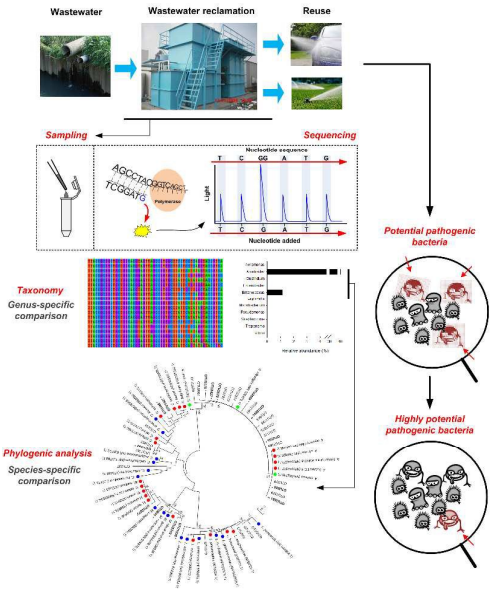
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances



Efficient diagnosis based on pyrosequencing can address the highly potential pathogenic bacteria in wastewater.

Occurrence and fate of potential pathogenic bacteria as revealed by pyrosequencing in a full-scale membrane bioreactor treating restaurant wastewater

Jinxing Ma[1], Zhiwei Wang[*1], Lili Zang[2], Jian Huang[1], Zhichao Wu[1]

[1] State Key Laboratory of Pollution Control and Resource Reuse, School of Environmental Science and Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, P.R. China

[2] Shanghai Zizheng Environm Technol Co Ltd, Shanghai 200437, P.R. China

Revised Manuscript Submitted to RSC Advances

*Corresponding Author. Tel./fax: +86-21-65980400 (Z. Wang); E-mail address: zwwang@tongji.edu.cn (Z. Wang)

Abstract

One of the primary concerns on wastewater reuse is the presence of pathogenic bacteria. Considering that indicator bacteria might only offer limited information, we applied high-throughput pyrosequencing in this study to reveal bacterial pathogen diversity in a full-scale membrane bioreactor (MBR) treating restaurant wastewater. The results showed that fecal indicator bacteria could provide a rough estimation rather than an accurate characterization of the potential pathogenic bacteria in wastewaters particularly from non-fecal sources. In general, MBR treatment had a good removal of potential pathogenic bacteria. The bacterial counts of *Arcobacter* was decreased by nearly seven orders of magnitude, from $(8.35 \pm 0.87) \times 10^7$ to <10 counts/mL, and *Aeromonas*, *Enterobacter*, *Enterococcus*, and *Pseudomonas* were not detected in the treated wastewater. The most dominant potential pathogens in activated sludge and treated wastewater were affiliated to the genera of *Legionella*, *Clostridium* and *Mycobacterium*. Species-specific comparison showed that only a small portion (0.0~1.6%) of the corresponding sequences had identities of > 99% to the neighbor pathogenic species, including *Arcobacter butzleri* and *Arcobacter cryaerophilus*. This study, therefore, provides insights into the occurrence and fate of potential bacterial pathogens in restaurant wastewater treatment and reclamation using MBRs.

Keywords: pyrosequencing; membrane bioreactor; pathogenic bacteria; restaurant wastewater

1. Introduction

During the last decade, catering industry has experienced an explosive growth in China and the business turnover was doubled from 2006 (≈$ 161 billion per year) to 2011 (≈ $ 332 billion per year) [1]. Wastewater streams discharged from restaurants are generally characterized by high content of oil, grease (O&G), suspended solid (SS) and detergent [2, 3]. High O&G is a tremendous burden to the municipal wastewater systems because these organic substances usually tend to clump together, causing drainage pipelines corrosion under anaerobic conditions. Therefore, appropriate treatment of the restaurant wastewater is necessary in order to reduce the adverse impacts of discharging [4]. As an option, the membrane bioreactor (MBR) is a fascinating and promising technology, which presents distinctive advantages such as high volumetric organic loading, small environmental footprint, and sound separation of emulsions that contain oil droplets with diameter less than 20 μm [5]. MBRs also offer the opportunity to spare the expenditure of

wastewater treatment since their superior effluent is more suitable for on-site reuse (e.g., flushing toilets) in the restaurants.

On condition that treated wastewater utilization is expected, contaminant removals should be sufficient to meet stringent regulatory standards, because of the public health concerns [6]. In MBRs, bacteria play an important role in the biochemical process, consuming nutrients and organic matters. To date, numerous studies have been conducted to improve the efficiency of the biochemical process, but most of them failed to attach importance to the potential hazard of the bacteria accordingly. One of the primary concerns on bacteria is the community of pathogenic bacteria originated from the excrement of disease-carrying humans and animals or other sources [7]. In many public places (e.g., general merchandise stores and restaurants), the outbreak of gastroenteritis or other infections due to access of reused water could actually be masked by the background levels of assumed sources, such as food-borne and community-based infections [8]. Microbial assessment of pathogenic bacteria in treated wastewater is thereby important in view of consequent health risks.

Historically, fecal indicator bacteria including total and fecal coliforms and enterococci have been widely used as a monitoring tool to predict the presence of potential bacterial, viral and protozoan pathogens [9]. The major drawback of fecal indicator bacteria arises from their poor correlation with pathogens, especially those from non-fecal sources [8, 10]. Moreover, membranes have a size-selective retention of different bacteria, and the abundance of pathogenic bacteria in the permeate could be underestimated or overestimated when referred to certain indicator bacteria [11]. In recent years, real-time qPCR assays have been proposed and these assays are now used in many diagnostic and reference laboratories for the detection of pathogenic bacteria in clinical fluids [8, 12, 13]. Compared to indicator bacteria methods, the qPCR assay enables quantitative and highly specific detection, which could target 16S ribosomal RNA, encoding genes or housekeeping genes of actual pathogens [12]. Nevertheless, the application of this technology is still hindered due to its limited throughput capacity. In environmental samples, bacterial pathogen diversity can be extremely high, as reflected by more than thirty phylogenetic genera and thousands of strains. Clueless one-by-one detection is definitely time-consuming, which might also miss the potential infectious risk. Therefore, illuminating solutions are urgently required for

elucidating bacterial pathogen diversity in MBRs and assessing full microbial risk of treated wastewater reuse in public places.

In this study, 454 high-throughput pyrosequencing was used to investigate the occurrence and fate of potential pathogenic bacteria in a full-scale MBR treating restaurant wastewater. Pyrosequencing is a high-throughput analytical method that generates a large amount of DNA reads through a massively parallel sequencing-by-synthesis approach, and this technology can provide an adequate resolution to the microbial diversity of different environmental samples [7, 14-16]. In the present work, 258,438 reads of the hypervariable V1~V3 regions of the bacterial 16S rRNA gene were obtained. Sequence subsets with the capacity of 10,000 and 100 were generated from the maternal gene libraries by a semi-random extraction method, and comprehensive comparison of these datasets was then carried out. Bacterial pathogen diversity was analyzed at genus level using Ribosomal Database Project (RDP) Classifier [17]. Alignment of the corresponding sequences to the known pathogen was further conducted by phylogenetic analysis.

