS study that examined bacteria in activated sludge WWTPs across Australia, including Perth (Ahmed et al., 2017). The study by McLellan et al. (2010) that compared V6 16S NGS bacterial profiles in WWTP influent, surface water and human faecal samples, also found that the most dominant bacterial phylum in the WWTPs was Proteobacteria (overall 59% sequence composition), and like our study, Gammaproteobacteria and Betaproteobacteria were the most abundant classes. McLellan et al. (2010) also found that Actinobacteria, Bacteroidetes and Firmicutes were dominant taxa in the WWTP influent, and sewage samples had high compositions of Firmicutes, particularly Clostridia (the human faecal samples were comprised mostly (98%) of Clostridia) and Bacilli (McLellan et al., 2010). In the present study, Firmicutes had the highest compositions in the influent of WWTPs 1, 3 and 4, ranging from 16.8-20.5%; Bacilli ranged from 5.6-11.9% and Clostridia ranged from 7.9-12.4% (Figure 2). *Bacteroides* is another faecal indicator bacterium (Kreader, 1995), and the sequence compositions for *Bacteroides* spp. in the present study ranged from 0.4% in the influent of WWTP 1 and 2 to 2.4% in the influent of WWTP 3, and sequence compositions were low (≤0.8%) or undetectable in the intermediate and effluent (Supplementary file B.2). *Faecalibacterium* is also associated with faeces (Zheng et al., 2009), and was detected in the influent of WWTP 1 (1.0%), 3 (1.7%) and 4 (1.3%), and had low sequence compositions (≤0.1%) or were not detected in the intermediate and effluent stages.

Nitrification is a fundamental process in the biological removal of nitrogen in WWTPs, and this two-step process is carried out by ammonia-oxidising bacteria (AOB) that convert ammonia to nitrite, then nitrite-oxidising bacteria (NOB) convert nitrite to nitrate (Bellucci and Curtis, 2011). *Nitrosomonas* and *Nitrospira* are two important genera of AOB in WWTPs, while *Nitrobacter* is a major NOB (Siripong et al., 2007). In the present study, *Nitrosomonas* and *Nitrobacter* were not detected, and *Nitrospira* spp. were only detected in the intermediate and effluent of WWTP 4 (sequence compositions 1.2% and 1.5%, respectively) (Table 5). Rhodocyclales are a widespread and abundant order of bacteria in WWTPs responsible for anaerobic nitrogen removal by denitrification (Yang et al., 2011). In the present study, 12 Rhodocyclales genera were identified: *Azoarcus* spp., *Azonexus* spp., *Azospira* spp., *Dechloromonas* spp., *Methyloversatilis* sp., *Propionivibrio* spp., *Rhodocyclaceae* spp., *Sterolibacterium* spp., *Sulfuritalea* sp., *Thauera* spp., *Uliginosibacterium* spp. and *Zoogloea* spp. (Table 5). In WWTP 1, Rhodocyclales were highest in the influent (*Thauera* spp. had the highest composition; 3.4%) and rare (≤0.1%) or not detected in the intermediate and effluent samples. Rhodocyclales compositions were low in the influent (0.8%) and effluent (1.0%) of WWTP 2. WWTP 3 had higher Rhodocyclales compositions in the intermediate (13.2%; most abundant was *Thauera* spp. at 6.0%)

compared to the influent (4.3%), and no Rhodocyclales sequences were detected in WWTP 3 effluent. In addition to denitrification, certain *Thauera* and *Dechloromonas* strains can degrade oil derivatives such as toluene (Shinoda et al., 2004; Chakraborty et al., 2005) and therefore may be important in reducing the ecological burden of these aromatic compounds, but we were unable to identify the species and strains of these genera based on V4 16S amplicons. Unlike the WSPs, the oxidation ditch plant WWTP 4 had high Rhodocyclales compositions in both the intermediate and effluent (7.7% and 7.1%, respectively). Members of the family Comamonadaceae are also denitrifiers and are responsible for aromatic degrading processes (Xu et al., 2018). The nine Comamonadaceae genera identified were *Aquabacterium* sp., *Brachymonas* (*Brachymonas denitrificans*), *Comamonas* sp., *Delftia* sp., *Flavobacterium* spp., *Hydrogenophaga* spp., *Polaromonas* spp., *Rhodoferax* spp. and *Rubrivivax* spp. Comamonadaceae compositions in WWTP 1 were similar to those observed for Rhodocyclales in this treatment plant, with the highest composition observed in the influent (7.4%; *Comamonas* sp. had the highest composition of 6.9%) and compositions were low in the intermediate and effluent (1.6% and 1.7%, respectively). Comamonadaceae compositions were much higher than Rhodocyclales in WWTP 2, which had 4.7% in the influent and 7.7% in the effluent, and *Flavobacterium* spp. had the greatest sequence compositions in both influent (2.8%) and effluent (4.7%). For WWTP 3, the Comamonadaceae compositions were similar to Rhodocyclales in the influent and intermediate, but unlike Rhodocyclales, were detected (mostly *Flavobacterium* spp. 7.6%) in the effluent. For WWTP 4, the Comamonadaceae were mostly comprised of *Comamonas* sp. in the influent (4.4%) and *Flavobacterium* spp. (6.5%) in the effluent, and the composition of Comamonadaceae was low in the intermediate (1.2%). Comamonadaceae, Rhodocyclaceae, Flavobacteriaceae and Pseudomonadaceae also play important roles in flocculation in activated sludge plants, and Comamonadaceae and Flavobacteriaceae are important for bulking and foaming (Shchegolkova et al., 2016).

5. Conclusions

In the present study, a total of 36 pathogenic or potentially pathogenic species were detected, but most could not be identified to species level. Of these, sequences belonging to 14 medically important genera that could possibly be from pathogens were identified primarily in the influent of WWTPs 1-4. In almost all cases, these bacteria were present in lower abundance in the effluent with the exception of *Aeromonas* sp. in the effluent of WWTP 1 (8.6%). The use of V4 16S NGS for bacterial pathogen identification has significant limitations for species level identification including the inability to differentiate Enterobacteriaceae genera that contain many important enteric pathogens of humans. Amplicon NGS is a useful tool for broad taxonomic surveys of bacteria, while tools such as quantitative PCR and droplet digital PCR could be used in follow-up studies to identify bacteria that could not be differentiated at the species or strain level. This would also allow quantification of pathogens before and after the wastewater treatment process. Future studies that aim for improved taxonomic resolution of bacterial pathogens in wastewater should consider sequencing full length 16S and more variable housekeeping genes such as *gyrB*, *rpoB* or *mdh* for differentiation of *E. coli* and *Shigella* strains. Shallow shotgun sequencing can also be used for pathogen identification and for gaining functional information that is important for public health.

Nitrifying, denitrifying and floc-forming bacteria could mostly be identified to the genus level. Only the activated sludge oxidation ditch plant showed the presence of an AOB, *Nitrospira* spp., for bacterial nitrification. However, both the lower technology WSPs and the activated sludge oxidation ditch plant showed the presence of Rhodocyclales, Comamonadaceae, Flavobacteriaceae and Pseudomonadaceae bacteria, which are responsible for anaerobic nitrogen removal by denitrification (i.e. conversion of nitrate to nitrogen gas). These bacteria are also important for WWTP performance since they assist floc formation. Our current work is examining the presence, diversity and relative abundances of bacterial communities responsible for the nitrification and denitrification cycle in WSPs (e.g. *Nitrobacter*, *Nitrosomonas*, *Nitrospira*, *Nitrosococcus* and *Nitrosomonas*) using functional genes that encode key enzymes (amoA, njfH, nirK, nosZ, norB, nxrB, narG, napA and nrfA). This will help us to better understand the correlations between the concentrations of selected nitrogenous species present in wastewater and their contribution to the nitrogen cycle in WSPs.

