Supplementary Information

Microbial composition and nitrogen removal pathways in a novel sequencing batch reactor integrated with semi-fixed biofilm carrier: Evidence from a pilot study for low- and high-strength sewage treatment

Bin Cui [a,*]

[a] Graduate School, Guangzhou University, Guangzhou, 510060, China;

[*] Corresponding Author, E-mail: cuibin@gzhu.edu.cn

Text S1 Scene photograph profile

Figure S1. Photograph from the abraded inner wall of a moving bed biofilm reactor. a) exposed steel grid on inner wall; b) enlarged view of exposed steel grid.

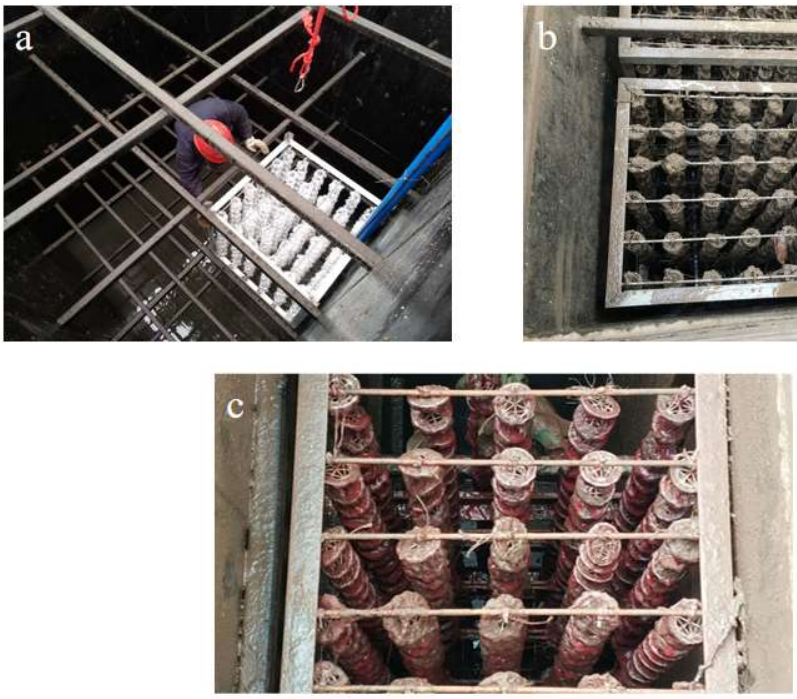

Figure S2. Photograph on the semi-fixed biofilm carrier during, a) initial stage, b) low-strength sewage load and c) high-strength sewage load.

Text S2 DNA extraction and bacterial community analysis

1. DNA extraction and sequencing

DNA was extracted using the E.Z.N.A.® Mag-Bind Soil DNA Kit (Omega Bio-tek, Norcross, GA, USA) according to the manufacturer's specifications. The V3–V4 region of the bacterial 16S rRNA gene was amplified and the polymerase chain reaction (PCR) conditions were as reported by Du *et al.* (2017) using a T100™ Thermal Cycler (Bio-Rad, Hercules, CA, USA). The PCR products were examined by 1% agarose gel electrophoresis and purified using MagicPure Size Selection DNA Beads (Transgen Biotech, Beijing, China). The purified PCR products were then quantified by Qubit 3.0 fluorometer (Life Technologies, cat. no. Q32866) before pyrosequencing. The amplicons were mixed at equal concentrations and sequenced on an Illumina MiSeq platform (Illumina, San Diego, USA) by Shanghai Sangon Biotech Co., Ltd.

2. Biodiversity analysis and phylogenetic classification

The sequence data generated was preprocessed using a series of steps. Briefly, Cutadapt (v.1.8.3) (Martin *et al.*, 2011) was used to remove adaptor sequences (-O 5 -m 50), PEAR (v.0.9.6) was used to merge reads (Zhang *et al.*, 2014), and PRINSEQ (v.0.0.4) (Schmieder *et al.*, 2011) was used to remove low-quality reads (-lc_method dust -lc_threshold 40 -min_len 200). The USEARCH (v. 11.0.667) script was used to performs additional quality filtering (-fastq_maxee 3.5) and conversion of a FASTQ file to FASTA format (Edgar *et al.*, 2013). Then, the FASTA sequences were screened to remove low-quality reads using MOTHUR software (v.1.42) following filtering

parameters, maxambig = 0, maxhomop = 8, minlength = 250, maxlength = 490 (Schloss *et al.*, 2009). The filtered high quality reads were dereplicated for noise filtering, chimeras identified and removed, and operational taxonomic units (OTUs) were identified at 97% similarity level by the USEARCH script with default parameters. Additional potential chimera sequences were identified and removed using Chimera slayer (Haas *et al.*, 2011) in Mothur with the Silva-based alignment of template file (v.132) (Quast *et al.*, 2013). The remaining OTUs representing sequences were assigned to the reference taxonomy using UCLUST in QIIME (v.1.9.1) against the Silva 132 reference database (Samarajeewa *et al.*, 2015). Then, the abundance table and taxonomy data were merged with the OTU IDs into an OTU table and then converted into BION (Biological Observation Matrix) format using biom-format software (v.2.1.6) (Mcdonald *et al.*, 2012). To eliminate the influence of sequencing depth across samples, a rarefied OTU table was generated at the same sequence number (30,300), and all following diversity and function analyses were performed using the rarefied-OTU table.

The OTU table data analysis were performed using the QIIME and Mothur software. Mothur software was used to calculate the following indexes (alpha diversity indexes) and generate the Venn diagram: the Shannon and Simpson indexes, representing the community diversity. The QIIME software was used to classify the taxonomy data into phylum, class, family and genus levels. Principal component analysis (PCA) was used to evaluate the relationship between the sample distribution and environmental factors using CANOCO v5.03 (Ter Braak *et al.*, 1989). The

heatmap of the top fifty genera was performed using the Pheatmap package (v.1.0.12) in R (Kolde *et al.*, 2015).

3. Functional prediction and analysis

To further explore the potential functional composition of sampled microbial communities in the suspension and biofilm, we used the OTU representative sequence and the rarefied OTU table to predict the bacterial metagenome by the PICRUSt2 algorithm (v.2.2.0-b) (Douglas *et al.*, 2019), and functional inferences were classified into the KEGG pathway, ENZYME and KO catalogs.

