

## BRIEF COMMUNICATION OPEN



# Evaluation of influent microbial immigration to activated sludge is affected by different-sized community segregation

Bing Guo<sup>1</sup>, Zhiya Sheng<sup>1</sup> and Yang Liu<sup>1</sup>✉

Activated sludge (AS) microbial communities were analyzed for seasonal variation, a disturbance-recovery event, and separated small aggregates (SAG) to study the influent immigration effect using both neutral immigration model and mass-balance model with operational parameters. SAG differed with AS, and higher immigration impact on SAG was confirmed by both models. Adding the SAG community segregation in the latter model to evaluate the contribution of influent immigration to community disturbance-recovery showed increased impact of immigration.

*npj Clean Water* (2021)4:20; <https://doi.org/10.1038/s41545-021-00112-7>

## INTRODUCTION

The activated sludge (AS) wastewater treatment plant provides a well-defined model system to study microbial ecology. The AS community is shaped by deterministic (Niche) factors and stochastic (Neutral) factors (e.g., neutral immigration from wastewater influent)<sup>1</sup>. Two types of numerical models have been used to quantify the influent immigration effect. One is the neutral community (immigration) model (NIM)<sup>2</sup>, employing the abundance and frequency of operational taxonomic units (OTUs) in metacommunities to calculate the community-level migration probability<sup>1,3–6</sup>. The other is mass-balance model (MBM) that defines the immigration as the proportion of reactor biomass derived from influent biomass  $\left(\frac{\text{influent biomass}}{\text{influent biomass} + \text{local growth}}\right)$ , using abundances of specific OTUs to calculate the immigration rates and net growth rates<sup>7</sup>. The NIM immigration describes the probability that a dead individual in the reactor is replaced by an immigrant from the influent, differentiated from the reproduction of local community. Intrinsically, the MBM bases on the fractionation of influent biomass and local growth, which also aims to differentiate the influent immigration and local community. The MBM has been increasingly used because of the quantitative measurements of the OTU growth activity<sup>8</sup>. Although steady-state metacommunities demonstrated that operational parameters (e.g., solid retention time, SRT) are important for the evaluation of immigration<sup>4,9</sup>, the operation parameters were not directly employed in the original MBM. Recently, Frigon and Wells<sup>10</sup> and Guo<sup>11</sup> developed the mass-flow immigration model, which is a MBM incorporating operational parameters (MBM-OP), to calculate the immigration rate of specific OTUs. By using operational parameters, this model is comprehensible for both wastewater engineers and microbiologists.

In wastewater treatment bioreactors, community variation has been shown in relation to the size and status (settled and unsettled) of the aggregates<sup>12,13</sup>. In conventional AS, the relative importance of influent immigration on different-sized aggregate communities (i.e., easily settleable large-size flocs and low-settleability small-size particles) has not been explored. Besides steady-state modeling, the bioreactors often experience disturbances<sup>9,14</sup>, such as sudden high flow rates which lead to a biomass washout. The non-steady-state community serves as a good example to study the community recovery and to compare

the contributions of local community growth and influent immigration.

