Supplementary Materials for

Taxonomic and functional variations in the microbial community during the upgrade process of a full-scale landfill leachate treatment plant – from conventional nitrification-denitrification to partial nitrification-denitrification

Binbin Sheng [a, b, d], Depeng Wang [a, b, *] Xianrong Liu [c], Guangxing Yang [c],

Wu Zeng [c], Yiqing Yang [c], Fangang Meng [a, b, *]

[a] School of Environmental Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China

[b] Guangdong Provincial Key Laboratory of Environmental Pollution Control and Remediation Technology (Sun Yat-sen University), Guangzhou 510275, China

[c] Guangzhou Environmental Protection Investment Group Co., Guangzhou 510330, China

[d] School of Life Sciences and Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou 510006, China

[*] **Corresponding author**

Depeng Wang, Ph. D., Email: wangdp6@mail.sysu.edu.cn

Fangang Meng, Ph. D., Email: mengfg@mail.sysu.edu.cn

Contents:

Method S2 The workflow of metagenomic analysis.

Table S1 Landfill leachate characteristics and the treatment performance of the LLTP biological process in the whole period of process transformation.

Table S2 Detailed information of raw and clean reads in each sample through metagenomic sequencing.

Table S3 Detailed information of the assembling scaffolds in each sample.

Figure S1 Alpha diversity indices of the bacterial communities from the two-stage A/O and MBR system.

Figure S2 The unique and shared OTUs among three phases illustrated by Venn diagram.

Figure S3 Welch's t-test with two-sided revealed the **s**ignificant differences of microbial community composition between CND and Trans phases (a) and between Trans and PND phases (b).

Figure S4 Metagenomic KEGG annotated analyses showed the abundance of denitrification functional genes carried by different bacteria.

Figure S5 Metagenomic KEGG annotated analyses showed the abundance of nitrogen assimilatory and fixation functional genes carried by different bacteria.

Figure S6 The abundance of metabolic pathways detected at the second level through KEGG annotation.

Figure S7 The abundance of carbohydrate metabolism active enzymes detected through CAZy annotation. (AAs: Auxiliary Activities; CBMs: Carbohydrate-Binding Modules; CEs: Carbohydrate Esterases; GHs: Glycoside Hydrolases; GTs: Glycosyl Transferases; PLs: Polysaccharide Lyases)

Method S1 The workflow of 16S rRNA amplicon sequencing data analysis.

Raw 16S rRNA amplicon sequencing data were firstly trimmed with Trimmomatic software (Bolger et al., 2014) and then processed and analyzed following the pipelines of MOTHUR and QIIME pipeline (Schloss et al., 2009; Caporaso et al., 2010). Chimeric sequences were removed as described by Luo et al (Luo et al., 2019) with UCHIME software (Edgar et al., 2011). In total, 1,238,047 high-quality sequences (25,155-38,957 per sample) were generated after quality control. As the minimum sequencing depth, 25155 clean sequences were randomly extracted from each sample for normalization. The resulting sequences were clustered into operational taxonomic units (OTUs) at the 97% identity threshold (Edgar, 2010). The taxonomic classifications of the representative sequences were conducted using the Ribosomal Database Project (RDP) Classifier via Silva SSU database (Release 123) with a confidence threshold of 80% (Wang et al., 2007). In addition, the relative abundance of bacteria at phylum and genus levels were identified by calculating the ratio of the number of the assigned reads versus the total number of the sequencing reads.

Method S2 The workflow of metagenomic analysis.

Raw metagenomic sequencing reads obtained in CND and PND phases were filtered using Trimmomatic software to remove low quality (score < 20) and ambiguous N bases (Bolger et al., 2014). The clean reads of each sample were separately assembled using the *de novo* algorithm of MEGAHIT software (Li et al., 2015) to obtain scaffolds, and the scaffolds with a minimum length of 500 bp were used for further analysis. Subsequently, the clean reads were mapped to scaffolds with bowtie2 software (Langmead and Salzberg, 2012), and un-using reads in mapping process of each sample were co-assembled to enhance the detection of low abundance of microorganisms. All

of the above scaffolds were performed for protein prediction using MetaGeneMark (Zhu et al., 2010), and the predicted open reading frames (ORFs) were then removed duplicates using CD-HIT (identity of 95% and coverage of 90%) (Fu et al., 2012), producing 2166049 nonredundant gene catalogues. The taxonomic classification of gene catalogues was aligned with NCBI-NR database (E value, 1E-5) and then annotated using MEGAN (Huson et al., 2007). The coverage of each nonredundant gene catalogue was calculated through BBmap scripts and the coverage value was used as the abundance of each nonredundant gene catalogue. Meanwhile, functional annotations were separately performed using an optimized Blast algorithm (E-value, 1E-5) with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and Carbohydrate-Active Enzymes (CAZy) database. Abundance of Each metabolic pathway was based on the sum of abundance of all nonredundant gene catalogue affiliated this metabolic pathway.

