

## Supplementary Information

### Microbiome assembly predictably shapes diversity across a range of disturbance frequencies in experimental microcosms

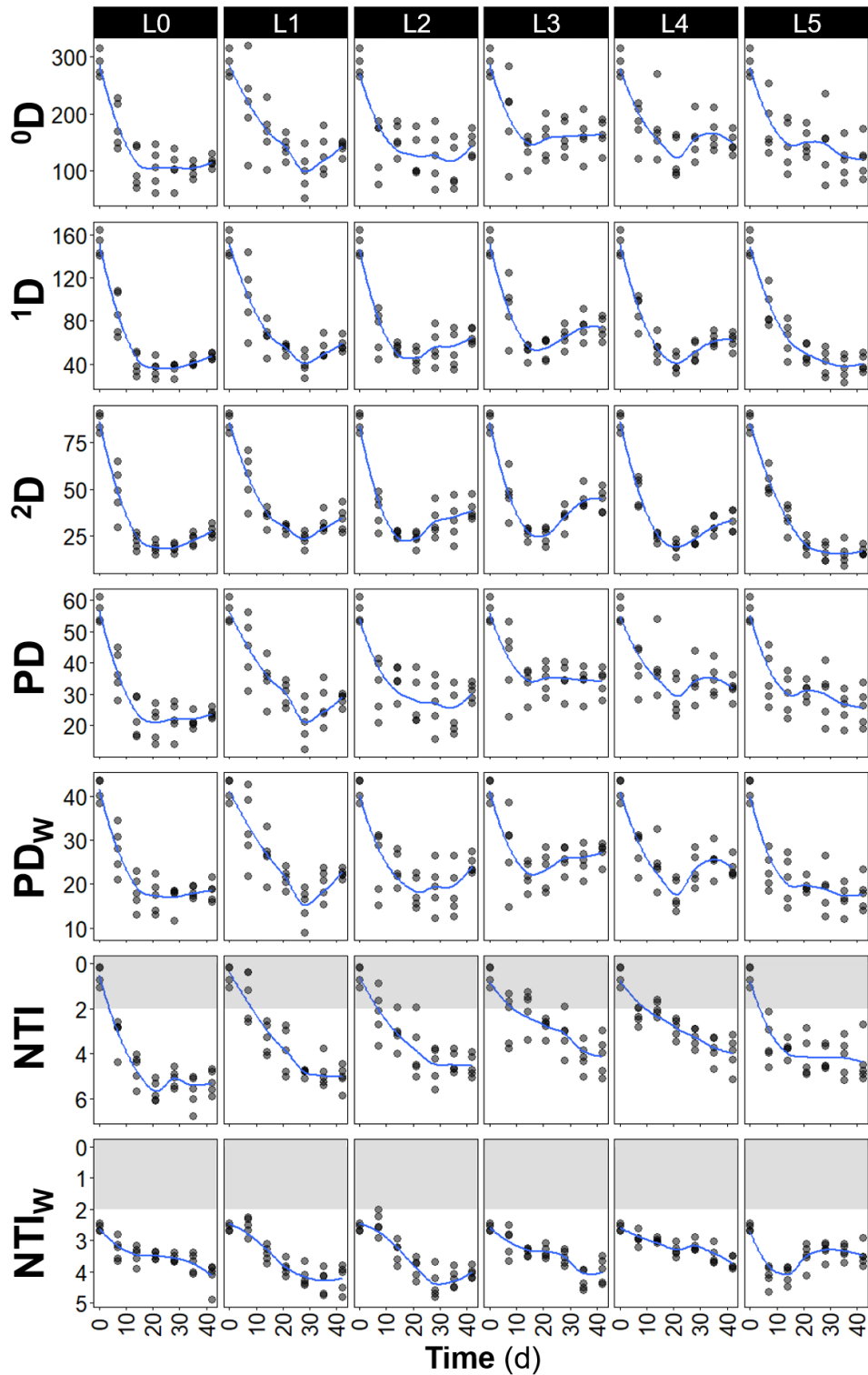
**Authors:** Ezequiel Santillan<sup>1</sup> and Stefan Wuertz<sup>1,2\*</sup>

