

## Transient-rare Bacterial Taxa Are Assembled Neutrally across Temporal Scales

SANG-HOON LEE<sup>1</sup>, TAEK-SEUNG KIM<sup>1</sup>, and HEE-DEUNG PARK<sup>1\*</sup>

<sup>1</sup>*School of Civil, Environmental and Architectural Engineering, Korea University, Seoul, South Korea, 02841*

(Received August 7, 2020—Accepted December 4, 2020—Published online February 10, 2021)

Despite the importance of microbial communities in ecosystem functions, the mechanisms underlying the assembly of rare taxa over time are poorly understood. It remains largely unknown whether rare taxa exhibit similar assembly processes to common taxa in local communities. We herein retrieved the 16S rRNA sequences of bacteria collected bimonthly for 2 years from the Pohang wastewater treatment plant. The transient-rare taxa showed different abundance distributions from the common taxa. Transient-rare taxon assemblages also exhibited higher temporal variations than common taxon assemblages, suggesting the distinct ecological patterns of the two assemblages. A multivariate analysis revealed that environmental parameters accounted for 25.3 and 61.6% of temporal variations in the transient-rare and common taxon assemblages, respectively. The fitting of all observed taxa to a neutral community model revealed that 96.4% of the transient-rare taxa (relative abundance, 71.4%) and 73.3% of the common taxa (relative abundance, 45.6%) followed the model, suggesting that stochastic mechanisms were more important than deterministic ones in the assembly of the transient-rare taxa. Collectively, the present results indicate that the transient-rare bacterial taxa at the Pohang wastewater treatment plant differed from the common taxa in ecological patterns, suggesting that dispersal is a key process in their assembly.

**Key words:** activated sludge, community assembly, neutral assembly, rare taxa

Local microbial communities influence ecosystem functions in their habitats. For example, local microbial communities are involved in plant growth in the soil (Nannipieri *et al.*, 2003), wastewater treatment in bioreactors (Briones and Raskin, 2003), and obesity in humans (Turnbaugh *et al.*, 2006). Microbial communities are typically characterized by species diversity (*e.g.*, richness or phylogenetic diversity) and relative abundance (community structure); however, these measures vary across spatiotemporal scales (Torsvik *et al.*, 2002). Despite the speculated significance of microbial community diversity in ecosystem functions (*e.g.*, see [Nemergut *et al.*, 2014]), the factors that drive and maintain the diversity dynamics of a community remain unclear. To better understand the mechanisms shaping local microbial communities, the ecological processes affecting the dynamic patterns of microbial diversity need to be evaluated.

Local microbes exhibit similar community structure patterns to larger organisms, namely, few taxa comprise a major part of the structure, while most are numerically rare. Therefore, the local community structure is strongly typified by rare microorganisms at both the micro- (Sogin *et al.*, 2006; Fuhrman, 2009) and macroecological scales (Grady *et al.*, 2019; Shade and Stopnisek, 2019; Morella *et al.*, 2020). Although rare species are often presumed to be inactive in local ecosystems, they contribute to ecosystem functions when environmental conditions become favorable for their growth (Shade and Gilbert, 2015). Nevertheless, the mechanisms underlying the assembly of rare taxa and their ecolog-

ical functions within local habitats remain largely unknown.

There are two main perspectives explaining species assembly in local communities across spatiotemporal scales: the niche and neutral theories. According to the niche theory, competition for resources among species and their different abilities to utilize resources (*i.e.*, niche differentiation) (Hubbell, 2001; Nemergut *et al.*, 2013) shape the structure of local communities. According to the neutral theory, species at the same trophic level are equivalent, and the community structure is shaped by stochastic mechanisms involving immigration, emigration, birth, death, speciation, and ecological drift (Hubbell, 2001; Nemergut *et al.*, 2013). These two contrasting views are unlikely to be mutually exclusive; the processes underpinning the two theories function synergistically to shape the local community structure (Ofițeru *et al.*, 2010; Ayarza and Erijman, 2011). Nevertheless, an outstanding question is how niche and neutral processes differently contribute to the dynamics of both rare and common taxa. The assembly of taxa with similar abundance (*e.g.*, rare *versus* common) or spatiotemporal distribution patterns (*e.g.*, persistent *versus* transient) may be driven by similar processes. Therefore, our overarching goal was to partition the contributions of niche and neutral processes to diverse community components, with a focus on processes governing the patterns of rarity.

Activated sludge bioreactors (Tchobanoglous and Burton, 1991) are the central units of biological wastewater treatment plants (WWTPs). In these bioreactors, naturally occurring microorganisms (mostly bacteria; >95%) are harnessed to degrade organic matter, oxidize inorganic nitrogen, and inactivate pathogens (Lee *et al.*, 2015). Microorganisms continuously enter and leave the bioreactor (<30 days of residence), in which they grow, die, and evolve. In this regard, each bioreactor may be viewed as an island at which ecological processes occur (Daims *et al.*, 2006). Therefore,

\* Corresponding author. E-mail: heedeung@korea.ac.kr;  
Tel: +82-2-3290-4861; Fax: +82-2-928-7656.

a bioreactor may serve as a platform for testing community assembly theories at a tractable temporal scale relevant to microbial growth and community turnover. We herein used an activated sludge bioreactor to evaluate our hypothesis that niche and neutral processes contribute differently to the local assembly of rare and common taxa. To test this hypothesis, we used a 2-year time series of microbial community sequencing from the Pohang WWTP in South Korea. We focused our analysis on a comparison of the temporal dynamics of rare and common taxa and examined the relative contributions of niche and neutral processes to community assembly.

## Materials and Methods

### *Sampling site, sample collection, and operational data*

The Pohang WWTP (36.00°N, 129.35°E) treats approximately 160,000 m<sup>3</sup> of domestic wastewater daily through a conventional activated sludge process (Rittmann and McCarty, 2001) with a low nutrient removal efficiency. Between May 2005 and April 2007, 12 samples were collected bimonthly at the same time of day from the bottom of the aeration tank of the treatment plant using a bucket grab sampler. Samples were immediately stored in an icebox and transferred to the laboratory for metacommunity DNA extraction. Plant operational data, including those related to biochemical oxygen demand, total nitrogen, total phosphorus, and mixed liquor, were collected. Suspended solids were analyzed using the standard method for the examination of water and wastewater (APHA *et al.*, 1989), and the data obtained are shown in Table S1. The flow rate, temperature, and dissolved oxygen level were monitored using installed online instruments.

### *DNA extraction, PCR, and pyrosequencing*

DNA from each sludge sample was extracted using the automated nucleic acid extractor (Obata *et al.*, 2001) Magstration System 6GC (PSS) following the manufacturer's protocol. Nucleic acid purity was assessed based on absorbance at 260 nm and concentration (ng L<sup>-1</sup>) quantified using ND-1000 (NanoDrop Technologies). The universal primers, 27F (5'-AGAGTTTGATCMTGGCT CAG-3') (Lane, 1991) and 518R (5'-ATTACCGCGGCTGCTG G-3') (Muyzer *et al.*, 1993), were used to amplify the bacterial 16S rRNA gene from each sample. Purified PCR amplicons from each sample were subjected to barcoded 454 pyrosequencing (Genome Sequencer FLX Titanium Series; 454 Life Science) at Macrogen (Seoul) following the manufacturer's protocol.