Other limitations include the misidentification of 16S sequences from chloroplasts as Cyanobacteria by the Greengenes database. Due to the discrepancies between taxonomic assignments with Greengenes and the NCBI nr/nt database, we recommend that future studies use the Greengenes database for 16S NGS taxonomic assignment with caution and compare OTU or ZOTU sequences with the NCBI nr/nt database to validate taxonomic assignments.

Acknowledgements

This work was supported by the Australian Research Council (ARC LP130100602), in collaboration with the Water Corporation of Western Australia and Water Research Australia. We thank Dr Elvina Lee, Annachiara Codello and Maninder Khurana for their assistance with NGS preparation. We thank Arron Lethorn and operational staff at the Water Corporation of Western Australia for project management and assisting with site visits and sample collection.

References

Ahmed W., Staley C., Sidhu J., Sadowsky M., Toze S., 2017. Amplicon-based profiling of bacteria in raw and secondary treated wastewater from treatment plants across Australia. Appl Microbiol Biotechnol. 101(3), 1253-1266.

Anonymous, 2009. Ponds for stabilising organic matter. www.water.wa.gov.au/__data/assets/pdf_file/0005/4100/84601.pdf

Apprill, A., McNally, S., Parsons, R., Weber, L., 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat. Microb. Ecol. 75, 129-137.

Baars, J.K., 1962. The use of oxidation ditches for treatment of sewage for small communities. Bull. World Health Organ. 26, 465-474.

Balvočiūtė, M., Huson, D.H., 2017. SILVA, RDP, Greengenes, NCBI and OTT—how do these taxonomies compare? BMC Genomics. 18, 114.

Beilfuss, H.A., Quig, D., Block, M.A., Schreckenberger, P.C., 2015. Definitive identification of *Laribacter hongkongensis* acquired in the United States. J. Clin. Microbiol. 53, 2385-2388.

Bellucci, M., Curtis, T.P., 2011. Ammonia-oxidizing bacteria in wastewater. Methods Enzymol. 496, 269-286.

Berlec, A., 2012. Novel techniques and findings in the study of plant microbiota: Search for plant probiotics. Plant Sci. 193-194,96-102.

Bhaduri, S., Wesley, I.V., Bush, E.J., 2005. Prevalence of pathogenic *Yersinia enterocolitica* strains in pigs in the United States. Appl Environ Microbiol. 71, 7117-7121.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., Caporaso, J.G., 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 6, 90.

Brosius, J., Palmer, M. L., Kennedy, P. J., Noller, H. F., 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci U S A. 75, 4801-4805.

Cabral, J.P., 2010. Water microbiology. Bacterial pathogens and water. Int J Environ Res Public Health. 7, 3657-3703.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight. R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 7, 335-336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Noah Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. USA. 108, 4516-4522.

Chakraborty, R. O'Connor, S.M. Chan, E. Coates, J.D., 2005. Anaerobic degradation of benzene, toluene, ethylbenzene, and xylene compounds by *Dechloromonas* strain RCB Appl. Environ. Microbiol. 71, 8649-8655.

Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods. 69, 330-339.

Connor, R., Renata, A., Ortigara, C., Koncagül, E., Uhlenbrook, S., Lamizana-Diallo, B.M., Zadeh, S.M., Qadir, M., Kjellén, M., Sjödin, J., Hendry, S., 2017. The United Nations world water development report. Wastewater: the untapped resource. Paris, UNESCO. https://reliefweb.int/sites/reliefweb.int/files/resources/247153e.pdf.

Cydzik-Kwiatkowska, A., Zielińska, M., 2016. Bacterial communities in full-scale wastewater treatment systems. World J Microbiol Biotechnol. 32, 66.

Devanga Ragupathi, N.K., Muthuirulandi Sethuvel, D.P., Inbanathan, F.Y., Veeraraghavan, B., 2017. Accurate differentiation of *Escherichia coli* and *Shigella serogroups*: challenges and strategies. New Microbes New Infect. 21, 58-62.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26, 2460-2461.

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv. 081257.

Eland, L.E., Davenport, R.J., Santos, A.B., Mota Filho, C.R., 2018. Molecular evaluation of microalgal communities in full-scale waste stabilisation ponds. Environ. Technol. In press. doi: 10.1080/09593330.2018.1435730.

Fukushima, M., Kakinuma, K., Kawaguchi, R., 2002. Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the gyrB gene sequence. J Clin Microbiol. 40, 2779-2785.

Garrido-Cardenas, J.A., Polo-López, M.I., Oller-Alberola, I., 2017. Advanced microbial analysis for wastewater quality monitoring: metagenomics trend. Appl. Microbiol. Biotechnol. 101, 7445-7458.

Goodwin, S., McPherson, J.D., McCombie, W.R., 2106. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 17, 333-351.

Guentzel, M.N., 1996. *Escherichia*, *Klebsiella*, *Enterobacter*, *Serratia*, *Citrobacter*, and *Proteus*. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307-321.

Haft, D.H., Tovchigrechko, A., 2012. High-speed microbial community profiling. Nat Methods. 9, 793-794.

Hillmann, B., Al-Ghalith, G. A., Shields-Cutler, R.R., Zhu, Q., Gohl, D.M., Beckman, K.B., Knight, R., Knights, D., 2018. Evaluating the information content of shallow shotgun metagenomics. mSystems. 3, e00069-00018.

Ho, L.T., Van Echelpoel, W., Goethals, P.L.M., 2017. Design of waste stabilization pond systems: a review. Water Res. 123, 236-248.

Ibal, J.C., Pham, H.Q., Park, C.E., Shin, J-H., 2019. Information about variations in multiple copies of bacterial 16S rRNA genes may aid in species identification. PLoS One. 14, e0212090.

Inaba, T., Hori, T., Aizawa, H., Ogata, A., Habe, H., 2017. Architecture, component, and microbiome of biofilm involved in the fouling of membrane bioreactors. NPJ Biofilms Microbiomes. 3, 5.

Katoh, K., Misawa, K., Kuma, K.-i., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066.

Kayman, T., Abay, S., Hizlisoy, H., Atabay, H.İ., Diker, K.S., Aydin, F., 2012. Emerging pathogen *Arcobacter* spp. in acute gastroenteritis: molecular identification, antibiotic susceptibilities and genotyping of the isolated arcobacters. J Med Microbiol. 61, 1439-1444.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28, 1647-1649.

Kembel, S.W., Wu, M., Eisen, J.A. and Green, J.L., 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS computational biology, 8(10), p.e1002743.

Kim, Y., Koh, I., Rho, M., 2015. Deciphering the human microbiome using next-generation sequencing data and bioinformatics approaches. Methods. 79–80, 52-59.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111-120.