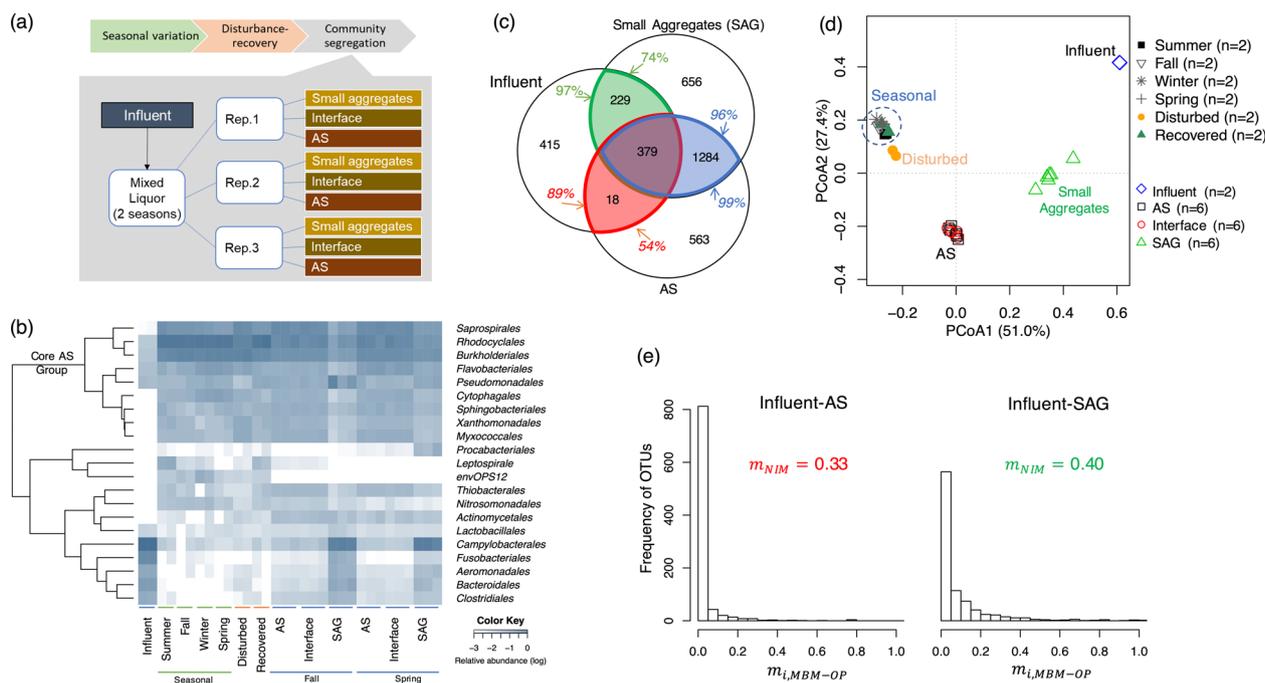
Our study aimed at monitoring the long-term variation of AS communities in different seasons and comparing the influent immigration effect on different-sized AS communities using NIM and MBM-OP models. In addition, the short-term non-steady-state (under operational high-flow disturbance) AS community's recovery attributed to local community growth and influent immigration was evaluated.

## AS community variation

The AS communities (between 2013 and 2019) showed that abundant microorganisms were relatively consistent during the long-term observation, and consisted of a core group at the order level (Fig. 1b) and at genus level (Supplementary Fig. S1). Seasonal changes and disturbance did not alter the core group<sup>14</sup>. The variation between small aggregates (SAG) and AS was clearly shown (Fig. 1b), being consistent in different seasons. The most abundant members in SAG were different from AS communities, but similar with the influent community's most abundant members, indicating that influent microbiome is an important origin for the SAG. The community richness (number of observed OTUs) was slightly higher in SAG than in AS (Supplementary Fig. S2), inferring that the SAG community was a union of AS and influent communities. The shared numbers and abundances of OTUs are indicators of immigration effect<sup>7,9,15</sup>. A large number of OTUs were shared among the influent, AS, and SAG (Fig. 1c). The shared OTUs between influent and AS took up  $89.1 \pm 0.1\%$  of the influent community total number of reads and  $54.0 \pm 4.3\%$  of the AS community total number of reads, which increased between influent and SAG ( $97.0 \pm 0.1\%$  and  $73.8 \pm 3.0\%$ , respectively), indicating higher immigration effect for influent–SAG than for influent–AS. The Bray-Curtis distances also showed higher similarity between influent and SAG than between influent and AS (Fig. 1d). Permutational multivariate analysis of variance (Bray-Curtis distance, permutation = 999) was performed to test the two-way factor effect of sampling time and aggregate size. Our results showed significant effect of aggregated size ( $p = 0.001$ ), but not sampling time ( $p = 0.911$ ) or their interaction ( $p = 0.866$ ).

