

## Supplementary Material

### Title

Wastewater treatment plant effluent introduces recoverable shifts in microbial community composition in receiving streams

### Author names and affiliations

Jacob R. Price<sup>a</sup>, Sarah H. Ledford<sup>b</sup>, Michael O. Ryan<sup>a</sup>, Laura Toran<sup>b</sup>, and Christopher M. Sales<sup>a\*</sup>

<sup>a</sup> Civil, Architectural, and Environmental Engineering, Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104, United States of America

<sup>b</sup> Earth and Environmental Science, Temple University, 1901 N. 13<sup>th</sup> St, Philadelphia, Pennsylvania, 19122, United States of America

\* Corresponding Author. E-mail: [chris.sales@drexel.edu](mailto:chris.sales@drexel.edu). Phone: (215) 895-2155. Fax: (215) 895-1363.

### Corresponding author

Christopher M. Sales  
Civil, Architectural, and Environmental Engineering  
Drexel University  
[chris.sales@drexel.edu](mailto:chris.sales@drexel.edu)  
Phone: (215) 895-2155  
Fax: (215) 525-4332

## **1. Supplemental results - indicator community**

### **1.1. Influence of effluent on the diversity and phylogenetic abundance of the indicator community**

Taxa within the orders of Bacteroidales, Bifidobacteriales, and Clostridiales were selected for additional study as indicators of human fecal contamination. As with the full microbial community (Figure 4), sites immediately downstream of WWTPs (Site 2 and 4) appear to have higher alpha diversity within the indicator subset in comparison to their upstream counterparts (Figure S8, Table S6). After filtering and preprocessing, as described in the main text, 171 taxa were annotated to these three orders and they account for approximately 0.5-2.0 % of the reads within the dataset (Figure S9); Bifidobacteriales was a minor contributor to this abundance. While Bacteroidales tended to be the most abundant order within the indicator community, Clostridiales was the dominant order in samples collected from Site 2 on the second sampling day and Site 4 on the first sampling day (samples 2.DSUG.2, 4.DSAmb.1); both of these sites are located immediately below an effluent source.

### **1.2. Exploratory ordination of the indicator community**

Variance stabilization transformation was applied to the indicator subset prior to exploratory ordination. The PCA plot for the indicator community displays the characteristics of a horseshoe effect (Figure S10), which implies that there is one dominant gradient in the dataset. The indicator subset is almost entirely composed of just two orders of Bacteria, and so this dominant gradient is along their abundances, or ratio of abundances. Horseshoe effect, as with arch effect, can make interpretation of ordination axis difficult. In spite of this, the analysis is still worthwhile, as, looking at the order along the horseshoe gradient, it can be seen that upstream samples are closer to the start of the gradient and their downstream pairs are further along the gradient. This difference translates into an increase in the ratio of Clostridiales to Bacteroidales (as reflected in Figure S9), and further reinforced by the application of DPCoA to the indicator subset (Figure S11).

### **1.3. Differential abundance within the microbial community up and downstream of effluent sources**

Differential abundance detection was carried out on the indicator community. 26 of the 171 taxa within the indicator subset were determined to be differentially abundant (Figure S12); 13 of the taxa fell within Clostridia, 12 within Bacteroidetes, and 1 within Bifidobacteriales. 11 of the 26 taxa matched those identified during the analysis for the whole community. The authors believe that removing potential noise induced by the non-indicator taxa, may have enabled the identification of the additional 15 taxa.

The indicator subset displayed patterns in increases in relative abundance of these taxa similar to those found for the whole microbial community, but were more variable at both sites and sampling dates (Figure S13, Table S7). The increase in relative abundance at the Upper Gwynedd WWTP was 9.7 times for the first sampling day and 140.3 for the second day (average of 75). At the Ambler WWTP, the increase was 18 times for the first sampling day and 5.2 for the second day (average of 11.6).

### **1.4. Environmental influences on the indicator community**

#### **1.4.1. Distance-based redundancy analysis on the indicator community**

The final db-RDA model for the indicator subset included the effects of (log-scaled)  $\text{Cl}^-$ , Si, and Mg, with an adjusted  $r^2$  of 0.3472 (unadjusted  $r^2 = 0.5252$ ), and was statistically significant ( $p = 0.001$ ) (Figure S14). Significance testing by term indicated that  $\text{Cl}^-$  ( $p = 0.001$ ) and Si ( $p = 0.014$ ) were both significant, while Mg was not significant ( $p = 0.104$ ) once the effects of  $\text{Cl}^-$  and Si were partialled out. This may be the result of Mg and Si being highly correlated, as observed in Figure S14.