Levican, A., Collado, L., Aguilar, C., Yustes, C., Diéguez, A.L., Romalde, J.L. Figueras, M.J., 2012. *Arcobacter bivalviorum* sp. nov. and *Arcobacter venerupis* sp. nov., new species isolated from shellfish. Syst Appl Microbiol. 35, 133-138.

Lu, X., Zhang, X.-X., Wang, Z., Huang, K., Wang, Y., Liang, W., Tan, Y., Liu, B., Tang, J., 2015. Bacterial pathogens and community composition in advanced sewage treatment systems revealed by metagenomics analysis based on high-throughput sequencing. PLoS One. 10, e0125549.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 6, 610.

McLellan, S., Huse, S., Mueller Spitz, S., Andreishcheva, E., Sogin, M., 2010. Diversity and population structure of sewage derived microorganisms in wastewater treatment plant influent. Environ. Microbiol. 12, 378-392.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R. B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H., 2018. Vegan: Community ecology package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan.

Parada, A.E., Needham, D.M., Fuhrman, J.A., 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol. 18, 1403-1414.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rajasulochana, R., Preethy, V., 2016. Comparison on efficiency of various techniques in treatment of waste and sewage water – a comprehensive review. Resource-Efficient Technologies. 2, 175-184.

Rodriguez-Manzano, J., Alonso, J.L., Ferrus, M.A., Moreno, Y., Amoros, I., Calgua, B., Hundesa, A., Guerrero-Latorre, L., Carratala, A., Rusinol, M., Girones, R., 2012. Standard and new faecal indicators and pathogens in sewage treatment plants, microbiological parameters for improving the control of reclaimed water. Water Sci. Technol. 66, 2517-2523.

Safton, S., Cooper, G., Harrison, M., Wright, L. and Walsh, P., 1999. *Neisseria canis* infection: a case report. Commun Dis Intell. 23, 221.

Sanapareddy, N., Hamp, T.J., Gonzalez, L.C., Hilger, H.A., Fodor, A.A., Clinton, S.M., 2009. Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. Appl Environ Microbiol. 75, 1688-1696.

Seng, P., Boushab, B.M., Romain, F., Gouriet, F., Bruder, N., Martin, C., Paganelli, F., Bernit, E., Le Treut, Y.P., Thomas, P., Papazian, L., 2016. Emerging role of *Raoultella ornithinolytica* in human infections: a series of cases and review of the literature. Int J Infect Dis. 45, 65-71.

Sharafi, K., Davil, M.F., Heidari, M., Almasi, A., Taheri, H. 2012. Comparison of conventional activated sludge system and stabilization pond in removal of chemical and biological parameters. Int. J. Environ. Health Eng. 1, 1-5.

Shchegolkova, N.M., Krasnov, G.S., Belova, A.A., Dmitriev, A.A., Kharitonov, S.L., Klimina, K.M., Melnikova, N.V., Kudryavtseva, A.V., 2016. Microbial community structure of activated sludge in treatment plants with different wastewater compositions Front Microbiol. 7, 90.

Shinoda, Y., Sakai, Y., Uenishi, H., Uchihashi, Y., Hiraishi, A., Yukawa, H., Yurimoto, H., Kato, N., 2004. Aerobic and anaerobic toluene degradation by a newly isolated denitrifying bacterium, *Thauera* sp. strain DNT-1 Appl. Environ. Microbiol. 70, 1385-1392.

Siripong, S. and Rittmann, B.E., 2007. Diversity study of nitrifying bacteria in full-scale municipal wastewater treatment plants. Water Res. 41, 1110-1120.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312-1313.

Verbyla, M., von Sperling, M., Maiga, Y., 2017. Waste Stabilization Ponds. In: J.B. Rose and B. Jiménez- Cisneros, (eds) Global Water Pathogens Project. http://www.waterpathogens.org (C. Haas, J.R. Mihelcic and M.E. Verbyla) (eds) Part 4 Management of risk from Excreta and Wastewater) www.waterpathogens.org/book/waste-stabilization-ponds Michigan State University, E. Lansing, MI, UNESCO.

Von Sperling, M., 2007. Waste stabilisation ponds. IWA Publishing. www.iwapublishing.com/sites/default/files/ebooks/9781780402109.pdf.

Xu, S., Yao, J., Ainiwaer, M., Hong, Y., Zhang, Y., 2018. Analysis of bacterial community structure of activated sludge from wastewater treatment plants in winter. Biomed Res Int. 2018, 8278970.

Yang, C., Zhang, W, Liu, R, Li, Q, Li, B, Wang, S, Song, C, Qiao, C, Mulchandani, A., 2011. Phylogenetic diversity and metabolic potential of activated sludge microbial communities in full-scale wastewater treatment plants. Environ Sci Technol. 45, 7408-7415.

Ye, L., Zhang, T., 2011. Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. Environ Sci Technol 45, 7173-7179.

Zhang, T., Shao, M.-F., Ye, L., 2012. 454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. ISME J. 6, 1137-1147.

Figure Legends

Figure 1. WWTP localities and different treatment stages sampled.

Figure 2. 16S NGS sequence percent composition plot of phyla (P) and classes (C) detected in different treatment stages of wastewater sampled from WWTPs 1-4. Treatment stages include influent (I), intermediate (INT) and effluent (E). Phyla with ≤10% overall sequence composition are grouped as "other".

Figure 3. Maximum likelihood tree of a 247 bp alignment (gaps excluded) of genomic 16S Enterobacteriaceae sequences trimmed to the V4 region. The seven Enterobacteriaceae ZOTU sequences derived from this study are in bold typeface. Values at nodes indicate Bootstrap values from 1,000 replicates. Outgroup of tree *Vibrio cholerae* (2614873) not shown.

Appendices

Appendix A. Supplementary Figures

Figure A.1. Alpha rarefaction plot of 16S sequencing depth and ZOTUs detected in WWTP samples prior to low read abundance (<0.05%) filtering.

Figure A.2. Alpha rarefaction plots of 16S sequencing depth and ZOTUs detected prior to low read abundance (<0.05%) filtering for WWTPs 1-4 and treatment stages.

Figure A.3. Alpha rarefaction plot of 16S sequencing depth and ZOTUs detected in WWTP samples after low read abundance (<0.05%) filtering.

Figure A.4. Alpha rarefaction plots of 16S sequencing depth and ZOTUs detected after low read abundance (<0.05%) filtering for WWTPs 1-4 and treatment stages.

Appendix B. Supplementary Data

Supplementary File B.1. List of 3,598 16S V4 region ZOTU sequences generated by this study.

Supplementary File B.2. Sequence totals and compositions.

Supplementary File B.3. Pairwise genetic distance matrix of the 247 bp alignment (gaps excluded) of genomic 16S Enterobacteriaceae sequences trimmed to the V4 region that was used to construct the phylogenetic tree in Figure 3.

Supplementary File B.4. Comparison of Greengenes and NCBI nr/nt database taxa to ZOTUs potentially from pathogenic bacteria.

Supplementary File B.5. Comparison of Greengenes and NCBI nr/nt database taxa to ZOTUs from nitrifying, denitrifying and floc-forming bacteria.

Supplementary File B.6. Chloroplast sequences in the Greengenes 99 OTU taxonomy file renamed with the kingdom “Chloroplast”.

Table 1. Rural wastewater treatment plant samples analysed in the present study.