2. Materials and methods

2.1. Sample collection and pyrosequencing

Sewage and sludge samples for pyrosequencing were taken from a full-scale MBR. The reactor, as schematically shown in Fig. S1 of the Supporting Information, was located in a general merchandise store (31.3°N 121.4°E) of Shanghai, China and has been in operation for over 6 years. The influent wastewater of the MBR includes 1) fresh food processing (FFP) wastewater, 2) restaurant wastewater generated from restaurants serving Chinese, Japanese and Western style food, 3) toilet flushing wastewater, 4) greywater from office region and washing basins, and 5) car washing wastewater. The raw wastewater passed through screens, a dissolved air flotation tank and an aerobic MBR tank. The treated wastewater was temporarily stored in an effluent tank and finally reused for toilet flushing, lawn watering and car washing. The MBR tank had an effective volume of 60 m^3. 600 poly(vinylidene fluoride) flat-sheet membrane modules (Zizheng Environm Technol Co. Ltd., Shanghai, China) with a mean pore size of 0.20 μm were installed in the tank. Details about MBR setup, characteristics of the influent and treated wastewater are summarized in the Supporting Information (Section I, Fig. S1 and Table S1 in the Supporting Information).

Influent wastewater, activated sludge and treated wastewater samples, termed as A1, A2 and A3 samples, were taken from the inlet pipe, aerobic tank and outlet pipe of the MBR, respectively

(see Fig. S1). After DNA extraction and PCR amplification (see Section II of the Supporting Information), amplicons from A1, A2 and A3 were mixed at the equal concentration, and the mixture was used for pyrosequencing on a Roche 454 FLX Titanium platform at Majorbio Bio-Pharm Technology Co., Ltd (Shanghai, China).

2.2. Read quality control and subset construction

After pyrosequencing, 258,438 raw reads (0.1 G) were obtained according to the unique match to the barcodes (Table 1). The results were deposited into the NCBI short reads archive database (Accession Number: SRA169387). To improve the validity of subsequent data processing, Qiime (version 1.17 http://qiime.org/) was applied to (1) check the completeness of 3' end of primers and adaptors; (2) remove reads containing ambiguous base ('N') or homologous run that was longer than 10-nucleotide; (3) enable sliding window test of quality scores (-w 50 and -s 20); and (4) remove reads shorter than 200 bps [18]. Barcodes and primers were also stripped from resulting sequences, and finally pyrosequencing produced 24,962 (A1), 113,131 (A2) and 54,525 (A3) high-quality V1-V3 tags of the 16S rRNA gene with an average length of 462 bp (Table 1).

Table 1 Statistical summary for pyrosequencing and microbial diversity analysis

Sample ID	Raw reads	High-quality reads	Assigned reads[a]	OTU	Chao	Shannon	*Fo*
A1[b]	34949	24962	19411	897	1323	3.70	0.92
B1	n.a.[c]	10000	8162	864	1886	3.88	0.89
C1	n.a.	100	89	40	164	2.90	0.64
A2	158938	113131	63243	1712	2026	4.85	0.90
B2	n.a.	10000	6163	1132	2664	5.04	0.81
C2	n.a.	100	63	46	187	3.65	0.42
A3	64551	54525	36644	1362	1670	5.20	0.89
B3	n.a.	10000	7132	1063	2092	5.34	0.83
C3	n.a.	100	75	59	190	3.95	0.37

a. Assigned reads are the reads that match the OTU in each sample. Some high-quality reads may not match any OTU for these reasons: (1) the read is chimeric, and (2) the read that has a singleton sequence is discarded.

b. A1, A2 and A3 represent pyrosequencing results of influent wastewater, activated sludge and treated wastewater samples; B1, B2 and B3 represent the subsets with 10000 reads extracted from A1, A2 and A3; C1, C2 and C3 represent the subsets with 100 reads extracted from B1, B2 and B3.

c. n.a. indicates the value is not available.

For a comprehensive understanding of the impacts of sequencing depth, subsets with the capacity of 10,000 and 100 were generated from the high-quality maternal sets of A1~A3 by a semi-random extraction method. Initially, the sub.samlpe command of MOTHUR program (http://www.mothur.org/wiki/Sub.sample) was used for A1~A3 to create 30 subsets comprised of

10,000 sequences, i.e., $B1_i = \{x_j \mid x_j \in A1, j = 1\sim1000\}$ ($i = 1\sim 10$), $B2_i = \{x_j \mid x_j \in A2, j = 1\sim1000\}$ ($i = 1\sim 10$) and $B3_i = \{x_j \mid x_j \in A3, j = 1\sim1000\}$ ($i = 1\sim 10$). Principal coordinates analysis (PCoA) with the Bray-Curtis index (R package, http://www.r-project.org/) was then performed to evaluate the relationship between A1~A3 and $B1_i$~$B3_i$ ($i = 1\sim10$), and the subsets with the highest homology were retained and specified as B1~B3 (Fig. S2 in the Supporting Information). Afterwards, a similar procedure was applied to create 30 subsets containing 100 sequences from B1~B3, and the subsets with the highest homology with B1~B3 were specified as C1~C3. Despite the debate that semi-random extraction is reliable enough compared to independent sequencing, this method is similar to pyrosequencing run in reverse; and in practice the final gene libraries (e.g., A1~A3) can be obtained based on the deficient datasets (e.g., B1~B3) by further sequencing of the amplicons. Nevertheless, evaluation of pathogenic bacteria diversity was mainly based on the original pyrosequencing results.

2.3. Phylogenetic classification and biodiversity analysis

Cluster of the high-quality reads into operational taxonomic units (OTUs) was performed using UPARSE pipeline (vsesion 7.1, http://drive5.com/uparse/) [19]. Briefly, abundance-sorted reads of the nine datasets (A1~A3, B1~B3 and C1~C3) were clustered by setting a minimum identity of 97%, and the uchime_ref command was used to filter out chimeras. The abundances of OTUs in each dataset were obtained by searching the reads as a query set against the OTU representative sequences. For the cluster files, alpha-diversity and rarefaction curves were generated in MOTHUR for each sample (version v.1.30.1, http://www.mothur.org). Functional organization indices (*Fo*) were calculated according to the standard method reported by Marzorati et al. [20]. Representative sequences from each OTU were assigned down to the phylum and genus level using the RDP Classifier with a set confidence threshold of 80% (https://rdp.cme.msu.edu/classifier/classifier.jsp, 16S rRNA training set 10)[17].