**Affiliations:**

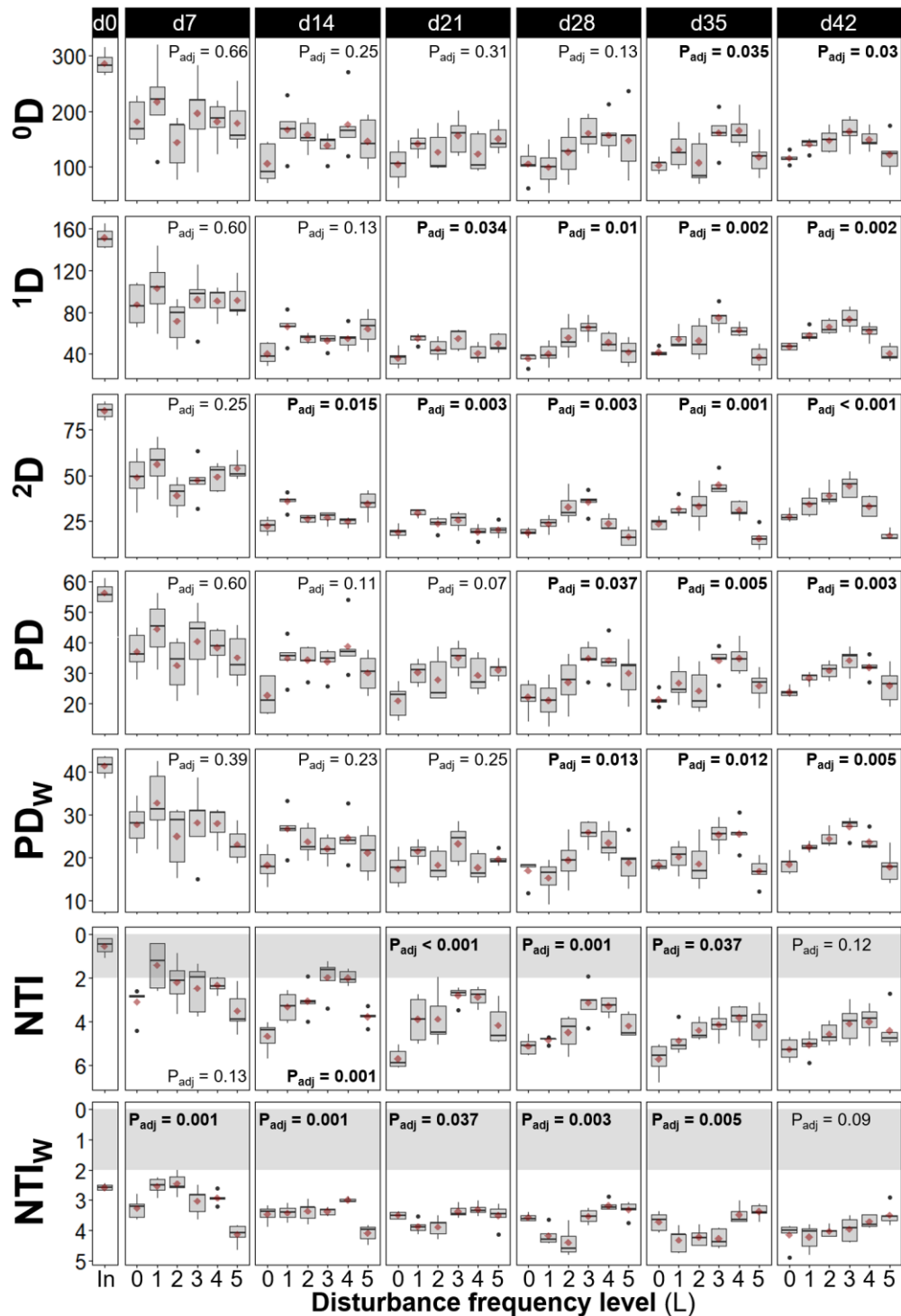
<sup>1</sup>Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, 637551, Singapore.

<sup>2</sup>School of Civil and Environmental Engineering, Nanyang Technological University, 639798, Singapore.

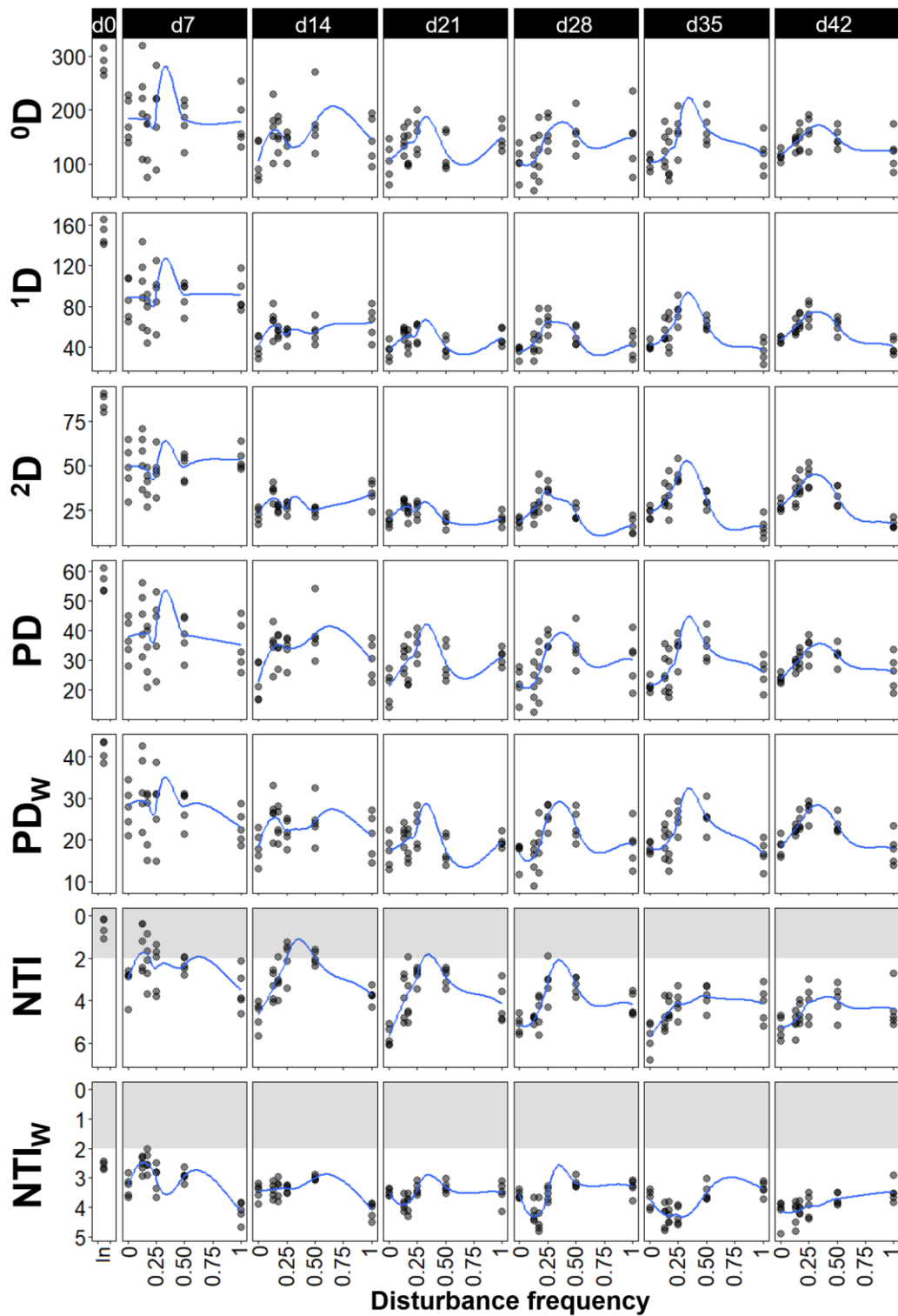
\*Correspondence to: Stefan Wuertz, [swuertz@ntu.edu.sg](mailto:swuertz@ntu.edu.sg)