4. Data deposition

Raw sequence data produced in this study was deposited in NCBI under the BioProject accession PRJNA605585.

Text S3 Water quality parameters

Table S1. The main characteristics of influent sewage

Parameter	Unit	Value	Parameter	Unit	Value
pH	—	7.3 ~ 7.7	NH_4^+ - N	mg L^{-1}	2.9 ~ 48
COD	mg L^{-1}	43 ~ 522	NO_3^- - N	mg L^{-1}	＜1
BOD	mg L^{-1}	40 ~ 160	TN	mg L^{-1}	5.8 ~ 58
SS	mg L^{-1}	80 ~ 150	Temperature	℃	9 ~ 30
TP	mg L^{-1}	0.7 ~ 59			

COD: chemical oxygen demand; BOD: biochemical oxygen demand; SS: suspended solids; TP: total phosphorus; TN: total nitrogen.

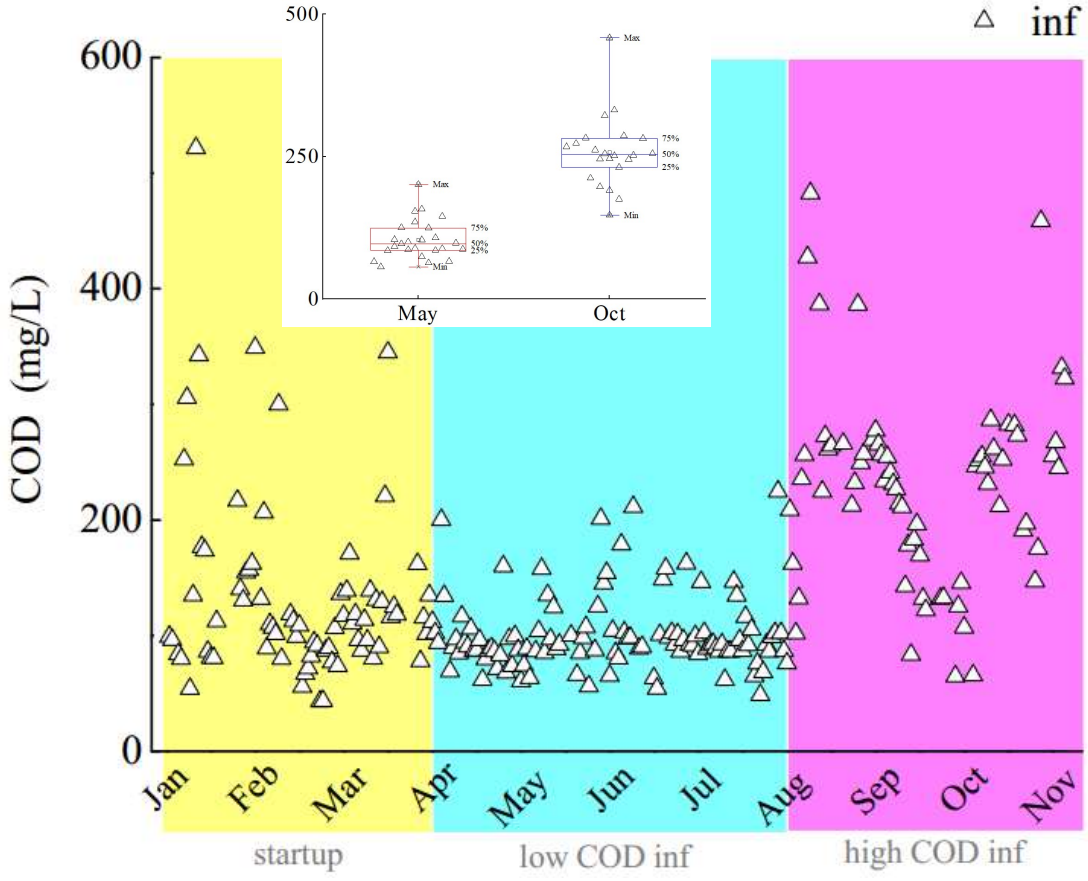

Figure S3. Chemical oxygen demand (COD) of influent (and the values of COD in May and October detailed in the insert box chart) at the novel pilot-scale SBR integrated with semi-fixed biofilm carrier. This novel SBR had been operated with three stages, including startup (from January to April), low COD influent (from April to August) and high COD influent (from August to November).