Venn diagrams with shared and unique OTUs were utilized to depict the similarity and difference between microbial communities. A pairwise statistical comparison of taxonomy at phylum level between maternal sets and subsets was carried out using STAMP [21]. Biological relevance between samples at genus level was evaluated using linear regression of SigmaPlot software (version 12.5, Systat Software, Inc., U.S.). Furthermore, LDA Effect Size (LEfSe) algorithm was introduced herein to identify taxa that characterize the differences among the three

environmental samples [22]. A1~A3, B1~B3 and C1~C3 were grouped according to the source (e.g., influent wastewater, activated sludge or treated wastewater sample), and each sample was firstly normalized to the sum of the values of 0.05 M. The parameters for data processing were set as follows: 'alpha value for the factorial Kruskal-Wallis test among classes' = 0.05, 'threshold on the logarithmic LDA score for discriminative features' = 2.7, and 'set the strategy for multi-class analysis' = all-against-all.

Alignment of microbial communities to pathogenic genera was firstly evaluated using the taxonomic results of RDP Classifier. The lists of known pathogenic genera summarized by Ye and Zhang and Biddy et al. were used as reference [7, 23]. Representative sequences from OTUs that were assigned into *Arcobacter*, *Clostridium*, *Legionella*, and *Mycobacterium* were further separated for phylogenetic analysis at species level. 16S rRNA gene of known pathogens and non-pathogens from the four genera were achieved from NCBI Genbank (Table S2 of the Supporting Information), and merged with the corresponding sequences of this study into a fasta file. ClustalW was used for aligning and bootstrapping of the phylogenetic tree, which was then viewed, edited and published with MEGA 6 [24]. Default settings were used. Furthermore, bacteria assigned to the families of Enterobacteriaceae and Enterococcaceae were regarded as the representative fecal indicators in this study.

2.4. Quantification of bacterial biomass using flow cytometer (FCM)

Bacteria biomass in wastewater and sludge samples was quantified using flow cytometer. Initially, influent wastewater, activated sludge and treated wastewater (A1, A2 and A3) were diluted 1:20 (v/v), 1:500 (v/v) and 1:1 (v/v) using 0.22-μm filtered phosphate-buffered-saline solution (0.84 ‰, pH = 7) to achieve optimal concentrations of bacteria for FCM analysis. Then the diluted mixtures (A1, A2 and A3) were subjected to ultrasonication treatment at power densities of 25, 80 and 0 kJ/L, respectively. After filtrated with 10-μm filters, samples were stained with SYBR Green I at a ratio of 100:1, incubated for 15 min in the dark at room temperature and finally processed to the flow cytometer (BD Accuri^{TM} C6, U.S.). Each sample was test in triplicate and total bacterial counts of A1, A2 and A3 were $(2.31 \pm 0.24) \times 10^8$, $(7.06 \pm 0.30) \times 10^9$ and $(3.35 \pm 0.82) \times 10^4$ counts/mL, respectively.

3. Results

3.1. Diversity and similarity analysis of microbial communities

By performing the alignment at α of 0.03 using UPARSE pipeline, 897, 1712 and 1362 OTUs were obtained from A1, A2 and A3 (Table 1). At a degraded and uniform library size of 10,000, Chao1 richness estimators of the three samples were 1886, 2664 and 2092, and Shannon diversity indices were 3.88, 5.04 and 5.34, respectively. Alpha-diversity analysis suggested that bacterial community from the influent wastewater sample had the lowest microbial richness and diversity. Moreover, *Fo* of the three samples were 0.89~0.92. It could be inferred that all the microbial communities were highly functionally organized [14, 20]. Pairwise comparison using Venn analysis showed that the similarity of A2-A3 was the highest, followed by that of A1-A2 and that of A1-A3 (Fig. 1). Notably, A2 and A3 had 707 shared OTUs that contained 65.0% and 77.8% of the reads, respectively. In contrast, only very few reads (0.5% and 0.1%) were classified into the OTUs that were shared by A1 and A3, which indicated that MBR treatment introduced a profound influence on the structure of microbial community in wastewater.

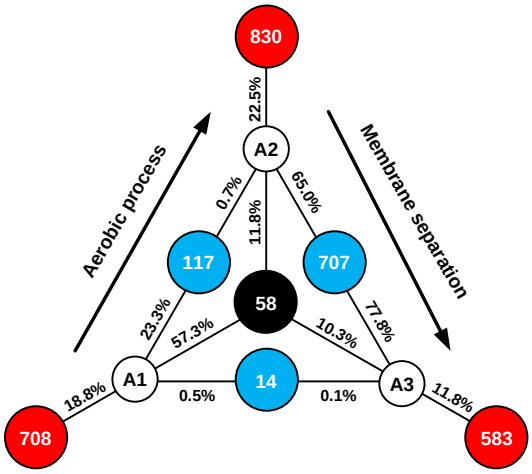


Fig. 1 Similarity analysis of the microbial communities (A1, A2 and A3) based on the clustering results at 3% distance cutoff. The numbers in the black circles represent the number of OTUs that is present in the core OTUs shared by the three samples. The numbers in the blue circles represent the OTUs shared by two samples. The numbers in the red circles represent the unique OTUs observed in only one sample. Percentages listed beside the branches indicate the percentages of reads of each sample assigned into the nearby OTUs groups.

3.2. Impacts of sequencing depth

Generally, microbial communities of environmental samples are highly diverse, and in this study rarefaction curves showed that new bacterial phylotypes continued to emerge even after 60,000 reads sampled (Fig. S3 in the Supporting Information). Addressing an appropriate

sequencing depth is crucial for high-throughput pyrosequencing to detect pathogenic bacteria at low abundance; an enlarged depth significantly increases the sequencing and processing cost, while a small library size could only provide insufficient resolution. Since 10,000 and 100 library sizes are always considered in pyrosequencing and conventional molecular biology studies, sequence subsets with corresponding capacities were generated from the maternal gene libraries by the semi-random extraction method.

Table 1 indicated that insufficient resolution reduced the accuracy of alpha-diversity analysis. For instance, at the sequencing depth of 100, only 40, 46 and 59 OTUs were predicted for the whole microbial communities in influent wastewater, activated sludge and treated wastewater, respectively. Furthermore, we compared the taxonomic results of the nine datasets at phylum and genus levels (Fig. 2). In total, 25 phyla were classified at the threshold of 80%. Proteobacteria was the most dominant phylum, accounting for 53.1~68.3% of total communities, respectively. Pairwise comparison using STAMP shows that there is no significant dissimilarity of taxonomic results between Aj and Cj (j = 1, 2, 3) at phylum level (Fig. 2a, Fig. 2c and Fig. 2e). However, the reliability was significantly declined with the taxonomy down to genus level. Most taxa of A1~A3 could not be predicted by the taxonomic results of C1~C3 at a 95% predication band. In contrast, linear regression showed that except for a few categories, B1~B3 supplied a credible characterization of the microbial communities of A1~A3 at genus level (Fig. 2b, Fig. 2d and Fig. 2f). The results suggested that compared to low-throughput sequencing methods pyrosequencing could provide a more valid estimation of the population structure of diverse communities especially at terminal taxonomic levels (e.g., genus level). Increasing the library size from 10,000 to 100,000, however, did not improve the taxonomic results as expected, probably due to the abundance of singletons at high sequencing depth, which were always discarded after denoising (Table 1).