This phenomenon of community segregation was observed in different forms of biomass in biofilm and granular sludge

<sup>1</sup>Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada. ✉email: yang.liu@ualberta.ca



**Fig. 1** Sampling strategy, microbial community composition and diversity, and modeled influent immigration. **a** Schematic chart of sample collection. **b** Assembly of the top 10 abundant bacterial orders from influent and AS. **c** Venn diagram of shared OTUs among influent, activated sludge (AS) and small aggregates (SAG), and the relative abundances of shared OTUs. **d** PCoA of Bray-Curtis distances among the set of seasonal AS communities and high-flow disturbed and recovered communities (pyrosequencing), and the size-separated bioreactor samples (Illumina Miseq sequencing); variation between the long-term data and current data is likely due to different sequencing technologies. **e** Quantified influent immigration to AS and SAG using neutral immigration model ( $m_{NIM}$ ) and mass-balance model with operational parameters ( $m_{i,MBM-OP}$ ).

reactors<sup>13,16,17</sup>. Less work has been reported for conventional AS. SAG have higher accessibility to substrates and lower retention time and are more subjected to influent immigration effect as revealed in our results. An ecological growth trait of community, weighted  $rm$  gene copy number per genome, showed that SAG has higher average  $rm$  than AS, implying possibly higher growth activity (Supplementary Fig. S3)<sup>14,18,19</sup>.

The influent immigration impact was quantified using the NIM<sup>2,3</sup> and MBM-OP (Eq. 14) models. The migration probability ( $m_{NIM}$ ) was higher for influent-SAG (0.40) than for influent-AS (0.33) (Fig. 1e). The distribution of immigration rates for each OTU based on MBM-OP ( $m_{i,MBM-OP}$ ) showed that influent-SAG shifted towards higher immigration rate compared to influent-AS (Fig. 1e), in accordance with the observation of  $m_{NIM}$ . These observations suggest a neglected fact that the influent immigration to bioreactors is associated with the aggregate size-differentiated community. The SAG community is less settleable compared to flocs and more subjected to washing out from the reactor, leading to higher dependency on the influent immigration. Moreover, the influent community that enters the reactor and the SAG community share a more similar suspension status compared to flocs, which may attribute to the higher community similarity (Fig. 1d) and immigration (Fig. 1e).

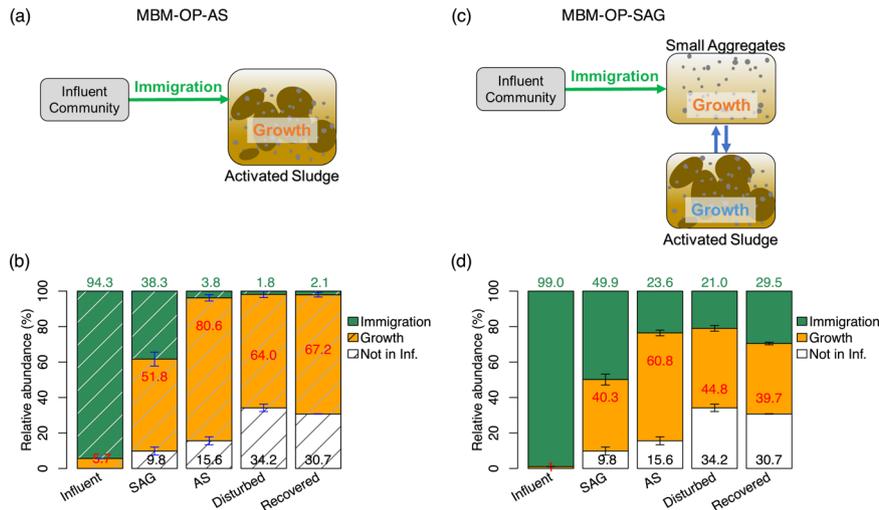
### Immigration contribution to community recovery

The AS community was disturbed by a high-flow event which led to a drop in biomass density, shifting away from the steady-state community and recovered after 3 days (Fig. 1d). The recovery could be attributed to different mechanisms, including growth of the indigenous microorganisms<sup>9,15</sup> and influent immigration<sup>1,2,9</sup>.

