Supplementary Information for:
Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly

Supplementary Methods

DNA Extraction and Amplification:

For each sample, 1.5ml aliquots of activated sludge collected from the inlet and outlet of the reactor were centrifuged at 10,000g to pellet biomass, rinsed twice in 1ml of TAE buffer, and extracted according to the MP Bio Soil DNA extraction kit protocol (MP Biomedicals, Santa Ana CA). Amplification was performed in two steps using the Fluidigm Biomark: Multiplex PCR strategy. First, the hypervariable V4 16s region of the rRNA gene was amplified using the Earth Microbiome Project primer set with forward primer CS1-515f (5'-GTGCCAGCMGCCGCGGTAA) and reverse primer CS2-806r (5'-GGACTACHVGGGTWTCTAAT). Two 20 μL independent PCR reactions were performed per DNA extract, using Epicentre Premix F (Epicentre, Madison WI), Expand Hi-Fidelity Taq (Roche Diagnostics, Indianapolis IN), 200 nM primer, and 100 ng of genomic DNA in a Biorad T100 thermal cycler (Bio-Rad, Hercules CA) at 95 °C for 5 minutes followed by 28 cycles of: 95 °C (30 s), 55 °C (45 s), and 68 °C (30 s) and a final elongation step at 68°C for 7 minutes. Amplicons from replicate PCRs were pooled for the second round of PCR and labeled with barcodes unique for each sample using Accuprime Supermix (ThermoFisher, Carslbad CA), 50 μM forward and reverse primers (Fluidigm, South San Francisco CA), and 1 μL of template from the first round of PCR at 95 °C for 5 minutes followed by 8 cycles of of: 95 °C (30 s), 60°C (30 s), and 68 °C (30 s). To minimize over-amplification, PCR products were checked via UV-Vis and gel electrophoresis for smearing of the bands. The total cycle count was kept to 36 cycles per the EMP guidelines. The resulting amplicons were processed with a Qiagen PCR purification kit and sequenced at the University of Illinois Chicago DNA Services Facility.

Statistical analyses:

Alpha and Beta diversity metrics were calculated in Quantitative Insights Into Microbial Ecology (QIIME 1.9.0-20140227). Ordination methods (Principal Coordinate Analysis and Redundancy Analysis) and multivariate analyses (ANOVA, PERMANOVA, ANOSIM) were performed in Python v2.7.11 using the "skbio" package (0.4.0). Alpha diversity was calculated using Faith's Phylogenetic Diversity (PD) and Shannon Diversity (Faith 1992, Hill 1973). PD is a commonly used alpha diversity metric that weights species diversity by phylogenetic distance between community members. Shannon Diversity takes into account both richness and evenness of species to calculate an effective species number. Absolute OTU abundance data were rarefied to the minimum sequence count ten times and both metrics were calculated for each rarefaction and averaged using QIIME's "alpha_rarefaction.py" script. The null hypothesis of no significant

difference between alpha diversity values between groups of samples by month or reactor was tested with Analysis of Variance (ANOVA). ANOVA is a multi-group extension of a t-test that compares within-group to between-group mean and variation. ANOVA tests were performed for both samples grouped by mean and reactor source.

To assess differences between samples in terms of composition, pairwise Weighted Unifrac distances were calculated from rarefied OTU abundance counts using QIIME's "beta_diversity.py" script. The Unifrac distance is defined as the fraction of a phylogenetic tree consisting of two communities that leads to taxa found in one community but not the other. Weighed Unifrac weights path lengths by OTU abundance before calculating similarity. Significance of visual differences between categorical groupings (e.g. month or reactor source) was assessed using Permutational Analysis of Variance (PERMANOVA). PERMANOVA is a multivariate statistical test that take a distance matrix as input and test the null hypothesis that the mean and variation between and within groups are the same. The strength of seasonal and linear temporal trends were investigated with Mantel tests using both linear and "seasonal" temporal distances between samples. Correlation between the Weighted Unifrac distance matrix and both temporal distance matrices were calculated and compared. Linear distance was the total number of days between samples whereas the seasonal temporal distance was calculated as the minimum number of days of the year separating two samples, regardless of calendar year.

Ordination methods were used to visualize differences in community structure and constrain these differences based on environmental and operational variables. Principal coordinate analysis (PCoA) was used to graphically represent the Weighted Unifrac distances between bacterial communities in two dimensions using QIIME's "principal_coordinates.py" script. PCoA summarizes data along a set of uncorrelated axes that explain the variance between samples. Two dimensions were selected to visualize the results based on the total percent of variance explained. In addition to the above unconstrained ordination, redundancy analysis (RDA) was used to relate beta diversity between communities to environmental variables. Redundancy analysis is a constrained ordination method that uses linear regression to correlate a distance matrix with explanatory variables. Potential explanatory variables included operating conditions, environmental variables and categorical dummy variables representing plant to plant differences. Partial RDA was performed by comparing the variance explained by RDA models containing two groups of parameters separately and together and subtracting the variance explained by each component individually to quantify the covariation between parameters and variation uniquely explained by each group.

Regional population synchrony was calculated from time series abundances of individual OTUs. Pearson correlation coefficients were calculated for populations of each OTU in each pair of reactors, i and j. The average of the correlation coefficient across all 15 reactor pairs was used to measure average regional synchrony for that OTU. Because OTU abundances are serially dependent, the significance of the regional synchrony values were assessed by random permutation. 10,000 pairs of OTU time-series in different reactors were selected with replacement at random and Pearson correlation coefficients were calculated for each pair.

References

Faith DP (1992). Conservation evaluation and phylogenetic diversity. *Biological conservation* **66**: 1-10.

Hill MO (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*. **54**: 427-432.