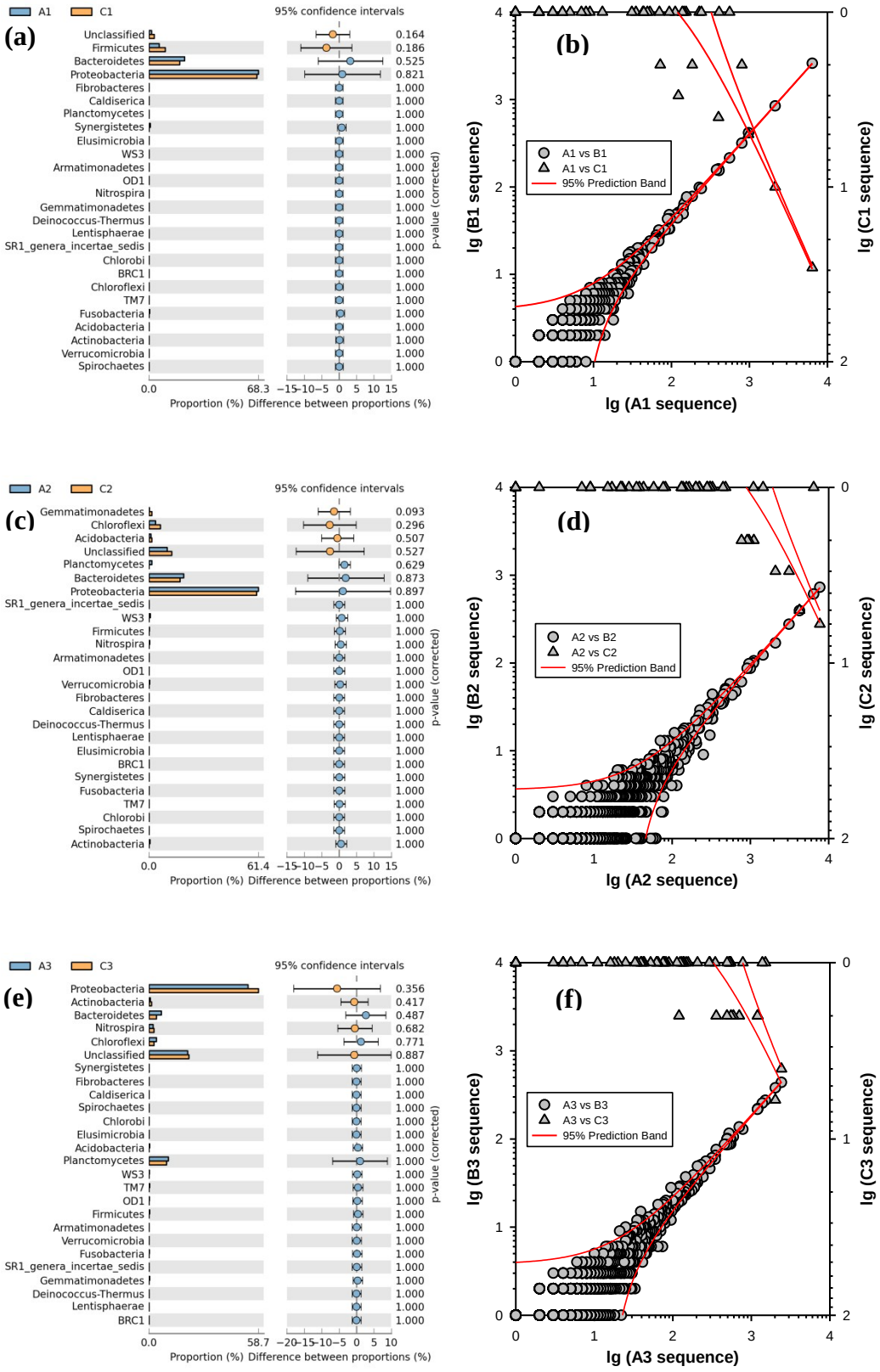


Fig. 2 Pairwise comparison of biological relevance of (a) A1 and C1 at phylum level, (b) A1, B1 and C1 at genus level, (c) A2 and C2 at phylum level, (d) A2, B2 and C2 at genus level, (e) A3 and C3 at phylum and (f) A3, B3 and C3 at genus level. Taxonomic results based on OTU clustering at a 3% distance were compared using STAMP at phylum level. A corrected *P*-value

lower than 0.05 is significant. Correlations of assignment results in each of the three samples were carried out at genus level. The horizontal and vertical axes in each subfigure (b, d and f) indicate the numbers of the corresponding genus sequences. The red lines represent the 95% prediction bands of linear regression.

3.3. Detection and characterization of the potential pathogenic bacteria

During MBR treatment, the structure of microbial community in wastewater changed in response to the environmental selective pressures, and taxa were differently enriched in different samples (Fig. S4 of the Supporting Information). Fig. 3, according to the alignment to the lists of known pathogenic genera [7, 23], shows the eleven genera of potential pathogenic bacteria found in the three samples. It could be noticed that only *Arcobacter* and *Clostridium* were ubiquitous in all the samples. In A1, *Arcobacter* was the most abundant genera, accounting for about 40% of the population. Expect for *Clostridium*, the other potential pathogens, including *Aeromonas*, *Enterobacter*, *Enterococcus* and *Treponema*, were present at very low abundances (0.026~0.031%). The number of sequences assigned into potential pathogenic bacteria was significantly decreased in the activated sludge sample (Table S3 of the Supporting Information). For example, *Arcobacter* were underrepresented in A2, with nearly two orders of magnitude difference in abundance compared to those found in A1. A similar decay was found for *Aeromonas*, *Enterobacter*, *Enterococcus* and *Treponema* at even lower abundances. Instead, a well-known source of infection, *Legionella* [25], became abundant among the potential pathogens. Furthermore, membrane retention induced a selective pressure on bacterial pathogen diversity. Several widely-reported pathogenic bacteria, including *Aeromonas*, *Enterobacter*, *Enterococcus*, and *Pseudomonas* [7, 8, 23, 26], were not found in A3, and the sequences assigned to potential pathogenic genera only accounted for 0.3% of the total population (Table S3). However, a gram-positive genus, *Clostridium*, was found to be the most abundant among the potential pathogenic phylotypes. Since no significant difference was noted by LEfSe analysis at this taxon (Fig. S4), it was possible to infer that *Clostridium* were more resistant to the treatment of MBR.

