ORIGINAL ARTICLE

Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly

James S Griffin[1] and George F Wells[2]

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA and
[2]Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL, USA

Seasonal community structure and regionally synchronous population dynamics have been observed in natural microbial ecosystems, but have not been well documented in wastewater treatment bioreactors. Few studies of community dynamics in full-scale activated sludge systems facing similar meteorological conditions have been done to compare the importance of deterministic and neutral community assembly mechanisms. We subjected weekly activated sludge samples from six regional full-scale bioreactors at four wastewater treatment plants obtained over 1 year to Illumina sequencing of 16S ribosomal RNA genes, resulting in a library of over 17 million sequences. All samples derived from reactors treating primarily municipal wastewater. Despite variation in operational characteristics and location, communities displayed temporal synchrony at the individual operational taxonomic unit (OTU), broad phylogenetic affiliation and community-wide scale. Bioreactor communities were dominated by 134 abundant and highly regionally synchronized OTU populations that accounted for over 50% of the total reads. Non-core OTUs displayed abundance-dependent population synchrony. Alpha diversity varied by reactor, but showed a highly reproducible and synchronous seasonal fluctuation. Community similarity was dominated by seasonal changes, but individual reactors maintained minor stable differences after 1 year. Finally, the impacts of mass migration driven by direct biomass transfers between reactors was investigated, but had no significant effect on community similarity or diversity in the sink community. Our results show that population dynamics in activated sludge bioreactors are consistent with niche-driven assembly guided by seasonal temperature fluctuations.

Introduction

Mixed-culture activated sludge systems are the most important form of modern wastewater treatment (Seviour *et al.*, 2010). Consistent system performance relies on a complex microbial community to remove organic carbon and nutrients in the face of dynamic environmental conditions. In addition to their importance for protecting environmental and public health, engineered environments such as activated sludge have been shown to be valuable environments to study fundamental microbial ecology phenomena (Daims *et al.*, 2006), in part due to their physical partitioning into ecological 'islands' with well-defined and monitored ecosystem functions. Wastewater microbiology has informed our understanding of community assembly mechanisms (Ofiteru *et al.*, 2010; Wells *et al.*, 2011; Kim *et al.*,

2012), resilience and stability (Werner *et al.*, 2011; Vuono *et al.*, 2014), and identified novel microbes and metabolic pathways (Strous *et al.*, 1999; van Kessel *et al.*, 2015). Translating this knowledge into effective strategies for 'microbial resource management' (Verstraete *et al.*, 2007) remains a challenge due to the high diversity and dynamics in activated sludge bioreactors, even during times of functional stability.

In both natural and engineered microbial ecosystems, the extent to which stochastic versus deterministic processes influence microbial community assembly is still debated. Traditional niche community assembly theory predicts that deterministic processes such as regional meteorological conditions and operational differences control assembly, whereas neutral theory predicts that trophically similar community members are ecologically similar and that stochastic processes such as immigration, birth and death lead to community differences (Hubbell, 2001; Woodcock *et al.*, 2007). Community assembly theories such as the metacommunity framework incorporate both niche processes, such as species sorting and dispersal limitations, as well as

Correspondence: G Wells, Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Tech A236, Evanston, IL 60208, USA.
E-mail: george.wells@northwestern.edu

neutral processes (Leibold *et al.*, 2004). Previous studies have shown that deterministic factors explain much of the variation in activated sludge community structure within the same bioreactor, but neutral models can explain some aspects of community assembly, such as rare taxa dynamics (Ofiteru *et al.*, 2010). However, the factors controlling the relative contributions of different community assembly mechanisms are still poorly understood.

Longitudinal studies of full-scale bioreactors have found that microbial communities appear to display continuous rather than cyclical succession patterns in activated sludge systems (Wells *et al.*, 2011; Ju *et al.*, 2013). Many of these studies focused on single reactors, making it challenging to identify whether observed associations were repeatable and generalizable. In contrast, seasonal population dynamics have been shown in a variety of marine and freshwater environments (Crump and Hobbie 2005; Gilbert *et al.*, 2009). The apparent differences between community assembly in natural and engineered environments (Shade *et al.*, 2013) have been ascribed to more highly controlled environments in engineered systems studied to date; however, seasonal differences in ecosystem function (performance), including nutrient removal are common (Ju *et al.*, 2013).

Spatially correlated population fluctuations or 'regional synchrony' (Eiler *et al.*, 2011) and synchronized shifts in community structure or 'temporal concordance' (Kent *et al.*, 2007) are hallmarks of deterministic community assembly mechanisms. Concordance and synchrony dictate how broadly we can generalize from community studies. Regional population synchrony (Bjørnstad *et al.*, 1999) is explained by similar regional weather patterns, known as the 'Moran Effect', immigration between communities and historical factors (Liebhold *et al.*, 2004). Synchrony has been demonstrated during the startup of lab-scale reactors (Vanwonterghem *et al.*, 2014), but whether the same processes control dynamics in full-scale reactors is an open question. Regional population synchrony would provide further evidence for deterministic control of wastewater treatment plant (WWTP) microbial community dynamics. To date, limited replication from full-scale bioreactors undergoing similar environmental conditions have made it difficult to identify synchrony in engineered systems.

The primary objectives of this work were to test whether deterministic factors such as seasonal variability drove microbial community assembly and whether assembly mechanisms affected rare and abundant operational taxonomic units (OTUs) differently in regional activated sludge bioreactors. We analyzed local communities within a plant that shared an immigration (influent) source, as well as geographically distinct communities within a region to determine the relative importance of extrinsic environmental factors and immigration. Temporal population synchrony was quantified as a means of identifying repeatable deterministic responses of individual OTUs to regional environmental dynamics. We examined synchrony at multiple scales, from whole-community metrics of alpha and beta diversity, to dynamics of individual OTU populations between reactors. We hypothesized that, similar to natural aquatic environments, seasonal variation would be a major determinant of community composition and diversity in activated sludge, and would outweigh stable differences between plants. Further, we hypothesized that abundant and rare OTUs would exhibit different dynamics, with abundant OTUs tending to exhibit synchronized seasonal blooms in different reactors due to seasonal factors and rare or transient OTUs displaying significantly less synchrony.

Materials and methods

Plants, sampling and sequencing
Activated sludge biomass samples were collected between October 2014 and September 2015. Six activated sludge bioreactors at four full-scale wastewater treatment plants (Hanover Park, Egan, Kirie and O'Brien Water Reclamation Plants) were sampled weekly for 1 year, except during plant shutdowns. All but one of the reactors operated as fully aerated activated sludge systems performing biochemical oxygen demand (BOD) removal and nitrification. The Hanover 1 reactor operated with a Modified Ludzack–Ettinger process with an anoxic zone for denitrification, followed by an aerobic zone for nitrification. Two independent reactors each at Hanover Park and Egan with separate return activated sludge systems and single reactors at Kirie and O'Brien were sampled, yielding a total of six independent activated sludge reactors sampled. Operational and environmental parameters were monitored according to standard methods (Eaton *et al.*, 1998; Table 1). Plants were located in the USA in Chicago, Illinois, USA suburbs with Hanover Park in Hanover Park, Egan in Schaumburg, Kirie in Des Plaines, and O'Brien in Skokie, IL, USA.