WWTP	**Treatment technology**	**Location**	**Climate**	**Sample ID**	**Wastewater treatment stage**	**Sample collection date; season**
WWTP 1	Stabilisation pond: Combined anaerobic and aerobic pond system, followed by two maturation ponds	Northwest Western Australia	Tropical climate. Wet and dry seasons.	WWTP 1-1	Influent	19-Feb-2015; Wet
				WWTP 1-2	Effluent (pre-chlorination)	
				WWTP 1-3	Effluent (post-chlorination)	
				WWTP 1-4	Influent	7-Sep-2015; Dry
				WWTP 1-5	Intermediate (post maturation pond 1)	
				WWTP 1-6	Intermediate (post maturation pond 2)	
				WWTP 1-7	Effluent (pre-chlorination)	
				WWTP 1-8	Effluent (post-chlorination)	
WWTP 2	Stabilisation	Wheatbelt,	Hot dry	WWTP 2-1	Influent	12-Feb-2015;

	pond: One facultative pond	Western Australia	summers and mild winters. Four distinct seasons.			Summer
				WWTP 2-2	Effluent (final pond)	
				WWTP 2-3	Effluent (storage basin)	
				WWTP 2-4	Influent	13-Jul-2015; Winter
				WWTP 2-5	Effluent (final pond)	
				WWTP 2-6	Effluent (storage basin)	
WWTP 3	Stabilisation pond: Two primary facultative ponds, and one secondary pond	Southwest Western Australia	Temperate climate. Four distinct seasons.	WWTP 3-1	Influent	23-Feb-2015; Summer
				WWTP 3-2	Intermediate (post-pond)	
				WWTP 3-3	Effluent	
				WWTP 3-4	Influent	14-July-2015; Winter
				WWTP 3-5	Intermediate (post-pond)	
				WWTP 3-6	Effluent	
WWTP 4	Activated sludge: Oxidation ditches	Southwest Western Australia	Temperate climate. Four distinct seasons.	WWTP 4-1	Influent	23-Feb-2015; Summer
				WWTP 4-2	Intermediate (oxidation	

	followed by sedimentation tanks				ditch)	
				WWTP 4-3	Effluent	
				WWTP 4-4	Influent	14-July-2015; Winter
				WWTP 4-5	Intermediate (oxidation ditch)	
				WWTP 4-6	Effluent	

Table 2. V4 16S NGS sequence statistics.

Statistics	**Raw (unprocessed)**	**Pre-processed**[a]	**Processed 16S sequences**[b]			
	Grand total (*n* = 34)		**Samples (*n* = 26)**	**Extraction controls (*n* = 6)**	**NTCs (*n* = 2)**	**Grand total (*n* = 34)**
Average	27,965	23,805	26,746	8	8	20,454
Standard deviation	27,254	24,239	20,608	7	2	21,314
Min	2,646	2	4,681	2	6	2
Max	182,113	95,135	85,305	21	9	85,305
Total	1,426,191	809,368	695,400	48	19	695,463

[a] Merged, quality filtered sequences with singletons and chimeras removed

[b] Merged, quality filtered sequences with singletons, chimeras, unassigned sequences and low abundance sequences (<0.05%) removed

Table 3. Enterobacteriaceae (Gammaproteobacteria: Enterobacteriales) ZOTUs Greengenes assigned taxonomy cross-checked against the NCBI nr/nt database.

			Greengenes results		**NCBI nr/nt results**			**Correct Greengenes taxonomy?**		
ZOTU no.	**Accession no.**	**Final taxonomy**	**Assigned taxonomy**	**Confidence scores**[a]	**GenBank® accession no.**	**Species**	**Percent identity**	**Family**	**Genus**	**Species**
28	MH892615	Enterobacteriaceae sp.	Enterobacteriaceae sp.	0.95	MH384426	*Enterobacter xiangfangensis*	100	✓	*	*
					MH190220	*Erwinia aphidicola*	100	✓	*	*
					MH411220	*Klebsiella pneumoniae*	100	✓	*	*
54	MH892622	Enterobacteriaceae sp.	*Escherichia coli*	0.96	MH396737	*Escherichia coli*	100	✓	✗	✗
					MH352164	*Salmonella enterica* subsp. *enterica*	100	✓	✗	✗
					MH371327	*Shigella flexneri*	100	✓	✗	✗
170	MH892637	Enterobacteriaceae sp.	*Citrobacter* sp.	0.95	NR156052	*Citrobacter europaeus*	100	✓	✗	*
					MH371322	*Citrobacter freundii*	100	✓	✗	*
					MH352205	*Salmonella enterica* subsp. *enterica*	100	✓	✗	*

183	MH892638	Enterobacteriaceae sp.	Enterobacteriaceae sp.	0.94	CP020089	*Enterobacter cloacae*	100	✓	*	*
					MF360016	*Klebsiella michiganensis*	100	✓	*	*
					MH196342	*Klebsiella oxytoca*	100	✓	*	*
417	MH892656	Enterobacteriaceae sp.	Enterobacteriaceae sp.	0.92	MG890203	*Leclercia adecarboxylata*	100	✓	*	*
					MG022656	*Raoultella electrica*	100	✓	*	*
					MG516115	*Raoultella ornithinolytica*	100	✓	*	*
546	MH892663	Enterobacteriaceae sp.	*Trabulsiella* sp.	1	MH085457	*Citrobacter amalonaticus*	100	✓	✗	*
					MF186607	*Citrobacter farmeri*	100	✓	✗	*
					MH169203	*Kosakonia oryzendophytica*	100	✓	✗	*
619	MH892672	Enterobacteriaceae sp.	Enterobacteriaceae sp.	1	MH141470	*Cronobacter sakazakii*	100	✓	*	*
					MH169205	*Kluyvera georgiana*	100	✓	*	*
					MG890202	*Pseudocitrobacter faecalis*	100	✓	*	*

*Taxon was unassigned, which was the correct choice based on BLAST results.

[a] Confidence scores are probabilities generated by the naïve Bayes algorithm implemented by QIIME 2 feature classifier (https://docs.qiime2.org/2018.6/tutorials/feature-classifier/).

Table 4. Sequence composition (%) of pathogens and possible pathogens in WWTPs 1-4 influent (I), intermediate (INT) and effluent (E) with taxonomy confirmed with Greengenes and NCBI nr/nt sequence databases.