### *Sequence analysis*

Low-quality (<Q20) reads were removed from raw 16S rRNA sequences. Adaptor sequences were trimmed using the custom Perl script trimBarcode.pl (Macrogen). In addition, short reads (<300 bp) and potential chimeric sequences were removed using Mothur (chimera.uchime) (Schloss *et al.*, 2009). After quality filtering, the phylogenetic affiliations of the sequence reads were assigned using the RDP Classifier (Wang *et al.*, 2007) for the taxonomic database (version 11.5). Operational taxonomic units (OTUs) were selected based on 97% sequence identity using Mothur (RDP Aligner, distance matrix, and UCLUST). Among all sequence reads, 2,988 (35,856 sequences and 4,999 OTUs from 12 samples) were randomly selected to reduce the error associated with the OTU richness estimation (Roesch *et al.*, 2007). Singleton OTUs (OTUs with only one sequence) were excluded from further statistical analyses (calculation of bacterial richness and generation of OTU rank abundance curves).

### *Species (OTU) abundance distribution analysis*

To evaluate how OTU abundance distribution patterns fit the

theoretical models (*e.g.*, neutral and/or niche models), histograms of power-of-2 abundance classes (octave classes) of OTUs included in the rare-transient and common cohorts were plotted and then fit to the abundance models. Log-normal and log-series algorithms (Krebs, 1989) were applied to fit the model curves to each dataset, and the chi-squared test was used to evaluate the significance of each dataset. All statistical analyses were performed using PAST (version 2.12; <http://folk.uio.no/ohammer/past/>), and plots were generated using SigmaPlot (version 10.0; Jandel Scientific). A stacked bar plot of the relative distribution of transient-rare and other taxa at the phylum and proteobacterial class levels in the 12 activated sludge samples was generated using SigmaPlot 10.0.

### *Non-metric multidimensional analysis (NMDS)*

Temporal variations in transient-rare and other bacterial taxa were analyzed using NMDS ordination in PRIMER 6 (version 6.1.13; Primer-E) based on weighted UniFrac distances (Lozupone and Knight, 2005). A weighted UniFrac distance matrix for each group was generated for NMDS ordination by running the beta\_diversity.py command on the equal-sampled OTU table in QIIME (version 1.9.1, [Caporaso *et al.*, 2010]). The dimensionality of the ordination was selected by comparing the final stress values in R.

### *Multivariate analyses*

To assess the relative importance of environmental parameters in explaining patterns across the 12 activated sludge samples, a simplified redundancy analysis (RDA) and canonical correspondence analysis (CCA) were performed as described previously (ter Braak and Verdonschot, 1995; Van der Gucht *et al.*, 2007) using Canoco 4.5 (ter Braak and Smilauer, 2002). RDA and CCA were used for the common and transient-rare taxa, respectively, based on the results of a detrended correspondence analysis in Canoco 4.5. In both analyses, the total variation in the bacterial community matrix under unique environmental components with the corresponding *P* values was decomposed. The significance of each operational factor was evaluated with a Monte Carlo permutation test (999 permutations under the null hypothesis) using the forward selection method (ter Braak and Verdonschot, 1995). The explanatory environmental (*E*) variable was employed to measure the degree of variation (computed as the percentage of total variation for axes 1 and 2) using the multivariate extension of the linear regression with the corresponding *R*<sup>2</sup> values. Unexplained variation was calculated as (1-[*E*]). To test our hypothesis by assessing differences in the community structure between the transient-rare and common taxa, a permutational multivariate analysis of variance (PERMANOVA) was performed based on weighted UniFrac distances. Additionally, to test the null hypothesis (no difference in dispersion between groups), a permutational multivariate dispersion (PERMDISP) test was conducted. PERMANOVA and the PERMDISP test were performed using the "adonis" and "betadisper" functions of the vegan package in R, respectively.

### *Sloan and beta-abundance null model fitting*

To evaluate the contribution of the neutral processes to community assembly, the OTU occurrence frequency and abundance data were fit to the Sloan neutral community model (Sloan *et al.*, 2006). This model was developed to predict the relationships between the occurrence frequencies and mean relative abundances of taxa sampled from an activated bioreactor at a given time point. The goodness-of-fit of the observed dataset to the model was calculated using the following equation: model fit = 1 - SS<sub>err</sub>/SS<sub>total</sub> (generalized *R*<sup>2</sup>; [Östman *et al.*, 2010]), where SS<sub>err</sub> is the sum of squares of residuals and SS<sub>total</sub> is the total sum of squares. Moreover, to establish whether incorporating drift and dispersal limitations improves the model fit, fitness was compared between the neutral model and a binomial distribution model beyond the randomly sampled source metacommunity (Sloan *et al.*, 2006). All

plots and statistical analyses were performed with the minpack.lm package in R (R Core Team, 2015) (Elzhov *et al.*, 2013) using the script provided by Burns *et al.* (2015). Moreover, 95% confidence intervals around model predictions were calculated using the HMisc package (Wilson score interval) in R (Brown *et al.*, 2001; Harrell Jr. and Dupont, 2006). The Student's *t*-test was used to compare the taxonomic compositions of assemblages using R.

Beta-abundance null model fitting was performed based on both Bray–Curtis and weighted UniFrac distances for each transient-rare and common taxon OTU matrix using custom R scripts provided by Tucker *et al.* (Tucker *et al.*, 2016). Abundance null deviation values closer to zero indicate neutral communities in which species are ecologically equivalent to one another (Tucker *et al.*, 2016; Lee *et al.*, 2017).

#### Dynamic simulation of the community assembly dataset

To confirm the effects of definition changes on the OTU classification at the detection limit using a dynamic simulation, 60 different dataset combinations (12 different relative abundances and 5 different occurrence frequencies) of various relative abundances and the occurrence frequencies of transient-rare and common taxa were used. The species abundance histograms of both transient-rare and common taxa were plotted from each dataset. Log-series and log-normal regression curves were then fit, and the goodness-of-fit (chi-squared test) of each regression was tested using both observed and expected regression values (Table S6). The closed box in Table S6 represents the dataset with the same transient-rare and common taxa.

#### Accession numbers

The 16S rRNA gene sequences obtained in the present study are deposited in GenBank under the accession numbers HQ462572 to HQ524318.