For each weekly sample, 50 ml mixed liquor grab samples were collected from near the inlet and outlet of the reactor, and combined as a single time point. Samples were transported on ice, centrifuged at 10 000g for 5 min and decanted, and the resulting biomass pellets were stored at $-80\,^\circ$C until DNA extraction. Genomic DNA was extracted from 1.5 ml of activated sludge using the FastDNA Spin Kit for Soil (MP Bio, Santa Ana, CA, USA). 16S ribosomal RNA (rRNA) V4 gene sequences were amplified, following the MiDAS Field Guide to Activated Sludge (McIlroy *et al.*, 2015) protocol originally published by the Earth Microbiome Project using the 515f and 806r primer set (Caporaso *et al.*, 2012). Two 20 μl independent PCR reactions were performed per DNA extract using 100 ng of genomic DNA in a Biorad T100 thermal cycler (Bio-Rad, Hercules, CA,

Table 1 Average operating parameters for the six different reactors studied

	Plant and reactor				Influent			Effluent				
Coordinates	Flow rate (mgd)	Solids retention time (days)	Mixed liquor suspended solids (mg l^{-1})	Sludge volume index (ml g^{-1})	BOD$_5$ (mg l^{-1})	Total Kjeldahl nitrogen (mgN l^{-1})	Total phosphorus (mgP l^{-1})	Suspended solids (mg l^{-1})	BOD$_5$ (mg l^{-1})	Total Kjeldahl nitrogen (mgN l^{-1})	NO_3^- and NO_2^- (mgN l^{-1})	Total phosphorous (mgP l^{-1})
Egan North 42.01N 88.02W	23.8 ± 4	6.4 ± 1.2	2243 ± 217	118 ± 28	164.9 ± 28	29.5 ± 5.4	5.8 ± 0.9	4.8 ± 1.2	2.2 ± 0.3	1.5 ± 0.3	15.9 ± 2.3	2.9 ± 0.8
Egan South[a] 42.01N 88.02W		11.5 ± 4.2	2284 ± 190	92 ± 12								
Hanover 1 42.00N 88.08W	9.3 ± 3	17.5 ± 4.8	2113 ± 282	206 ± 86	146.8 ± 31.8	31.3 ± 8.3	4.6 ± 1.1	5.1 ± 2.1	5.5 ± 3.4	2.3 ± 0.3	Not avail.	2.8 ± 0.7
Hanover 7[a] 42.00N 88.08W		14.2 ± 4.1	2346 ± 210	115 ± 24								
Kirie 42.01N 87.56W	30.7 ± 12	10.2 ± 2.1	2806 ± 290	76 ± 15	138.3 ± 29.4	28.43 ± 7.6	4.2 ± 0.8	2.2 ± 0.3	2.7 ± 0.8	1.6 ± 0.5	9.2 ± 3.1	1 ± 0.6
O'Brien 42.01N 87.42W	218.8 ± 37	9.7 ± 1.1	2183 ± 155	89 ± 12	126.3 ± 39.9	25.9 ± 7.0	3.9 ± 1.1	7.1 ± 2.2	9.9 ± 5	1.5 ± 0.3	9.7 ± 1.2	1.5 ± 0.3

Average operating data for six full-scale activated sludge bioreactors used in this study. Plant-wide variables (influent and effluent) are shown for one reactor per plant.
[a]Egan and Hanover plant-wide, influent and effluent values are the same for both sampled reactors.

USA) at 95 °C for 5 min 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 min. (For details, see Supplementary Methods.) DNA sequencing was performed using a Miseq V2 sequencer (Illumina, San Diego, CA, USA) at the University of Illinois Chicago DNA Services Facility. Sequences can be accessed on Genbank (PRJNA317773).

Amplicon sequence processing and quality control
Paired-end Illumina V4 16S rRNA sequences were processed using Vsearch 1.9.1 (Rognes 2015). After merging paired-end reads, sequences with more than one expected error, longer than 300 bp, or with any unknown nucleotides were discarded. Singletons and likely chimeras were also discarded using default settings in Vsearch. Samples with fewer than 5000 reads ($<$0.5% of mean) were discarded because of low coverage. Representative sequences from each OTU were aligned using the Greengenes imputed core reference align and PyNast implemented in Quantitative Insights Into Microbial Ecology (QIIME 1.9.0-20140227; DeSantis *et al.*, 2006; Caporaso *et al.*, 2010a, b). After filtering the alignment to remove gaps and hypervariable regions, a phylogenetic tree was built using FastTree (Price *et al.*, 2010). Samples were rarefied to the minimum sequencing depth 10 times, and relative abundances were averaged between rarefactions before diversity metrics were calculated. Of the 271 samples collected over 49 weeks, 22 551 652 unique sequences passed quality filtering, and were dereplicated and clustered into 19 171 OTUs at 97% similarity. After chimera filtering, 17 666 983 of the initial sequences (78%) were mapped back to OTUs. Alpha and beta diversity metrics were calculated in QIIME using an OTU table rarefied to the lowest sequencing depth (11 542 sequences per sample). A phylogenetic alpha diversity metric, Faith's Phylogenetic Diversity (PD) that weights changes in observed taxa by their phylogenetic distance to the nearest neighbor, and Shannon diversity, ^1D, which measures community evenness as well as richness, were used to quantify ecosystem diversity over time in all six reactors. Weighted unifrac was used to calculate beta diversity between samples (Lozupone and Knight, 2005). Taxonomy was assigned to OTU representative sequences using uclust and the Greengenes sequence database (Edgar, 2010).

Statistical analysis
All statistical analyses were performed in Python (2.7) (python.org) using the skbio package (0.4.0) (scikit-bio.org) and R (3.2) using the vegan package (3.2) (Dixon, 2003). Significance of observed alpha diversity trends were assessed using analysis of variance (ANOVA). Then, multiple linear regression was used to identify factors that best explained changes in alpha diversity. Measured environmental

and operational variables were normalized to unit variance. These were used alongside categorical variables for each plant, as regressors to predict alpha diversity. Because the regressors were not independent of each other, relative importance was calculated by averaging sequential sum-squared error over all orderings of regressors (Lindeman et al., 1980) in the R package, relaimpo (Grömping, 2006) resulting in estimates of the variation uniquely explained by individual regressors.

Two distinct methods were used to evaluate the effect of persistent reactor specific differences, as well as seasonal and continual change on community similarity (Beta diversity). Permutational ANOVA (PERMANOVA) was used to evaluate the significance of categorical groupings based on month and season. Mantel tests were used to evaluate the correlation between absolute temporal distance and beta diversity, as well as seasonal temporal distance and beta diversity. Ordination methods (Principal Coordinate Analysis and Redundancy Analysis) were used to visualize community similarity and calculate the fraction of variation in community structure explained by environmental gradients or influent differences. The preceding statistical methods are described in more detail in the Supplementary Methods.