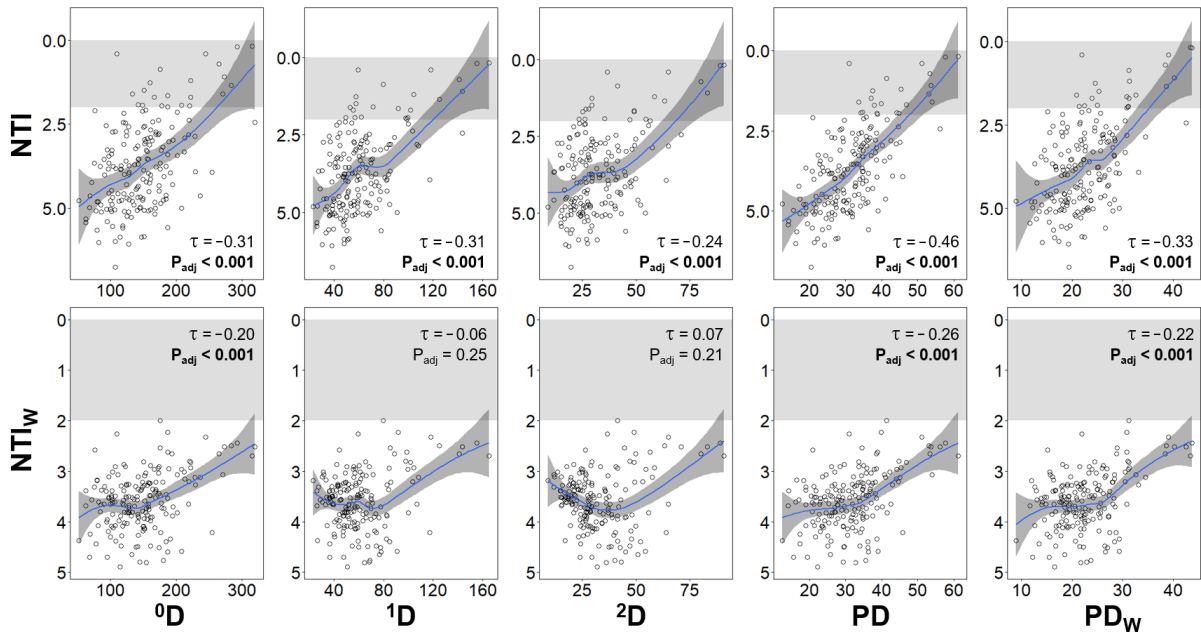
**Supplementary Figure 1** – Temporal dynamics of richness ( ${}^0D$ ), Hill  $\alpha$ -diversity of 1<sup>st</sup> ( ${}^1D$ ) and 2<sup>nd</sup> ( ${}^2D$ ) order, phylogenetic diversity unweighted (PD) and abundance-weighted ( $PD_w$ ), nearest taxon index unweighted (NTI) and abundance-weighted ( $NTI_w$ ), from bacterial ASV data for each frequency of organic loading disturbance ( $n = 5$ , except for day 0 where  $n = 4$ ). Disturbance frequency levels (L): 0 (undisturbed), 1-4 (intermediately disturbed), 5 (press disturbed). Blue line represents locally estimated scatterplot smoothing regression (loess). Note the inverted y-axis for both NTI and  $NTI_w$ , as values closer to zero indicate a higher relative contribution of stochastic assembly. Shaded in grey is the zone of significant stochastic phylogenetic dispersion,  $|NTI| < 2$  and  $|NTI_w| < 2$ .



