Environmental Science & Technology

pubs.acs.org/est

Article

Predicting Anaerobic Membrane Bioreactor Performance Using Flow-Cytometry-Derived High and Low Nucleic Acid Content Cells

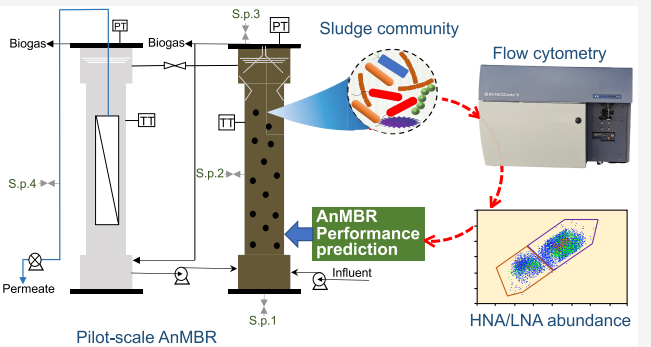
Hong Cheng,* Julie Sanchez Medina, Jianqiang Zhou, Eduardo Machado Pinho, Rui Meng, Liuwei Wang, Qiang He, Xosé Anxelu G. Morán, and Pei-Ying Hong*

Cite This: *Environ. Sci. Technol.* 2024, 58, 2360−2372

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Having a tool to monitor the microbial abundances rapidly and to utilize the data to predict the reactor performance would facilitate the operation of an anaerobic membrane bioreactor (AnMBR). This study aims to achieve the aforementioned scenario by developing a linear regression model that incorporates a time-lagging mode. The model uses low nucleic acid (LNA) cell numbers and the ratio of high nucleic acid (HNA) to LNA cells as an input data set. First, the model was trained using data sets obtained from a 35 L pilot-scale AnMBR. The model was able to predict the chemical oxygen demand (COD) removal efficiency and methane production 3.5 days in advance. Subsequent validation of the model using flow cytometry (FCM)-derived data (at time t − 3.5 days) obtained from another biologically independent reactor did not exhibit any substantial difference between predicted and actual measurements of reactor performance at time t. Further cell sorting, 16S rRNA gene sequencing, and correlation analysis partly attributed this accurate prediction to HNA genera (e.g., *Anaerovibrio* and unclassified Bacteroidales) and LNA genera (e.g., *Achromobacter*, *Ochrobactrum*, and unclassified Anaerolineae). In summary, our findings suggest that HNA and LNA cell routine enumeration, along with the trained model, can derive a fast approach to predict the AnMBR performance.



KEYWORDS: *anaerobic membrane bioreactor, flow cytometry, HNA and LNA cells, microbial diversity, predictive model*

1. INTRODUCTION

An anaerobic membrane bioreactor (AnMBR) is increasingly being viewed as a promising biotechnology for sustainable municipal wastewater treatment.[1,2] This bioreactor, which combines membrane-based filtration with anaerobic fermentation, has several advantages over conventional aerobic processes that are currently used in most municipal wastewater treatment plants. For instance, an AnMBR eliminates the need for aeration and reduces the energy consumption rate from ca. 2 kWh/m^3 in an aerobic membrane bioreactor (MBR) to ca. 0.8 kWh/m^3.[3] In addition, anaerobic fermentation converts organic carbon in municipal wastewater into methane, an energy source, while retaining total nitrogen and phosphorus in the effluent, which can be beneficial when used to irrigate certain types of crops.[4] However, anaerobic fermentation is highly reliant on symbiosis among fermenters (e.g., *Acinetobater* and *Clostridium*), syntrophs (e.g., *Syntrophomonas*), and methanogens (e.g., *Methanosaeta*).[5,6] Therefore, tools that monitor the abundance or relative proportions of microbial populations, which are associated with the reactor performance, would be beneficial during the startup and routine operation of this process.

Molecular-based tools, e.g., quantitative polymerase chain reaction (qPCR), can quantify the gene copies or cell numbers of bacteria but would require DNA extraction prior to qPCR (which collectively takes ca. 2−3 h from start until completion) and knowledge on which microbial targets to monitor.[7] 16S rRNA gene-based amplicon sequencing can help determine the relative abundance of microbial populations without the need for prior knowledge on microbial targets. However, this method requires DNA extraction, library preparation, the deployment of samples on the next-generation sequencing platform, and the subsequent bioinformatic analysis of raw data. With all of these required steps, 16S rRNA gene-based amplicon sequencing can take a long time to determine the existing microbial populations and their effect on reactor operation. Considering the existing limitations of qPCR and amplicon sequencing, flow cytometry (FCM) may

Received: September 18, 2023
Revised: December 20, 2023
Accepted: December 20, 2023
Published: January 23, 2024

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

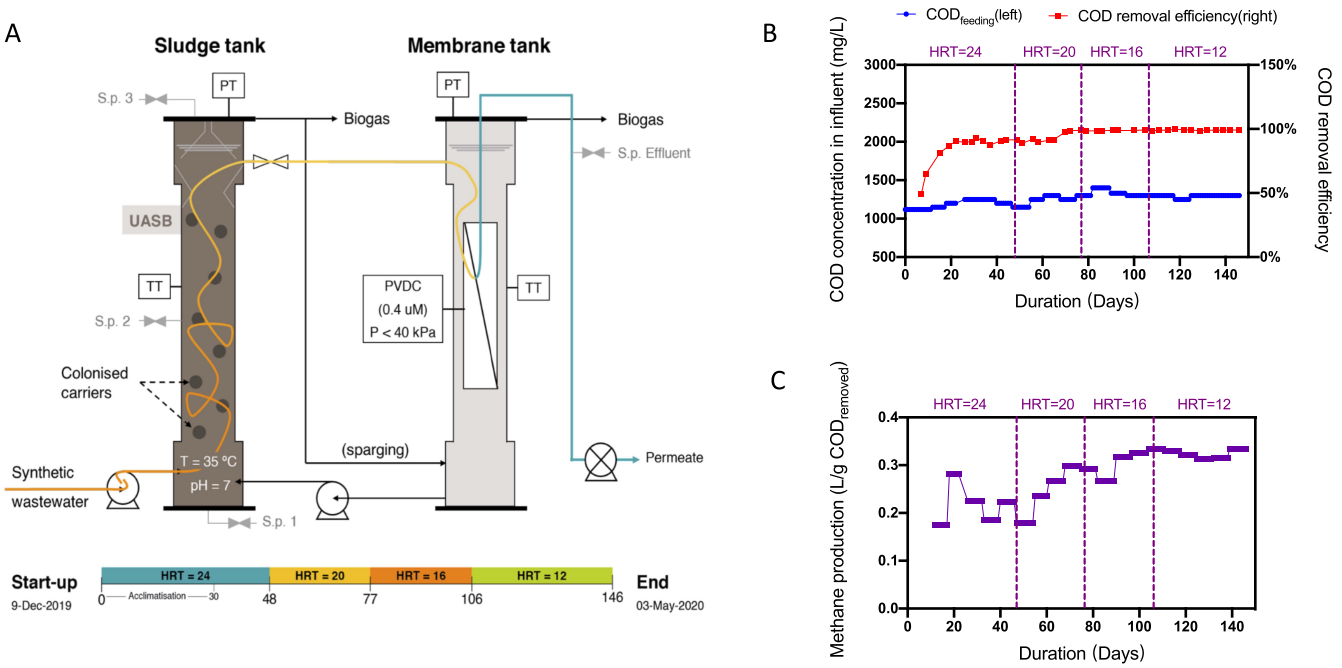


Figure 1. Pilot AnMBR and its performance considering the variation in the HRT: (A) different operational phases conducted and sample points for this study, (B) COD removal efficiency, and (C) methane production.

be a good alternative method to rapidly enumerate microbial abundance without the need to perform DNA extraction and can be performed at a low cost of <0.1 USD per sample.[8]

FCM detects discrete particles suspended in a sheath fluid in two ways. First, the light signals of the laser beam are scattered off upon hitting the target particle at different angles, depending upon the cell size and density. These signals are detected as forward (FSC) and side (SSC) light scatter, respectively. Second, fluorescence signals, including autofluorescence emitted by the cells (e.g., chlorophyll in algae and microalgae and coenzyme F420 in methanogens) and/or that emitted from the nucleic acid stains (e.g., SYBR- and SYTO-based stains), are detected on the basis of the appropriate excitation and emission wavelengths of the laser. FCM has already been used on drinking water or surface water to acquire information on total cell concentrations.[9−11] In addition to enumerating cell numbers, cells with a high nucleic acid (HNA) content can be differentiated from those with a low nucleic acid (LNA) content on a flow cytogram using a nuclear stain, because the former would emit higher signal fluorescence as a result of a larger genome size or differences in phylogenetic composition.[12−14]

HNA population densities were previously observed to correlate with perturbations. To exemplify, Zolkefli et al. tracked changes in the proportion of river HNA to LNA microbial communities in response to the palm oil mill effluent discharging and observed that HNA cells may be much more susceptible to external nutrients than LNA cells.[15] Therefore, by determining the ratio of HNA/LNA and observing the deviation from the baseline ratio, one can potentially relate these changes to aberrations in treatment conditions or environmental stressors that resulted in a spike in either group of bacteria (i.e., HNA or LNA). On the basis of observations from earlier studies,[14−17] it was hypothesized that the abundance ratio of HNA and LNA monitored by FCM could potentially be used as a tool to rapidly predict the AnMBR reactor performance.