Regional synchrony calculations
Regional OTU synchrony, defined as synchronous changes in individual OTU population in separate communities, was calculated using Pearson correlation (Liebhold et al., 2004). For each OTU, N, the Pearson correlation $\rho_{i,j}^{N}$ of its abundance time series in each reactor pair, (i, j), was calculated. Regional synchrony values for each OTU were calculated by averaging $\rho_{i,j}^{N}$ across all reactor pairs. Owing to the compositional effects (Aitchison, 1982), and the autocorrelation of populations over time (Liebhold et al., 2004) correlation coefficients will not necessarily be zero for uncorrelated OTUs. To test whether OTU populations in different reactors were more synchronized with each other than expected for independent OTUs, synchrony values were compared with a bootstrap confidence interval (CI) obtained from sampling 10 000 random pairs of OTU time series.

Results

Ecosystem function and operating conditions
Operating data, environmental conditions and effluent quality for all six reactors are shown in Table 1. Influent flow rate varied from 9 million gallons per day (MGD) at Hanover to 219 MGD at O'Brien. Influent composition, nutrient concentrations and effluent quality were similar in all four plants; however, average solids retention time (SRT) varied from 6.4 days in Egan North to 17.5 days in Hanover

1. Average sludge volume index (SVI), a measure of settleability and indicator for filamentous bulking problems, varied from $76\ ml\ g^{-1}$ in Kirie to $206\ ml\ g^{-1}$ in Hanover 1. Monthly summaries of environmental and operating parameters are shown in Supplementary Table 1. Water temperature varied seasonally from a low of 13 °C in March 2015 to 20 °C in October 2014. Influent total Kjeldahl nitrogen varied seasonally with a peak concentration of $32.8\ mgN\ l^{-1}$ in November and a minimum concentration of $25.9\ mgN\ l^{-1}$ in May, but nitrification efficiency remained above 94% throughout the year. Average SVI was highest in February, March and June, when one or more plants experienced bulking events.

Taxa showed consistent seasonal dynamics across plants
Despite relatively stable performance throughout the year, core OTU populations in different reactors were highly dynamic and regionally synchronized. OTU abundances were highly skewed, with a small number of OTUs constituting most of the sequences recovered (Supplementary Figure 1). A 'core' community of 134 OTUs was found in all reactors at all time points. These core OTUs made up 51% of the total reads. A total of 599 OTUs were constitutively present in one or more plants but not all, suggesting that individual plants harbored some stable differences in composition throughout the year.

Figure 1a shows synchrony coefficients for core OTUs compared with a bootstrap distribution of randomly selected OTU populations. Core OTUs were much more synchronized than randomly compared taxa (Student's t-test = 13.1, $P < 0.001$), consistent with niche-driven selection influenced by regional environmental factors. Non-core OTUs present in at least 20% of samples displayed enriched synchrony as well (Student's t-test = 31, $P < 0.001$). In addition, there was a significant correlation between synchrony and OTU occurrence frequency (Pearson $R = 0.55$, $P < 0.001$) and average rank abundance (Spearman $\rho = -0.61$, $P < 0.001$; Figures 1b and c). Common OTUs displayed synchronized changes in abundance between reactors, whereas rare OTUs tended to be uncorrelated. The apparent discrepancy between abundant and rare OTU synchrony suggests that seasonal effects may have a larger influence on shaping abundant OTU dynamics, whereas operating conditions or neutral factors may play a larger role in shaping the rare microbiome of full-scale activated sludge bioreactors.

We further examined whether populations linked by plant specific factors (for example, nutrient load and operational set points) were more synchronized than those that shared regional environmental variables only (for example, water temperature and precipitation). Mean synchrony coefficients for core OTU populations in the two pairs of bioreactors
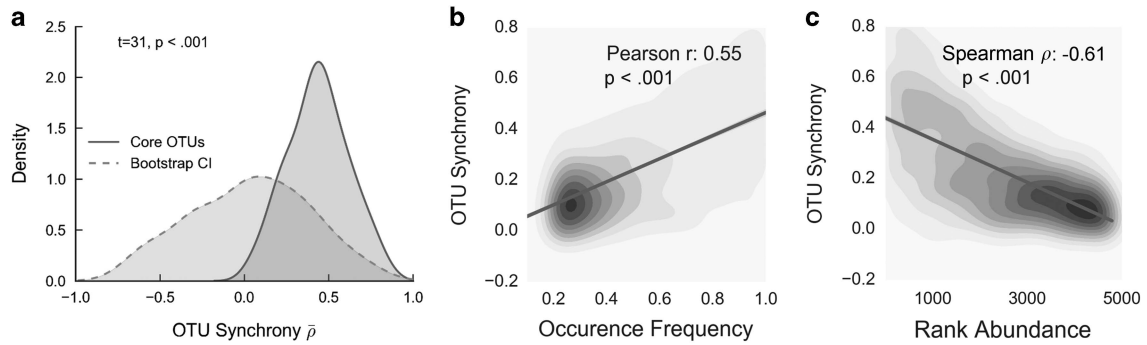
Figure 1 Regional OTU synchrony in activated sludge communities. (**a**) Density plot of regional synchrony for 'core' OTUs (present in all samples) vs a bootstrap sample of randomly selected OTU time-series pairs. Density plots of synchrony as a function of observed frequency and average abundance are shown in **b** and **c**. Most OTUs were relatively rare and uncorrelated between plants, but a small number of frequent OTUs were highly synchronized.
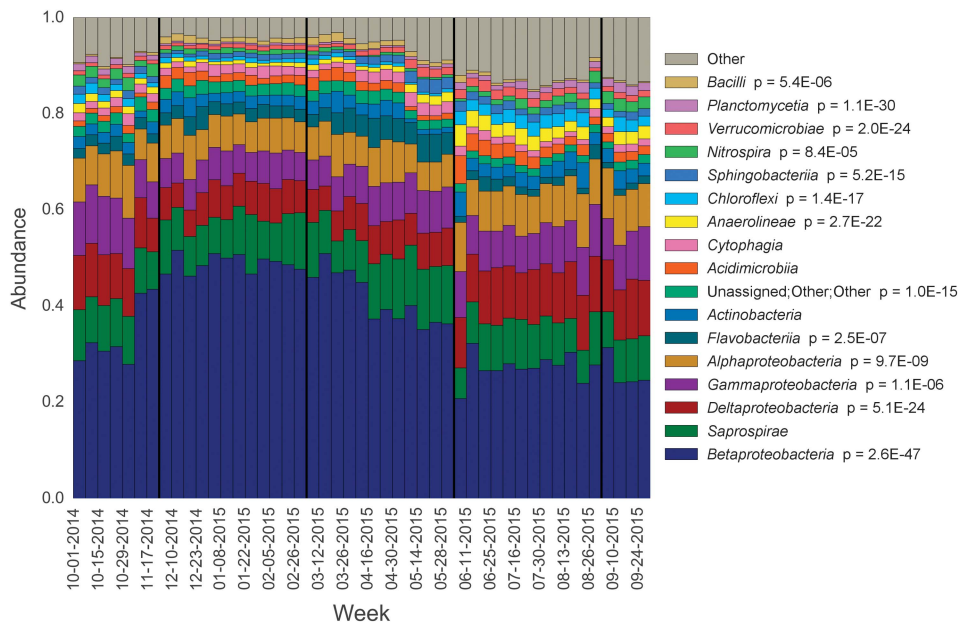