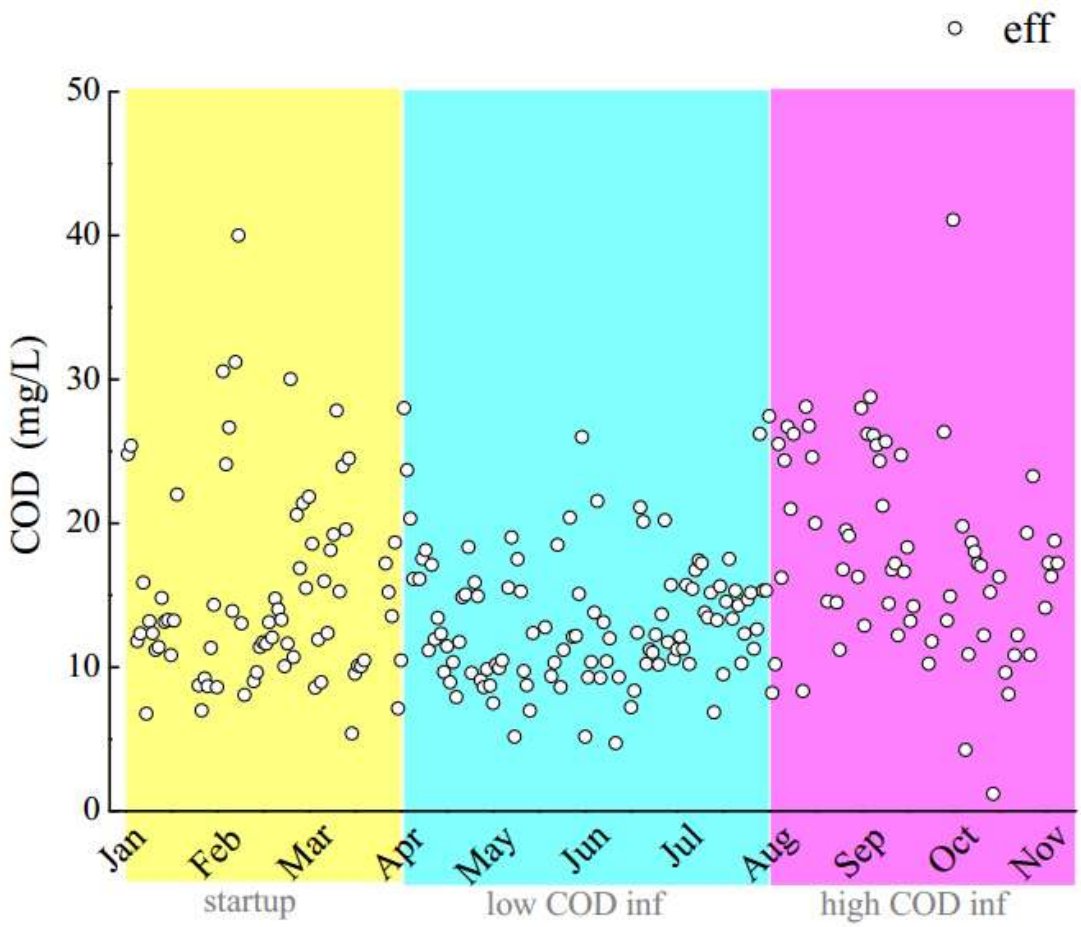

Figure S4. Chemical oxygen demand (COD) of effluent at the novel pilot-scale SBR integrated with semi-fixed biofilm carrier. This novel SBR had been operated with three stages, including startup (from January to April), low COD influent (from April to August) and high COD influent (from August to November).

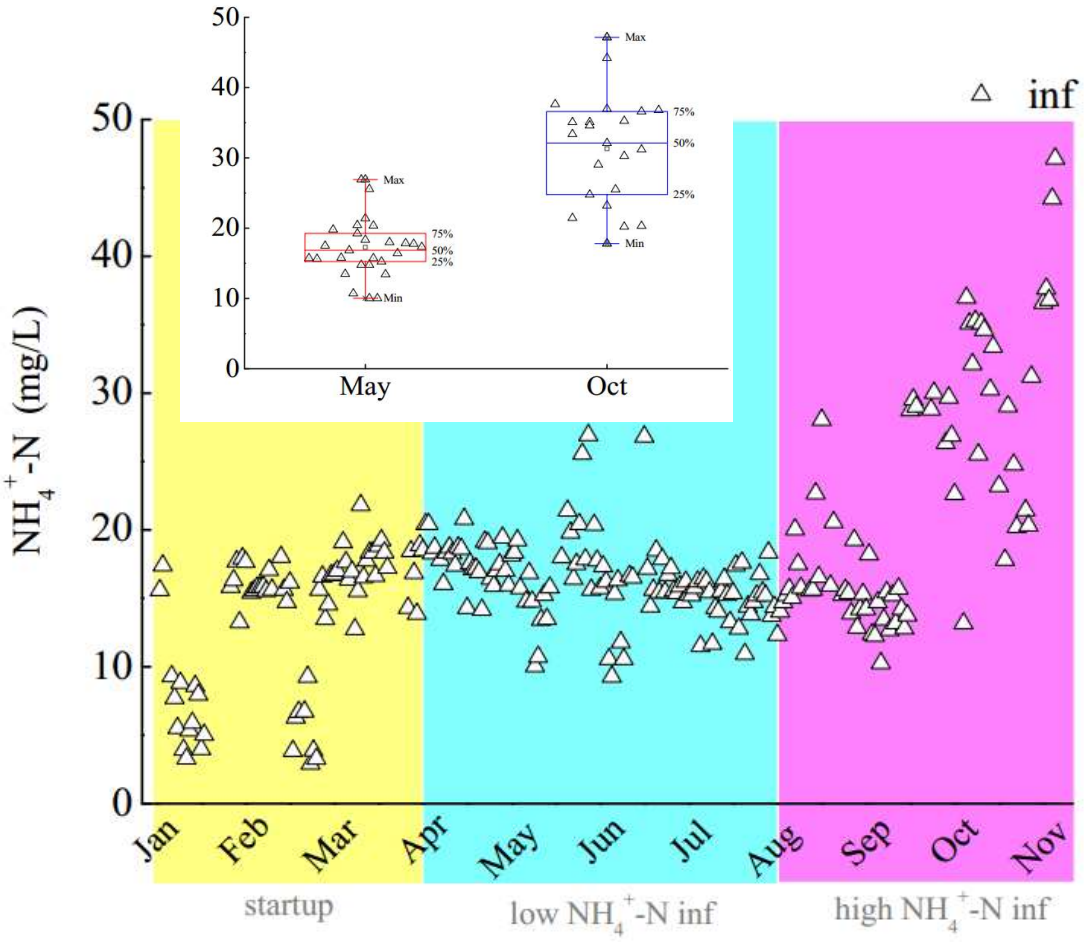

Figure S5. Ammonium-nitrogen (NH_4^+-N) concentrations in influent (and the values of NH_4^+-N in May and October detailed in the insert box chart) at the novel pilot-scale SBR integrated with semi-fixed biofilm carrier. This novel SBR had been operated with three stages, including startup (from January to April), low NH_4^+-N influent (from April to August) and high NH_4^+-N influent (from August to November).