Herein, a pilot-scale AnMBR was initiated at an initial hydraulic retention time (HRT) of 24 h, which was then gradually decreased to 12 h. The corresponding organic loading rate (OLR) therefore increased from 1.15 to 2.67 g L^{-1} day^{-1}. The correlation between the performance of AnMBR, α microbial diversity, and changes in HNA and LNA abundances as determined by FCM were analyzed. A linear regression model was then fitted using FCM-derived data to correlate with the reactor performance and further tested for its accuracy using samples obtained from an AnMBR that was operated in a biologically independent manner. In addition, selected HNA and LNA microbiota that were successfully sorted through FCM were further identified for their phylogenetic identities. This study aims to demonstrate the feasibility of using FCM-derived data within a regression model that can predict AnMBR performance in advance.

2. MATERIALS AND METHODS

2.1. Pilot Reactor Configuration and Operating Condition. A pilot AnMBR was set up and operated over 5 months since Dec 9, 2021 (Figure 1A). The pilot AnMBR included two tanks (i.e., a 35 L sludge tank and a 35 L membrane tank). The sludge tank was filled with plastic bio-ball carriers (2.5 cm diameter) and operated in up-flow mode, and the membrane tank had a 0.5 m^2 submerged cross-flow flat sheet polyvinylidene difluoride (PVDF) microfiltration membrane (nominal pore size of 0.4 μm, PHILOS membrane, Korea). The inoculated anaerobic sludge concentration included a mixed liquor suspended solid (MLSS) of approximately 4.18 g/L. The AnMBR was then maintained at 35 °C and fed with synthetic wastewater, with constituents listed in Table S1 of the Supporting Information.[18] The synthetic wastewater included a chemical oxygen demand (COD) of 1247 ± 77 mg/L and was adjusted to pH 7. Herein, four HRTs were considered. The initial HRT was set at 24 h for the first 48 days (i.e., phase 1). The HRT was then decreased to 20 h in the following 29 days (i.e., phase 2) and

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

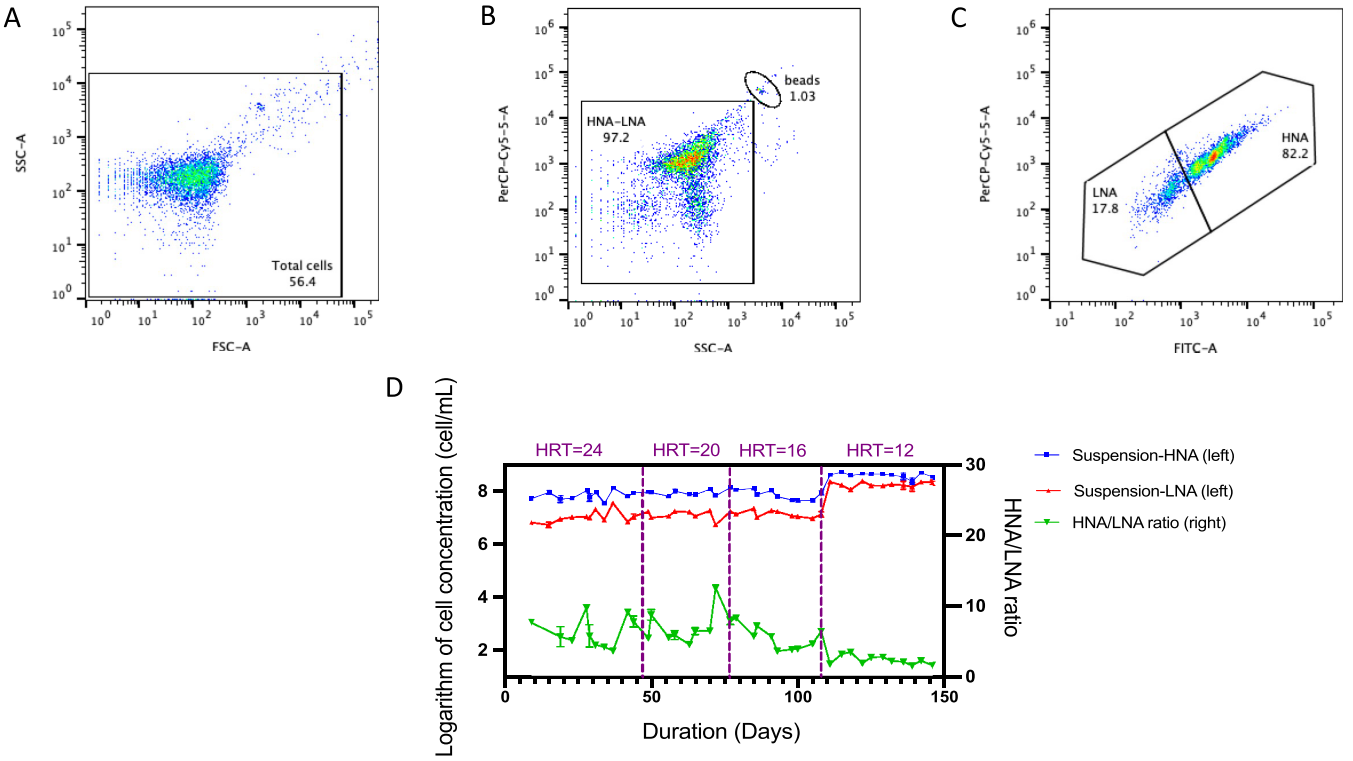


Figure 2. Gating strategy to quantify the HNA and LNA populations in sludge samples and the monitoring of their actual abundance: (A) removal of events on the edges of SSC/FSC, (B) selection of the cell population for HNA/LNA analysis with FITC/SSC, (C) establishing gates between HNA and LNA subpopulations, and (D) number of HNA and LNA cells and the ratio of HNA/LNA abundance at different HRTs.

then 16 h for an additional 29 days (i.e., phase 3) and last was reduced to 12 h for the remaining 40 days (i.e., phase 4). The variations in HRTs resulted in the corresponding OLRs from 1.15 to 2.67 g of COD L^{-1} day^{-1}. The change in HRTs was made when the reactor achieved a steady state at the earlier HRT (i.e., more than 91% COD removal for over 10 days and with no further increase in COD removal). Throughout the bioreactor operation, the membrane was not cleaned or changed; however, the transmembrane pressure was monitored to maintain below 40 kPa. The sampling points are shown in Figure 1A. Three 60 mL samples are individually taken from the bottom, middle, and top sections of the sludge tank (i.e., S.p 1, S.p 2, and S.p 3, respectively) twice per week throughout this study. Thus, the solid retention time (SRT) was considered as approximately 680 days. At the same time, 60 mL of effluent was sampled after membrane filtration (i.e., S.p effluent; Figure 1A).

2.2. Biogas and Methane Measurement. Biogas produced from AnMBR was captured continuously in gas bags from the headspace of the reactor. The produced biogas volume was measured using a wet drum chamber gas meter (Ritter Apparatebau, Germany). The biogas composition (i.e., CH_4, O_2, N_2, and H_2) was measured using a SRI model 301C gas chromatograph (SRI Instruments, Torrance, CA, U.S.A.) with a molecular sieve column using argon as a carrier gas, followed by a thermal conductivity detector (TCD). CO_2 was measured using another SRI model 301C gas chromatograph with a silica column using helium as the carrier gas, followed by a TCD. Each biogas sample was measured considering technical triplicates.

2.3. Flow Cytometry and Flow Cytogram Analysis. The samples were analyzed via flow cytometry, as described in one of the previous studies with slight modifications.[19] In summary, 2 mL of each sample retrieved from the sludge tank was diluted 10 000-fold with sterile water. All of the diluted samples were incubated in the dark for 10 min at 37 °C and then stained using 1× SYBR Green I (Invitrogen Corp., Carlsbad, CA, U.S.A.). After staining, 400 μL of the stained samples with 10 μL of 10^6 latex beads mL^{-1} (1.002 μm, Polysciences Europe) solution was enumerated by a FACSCanto II instrument equipped with a blue 488 nm Sapphire 488-20 (20 mW, Coherent). Molecular Probes fluorescent latex beads were used as an internal standard for fluorescence measurement and particle sizing. The 488 nm lasers were used for the analysis of the forward scattering channel (FSC, 488/10), side scattering channel (SSC, trigger signal, 488/10), SYBR Green I fluorescence (FITC, 530/30), and red fluorescence (PerCP-Cy5.5, >670). The samples were individually analyzed at a high flow rate (i.e., ca. 117.9 μL min^{-1}) until 10 000 events were performed. Measurements were conducted in technical triplicates. After evaluation, FlowJo V10 (FlowJo, Ashland, OR, U.S.A.) was used to analyze the flow cytogram accordingly.[20] First, the events were viewed on SSC-A versus FSC-A channels, and abnormal events located on the edge of the plots (i.e., abnormal size or morphological complexities) were removed (Figure 2A). Then, the evaluation of the FITC-A versus SSC-A channels allowed for the visualization of the region associated with bacterial cells, identified on the basis of the size comparison to latex beads (Figure 2B). Furthermore, when viewed on the PerCP-CyTM5.5 versus FITC-A channels, HNA and LNA clusters can be distinguished as two clusters because of the difference in the microbial nucleic acid content (Figure 2C). The

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

abundance of HNA and LNA cells within each cluster was enumerated accordingly.

2.4. HNA and LNA Cell Sorting. To identify the HNA and LNA cells, BD FACS Aria III (BD Bioscience, Franklin Lakes, NJ, U.S.A.) was further applied to sort and enrich HNA and LNA cells similar to previous studies.[20,21] Samples stained by 1× SYBR Green I (Invitrogen Corp., Carlsbad, CA, U.S.A.) were HNA and LNA cells determined on the basis of the gating strategy described in section 2.3. Sheath fluid was applied as the negative control at each sorting. A total of 1×10^7 events were sorted using a 70 μm nozzle for each HNA or LNA group that was present in each sample, with the HNA and LNA cells from each sample kept separately in sterile microcentrifuge tubes and frozen at −80 °C prior to DNA extraction. Sheath fluid was also collected at the end of every sorting day as a negative control to detect contaminant DNA. Data files were analyzed using FlowJo V10 (FlowJo, Ashland, OR, U.S.A.).