Figure 2 Average abundance of the top 15 most abundant bacterial classes at each time point. Sampling date is shown on the *x* axis. Black lines represent cutoffs for different seasons starting from Fall 2014. *Betaproteobacteria* were the most dominant class at all time points, but increased in relative abundance between winter and summer. *P*-values for classes with significantly different abundance in winter and summer are shown in the legend.

sharing influent, Egan South and North and Hanover 1 and 7, were compared with average population synchrony values across all 15 reactor pairs. Egan populations were the most synchronized (average $\rho = 0.62$ for Egan compared with 0.44 across all reactors). Hanover populations were less synchronized (average $\rho = 0.42$) than the average across all regional reactors, likely owing to environmental differences imposed by the anoxic selector present in Hanover 1.

Next, we investigated whether synchronous OTU dynamics gave rise to similar shifts in broader phylogenetic groups. Class level abundances averaged by season and plant are shown in Figure 2. Classes with significantly different abundances in winter (December–February) and summer (June–August) are indicated by *P*-values in the legend.

Coefficients of variation of class abundances within weeks are shown in Supplementary Table 2. Dominant taxa were consistent between plants, and within-week variation between reactors was low for most classes. However, class abundances varied throughout the year, in agreement with high rates of temporal OTU synchrony. *Betaproteobacteria* were the most abundant class throughout the year, followed by *Saprospirae*, *Delta-*, *Alpha-* and *Gammaproteobacteria*. *Betaproteobacteria* were most dominant during winter, making up on average 44% of the total reads, but were significantly less abundant (Student's *t*-test = 22.1, $P < 0.001$) and accounted for only 27% of total reads in summer. Student's *t*-tests comparing class abundances in samples collected in summer and winter revealed that 11 of the 15 most abundant classes were

significantly more abundant during summer than winter after Bonferroni Correction (Student's *t*-test, all $P < 0.0001$). Among these, *Nitrospira* were on average 84% more abundant in summer than winter samples.

Alpha diversity is seasonally synchronized and primarily influenced by temperature

Population synchrony and shifts in broad phylogenetic abundance were accompanied by repeatable and putatively seasonal alpha diversity patterns. Shannon Diversity and Faith's phylogenetic diversity (PD) were used to quantify richness and evenness, and PD throughout the sampling period (Figure 3). One-way ANOVA revealed statistically significant differences between seasons for both phylogenetic ($R_{ANOVA} = 93.23$, $P < 0.001$) and Shannon ($R_{ANOVA} = 42.84$, $P < 0.001$) diversity. For both metrics, diversity peaked in fall between September and November, and reached a minimum in all reactors in March or December. Diversity in all plants gradually increased during spring and summer, such that there was no statistically significant difference in Shannon Diversity (Student's *t*-test = 0.98, $P = 0.33$) between September 2015 and October 2014 samples. Pairwise Pearson correlation coefficients of alpha diversity between plants confirmed synchronized changes in community diversity over time based on both Faith's PD (mean $R = 0.65 \pm 0.11$, all $P < .001$) and Shannon Diversity (mean $R = 0.54 \pm 0.16$, all $P < 0.001$). In addition to reproducible and synchronized seasonal variation in alpha diversity, ANOVA revealed that individual reactors harbored stable differences in diversity throughout the year ($R_{ANOVA} = 11.4$, $P < 0.001$). Kirie was the most diverse community in 29 time points (60%), and Egan South and O'Brien were the least diverse in 31% and 39% of all samples, respectively.

Multiple linear regression was used to identify the unique fraction of variation in Shannon Diversity that could be linked to environmental and operational factors (Supplementary Table 3). Regressors included regional environmental gradients (influent water temperature, chloride concentration and precipitation) that varied seasonally, local environmental conditions (BOD, N and P influent concentration) and plant location, operational parameters (SVI, SRT and MLSS) and performance indicators (effluent nitrite concentration, BOD, N and P removal). The relative importance of each variable was assessed by averaging sequential sum-squared error over all orderings of regressors to account for potential colinearity between regressors. Together, these factors explained 49% of the variation in Shannon Diversity.

Plant identity and temperature were the most important predictors of Shannon Diversity, and explained 25% and 12% of the variation, respectively. One caveat of interpreting relative importance of regressors is that predictors with higher variance tend to have inflated effect sizes (Kruskal, 1987). Despite this, temperature is an important driver of community structure in many microbial ecosystems and is likely to be important in nutrient- and oxygen-rich environments like activated sludge. In addition, increased SVI, which is typically associated with enrichment of filamentous bacteria, was linked to decreased Shannon Diversity. Operational parameters that could be most easily controlled, SRT and suspended solids concentration, accounted for <4% of total variation.

Community similarity was controlled by region-wide seasonal factors and local plant differences

Principal coordinates analysis of weighted unifrac community distance revealed seasonal clustering of communities (Supplementary Figure 2). Mantel correlograms (Figure 4) comparing weighted unifrac distance and time lag between samples were created to assess periodic changes in community structure, as well as conserved differences between reactors. Dissimilarity was plotted as a function of time for three groups of samples: those originating from the same reactor, those taken from separate reactors in the same plant and those taken from different plants. At low temporal lag, communities from the same reactor were more autocorrelated than communities
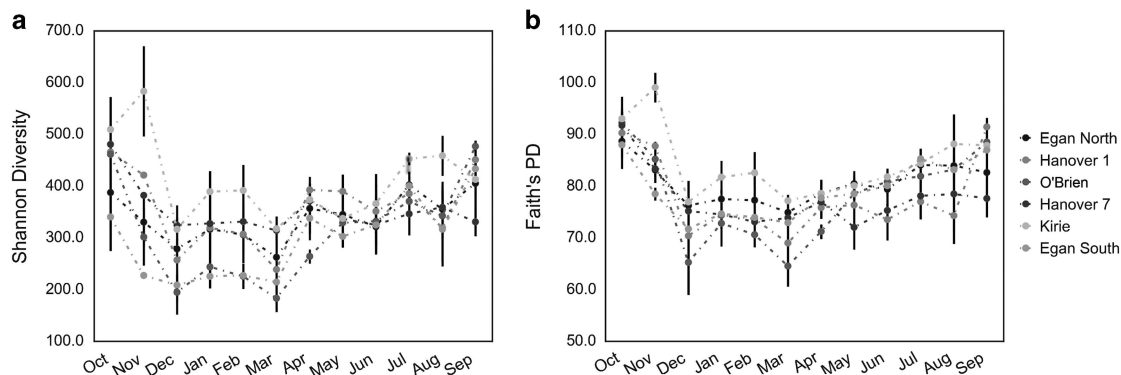