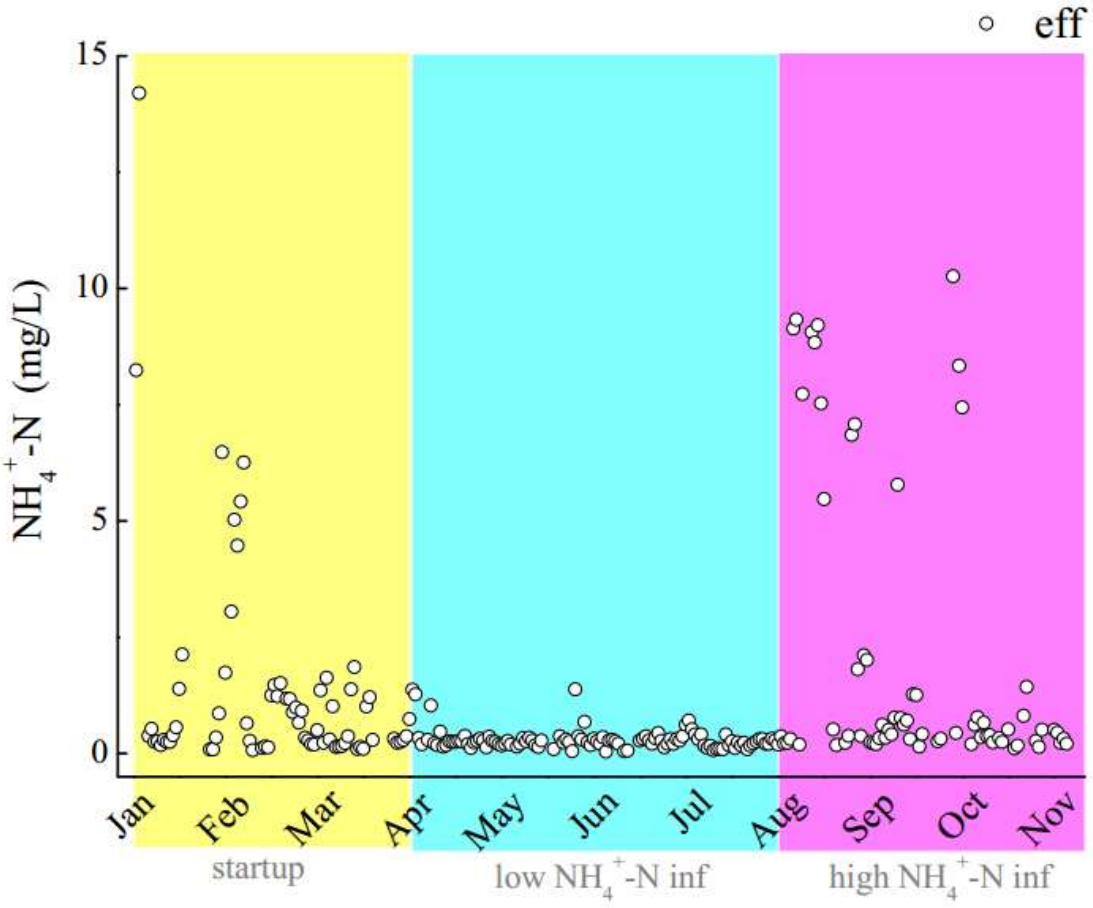

Figure S6. Ammonium-nitrogen (NH_4^+-N) effluent species at the novel pilot-scale SBR integrated with semi-fixed biofilm carrier. This novel SBR had been operated with three stages, including startup (from January to April), low NH_4^+-N influent (from April to August) and high NH_4^+-N influent (from August to November).

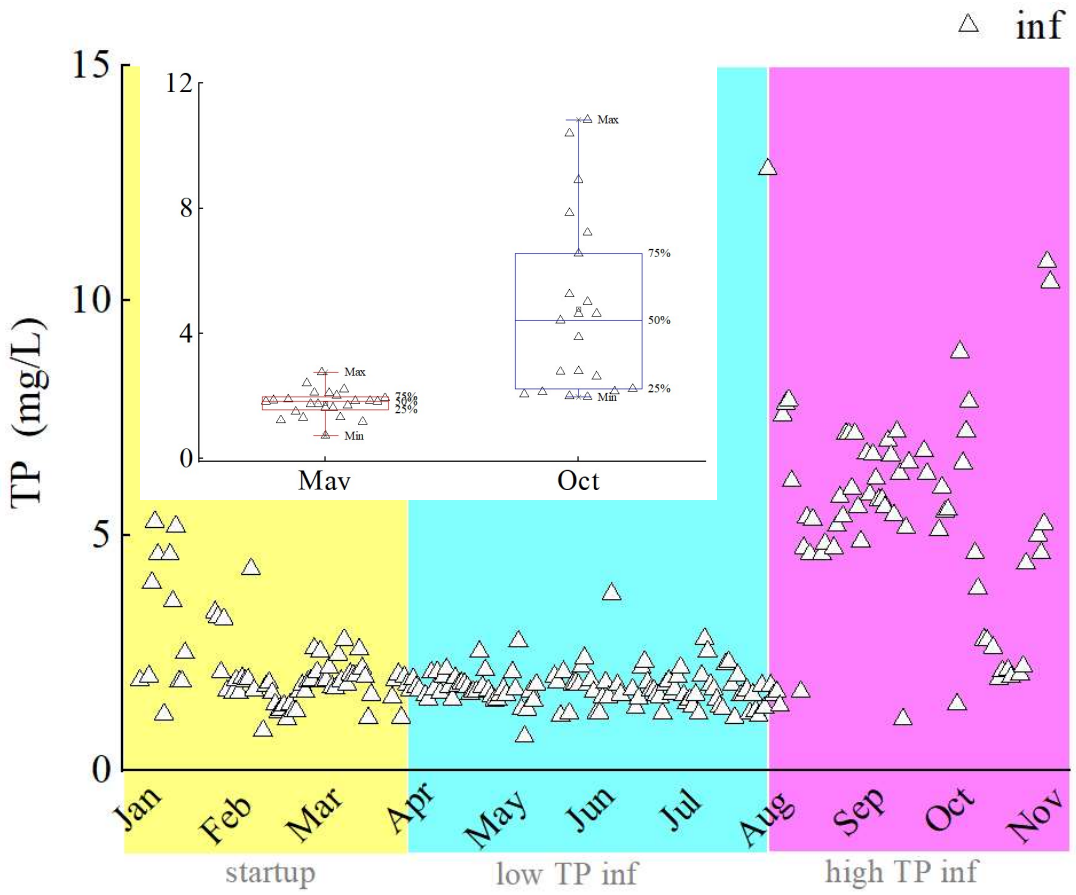

Figure S7. Total phosphates (TP) concentration in influent (and the values of TP in May and October detailed in the insert box chart) at the novel pilot-scale SBR integrated with semi-fixed biofilm carrier. This novel SBR had been operated with three stages, including startup (from January to April), low TP influent (from April to August) and high TP influent (from August to November).