2.5. DNA Extraction, 16S rRNA Gene Amplicon Sequencing, and Data Processing. The microbial richness, the diversity of the sampled sludge, and the sorted HNA and LNA groups were analyzed via high-throughput sequencing in accordance with the previous report.[22] The genomic DNA was extracted using the PowerSoil DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, U.S.A.) with slight modifications.[23] The quality and quantity of extracted DNA were checked by a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, U.S.A.). 16S rRNA gene fragments (V3−V4) of the extracted DNA from each sample were amplified using primers 515F (5′-Illumina overhang-GTGYCAGCMGCCGCGGTAA-3′) and 907R (5′-Illumina overhang- CCCCGYCAATTCMTTTRAGT-3′). The amplicons were of the anticipated size of approximately 550 bp, and the negative control had no amplification. Polymerase chain reaction (PCR) amplicons were cleaned using AMPure XP beads (Beckman Coulter, Brea, CA, U.S.A.). Hereafter, index PCR was conducted on the PCR amplicons derived from each sample to attach dual indices provided by the Nextera XT Index Kit (Illumina, Inc., San Diego, CA, U.S.A.), which allowed us to bioinformatically differentiate each sample after sequencing was completed. Indexed PCR amplicons were cleaned by AMPure XP beads (Beckman Coulter, Brea, CA, U.S.A.). The indexed PCR amplicons from the samples were mixed in equimolar concentrations and then submitted to the King Abdullah University of Science and Technology (KAUST) Core Lab for Illumina MiSeq sequencing. All high-throughput sequencing raw data were deposited in the short read archive of the European Nucleotide Archive under study accession number PRJEB54690. After sequencing, the raw amplicon sequences obtained in this study were analyzed on the QIIME 2 platform (version 2020.2).[24] The sequencing data were first processed using DADA2,[25] including quality filtering, denoising, and chimera removal. Amplicon sequence variants, which are analogous to operational taxonomic units (OTUs), were generated by DADA2. Taxonomy was assigned to OTUs using q2-feature-classifier[26] (the classify-sklearn option) against Silva SSU database release 132.[27] To further illustrate the dynamics of microbial diversity under different HRTs, the α diversity of sludge (i.e., Chao1, Pielou, and Shannon indexes) was further evaluated through QIIME 2. The previous study reported that the Chao 1 index could indicate the richness of the microbial community, the Pielou index could provide information about the equity in species abundance in each sample, and the Shannon index could offer information about both richness and evenness.[28] For sorted HNA and LNA identification, to minimize errors incurred during phylogenetic identification because cell sorting cannot be performed precisely to ensure perfect separation and sorting of the HNA and LNA cells, we define those genera that were consistently present and in considerably higher relative abundance in HNA-sequenced samples as HNA cells. In contrast, the genera that were consistently present and in considerably higher relative abundance in LNA-sequenced samples were the LNA cells. All other genera that did not match these criteria are not considered herein.

2.6. Correlation Analysis. To visualize the correlations in a network interface, all possible pairwise Spearman's rank correlations between reactor performances (i.e., COD removal efficiency, methane production, and OLR), α diversity of the microbial community, and FCM-derived results (i.e., HNA cell abundance, LNA cell abundance, and HNA/LNA ratio) were calculated to construct a correlation matrix. Spearman's rank correlation coefficient was chosen because it is a nonparametric measure of association that is suitable for assessing the strength and direction of the monotonic relationship between two ranked variables, especially when assumptions of linearity and homoscedasticity are not met or when data are ordinal or have outliers. The correlation coefficient and its p value were calculated using the Himisc package in RStudio. The matrix was further visualized using corrplot in RStudio.

2.7. Linear Model Fitting without Time Lagging. Herein, a linear model was constructed to determine a mathematical relationship between a set of predictor data (which are FCM-derived data at time t) and responses [either the COD removal efficiency or methane production measured at the same time (i.e., t), respectively]. On the basis of the existing measured values of different variables and their correlative analysis in section 2.6, five FCM-derived variables (HNA abundance, LNA abundance, HNA/LNA ratio, log HNA, and log LNA) were identified to significantly affect the AnMBR performance (i.e., either COD removal efficiency or methane production, respectively), with all having $p < 0.05$, and thus were selected for the model development. Given that there is a temporal dependency between the variables, a time-series cross-validation method was performed herein. This method splits the data into multiple training and test sets based on a specific time point. In contrast to traditional cross-validation methods (e.g., k-fold cross-validation), time-series cross-validation could fully consider the theoretical problems with respect to temporal evolutionary effects and dependencies within the data as well as practical problems regarding missing values.[29] The data were partitioned into two subsets for training and test, with the condition that the test set is always ahead of the training set. At the first iteration, the initial training set comprised 20 samples × 5 variables and the test set contained 3 samples × 5 variables. Then, the linear model was fitted and evaluated through the training and test sets. For the subsequent iteration, the training set was gradually expanded with 1 step size to include more data points, while the test set synchronously moved forward but kept a constant number of data points. Repeated iterations were performed for the linear model fitting and evaluation until all data are covered in the model. To further determine model fitting, stepwise model selection based on the Akaike information criterion was applied using the MASS package in R.

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

To determine the accuracy of the regression model, different factors, such as the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE), were calculated using the tidyverse package in R. The definitions of R^2, RMSE, and MAE are further defined as follows:

(a) The coefficient of determination (or R^2) represents the proportion of the variance in the dependent variable, which is explained by the linear regression model.

$$R^2 = 1 - \frac{\sum (y_i - y_{\text{pred}})^2}{\sum (y_i - y_{\text{means}})^2} \quad (1)$$

(b) The RMSE represents the error associated with the model and can be computed as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - y_{\text{pred}})^2} \quad (2)$$

(c) The MAE represents the average of the absolute difference between the actual and predicted values in the data set. It measures the average of the residuals in the data set

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - y_{\text{pred}}| \quad (3)$$

where y_{pred} and y_{means} represent the model computed and measured values of the variable and N represents the number of observations.

2.8. Linear Model Fitting with Time Lagging. Considering that the reactor performance is highly reliant on functional microbiota,[5,6] there may be a temporal lag between microbial community variation and reactor performance response. Therefore, we attempted to develop the linear model by time lagging the reactor performance data (time t) with the FCM-derived data from samples obtained beforehand at lags 1 (3.5 days before time t, i.e., $t - 3.5$), 2 (7 days before time t, i.e., $t - 7$), and 3 (10.5 days before time t, i.e., $t - 10.5$) intervals in comparison to that without any time lag (i.e., lag 0). These data were refitted and evaluated on the basis of the procedures mentioned in section 2.7. After comparison of R^2, RSME, and MAE values, the best fitted model was proposed.

2.9. Model Verification. To verify the applicability of the proposed linear regression model, a separate 35 L AnMBR that was continuously operated to treat real municipal wastewater since Dec 21, 2020 was sampled on a routine basis to provide biologically independent data sets. The reactor was monitored in two separate stages: from Jan 12, 2021 to Nov 7, 2021 (i.e., stage 1) and from March 26, 2023 to May 15, 2023 (i.e., stage 2). Similar to section 2.1, four HRTs were applied during this new operational phase. The other operational parameters were kept consistent. The two stages maintained the same SRTs (i.e., 680 days). The suspension MLSS was evaluated. A total of 20 sludge samples in stage 1 and 18 sludge samples in stage 2 were assessed via FCM, as described in section 2.3. The FCM data sets, which were collected 3.5 days in advance of time t (i.e., lag time of t = 3.5 days) were input into the linear regression models to predict COD removal efficiency or methane production at time t. The coefficient of determination (R^2), RMSE, and MAE were collated to determine the accuracy of the model. The higher the value of R^2 but with lower RSME and MAE values, the more accurate the model would be. However, for the methane production, we failed to gather precise methane production in stage 1 to accurately verify the model as a result of gas leakage.

2.10. Contribution of Identified HNA and LNA Genera to the Reactor Performance. To evaluate the contribution of the identified HNA and LNA genera to the reactor performance, the relative and absolute abundance of these HNA and LNA genera were first correlated with the reactor performance (i.e., COD removal efficiency and methane production) through the Himisc package in RStudio. The absolute abundance was calculated on the basis of multiplying the FCM total cell count results with the relative abundance obtained from the 16S rRNA gene-based amplicon sequencing result.[30] In addition, the relative and absolute abundance of these HNA and LNA genera were, respectively, correlated to the reactor performance (i.e., COD removal efficiency and methane production) using the random forest model, a R package, with 70% of the samples randomly selected for the training set and the others used as the test set. From these full models, the individual genus was ranked by its importance in contributing to the accuracy of the COD removal efficiency and methane production prediction, respectively. This process is performed by permuting the relative abundance levels for a genus and calculating the increase in the mean-squared error (i.e., %incMSE) of the model. The %incMSE refers to the increase in the mean-squared error of predictions as a result of the variable (in this case, the relative abundance of particular genera) being permuted (with values randomly shuffled). A lower %incMSE indicates that the random permuted form of that variable worked better in the model, which then indicates that the said variable is not as important as others. Hence, the higher the value of %incMSE, the more important that variable would be.

2.11. Statistical Analysis. All significance testing was analyzed by either a two-tailed t test or one-way analysis of variance (ANOVA), available in Microsoft Excel. The null hypothesis was rejected at a confidence level of 95% (i.e., p value of ≤ 0.05).