Table S1 Landfill leachate characteristics and the treatment performance of the LLTP biological process in the whole period of process transformation. [a]

Full-name of environmental factors		MLSS [b]	MLVSS [b]	SV [b]	SVI [b]	Temperature[b]	DO [c]	F/M	Influent COD	Influent BOD	Influent TN	Influent NH_4^+-N	Influent NO_3^--N	Influent TP	Influent pH	Influent Alkalinity
Short-name of environmental factors	Sampling time	MLSS	MLVSS	SV	SVI	T	DO	F/M	Inf.COD	Inf.BOD	Inf.TN	Inf.NH_4^+	Inf.NO_3^-	Inf.TP	Inf.pH	Inf.Alkali.
Sample	(D/M/Y)	g/L	(g/L)		(mL/g)	(°C)	(mg/L)		(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)		
CND.Sep	13/9/2016	26.25	17.30	22	83.80	34.3	2.10	11.79	6480	2810	3883.90	3563.90	240.20	36.10	8.320	21100
CND.Sep II	28/9/2016	26.97	17.60	34	126.10	34.4	2.86	11.31	4800	2510	5256.40	3959.90	94.80	41.30	8.150	20850
CND.Oct.	17/10/2016	29.94	19.00	33	122.40	34.3	2.18	9.62	8800	2420	5872.20	4071.30	77.70	40.70	8.210	20500
CND.Nov.	16/11/2016	21.62	13.10	13	60.10	32.8	1.85	7.59	13600	1820	4633.90	3959.90	118.00	40.00	8.310	20800
Trans.Dec.	15/12/2016	26.48	17.40	32	120.90	32.4	3.43	46.78	27480	10010	5514.50	4813.80	221.80	43.40	8.349	20600
Trans.Jan.	16/1/2017	26.31	17.12	32	121.63	34.1	0.23	41.39	16920	9010	3893.44	3205.06	89.71	22.70	7.963	15700
Trans.Feb.	17/2/2017	31.01	19.97	25	80.61	35.0	0.06	18.37	8320	5010	4213.65	2648.20	197.30	32.14	7.980	17500
Trans.Mar.	16/3/2017	27.59	18.29	19	68.88	34.4	0.14	20.47	15880	5210	3653.29	3365.93	197.30	21.29	6.755	4300
PND.Apr.	17/4/2017	27.97	19.22	18	64.36	34.8	0.14	16.88	12720	4020	5634.58	4541.53	129.35	38.16	8.148	23050
PND.May.	17/5/2017	24.91	16.65	23	92.32	33.5	0.34	32.64	14960	10010	3253.03	3001.00	121.80	36.76	7.924	21350
PND.Jun.	14/6/2017	27.24	17.76	16	58.73	34.3	0.40	23.72	8480	5020	4153.61	3878.62	134.82	39.17	8.064	52350
PND.Jul.	14/7/2017	28.95	17.41	12	41.46	35.2	0.11	39.07	26320	10010	3393.12	1873.54	165.71	31.74	7.526	13950
PND.Aug.	15/8/2017	30.06	17.55	12	39.93	35.3	0.66	11.67	9600	4000	1351.79	860.66	47.49	18.48	8.051	3800

Full-name of environmental	Influent Glucose	Effluent COD	Effluent BOD	Effluent TN	Effluent NH_4^+-N	Effluent NO_2^--N	Effluent NO_3^--N	Effluent TP	Effluent pH	Effluent Alkalinity	Removal Efficiency	Removal Efficiency	Removal Efficiency	Removal Efficiency	Removal Efficiency	Removal Efficiency