Figure 3 Time series of (**a**) Shannon and (**b**) Faith's PD alpha diversity for all six reactors. Data shown are average of all samples taken during each month. Alpha diversity was highest in October and November, and lowest in December and March.
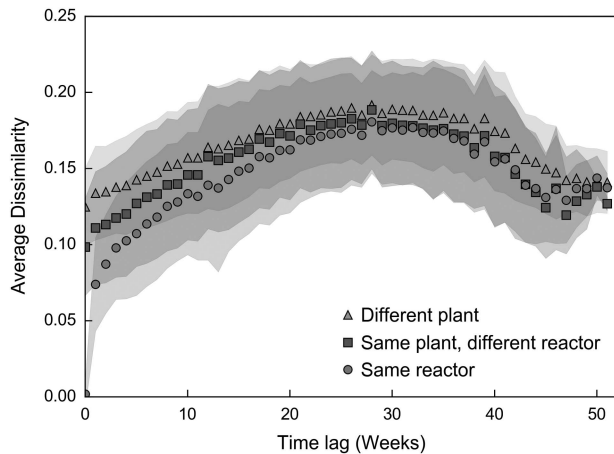
Figure 4 Temporal correlograms of time lag between samples versus dissimilarity (weighted unifrac distance) showing the mean and standard deviation (shaded area) for each week. Colors represent whether the compared samples originate from the same reactor (red circles), different reactors at the same plant (blue squares) or different plants (green triangles). The day of year lag between samples correlated well with weighted unifrac distance ($R_{pearson} = 0.47$, $P < 0.001$). A full color version of this figure is available at the *ISME* journal online.

from other reactors or plants. As time lag between samples began to increase, average dissimilarity increased for all groups of reactor pairs. However, the beta diversity difference between intra- and inter-reactor sample pairs diminished. Average dissimilarity peaked at a lag of 28 weeks. The mean dissimilarity between samples taken 6 and 51 weeks apart was comparable, indicating that both continual succession, as well as seasonal effects controlled overall community dynamics. Samples taken from different reactors within the same plant were on average more similar to each other than those taken from different plants at every time point, suggesting that influent composition or plant-to-plant difference in operating set points also influenced community structure. Mantel tests were used to calculate the correlation between weighted unifrac distance and temporal distance. The total time between samples and the day number lag between samples (182 days for samples half a year apart and 0 for samples exactly a year apart) were calculated for each pair of samples. The day number lag was more correlated with differences in community structure between all communities ($R_{pearson} = 0.47$, $P < 0.001$) than absolute time difference ($R = 0.36$, $P < 0.001$).

To better visualize the seasonal community structure, weighted unifrac trajectories for each reactor were plotted individually, with points colored by season of origin and connected over time by lines (Figure 5). In general, samples from the same season grouped together, with communities tending to be highly similar to one another in winter and summer, and transition between these clusters in fall and spring. The overall placement of the seasonal clusters along the principal coordinates was similar across reactors. In addition, a strong annual cycle

was observed for all of the reactors studied, and samples taken in fall of 2015 closely resembled those taken from the fall of 2014. PERMANOVA was used to test the significance of bacterial community groupings by reactor and season. Grouping by both season (Pseudo-$F_{PERMANOVA} = 49.15$, $P < 0.001$) and reactor (Pseudo-$F_{PERMANOVA} = 11.61$, $P < 0.001$) showed significantly different community structure.

Using partial redundancy analysis, the beta diversity that could be uniquely explained by regional, local, operation and performance regressors was calculated (Supplementary Table 4). Weighted unifrac distances were constrained by each set of explanatory variables after partialing out the other sets of variables. Variance that was explained by one or more set of variables is listed as covariation. Together, 52% of total variation was explained. Similar to alpha diversity, temperature was the largest individual driver of community dissimilarity and accounted for above 10% of the total variation in beta diversity. Over half of the total explained variance (28%) was controlled by covariation between one more variables. The largest portions of covariation were between temperature, local and performance factors due to seasonal fluctuations in influent nutrient load and effluent quality (Supplementary Table 1). Taken together with the PERMANOVA results, reactor communities in the system were highly controlled by deterministic factors.

Finally, direct sludge transfers between plants were used to quantify the impact of mass effects on community similarity. During the sampling period, waste activated sludge was transferred between Kirie, Egan North and Egan South on six occasions in an attempt to improve settling during SVI upsets. We compared weighted unifrac distance between source and sink communities collected the week before and after transfers to find evidence for colonization of the sink community by the source. Transfer sludge source and sink reactor, mass, and pre- and post-transfer weighted unifrac distance are shown in Supplementary Table 5. No significant change in similarity in source and sink communities was evident before and after reseeding (Student's t-test = 0.39, $P = 0.71$), suggesting that sink communities were resistant to colonization or invasion at tested levels of immigration or effects were too transient to see with our sampling strategy.

Seasonal changes in ecosystem function
Lastly, we examined whether synchronous population shifts could explain seasonal differences in nitrification performance and bulking event frequency observed in these plants. Operating data showed that several plants experienced seasonal increases in effluent nitrite residual during winter between 2010 and 2014 (Supplementary Figure 3). Nitrification is traditionally thought to be a two-step process, in which ammonia oxidizing bacteria such as *Nitrosomonas* convert ammonia to nitrite and
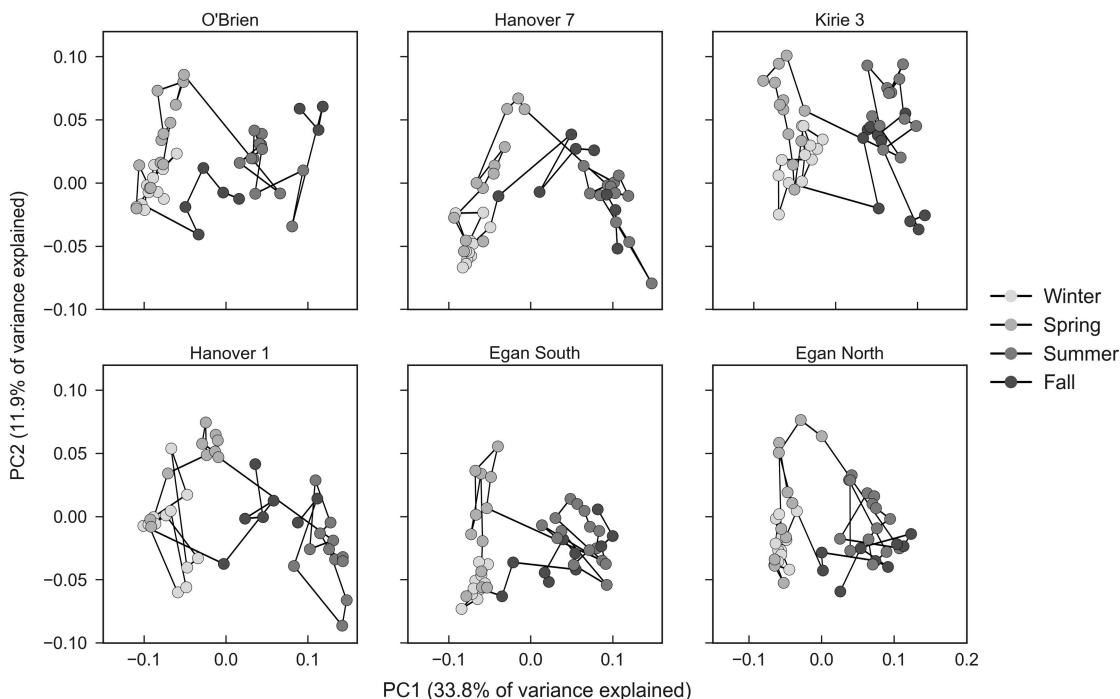
Figure 5 Principal coordinate analysis of weighted unifrac distances between samples. Colors represent season of origin and lines connect samples taken on consecutive sampling trips.