						WWTP 1			**WWTP 2**		**WWTP 3**			**WWTP 4**		
Class	**Order**	**Family**	**Taxonomic assignment**[a]	**ZOTU no.**	**Accession no.**	**I**	**INT**	**E**	**I**	**E**	**I**	**INT**	**E**	**I**	**INT**	**E**
Actinobacteria																
ctinobacteria	ctinomycetales	orynebacteriaceae	*orynebacterium* sp.	603	H892704	0.1	-	-	-	-	-	-	-	-	-	-
		ycobacteriaceae	*ycobacterium* spp.	6; 332; 588; 1469; 1651; 1756; 1801	H892617; MH892651; MH892668; MH892692; MH892696; MH8	-	2.5	2.2	0.1	-	-	0.1	0.3	-	0.2	0.1

					92698; MH892699											
	Chlamydiae															
hlamydiia	hlamydiales	arachlamydiaceae	arachlamydiaceae sp.	459	H892691	-	-	-	-	-	-	-	-	-	0.1	0.1
		habdochlamydiaceae	*andidatus* Rhabdochlamydia sp.	044	H892708	-	-	-	-	-	-	-	-	-	-	<0.1
			hlamydiales spp.	597; 2741; 3295; 3540	H892695; MH892705; MH892711; MH892712	-	-	-	-	-	-	-	-	-	-	0.2
			hlamydiia spp.	035; 3270	H892707; MH89271	-	-	-	-	-	-	-	-	-	-	0.1

					0											
Firmicutes																
acilli	actobacillales	treptococcaceae	actobacillales spp.	1; 134; 443	H892616; MH892632; MH892657	1.5	-	-	-	-	2.3	-	-	1.9	0.2	0.2
			treptococcaceae sp.	10	H892678	0.1	-	-	-	-	-	-	-	-	-	-
			treptococcus spp.	; 559	H892611; MH892665	12.1	-	0.2	-	-	2.6	-	-	3.2	0.4	0.2
lostridia	lostridiales	lostridiaceae	lostridiaceae spp.	57; 1161	H892652; MH892684	-	-	0.2	0.1	-	-	-	-	-	-	-
			uminococcaceae sp.	387	H892688	<0.1	-	-	-	-	-	-	-	-	-	-
						0.1	0.4	-	-	-	0.1	-	-	0.1	-	-

			lostridiales spp.	59; 460; 1119; 1967	H892653; MH892659; MH892683; MH892701											
Bacilli	Lactobacillales	nterococcaceae	*nterococcus* spp.	86; 288; 1246	H892646; MH892648; MH892686	0.2	-	0.1	-	-	0.2	-	-	0.3	-	-
roteobacteria																
etaproteobacteria	urkholderiales	lcaligenaceae	lcaligenaceae sp.	9	H892619	-	-	<0.1	0.2	2.0	-	-	-	-	-	-
	eisseriales	eisseriaceae	*aribacter hongkongensis*	5	H892623	0.8	-	-	-	-	0.9	0.1	-	0.7	0.1	0.1

			eisseria canis	4	H892626	1.1	-	-	-	-	0.6	-	-	0.3	<0.1	0.1
			eisseria sp.	251	H892687	<0.1	-	-	-	-	-	-	-	-	-	-
			eisseriaceae spp.	9; 197; 287; 960	H892614; MH892639; MH892647; MH892679	0.9	-	-	0.1	-	4.2	0.3	-	4.0	0.2	0.1
			itreoscilla spp.	7; 169; 466	H892629; MH892636; MH892660	0.2	-	-	-	-	1.0	0.1	-	1.0	-	-
psilonproteobacteria	ampylobacterales	ampylobacteraceae	*rcobacter* spp.	18; 289; 598; 1174	H892642; MH892649;	0.3	-	-	-	-	0.6	-	-	0.1	-	-

					MH892670; MH892685											
			rcobacter venerupis	14	H892641	0.7	-	-	-	-	-	-	-	-	-	-
			ampylobacteraceae spp.	; 37; 59; 158; 229	H892609; MH892618; MH892624; MH892635; MH892643	13.4	0.2	1.6	0.9	0.1	14.9	1.9	-	20.7	3.8	0.7
ammaproteobacteria	eromonadales	eromonadaceae	eromonadaceae spp.	83; 639; 1476	H892661; MH892673; MH892693	0.2	-	-	-	-	0.1	-	-	-	-	-

			eromonas sp.		H892610	6.0	2.2	8.6	0.4	0.2	4.7	1.1	4.8	2.8	0.5	0.8
	nterobacteriales	nterobacteriaceae	nterobacteriaceae spp.	8; 54; 170; 183; 417; 546; 619	H892615; MH892622; MH892637; MH892638; MH892656; MH892663; MH892672	3.7	0.2	0.6	-	-	2.6	-	0.6	2.4	0.1	0.3
	egionellales	oxiellaceae	oxiellaceae sp.	92	H892677	-	-	-	-	-	-	-	-	-	-	0.4
			egionellales sp.	079	H892677	-	-	-	-	<0.1	-	-	-	-	-	-
		egion	*egion*	554	H892	-	-	-	-	-	-	-	-	-	-	0.1

		ellaceae	*ella* sp.		703											
	seudomonadales	oraxellaceae	*cinetobacter* spp.	0; 16; 41; 45; 75; 101; 283; 317; 564; 584; 886; 991; 992	H892612; MH892613; MH892620; MH892621; MH892627; MH892630; MH892645; MH892650; MH892666; MH892667; MH892676;	1.0	-	0.1	-	-	8.5	17.1	-	12.8	0.7	0.7

					MH892681; MH892682											
		seudomonadaceae	*seudomonas* spp.	5; 77; 109; 147; 151; 210; 233; 361; 402; 515; 556; 607; 678; 722; 1402	H892625; MH892628; MH892631; MH892633; MH892634; MH892640; MH892644; MH892654; MH892655; MH89266	1.1	-	1.3	0.1	1.3	0.3	3.1	1.4	0.6	-	.1

					2; MH892664; MH892671; MH892674; MH892675; MH892689											
			ammaproteobacteria spp.	440; 1709	H892690; MH892697	-	-	-	-	-	-	-	-	-	0.2	0.1
	ibrionales	seudoalteromonadaceae	*ibrio* sp.	89	H892669	-	-	-	-	0.1	-	-	-	-	-	-
	Spirochaetes															
eptospirae	eptospirales	eptospiraceae	eptospiraceae sp.	146	H892702	-	-	-	-	-	-	-	-	-	0.1	-
						-	-	-	-	-	-	-	-	-	<0.1	-

			pirochaetes sp.	988	H892706											
			urneriella sp.	946	H892700	-	-	-	-	-	-	-	-	-	0.1	0.1
pirochaetia	pirochaetales	pirochaetaceae	pirochaetaceae spp.	45; 965; 1564	H892658; MH892680; MH892694	-	0.1	0.1	0.1	-	-	-	-	-	-	-

[a]Most specific level of taxonomy designated after comparing ZOTUs to Greengenes and NCBI nr/nt databases.

Table 5. Sequence composition (%) of nitrifying, denitrifying and floc-forming bacteria in WWTPs 1-4 influent (I), intermediate (INT) and effluent (E) with taxonomy confirmed with Greengenes and NCBI nr/nt sequence databases.