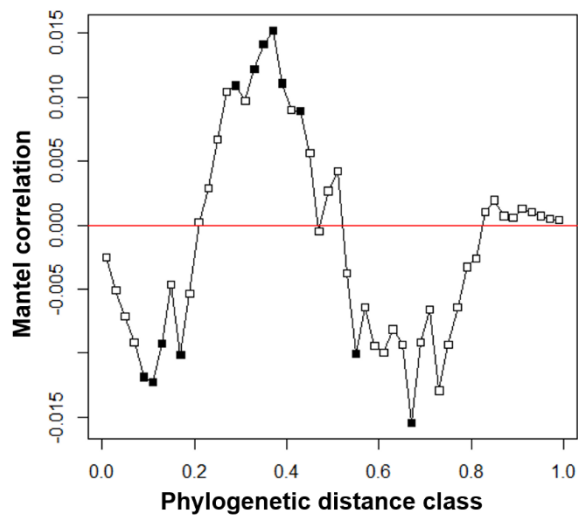
**Supplementary Figure 2** – Community structure and assembly assessed via richness ( ${}^0D$ ), Hill  $\alpha$ -diversity of 1<sup>st</sup> ( ${}^1D$ ) and 2<sup>nd</sup> ( ${}^2D$ ) order, phylogenetic diversity unweighted (PD) and abundance-weighted (PD<sub>w</sub>), nearest taxon index unweighted (NTI) and abundance-weighted (NTI<sub>w</sub>), from bacterial ASV data for different frequencies of organic loading disturbance ( $n = 5$ ). Disturbance frequency levels (L): 0 (undisturbed), 1-4 (intermediately disturbed), 5 (press disturbed). In: sludge inoculum (day 0,  $n = 4$ ). Each panel represents a sampling day, red diamonds display mean values. Welch's ANOVA P-values adjusted at 5% FDR shown within panels. Note the inverted y-axis for both NTI and NTI<sub>w</sub>, as values closer to zero indicate a higher relative contribution of stochastic assembly. Shaded in grey is the zone of significant stochastic phylogenetic dispersion,  $|NTI| < 2$  and  $|NTI_w| < 2$ . The box bounds the interquartile range (IQR) divided by the median, and Tukey-style whiskers extend to a maximum of 1.5 times the IQR beyond the box.



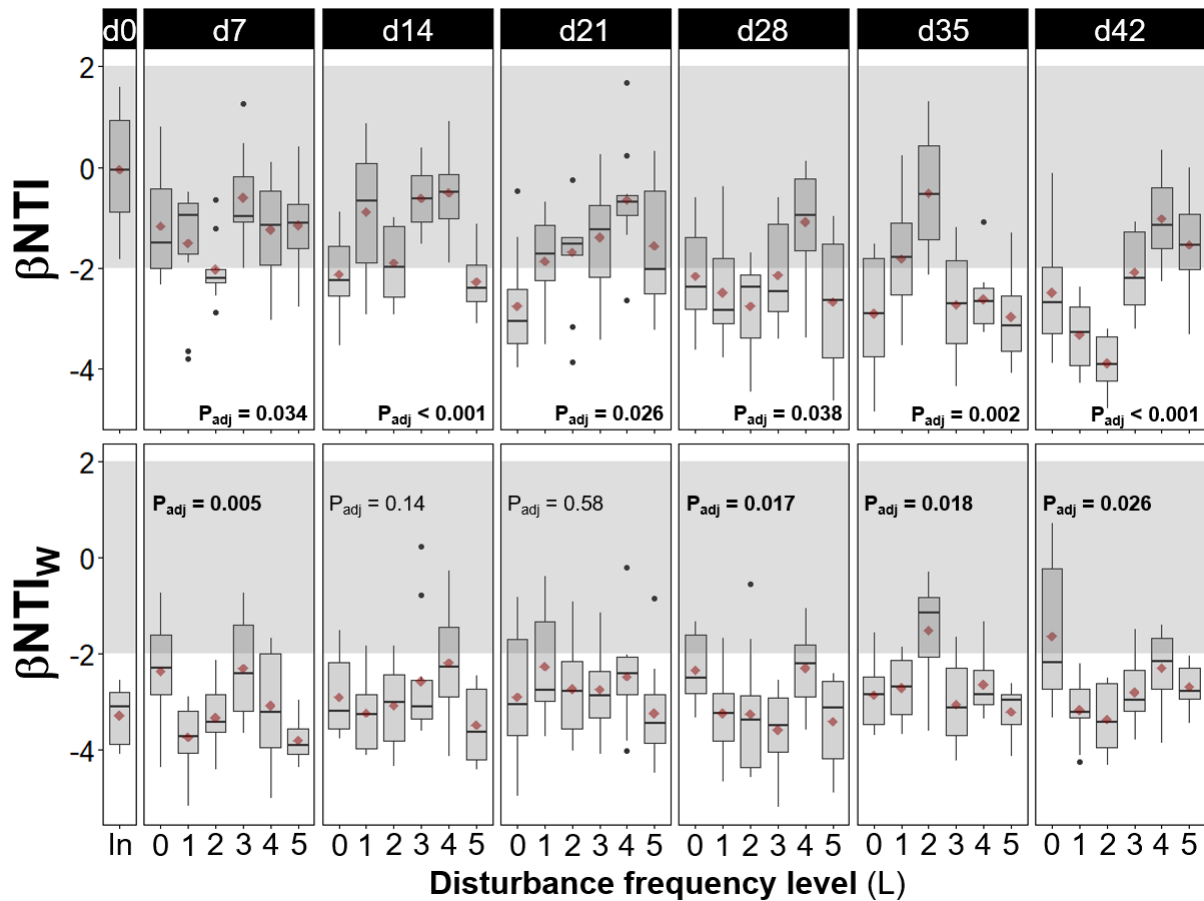
**Supplementary Figure 3** – Community structure and assembly assessed via richness ( ${}^0D$ ), Hill  $\alpha$ -diversity of 1<sup>st</sup> ( ${}^1D$ ) and 2<sup>nd</sup> ( ${}^2D$ ) order, phylogenetic diversity non-weighted (PD) and abundance-weighted ( $PD_w$ ), nearest taxon index unweighted (NTI) and abundance-weighted ( $NTI_w$ ), from bacterial ASV data for different frequencies of organic loading disturbance ( $n = 5$ ). Disturbance frequency values were calculated from the frequency of the high organic loading at each disturbance level. In: sludge inoculum (day 0,  $n = 4$ ). Each panel represents a sampling day. Blue line represents locally estimated scatterplot smoothing regression (loess). Note the inverted y-axis for both NTI and  $NTI_w$ , as values closer to zero indicate a higher relative contribution of stochastic assembly. Shaded in grey is the zone of significant stochastic phylogenetic dispersion,  $|NTI| < 2$  and  $|NTI_w| < 2$ .