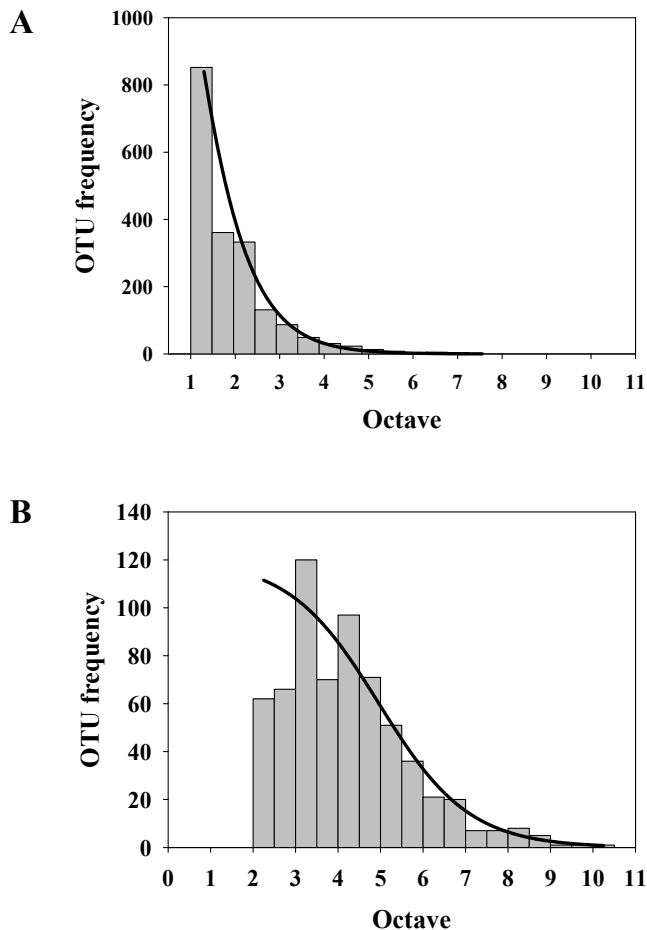
## Results

### Definition of rare taxa

We performed 16S rRNA pyrosequencing to assess microbial community diversity. After removing singleton sequence reads (reads with a single sequence occurring only once in the dataset), 2,544 unique OTUs were retrieved at a 3% identity cut-off from 35,856 sequence reads of 12 samples collected bimonthly over 2 years (Table S2). Ranked taxon abundance distribution (Fig. S1) showed few abundant (*i.e.*, common) taxa and many rare taxa, similar to the commonly reported distributions of microorganisms in the soil, sediment, freshwater, and seawater (Nemergut *et al.*, 2011; Kim *et al.*, 2013; Mariadassou *et al.*, 2015; Roguet *et al.*, 2015). Rare taxa were defined as OTUs comprising less than 1% of all sequence reads (<334 reads), and were further classified as transient- (<4 occurrences [33%]) and persistent-rare taxa (>8 occurrences [66%]). Transient-rare taxa constituted 74.7% of the observed OTUs and 28.7% of the total sequence reads, but were numerically less important than the common taxa (Table S4). Persistent-rare taxa constituted 1.5% of the observed OTUs and 6.2% of the total sequence reads. Thereafter, analyses were focused on comparisons between the transient-rare and common taxa.

### Ecological patterns of transient-rare taxa

OTU abundance distributions were analyzed based on the logarithmically binned abundance of the transient-rare and common taxa (Fig. 1). In the bioreactor, the two groups of taxon assemblages demonstrated very different patterns. The

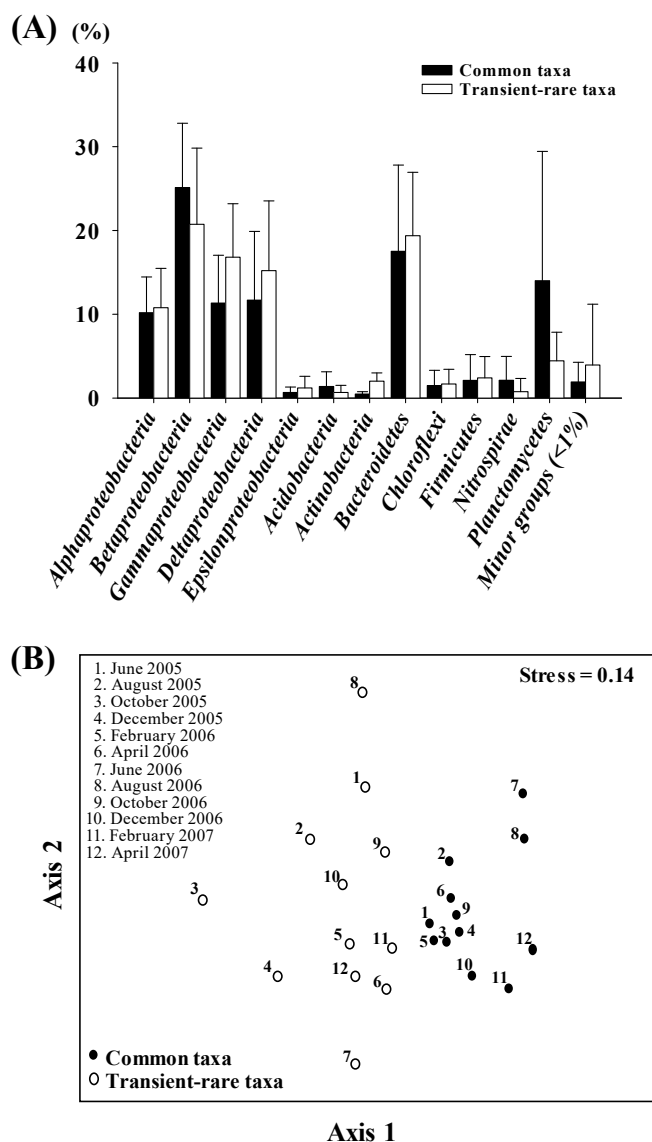


**Fig. 1.** Operational taxonomic unit abundance distribution of (A) transient-rare and (B) common taxa. The fitted lines predicting the frequency and abundance of each taxon were obtained based on log-normal and log-series models. The octaves represent power-of-2 abundance classes.

transient-rare taxa followed a log-series distribution ( $\chi^2=29.14$ ,  $P<0.0001$ ) (Fig. 1A), whereas the common taxa followed a near log-normal distribution ( $\chi^2=843.6$ ,  $P<0.0001$ ) (Fig. 1B). These results remained unchanged even when the definition of the transient-rare taxa was changed (relative abundance of 0.05–4.00% and occurrence frequency of 2–6; Table S6). Since species abundance distribution is affected by metacommunity diversity and the immigration rate (Brown and Kodric-Brown, 1977; Shmida and Wilson, 1985) and the taxon groups defined in this study were assembled from the same regional species pool, these two groups of taxa were expected to show similar abundance distributions. Accordingly, differences in their abundance distributions may be indicative of different contributions of niche and neutral processes in the bioreactor.