factors											of COD	of BOD	of TN	of NH_4^+-N	of TP	of Alkalinity
Short-name of environmental factors	Inf.Glu.	Eff.COD	Eff.BOD	Eff.TN	Eff.NH_4^+	Eff.NO_2^-	Eff.NO_3^-	Eff.TP	Eff.pH	Eff.Alkali.	RE.COD	RE.BOD	RE.TN	RE.NH_4^+	RE.TP	RE.Alkali.
Sample	t/month	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)			%	%	%	%	%	%
CND.Sep	44.22	467	110	641.26	17.96	8.66	727.29	8.49	7.205	1877	92.80	96.09	83.49	99.50	76.51	91.11
CND.Sep II	44.22	416	110	78.71	5.82	2.96	70.86	2.43	7.750	2595	91.33	95.62	98.50	99.85	94.11	87.55
CND.Oct.	0.00	1016	20	117.42	16.58	0.46	29.91	2.23	7.850	2630	88.45	99.17	98.00	99.59	94.51	87.17
CND.Nov.	12.85	848	120	321.30	11.01	7.57	263.36	1.61	7.875	1795	93.76	93.41	93.07	99.72	95.97	91.37
Trans.Dec.	21.40	1688	0	649.51	10.39	1.76	591.79	7.03	7.122	750	93.86	100.00	88.22	99.78	83.79	96.36
Trans.Jan.	10.10	1400	0	305.29	8.66	1.30	293.56	3.02	7.725	1335	91.73	100.00	92.16	99.73	86.71	91.50
Trans.Feb.	19.78	1056	0	361.33	124.37	78.55	42.53	5.02	8.040	3170	87.31	100.00	91.42	95.30	84.37	81.89
Trans.Mar.	0.59	864	0	45.12	26.36	0.11	34.98	2.61	7.966	1280	94.56	100.00	98.76	99.22	87.72	70.23
PND.Apr.	0.00	1416	0	113.17	20.42	8.20	78.39	2.82	7.874	2840	88.87	100.00	97.99	99.55	92.62	87.68
PND.May.	0.00	872	0	85.15	22.03	0.14	36.86	4.22	7.909	2710	94.17	100.00	97.38	99.27	88.52	87.31
PND.Jun.	0.00	888	0	149.19	16.04	32.97	63.50	8.04	7.851	6645	89.53	100.00	96.17	99.59	79.48	87.31
PND.Jul.	0.00	2624	110	63.13	1.91	3.59	109.65	2.41	8.005	3165	90.03	98.90	98.14	99.90	92.40	77.31
PND.Aug.	0.00	920	100	129.18	1.92	4.53	78.19	3.22	7.918	1300	90.42	97.50	90.44	99.78	82.59	65.79

a. Effluent and removal efficiency here are refer to the biological treatment system (i.e., MBR).

b.The characteristics of the activated sludge from the post nitrification tank.

c.The DO concentration in the prenitrification tank which was the main regulatory parameter for the process transformation.

Table S2 Detailed information of raw and clean reads in each sample through metagenomic sequencing.

Sample	RawReads	RawBases	CleanReads	CleanBases	Effective (%)
CND.Nov	54082980	8112447000	46857154	6853096622	84.48
CND.Oct	51172850	7675927500	44281918	6493639956	84.6
CND.Sep	45847070	6877060500	40448578	5899099543	85.78
PND.Aug	52448674	6877060500	47802530	7089281471	90.11
PND.Jul	53296528	7994479200	47488248	6940839104	86.82
PND.Jun	60437962	9065694300	53710730	7861788675	86.72

Table S3 Detailed information of the assembling scaffolds in each sample.

Sample	Total number	Total len (bp)	Average len (bp)	Max len (bp)	N50 len (bp)	Assembly Rate（%）
CND.Nov	255800	366037020	1430.95	251320	1852	12.68%
CND.Oct	249537	362470365	1452.57	463632	1913	12.79%
CND.Sep	218373	304443464	1394.14	412044	1742	12.73%
PND.Aug	258776	381590262	1474.6	500479	1890	13.08%
PND.Jul	274044	400204293	1460.37	329223	1933	12.85%
PND.Jun	305370	434052554	1421.4	341440	1783	12.80%
unmap_samples	794274	708214127	891.65	14922	891	12.68%