**Supplementary Figure 4** – Richness ( ${}^0D$ ), Hill  $\alpha$ -diversity of 1<sup>st</sup> ( ${}^1D$ ) and 2<sup>nd</sup> ( ${}^2D$ ) order, unweighted (PD) and abundance-weighted phylogenetic diversity ( $PD_w$ ), correlated against unweighted (NTI, upper panels) and abundance-weighted nearest taxon index ( $NTI_w$ , lower panels), from bacterial ASV data for all frequency levels and time points evaluated in this study ( $m = 184$ ). Kendall correlation  $\tau$ - and P-values adjusted at 5% FDR are indicated within the panels. Blue line represents locally estimated scatterplot smoothing regression (loess) with confidence interval in dark-grey shading. Note the inverted y-axis for both NTI and  $NTI_w$ , as values closer to zero indicate a higher relative contribution of stochastic assembly. Shaded in grey is the zone of significant stochastic phylogenetic dispersion,  $|NTI| < 2$  and  $|NTI_w| < 2$ .

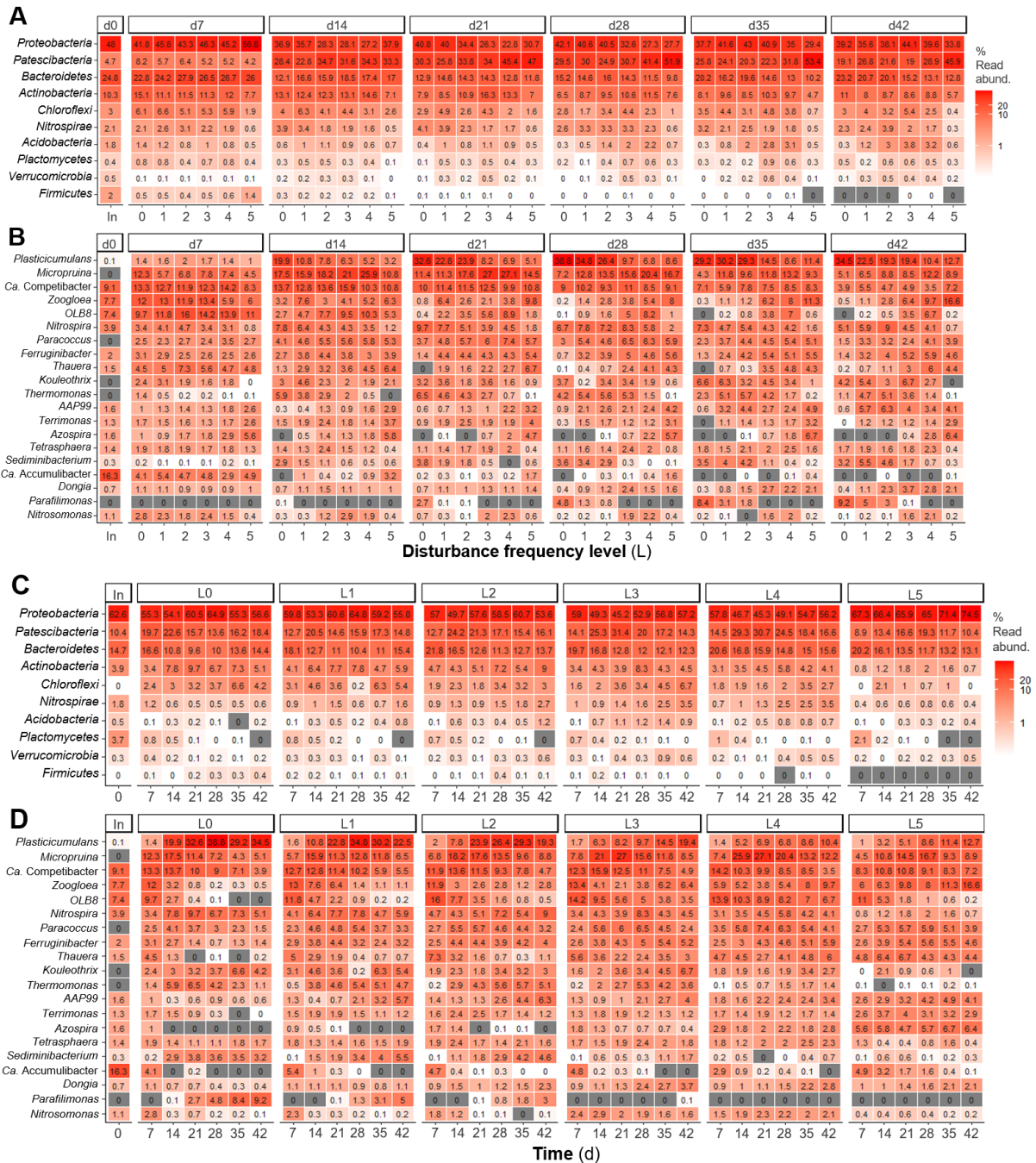


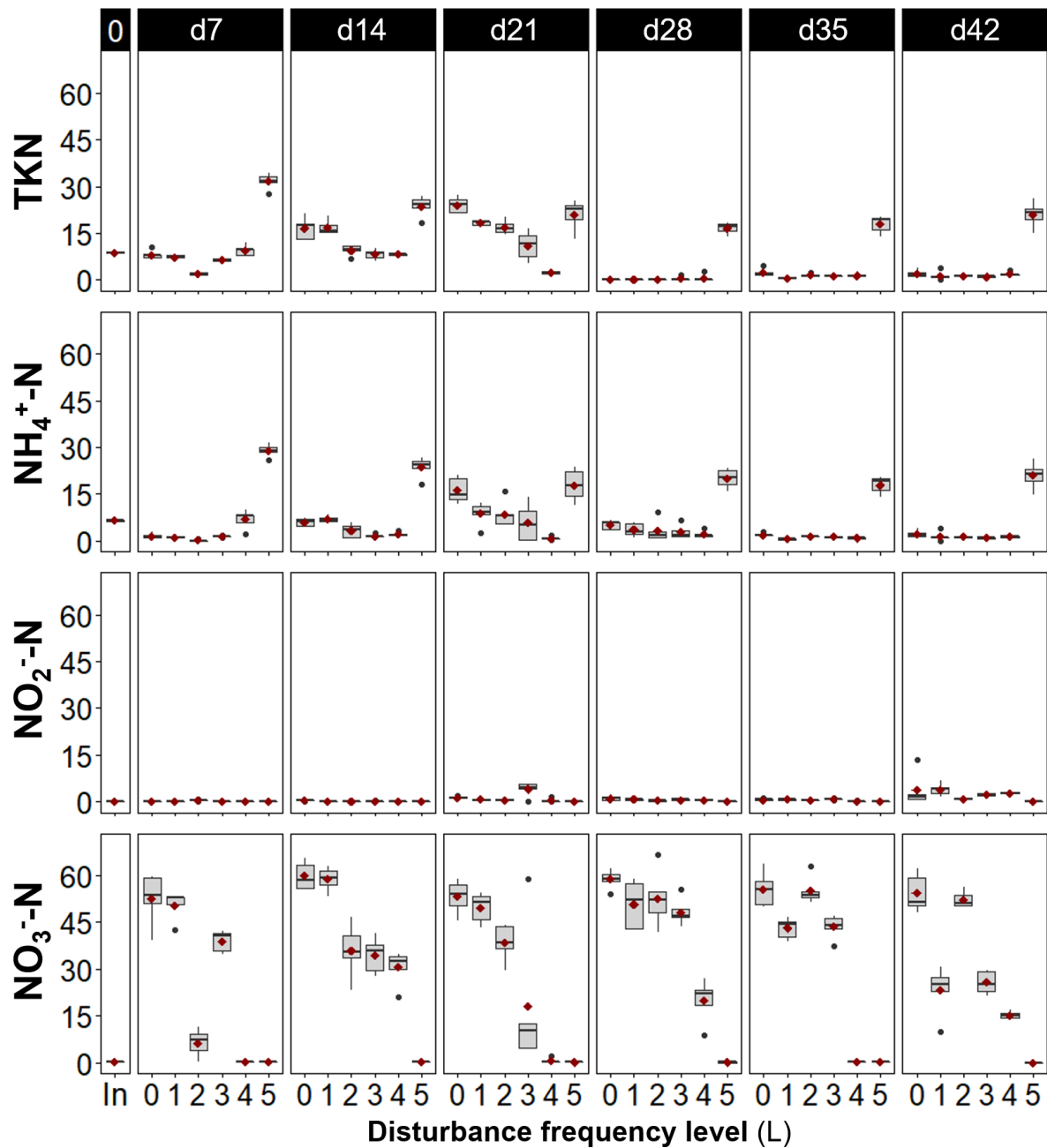
**Supplementary Figure 5** - Phylogenetic Mantel correlogram evaluating phylogenetic signal in the sludge bioreactor communities sampled in this study. The plot relates between-ASV niche differences to between-ASV phylogenetic distances across a given phylogenetic distance. Significant correlations ( $P_{adj} < 