### Temporal variations in transient-rare taxon assembly

We assessed the community composition of the transient-rare and common taxa by taxonomically classifying the observed OTUs at the phylum and proteobacterial class levels (Fig. 2A). Overall, the taxonomic compositions of the two assemblages were very similar (*t*-test,  $P=1.000$ ). *Betaproteobacteria* and *Bacteroidetes* were the two major groups, followed by *Gammaproteobacteria*, *Alphaproteobacteria*,



**Fig. 2.** Taxonomic affiliations of transient-rare taxa and their temporal variability. (A) Classification of the transient-rare and common taxa according to the phylum and proteobacterial class levels. The error bars indicate the SD of 12 samples. (B) Non-metric multidimensional analysis ordination based on the weighted UniFrac distances for the assemblages of the transient-rare and common taxa.

and *Deltaproteobacteria*. Nevertheless, a few groups (e.g., *Epsilonproteobacteria*, *Nitrospirae*, and *Actinobacteria*) showed significant differences between the transient-rare and common taxa ( $P < 0.05$ ). The composition of the two cohorts was evidently different at the family level. A Venn diagram analysis demonstrated that specific families belonged to each cohort. Therefore, among 162 families, 60 and 11 families were only observed in the transient-rare and common cohorts, respectively (Fig. S2). Furthermore, the temporal community dynamics of the transient-rare and common taxa were compared across the samples. The relative abundance of the transient-rare taxa ranged between 18.4 and 40.9% and that of the common taxa ranged between 58.5 and 85.4%; however, the overall community composition of the two cohorts remained relatively stable over time (Fig. S3). The transient-rare taxa were dominated

by *Betaproteobacteria* (average and standard deviation:  $6.0 \pm 3.0\%$ ) and *Gammaproteobacteria* (average and standard deviation:  $5.0 \pm 2.5\%$ ), while the common taxa were dominated by *Betaproteobacteria* (average and standard deviation:  $17.7 \pm 5.2\%$ ) and *Bacteroidetes* (average and standard deviation:  $12.3 \pm 6.6\%$ ). However, the relative abundance of the transient-rare and common taxa did not significantly differ across the samples over time ( $P > 0.05$ ). Moreover, temporal variations were investigated using NMDS ordination based on weighted UniFrac distances (Lozupone and Knight, 2005) between samples (Fig. 2B), which demonstrated that the phylogenetic distances between assemblages of the transient-rare taxa (average, 0.366 in 12 samples) were greater than those between assemblages of the common taxa (average, 0.275 in 12 samples). This result indicates a higher temporal variation in the transient-rare taxa than in the common taxa with respect to the phylogenetic distance (PERMDISP,  $F = 8.011$  and  $P = 0.013$ ). In addition, distinct clusters of the assemblages of the transient-rare and common taxa demonstrated that the two assemblages were phylogenetically distinct (PERMANOVA,  $F = 8.011$  and  $P < 0.001$ ) at the OTU level (97% sequence identity), which was not evident at the phylum and proteobacterial class levels (Fig. 2A). PERMANOVA and the PERMDISP test for the two cohorts based on unweighted UniFrac distances also supported our inference.

#### *Deterministic and stochastic processes contributing to the community assembly*

The effects of explanatory variables related to the bioreactor (i.e., environmental and operational parameters) on temporal variations in the two assemblages were assessed by quantifying their explanatory power using multivariate analyses (CCA for the transient-rare taxa and RDA for the common taxa) (Kelley, 1940). Twelve deterministic factors (temperature, dissolved oxygen, pH, hydraulic retention time, solid retention time, and mixed liquor suspended solids in the bioreactor and the biochemical oxygen demand, total nitrogen, and total phosphorus of the influent and effluent; Table S3) were tested, and the significance of differences in community assembly and environmental factors was estimated (Table S3). Collectively, these parameters may explain 23.5 and 63.4% of the variations in the transient-rare and common taxa, respectively. The significance of the results of RDA and CCA was also calculated for both the transient-rare and common taxa (Table S3). While the scores for both axes were significant for the transient-rare taxa (both  $P < 0.01$ ), the score was significant for axis 1 alone for the common taxa (for axes 1 and 2,  $P = 0.04$  and 0.382, respectively).

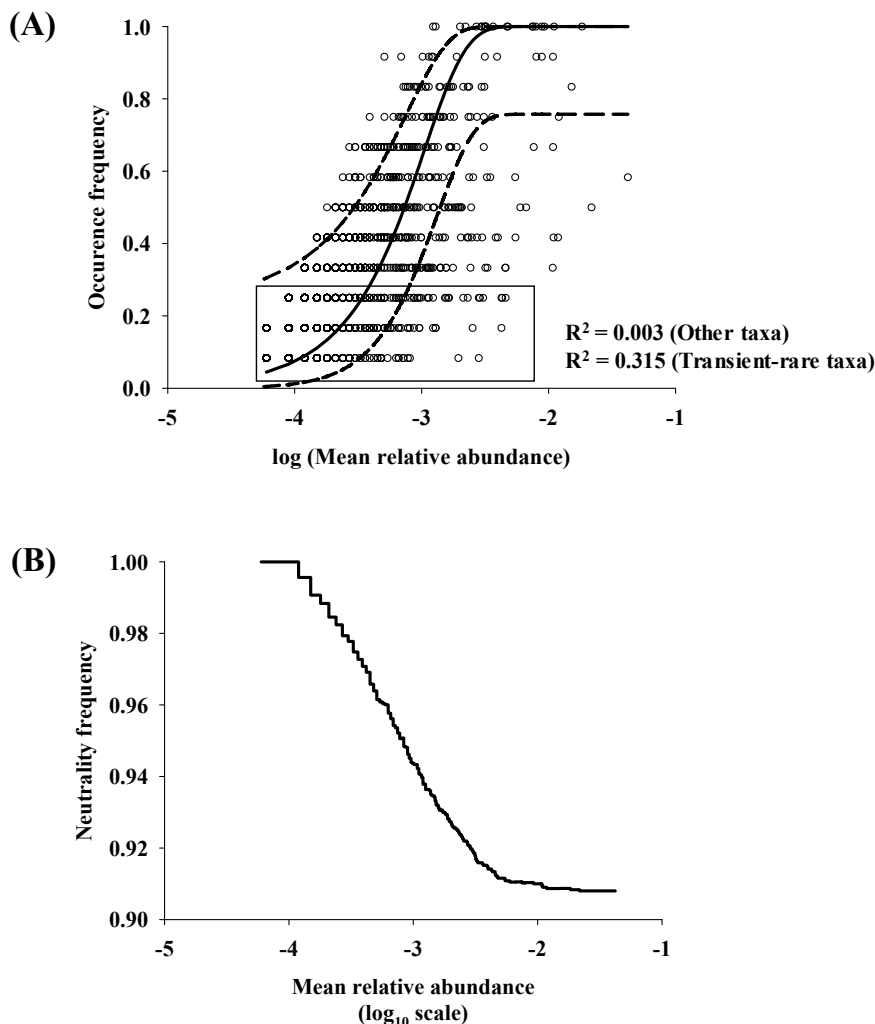
Furthermore, the contribution of neutral processes to the assembly of the transient-rare taxa was tested by fitting the OTU occurrence frequency and abundance data to the Sloan neutral community model (Sloan *et al.*, 2006). This model was developed to explain the occurrence and abundance patterns of prokaryotic communities based on dispersal and ecological drift and successfully described the neutral assembly of bacterial communities in activated sludge bioreactors (Ofiteru *et al.*, 2010), lakes (Roguet *et al.*, 2015), human lungs (Venkataraman *et al.*, 2015), trees (Woodcock

*et al.*, 2007), and zebrafish guts (Burns *et al.*, 2015). Fig. 3A shows the goodness-of-fit scores of the neutral model to all OTUs observed in the present study (solid line) with 95% confidence intervals of the prediction (dashed lines). The proportion of taxa (inserted box in Fig. 3A) within 95% confidence intervals was 96.4% (relative abundance, 71.4%) for the transient-rare taxa and 73.3% (relative abundance, 45.6%) for the common taxa, and both assemblages followed the neutral community model (Table S4). The neutrality fraction (defined as the fraction of OTUs within 95% confidence intervals of the neutral model) was inversely proportional to the mean relative abundance (Fig. 3B), and this result remained unchanged even when the definition of the transient-rare taxa was changed (relative abundance of 0.05–4.00% and occurrence frequency of 2–6; Table S6). Therefore, when taxa are relatively rare, neutral processes are more important for explaining their patterns.