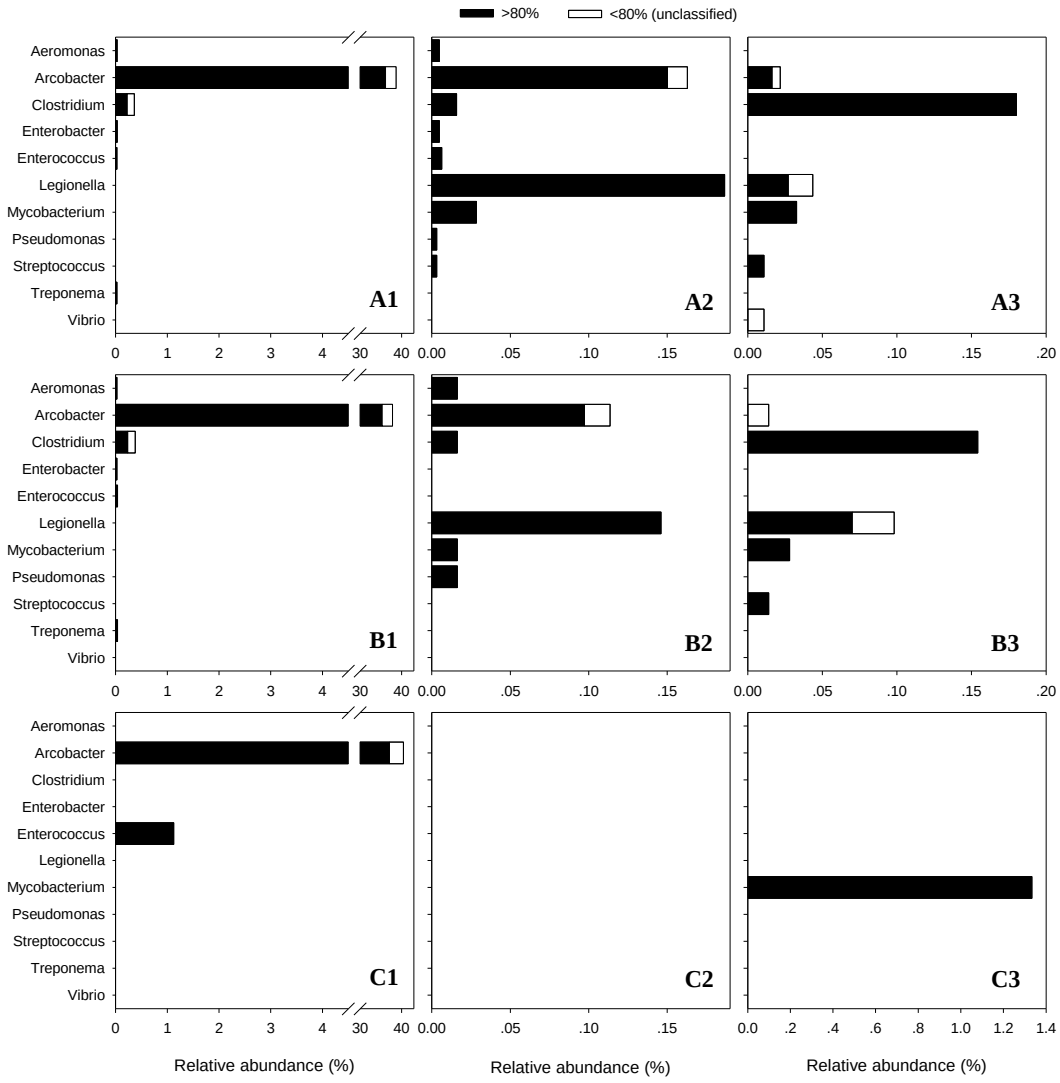


Fig. 3 Relative abundances of potential pathogenic genera in influent wastewater (A1), activated sludge (A2), treated wastewater (A3), subsets with 10000 reads (B1, B2 and B3) and subsets with 100 reads (C1, C2 and C3). Relative abundance is defined as the percentage of a pathogenic genus in total population. The different colors represent the percentages of sequences in the corresponding confidence ranges.

In this study, the depth of pyrosequencing also had a significant influence on detection and characterization of potential pathogenic bacteria. For B1~B3, the dominant pathogenic genera could be well identified, while those with low abundances (e.g., *Enterobacter*, *Enterococcus* and *Vibrio*) were neglected (Fig. 3). It is worthy noting that sequencing failed to reveal the majority of potential pathognic bacteria in environmental samples when the library size was declined to 100. Specifically, no pothenigc bacteria was detected in C2 and *Mycobacterium* were obviously overestimated in C3 (Fig. 3).

Overall, the corresponding sequences got a good alignment with the potential pathogenic genera by using the RDP Classifier, and most had a bootstrap confidence over 80% (Fig. 3 and Table S3). Since species-specific comparison with known pathogenic bacteria could give a more accurate estimation of the potential pathogens in the samples, representative sequences of concerned OTUs assigned into *Arcobacter*, *Clostridium*, *Legionella* and *Mycobacterium* genera (Fig. 3) were retrieved from the datasets. Phylogenetic analysis was then conducted by building a library with representative 16S rRNA gene sequences of pathogenic and non-pathogenic bacterial species. As shown in Fig. 4, 15 OTUs of the total (37 OTUs) had an identity over 95% with neighbor pathogens, including 8 OTUs assigned into *Arcobacter*, 3 OTUs into *Clostridium* and 4 OTUs into *Mycobacterium* (Table S4 of the Supporting Information). In *Arcobacter* genus, the most abundant taxon OTU2964, accounting for 32.9% of the population in A1, showed a low alignment with known pathogens. Without additional information, it could not be concluded whether these sequences referred to nonpathogenic strains, because variants might be also associated with disease but not yet identified [23]. Of particulate importance is that only two strains (OTU2202 and OTU2091) were recognized as potential pathogenic species by species-specific comparison. It could be deduced that phylogenetic analysis at the genus level might lead to an overestimation of the pathogenic bacteria in environmental samples.