								W					W			W
						WTP 1			WTP 2		WTP 3			WTP 4		
Class	**Order**	**Family**	**OTU no.**	**Accession no.**	**Species**	**I**	**NT**	**E**	**I**	**E**		**INT**	**E**	**I**	**INT**	**E**
Bacteroidetes																
Flavobacteria	Flavobacteriales	Flavobacteriaceae	35; 44; 57; 103; 153; 155; 172; 178; 263; 365; 474; 660; 766; 970	MH892717; MH892718; MH892720; MH892728; MH892733; MH892734; MH892737; MH892738; MH892745; MH892752; MH892763; MH892771;	*Flavobacterium* spp.	-	0.8	0.4	2.8	4.7	-	0.2	7.6	0.1	0.4	6.5

			; 1142; 1241; 1883; 1912; 2042; 2242; 2349; 2374; 2375; 2493; 2649; 2905; 3231; 3371	MH892779; MH892792; MH892798; MH892800; MH892811; MH892813; MH892815; MH892816; MH892817; MH892818; MH892819; MH892820; MH892821; MH892823; MH892825; MH892826												
Nitrospirae																
Nitrospira	Nitrospirales	Nitrospira	404;	MH892758;	*Nitrospira* spp.	-	-	-	-	-	-	-	-	-	1.2	1.5

			614; 1574; 2690	MH892767; MH892808; MH892822												
Proteobacteria																
Betaproteobacteria	Burkholderiales	Comamonadaceae	27	MH892715	*Aquabacterium* sp.	-	0.3	0.5	1.3	0.2	0.1	3.8	0.1	0.1	0.1	0.1
			647	MH892769	*Brachymonas denitrificans*	0.1	-	-	-	-	-	-	-	-	-	-
			94; 677; 356; 492; 776; 580; 926; 1619; 3101; 855	MH892726; MH892772; MH892751; MH892765; MH892780; MH892766; MH892790; MH892809; MH892824; MH892785	Comamonadaceae spp.	0.3	0.1	0.1	-	0.3	0.6	0.6	-	0.4	0.6	1.1
			11	MH892713	Comamonas sp.	6.9	-	0.1	0.1	-	4.4	1.4	-	4.4	0.2	0.2
			133	MH892	Delftia sp.	-	-	-	-	-	-	-	-	-	-	-

			6	802												
			73; 85; 100; 184; 225; 275; 319; 426; 449; 454; 1420	MH892723; MH892724; MH892727; MH892741; MH892744; MH892747; MH892749; MH892759; MH892761; MH892762; MH892805	*Hydrogenophaga* spp.	-	0.2	0.6	0.4	2.4	-	2.8	0.3	<0.1	-	<0.1
			1109; 1403	MH892796; MH892804	*Polaromonas* spp.	-	-	-	0.1	<0.1	-	<0.1	-	-	-	0.2
			716; 904	MH892774; MH892788	*Rhodoferax* spp.	-	0.2	-	-	-	-	-	-	-	-	0.5
			762; 3498	MH892778; MH892828	*Rubrivivax* spp.	-	-	<0.1	-	0.1	-	-	-	-	-	-

	Rhodocyclales	Rhodocyclaceae	171; 823; 1898	MH892736; MH892782; MH892812	*Azoarcus* spp.	-	-	-	-	0.6	-	-	-	-	-	-
			387; 980	MH892755; MH892794	*Azonexus* spp.	<0.1	-	-	-	0.2	-	-	-	-	-	-
			193; 863	MH892743; MH892786	*Azospira* spp.	0.1	<0.1	-	-	-	-	-	-	-	2.3	1.5
			11; 30; 71; 302; 490	MH892713; MH892716; MH892722; MH892748; MH892764	*Dechloromonas* spp.	1.8	-	0.1	0.1	-	0.4	5.6	-	0.3	1.0	1.2
			1462	MH892806	*Methyloversatilis* sp.	-	-	-	-	<0.1	-	-	-	-	-	-
			49; 70; 104; 389; 717; 847	MH892719; MH892721; MH892729; MH892756; MH892	*Propionivibrio* spp.	2.2	-	-	0.2	-	2.1	0.9	-	1.8	0.3	0.1

			; 901	775; MH892784; MH892787												
			623; 1139; 1305; 1367; 1546; 148; 269; 373; 754; 1151	MH892768; MH892797; MH892801; MH892803; MH892807; MH892732; MH892746; MH892753; MH892777; MH892799	Rhodocyclaceae spp.	0.6	<0.1	-	0.3	0.1	0.4	0.1	-	0.6	0.3	0.4
			787; 385; 977	MH892781; MH892754; MH892793	*Sterolibacterium* spp.	-	-	-	-	-	-		-	-	1.3	1.2
			1796	MH892810	*Sulfuritalea* sp.	-	<0.1	-	-	-	-	-	-	-	-	-
			23; 91;	MH892714;	*Thauera* spp.	3.4	-	<0.1	-	0.2	0.5	6.0	-	0.4	-	0.2

			126; 924	MH892725; MH892731; MH892789												
			191; 1974	MH892742; MH892814	*Uliginosibacterium* spp.	-	-	<0.1	-	-	-	-	-	-	2.4	1.4
			120; 180; 181; 335	MH892730; MH892739; MH892740; MH892750	*Zoogloea* spp.	0.7	-	<0.1	0.1	-	0.8	0.6	-	0.5	0.1	1.0
	Unclassified	Unclassified	165; 394; 441; 655; 693; 728; 825; 938; 101	MH892735; MH892757; MH892760; MH892770; MH892773; MH892776; MH892783; MH892791; MH892	*Candidatus* Accumulibacter spp.[b]	-	-	-	-	-	-	-	-	-	5.0	5.9

			5; 3404	795; MH892827												

[a]Most specific level of taxonomy designated after comparing ZOTUs to Greengenes and NCBI nr/nt sequences.

[b]*Candidatus* Accumulibacter spp. was assigned by Greengenes to the family Rhodocyclaceae, but is a recently discovered bacterium that has not yet been classified to an order or family.

Highlights

- Waste water samples were screened with bacterial 16S next-generation sequencing
- The V4 region of 16S could not differentiate Enterobacteriaceae
- Only three pathogens could be identified to the species level
- Erroneous taxa in the 16S Greengenes database were identified
- NCBI nr/nt database comparisons provided more accurate taxonomic assignments

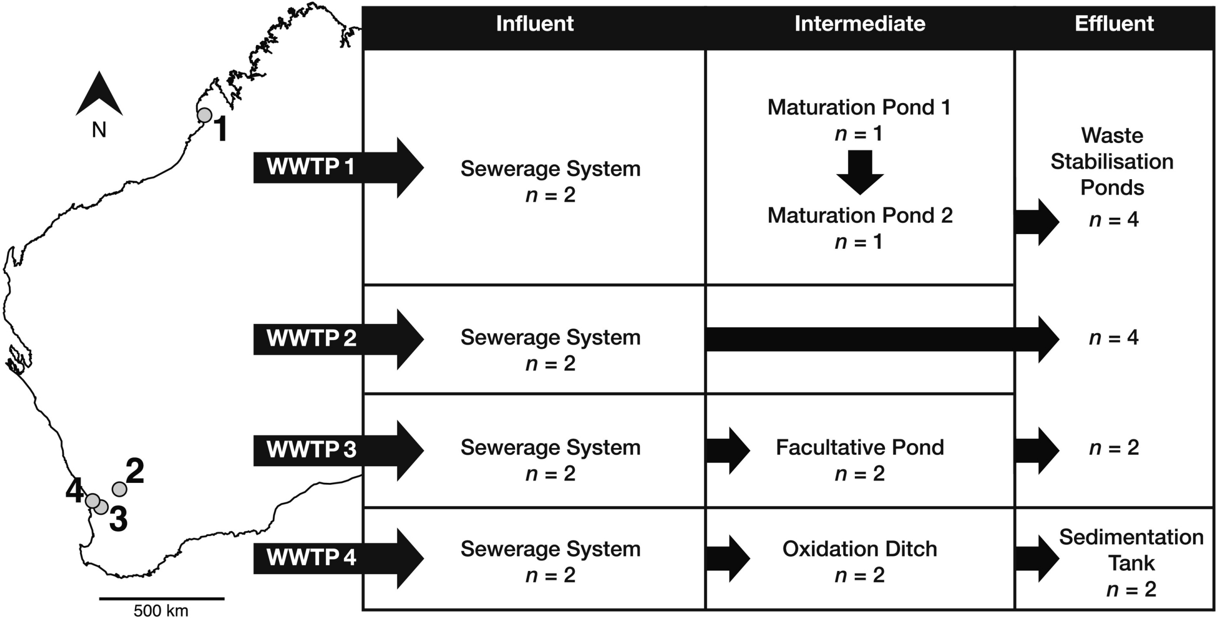
N
1
2
3
4
500 km
Influent
Intermediate
Effluent
WWTP 1
Sewerage System
n = 2
Maturation Pond 1
n = 1
Maturation Pond 2
n = 1
Waste Stabilisation Ponds
n = 4
WWTP 2
Sewerage System
n = 2
n = 4
WWTP 3
Sewerage System
n = 2
Facultative Pond
n = 2
n = 2
WWTP 4
Sewerage System
n = 2
Oxidation Ditch
n = 2
Sedimentation Tank
n = 2