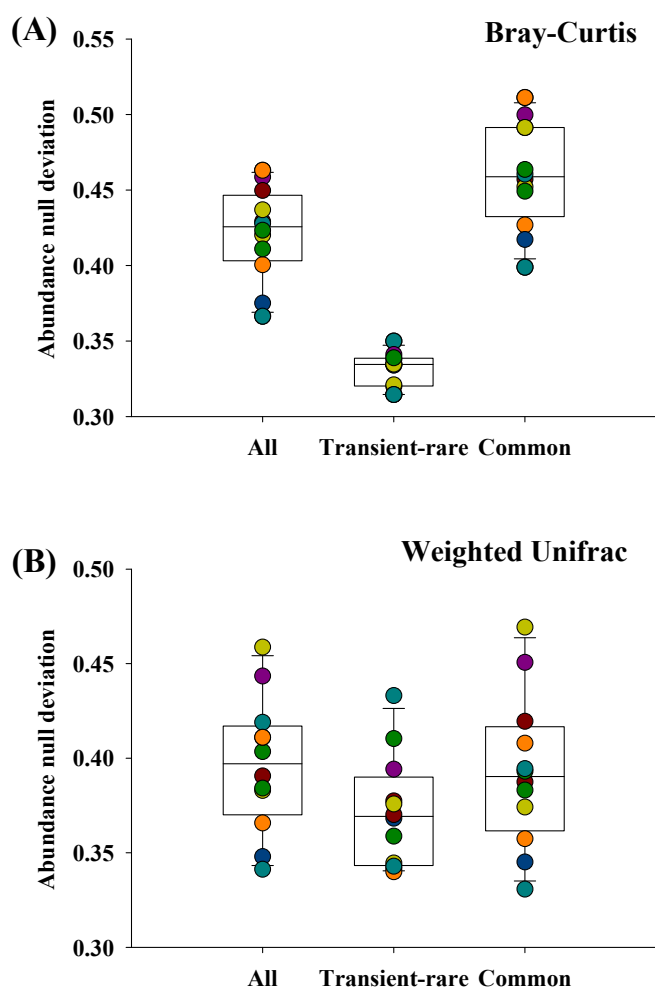
Additionally, to understand the deviation of the observed differences in beta-diversity from null expectations, abundance-based beta-null approaches were used to distinguish between the niche and neutral processes, as described by Tucker *et al.* (Tucker *et al.*, 2016) and Lee *et al.* (Lee *et*

*al.*, 2017). In this comparative approach, deviations to and from a permuted null expectation (neutral) were used to interpret the relative contributions of the neutral and niche processes, respectively. Transient-rare communities deviated from the null expectation, with the transient-rare taxa occurring closer to the null expectation than the common taxa (Fig. 4). In addition, the common taxa showed a significantly higher beta-null deviation than the transient-rare taxa ( $P < 0.05$ ).

In the present study, the transient-rare taxa were defined as those with an occurrence frequency of  $< 4$  and a relative abundance of  $< 1\%$ . However, even minor changes in OTU abundance may change the classification of the transient-rare and common taxa. These noises are frequently observed in the OTUs around the detection limit. When both occurrence frequency and relative abundance conditions were applied together to construct the dataset, the relative abundance of the transient-rare taxa was  $< 0.5\%$  and the occurrence frequency ranged between  $< 3$  and  $< 5$ . Slightly different combinations of the transient-rare and common taxa were observed in this dataset, with a relative abundance of  $< 0.8\%$  and an occurrence frequency ranging between  $< 6$  and  $< 7$ .



**Fig. 3.** Evaluation of neutral community assembly. (A) Fitting of the operational taxonomic units (OTUs) observed in the bioreactor to the Sloan neutral community model. The solid line indicates the best fitting line, and dashed lines represent the 95% confidence intervals of the model. The transient-rare taxa are indicated in the inserted box. (B) Neutrality fraction of the observed OTUs as a function of mean relative abundance.



**Fig. 4.** Relative changes in niche and neutral processes assessed based on deviations from abundance-weighted beta-null models. Abundance null deviations of taxon assembly based on (A) Bray–Curtis and (B) weighted UniFrac distances. Colored circles indicate individual samples. In both Bray–Curtis and weighted UniFrac resemblances, the transient-rare and common taxa showed distinct null deviations ( $P < 0.05$ ).

In all tested dataset combinations, the transient-rare taxa mostly fit well to the log-series regression curves (red area in Table S6A), except in 10 combinations (blue area in Table S6A). In contrast, the common taxa fit better to the log-normal regression curves than to the log-series regression curves (blue area in Table S6B). The highest log-normal goodness-of-fit score for the common taxa was observed in a dataset with an occurrence frequency of  $>6$  and a relative abundance ranging between  $>0.075$  and  $0.8\%$  ( $\chi^2=10.8$ ); however, a relatively higher log-series goodness-of-fit score was obtained for the transient-rare taxa in the same dataset ( $\chi^2=603.8$ ). The lowest goodness-of-fit score ( $\chi^2=89.4$ ) for the transient-rare taxa was obtained in a dataset with an occurrence frequency of  $>4$  and relative abundance of  $>1\%$ ; however, the log-normal goodness-of-fit score for the common taxa ( $\chi^2=32.8$ ) was slightly higher than the lowest value for the transient-rare taxa ( $\chi^2=10.8$ ). Therefore, based on the best results of the chi-squared test on various combinations of the community dataset (occurrence frequency of 1–12 and relative abundance of 0.05–4.00%), we defined the transient-rare taxa as those with an

occurrence frequency of  $<4$  and relative abundance of  $<1\%$ . Collectively, these results suggest that the species abundance distributions of the transient-rare and common taxa did not shift due to sampling artifacts.