Using the MBM-OP immigration model, the influent community can be partitioned to growth and neutral immigration (non-growth) (Fig. 2a). A conceptual model MBM-OP-SAG was constructed (Fig. 2c). In both models, OTUs were classified into

three groups based on their relative abundances in the influent and in the AS: (i) not in influent but in reactor community, (ii) growth group (calculated net growth rate  $u_{i,MBM-OP} > 0$ , Eq. 16), (iii) complete immigration group (calculated net growth rate  $u_{i,MBM-OP} \leq 0$ , Eq. 16). The relative abundances (Fig. 2b, d) showed that the immigration group was using the MOM-OP-SAG model more (Fig. 2d) than using the MBM-OP-AS model (Fig. 2b). For the disturbed and recovered AS communities, the MBM-OP-AS model showed similar abundances of immigration group (Fig. 2b). The MBM-OP-SAG model estimated higher contribution of immigration in recovered than in disturbed community (Fig. 2d). The two models differed in certain taxa's growth rates and immigration rates, resulting in a list of genera classified into growth group by MBM-OP-AS model while into immigration group by MBM-OP-SAG model (Supplementary Table S1). Microorganisms such as *Methanobacterium*, *Methanobrevibacter*, *Clostridiales*, *Synergistales*, and *Bifidobacterium* are known anaerobic microorganisms and likely to be carried by wastewater from sewer<sup>20</sup> and not to grow in AS. The MBM-OP-SAG model was successful to identify these genera as influent immigration microorganisms.

Our results demonstrated AS community variation in different-sized aggregates. NIM and MBM-OP models were compared, both resulted in higher immigration impact for the small aggregate (SAG) community than for the AS community. The different-sized community variation led to an approach (MBM-OP-SAG model) of modeling influent and segregated communities for immigration effects, which was applied to investigate the community recovery. The MBM-OP-SAG model showed higher immigration contribution to recovery than the MBM-OP-AS model. This approach could be applied to other bioreactor systems to increase modeling accuracy of microbial growth and immigration, compared to the conventional NIM and MBM-OP models.



**Fig. 2 Contributions of immigration and growth for community recovery after disturbance.** **a** Influent to AS immigration model. **b** Proportions of the immigration and growth groups calculated by MBM-OP-AS model using the relative abundances of OTUs in influent and AS. **c** Proposed influent to small aggregates immigration model. **d** Proportions of the immigration and growth groups calculated by MBM-OP-SAG model using the relative abundances of OTUs in influent and small aggregates. Error bar represents standard deviation of  $n$  samples ( $n = 2$  for influent;  $n = 6$  for SAG and AS;  $n = 2$  for disturbed and recovered).

## METHODS

### AS and influent biomass sampling from full-scale wastewater treatment plant

Influent (treated by primary clarifier) and AS samples under various flow and seasonal conditions were taken from a full-scale biological nutrient removal wastewater treatment plant with a capacity of treating 310 MLD in Edmonton, Canada. For long-term seasonal comparison, AS samples were collected spanning 6 years (2013–2019), which covers seasonal variations and various flow conditions (240–310 MLD). A high-flow event resulted in a decrease of mixed liquor volatile suspended solids from 2840 to 2200 mg/L. The most recent two AS samples (Fall 2018 and Spring 2019) were used to analyze different-sized AS communities. AS samples were taken from the aeration tank to the on-site lab and immediately settled for 30 min using a stir tank at 5 rpm. The supernatant (containing SAG), the interface, and the settled AS were transferred to 50 mL sterile centrifuge tubes and stored on ice before centrifugation (within 1 h). The particle size in the supernatant was determined using Zetasizer (model ZEN 1600, Malvern Instruments Ltd.), resulting in a median size of  $25 \pm 4.5 \mu\text{m}$  and range of 1.9–29.0  $\mu\text{m}$ .

Influent samples and small aggregate samples were centrifuged in 50 mL tubes at 5000g for 10 min to collect the pellets. The AS samples were centrifuged in 2 mL tubes at 13,000g for 2 min. The samples were stored at  $-20^\circ\text{C}$  until DNA extraction.