Figure 1

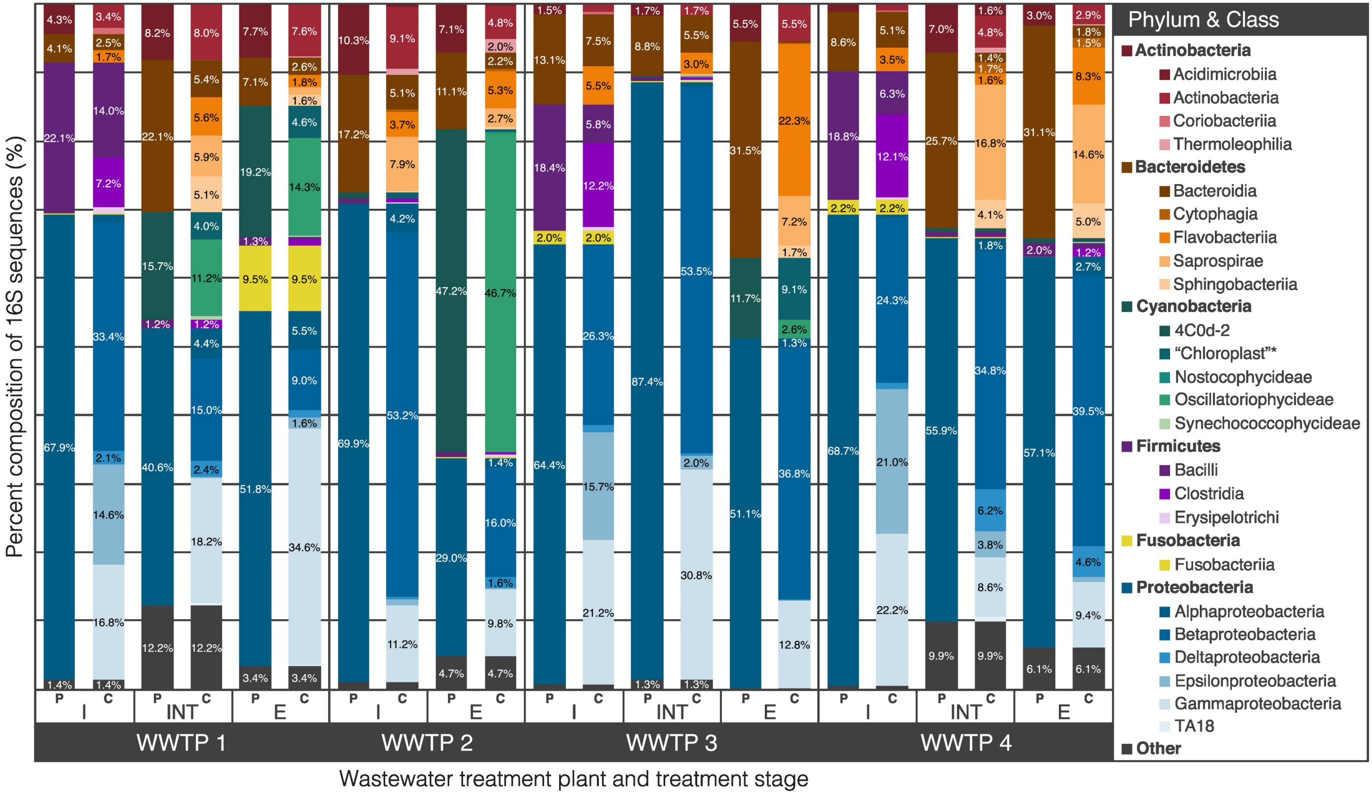
Percent composition of 16S sequences (%)
Wastewater treatment plant and treatment stage
WWTP 1
WWTP 2
WWTP 3
WWTP 4
I
INT
E
P
C
4.3%
4.1%
22.1%
67.9%
1.4%
3.4%
2.5%
1.7%
14.0%
7.2%
33.4%
2.1%
14.6%
16.8%
1.4%
8.2%
22.1%
15.7%
1.2%
40.6%
12.2%
8.0%
5.4%
5.6%
5.9%
5.1%
4.0%
11.2%
1.2%
4.4%
15.0%
2.4%
18.2%
12.2%
7.7%
7.1%
19.2%
1.3%
9.5%
51.8%
3.4%
7.6%
2.6%
1.8%
1.6%
4.6%
14.3%
9.5%
5.5%
9.0%
1.6%
34.6%
3.4%
10.3%
17.2%
69.9%
9.1%
5.1%
3.7%
7.9%
4.2%
53.2%
11.2%
7.1%
11.1%
47.2%
29.0%
4.7%
4.8%
2.0%
2.2%
5.3%
2.7%
46.7%
1.4%
16.0%
1.6%
9.8%
4.7%
1.5%
13.1%
18.4%
2.0%
64.4%
7.5%
5.5%
5.8%
12.2%
2.0%
26.3%
15.7%
21.2%
1.7%
8.8%
87.4%
1.3%
1.7%
5.5%
3.0%
53.5%
2.0%
30.8%
1.3%
5.5%
31.5%
11.7%
51.1%
5.5%
22.3%
7.2%
1.7%
9.1%
2.6%
1.3%
36.8%
12.8%
8.6%
18.8%
2.2%
68.7%
5.1%
3.5%
6.3%
12.1%
2.2%
24.3%
21.0%
22.2%
7.0%
25.7%
55.9%
9.9%
1.6%
4.8%
1.4%
1.7%
1.6%
16.8%
4.1%
1.8%
34.8%
6.2%
3.8%
8.6%
9.9%
3.0%
31.1%
2.0%
57.1%
6.1%
2.9%
1.8%
1.5%
8.3%
14.6%
5.0%
1.2%
2.7%
39.5%
4.6%
9.4%
6.1%
Phylum & Class
Actinobacteria
Acidimicrobiia
Actinobacteria
Coriobacteriia
Thermoleophilia
Bacteroidetes
Bacteroidia
Cytophagia
Flavobacteriia
Saprospirae
Sphingobacteriia
Cyanobacteria
4C0d-2
"Chloroplast"*
Nostocophycideae
Oscillatoriophycideae
Synechococcophycideae
Firmicutes
Bacilli
Clostridia
Erysipelotrichi
Fusobacteria
Fusobacteriia
Proteobacteria
Alphaproteobacteria
Betaproteobacteria
Deltaproteobacteria
Epsilonproteobacteria
Gammaproteobacteria
TA18
Other