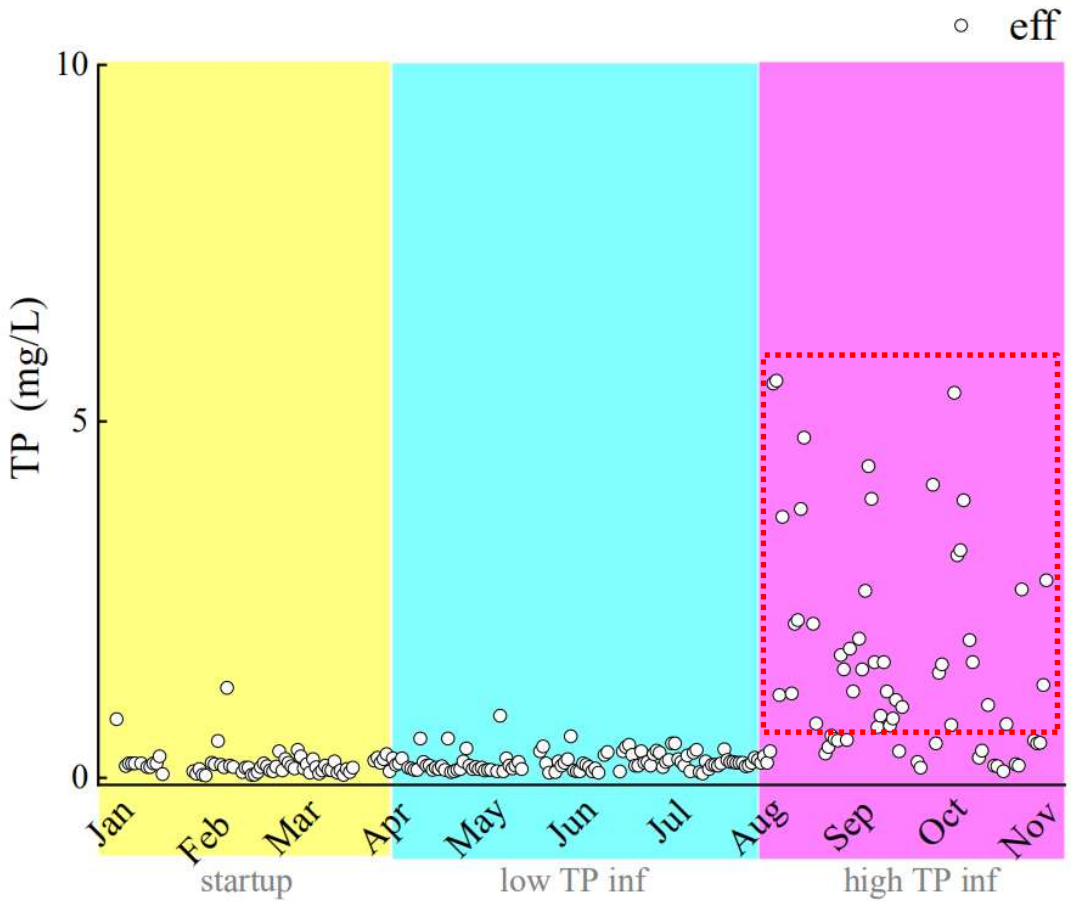

Figure S8. Total phosphates (TP) concentration in effluent at the novel pilot-scale SBR integrated with semi-fixed biofilm carrier. This novel SBR had been operated with three stages, including startup (from January to April), low TP influent (from April to August) and high TP influent (from August to November). The rectangle in red denoted the absence of dosing poly aluminum chloride during August to October, leading the over discharge for TP.

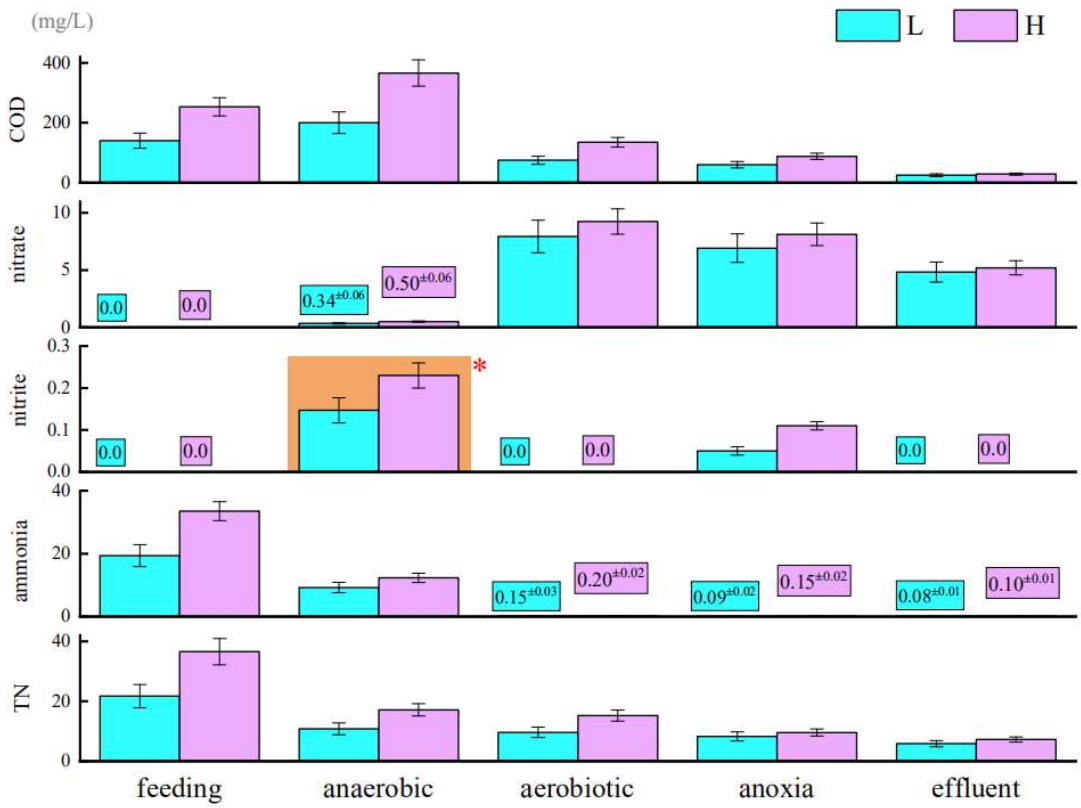

Figure S9. Value of COD, nitrate, nitrite, ammonia and total nitrogen (TN) concentrations throughout entire process in the novel SBR during low- and high- strength sewage load, respectively. Asterisk in red indicated higher nitrite concentration presenting in high- strength sewage load.