## Discussion

Recent advances in high-throughput sequencing technologies, together with increased computational performance, have enabled us to uncover the members of rare biospheres in diverse ecosystems (Sogin *et al.*, 2006; Pedrós-Alió, 2012). Although this has further opened avenues for characterizing and understanding the microbial rare biosphere, how and why numerous species are rare remain unclear (Pedrós-Alió, 2012). Accordingly, the present study demonstrated the contributions of the transient-rare taxa to local community dynamics (in an activated sludge bioreactor). Environmental conditions and operational parameters minimally accounted for the variations observed in the transient-rare taxa in the bioreactor (25.2%), implying that abiotic deterministic factors did not play a prominent role in driving the assembly of transient-rare taxa. The remaining high proportion of the unexplained variation (76.5%) in the transient-rare taxa may be explained by unmeasured deterministic factors, microbial interactions (*e.g.*, viral infection), and neutral factors. However, neutral processes may explain most of the variations in the transient-rare taxa, as evidenced by the Sloan neutral community model (Fig. 3A). Neutral processes describe the occurrence of the transient-rare taxa as the stochastic replacement of open sites (generated by random death or emigration) by other taxa in the inflow or within the bioreactor. These findings indicate that neutral processes are likely to be more important than deterministic ones for the assembly of transient-rare taxa in the bioreactor. Similar findings have been reported in several non-bacterial ecosystems. Based on a study of six rare and six abundant amoeba species in a soil ecosystem, Finlay *et al.* (Finlay *et al.*, 2001) demonstrated that rare species followed a Poisson distribution (*i.e.*, random distribution). In addition, Magurran and Henderson (Magurran and Henderson, 2003) examined an estuarine fish community using a 21-year dataset and demonstrated that the abundance of occasional species was low, and they followed a random distribution in a headland ecosystem. These findings, together with the present results, suggest that the neutral assembly of the transient-rare or occasional species may be a common phenomenon in various organisms and ecosystems.

Diverse bacterial species from the influent wastewater and atmosphere continuously arrive in activated sludge bioreactors (Lee *et al.*, 2015), which may affect the bacterial community compositions of these bioreactors. Some bacteria arriving in the bioreactors are selected by deterministic factors and, in turn, these proliferate in the bioreactors, whereas the other bacteria are not selected. The transient-rare taxa arrive in the bioreactors infrequently and at low numbers, but are not selected by deterministic factors. However, it remains unclear which factors (deterministic or neutral) drive the assembly of persistent-rare taxa. Such taxa may be assembled similar to the transient-rare taxa (*i.e.*, neutral assembly), and they may frequently immigrate to the

bioreactor, but are not selected, suggesting that their standing populations are supported by continuous dispersal (e.g., the mass effect [Leibold *et al.*, 2004]). Another possibility is that abiotic and/or biotic conditions maintain these taxa in the bioreactor, but grow markedly slower than the abundant taxa (Fuhrman, 2009). We also fit our cohort of persistent-rare taxa to the Sloan neutral community model (Fig. 3A and Table S8). Notably, their percentage was markedly lower than that of the transient-rare taxa (96.4%), suggesting that the second scenario explains the assembly of the persistent-rare taxa. Additionally, we performed beta-null deviation analyses and found that the transient-rare taxa occurred closer to the neutral expectations than the common taxa (Fig. 4). In addition, the common taxa showed significantly higher beta-null deviations than the transient-rare taxa ( $P < 0.05$ ). Furthermore, the phylogenetic composition of the transient- and persistent-rare taxa was distinct. The persistent-rare taxa included significantly higher proportions of *Betaproteobacteria* and *Planctomycetes*, but lower proportions of *Alphaproteobacteria*, *Gammaproteobacteria*, and *Bacteroidetes* than the transient-rare ones (Fig. 5). In our simulation test, the persistent-rare taxa with a relative abundance of  $>1\%$  and an occurrence frequency of  $>10$  showed a better goodness-of-fit for the log-normal distribution than the log-series distribution, indicating that the distribution of the persistent-rare taxa was affected by deterministic factors of the Pohang WWTP (Table S8). Overall, these results suggest that the two rare taxon assemblages differed in terms of their ecological roles in the bioreactor.

Difficulties are associated with predicting community assembly processes based on ecological patterns because a similar pattern may be generated by interactions among multiple processes (Hanson *et al.*, 2012) and different processes may predict similar patterns (Harpole, 2010). Never-

theless, community assembly processes may be projected in ecological patterns in a certain manner (Fargione *et al.*, 2003). In this regard, different ecological patterns observed between the transient-rare and common taxa (e.g., taxa abundance distribution in Fig. 1) suggest that different community assembly processes acted on these two groups of assemblages. Similar to the present results, the transient-rare and common taxa showed a log-series and log-normal distribution, respectively, in other assemblages, including marine prokaryotes (Galand *et al.*, 2009), estuarine fish (Magurran and Henderson, 2003), and terrestrial insects (Southwood, 1996). However, the patterns of community similarity decay over time (“time-decay” [Korhonen *et al.*, 2010]) were not significant for both cohorts (Fig. 6).

In the present study, the transient-rare taxa were defined as those with an occurrence frequency of  $<4$  and relative abundance of  $<1\%$ . To confirm the effects of changes in this definition on the OTU classification at the detection limit, a dynamic simulation of 60 different dataset combinations was performed. The simulation results indicated that sampling artifacts were unlikely to shift the species abundance distributions of the transient-rare and common taxa (Table S6). In addition, we evaluated the neutrality of both transient-rare and common taxon assemblages in response to the dynamic OTU abundance and occurrence frequency based on the Sloan neutral community model fit (Table S7). The results obtained showed that even when relative abundance and occurrence frequency were changed, up to 90% OTUs belonging to the transient-rare taxa showed stronger neutrality than the common taxa. Specifically, the transient-rare taxa with an occurrence frequency of 2–5 showed significant neutrality in this analysis, indicating that the overarching patterns remain the same regardless of nuances in the thresholds applied.

Several explanations and conceptual models have been proposed for the roles of rare microbial taxa. The most common explanation is that rare microbial taxa do not actively grow in local communities (Pedrós-Alió, 2012), and, in turn, simply pass through the local systems (Shade *et al.*, 2014). Jones and Lennon (2010) investigated the activity of bacte-

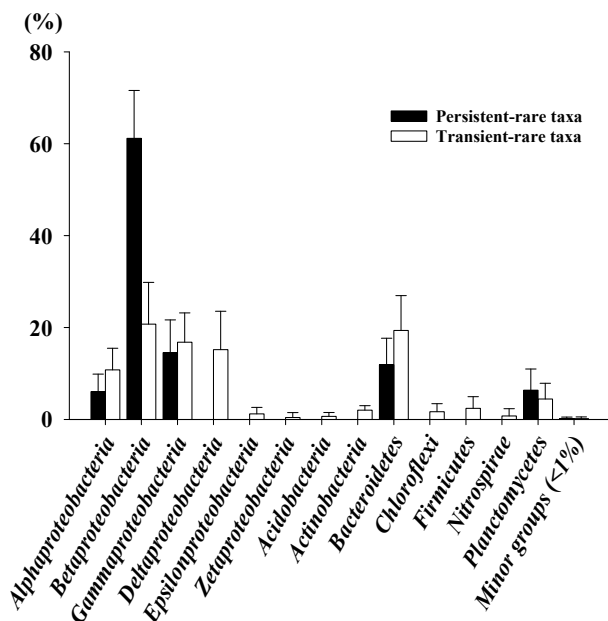


Fig. 5. Classification of transient-rare (open bars) and persistent-rare (solid bars) taxa at phylum and proteobacterial class levels. The error bars indicate the SD of 12 samples.

