Supporting Information

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

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

Contents

Section I MBR setup

Briefly, the MBR tank was divided into a riser zone and two down-comer zones by two baffle plates, which could enhance the recirculation of mixed liquor and thus increase the cross-flow velocity (CFV) according to the theory of internal-loop-airlift reactor. Air diffuser was placed at the bottom of the riser zone to aerate the mixed liquors and induce a CFV along membrane surfaces. Due to the rapid recirculation of mixed liquors between the riser and down-comer zones, a relatively similar dissolved oxygen level (1~3 mg/L) was maintained in the whole reactor. The membrane-filtered effluent was obtained by suction using a peristaltic pump connected to the modules. The effluent flow rate and trans-membrane pressure (TMP) were monitored by a flowmeter and a pressure gauge, respectively. Intermittent operation of the suction pumps (2 min pause for every 12 min of operation) was employed to mitigate membrane fouling. Chemical cleaning-in-place procedure (0.5% (w/v) NaClO solution, 2 h duration) would be carried out if the TMP reached about 30 kPa during the operation.

Sludge was periodically wasted from the tank to maintain a solid retention time (SRT) of ~30 d. The hydraulic retention time (HRT) of the MBR was adjusted from 4~9 h according to the influent wastewater. The actuation of pumps and meters in the system was controlled through a programmable logic controller (PLC).

Section II Procedures of DNA extraction and PCR amplification

Initially, 500 mL of A1 was filtered through 0.45-μm filter membrane (Supor®-450, Pall Corporation, U.S.). 20 mL of A2 was centrifuged at 6000 rpm for 10 min at 4 °C, and the pellets were recovered through decantation of the supernatant. 2000 mL of A3 was filtered using an ultrafiltration filter with nominal molecular weight cut-off (MWCO) of 50 kDa (Millipore Corporation, MA, U.S.). Extraction of DNA from the microbial cells collected from filter membranes (A1 and A3) and pellets (A2) was then conducted using the E.Z.N.A.® Soil DNA kit (Omega Bio-Tek, Inc., Norcross, GA, U.S.). Afterwards, the quality of DNA fragments was assessed using a 2.0 % (w/v) agarose gel electrophoresis.

Bacterial DNA from A1, A2 and A3 samples was amplified by PCR using the primer set 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 533R (5'-TTACCGCGGCTGCTGGCAC-3') targeting the V1-V3 region of the 16S rRNA gene [1].10-nucleotide barcodes were incorporated

between the 454 adaptor and the fused 27F primer, which allowed sample multiplexing during pyrosequencing in a single GS-FLX run. A 20 μL RCR reaction solution was prepared for each sample, containing 4 μL of 5 × FastPfu Buffer, 2 μL of 2.5 mM dNTPs, 0.4 μL of each primer (5 μM), 10 ng of template DNA and 0.4 μL of FastPfu Polymerase (TransGen AP221-02, Beijing, China). The PCR amplification was conducted in a GeneAmp® 9700 under the following thermocycling steps: initial denaturation at 95 °C for 2 min, followed by 25 cycles at 95°C for 30 s, 55°C for 30 s and 72 °C for 30 s, and a final extension at 72°C for 5 min and at 10 °C until halted by user. To minimize the adverse impact of potential early round errors, PCR amplicon libraries were prepared by combining 3 independent products for each sample [2]. After purification from agarose gels using AxyPrep DNA gel extraction kit (Axygen Biosciences, CA, U.S.) and elution using Tris_HCl, the concentrations of PCR products were measured using PicoGreen® dsDNA quantitation reagent (Life Technologies, NY, U.S.) in a QuantiFluor™-ST system (Promega Corporation, WI, U.S.).

Section III The enumeration method for *Arcobacter*

Quantification of potential pathogens was conducted based on results of pyrosequencing and FCM. The *Arcobacter* counts were determined as follows:

$$\text{The } \textit{Arcobacter} \text{ counts} = \text{the total bacterial counts} \times r$$

Where *r,* i.e. relative abundance, is defined as the number of sequences affiliated with that taxon divided by the total number of sequences per sample.