### DNA extraction, 16S rRNA gene amplicon sequencing

DNA was extracted from influent and bioreactor samples using DNeasy PowerSoil Kit (Qiagen Inc., Toronto, Canada) according to the manufacturer's protocol. The long-term monitoring of the community was conducted using pyrosequencing of the 16S rRNA genes amplified by 28F (5'-GAGTTT-GATCNTGGCTCAG-3') and 519R (5'-GTNTTACNGCGGCKGCTG-3'). The different-sized AS and influent communities were analyzed using Illumina sequencing. The extracted DNA was used for PCR with modified primers with adaptors: 515 forward primer (5'-ACA CTG ACG ACA TGG TTC TAC AGT GYC AGC MGC CGC GGT AA-3') and 806 reverse primer (5'-TAC GGT AGC AGA GAC TTG GTC TGG ACT ACN VGG GTW TCT AAT-3')<sup>21,22</sup>, targeting the V4 region of 16S rRNA genes. The PCR was performed under the following conditions: primer concentration at 0.3  $\mu\text{M}$ ,  $\text{Mg}^{2+}$  at 1.5 mM, dNTP at 200  $\mu\text{M}$ , Taq DNA polymerase at 1.25 U/50  $\mu\text{L}$  reaction (New England Biolabs, Whitby, ON, Canada), initial denaturation at  $94^\circ\text{C}$  for 3 min, then 25 cycles of denaturation at  $94^\circ\text{C}$  for 30 s, annealing at  $62^\circ\text{C}$  for 45 s and elongation at  $68^\circ\text{C}$  for 1 min, then final elongation at  $68^\circ\text{C}$  for 10 min. The PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Toronto, ON, Canada), then sent for barcoding and sequencing at McGill University and G enome Qu ebec Innovation Centre (Montr eal, QC, Canada) using Illumina MiSeq PE250 platform. Sequence data were deposited to the National Center

for Biotechnology Information (NCBI) GenBank (Bio Project Accession number PRJNA580466).

The raw sequences ( $2 \times 250$  nt) were paired, quality-filtered using DADA2<sup>23</sup> in Qiime2 pipelines<sup>24</sup>. The OTUs were defined using the default 100% similarity. The taxonomic identification was assigned using GreenGenes (version gg\_13\_8) reference database at 99% similarity<sup>25,26</sup>. Microbial community diversities (alpha-diversity, beta-diversity) were analyzed using R "vegan" package<sup>27</sup>. The weighted *rrm* copy number was estimated by the normalized OTU table using PICRUSt<sup>28</sup> with metagenome references in Integrated Microbial Genomes (IMG) system<sup>29</sup>.

$$\text{Weighted } rrm \text{ copy number} = \sum f_{16S\_norm,i} \times N_{rrm,i} \quad (1)$$

where  $f_{16S\_norm,i}$  is the relative abundance of the  $i$ th OTU in total 16S rRNA gene amplicon sequence reads normalized by PICRUSt;  $N_{rrm,i}$  is the *rrm* copy number of the  $i$ th OTU in the IMG database.

### Neutral immigration model

The probability of influent immigration was defined by Sloan et al.<sup>2</sup> and calculated using the neutral immigration model. The model described that in a microbial community with  $N_T$  individuals, an individual die or leave the system and is immediately replaced by an immigrant with probability of  $m$ , or by reproduction of a local member of the community with probability of  $1-m$ .

For the  $i$ th species (initial number  $N_i$ ), the probability (Pr) of increase by one, stay the same, or decrease by one is expressed, respectively, by the following three equations:

$$\Pr\left(N_i + \frac{1}{N_i}\right) = \left(\frac{N_T - N_i}{N_T}\right) \left[mp_i + (1 + \alpha_i)(1 - m)\left(\frac{N_i}{N_T - 1}\right)\right] \quad (2)$$

$$\Pr(N_i/N_i) = \frac{N_i}{N_T} \left[mp_i + (1 - m)\left(\frac{N_i - 1}{N_T - 1}\right)\right] + \left(\frac{N_T - N_i}{N_T}\right) \left[m(1 - p_i) + (1 - m)\left(\frac{N_T - N_i - 1}{N_T - 1}\right)\right] \quad (3)$$

$$\Pr\left(N_i - \frac{1}{N_i}\right) = \frac{N_i}{N_T} \left[m(1 - p_i) + (1 - \alpha_i)(1 - m)\left(\frac{N_T - N_i}{N_T - 1}\right)\right] \quad (4)$$

where  $p_i$  is the relative abundance of the  $i$ th species,  $\alpha_i$  is the advantage parameter of the  $i$ th species over the other species.