#### **1.4.2. BIOENV procedure on the indicator community**

For the indicator subset, BIOENV selected a model including (log-transformed)  $\text{Cl}^-$ ,  $\text{NO}_2^-$ -N, Mg, and Si which had a BIOENV correlation of 0.6804 (Table S9). As with the full community,  $\text{Cl}^-$  was the single best parameter for explaining variation in OTU abundance and was found in each of the models generated (Table S9).

## 2. Supplemental Tables

Table S1: Discharge and concentration data from 2016 Delaware River Basin Commission Annual Effluent Monitoring Reports; nr=not reported.

<b>WWTP</b>	<b>Permitted Discharge (MGD)</b>	<b>Annual Average Discharge (MGD)</b>	<b>Average TP (mg/L)</b>	<b>Average NO<sub>3</sub>+NO<sub>2</sub> (mg N/L)</b>	<b>Average Fecal Coliform (cfs/100 mL)</b>	<b>Approximate length of contributing sewers (km)</b>
Ambler	6.5	3.64	nr	nr	8.4	263
Upper Gwynedd	5.7	2.31	nr	nr	11.8	136
Upper Dublin	1.1	0.76	nr	nr	4.7	73
Abington	3.91	2.60	1.61	12.46	15.2	227

Reference for this table:

Delaware River Basin Commission Annual Effluent Monitoring Reports. 2016. Available by request to the DRBC: [www.state.nj.us/drbc/contact/general/](http://www.state.nj.us/drbc/contact/general/)

Table S2: Land use statistics for the watersheds contributing to the flow at each of the sampling sites

	<b>Contributing Area (km<sup>2</sup>)</b>	<b>Developed, Open Space<sup>1</sup></b>	<b>Developed, Low Intensity<sup>1</sup></b>	<b>Developed, Medium Intensity<sup>1</sup></b>	<b>Developed, High Intensity<sup>1</sup></b>	<b>Deciduous Forest<sup>1</sup></b>	<b>Shrub/scrub<sup>1</sup></b>	<b>Pasture/hay<sup>1</sup></b>	<b>Other<sup>1,2</sup></b>
<b>1.USUG</b>	18.98	6.14 (32.4%)	3.84 (20.2%)	2.08 (10.9%)	1.00 (5.3%)	4.01 (21.1%)	0.42 (2.2%)	0.88 (4.6%)	0.61 (3.2%)
<b>2.DSUG</b>	19.38	6.23 (32.1%)	3.89 (20.1%)	2.08 (10.7%)	1.00 (5.2%)	4.08 (21.0%)	0.45 (2.3%)	0.93 (4.8%)	0.72 (3.7%)
<b>3.USAmb</b>	63.59	21.72 (34.2%)	10.48 (16.5%)	4.48 (7.1%)	1.52 (2.4%)	15.06 (23.7%)	2.40 (3.8%)	4.52 (7.1%)	3.40 (5.4%)
<b>4.DSAmb</b>	67.94	23.21 (34.2%)	11.55 (17.0%)	5.02 (7.4%)	1.61 (2.4%)	15.77 (23.2%)	2.44 (3.6%)	4.68 (6.9%)	3.66 (5.4%)
<b>5.BC</b>	105.11	36.36 (34.6%)	21.39 (20.4%)	8.11 (7.7%)	2.26 (2.1%)	22.33 (21.2%)	3.21 (3.1%)	6.35 (6.0%)	5.09 (4.8%)
<b>6.SR</b>	31.29	11.61 (37.1%)	9.09 (29.1%)	2.63 (8.4%)	0.58 (1.9%)	5.19 (16.6%)	0.53 (1.7%)	0.74 (2.4%)	0.91 (2.9%)

<sup>1</sup>All classifications are from the 2011 National Land Cover Database (Homer et al., 2015). The first number is the km<sup>2</sup> of that land cover in the contributing area, followed in parentheses by the percentage cover.

<sup>2</sup>Other includes: open water, barren, evergreen forest, mixed forest, shrub/scrub, grassland/herbaceous, cultivated crops, woody wetlands and emergent herbaceous wetlands

Table S3: Real-time PCR Primers and Targets

Assay Target	Target Taxa		Primer Sequence (5' to 3')
Human Specific <i>Bacteroides</i>	<i>Bacteroides dorei</i>	Forward	GGCAGCATGGTCTTAGCTTG
		Reverse	CGGACATGTGAACTCATGAT
		Probe*	GCTGATGG
Universal <i>Bacteroides</i> spp.	<i>Bacteroides</i> spp.	Forward	GGGGTTCTGAGAGGAAGGT
		Reverse	CCGTCATCCTTCACGCTACT
		Probe*	CAATATTCCTCACTGCTGCCTCCCGTA

\* Probe was not used for this study.