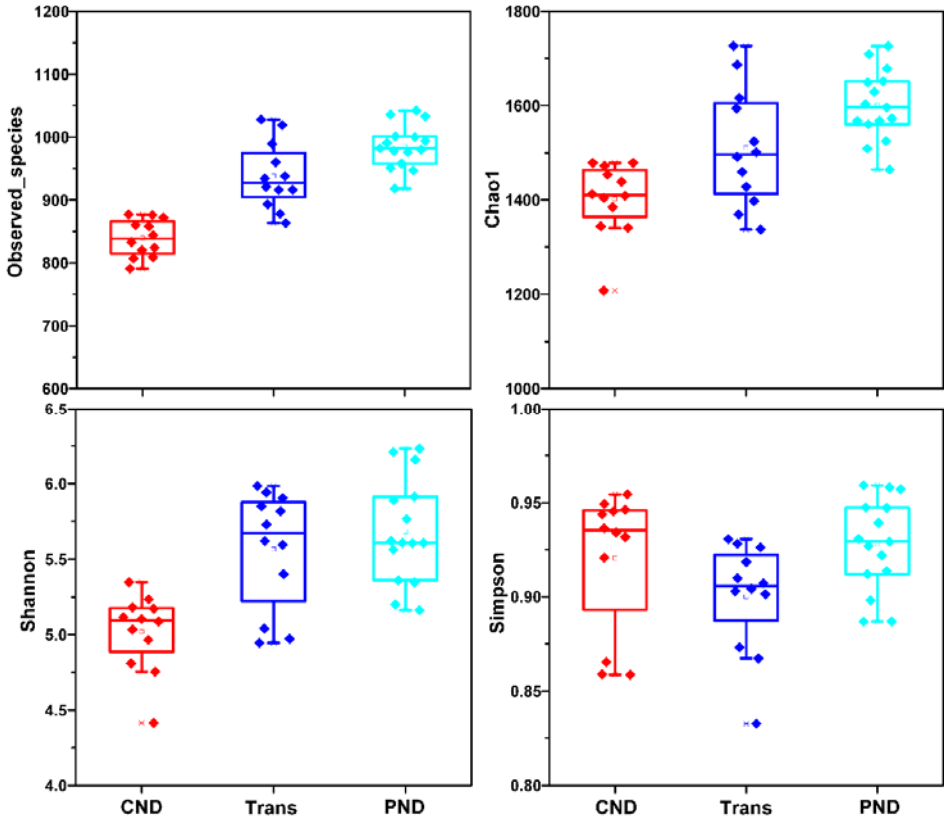

Figure S1 Alpha diversity indices of the bacterial communities from the two-stage A/O and MBR system.

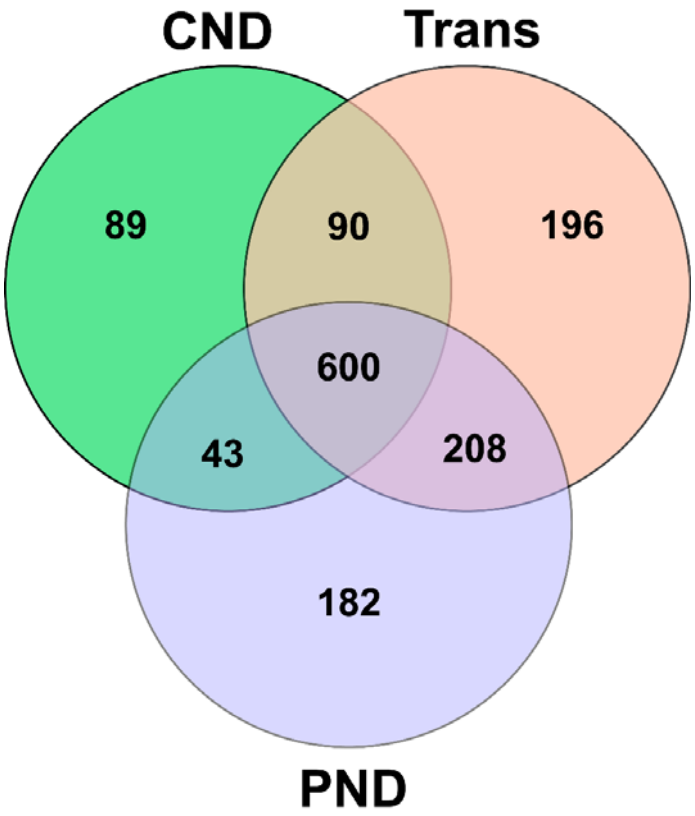

Figure S2 The unique and shared OTUs among three phases illustrated by Venn diagram.