The *Arcobacter* counts in influent wastewater (A1):

$$(2.31 \pm 0.24) \times 10^8 \times 36.14\% = (8.35 \pm 0.87) \times 10^7 \text{ counts/mL}$$

The *Arcobacter* counts in activated sludge (A2):

$$(7.06 \pm 0.30) \times 10^9 \times 0.16\% = (1.15 \pm 0.05) \times 10^7 \text{ counts/mL}$$

The *Arcobacter* counts in treated wastewater (A3):

$$(3.35 \pm 0.82) \times 10^4 \times 0.02\% = <10 \text{ counts/mL}$$

The OTU2091 and OTU2202 counts in influent wastewater (A1):

$$(2.31 \pm 0.24) \times 10^8 \times 118/19411 = (1.40 \pm 0.15) \times 10^6 \text{ counts/mL}$$

The OTU2091 and OTU2202 counts in activated sludge (A2):

$$(7.06 \pm 0.30) \times 10^9 \times 61/63243 = (6.81 \pm 0.29) \times 10^6 \text{ counts/mL}$$

The OTU2091 and OTU2202 counts in treated wastewater (A3):

$$(3.35 \pm 0.82) \times 10^4 \times 1/36644 = \sim 1 \text{ counts/mL}$$

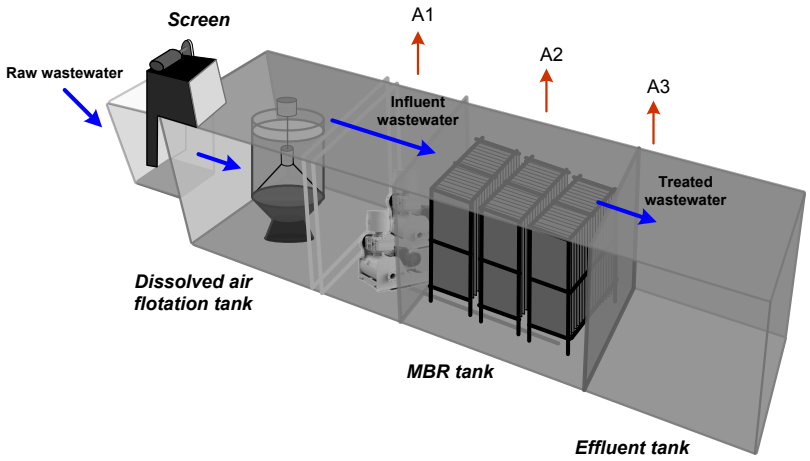

Fig. S1 Flow diagram of the full-scale MBR.

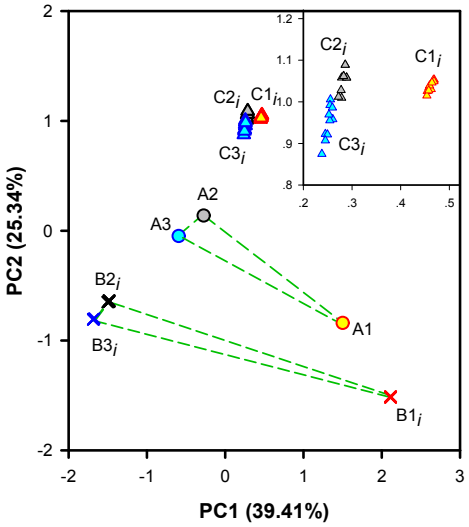

Fig. S2 PCoA of the maternal datasets (A1~A3) and the subsets (B1$_i$~B3$_i$ and C1$_i$ ~C3$_i$).

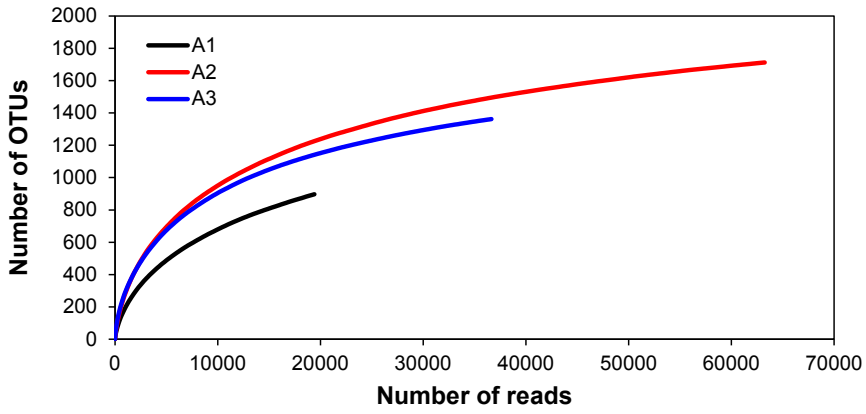

Fig. S3 Rarefaction curves of A1, A2 and A3.