3. RESULTS

3.1. Reactor Performance. With the decrease in HRT and, hence, increase in the level of the OLR, the efficiency of COD removal by the AnMBR increased accordingly (Figure 1B). The average COD removal efficiency was 83.6% at a HRT of 24 h and gradually increased to 87.8% at a HRT of 20 h. In contrast, the average COD removal efficiency was more than 93% at HRTs of 16 and 12 h (i.e., 93.2% and 95.9%, respectively). Further standard deviation analysis indicated that, at a HRT of 12 h, the variation in COD removal efficiency was the lowest (i.e., 0.0091; Figure S1A of the Supporting Information). In addition, the contribution of membrane filtration to COD removal was calculated on the basis of the COD difference between the suspension in the anaerobic digestor and final effluent (post-filtration). Membrane filtration contributed to less than 10% of the total COD removal (Figure S2 of the Supporting Information), and hence, the majority of the COD removal in this system is through anaerobic fermentation. In addition to COD removal, other water quality parameters (i.e., ammonium, nitrite, nitrate, and phosphate) were monitored (Figure S3 of the Supporting Information), and there was no considerable difference in ammonium and phosphate concentrations at different HRTs (ANOVA, with $p > 0.05$).

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

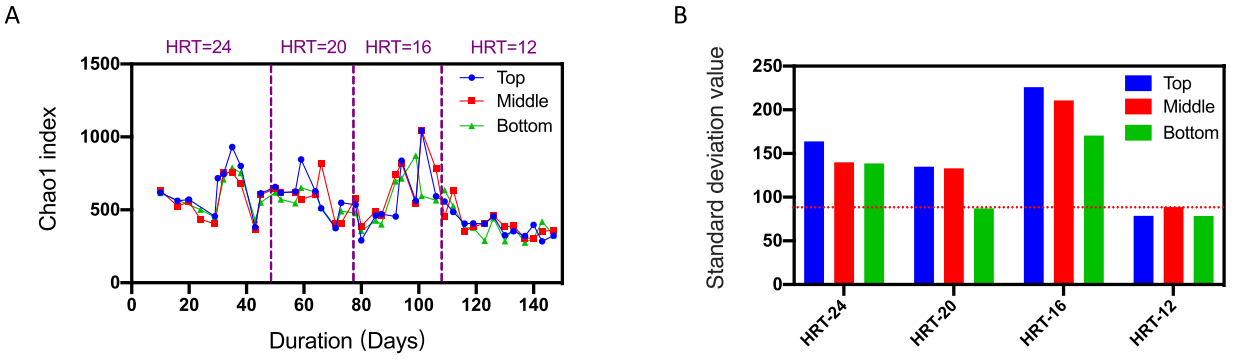


Figure 3. Microbial diversity of the anaerobic microbial consortium derived from 16S rRNA gene-based amplicon sequencing: (A) microbial richness represented on the basis of the Chao 1 index at different HRTs and (B) associated standard deviation in Chao 1 index values of each examined HRT.

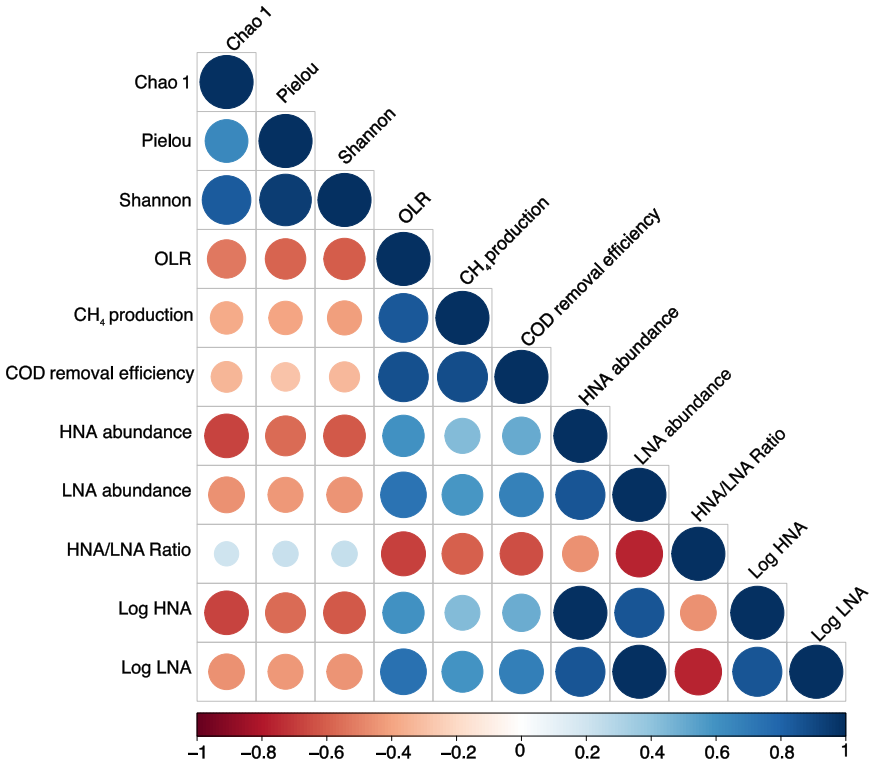


Figure 4. Spearman correlation matrix of the relationships among microbial diversity, FCM-derived fingerprint data, and reactor performance. The colors and sizes of the dot indicate the correlation coefficient. The darkest blue color indicates a perfect positive correlation ($r = 1$); the darkest red indicates a perfect negative correlation ($r = -1$); and colors of different gradients indicate a gradual loss in correlation from both spectrum ends.

Moreover, this study evaluated the biogas content and methane production at different HRTs. The biogas content result exhibited lower percentages of CH_4 and CO_2 at a HRT of 24 h; however, the content gradually increased to more than 70% in CH_4 and 10% in CO_2 as HRT decreased (Figure S4 of the Supporting Information). In accordance with COD removal efficiency, the average assessed CH_4 production accounted for 62.4% of the theoretical methane production in the initial stage (i.e., HRT of 24 h) and then increased to 69.9% at a HRT of 20 h, 86.9% at a HRT of 16 h, and finally 92.7% at a HRT of 12 h, respectively (Figure 1C). The variation in CH_4 volume production was the lowest at a HRT of 12 h compared to those at other HRTs (Figure S1B of the Supporting Information).

3.2. Variation of the FCM Fingerprint in HNA and LNA Cells. The abundance of HNA and LNA cells (in terms of $\log_{10}$ scale) was 7.8 and 7.0 at a HRT of 24 h and gradually increased to 7.9 and 7.1 at a HRT of 16 h, respectively. In comparison to the initial stage (i.e., HRT of 24 h), the logarithm of HNA and LNA cell abundance during the HRT period of 12 h considerably increased by 7.0% and 10.0% (t test, with $p = 0.05$ and 0.02, respectively). The gating regions of HNA and LNA in the flow cytogram are defined and exemplified in panels A−C of Figure 2. Although HNA and LNA log abundance increased with a decrease in HRT, a decrease in the HNA/LNA ratio was observed (Figure 2D). For example, the average HNA/LNA ratio was maintained at 6.8 and 7.1 for the first two examined HRTs (i.e., 24 and 20 h); however, the average ratio declined to 5.9 at a HRT of 16 h and then further declined to 2.9 at a HRT of 12 h (Figure 2D). This average HNA/LNA ratio at a HRT of 12 h was

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

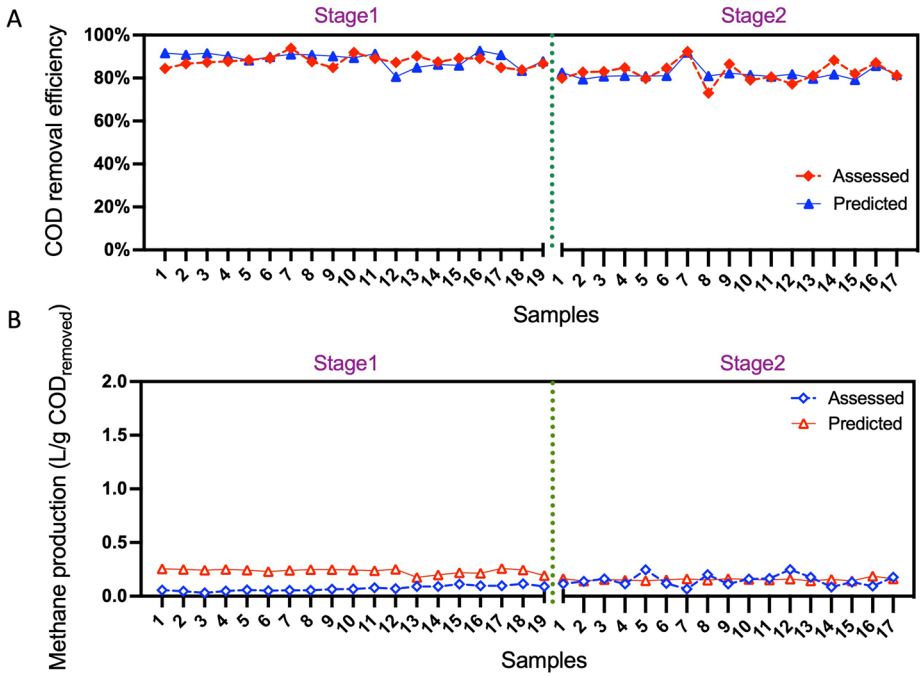


Figure 5. Comparison of the predicted and measured (A) COD removal efficiency and (B) methane production based on the corresponding proposed linear regression models with two separated stages.

significantly lower than the ratios noted at the other three HRTs (ANOVA, with $p < 0.01$).