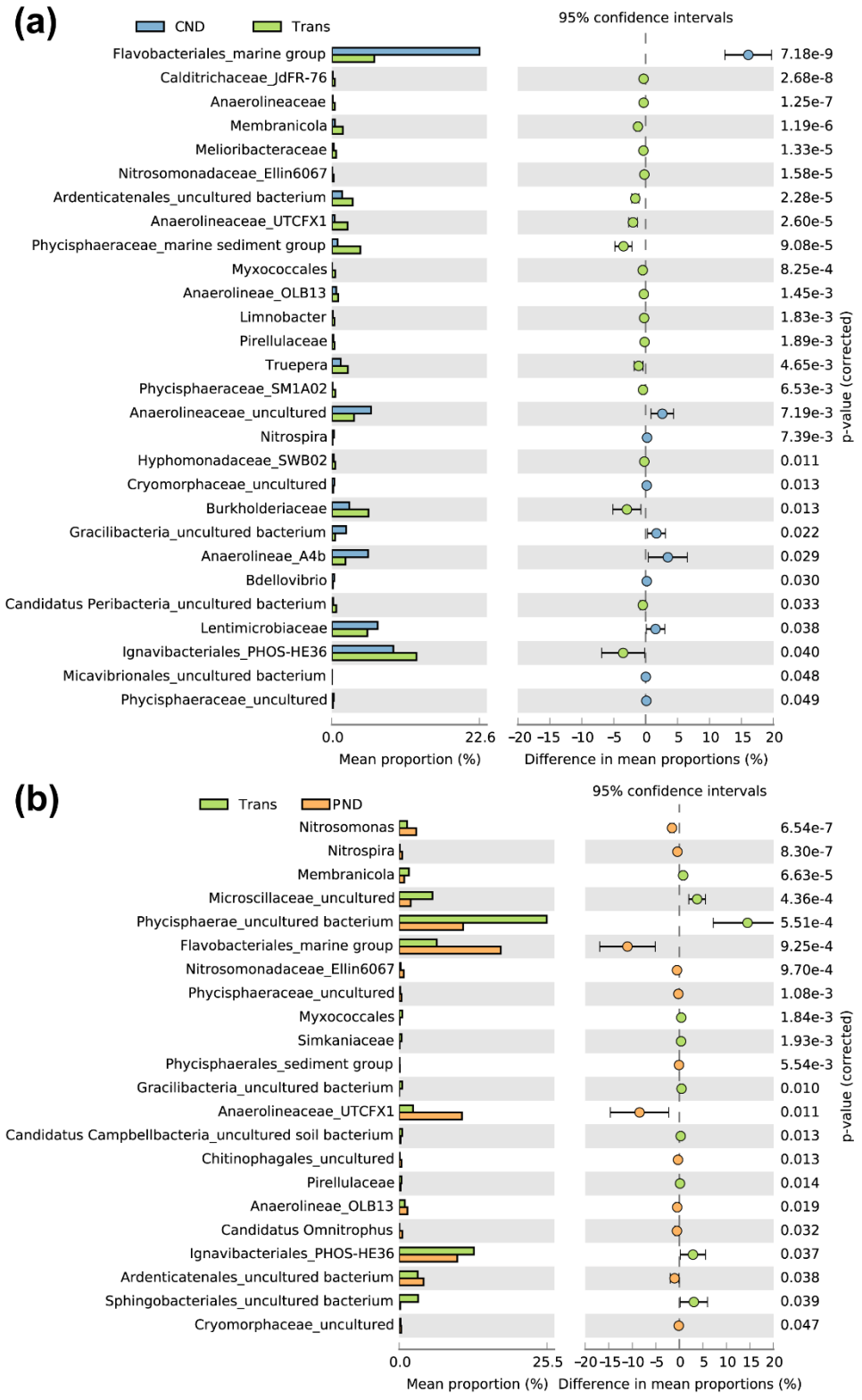

Figure S3 Welch's t-test with two-sided revealed the significant differences of microbial community composition between CND and Trans phases (a) and between Trans and PND phases (b).