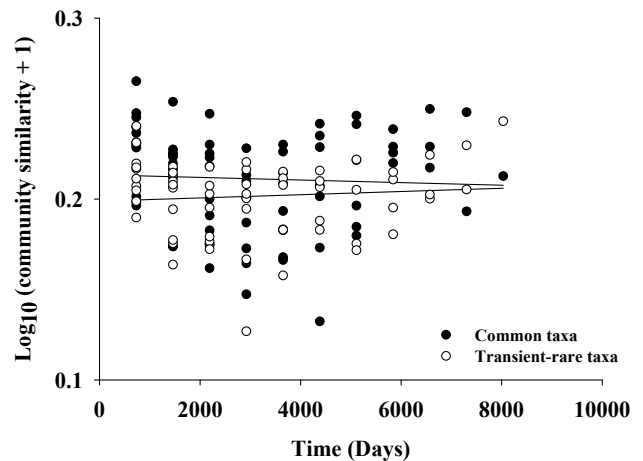


Fig. 6. Time-decay relationships for transient-rare and common taxa. Data were fit to a model describing the relationship between decayed richness and time based on the calculation method described by Korhonen *et al.* (2010).

rial taxa (relative recovery of 16S rRNA relative to 16S rDNA) across abundance ranks in two lakes and observed that many rare bacterial taxa showed high relative recovery of 16S rRNA. Based on this observation, they claimed that rare bacterial taxa serve as a seedbank (Lennon and Jones, 2011) and respond to environmental changes. Furthermore, Shade *et al.* (2014) demonstrated that bacterial and archaeal rare taxa occasionally become abundant (conditionally rare taxa) in diverse ecosystems and then contribute to temporal changes in bacterial and archaeal diversity. This analysis included a distinction between taxa that were observed only once and those that were more persistent within a community. In this regard, although the transient-rare taxa defined in the present study had low abundance and were generally insensitive to the available resources in the bioreactor, they may become abundant in response to perturbation or changing conditions in the future. We observed a portion of the taxa within the present study that may be classified as exhibiting “conditional rarity” (2.0%), and all these conditionally rare taxa belonged to the persistent-rare cohort and not to the transient-rare cohort. They were occasionally rare, but remained prevalent during some periods. Notably, the neutral models often fit well to microbial communities (Sloan *et al.*, 2006; Woodcock *et al.*, 2007; Roguet *et al.*, 2015; Venkataraman *et al.*, 2015), which may have been due to the high proportion of transient-rare taxa in microbial communities because of high dispersal (Shade *et al.*, 2014). Nevertheless, the present study suggests that the transient-rare taxa may arguably not contribute to the key functional processes within the community in an activated sludge bioreactor.

It currently remains unclear whether the transient-rare taxa contribute to key functional processes within the bioreactor. However, previous studies suggested that these taxa are important for serving novel or redundant functions when the environment changes (Shade *et al.*, 2014). As a “microbial seedbank”, rare species are important because they contribute to the genetic diversity of the microbial community (Fuhrman, 2009) and potentially contribute to ecological functions if they exhibit blooming dynamics (Fuhrman, 2009; Shade *et al.*, 2014). Accordingly, rare taxa may play pivotal ecological roles in local communities when the environmental conditions in the bioreactor change. The present results demonstrating the importance of neutral processes in the assembly of the transient-rare taxa within a local community deepen our understanding of the various roles and dynamics of members of the microbial rare biosphere. In studies on communities, the transient-rare taxa impart significant noise due to their stochastic behavior instead of interactions with abiotic and biotic factors; however, they have been proposed to be important after awakening from dormancy (Jones and Lennon, 2010), and their neutrality is not considered to be permanent because of their existence at the limit of detection (Shade *et al.*, 2014; Shade and Gilbert, 2015). Therefore, if we remove these taxa from such studies, we may be able to further our understanding of the standing community and better explain deterministic drivers. Furthermore, the present results showing that the relative importance of diverse assembly processes is specific and that the composition of the transient- and persistent-rare

taxa is distinct suggest that rare taxa with various occurrence patterns play ecological roles within their local communities (Shade and Gilbert, 2015).

The present results indicate that the community assembly of the transient-rare taxa is highly diverse and mostly affected by stochastic processes in the bioreactor. In contrast, the common taxa show relatively low diversity in an activated sludge bioreactor. Moreover, the common taxa may play important ecological roles in a WWTP because their distribution patterns are affected more by the operational conditions in an activated bioreactor (niche processes); therefore, the common taxa may be used as indicators of performance when evaluating bioreactor function.

In conclusion, microbial communities are dynamic in space and time, and their collective dynamics are underpinned by changes in discrete microbial populations, the relative contributions of which to the community fluctuate. The present study focused on a prominent subset of microbial populations within their community—the transient-rare taxa. Since microbial communities are species-rich, rare taxa often comprise the majority of the total observed taxa within these communities. We found that the neutral patterns of assembly best described the dynamics of these transient-rare taxa, supporting the notion that these taxa are not the contributing members of the community, but are rather driven by stochastic forces. Therefore, the dynamics of the transient-rare taxa may be explained by their patterns of dispersal in regional metacommunities.

#### Acknowledgements

We thank Prof. Hwang at POSTEC for providing DNA samples of the Pohang WWTP. We also thank Prof. James Tiedje and Prof. Ashley Shade at MSU for their critical reading of the manuscript and their advice. This study was supported by the National Research Foundation of Korea (NRF-2018R1A2B2002110).

#### References

- APHA, AWWA, and WPCF. (1989) *Standard Methods for the Examination of Water and Wastewater*. Washington, D.C.: American Public Health Association, American Water Works Association, and Water Environment Federation.
- Ayarza, J.M., and Erijman, L. (2011) Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. *Microb Ecol* **61**: 486–495.
- Briones, A., and Raskin, L. (2003) Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *Curr Opin Biotechnol* **14**: 270–276.
- Brown, J.H., and Kodric-Brown, A. (1977) Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology* **58**: 445–449.
- Brown, L.D., Cai, T.T., and DasGupta, A. (2001) Interval estimation for a binomial proportion. *Stat Sci* **16**: 101–117.
- Burns, A.R., Stephens, W.Z., Stagaman, K., Wong, S., Rawls, J.F., Guillemin, K., and Bohannan, B.J.M. (2015) Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J* **10**: 655–664.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Daims, H., Taylor, M.W., and Wagner, M. (2006) Wastewater treatment: a model system for microbial ecology. *Trends Biotechnol* **24**: 483–489.