3.3. Dynamics in Microbial Diversity. The α diversity of the total microbial community at each HRT period was calculated on the basis of the 16S rRNA gene-based amplicon sequencing results. The Chao 1 index ranged from 355 to 1044 among the samples collected at different time points of HRTs of 24, 20, and 16 h (Figure 3A). The standard deviations at HRTs of 24, 20, and 16 h (which considerably varied) were 147.4, 118.3, and 202.3, respectively (Figure 3B). In contrast, the Chao 1 index had an average value of 393 and exhibited a relatively lower variation across the different samples collected at a HRT of 12 h (Figure 3A). The standard deviations at HRTs of 24, 20, and 16 h were 1.8, 1.4, and 2.5 times that determined at 12 h. A similar result for the Pielou and Shannon indexes was obtained as well (Figure S5 of the Supporting Information).

3.4. Correlation among Microbial Diversity, Reactor Performance, and FCM-Derived Data. The HNA and LNA abundances (log scale) were negatively correlated with microbial diversity (i.e., Chao 1, Pielou, and Shannon indices). In addition, HNA and LNA abundances were positively correlated with the AnMBR performance (i.e., COD removal efficiency and CH_4 production). In contrast, the HNA/LNA ratio exhibited a significant negative correlation with reactor performance ($p < 0.05$; Figure 4).

3.5. Fitting of the Linear Regression Model. The results obtained via the correlation analysis described in section 3.4 indicated that the reactor performance was substantially correlated with the FCM-derived data. Thus, this study further used the linear regression model to explore the relationship between FCM-derived data and reactor performance (i.e., COD removal efficiency and CH_4 production, respectively). COD removal efficiency and CH_4 production were chosen as the representative reactor performance, whereas FCM-derived parameters were chosen as the variables. Before deciding on the linear regression model for further development, this study conducted three other kinds of regressions (i.e., ridge regression, lasso regression, and partial least squares regression), and the linear model exhibited best fitting over others (Figure S6 of the Supporting Information). Thus, only the linear model was validated and is discussed hereafter. Next, as a result of the potential time-lagging effect between microbial community variation and reactor performance response, this study lagged the reactor performance data with three iterations of the sample period. As indicated in Figure S7 of the Supporting Information, lag 1 iteration (i.e., lag 3.5 days) exhibited the best fitting, with a maximal R^2 value (i.e., 0.57 and 0.45 in COD removal efficiency and CH_4 production, respectively) but with minimal RSME and MAE values (Figure S7 of the Supporting Information).

Hence, two linear models (with lag 1) were proposed after fitting (i.e., COD removal efficiency and CH_4 production, respectively).

For COD removal efficiency, after stepwise analysis and optimization, two parameters remained as variables of significance, and the linear model can be expressed as

$$\text{COD removal efficiency}_t = a + b1 \times \log \text{LNA}_{t-3.5} + b2 \times \text{HNA/LNA ratio}_{t-3.5} \quad (4)$$

where COD removal efficiency denotes the dependent variable, a denotes the intercept, and $b1$ and $b2$ denote the polynomial coefficients.

Further calculation was performed to determine the best fit and constant factors. This resulted in the equation as follows:

$$\text{COD removal efficiency}_t = 0.49 + 0.067 \times \log \text{LNA}_{t-3.5} - 0.0038 \times \text{HNA/LNA ratio}_{t-3.5} \quad (5)$$

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

Table 1. HNA and LNA Genera Identified in This Study[a]

		COD removal efficiency						CH_4 production					
		absolute abundance			relative abundance			absolute abundance			relative abundance		
genus	classification	correlation	*p* value	%incMSE	correlation	*p* value	%incMSE	correlation	*p* value	%incMSE	correlation	*p* value	%incMSE
Anaerovibrio	HNA	**0.74**	**0.00**	3.39	**0.84**	**0.00**	4.56	**0.66**	**0.00**	1.07	**0.79**	**0.00**	1.53
unclassified Veillonellaceae	HNA	0.15	0.47	0.94	−0.78	0.00	4.49	0.08	0.21	1.25	−0.71	0.00	1.56
Enterococcus	HNA	0.40	0.29	1.33	−0.44	0.00	3.21	0.12	0.00	1.10	−0.49	0.00	0.98
Elizabethkingia	HNA	−0.29	0.23	4.52	−0.61	0.00	3.07	−0.14	0.01	4.98	−0.56	0.00	6.53
unclassified Rikenellaceae	HNA	−0.11	0.34	3.03	−0.73	0.00	2.97	−0.11	0.34	0.03	−0.66	0.00	5.29
unclassified Enterobacteriaceae	HNA	−0.79	0.00	8.84	−0.78	0.00	2.24	−0.80	0.00	7.81	−0.73	0.00	7.44
Burkholderiales bacterium Beta_02	HNA	−0.01	0.98	0.31	−0.55	0.00	2.17	0.00	0.92	0.31	−0.40	0.00	1.89
HN-HFO75	HNA	−0.01	0.89	2.58	−0.23	0.01	2.12	−0.02	0.96	1.31	−0.13	0.16	1.38
unclassified Bacteroidales	HNA	**0.81**	**0.00**	1.61	**0.76**	**0.00**	2.08	**0.74**	**0.00**	3.01	**0.74**	**0.00**	1.89
Leucobacter	HNA	0.36	0.04	1.69	−0.42	0.00	1.92	0.23	0.00	1.74	−0.42	0.00	−0.70
Lactococcus	HNA	−0.01	0.00	0.99	−0.68	0.00	1.89	−0.33	0.93	−0.19	−0.75	0.00	1.30
Prevotella 9	HNA	−0.53	0.00	4.22	−0.62	0.00	1.85	−0.51	0.00	1.25	−0.58	0.00	−1.60
Achromobacter	LNA	**0.41**	**0.00**	2.18	**0.19**	**0.04**	1.57	**0.50**	**0.00**	3.79	**0.29**	**0.00**	1.36
Clostridium sensu stricto 1	HNA	0.61	0.00	0.74	−0.31	0.00	1.51	0.44	0.00	1.09	−0.39	0.00	2.18
Bacteroides	HNA	−0.19	0.78	1.17	−0.62	0.00	1.39	−0.03	0.10	2.30	−0.54	0.00	1.11
Stenotrophomonas	LNA	−0.28	0.20	1.78	−0.44	0.00	1.39	−0.15	0.02	−1.38	−0.43	0.00	2.24
Clostridium sensu stricto 11	HNA	0.14	0.64	−0.93	−0.51	0.00	1.33	0.05	0.24	1.57	−0.52	0.00	3.57
unclassified Rikenellaceae	HNA	0.01	0.21	0.32	−0.68	0.00	1.18	−0.15	0.93	3.28	−0.67	0.00	0.91
unclassified Anaerolineae	LNA	**0.59**	**0.00**	4.07	**0.52**	**0.00**	1.00	**0.38**	**0.00**	1.54	**0.50**	**0.00**	0.25
Clostridium sensu stricto 10	HNA	0.45	0.00	1.15	0.10	0.27	0.78	0.49	0.00	1.16	0.26	0.00	3.97
Sphingobacterium	HNA	−0.15	0.50	−0.13	−0.59	0.00	0.71	0.08	0.20	1.51	−0.48	0.00	4.28
Prevotella	HNA	0.14	0.73	−1.27	**0.19**	**0.04**	0.42	−0.04	0.23	1.42	0.07	0.43	1.86
Acinetobacter	HNA	0.31	0.00	1.51	−0.56	0.00	0.41	0.45	0.01	−0.49	−0.45	0.00	0.11
Pseudomonas	LNA	0.09	0.40	1.53	−0.46	0.00	0.34	−0.10	0.46	1.79	−0.54	0.00	4.37
Ochrobactrum	LNA	**0.69**	**0.00**	3.63	**0.70**	**0.00**	0.34	**0.64**	**0.00**	2.65	**0.71**	**0.00**	−1.15
Christensenellaceae R-7 group	HNA	0.20	0.97	0.91	−0.63	0.00	0.29	0.00	0.08	0.60	−0.61	0.00	−0.02
Flexilinea	HNA	0.04	0.53	0.28	−0.55	0.00	0.18	−0.07	0.73	−1.41	−0.50	0.00	1.07
Alcaligenes	LNA	**0.43**	**0.00**	1.02	**0.23**	**0.01**	0.00	**0.46**	**0.00**	1.94	**0.28**	**0.00**	0.00
Thauera	LNA	0.14	0.11	−0.04	0.08	0.38	0.00	0.18	0.21	1.15	0.12	0.22	0.00
unclassified Micrococcales	LNA	0.01	NA		0.03	0.72	0.00	NA	NA	NA	0.03	0.78	0.00
unclassified Selenomonadales	HNA	−0.02	0.30	2.31	−0.51	0.00	−0.18	−0.12	0.86	1.67	−0.52	0.00	0.94
Comamonas	HNA	0.21	0.00	0.18	−0.39	0.00	−0.35	0.41	0.06	3.57	−0.32	0.00	5.26
Rhodococcus	LNA	0.28	0.12	−0.95	0.14	0.13	−1.02	**0.18**	**0.01**	1.04	**0.20**	**0.03**	−0.85
Pseudochrobactrum	HNA	0.20	0.02	0.54	−0.55	0.00	−1.42	0.26	0.08	0.69	−0.47	0.00	0.69

[a]Each genus is demonstrated with their corresponding correlation value and their importance to COD removal efficiency or CH_4 production. Their importance in COD removal efficiency or CH_4 production is indicated by the %incMSE value; the higher the %incMSE value, the more importance associated with that particular genus.

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

For this model, a R^2 value of 0.57 was obtained, indicating that approximately 57% of the dependent variable (i.e., y) could be explained by x.[31] The RMSE and MAE values were 0.0040 and 0.0038, respectively.