nitrite-oxidizing bacteria such as *Nitrospira* further oxidize nitrite to nitrate (Wagner *et al.*, 2002). We found that *Nitrospira* abundance decreased by nearly 70% from October to April (~2.3–0.75%; Supplementary Figure 4A) before rebounding in the summer. Hanover 1 was the only plant that did not experience a decrease in *Nitrospira*, possibly due to higher SRT preventing washout. In contrast, the dominant genus of ammonia oxidizing bacteria in each reactor, *Nitrosomonas*, showed little seasonal variability (Supplementary Figure 4B).

During our time-series monitoring, several bulking events lasting longer than a week were investigated. Bulking sludge is a common process disturbance caused by blooms of filamentous or bulking bacteria such as *Microthrix* or foaming *Actinomycetes* (Martins *et al.*, 2004). During the startup phase of modified ludzack-ettinger operation, Hanover 1 had a 3-day average SVI > 150 for a period of 8 months from December 2014 to August 2015. Hanover 7 had several instances of elevated SVI in November 2014, and May, June and July 2015. Egan North had a bulking event from February to mid March 2015 and again from May to June 2015. On the basis of bulking and filamentous bacteria (BFB) previously identified using 16S rRNA amplicon-based sequencing (Guo and Zhang 2012), we identified 11 genera putatively linked to poor settling characteristics with abundance > 1% in at least one sample, including *Microthrix, Thiothrix, Caldilinea, Trichococcus, Rhodococcus, Haliscomenobacter, Gordonia, Kouleothrix, Mycobacterium, Tetrasphaera* and *Isosphaera*.

The most common taxa in terms of average abundance were *Thiothrix* (2.2%), *Kouleothrix* (1.3%), *Caldilinea* (1.2%) and *Microthrix* (0.8%); however, the BFB were highly dynamic (Supplementary Figure 5). Of the observed bulking bacteria, *Caldilinea, Mycobacterium, Haliscomenobacter* and *Gordonia* primarily displayed seasonal variation and synchrony between plants, but did not correlate with SVI. Elevated *Microthrix* abundance was linked to increased SVI ($R_{pearson} = 0.44$, $P < 0.001$); however, *Microthrix* (S5A) and *Thiothrix* (S5B) were inversely correlated with one another ($R_{pearson} = -0.37, P < 0.001$). *Microthrix* tended to dominate in reactors with elevated SVI, whereas *Thiothrix* were more abundant in the three reactors that did not experience bulking events. In general, bulking bacteria tended to be more diverse in fall and dominated by a few taxa during the winter and spring.

Discussion

Dominant taxa were regionally synchronized
In this study, community dynamics in regional activated sludge bioreactors were monitored to identify the relative contribution of different assembly mechanisms. Our results show that OTU populations are highly regionally synchronized between reactors (Figure 1) across several scales. Synchrony in macroscopic ecological systems is explained by deterministic assembly mechanisms, such as spatially correlated environmental conditions (Moran, 1953) and trophic interactions, such as predation

(Ydenberg and Dill, 1986). In microbial communities, synchrony has been observed in rivers (Crump and Hobbie, 2005) and dispersal-limited lakes (Kent *et al.*, 2007), but this effect has not previously been observed in full-scale activated sludge bioreactors.

The observation of synchrony in dispersal-limited reactors combined with the strong effect of temperature on community structure (Supplementary Table 4) supports regional temperature fluctuations as a primary driver of community assembly. In addition to temperature, seasonal nutrient variation is typically a driver of regional synchrony in aquatic microbial ecosystems (Andersson *et al.*, 2009; Eiler *et al.*, 2011), but we found that influent nutrient concentrations in the studied WWTPs were stable over the course of the year and contributed little to observed variation. While all reactor pairs showed significant synchrony, this effect was highest for core OTU populations in the two Egan bioreactors that shared the same influent as a potential immigration source. Interestingly, population dynamics in the Hanover bioreactors that also shared the same influent were less similar than between plants, likely due to environmental differences caused by the anoxic selector in Hanover 1. Previous studies investigating the impact of immigration from raw influent on activated sludge community structure have found highly variable estimates for the fraction of OTUs shared between influent and activated sludge, from 5-10% OTU overlap (Lee *et al.,* 2015) to 35% (Saunders *et al.*, 2015). Communities from linked processes within a single WWTP also show evidence for within-plant immigration (Wells *et al.*, 2014) further suggesting that both selection and mass effect mechanisms influence community assembly. Although we did not find strong evidence that local immigration affected synchrony, it is possible that synchrony could be explained in part by region-wide changes in influent community structure, in addition to direct effects on the activated sludge community.

Despite high PD and stable nutrient removal performance, functionally redundant taxa did not appear to randomly dominate different bioreactors. Instead, individual taxa displayed synchronized patterns across multiple plants. High rates of synchrony have previously been observed in the startup of replicate lab-scale anaerobic digesters (Vanwonterghem *et al.*, 2014), but it was unclear whether similar assembly mechanisms would drive dynamics in full-scale open-engineered environments. A high density of microbial interactions has previously been reported in activated sludge (Ju *et al.*, 2014). In principle, these ecological interactions may compound seasonal environmental changes and increase synchrony.

Few studies have quantified differences in synchrony between rare and abundant taxa in activated sludge or other environments. Population synchrony was strongest for core and highly abundant OTUs, and was positively correlated with OTU frequency (Figure 1). By analyzing synchrony along a continuous gradient of average OTU abundance and observed frequency, our results show that community assembly mechanisms differ along a continuum between 'core' and 'rare' taxa. Previously, rare taxa have demonstrated higher variability and turnover rates compared with general taxa within activated sludge (Kim *et al.,* 2012). In another study, 10% of the total reads found in WWTPs came from OTUs with negative growth rates in activated sludge that are likely inactive or slow growing (Saunders *et al.*, 2015). Rare OTUs may be more impacted by dispersal or immigration from influent sources rather than environmental pressure within the reactors.