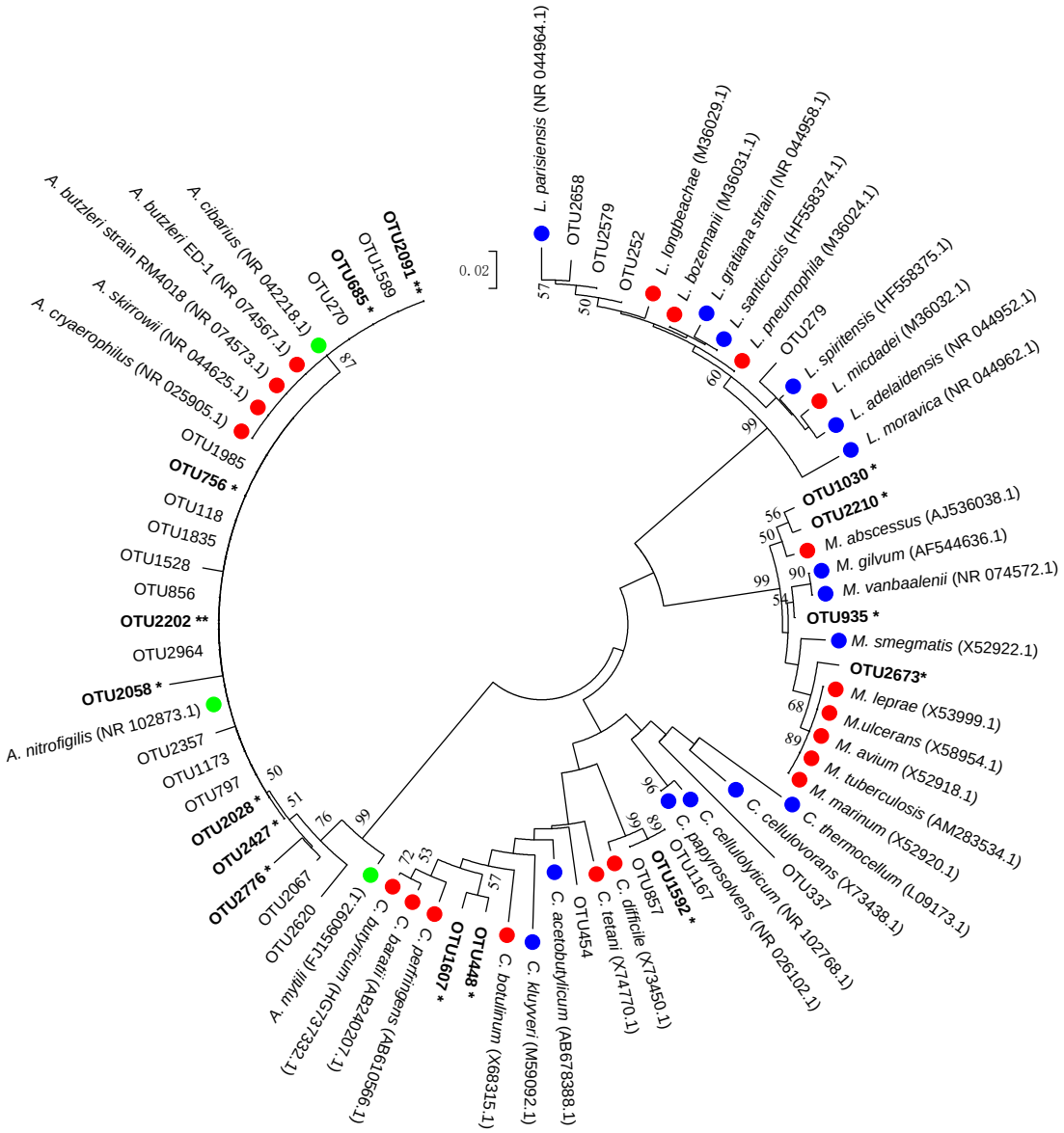


Fig. 4 Phylogenic tree of concerned OTUs from A1~A3 that were assigned into potential pathogenic genera (*Arcobacter*, *Clostridium*, *Legionella* and *Mycobacterium*). Representative sequences from these OTUs were reahieved for alignment and phylogenic analysis. The number of OTU (e.g., OTU448) only indicates the logical order in OTU clustering. OTUs with satisified identities to neighbor pathogens are bolded and marked with * (95~99%) or ** (>99%). Bootstrap values are calculated by 1000 repetitions, and values >50% are given. ● indicate the non-pathogenic species, ● the pathogenic species and ● the vague species.

Since fecal indicator bacteria are still widely used to predict the presence of bacterial, viral and protozoan pathogens [9], the abundances of typical indicators were further evaluated based on the taxonomic results in the present work. Because there is no full taxonomic definition of fecal

indicator yet, bacteria assigned to the families of Enterobacteriaceae and Enterococcaceae were regarded as the representative fecal indicators herein. As shown in Table 2, 41 sequences from A1 were classified, which contributed 0.21% of the dataset. In A2, only 10 sequences got a valid match by RDP Classifier, including 6 sequences assigned into Enterobacteriaceae and 4 sequences into Enterococcaceae. Probably due to the sound separation of 0.20-μm poly(vinylidene fluoride) membranes, no Enterobacteriaceae or Enterococcaceae was detected in the treated wastewater (A3) in this study.

Table 2 Summary of sequences assigned to Enterobacteriaceae and Enterococcaceae [a]

	A1		A2		A3	
	Number of sequences	*r*, %[b]	Number of sequences	*r*, %	Number of sequences	*r*, %
Enterobacteriaceae	36	0.185	6	0.009	0	0
Enterococcaceae	5	0.026	4	0.006	0	0
Total	41	0.211	10	0.015	0	0

a. based on the taxonomic results of RDP Classifier

b. *r* indicates the relative abundance of sequences.

4. Discussion

In this study, a group of predominant potential pathogens, *Arcobacter*, were differently abundant in the influent wastewater compared to other samples (Fig. 3 and Fig. S4). The genus *Arcobacter* belonging to the RNA Superfamily VI of Proteobacteria was proposed in 1991, and the International Commission on Microbiological Specification for Foods has considered *Arcobacter* to be one of the most frequently notified food-borne infectious agents [27]. Full understanding of its occurrence and fate during wastewater reclamation is, thereby, very important, especially for a rapid and accurate diagnosis of the infection source of outbreaks (e.g., acute enteric disease) in public places. In the present work, RDP Classifier indicated that 21 OTUs from A1 were classified into *Arcobacter* genus, which accounted for 36.5% of the total population. The microbial composition might be a typical pathogenic characteristic of restaurant wastewater, because raw and undercooked meat and poultry products have been recognized as the sources of *Arcobacter* [28]. Overall, the results showed that the hybrid MBR system presented a good removal of *Arcobacter*; ~87% of the influent *Arcobacter* were eliminated in the activated sludge and <10 counts/mL were detected in the treated wastewater (Section III of the Supporting Information). Furthermore, phylogenic analysis suggested that the genus-specific comparison could result in an

overestimation of pathogenic bacteria, since more than 90% of *Arcobacter* had highest homology with a free-living nitrogen-fixing bacterium, *A. nitrofigilis*. Notably, 2 strains (OTU2091 and OTU2202) had an identity over 99% with known pathogenic *Arcobacter* species (e.g., *A. butzleri*), which were estimated to be at concentrations of $(1.40 \pm 0.15) \times 10^6$, $(6.81 \pm 0.29) \times 10^6$ and ~1 counts/mL in influent wastewater, activated sludge and treated wastewater, respectively (Section III of the Supporting Information).