Similarly, for methane production, after stepwise analysis and optimization, only two parameters remained as variables of significance, and the linear model can be expressed as

$$\text{CH}_4\ \text{production}_t = a + b1 \times \log \text{LNA}_{t-3.5} + b2 \times \text{HNA/LNA ratio}_{t-3.5} \quad (6)$$

where CH_4 production denotes the dependent variable, a denotes the intercept, and $b1$ and $b2$ denote the polynomial coefficients.

Further calculations were performed to determine the best fit and constant factors. This resulted in the equation as follows:

$$\text{CH}_4\ \text{production}_t = -0.0067 + 0.042 \times \log \text{LNA}_{t-3.5} - 0.0050 \times \text{HNA/LNA ratio}_{t-3.5} \quad (7)$$

For this model, a R^2 value of 0.45 was obtained. The RMSE and MAE values were 0.0071 and 0.0069, respectively.

3.6. Verification of the Lag 1 Fitted Linear Model. To further verify that the proposed linear regression model could predict one of the two reactor performance indicators, FCM-derived data (at time $t - 3.5$) obtained from another independently operated pilot-scale AnMBR fed with real municipal wastewater at two operational stages were input into the model to predict the COD removal efficiency (t) in a scale of 0−100%. As depicted in Figure S8 of the Supporting Information, the MLSS content in the AnMBR operated in stage 1 ranged from 320 to 950 mg/L. In contrast, the MLSS content in the AnMBR operated in stage 2 was higher than that at stage 1 and ranged from 1800 to 3770 mg/L (Figure S8A of the Supporting Information). In addition, the water quality (i.e., ammonium, nitrite, nitrate, and phosphate) of permeate remained stable during sampling (Figure S8B of the Supporting Information). Despite differences in the MLSS content, as shown in Figure 5A, the predicted values approximated the tested values of the COD removal efficiency for both stages. Further calculation indicated that the predicted values (i.e., 89.1 ± 3.3%) in stage 1 were not significantly different from the observed measurements (i.e., 87.9 ± 2.5%; ANOVA, with $p = 0.23$). Furthermore, the R^2, RMSE, and MAE values were 0.72, 0.048, and 0.041, respectively, indicating that the model could well predict the COD removal efficiency in this case (Figure 5A and Figure S8C of the Supporting Information). A similar observation could be obtained in stage 2.

However, for methane production, we failed to gather precise methane production to accurately verify the model as a result of potential gas leakage in stage 1. The observed methane production was 0.073 ± 0.024 L/g of $COD_{removed}$, which is much lower than the predicted methane production based on the proposed model (i.e., 0.23 ± 0.023 L/g of $COD_{removed}$; Figure 5B). Correspondingly, the R^2, RMSE, and MAE values were 0.20, 0.17, and 0.16, respectively. In contrast, at stage 2, the predicted values (i.e., 0.16 ± 0.011 L/g of $COD_{removed}$) exhibited no significant difference with the observed measurements (i.e., 0.15 ± 0.052 L/g of $COD_{removed}$; ANOVA, with $p = 0.56$), indicating that the model could well predict the methane production in this case (Figure 5B).

3.7. Contribution of HNA and LNA to the Reactor Performance. Considering the correlation between HNA and LNA abundance with the AnMBR performance, this study sorted the HNA and LNA cells via FCM and sequenced the sorted cells to determine their identities. Multivariate analysis revealed that HNA and LNA microbiota were considerably different after cell sorting (Figure S9 of the Supporting Information). Further comparison indicated that there were 34 genera identified in this study, including 25 HNA and 9 LNA cells, and no syntrophs or methanogens were identified (Table 1). As summarized in Table 1, all of the LNA cells (in both absolute and relative abundance) except *Pseudomonas* and *Stenotrophomonas* positively correlated with COD removal efficiency and methane production. For instance, unclassified Anaerolineae and *Ochrobactrum* were strongly and positively correlated with COD removal efficiency and methane production ($R > 0.5$; $p < 0.01$) and with %incMSE value of >0. In contrast to LNA, 80% of HNA cells (i.e., 20 HNA cells/25 HNA cells) were negatively correlated with the COD removal efficiency (Table 1). This is with the exception of *Anaerovibrio* and unclassified Bacteroidales, which were significantly correlated with COD removal efficiency and methane production (all $p < 0.05$). Consistently, random forest analysis indicated that, among HNA, *Anaerovibrio* and unclassified Bacteroidales contributed to the COD removal efficiency and methane production (Table 1).

4. DISCUSSION

A healthy and functioning microbiome, particularly one that plays a keystone role in hydrolysis, fermentation, and methanogenesis, is considered crucial to the overall performance of AnMBRs in removing COD and generating methane as an energy source.[32,33] Flow cytometry can rapidly profile the dynamics of microbial communities in relation to different operational parameters or disturbances. To exemplify this, Dhoble et al. utilized FCM to detect autofluorescence emitted by microbial cells in two anaerobic digesters (ADs), each operated with different HRTs. On the basis of clustering analysis, Dhoble et al. differentiated microbial profiles in both reactors.[34] Similarly, De Vrieze et al. compared 16S rRNA gene-based amplicon sequencing, metaproteomics, metabolomics, and flow cytomics and obtained a similar clustering pattern of the microbiomes using the four approaches.[35] This indicates that FCM can be used for faster process monitoring than other approaches (i.e., 16S rRNA gene-based amplicon sequencing, metaproteomics, and metabolomics) without compromising the accuracy specifically, because the latter approaches would require various forms of sample processing (e.g., extraction, library preparation, sequencing, and bioinformatics), which are more time-consuming than that required by FCM.

Despite showing the overall feasibility of FCM as a monitoring tool for AD, the previous study did not attempt to correlate the FCM-derived data with the functional performance of the reactors and did not use these data to predict the reactor performance in advance. To address this knowledge gap, this study focused on differentiating and enumerating two groups, the HNA and LNA cells, as the HRT of the AnMBR varied. As the HRT decreased and the reactor trended toward better stability in terms of COD removal and methane production, the abundance of HNA and LNA

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

increased; however, the ratio of HNA/LNA cells decreased (Figures 1 and 2). Furthermore, the abundance of HNA and LNA negatively correlated with α diversity (Figure 4), indicating the enrichment of selected keystone groups, particularly that of the LNA groups, which would contribute more toward the performance of the AnMBR. These observations agreed with an earlier metadata analysis, which indicated that the presence of 20 core microbial genera examined from 13 different AnMBRs potentially accounted for the COD removal in AnMBRs.[6]

The LNA group was increasing at a higher rate than the HNA population when the AnMBR trended toward better performance, suggesting that the LNA group may play a more important role than the HNA group in COD removal within the AnMBR. In aquatic systems, the densities of the HNA group were observed to correlate positively with heterotrophic productivity (i.e., incorporation of C into biomass over time);[14,16,17] hence, it appeared counterintuitive that the LNA in the AnMBR was contributing more toward COD removal instead of the HNA. However, most of our prior knowledge regarding the impact of the HNA and LNA groups in response to a treatment factor or stressor was relatively well-characterized for the aquatic systems[13,21,36] but less so for AnMBR systems. It may be possible that the current known roles and phylogenetic identities associated with the LNA group in aquatic systems are not the same as those for the anaerobic fermentation process.

We further sorted HNA and LNA cells to sequence for their phylogenetic identities (Figure S9 of the Supporting Information and Table 1). The result indicated that 25 and 9 genera were assigned to HNA and LNA, respectively. Specifically, there were two HNA genera (i.e., *Anaerovibrio* and unclassified Bacteroidales) and four LNA genera (i.e., *Achromobacter*, *Ochrobactrum*, unclassified Anaerolineae, and *Alcaligenes*), which were consistently positively correlated with COD removal efficiency and methane production (Table 1). *Anaerovibrio* consists of species that are lipid hydrolytic bacteria (e.g., glycerides) and/or fermenters of glycerol.[37] Moreover, *Achromobacter* was reported to produce the cell wall hydrolytic enzyme,[38] and certain species of *Ochrobactrum* are biodelignified bacteria that degrade lignin to release trapped cellulose and hemicellulose.[39] For unclassified Anaerolineae, Cai et al. found that some unknown genera of Anaerolineae possessed the metabolic capacity of carbohydrate hydrolysis and proteolysis.[40] Although *Alcaligenes* exhibited a weak correlation with COD removal efficiency and methane production, the random forest model analysis indicated that *Alcaligenes* in absolute cell number contributed to the COD removal efficiency and methane production (i.e., %incMSE = 1.02 and 1.94, respectively; Table 1). It is assumed that hydrolysis is the rate-limiting step of anaerobic digestion,[41] which may explain the reason for higher COD removal efficiency and methane production with an increase in the abundance of hydrolytic LNA populations along with an increase in the HNA fermenter populations (Figures 1 and 2). Moreover, this study observed that the genus *Prevotella* positively correlated with only the COD removal efficiency but not methane production (Table 1). *Prevotella* is a lignocellulolytic bacteria, which can digest lignocellulosic biomass into volatile fatty acids (VFAs) through producing cellulolytic enzymes;[42] however, it would reduce the amount of hydrogen available for methanogenesis, mitigating methane production.[43]

Given the importance of HNA and LNA populations in the AnMBR performance, this study subsequently conducted a linear regression model fitting to predict the AnMBR performance (i.e., COD removal efficiency and methane production, respectively) using 72 sets of FCM-derived data. An earlier study used 21−185 sets of data, which comprised pH, temperature, influent COD, and flux, to train a deep learning model that aimed to predict the reactor performance.[44] However, the previous study did not demonstrate that the model could predict the reactor performance in advance. Herein, we used FCM-derived data as an input data set for the predictive model. Considering that there is most likely a time lag between any changes in the microbial community and the corresponding reactor performance, this allows us to develop a model that incorporates the time component. Specifically, we found that our proposed model could predict the reactor performance 3.5 days in advance (Figure S7 of the Supporting Information). However, owing to the limitation of sampling frequency (i.e., twice per week), the prediction time may not be exactly 3.5 days and would need to be further verified. In addition, although the proposed model predicted the COD removal efficiency of AnMBR, the model only used FCM-derived data that were obtained when AnMBR was operating in a stable steady state. It is uncertain if the model can predict the reactor performance in advance when sudden aberrations occur, which can drastically change the HNA and LNA ratio. Further study would be needed to validate the accuracy of this model in predicting such fluctuating conditions. Lastly, considering different treatment efficacies and, hence, microbial communities that might arise in pilot- and full-scale AnMBR reactors, further studies would be needed to verify the applicability of our predictive model on larger or even full-scale AnMBR systems.