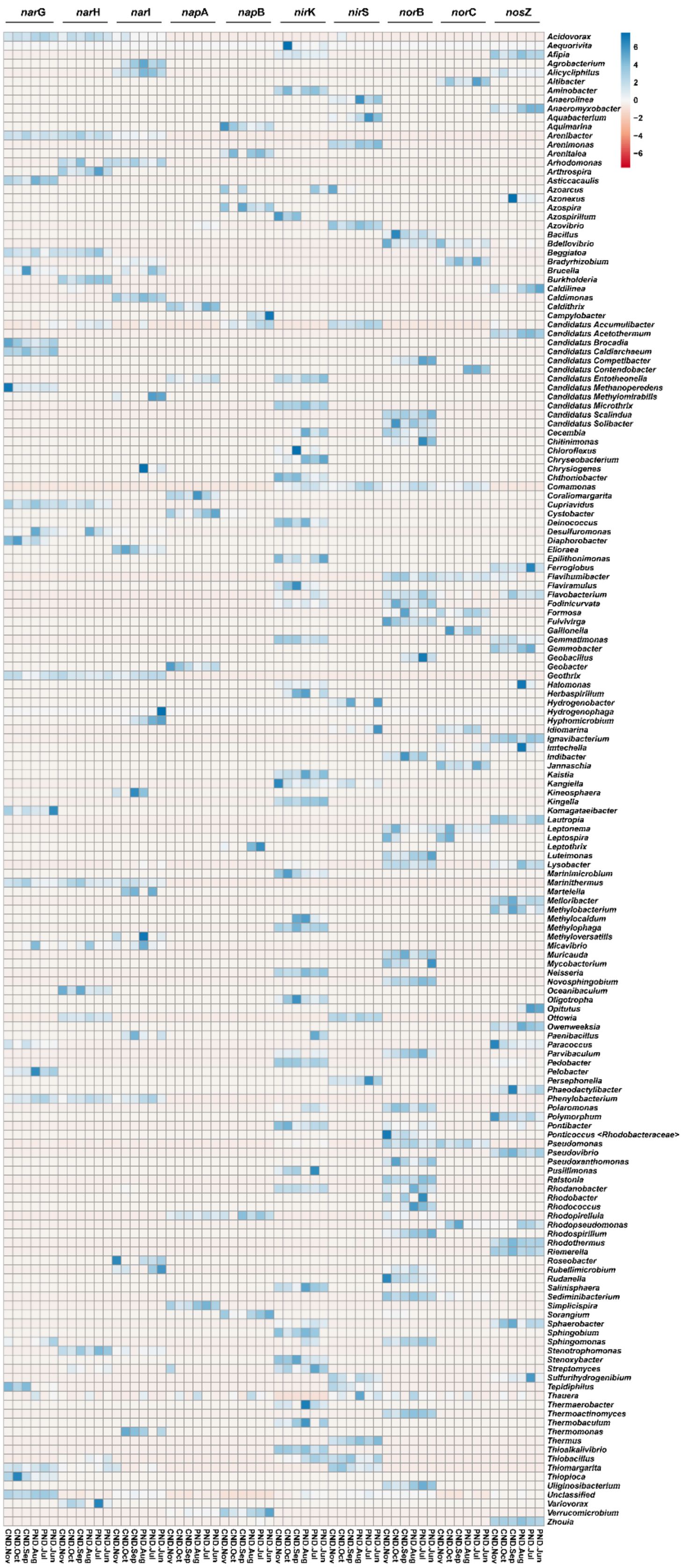

Figure S4 Metagenomic KEGG annotated analyses showed the abundance of denitrification functional genes carried by different bacteria.

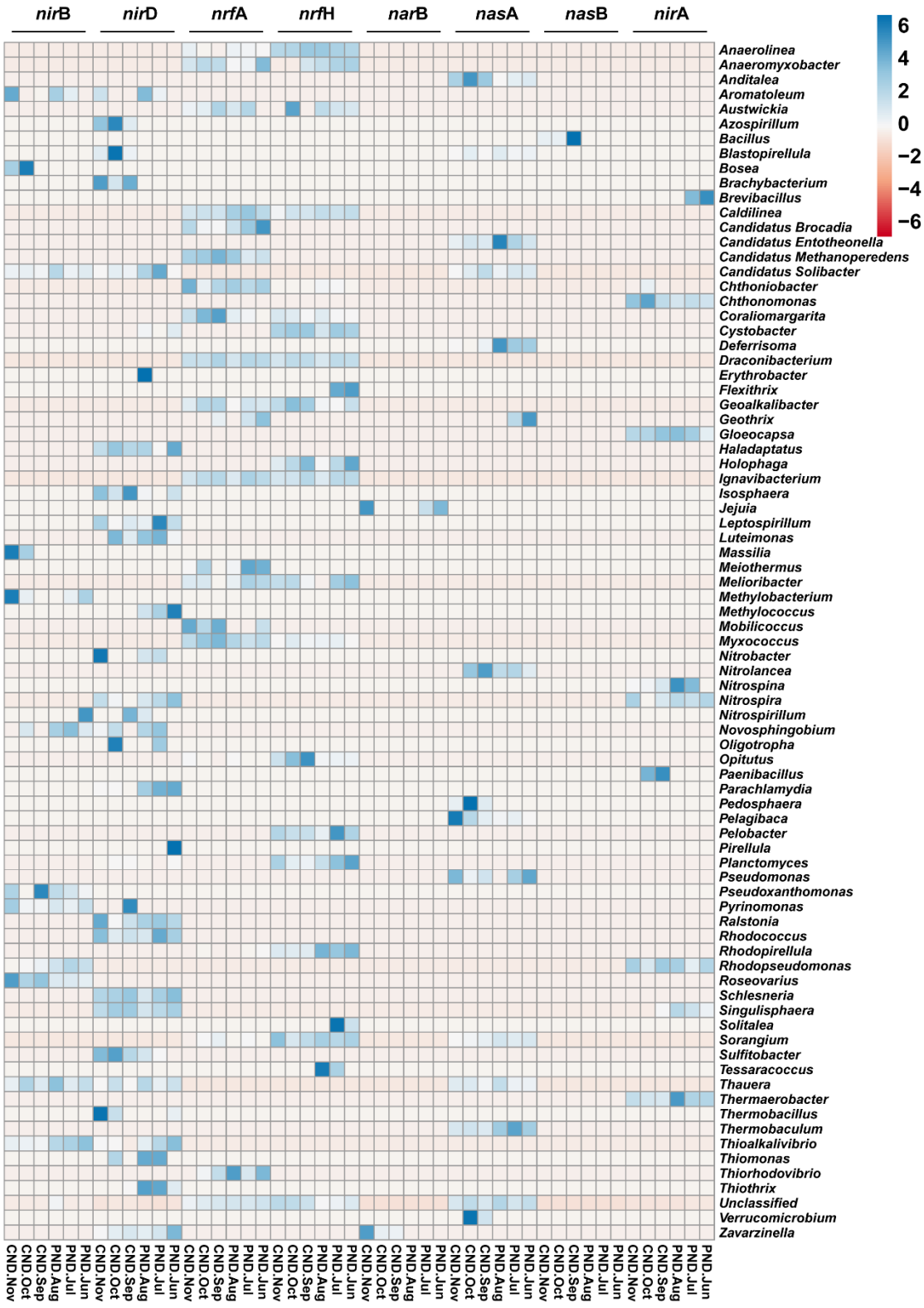

Figure S5 Metagenomic KEGG annotated analyses showed the abundance of nitrogen assimilatory and fixation functional genes carried by different bacteria.