The relative abundance can be expressed as  $x_i = \frac{N_i}{N_T}$ . Assume that  $N_T$  is large enough and  $x_i$  becomes continuous, and  $\alpha_i = 0$  so the species has no advantages, then the probability density function of  $x_i$ ,  $\phi_i = \phi_i(x_i, t)$  can be expressed as (for more details please see Sloan et al.<sup>2</sup>):

$$\phi_i(x_i; N_T, p_i, m) = c(1 - x_i)^{N_T m(1 - p_i) - 1} x_i^{N_T m p_i - 1} \quad (5)$$

where  $c$  is a constant so that  $\int_0^1 \phi_i(x_i) dx_i = 1$ .

$$c = \frac{\Gamma(N_T m)}{\Gamma[N_T m(1 - p_i)] \Gamma(N_T m p_i)} \quad (6)$$

The probability of the  $i$ th species observed in any local community is:

$$\Pr(\text{species } i \text{ is present with a relative abundance } > d) = \int_d^1 \phi_i(x_i; N_T, p_i, m) dx \quad (7)$$

where  $d$  is the detection threshold.

This equation can be used to calibrate community data (frequency and mean relative abundance plot) and solve for  $m$ .

AS and SAG communities included six replicates each. The immigration rate  $m$  was estimated for influent and AS communities and influent and SAG communities, using the R codes from Burns et al.<sup>3</sup> with 100 permutations (R code in Supplementary Information).

### Mass-balance model with operational parameters

The MBM with operational parameters (mass-flow immigration model) calculation for net growth rates was described by Guo<sup>11</sup>.

The mixed liquor heterotrophic biomass is contributed from the influent biomass and the microbial growth from the consumption of substrate resources. The immigration is defined as the fraction of influent biomass contribution:

$$m = \frac{\text{influent biomass}}{\text{influent biomass} + \text{local growth}} = \frac{f_{\text{OHO,Capt}} * X_{\text{OHO,Inf}}}{f_{\text{OHO,Capt}} * X_{\text{OHO,Inf}} + Y_{\text{OHO}} * S_{\text{consumed}}} \quad (8)$$

where  $m$  is the mass-flow immigration rate;  $X_{\text{OHO,Inf}}$  and  $X_{\text{OHO,ML}}$  are the active heterotrophic biomass in influent and mixed liquor;  $f_{\text{OHO,Capt}}$  is the fraction of influent biomass captured by the mixed liquor which is default 1;  $Y_{\text{OHO}}$  is the heterotrophic growth yield;  $S_{\text{consumed}}$  is the consumed substrates.

For continuous stirred tank reactors, the heterotrophic biomass based on mass-balance calculation<sup>30</sup> is:

$$X_{\text{OHO,ML}} = \frac{\theta_x \cdot (Y_{\text{OHO}} * S_{\text{consumed}})}{\theta \cdot (1 + b_{\text{OHO}} * \theta_x)} \quad (9)$$

Rearrangement gives:

$$Y_{\text{OHO}} * S_{\text{consumed}} = X_{\text{OHO,ML}} (\theta / \theta_x) (1 + b_{\text{OHO}} \theta_x) \quad (10)$$

Under the assumption that  $f_{\text{OHO,Capt}} * X_{\text{OHO,Inf}} \ll Y_{\text{OHO}} * S_{\text{consumed}}$ , substitution of Eq. 10 to Eq. 8 gives:

$$m = \frac{f_{\text{OHO,Capt}} * X_{\text{OHO,Inf}}}{X_{\text{OHO,ML}} (\theta / \theta_x) (1 + b_{\text{OHO}} * \theta_x)} \quad (11)$$