In the view of pathology, it is of great concern to focus on the pathogenic bacteria emerging in the aerobic tank of MBR because not only these tolerant microorganisms got a competitive advantage with the biomarkers involved in contaminant degradation (e.g., *Zoogloea* and *Dechloromonas* as shown in Fig. S4) but also aerosols containing pathogens could be generated from the aeration tank and further transported and dispersed by wind. A genus of gram-negative coccobacilli, *Legionella*, was well recognized in the activated sludge sample (A2). It has been reported that *Legionella* prefer to inhabit in man-made aquatic environments where the water temperature is higher than ambient temperature, and that the growth of *Legionella* spp. can be aided by co-existing micro-organisms (e.g., protozoa) [25, 29]. Although this bacterial genus was enriched in activated sludge, *Legionella* found herein seemed non-pathogenic; all the representative sequences had a low alignment (87~92%) with the foremost pathogenic species *L. pneumophila*, *L. longbeachae*, *L. micdadei* and *L. bozemanii*. Fig. 3 indicates that MBR removal of *Legionella* from restaurant wastewater was mainly attributed to membrane retention that could efficiently eliminate the hosts (e.g., amebae) in the treated wastewater. Moreover, a recent study on bacterial pathogen diversity in biosolids (digested sludge) using pyrosequencing has revealed that most of the pathogenic sequences belonged to the genera of *Mycobacterium* and *Clostridium* [23]. In the present work, our results showed that despite low relative abundances, all sequences belonging to *Mycobacterium* genus had more than 95% similarity to a 'freak' pathogenic species, *M. abscessus*. The gene order phylogeny of *M. abscessus* groups the organisms into rapid and slow-growers [30, 31]. *M. abscessus* is closer to the non-pathogens in terms of its growth characteristics and is placed away from the pathogens (Fig. 4), which could lead to taxonomic bias based on 16S rRNA gene pyrosequencing. As a result, virulence assays that target the functional genes are further required to convince the relevant conclusions.

This study also reinvigorates the debate that the indicator bacteria are inefficient in representing the potential pathogenic bacteria from non-fecal sources. The relative abundance of fecal indicator bacteria did not show a good relationship with that of potential pathogenic species, though providing a rough evaluation on the occurrence of potential pathogens in the restaurant wastewater (Table 2 and Table S4). Notably, Enterobacteriaceae and Enterococcaceae were not detected in A3 but 97 of total 36644 sequences were classified into the pathogenic genera, including 25 sequences with identities of > 95% with *M. abscessus*, *C. difficile* and *C. botulinum*. Furthermore, *Clostridium* genus was an important group in A3, which was resistant to the MBR treatment (Fig. 3). In this study, the ambiguously defined taxon contained *Clostridium* cluster sensu stricto (Clostridiaceae 1), *Clostridium* cluster IV (Ruminococcaceae) and *Clostridium* cluster XI (Peptostreptococcaceae). Three strains from *Clostridium* cluster XI and *Clostridium* cluster sensu stricto had a good phylogenetic alignment with *C. difficile* and *C. botulinum*, respectively (Fig. 4 and Table S4). Since *Clostridia* (spores) are highly resistant to chlorination [32], disinfection efficiency could be easily overestimated when referred to the elimination of intolerant indicator bacteria (e.g., Enterobacteriaceae).

In this study, 454 pyrosequencing was introduced for a comprehensive understanding of bacterial pathogens in the restaurant wastewater. Compared to conventional culture-based methods and qPCR assays, this technology is high-throughput for mining potential pathogenic bacteria in environmental samples, which avoids the misestimation of pathogens by using a certain group of indicator bacteria. Molecular biology methods that offer ≈100 tags could only provide rough information on the structure of microbial communities at phylum level (Fig. 2). By contrast, the 10,000-sequence datasets were generally valid in forecasting individuals within microbial communities, but overrepresentation and underrepresentation were still noted regarding the highly-potential bacterial pathogens (Table S4). For accurately reaping the rare strains, exponential growth of the library size (1~2 orders of magnitude) might be unwise since a large number of singletons were generated at a sequencing depth of 30,000~150,000 (Table 1). DNA fragment pretreatment (e.g., the use of multiple genus level PCR primers) should be thereby considered in pathogenic studies. Furthermore, 16S rRNA gene pyrosequencing provided the opportunity to discover the important strains that have not been cultured yet (e.g., OTU2964). Short-gun metagenomic and metatranscriptomic sequencing could be used to predict their

functions. Phylogenic analysis of concerned pathogenic and non-pathogenic bacterial species herein gave a more accurate evaluation of the abundance and diversity of bacterial pathogens. Virulence of relevant communities could be further analyzed using qPCR or microarrays that target the functional genes. Overall, the present work showed that restaurant wastewater was suitable for reclamation using MBR technology. Pathogenic bacteria were efficiently removed in the hybrid systems, and membrane filtration process retained the communities that were resistant to biological treatment. The tolerant bacterial pathogens in treated wastewater revealed by pyrosequencing provide insights into the selection of specific tertiary treatment and proper disinfection methods as well.

5. Conclusions

In the present work, high-throughput pyrosequencing was used to characterize the potential pathogenic bacteria in a full-scale MBR treating restaurant wastewater. The results indicated that the influent pathogenic community might be highly diverse and that 39.2% of the population was assigned into the pathogenic genera. Overall, MBR treatment had a good removal of *Aeromonas*, *Arcobacter*, *Enterobacter*, *Enterococcus* and *Treponema*, and in the treated wastewater the bacterial count of *Arcobacter* was decreased to <10 counts/mL. The most dominant potential pathogens in activated sludge and treated wastewater were affiliated to the genera of *Legionella*, *Clostridium* and *Mycobacterium*. Nevertheless, species-specific comparison showed that only a small portion (0.0~1.6%) of the corresponding sequences had identities of > 99% to the neighbor pathogenic species, suggesting that phylogenetic analysis at the genus level might lead to an overestimation of the potential pathogens. This study provided insights into assessing pathogenic bacteria risk in wastewater purification and reclamation.

Acknowledgments

The work is financially supported by National Natural Science Foundation of China (51422811), Shanghai Rising-Star Program (14QA1403800) and the Shanghai Science & Technology Commission Program (13231202002).

References

1. Chinese catering information center. http://www.canyin.com/a/zixunzhongxin

/xingyexinwen/20120724/991.html. (accessed April 18, 2014).