0.05$ ; closed symbols) indicate significant phylogenetic signal in ASV ecological niches within the associated phylogenetic distance class. The analysis shows a significant phylogenetic signal but mostly across relatively short phylogenetic distances.



**Supplementary Figure 6** – Temporal dynamics of community assembly via  $\beta$ -diversity null modelling of phylogenetic turnover across samples. Within-treatment pairwise values of the  $\beta$ -nearest taxon index, unweighted ( $\beta\text{NTI}$ , upper panels) and abundance-weighted ( $\beta\text{NTI}_w$ , lower panels), from bacterial ASV data for different frequencies of organic loading disturbance ( $n = 10$ ). Disturbance frequency levels (L): 0 (undisturbed), 1-4 (intermediately disturbed), 5 (press disturbed). In: sludge inoculum (day 0,  $n = 6$ ). Each panel represents a sampling day, red diamonds display mean values. Welch's ANOVA P-values adjusted at 5% FDR shown within panels. Shaded in grey is the zone where stochastic processes significantly dominate,  $|\beta\text{NTI}| < 2$ .  $\beta\text{NTI}$  values closer to zero indicate a higher relative contribution of stochastic assembly. The box bounds the interquartile range (IQR) divided by the median, and Tukey-style whiskers extend to a maximum of 1.5 times the IQR beyond the box.