One limitation of our approach to quantify synchrony and dynamics is the use of 16S rRNA gene sequences, which do not directly inform community functional capabilities. Further, 16S rRNA amplicon and shotgun metagenomic studies do not always yield similar estimates of community diversity and dynamics (Poretsky *et al.*, 2014). However, amplicon sequencing enables higher temporal coverage of communities due to its lower cost. Phylogenetically related organisms tend to be functionally similar (Zaneveld *et al.*, 2010) and this has been used to predict functional capabilities for communities from phylogenetic studies (Langille *et al.*, 2013). A combined metagenomic and 16S rRNA gene sequencing study of activated sludge revealed that the community function was less seasonally variable than phylogenetic abundance (Ju *et al.*, 2013).

Diversity and class abundance varied consistently with temperature
Phylogenetic (Faith) and non-phylogenetic (Shannon) measures of alpha diversity were highly synchronized between reactors (Figure 3). Notably, in all six reactors surveyed, diversity followed a cyclical trend, dropping sharply near the end of the year and increasing again in the spring. Averaging over orderings of regressors revealed that temperature was the single most important predictor of diversity (Supplementary Table 3). Measured environmental and operational parameters explained nearly twice as much variation as plant location. Kim *et al.* (2012) demonstrated a similar decrease in alpha diversity in a single activated sludge bioreactor in winter, although the opposite effect was found in marine systems (Gilbert *et al.*, 2009).

Diversity has been shown to be critical for productivity, resilience to disturbance and functional stability in so-called 'biodiversity ecosystem function relationships' (Wittebolle *et al.*, 2009). We found that 9 of the top 15 classes reproducibly decreased in abundance across all reactors surveyed, although dominance by *Betaproteobacteria* increased during winter (Figure 2). These broad-scale shifts in diversity were accompanied by a

decrease in nitrifier populations and an increase in bulking and filamentous bacteria (BFB). Shannon diversity index was positively correlated with removal efficiency of BOD, nitrogen and phosphorous, but negatively correlated with SVI, suggesting that a positive biodiversity ecosystem function relationship exists for treatment communities. Elevated SVI is a process disturbance caused by blooms of BFB such as *Microthrix* or foaming *Actinomycetes* (Martins *et al.*, 2004). Decreased alpha diversity is symptomatic of bulking, as BFB accounted for >10% of all sequencing reads during some of the bulking episodes. It may also indirectly lead to susceptibility to bulking by opening niches for BFB, which typically bloom under low-nutrient availability, temperature and dissolved oxygen conditions (De los Reyes 2010).

Nitrospira—a key nitrite-oxidizing bacteria—displayed much higher abundance in summer than in winter in all activated sludge bioreactors. Seasonal decline of *Nitrospira* may be related to increased effluent nitrite observed across several previous winters in the plants we surveyed (Supplementary Figure 3), although effluent nitrite accumulation was not apparent during the time series described here. Nitrifier growth rates are strongly influenced by temperature and higher nitrite residuals have been observed in winter in nitrifying activated sludge bioreactors in previous studies (Randall and Buth, 1984).

Community similarity was driven by temperature and season
In addition to cyclical alpha diversity, community similarity was temporally concordant. Mantel Correlograms (Figure 4) revealed a cyclical trend in weighted unifrac distance between samples. Strikingly, beta diversity peaked just over half a year apart, and communities sampled 1 year apart were as similar as those sampled 6 weeks apart on average. Differences between plants were evident at low temporal distance, but disappeared in samples from different seasons. These differences reemerged for samples taken a year apart, and it is possible that plant-wide idiosyncrasies such as influent composition or distinct operational conditions (for example, SRT) lead to consistent year after year differences in population structure. More longitudinal studies are necessary to test this hypothesis. Constrained multivariate analyses revealed that, similar to alpha diversity, temporal variation in community structure was primarily driven by changes in temperature (Supplementary Table 4).

Seasonal community succession appears to be a general feature of aquatic microbial communities (Shade *et al.*, 2013). It has been shown across multiple years in marine (Fuhrman *et al.*, 2006) and freshwater environments (Shade *et al.*, 2007). However, in the limited studies available continual succession has been more frequently reported for activated sludge systems (Ju and Zhang, 2014). The authors speculated that due to the controlled nature of WWTPs, intrinsic factors play a stronger role than extrinsic factors in shaping communities. Contrary to previous studies of activated sludge communities, continual drift accounted for only a minor portion of variation in community composition in our study. It should be noted, however, that deconvoluting short-term fluctuations and long-term seasonal variation is not possible with a single year of data, and more work is needed to identify whether synchronous OTU patterns repeat in following years, similar to multi-year patterns observed in other environments (Fuhrman *et al.*, 2015).

One mechanism that could explain observed regional synchrony and seasonal variability in WWTPs is large temperature fluctuations combined with biomass wasting. Because solids are continuously wasted during operation, bacteria must maintain a growth rate greater than the inverse of the SRT to be maintained within the reactor. Lower temperatures reduce prokaryotic growth rates and may washout slow-growing microbes at low SRT. Cold weather plants can operate at SRTs as high as 30 days (Saunders *et al.*, 2015) and it is possible that the relatively short SRTs in our reactors (8.0–13.9 days on average) contributed to the strong seasonal taxonomic shifts.

Conclusion

Regionally synchronized patterns of biogeography have been observed in a variety of microbial ecosystems, but the importance of these processes in highly managed ecosystems such as activated sludge is not well known. Our primary objective in this study was to quantify microbial population synchrony as a means of clarifying the relative importance of deterministic seasonal environmental factors and stochastic processes on bacterial community succession in activated sludge. A clear seasonal pattern of microbial community structure was evident, from community-wide diversity metrics to dynamics of dominant OTUs between plants, and temperature was the primary driver of changes in alpha and beta diversity. In contrast, neutral factors such as immigration from continuous shared influent and intentional waste activated sludge reseeding attempts did not significantly alter beta diversity between communities. Core OTUs present in every sample were highly abundant and strongly synchronized between reactors, whereas less common OTUs tended to fluctuate more randomly, suggesting they were driven by local differences in operating conditions or stochastic processes. Key functional groups such as nitrite-oxidizing bacteria displayed repeatable seasonal differences in fully aerated systems. Our results demonstrate the importance of seasonal variability on microbial consortia and the influence of deterministic community assembly mechanisms in wastewater treatment bioreactors.

510

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Aitchison J. (1982). The statistical analysis of compositional data. *J R Stat Soc B* **44**: 139–177.

Andersson AF, Riemann L, Bertilsson S. (2009). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4**: 171–181.

Bjørnstad ON, Ims RA, Lambin X. (1999). Spatial population dynamics: analyzing patterns and processes of population synchrony. *Trends Ecol Evol* **14**: 427–432.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266–267.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N *et al.* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**: 1621–1624.