Jeong et al. proposed deep learning models using 16 operating parameters (e.g., SRT, temperature, suspended solid concentration, etc.) to predict biogas production in the anaerobic co-digestion of organic wastes, and the previous study predicted a coefficient of determination (R^2) of 0.68.[45] The lower R^2 value obtained in this study (i.e., 0.45) might be partly due to the limited data set. In addition, different input data and missing FCM-derived data related to syntrophs and methanogens could also result in a low R^2 value. The previous study indicated that methanogens and syntrophs considerably contribute to COD removal and methane production.[46] However, herein, no methanogens or syntrophs were identified as HNA or LNA cells (Table 1). Because methanogens and syntrophs are present in low abundance in an AD system,[47] FCM may not have the needed detection sensitivity to detect these low-abundant populations if the samples were stained with a non-specific fluorescence reporter like that of SYBR Green or 4′,6-diamidino-2-phenylindole (DAPI). Instead, because methanogens emit autofluorescence as a result of cofactor F_{420}, which is an essential hydride carrier in hydrogenotrophic methane synthesis,[48−50] Lambrecht et al. demonstrated the use of autofluorescence-based FCM to enumerate methanogenic archaea. The methanogens were denoted to be present in a range from 3.7×10^8 to 1.8×10^9 cells/mL in methanogenic enrichment cultures and digester samples,[51] suggesting the feasibility to use FCM to detect them based on autofluorescence. Such approaches may not be applicable for low-abundant syntrophs, because they are not known to emit autofluorescence. Further studies would have to

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372

be conducted to enumerate the autofluorescence portions via FCM and evaluate their correlation to CH_4 production.

This study demonstrated the use of FCM to monitor HNA and LNA populations, which were positively correlated with the AnMBR performance (i.e., COD removal efficiency and methane production). Cell sorting, 16S rRNA gene-based amplicon sequencing, and correlation analysis further determined HNA genera (i.e., *Anaerovibrio* and unclassified Bacteroidales) and LNA genera (i.e., *Achromobacter*, *Ochrobactrum*, and unclassified Anaerolineae) to substantially contribute to COD removal and methane production. The use of LNA log abundance and the ratio of HNA/LNA abundance in a linear regression model enabled the prediction of COD removal efficiency and methane production approximately 3.5 days in advance. Overall, this study demonstrated the feasibility of using FCM-derived data to monitor and predict the AnMBR performance.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.3c07702.

Additional supplementary table that denotes composition of synthetic wastewater feed (Table S1) and additional supplementary figures that denote the associated standard deviation at each examined HRT (Figure S1), contribution of the membrane toward COD removal (Figure S2), water quality of the pilot-scale AnMBR permeate at different HRTs (Figure S3), composition of the biogas produced by the AnMBR (Figure S4), microbial richness (Figure S5), coefficient of determination (R^2) through different regression models (Figure S6), composition of the linear regression models with different lag iterations (Figure S7), characterizations of the validated pilot AnMBR performance (Figure S8), and principal coordinate analysis of the HNA and LNA microbiota (Figure S9) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Pei-Ying Hong − *Environmental Science and Engineering Program, Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; Water Desalination and Reuse Center (WDRC), Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia;* orcid.org/0000-0002-4474-6600; Phone: +966-12-8082218; Email: peiying.hong@kaust.edu.sa

Hong Cheng − *Key Laboratory of Eco-environments in Three Gorges Reservoir Region, Ministry of Education, College of Environment and Ecology, Chongqing University, Chongqing 400044, People's Republic of China; Environmental Science and Engineering Program, Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia;* orcid.org/0000-0002-4943-9060; Phone: +86-23-65120980; Email: hong.cheng@cqu.edu.cn

Authors

Julie Sanchez Medina − *Environmental Science and Engineering Program, Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; Water Desalination and Reuse Center (WDRC), Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*

Jianqiang Zhou − *Environmental Science and Engineering Program, Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; State Power Investment Corporation Research Institute, Beijing 102209, People's Republic of China*

Eduardo Machado Pinho − *Environmental Science and Engineering Program, Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; Department of Bioengineering, Faculty of Engineering, University of Porto, 4099-002 Porto, Portugal*

Rui Meng − *Lawrence Berkeley National Laboratory, Berkeley, California 94301, United States; Amazon, Incorporated, Palo Alto, California 94301, United States*

Liuwei Wang − *Systems Medicine of Infectious Disease (P5), Robert Koch Institute, 13353 Berlin, Germany; Department of Mathematics and Computer Science, Freie Universität Berlin, 10117 Berlin, Germany*

Qiang He − *Key Laboratory of Eco-environments in Three Gorges Reservoir Region, Ministry of Education, College of Environment and Ecology, Chongqing University, Chongqing 400044, People's Republic of China*

Xosé Anxelu G. Morán − *Red Sea Research Center, Biological and Environmental Science & Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.3c07702

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study is supported by KAUST WDRC CCF funding FCC/1/1971-32-01 awarded to Pei-Ying Hong. The authors thank Dr. Luis Ribeiro da Silva for his guidance on flow cytometry and gating analysis and Dr. Haleem Shah for his technical assistance in flow cytometry during the COVID-19 lab lockdown period. The authors are grateful to the KAUST Red Sea Research Center for granting access to flow cytometry.

■ REFERENCES

(1) Lew, B.; Tarre, S.; Beliavski, M.; Dosoretz, C.; Green, M. Anaerobic membrane bioreactor (AnMBR) for domestic wastewater treatment. *Desalination* **2009**, *243* (1−3), 251−257.

(2) Ozgun, H.; Dereli, R. K.; Ersahin, M. E.; Kinaci, C.; Spanjers, H.; van Lier, J. B. A review of anaerobic membrane bioreactors for municipal wastewater treatment: Integration options, limitations and expectations. *Sep. Purif. Technol.* **2013**, *118*, 89−104.

(3) Martin, I.; Pidou, M.; Soares, A.; Judd, S.; Jefferson, B. Modelling the energy demands of aerobic and anaerobic membrane bioreactors for wastewater treatment. *Environ. Technol.* **2011**, *32* (9), 921−932.