Figure 2

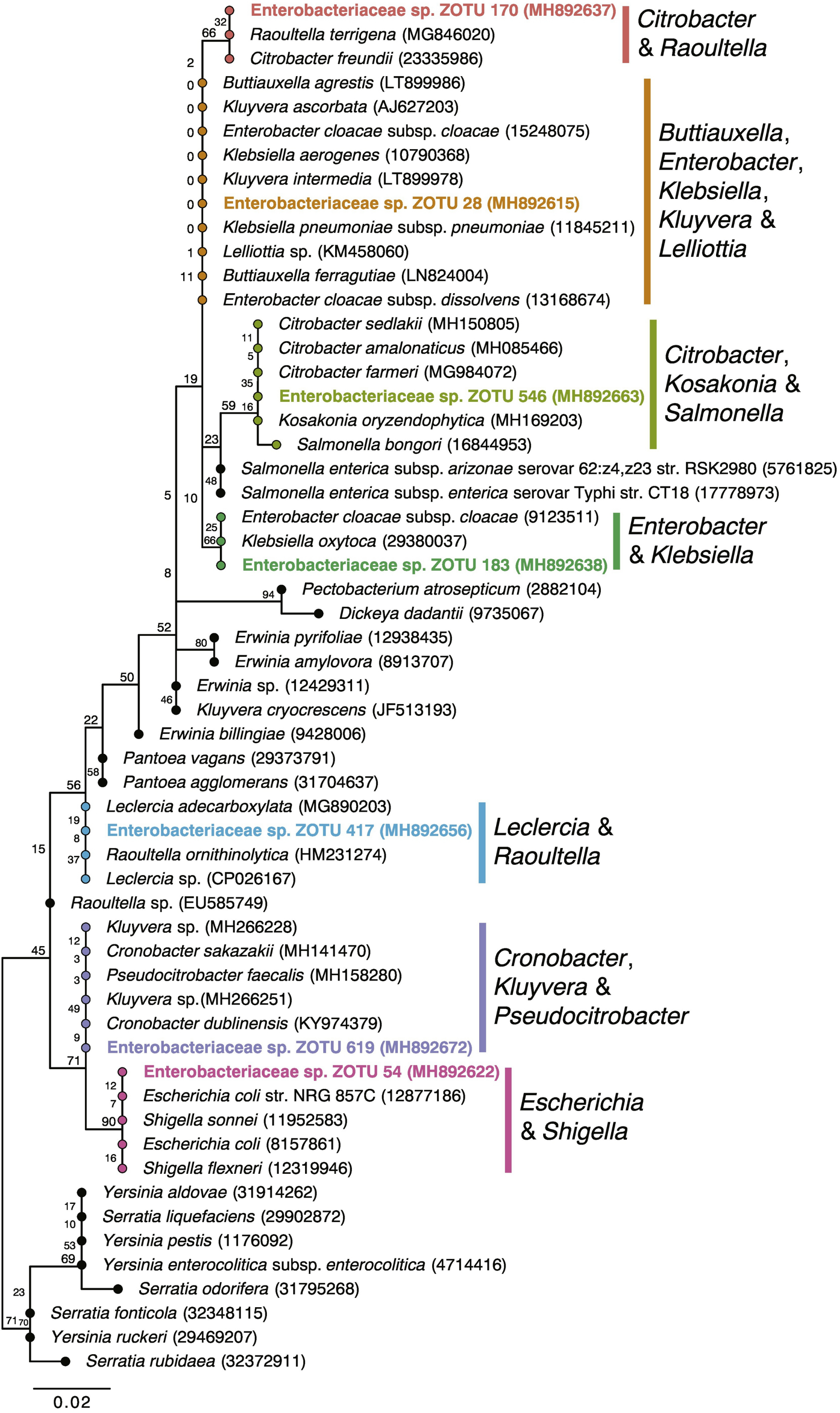
Enterobacteriaceae sp. ZOTU 170 (MH892637)
Raoultella terrigena (MG846020)
Citrobacter freundii (23335986)
Citrobacter & Raoultella
Buttiauxella agrestis (LT899986)
Kluyvera ascorbata (AJ627203)
Enterobacter cloacae subsp. cloacae (15248075)
Klebsiella aerogenes (10790368)
Kluyvera intermedia (LT899978)
Enterobacteriaceae sp. ZOTU 28 (MH892615)
Klebsiella pneumoniae subsp. pneumoniae (11845211)
Lelliottia sp. (KM458060)
Buttiauxella ferragutiae (LN824004)
Enterobacter cloacae subsp. dissolvens (13168674)
Buttiauxella, Enterobacter, Klebsiella, Kluyvera & Lelliottia
Citrobacter sedlakii (MH150805)
Citrobacter amalonaticus (MH085466)
Citrobacter farmeri (MG984072)
Enterobacteriaceae sp. ZOTU 546 (MH892663)
Kosakonia oryzendophytica (MH169203)
Salmonella bongori (16844953)
Citrobacter, Kosakonia & Salmonella
Salmonella enterica subsp. arizonae serovar 62:z4,z23 str. RSK2980 (5761825)
Salmonella enterica subsp. enterica serovar Typhi str. CT18 (17778973)
Enterobacter cloacae subsp. cloacae (9123511)
Klebsiella oxytoca (29380037)
Enterobacteriaceae sp. ZOTU 183 (MH892638)
Enterobacter & Klebsiella
Pectobacterium atrosepticum (2882104)
Dickeya dadantii (9735067)
Erwinia pyrifoliae (12938435)
Erwinia amylovora (8913707)
Erwinia sp. (12429311)
Kluyvera cryocrescens (JF513193)
Erwinia billingiae (9428006)
Pantoea vagans (29373791)
Pantoea agglomerans (31704637)
Leclercia adecarboxylata (MG890203)
Enterobacteriaceae sp. ZOTU 417 (MH892656)
Raoultella ornithinolytica (HM231274)
Leclercia sp. (CP026167)
Leclercia & Raoultella
Raoultella sp. (EU585749)
Kluyvera sp. (MH266228)
Cronobacter sakazakii (MH141470)
Pseudocitrobacter faecalis (MH158280)
Kluyvera sp.(MH266251)
Cronobacter dublinensis (KY974379)
Enterobacteriaceae sp. ZOTU 619 (MH892672)
Cronobacter, Kluyvera & Pseudocitrobacter
Enterobacteriaceae sp. ZOTU 54 (MH892622)
Escherichia coli str. NRG 857C (12877186)
Shigella sonnei (11952583)
Escherichia coli (8157861)
Shigella flexneri (12319946)
Escherichia & Shigella
Yersinia aldovae (31914262)
Serratia liquefaciens (29902872)
Yersinia pestis (1176092)
Yersinia enterocolitica subsp. enterocolitica (4714416)
Serratia odorifera (31795268)
Serratia fonticola (32348115)
Yersinia ruckeri (29469207)
Serratia rubidaea (32372911)
32
66
2
0
0
0
0
0
0
0
1
11
11
5
35
59
16
19
23
48
5
10
25
66
8
94
52
80
50
46
22
58
56
19
8
37
15
45
12
3
3
49
9
71
12
7
90
16
17
10
53
69
23
71
70
0.02

Figure 3