Text S4 Profile of microbial communities in suspended aggregates and biofilms

Table S2. Sequence reads, diversity, richness indices, coverage and operational taxonomic units (OTUs) at 97% sequence identity.

Sample	Sequence Reads		OTUs	Diversity indices		Richness indices		Good's coverage
	Raw	Filtered		Shannon	Simpsons	ACE	Chao1	
L_SU	41851	39317	1128	5.99	0.01	1161.77	1167.80	0.997
L_B	48404	45127	1144	5.95	0.01	1186.66	1195.53	0.997
H_SU	57986	48720	1095	5.73	0.01	1150.41	1159.91	0.996
H_B	57100	48277	1110	5.77	0.01	1181.44	1203.56	0.995

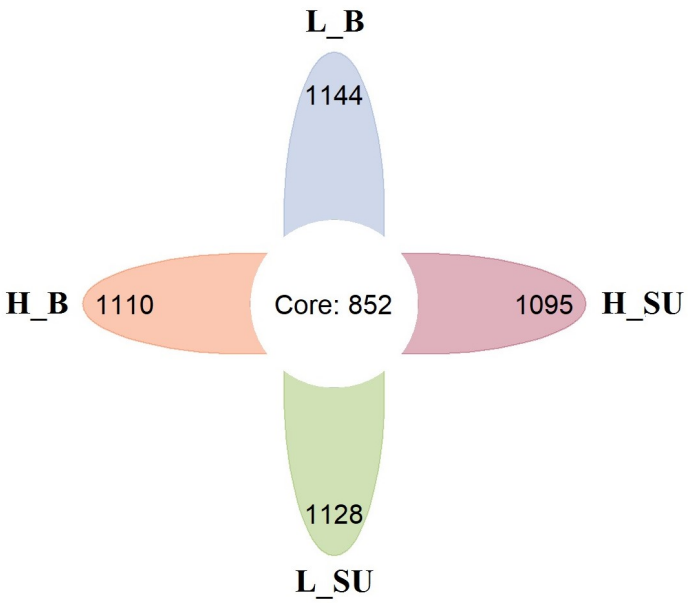

Figure S10. Venn diagram of species richness measures shows the total shared and unique operational taxonomic units (OTUs) between suspended aggregates and semi-fixed biofilm carrier.

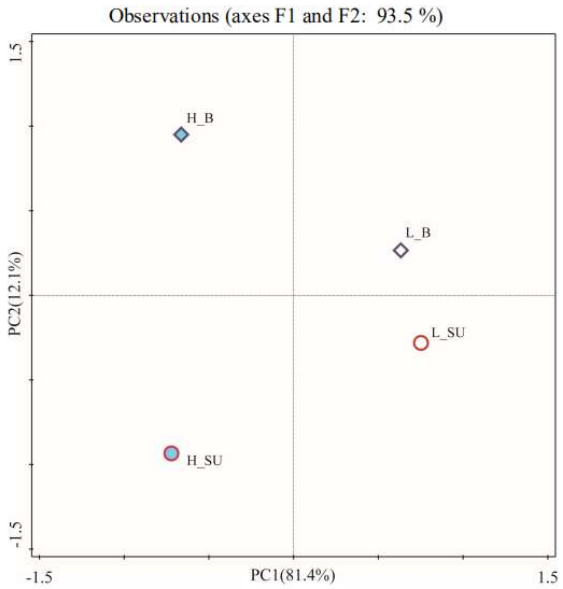

Figure S11. Principal component analysis (PCA) of microbial communities of 4 specimens based on metagenomes. The first two principal components (PC1 and PC2) explained 93.5% of the data variance.

Text S5 Analysis of the KEGG pathway and genes associated with nitrogen metabolism