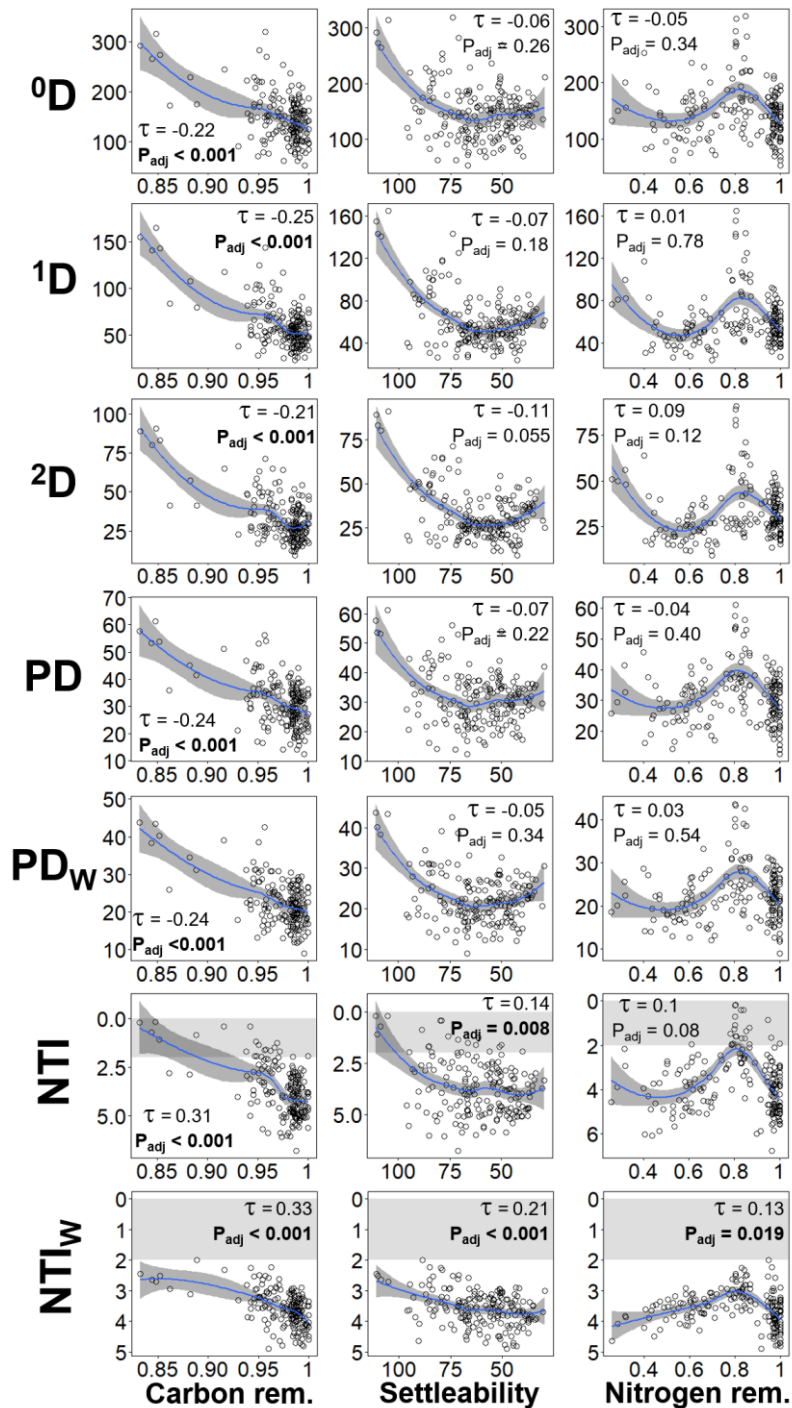




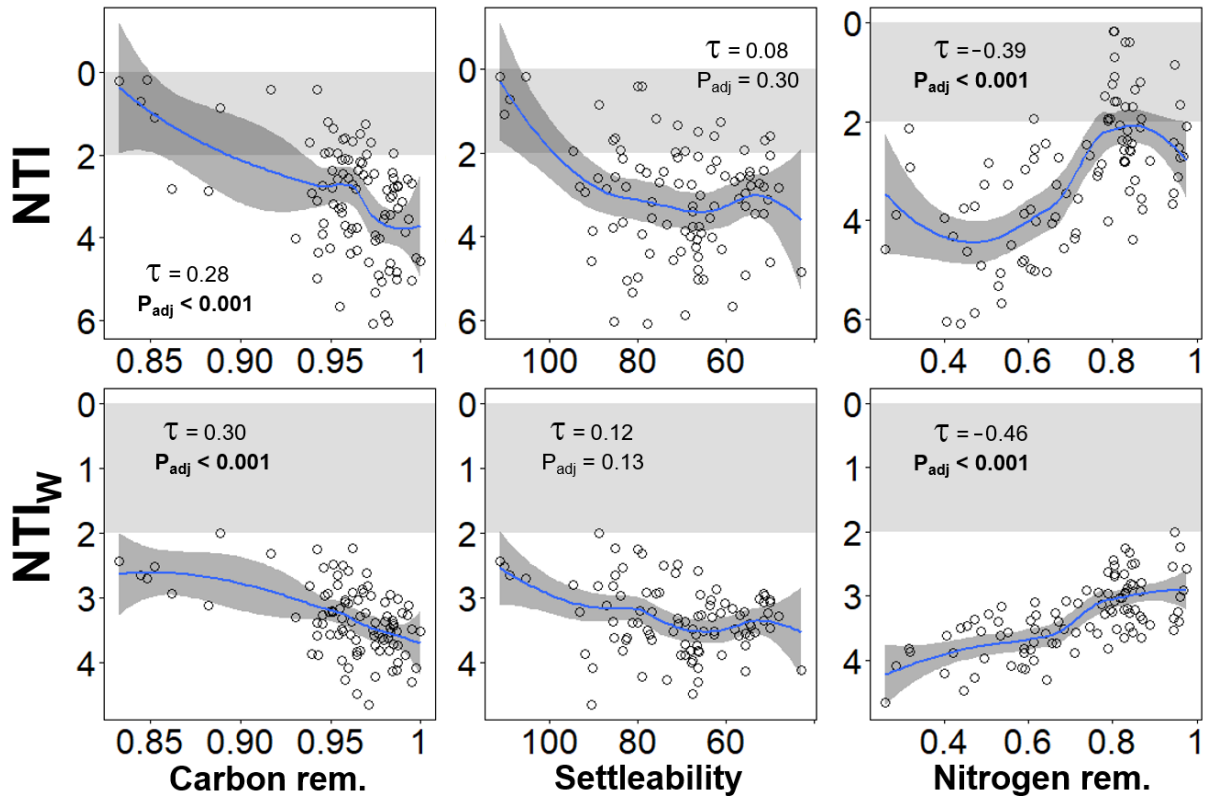


**Supplementary Figure 8** – Effluent values of total Kjeldahl nitrogen (TKN), ammonia ( $\text{NH}_4^+\text{-N}$ ), nitrite ( $\text{NO}_2^-\text{-N}$ ) and nitrate ( $\text{NO}_3^-\text{-N}$ ) as nitrogen for different frequencies of organic loading disturbance ( $n = 5$ ). Disturbance frequency levels (L): 0 (undisturbed), 1-4 (intermediately disturbed), 5 (press disturbed). In: sludge inoculum (day 0,  $n = 4$ ). Each panel represents a sampling day, and red diamonds display mean values. The box bounds the interquartile range (IQR) divided by the median, and Tukey-style whiskers extend to a maximum of 1.5 times the IQR beyond the box.





**Supplementary Figure 9** – Community function assessed via influent chemical oxygen demand removal (carbon removal, left panels), sludge volume index (sludge settleability, middle panels), and influent total Kjeldahl nitrogen removal (nitrogen removal, right panels), correlated against richness (<sup>0</sup>D), Hill  $\alpha$ -diversity of 1<sup>st</sup> (<sup>1</sup>D) and 2<sup>nd</sup> (<sup>2</sup>D) order, unweighted (PD) and abundance-weighted phylogenetic diversity (PD<sub>w</sub>), unweighted (NTI, upper panels) and abundance-weighted nearest taxon index (NTI<sub>w</sub>, lower panels), from bacterial ASV data for all frequency levels and time points evaluated in this study ( $m = 184$ ). Kendall correlation  $\tau$ - and P-values adjusted at 5% FDR are indicated within the panels. Blue line represents locally estimated scatterplot smoothing regression (loess) with confidence interval in dark-grey shading. Shaded in grey is the zone of significant stochastic phylogenetic dispersion,  $|NTI| < 2$  and  $|NTI_w| < 2$ . Note the inverted axis for sludge settleability, as it improves with decreasing SVI values, and for both NTI and NTI<sub>w</sub>, since values closer to zero indicate a higher relative contribution of stochastic assembly.