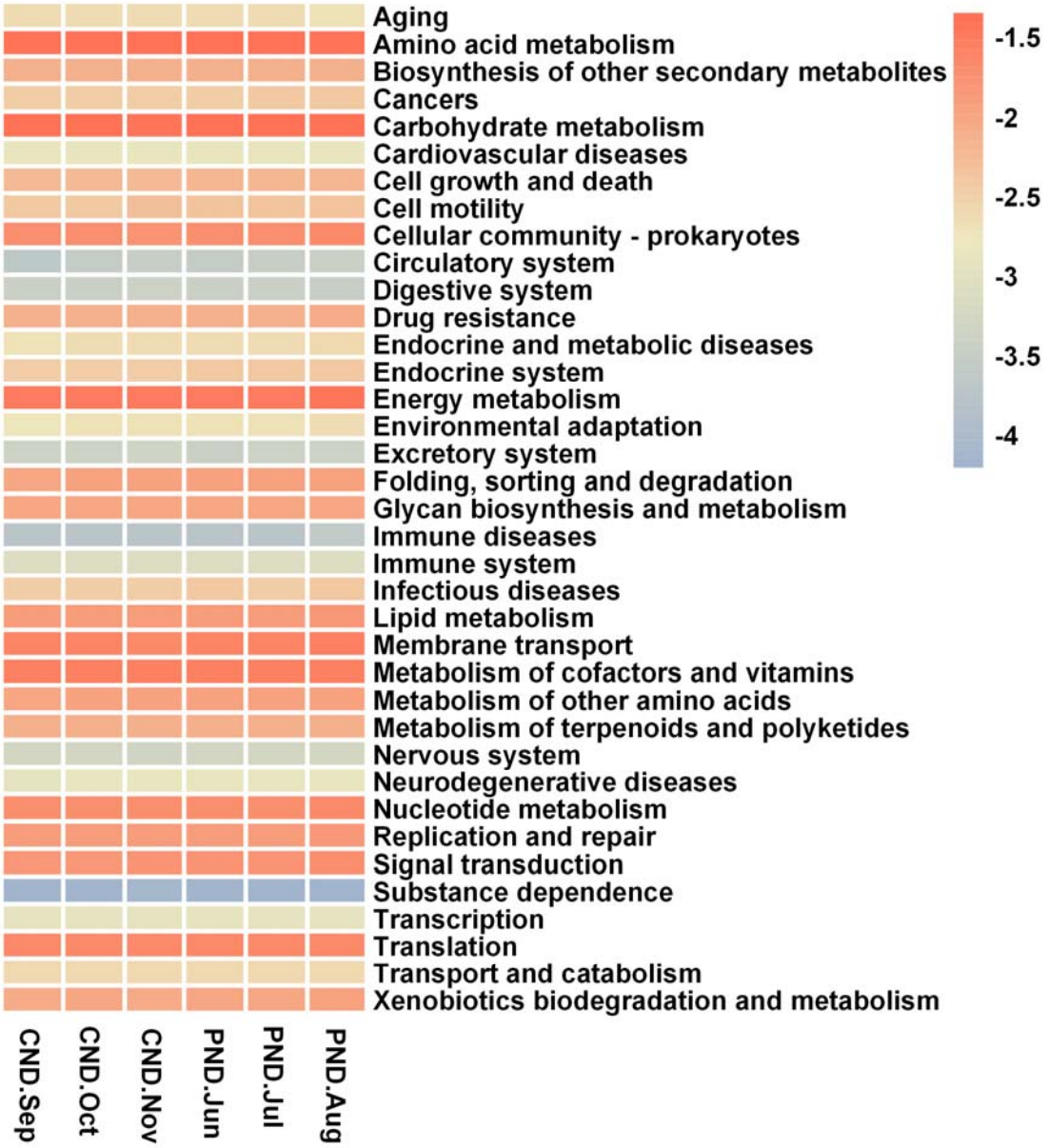

Figure S6 The abundance of metabolic pathways detected at the second level through KEGG annotation.