Phylogenetic complexity of bacterial communities

LEfSe analysis was used to identify the predominant taxa that represented the differences. In this study, 68 bacterial clades showed statistically significant and biologically consistent differences, and 42 clades with linear discriminant analysis (LDA) score higher than 1% of the dataset size were then retained. Specifically, the most differently abundant genera in influent wastewater belong to the orders: Neisseriales, Desulfuromonadales, Campylobacterales and Bacteroidales, including environmental organisms from Prevotellaceae and Porphyromonadaceae clades. In the activated sludge sample, Betaproteobacteria were notably enriched, with a relative abundance higher than 45%. The overrepresented genera, including *Zoogloea*, *Dechloromonas* and *Aquabacterium*, are prevalent in activated sludge samples and believed to play an important role in wastewater treatment [1, 3]. As shown in Fig. S4, the structure of microbial community also varied due to membrane retention, and bacteria assigned into Nitrospira, Phycisphaerae, and Alphaproteobacteria classes became differently abundant in treated wastewater. Also of note is that the microorganisms from these overrepresented genera (e.g., *Nitrospira*, *Phycisphaera* and *Bradyrhizobium*) are always considered versatile in nitrogen metabolism [4-6]. Restaurant wastewater is always characterized by high carbon to nitrogen ratio (e.g., high COD/N ratio), and our previous study showed that nitrifiers could be outcompeted by the heterotrophs under such a copiotrophic environment [7]. However, it seemed that these dominant bacteria of activated sludge

could not easily pass through membranes, and consequently some successors (e.g., *Nitrospira*) were enriched in the oligotrophic treated wastewater (Fig. S4).

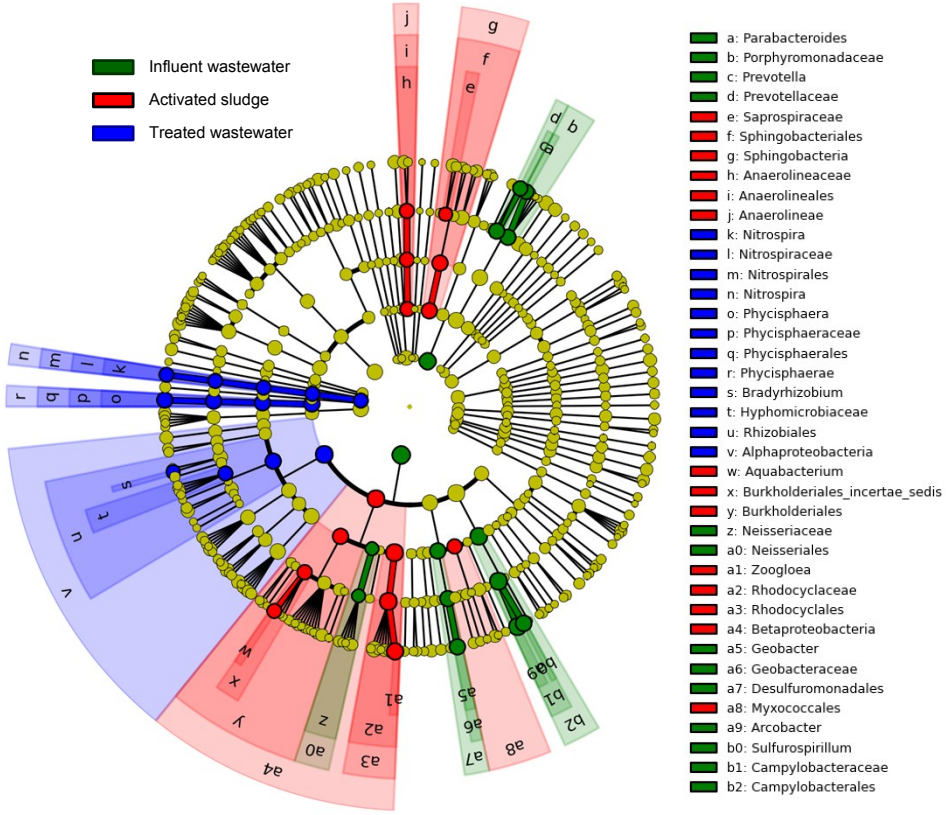

Fig. S4 Taxonomic representation of statistically and biologically consistent differences between influent wastewater, activated sludge and treated wastewater samples. Differences are represented in the color of the most abundant biomarkers (green indicating influent wastewater, red activated sludge, blue treated wastewater, and yellow non-significant). Each circle's diameter is proportional to the relative abundance of taxa.