**Supplementary Figure 10** – Community function assessed via influent chemical oxygen demand removal (carbon removal, left panels), sludge volume index (sludge settleability, middle panels), and influent total Kjeldahl nitrogen removal (nitrogen removal, right panels), correlated against unweighted (NTI, upper panels) and abundance-weighted nearest taxon index (NTI<sub>w</sub>, lower panels), from bacterial ASV data at initial stages of succession (d0 to d21, m = 94). Kendall correlation  $\tau$ - and P-values adjusted at 5% FDR are indicated within the panels. Blue line represents locally estimated scatterplot smoothing regression (loess) with confidence interval in dark-grey shading. Shaded in grey is the zone of significant stochastic phylogenetic dispersion,  $|NTI| < 2$  and  $|NTI_w| < 2$ . Note the inverted axis for sludge settleability, as it improves with decreasing SVI values, and for both NTI and NTI<sub>w</sub>, since values closer to zero indicate a higher relative contribution of stochastic assembly.



**Supplementary Table 1.** Multivariate tests on bacterial community structure<sup>Ψ</sup> across disturbance frequency levels.

Time (d)	No of levels*	$n^{\ddagger}$	df <sup>§</sup> res	PERMANOVA <sup>†</sup>		PERMDISP <sup>†</sup>	
				F	P <sub>adj</sub> <sup>¶</sup>	F	P <sub>adj</sub> <sup>¶</sup>
7	6	5	24	2.64	<0.001	1.44	0.50
14	6	5	24	5.19	<0.001	1.88	0.45
21	6	5	24	8.13	<0.001	2.66	0.13
28	6	5	24	9.75	<0.001	2.70	0.13
35	6	5	24	9.23	<0.001	0.67	0.92
42	6	5	24	11.9	<0.001	0.52	0.90

<sup>Ψ</sup>The Bray-Curtis dissimilarity metric was used on square-root transformed ASV relative abundance data.

\* Factor levels (L): 0 (undisturbed), 1-4 (intermediately disturbed), 5 (press disturbed).

<sup>†</sup> Number of permutations used was 9,999

<sup>‡</sup> Number of replicates per level

<sup>§</sup> Degrees of freedom of the residual

<sup>¶</sup> P-values after correction for multiple comparisons at 5% FDR, via the Benjamini-Hochberg's method.