- Elzhov, T.V., Mullen, K.M., and Bolker, B. (2013) minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds. URL <http://CRAN.R-project.org/package=minpack.lm>
- Fargione, J., Brown, C.S., and Tilman, D. (2003) Community assembly and invasion: an experimental test of neutral versus niche processes. *Proc Natl Acad Sci U S A* **100**: 8916–8920.
- Finlay, B.J., Esteban, G.F., Clarke, K.J., and Olmo, J.L. (2001) Biodiversity of terrestrial protozoa appears homogeneous across local and global spatial scales. *Protist* **152**: 355–366.
- Fuhrman, J.A. (2009) Microbial community structure and its functional implications. *Nature* **459**: 193–199.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* **106**: 22427–22432.
- Grady, K.L., Sorensen, J.W., Stopnisek, N., Guittar, J., and Shade, A. (2019) Assembly and seasonality of core phyllosphere microbiota on perennial biofuel crops. *Nat Commun* **10**: 1–10.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- Harpole, W. (2010) Neutral theory of species diversity. *Nature Education Knowledge Project* **1**: 31.
- Harrell Jr., F.E., and Dupont, M.C. (2006) The Hmisc Package. *R package version*: 3.0-12. URL <https://cran.r-project.org/web/packages/Hmisc/index.html>
- Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*, vol. 32. Princeton, NJ: Princeton University Press.
- Jones, S.E., and Lennon, J.T. (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci U S A* **107**: 5881–5886.
- Kelley, T.L. (1940) *Talents and Tasks: Their Conjunction in a Democracy for Wholesome Living and National Defense*. Cambridge, MA: Harvard Graduate School of Education.
- Kim, T-S, Jeong, J-Y, Wells, G.F., and Park, H-D. (2013) General and rare bacterial taxa demonstrating different temporal dynamic patterns in an activated sludge bioreactor. *Appl Microbiol Biotechnol* **97**: 1755–1765.
- Korhonen, J.J., Soininen, J., and Hillebrand, H. (2010) A quantitative analysis of temporal turnover in aquatic species assemblages across ecosystems. *Ecology* **91**: 508–517.
- Krebs, C.J. (1989) *Ecological Methodology*. New York, NY: Harper & Row.
- Lane, D.J. 16S/23S rRNA Sequencing. (1991) In *Nucleic Acid Techniques in Bacterial Systematics*. Stackebrandt, E., and Goodfellow, M. (eds). New York, NY: John Wiley & Sons, pp. 115–175.
- Lee, S-H., Kang, H-J., and Park, H-D. (2015) Influence of influent wastewater communities on temporal variation of activated sludge communities. *Water Res* **73**: 132–144.
- Lee, S.H., Sorensen, J.W., Grady, K.L., Tobin, T.C., and Shade, A. (2017) Divergent extremes but convergent recovery of bacterial and archaeal soil communities to an ongoing subterranean coal mine fire. *ISME J* **11**: 1447–1459.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., et al. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* **7**: 601–613.
- Lennon, J.T., and Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* **9**: 119–130.
- Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.
- Magurran, A.E., and Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**: 714–716.
- Mariadassou, M., Pichon, S., and Ebert, D. (2015) Microbial ecosystems are dominated by specialist taxa. *Ecol Lett* **18**: 974–982.
- Morella, N.M., Weng, F.C-H., Joubert, P.M., Metcalf, C.J.E., Lindow, S., and Koskella, B. (2020) Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection. *Proc Natl Acad Sci U S A* **117**: 1148–1159.
- Muyzer, G., de Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695–700.
- Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G., and Renella, G. (2003) Microbial diversity and soil functions. *Eur J Soil Sci* **54**: 655–670.
- Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., et al. (2011) Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**: 135–144.
- Nemergut, D.R., Schmidt, S.K., Fukami, T., O’Neill, S.P., Bilinski, T.M., Stanish, L.F., et al. (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* **77**: 342–356.
- Nemergut, D.R., Shade, A., and Violle, C. (2014) When, where and how does microbial community composition matter? *Front Microbiol* **5**: 497.
- Obata, K., Segawa, O., Yakabe, M., Ishida, Y., Kuroita, T., Ikeda, K., et al. (2001) Development of a novel method for operating magnetic particles, Magtration Technology, and its use for automating nucleic acid purification. *J Biosci Bioeng* **91**: 500–503.
- Ofiteru, I.D., Lunn, M., Curtis, T.P., Wells, G.F., Criddle, C.S., Francis, C.A., and Sloan, W.T. (2010) Combined niche and neutral effects in a microbial wastewater treatment community. *Proc Natl Acad Sci U S A* **107**: 15345–15350.
- Östman, Ö., Drakare, S., Kritzberg, E.S., Langenheder, S., Logue, J.B., and Lindström, E.S. (2010) Regional invariance among microbial communities. *Ecol Lett* **13**: 118–127.
- Pedrós-Alió, C. (2012) The rare bacterial biosphere. *Ann Rev Mar Sci* **4**: 449–466.
- Rittmann, B.E., and McCarty, P.E. (2001) *Environmental Biotechnology: Principles and Applications*. New York, NY: McGraw-Hill Higher Education.
- Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., et al. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**: 283–290.
- Roguet, A., Laigle, G.S., Theriault, C., Bressy, A., Soullignac, F., Catherine, A., et al. (2015) Neutral community model explains the bacterial community assembly in freshwater lakes. *FEMS Microbiol Ecol* **91**: fiv125.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J.A. (2014) Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* **5**: e01371-14.
- Shade, A., and Gilbert, J.A. (2015) Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends Microbiol* **23**: 335–340.
- Shade, A., and Stopnisek, N. (2019) Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr Opin Microbiol* **49**: 50–58.
- Shmida, A., and Wilson, M.V. (1985) Biological determinants of species diversity. *J Biogeogr* **12**: 1–20.
- Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., and Curtis, T.P. (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**: 732–740.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., et al. (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci U S A* **103**: 12115–12120.
- Southwood, T.R.E. (1996) The Croonian lecture, 1995 Natural communities: structure and dynamics. *Philos Trans R Soc, B* **351**: 1113–1129.
- Tchobanoglous, G., and Burton, F.L. (1991) Wastewater engineering. *Management* **7**: 1–4.
- ter Braak, C.J.F., and Verdonschot, P.F.M. (1995) Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat Sci* **57**: 255–289.
- ter Braak, C.J.F., and Smilauer, P. (2002) *Canoco for Windows version 4.5*. Wageningen, Netherlands: Biometris–Plant Research International. URL <https://canoco-for-windows.software.informer.com/4.5/>

- Torsvik, V., Øvreås, L., and Thingstad, T.F. (2002) Prokaryotic diversity —magnitude, dynamics, and controlling factors. *Science* **296**: 1064–1066.
- Tucker, C.M., Shoemaker, L.G., Davies, K.F., Nemergut, D.R., and Melbourne, B.A. (2016) Differentiating between niche and neutral assembly in metacommunities using null models of  $\beta$ -diversity. *Oikos* **125**: 778–789.
- Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.
- Van der Gucht, K., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., *et al.* (2007) The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proc Natl Acad Sci U S A* **104**: 20404–20409.
- Venkataraman, A., Bassis, C.M., Beck, J.M., Young, V.B., Curtis, J.L., Huffnagle, G.B., and Schmidt, T.M. (2015) Application of a neutral community model to assess structuring of the human lung microbiome. *mBio* **6**: e02284-14.
- Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Woodcock, S., van der Gast, C.J., Bell, T., Lunn, M., Curtis, T.P., Head, I.M., and Sloan, W.T. (2007) Neutral assembly of bacterial communities. *FEMS Microbiol Ecol* **62**: 171–180.