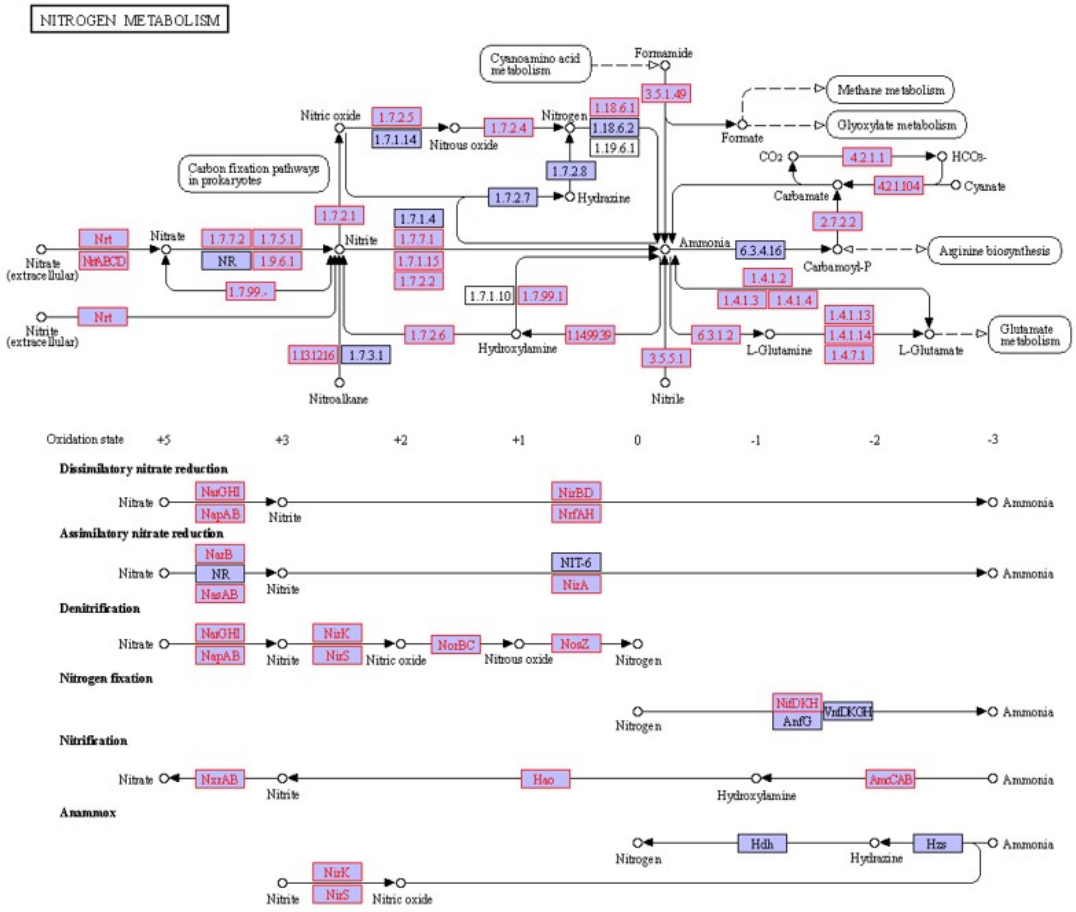

Figure S12. Nitrogen metabolism of microorganisms in the sequencing batch reactor. Based on the reference pathway (KO), The different gene-encoding enzymes and proteins relating to the nitrogen metabolism pathways in the bacterial communities were marked in red.

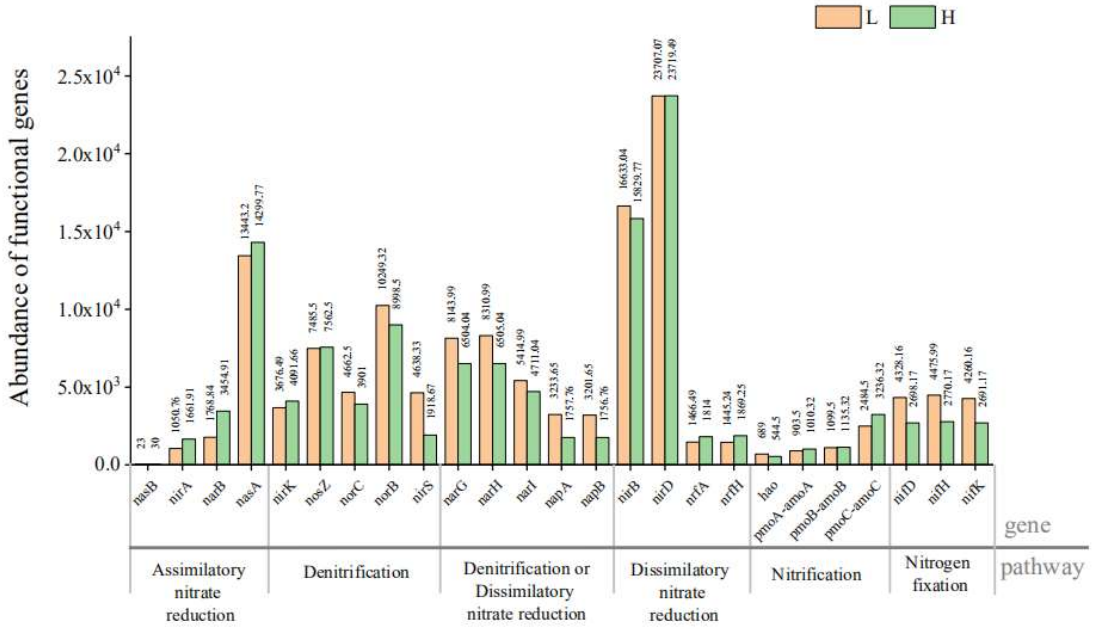

Figure S13. Abundances of functional genes related to the core nitrogen metabolism pathways.

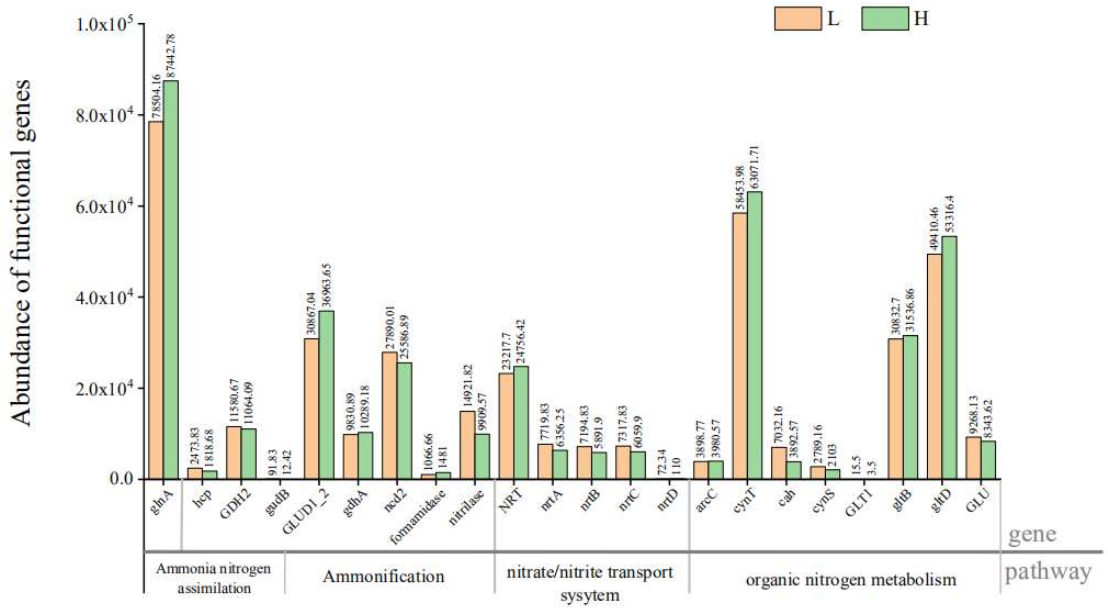

Figure S14. Abundances of functional genes related to the subordinate nitrogen metabolism pathways.