2. X. Zhu, Z. Wang and Z. Wu, *Process Biochem.* , 2011, **46**, 1001-1009.

3. B. Yang, G. Chen and G. Chen, *Sep. Purif. Technol.* , 2012, **88**, 184-190.

4. J.-x. Kang, L. Lu, W. Zhan, B. Li, D.-S. Li, Y.-Z. Ren and D.-q. Liu, *J. Hazard. Mater.* , 2011, **186**, 849-854.

5. S. Judd, *Trends Biotechnol.* , 2008, **26**, 109-116.

6. M. E. Verbyla, S. M. Oakley and J. R. Mihelcic, *Environ. Sci. Technol.*, 2013, **47**, 3598-3605.

7. L. Ye and T. Zhang, *Environ. Sci. Technol.*, 2011, **45**, 7173-7179.

8. W. Ahmed, H. Brandes, P. Gyawali, J. P. S. Sidhu and S. Toze, *Water Res.* , 2014, **53**, 361-369.

9. O. Savichtcheva and S. Okabe, *Water Res.* , 2006, **40**, 2463-2476.

10. W. Ahmed, S. Sawant, F. Huygens, A. Goonetilleke and T. Gardner, *Water Res.* , 2009, **43**, 4918-4928.

11. K. Zhang and K. Farahbakhsh, *Water Res.* , 2007, **41**, 2816-2824.

12. P. Bourhy, S. Bremont, F. Zinini, C. Giry and M. Picardeau, *J. Clin. Microbiol.* , 2011, **49**, 2154-2160.

13. D.-Y. Lee, H. Lauder, H. Cruwys, P. Falletta and L. A. Beaudette, *Sci. Total Environ.* , 2008, **398**, 203-211.

14. J. Ma, Z. Wang, Y. Yang, X. Mei and Z. Wu, *Water Res.* , 2013, **47**, 859-869.

15. T. Zhang, M. F. Shao and L. Ye, *Isme Journal*, 2012, **6**, 1137-1147.

16. A. S. Laufer, J. P. Metlay, J. F. Gent, K. P. Fennie, Y. Kong and M. M. Pettigrew, *Mbio*, 2011, **2**.

17. Q. Wang, G. M. Garrity, J. M. Tiedje and J. R. Cole, *Appl. Environ. Microbiol.*, 2007, **73**, 5261-5267.

18. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Tumbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld and R. Knight, *Nat. Methods* 2010, **7**, 335-336.

19. R. C. Edgar, *Nat. Methods* 2013, **10**, 996-+.
20. M. Marzorati, L. Wittebolle, N. Boon, D. Daffonchio and W. Verstraete, *Environ. Microbiol.* , 2008, **10**, 1571-1581.
21. D. H. Parks and R. G. Beiko, *Bioinformatics*, 2010, **26**, 715-721.
22. N. Segata, J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett and C. Huttenhower, *Genome Biol.*, 2011, **12**.
23. K. Bibby, E. Viau and J. Peccia, *Water Res.* , 2010, **44**, 4252-4260.
24. K. Tamura, G. Stecher, D. Peterson, A. Filipski and S. Kumar, *Mol. Biol. Evol.*, 2013, **30**, 2725-2729.
25. J. M. Blatny, B. A. P. Reif, G. Skogan, O. Andreassen, E. A. Hoiby, E. Ask, V. Waagen, D. Aanonsen, I. S. Aaberge and D. A. Caugant, *Environ. Sci. Technol.*, 2008, **42**, 7360-7367.
26. H. de Man, M. Bouwknegt, E. van Heijnsbergen, E. J. T. M. Leenen, F. van Knapen and A. M. de Roda Husman, *Water Res.* , 2014, **54**, 254-261.
27. I. Gonzalez, T. Garcia, S. Fernandez and R. Martin, *Food Anal. Methods* 2012, **5**, 956-968.
28. M. Mor-Mur and J. Yuste, *Food Bioprocess Technol.* , 2010, **3**, 24-35.
29. H. Y. Lau and N. J. Ashbolt, *J. Appl. Microbiol.* , 2009, **107**, 368-378.
30. H. Medjahed, J.-L. Gaillard and J.-M. Reyrat, *Trends Microbiol.*, 2010, **18**, 117-123.
31. A. N. Prasanna and S. Mehra, *Plos One*, 2013, **8**.
32. T. Karpova, P. Pekonen, R. Gramstad, U. Ojstedt, S. Laborda, H. Heinonen-Tanski, A. Chavez and B. Jimenez, *Water Sci. Technol.* , 2013, **68**, 2090-2096.

Figure captions

Fig. 1 Similarity analysis of the microbial communities (A1, A2 and A3) based on the clustering results at 3% distance cutoff. The numbers in the black circles represent the number of OTUs that is present in the core OTUs shared by the three samples. The numbers in the blue circles represent the OTUs shared by two samples. The numbers in the red circles represent the unique OTUs observed in only one sample. Percentages listed beside the branches indicate the percentages of reads of each sample assigned into the nearby OTUs groups.

Fig. 2 Pairwise comparison of biological relevance of (a) A1 and C1 at phylum level, (b) A1, B1 and C1 at genus level, (c) A2 and C2 at phylum level, (d) A2, B2 and C2 at genus level, (e) A3 and C3 at phylum and (f) A3, B3 and C3 at genus level. Taxonomic results based on OTU clustering at a 3% distance were compared using STAMP at phylum level. A corrected *P*-value lower than 0.05 is significant. Correlations of assignment results in each of the three samples were carried out at genus level. The horizontal and vertical axes in each subfigure (b, d and f) indicate the numbers of the corresponding genus sequences. The red lines represent the 95% prediction bands of linear regression.

Fig. 3 Relative abundances of potential pathogenic genera in influent wastewater (A1), activated sludge (A2), treated wastewater (A3), subsets with 10000 reads (B1, B2 and B3) and subsets with 100 reads (C1, C2 and C3). Relative abundance is defined as the percentage of a pathogenic genus in total population. The different colors represent the percentages of sequences in the corresponding confidence ranges.

Fig. 4 Phylogenic tree of concerned OTUs from A1~A3 that were assigned into potential pathogenic genera (*Arcobacter*, *Clostridium*, *Legionella* and *Mycobacterium*). Representative sequences from these OTUs were reahieved for alignment and phylogenic analysis. The number of OTU (e.g., OTU448) only indicates the logical order in OTU clustering. OTUs with satisified identities to neighbor pathogens are bolded and marked with * (95~99%) or ** (>99%). Bootstrap values are calculated by 1000 repetitions, and values >50% are given. ● indicate the non-pathogenic species, ● the pathogenic species and ● the vague species.

Table Captions

Table 1 Statistical summary for pyrosequencing and microbial diversity analysis

517 **Table 2** Summary of sequences assigned to Enterobacteriaceae and Enterococcaceae

518