Table S1 Characteristics of the influent and treated wastewater (unit: mg/L).

Item	COD	TN	NH_3-N	TP	SS	Oil
Influent wastewater	1020~2490	20.3~43.4	9.3~26.0	8.6~17.0	125~493	13.4~20.0
Treated wastewater	44~170	3.8~5.1	0.4~2.0	0.5~3.4	n.d.a	0.1~0.2

a. n.d. indicates the value is not detetable.

Table S2 Summary of the pathogenic and non-pathogenic species.

	Genus	Species	Accession number
Pathogenic species	*Arcobacter*	*A.cryaerophilus* strain_A_169/B	NR_025905.1
		A.skirrowii	NR_044625.1
		A.butzleri strain_RM4018	NR_074573.1
		A.butzleri ED-1	NR_074567.1
	Clostridium	*C.botulinum* type_C	X68315.1
		C.tetani	X74770.1
		C.perfringens	AB610566.1
		C.baratii strain:_T8	AB240207.1
		C.butyricum strain_NEC8	HG737332.1
		C.difficile strain_DSM_11209	X73450.1
	Legionella	*L.pneumophila*	M36024.1
		L.micdadei	M36032.1
		L.longbeachae	M36029.1
		L.bozemanii	M36031.1
	Mycobacterium	*M.abscessus*	AJ536038.1
		M.leprae	X53999.1
		M.ulcerans	X58954.1
		M.avium	X52918.1
		M.tuberculosis isolate_TB36	AM283534.1
		M.marinum	X52920.1
Non-pathogenic species	*Clostridium*	*C.acetobutylicum* strain:_JCM_8021	AB678388.1
		C.thermocellum DSM_1237	L09173.1
		C.cellulovorans strain_DSM_3052	X73438.1
		C.kluyveri	M59092.1
		C.ellulolyticum strain_H10	NR_102768.1
		C.papyrosolvens DSM_2782	NR_026102.1
	Legionella	*L.adelaidensis* strain_NCTC_12735	NR_044952.1

		L.gratiana strain_NCTC_12388	NR_044958.1
		L.moravica strain_NCTC_12239	NR_044962.1
		L.parisiensis strain_NCTC_11983	NR_044964.1
		L.santicrucis strain_SC-63-C7	HF558374.1
		L.spiritensis strain_Bibb_HSH-9	HF558375.1
	Mycobacterium	*M.smegmatis*	X52922.1
		M.gilvum isolate_VM0442	AF544636.1
		M.vanbaalenii strain_PYR-1	NR_074572.1
Vague species[a]	*Arcobacter*	*A.cibarius* strain_LMG_21996	NR_042218.1
		A.mytili strain_T234	FJ156092.1
		A.nitrofigilis strain_DSM_7299	NR_102873.1

a. Vague species indicates the unclear species that is pathogenic or non-pathogenic.

Table S3 Summary of the confidence of corresponding sequences in A1, A2 and A3.

Confidence to potential pathogenic genus	A1 Number of sequences	*r*, %[a]	A2 Number of sequences	*r*, %	A3 Number of sequences	*r*, %
>80%	7077	36.46	225	0.40	98	0.27
<80% (unclassified)	516	2.66	8	0.01	12	0.03

r indicates the relative abundance of sequences in the corresponding confidence range.

Table S4 Alignment of OTUs to the neighbor pathogenic or non-pathogenic species with highest identity.