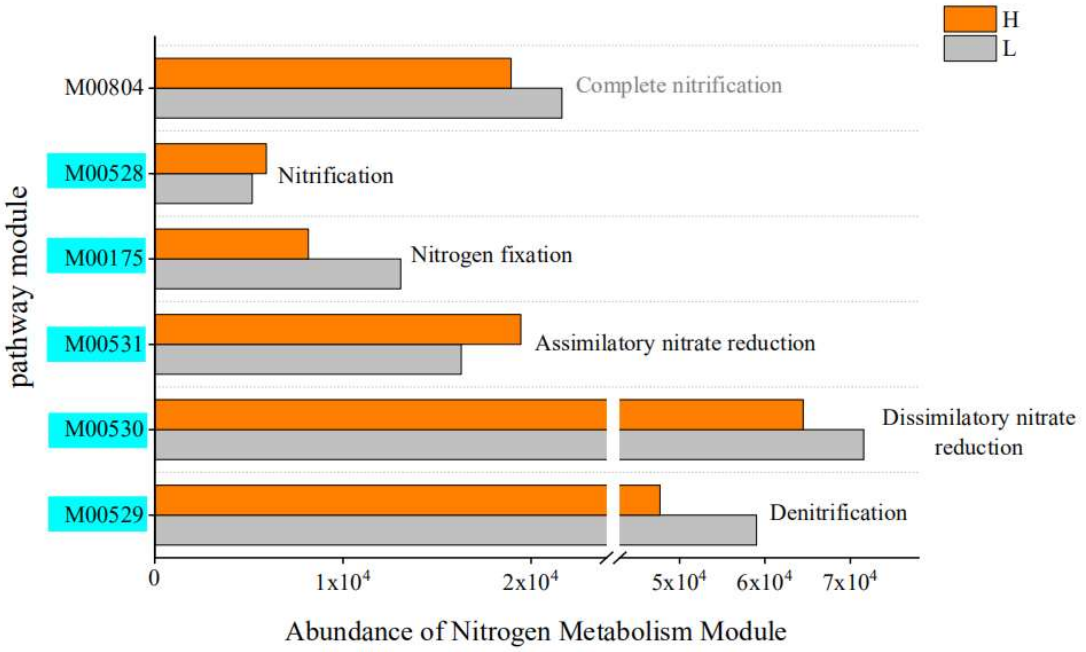

Figure S15.Abundances of KEGG module classifications for each treatment in the nitrogen metabolism pathway.

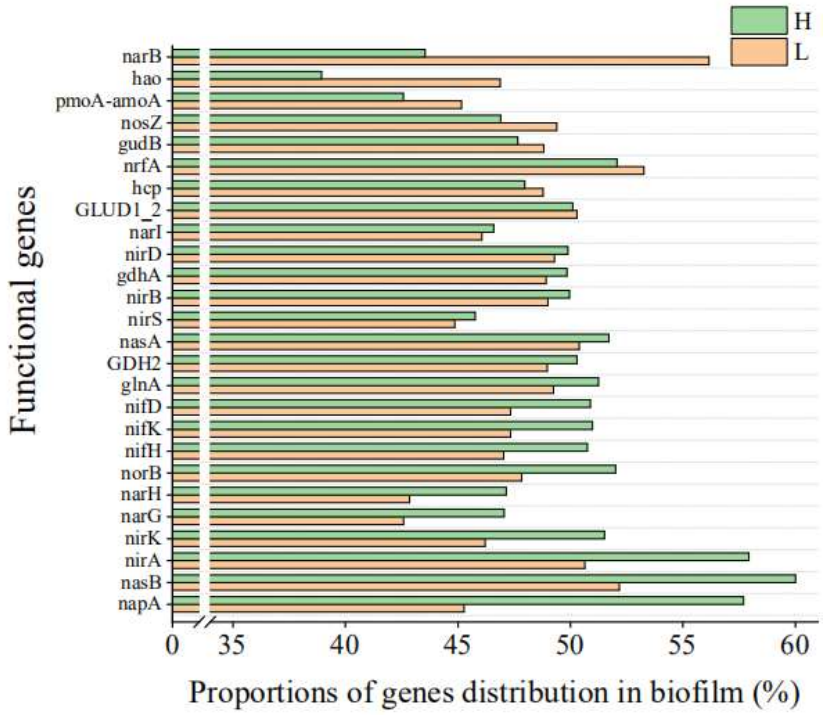

Figure S16. Proportions of the key functional genes in biofilm samples for each treatment related to the main nitrogen metabolism pathways. H: high-strength sewage load; L: low-strength sewage load.

References

Douglas, G.M., Maffei, V.J., Zaneveld, J., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., Langille, M.G., 2019. PICRUSt2: An improved and extensible approach for metagenome inference. BioRxiv, 672295.

Du, L., Chen, Q.R., Liu, P.P., Zhang, X., Wang, H.H., Zhou, Q.H., Xu, D., Wu, Z.B., 2017. Phosphorus removal performance and biological dephosphorization process in treating reclaimed water by Integrated Vertical-flow Constructed Wetlands (IVCWs). Bioresource Technology, 243:204-211.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods, 10(10):996–998.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methe, B.A., Desantis, T.Z., Petrosino, J.F., Knight, R., Birren, B.W., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research, 21(3):494-504.

Kolde, R., 2015. Pheatmap: pretty heatmaps.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17(1):10-12.

Mcdonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D.,Wilke, A., Huse, S.M., Hufnagle, J., Meyer, F., Knight, R., Caporaso, J. G., 2012. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience, 1(1):7-7.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glockner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41:590-596.

Samarajeewa, A.D., Hammad, A., Masson, L., Khan, I.U.H., Scroggins, R., Beaudette, L.A., 2015. Comparative assessment of next-generation sequencing, denaturing gradient gel electrophoresis, clonal restriction fragment length polymorphism and cloning-sequencing as methods for characterizing commercial microbial consortia. Journal of Microbiological Methods, 108:103-111.

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., Van Horn, D.J., 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.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics, 27(6):863-864.

Ter Braak, C.J.F., 1989. CANOCO—an extension of DECORANA to analyze species-environment relationships. Hydrobiologia, 184(3):169-170.

Zhang, J.J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5):614-620.