(4) Zaouri, N.; Cheng, H.; Khairunnisa, F.; Alahmed, A.; Blilou, I.; Hong, P.-Y. A type dependent effect of treated wastewater matrix on seed germination and food production. *Sci. Total Environ.* **2021**, *769*, 144573.
(5) Cheng, H.; Hong, P.-Y. Nanoparticles applied in membrane bioreactors: Potential impact on reactor performance and microbial communities. *Contaminants of Emerging Concern in Water and Wastewater*; Elsevier: Amsterdam, Netherlands, 2020; Chapter 7, pp 207−236, DOI: 10.1016/B978-0-12-813561-7.00007-9.
(6) Cheng, H.; Cheng, D.; Mao, J.; Lu, T.; Hong, P.-Y. Identification and characterization of core sludge and biofilm microbiota in anaerobic membrane bioreactors. *Environ. Int.* **2019**, *133*, 105165.
(7) Jiang, Q.; Fu, B.; Chen, Y.; Wang, Y.; Liu, H. Quantification of viable but nonculturable bacterial pathogens in anaerobic digested sludge. *Appl. Microbiol. Biotechnol.* **2013**, *97* (13), 6043−6050.
(8) Besmer, M. D.; Weissbrodt, D. G.; Kratochvil, B. E.; Sigrist, J. A.; Weyland, M. S.; Hammes, F. The feasibility of automated online flow cytometry for in-situ monitoring of microbial dynamics in aquatic ecosystems. *Front. Microbiol.* **2014**, *5*, 265.
(9) Prest, E.; Hammes, F.; Kötzsch, S.; van Loosdrecht, M. C.; Vrouwenvelder, J. S. Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Res.* **2013**, *47* (19), 7131−7142.
(10) Props, R.; Rubbens, P.; Besmer, M.; Buysschaert, B.; Sigrist, J.; Weilenmann, H.; Waegeman, W.; Boon, N.; Hammes, F. Detection of microbial disturbances in a drinking water microbial community through continuous acquisition and advanced analysis of flow cytometry data. *Water Res.* **2018**, *145*, 73−82.
(11) Harry, I. S. K.; Ameh, E.; Coulon, F.; Nocker, A. Impact of treated sewage effluent on the microbiology of a small brook using flow cytometry as a diagnostic tool. *Water, Air, Soil Pollut.* **2016**, *227* (2), 57.
(12) Gasol, J. M.; Zweifel, U. L.; Peters, F.; Fuhrman, J. A.; Hagstrom, A. Significance of size and nucleic acid content heterogeneity as measured by flow cytometry in natural planktonic bacteria. *Appl. Environ. Microbiol.* **1999**, *65* (10), 4475−4483.
(13) Vila-Costa, M.; Gasol, J. M.; Sharma, S.; Moran, M. A. Community analysis of high-and low-nucleic acid-containing bacteria in NW Mediterranean coastal waters using 16S rDNA pyrosequencing. *Environ. Microbiol.* **2012**, *14* (6), 1390−1402.
(14) Bouvier, T.; Del Giorgio, P. A.; Gasol, J. M. A comparative study of the cytometric characteristics of high and low nucleic-acid bacterioplankton cells from different aquatic ecosystems. *Environ. Microbiol.* **2007**, *9* (8), 2050−66.
(15) Zolkefli, N.; Ramli, N.; Mohamad-Zainal, N. S. L.; Mustapha, N. A.; Yusoff, M. Z. M.; Hassan, M. A.; Maeda, T. Alcaligenaceae and Chromatiaceae as pollution bacterial bioindicators in palm oil mill effluent (POME) final discharge polluted rivers. *Ecol. Indic.* **2020**, *111*, 106048.
(16) Zubkov, M. V.; Fuchs, B. M.; Burkill, P. H.; Amann, R. Comparison of cellular and biomass specific activities of dominant bacterioplankton groups in stratified waters of the Celtic Sea. *Appl. Environ. Microbiol.* **2001**, *67* (11), 5210−5218.
(17) Belzile, C.; Brugel, S.; Nozais, C.; Gratton, Y.; Demers, S. Variations of the abundance and nucleic acid content of heterotrophic bacteria in Beaufort Shelf waters during winter and spring. *J. Mar. Syst.* **2008**, *74* (3−4), 946−956.
(18) Nopens, I.; Capalozza, C.; Vanrolleghem, P. A. *Stability Analysis of a Synthetic Municipal Wastewater*; Department of Applied Mathematics Biometrics and Process Control, University of Gent, Gent, Belgium, 2001.
(19) Al-Otaibi, N.; Huete-Stauffer, T. M.; Calleja, M. L.; Irigoien, X.; Morán, X. A. G. Seasonal variability and vertical distribution of autotrophic and heterotrophic picoplankton in the Central Red Sea. *PeerJ.* **2020**, *8*, No. e8612.
(20) Gasol, J. M.; Morán, X. A. G. Flow cytometric determination of microbial abundances and its use to obtain indices of community structure and relative activity. In *Hydrocarbon and Lipid Microbiology Protocols*; McGenity, T. J., Timmis, K. N., Nogales, B., Eds.; Springer: Berlin, Germany, 2015; pp 159−187, DOI: 10.1007/8623_2015_139.
(21) Wang, Y.; Hammes, F.; Boon, N.; Chami, M.; Egli, T. Isolation and characterization of low nucleic acid (LNA)-content bacteria. *ISME J.* **2009**, *3* (8), 889−902.
(22) Cheng, H.; Zhou, J.; Hong, P.-Y. Attached-growth configuration outperforms continuously stirred tank anaerobic membrane bioreactors in alleviating membrane biofouling. *Environ. Res.* **2021**, *199*, 111272.
(23) Hong, P.-Y.; Wheeler, E.; Cann, I. K.; Mackie, R. I. Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA-based pyrosequencing. *ISME J.* **2011**, *5* (9), 1461−1470.
(24) 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. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7* (5), 335−336.
(25) Callahan, B. J.; McMurdie, P. J.; Rosen, M. J.; Han, A. W.; Johnson, A. J. A.; Holmes, S. P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13* (7), 581−583.
(26) Bokulich, N. A.; Kaehler, B. D.; Rideout, J. R.; Dillon, M.; Bolyen, E.; Knight, R.; Huttley, G. A.; Caporaso, J. G. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **2018**, *6* (1), 90.
(27) Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F. O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2012**, *41* (D1), D590−D596.
(28) Loiseau, N.; Gaertner, J. C. Indices for assessing coral reef fish biodiversity: The need for a change in habits. *Ecol. Evol.* **2015**, *5* (18), 4018−4027.
(29) Bergmeir, C.; Benítez, J. M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **2012**, *191*, 192−213.
(30) Gloor, G. B.; Macklaim, J. M.; Pawlowsky-Glahn, V.; Egozcue, J. J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **2017**, *8*, 2224.
(31) Singh, K. P.; Basant, N.; Malik, A.; Jain, G. Modeling the performance of "up-flow anaerobic sludge blanket" reactor based wastewater treatment plant using linear and nonlinear approaches—a case study. *Anal. Chim. Acta* **2010**, *658* (1), 1−11.
(32) Werner, J. J.; Knights, D.; Garcia, M. L.; Scalfone, N. B.; Smith, S.; Yarasheski, K.; Cummings, T. A.; Beers, A. R.; Knight, R.; Angenent, L. T. Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (10), 4158−4163.
(33) Briones, A.; Raskin, L. Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *Curr. Opin. Biotechnol.* **2003**, *14* (3), 270−276.
(34) Dhoble, A. S.; Bekal, S.; Dolatowski, W.; Yanz, C.; Lambert, K. N.; Bhalerao, K. D. A novel high-throughput multi-parameter flow cytometry based method for monitoring and rapid characterization of microbiome dynamics in anaerobic systems. *Bioresour. Technol.* **2016**, *220*, 566−571.
(35) De Vrieze, J.; Heyer, R.; Props, R.; Van Meulebroek, L.; Gille, K.; Vanhaecke, L.; Benndorf, D.; Boon, N. Triangulation of microbial fingerprinting in anaerobic digestion reveals consistent fingerprinting profiles. *Water Res.* **2021**, *202*, 117422.
(36) Props, R.; Schmidt, M. L.; Heyse, J.; Vanderploeg, H. A.; Boon, N.; Denef, V. J. Flow cytometric monitoring of bacterioplankton phenotypic diversity predicts high population-specific feeding rates by invasive dreissenid mussels. *Environ. Microbiol.* **2018**, *20* (2), 521−534.

(37) Schauder, R.; Schink, B. Anaerovibrio glycerini sp. nov., an anaerobic bacterium fermenting glycerol to propionate, cell matter, and hydrogen. *Arch. Microbiol.* **1989**, *152* (5), 473−478.
(38) Li, S.; Norioka, S.; Sakiyama, F. Purification, characterization, and primary structure of a novel cell wall hydrolytic amidase, CwhA, from Achromobacter lyticus. *J. Biochem.* **2000**, *127* (6), 1033−1039.
(39) Tsegaye, B.; Balomajumder, C.; Roy, P. Biodegradation of wheat straw by Ochrobactrum oryzae BMP03 and Bacillus sp. BMP01 bacteria to enhance biofuel production by increasing total reducing sugars yield. *Environ. Sci. Pollut. Res.* **2018**, *25* (30), 30585−30596.
(40) Cai, C.; Li, L.; Hua, Y.; Liu, H.; Dai, X. Ferroferric oxide promotes metabolism in Anaerolineae other than microbial syntrophy in anaerobic methanogenesis of antibiotic fermentation residue. *Sci. Total Environ.* **2021**, *758*, 143601.
(41) Li, Y.; Chen, Y.; Wu, J. Enhancement of methane production in anaerobic digestion process: A review. *Appl. Energy* **2019**, *240*, 120−137.
(42) Nguyen, L. N.; Nguyen, A. Q.; Johir, M. A. H.; Guo, W.; Ngo, H. H.; Chaves, A. V.; Nghiem, L. D. Application of rumen and anaerobic sludge microbes for bio harvesting from lignocellulosic biomass. *Chemosphere* **2019**, *228*, 702−708.
(43) Aguilar-Marin, S. B.; Betancur-Murillo, C. L.; Isaza, G. A.; Mesa, H.; Jovel, J. Lower methane emissions were associated with higher abundance of ruminal Prevotella in a cohort of Colombian buffalos. *BMC Microbiol.* **2020**, *20* (1), 364.
(44) Li, G.; Ji, J.; Ni, J.; Wang, S.; Guo, Y.; Hu, Y.; Liu, S.; Huang, S.-F.; Li, Y.-Y. Application of deep learning for predicting the treatment performance of real municipal wastewater based on one-year operation of two anaerobic membrane bioreactors. *Sci. Total Environ.* **2022**, *813*, 151920.
(45) Jeong, K.; Abbas, A.; Shin, J.; Son, M.; Kim, Y. M.; Cho, K. H. Prediction of biogas production in anaerobic co-digestion of organic wastes using deep learning models. *Water Res.* **2021**, *205*, 117697.
(46) Adekunle, K. F.; Okolie, J. A. A review of biochemical process of anaerobic digestion. *Adv. Biosci. Biotechnol.* **2015**, *6* (3), 205.
(47) Mei, R.; Nobu, M. K.; Narihiro, T.; Yu, J.; Sathyagal, A.; Willman, E.; Liu, W.-T. Novel Geobacter species and diverse methanogens contribute to enhanced methane production in media-added methanogenic reactors. *Water Res.* **2018**, *147*, 403−412.
(48) Graham, D. E.; White, R. H. Elucidation of methanogenic coenzyme biosyntheses: From spectroscopy to genomics. *Nat. Prod. Rep.* **2002**, *19* (2), 133−147.
(49) Greening, C.; Ahmed, F.; Mohamed, A.; Lee, B.; Pandey, G.; Warden, A.; Scott, C.; Oakeshott, J.; Taylor, M.; Jackson, C. F420-and Fo-dependent redox reactions: Physiology, biochemistry, and applications. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 451−493.
(50) Johnson, E. F.; Mukhopadhyay, B. A new type of sulfite reductase, a novel coenzyme F420-dependent enzyme, from the methanarchaeon Methanocaldococcus jannaschii. *J. Biol. Chem.* **2005**, *280* (46), 38776−38786.
(51) Lambrecht, J.; Cichocki, N.; Hübschmann, T.; Koch, C.; Harms, H.; Müller, S. Flow cytometric quantification, sorting and sequencing of methanogenic archaea based on F420 autofluorescence. *Microb. Cell Fact.* **2017**, *16* (1), 180.

https://doi.org/10.1021/acs.est.3c07702
Environ. Sci. Technol. 2024, 58, 2360−2372