The immigration rate of a specific  $i$ th OTU is:

$$m_i = \frac{f_{\text{OHO,Capt}} * X_{\text{OHO,Inf},i}}{X_{\text{OHO,ML},i} \left(\frac{\theta}{\theta_x}\right) (1 + b_{\text{OHO},i} \theta_x)} \quad (12)$$

The proportion of the  $i$ th OTU in the biomass sample ( $f_{16S,i}$ ) can be quantified from the 16S rRNA amplicon sequencing data. The biomass concentration of the  $i$ th OTU can be defined by:

$$X_{\text{OHO},i} = X_{\text{Tot}} Y_{\text{DNA}} f_{16S,i} \quad (13)$$

where  $X_{\text{OHO},i}$  is the biomass concentration of the  $i$ th OTU;  $X_{\text{Tot}}$  is the total concentration of solids;  $Y_{\text{DNA}}$  is the mass of DNA per solids;  $f_{16S,i}$  is the proportion of the  $i$ th genus in total 16S rRNA gene amplicon sequence reads.

Therefore, the immigration rate of a specific  $i$ th OTU is expressed using 16S rRNA amplicon sequencing data:

$$m_i = \frac{\theta_x}{\theta} \cdot \frac{f_{\text{OHO,Capt}} \cdot X_{\text{Tot,Inf}} \cdot Y_{\text{DNA,Inf}} \cdot f_{16S,Inf,i}}{(1 + b_{\text{OHO},i} \theta_x) \cdot X_{\text{Tot,ML}} \cdot Y_{\text{DNA,ML}} \cdot f_{16S,ML,i}} \quad (14)$$

Note that the maximum value of  $m_i$  is 1, so the calculated  $m_i$  values when greater than 1 were set to 1. When  $m_i = 1$  (i.e., the growth rate  $[\mu_{\text{OHO},i}]$  is 0):

$$b_{\text{OHO},i} = \frac{f_{\text{OHO,Capt}} \cdot X_{\text{Tot,Inf}} \cdot Y_{\text{DNA,Inf}} \cdot f_{16S,Inf,i}}{X_{\text{Tot,ML}} \cdot Y_{\text{DNA,ML}} \cdot f_{16S,ML,i}} \cdot \frac{1}{\theta} - \frac{1}{\theta_x} \quad (15)$$

The net growth rate equals the growth rate minus decay rate, therefore the net growth rate of the  $i$ th OTU is:

$$\mu_{\text{OHO,Net},i} = 0 - b_{\text{OHO},i} = \frac{1}{\theta_x} - \frac{f_{\text{OHO,Capt}} \cdot X_{\text{Tot,Inf}} \cdot Y_{\text{DNA,Inf}} \cdot f_{16S,Inf,i}}{X_{\text{Tot,ML}} \cdot Y_{\text{DNA,ML}} \cdot f_{16S,ML,i}} \cdot \frac{1}{\theta} \quad (16)$$

The OTUs then can be classified into three groups:

- (i) not in influent but in reactor community
- (ii) growth group (calculated net growth rate  $\mu_{\text{OHO,Net},i} > 0$ )
- (iii) complete immigration group (calculated net growth rate  $\mu_{\text{OHO,Net},i} \leq 0$ ).

### DATA AVAILABILITY

The raw sequence data of 16S rRNA gene amplicons can be accessed at GenBank (Bio Project Accession number [PRJNA580466](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=PRJNA580466)).

Received: 23 September 2020; Accepted: 12 January 2021;