Crump BC, Hobbie JE. (2005). Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnol Oceanogr* **50**: 1718–1729.

Daims H, Taylor MW, Wagner M. (2006). Wastewater treatment: a model system for microbial ecology. *Trends Biotechnol* **24**: 483–489.

De los Reyes FL. (2010). *Microbial Ecology of Activated Sludge*. IWA Publishing: London, 215–259.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Dixon P. (2003). VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**: 927–930.

Eaton AD, Clesceri LS, Greenberg AE, Franson MAH. (1998). *Standard Methods for the Examination of Water and Wastewater*. 20th edn. American Public Health Association: Washington, DC, USA.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Eiler A, Heinrich F, Bertilsson S. (2011). Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J* **6**: 330–342.

Fuhrman JA, Cram JA, Needham DM. (2015). Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* **13**: 133–146.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci* **103**: 13104–13109.

Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T *et al.* (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11**: 3132–3139.

Grömping U. (2006). Relative importance for linear regression in R: the package relaimpo. *J Stat software* **17**: 1–27.

Guo F, Zhang T. (2012). Profiling bulking and foaming bacteria in activated sludge by high throughput sequencing. *Water Res* **46**: 2772–2782.

Hubbell SP. (2001). *The Unified Neutral Theory of Biodiversity and Bogeography (MPB-32)* vol. **32**. Princeton University Press: Princeton, NJ, USA.

Ju F, Guo F, Ye L, Xia Y, Zhang T. (2013). Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environ Microbiol Rep* **6**: 80–89.

Ju F, Xia Y, Guo F, Wang Z, Zhang T. (2014). Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environ Microbiol* **16**: 2421–2432.

Ju F, Zhang T. (2014). Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J* **9**: 683–695.

Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD. (2007). Synchrony in aquatic microbial community dynamics. *ISME J* **1**: 38–47.

Kim T-S, Jeong J-Y, Wells GF, Park H-D. (2012). General and rare bacterial taxa demonstrating different temporal dynamic patterns in an activated sludge bioreactor. *Appl Microbiol Biotechnol* **97**: 1755–1765.

Kruskal W. (1987). Relative importance by averaging over orderings. *Am Stat* **41**: 6–10.

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA *et al.* (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**: 814–821.

Lee S-H, Kang H-J, Park H-D. (2015). Influence of influent wastewater communities on temporal variation of activated sludge communities. *Water Res* **73**: 132–144.

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* **7**: 601–613.

Liebhold A, Koenig WD, Bjørnstad ON. (2004). Spatial synchrony in population dynamics. *Annu Rev Ecol Evol Syst*, 467–490.

Lindeman RH, Merenda PF, Gold RZ. (1980). *Introduction to Bivariate and Multivariate Analysis*. Scott Foresman & Co.: Glenview, Ill.

Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

Martins AMP, Pagilla K, Heijnen JJ, van Loosdrecht MCM. (2004). Filamentous bulking sludge—a critical review. *Water Res* **38**: 793–817.

McIlroy SJ, Saunders AM, Albertsen M, Nierychlo M, McIlroy B, Hansen AA *et al.* (2015). MiDAS: the field guide to the microbes of activated sludge. *Database* **2015**: bav062–bav068.

Moran P. (1953). The statistical analysis of the Canadian lynx cycle. *Aust J Zoo* **1**: 291–298.

Ofiţeru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA *et al.* (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proc Natl Acad Sci USA* **107**: 15345–15350.

Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS ONE* **9**: e93827.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2– approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Randall CW, Buth D. (1984). Nitrite build-up in activated sludge resulting from temperature effects. *Water Pollut Control Fed J* **56**: 1039–1044.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. (2015). VSEARCH: a versatile open source tool for metagenomics (No. e2409v1). *Peer J Preprints*. Available at: https://doi.org/10.7287/peerj.preprints.2409v1 (accessed on 19 November 2015).

Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. (2015). The activated sludge ecosystem contains a core community of abundant organisms. *ISME J* **10**: 11–20.

Seviour R, Halkjær P, Nielsen R. (2010). *Microbial Ecology of Activated Sludge* vol. 9. IWA publishing.

Shade A, Caporaso JG, Handelsman J, Knight R, Fierer N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J* **7**: 1493–1506.

Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD. (2007). Interannual dynamics and phenology of bacterial communities in a eutrophic lake. . *Limnol Oceanogr* **52**: 483–494.

Strous M, Fuerst JA, Kramer EH, Logemann S, Muyzer G, van de Pas-Schoonen KT *et al.* (1999). Missing lithotroph identified as new planctomycete. *Nature* **400**: 446–449.

van Kessel MA, Speth DR, Albertsen M, Nielsen PH, den Camp HJO, Kartal B *et al.* (2015). Complete nitrification by a single microorganism. *Nature* **528**: 555–559.

Vanwonterghem I, Jensen PD, Dennis PG, Hugenholtz P, Rabaey K, Tyson GW. (2014). Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J* **8**: 2015–2028.

Verstraete W, Wittebolle L, Heylen K, Vanparys B, de Vos P, van de Wiele T *et al.* (2007). Microbial Resource Management: The Road To Go for Environmental Biotechnology. *Eng Life Sci* **7**: 117–126.

Vuono DC, Benecke J, Henkel J, Navidi WC, Cath TY, Munakata-Marr J *et al.* (2014). Disturbance and temporal partitioning of the activated sludge metacommunity. *ISME J* **9**: 425–435.

Wagner M, Loy A, Nogueira R, Purkhold U, Lee N, Daims H. (2002). Microbial community composition and function in wastewater treatment plants. *Antonie Van Leeuwenhoek* **81**: 665–680.

Wells GF, Park H-D, Eggleston B, Francis CA, Criddle CS. (2011). Fine-scale bacterial community dynamics and the taxa-time relationship within a full-scale activated sludge bioreactor. *Water Res* **45**: 5476–5488.

Wells GF, Wu CH, Piceno YM, Eggleston B, Brodie EL, DeSantis TZ *et al.* (2014). Microbial biogeography across a full-scale wastewater treatment plant transect: evidence for immigration between coupled processes. *Appl Microbiol Biotechnol* **98**: 4723–4736.

Werner JJ, Knights D, Garcia ML, Scalfone NB, Smith S, Yarasheski K *et al.* (2011). Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc Natl Acad Sci USA* **108**: 4158–4163.

Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K *et al.* (2009). Initial community evenness favours functionality under selective stress. *Nature* **458**: 623–626.

Woodcock S, Van Der Gast CJ, Bell T, Lunn M, Curtis TP, Head IM *et al.* (2007). Neutral assembly of bacterial communities. *FEMS Microbiology Ecology* **62**: 171–180.

Ydenberg RC, Dill LM. (1986). The economics of fleeing from predators. *Adv Study Behav* **16**: 229–249.

Zaneveld JR, Lozupone C, Gordon JI, Knight R. (2010). Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Res* **38**: 3869–3879.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)