	Number of sequences									Neighbor known species with highest identity	
	A1	B1	C1	A2	B2	C2	A3	B3	C3	Species	Identity
OTU118	56	20	0	2	0	0	0	0	0	*A. nitrofigilis* (NR	95%

										102873.1)	
OTU270	9	7	1	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	92%
OTU685	2	1	0	0	0	0	0	0	0	*A. cryaerophilus* (NR 025905.1)	96%
OTU756	36	12	0	0	0	0	0	0	0	*A. butzleri* strain RM4018 (NR 074573.1)	98%
OTU797	2	2	0	0	0	0	0	0	0	*A. mytili* (FJ156092.1)	89%
OTU856	4	1	0	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	94%
OTU1173	29	15	0	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	91%
OTU1528	2	1	1	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	92%
OTU1589	4	2	0	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	90%
OTU1835	157	77	0	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	91%
OTU1985	10	5	0	0	0	0	0	0	0	*A. mytili* (FJ156092.1)	91%
OTU2028	5	2	0	0	0	0	0	0	0	*A. butzleri* strain RM4018 (NR 074573.1)	96%
OTU2058	2	1	0	0	0	0	0	0	0	*A. butzleri* strain RM4018 (NR 074573.1)	97%
OTU2067	138	54	1	2	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	95%
OTU2091	94	48	1	1	0	0	0	0	0	*A. cryaerophilus* (NR 025905.1)	99%
OTU2202	24	13	0	0	0	0	0	0	0	*A. butzleri* strain RM4018 (NR 074573.1)	99%

OTU2357	2	1	0	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	93%
OTU2427	17	8	0	0	0	0	0	0	0	*A. butzleri* strain RM4018 (NR 074573.1)	98%
OTU2620	9	3	0	0	0	0	0	0	0	*A. nitrofigilis* (NR 102873.1)	96%
OTU2776	13	7	0	0	0	0	0	0	0	*A. butzleri strain* RM4018 (NR 074573.1)	95%
OTU2964	6394	2596	29	90	6	0	6	0	0	*A. nitrofigilis* (NR 102873.1)	94%
OTU337	3	1	0	0	0	0	0	0	0	*C. cellulolyticum* (NR 102768.1)	85%
OTU448	1	0	0	1	0	0	11	3	0	*C. botulinum* (X68315.1)	95%
OTU454	0	0	0	0	0	0	6	1	0	*C. botulinum* (X68315.1)	93%
OTU857	2	1	0	0	0	0	4	0	0	*C. difficile* (X73450.1)	94%
OTU1167	20	9	0	5	1	0	44	6	0	*C. difficile* (X73450.1)	94%
OTU1592	0	0	0	1	0	0	1	1	0	*C. difficile* (X73450.1)	95%
OTU1607	12	7	0	3	0	0	0	0	0	*C. botulinum* (X68315.1)	98%
OTU252	0	0	0	5	0	0	1	0	0	*L. parisiensis* (NR 044964.1)	96%
OTU279	0	0	0	79	8	0	7	4	0	*L. adelaidensis* (NR 044952.1)	93%
OTU2579	0	0	0	30	0	0	0	0	0	*L. parisiensis* (NR 044964.1)	96%
OTU2658	0	0	0	4	1	0	1	0	0	*L. parisiensis* (NR 044964.1)	95%
OTU935	0	0	0	7	0	0	1	0	0	*M. abscessus* (AJ536038.1)	96%
OTU1030	0	0	0	10	1	0	1	1	1	*M. abscessus*	95%

										(AJ536038.1)	
OTU2210	0	0	0	0	0	0	3	0	0	*M. abscessus* (AJ536038.1)	95%
OTU2673	0	0	0	1	0	0	7	1	0	*M. abscessus* (AJ536038.1)	95%

References

1. J. Ma, Z. Wang, Y. Yang, X. Mei and Z. Wu, *Water Res.* , 2013, **47**, 859-869.
2. L. Ye and T. Zhang, *Environ. Sci. Technol.*, 2011, **45**, 7173-7179.
3. T. Zhang, M. F. Shao and L. Ye, *Isme Journal*, 2012, **6**, 1137-1147.
4. Y. Fukunaga, M. Kurahashi, Y. Sakiyama, M. Ohuchi, A. Yokota and S. Harayama, *J. Gen. Appl. Microbiol.* , 2009, **55**, 267-275.
5. L. Ye, T. Zhang, T. Wang and Z. Fang, *Environ. Sci. Technol.*, 2012, **46**, 13244-13252.
6. N. Delmotte, C. H. Ahrens, C. Knief, E. Qeli, M. Koch, H. M. Fischer, J. A. Vorholt, H. Hennecke and G. Pessi, *Proteomics*, 2010, **10**, 1391-1400.
7. J. Ma, Z. Wang, C. Zhu, S. Liu, Q. Wang and Z. Wu, *PLoS ONE*, 2013, **8**, e63059.