Published online: 22 March 2021

### REFERENCES

1. Ofiteru, I. D. et al. Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl Acad. Sci. USA* **107**, 15345–15350 (2010).
2. Sloan, W. T. et al. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* **8**, 732–740 (2006).
3. Burns, A. R. et al. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J.* **10**, 655–664 (2016).
4. Yuan, H., Mei, R., Liao, J. & Liu, W. T. Nexus of stochastic and deterministic processes on microbial community assembly in biological systems. *Front. Microbiol.* **10**, 1536 (2019).
5. Mei, R., Kim, J., Wilson, F. P., Bocher, B. T. W. & Liu, W. T. Coupling growth kinetics modeling with machine learning reveals microbial immigration impacts and identifies key environmental parameters in a biological wastewater treatment process. *Microbiome* **7**, 65 (2019).
6. Ning, D., Deng, Y., Tiedje, J. M. & Zhou, J. A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl Acad. Sci. USA* **116**, 16892–16898 (2019).
7. Saunders, A. M., Albertsen, M., Vollertsen, J. & Nielsen, P. H. The activated sludge ecosystem contains a core community of abundant organisms. *ISME J.* **10**, 11–20 (2016).
8. Mei, R. & Liu, W. T. Quantifying the contribution of microbial immigration in engineered water systems. *Microbiome* **7**, 144 (2019).
9. Vuono, D. C., Munakata-Marr, J., Spear, J. R. & Drewes, J. E. Disturbance opens recruitment sites for bacterial colonization in activated sludge. *Environ. Microbiol.* **18**, 87–99 (2016).
10. Frigon, D. & Wells, G. Microbial immigration in wastewater treatment systems: analytical considerations and process implications. *Curr. Opin. Biotechnol.* **57**, 151–159 (2019).
11. Guo, B. Cellular Metabolic Markers and Growth Dynamics Definition of Functional Groups in Activated Sludge Wastewater Treatment Heterotrophic Population. Doctor of Philosophy thesis, McGill Univ. (2019).
12. Sun, X., Sheng, Z. & Liu, Y. Effects of silver nanoparticles on microbial community structure in activated sludge. *Sci. Total Environ.* **443**, 828–835 (2013).
13. Ali, M. et al. Importance of species sorting and immigration on the bacterial assembly of different-sized aggregates in a full-scale aerobic granular sludge plant. *Environ. Sci. Technol.* **53**, 8291–8301 (2019).
14. Perez, M.V., Guerrero, L.D., Orellana, E., Figuerola, E.L. & Erijman, L. Time series genome-centric analysis unveils bacterial response to operational disturbance in activated sludge. *mSystems* **4**(4), e00169–19, <https://doi.org/10.1128/mSystems.00169-19> (2019).
15. Lee, S. H., Kang, H. J. & Park, H. D. Influence of influent wastewater communities on temporal variation of activated sludge communities. *Water Res.* **73**, 132–144 (2015).
16. Laurenzi, M. et al. Biomass segregation between biofilm and flocs improves the control of nitrite-oxidizing bacteria in mainstream partial nitrification and anammox processes. *Water Res.* **154**, 104–116 (2019).
17. Wells, G. F. et al. Microbial biogeography across a full-scale wastewater treatment plant transect: evidence for immigration between coupled processes. *Appl. Microbiol. Biotechnol.* **98**, 4723–4736 (2014).
18. Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**, 1328–1333 (2000).

19. Nemergut, D. R. et al. Decreases in average bacterial community rRNA operon copy number during succession. *ISME J.* **10**, 1147–1156 (2016).
20. Guo, B., Liu, C., Gibson, C. & Frigon, D. Wastewater microbial community structure and functional traits change over short timescales. *Sci. Total Environ.* **662**, 779–785 (2019).
21. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
22. Apprill, A., McNally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **75**, 129–137 (2015).
23. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
24. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
25. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
26. Werner, J. J. et al. Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J.* **6**, 94–103 (2012).
27. Oksanen, F.J. et al. *vegan: Community Ecology Package*. R package Version 2.4-3. <https://CRAN.R-project.org/package=vegan> (2017).
28. Langille, M. G. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).
29. Markowitz, V. M. et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
30. Grady, C. P. L. & Grady, C. P. L. *Biological Wastewater Treatment*. 3rd edn. (Taylor & Francis, 2011).

## ACKNOWLEDGEMENTS

This work was funded by the Canada Research Chair (CRC) in Future Water Services (Y.L.) and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (Y.L.).

## AUTHOR CONTRIBUTIONS

B.G. collected and analyzed the samples, carried out data analysis, modeling and wrote the manuscript. Z.S. analyzed the samples and carried out analysis. Y.L.

conceived and guided the study and revised the manuscript. All authors read and approved the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41545-021-00112-7>.

**Correspondence** and requests for materials should be addressed to Y.L.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021