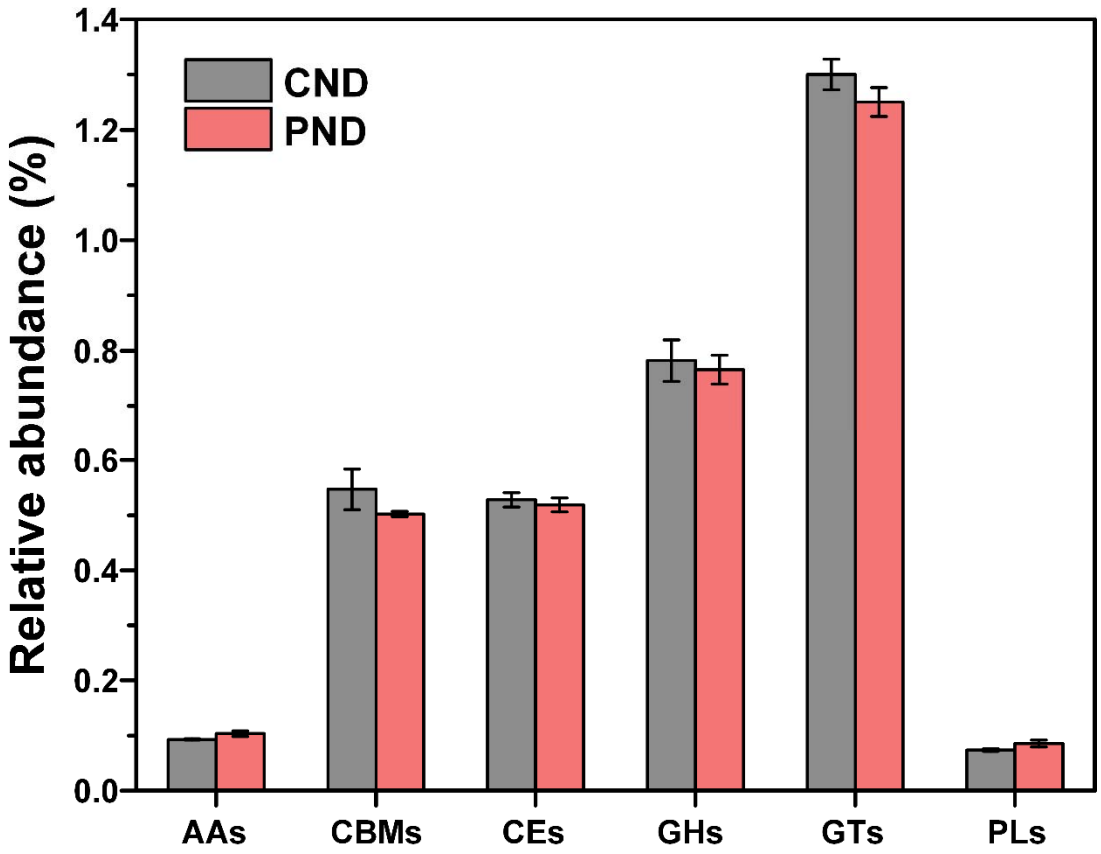

Figure S7 The abundance of carbohydrate metabolism active enzymes detected through CAZy annotation. (AAs: Auxiliary Activities; CBMs: Carbohydrate-Binding Modules; CEs: Carbohydrate Esterases; GHs: Glycoside Hydrolases; GTs: Glycosyl Transferases; PLs: Polysaccharide Lyases)

References

Bolger A M, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15): 2114-2120

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. Nature Methods, 7(5): 335-336

Edgar R C (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19): 2460-2461

Edgar R C, Haas B J, Clemente J C, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics, 27(16): 2194-2200

Fu L, Niu B, Zhu Z, Wu S, Li W (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23): 3150-3152

Huson D H, Auch A F, Qi J, Schuster S C (2007). MEGAN analysis of metagenomic data. Genome Research, 17(3): 377-386

Langmead B, Salzberg S L (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4): 357-359

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10): 1674-1676

Luo X, Shen L, Meng F (2019). Response of Microbial Community Structures and Functions of Nitrosifying Consortia to Biorefractory Humic Substances. ACS Sustainable Chemistry & Engineering, 7(5): 4744-4754

Schloss P D, Westcott S L, Ryabin T, Hall J R, Hartmann M, Hollister E B, Lesniewski R A, Oakley B B, Parks D H, Robinson C J, Sahl J W, Stres B, Thallinger G G, Horn D J V, Weber C F (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology, 75(23): 7537-7541

Wang Q, Garrity G M, Tiedje J M, Cole J R (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology, 73(16): 5261-5267

Zhu W, Lomsadze A, Borodovsky M (2010). Ab initio gene identification in metagenomic sequences. Nucleic Acids Research, 38(12): e132-e132