Table S4: Accession Identifiers for the raw reads uploaded to the National Center for Biotechnology Information (NCBI) Sequencing Read Archive (SRA).

Sequencing Read Archive Study Accession ID	SRP103534
BioProject Accession ID	PRJNA382371
Sample Name	BioSample Accession ID
1.USUG.1	SAMN06705191
1.USUG.2	SAMN06705192
2.DSUG.1	SAMN06705193
2.DSUG.2	SAMN06705194
3.USAmb.1	SAMN06705195
3.USAmb.2	SAMN06705196
4.DSAmb.1	SAMN06705197
4.DSAmb.2	SAMN06705198
5.BC.1	SAMN06705199
5.BC.2	SAMN06705200
6.SR.1	SAMN06705201
6.SR.2	SAMN06705202

Table S5: Counts of raw reads and reads passing filter

Sample Name	Raw Reads [count]	Reads Passing Filter [count]
1.USUG.1	106,065	51,118
1.USUG.2	91,569	67,947
2.DSUG.1	104,740	61,675
2.DSUG.2	121,143	82,949
3.USAmb.1	72,049	31,568
3.USAmb.2	105,844	84,543
4.DSAmb.1	74,737	41,753
4.DSAmb.2	88,673	65,797
5.BC.1	109,071	63,472
5.BC.2	102,176	74,947
6.SR.1	110,097	71,458
6.SR.2	119,847	83,476



Table S6: Alpha diversity metrics for the full microbial community and the indicator community.

Sample Name	Microbial Community				Indicator Community				
	Observed	Chao1	(Chao1) Std. Error	Shannon	Observed	Chao1	(Chao1) Std. Error	Shannon	
1.USUG.1	3,327	4,014	58.6	5.2	111	171	21.7	4.0	
1.USUG.2	3,339	3,970	55.6	5.2	84	114	13.4	3.7	
2.DSUG.1	4,112	4,673	50.1	6.4	142	172	11.6	4.5	
2.DSUG.2	3,923	4,566	56.1	6.0	180	211	13.0	4.4	
3.USAmb.1	2,755	3,892	89.7	5.2	96	137	16.2	4.3	
3.USAmb.2	2,871	3,639	65.6	4.6	89	117	12.0	4.0	
4.DSAmb.1	3,084	4,031	77.9	5.7	138	183	17.1	4.3	
4.DSAmb.2	2,889	3,725	70.0	4.8	95	126	12.5	4.2	
5.BC.1	3,854	4,640	62.1	5.2	141	174	12.6	4.6	
5.BC.2	3,506	4,335	65.9	4.9	131	163	13.0	4.5	
6.SR.1	4,214	4,721	47.8	6.1	195	218	10.5	4.9	
6.SR.2	4,023	4,538	47.3	5.5	183	212	12.0	4.7	

Table S7: Relative abundance of taxa determined to be differentially abundant

Sample	Decimal		Percentage	
	Full Microbial Community	Indicator Community	Full Microbial Community	Indicator Community
1.USUG.1	2.69E-02	2.42E-04	2.685%	0.024%
1.USUG.2	2.60E-02	7.58E-05	2.596%	0.008%
2.DSUG.1	1.69E-01	2.33E-03	16.904%	0.233%
2.DSUG.2	2.03E-01	1.06E-02	20.316%	1.064%
3.USAmb.1	1.40E-02	4.55E-04	1.398%	0.046%
3.USAmb.2	8.52E-03	1.56E-04	0.852%	0.016%
4.DSAmb.1	2.46E-01	8.19E-03	24.649%	0.819%
4.DSAmb.2	4.81E-02	8.03E-04	4.808%	0.080%
5.BC.1	7.06E-02	1.13E-03	7.055%	0.113%
5.BC.2	2.93E-02	8.28E-04	2.930%	0.083%
6.SR.1	2.05E-01	4.10E-03	20.544%	0.410%
6.SR.2	6.32E-02	2.72E-03	6.319%	0.272%

Table S8: Environmental parameter models and their correlations with the microbial community, as determined by BIOENV.

Model	Model Size	Correlation
$\log(\text{Cl}^-)$	1	0.5131
$\log(\text{Cl}^-) + \log(\text{SO}_4^{2-})$	2	0.6003
$\log(\text{Cl}^-) + \log(\text{SO}_4^{2-}) + \log(\text{Na})$	3	0.5841
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{Si}) + \log(\text{Sr})$	4	0.5815
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{NO}_3^-) + \log(\text{Si}) + \log(\text{Sr})$	5	0.5885
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{NO}_3^-) + \log(\text{K}) + \log(\text{Si}) + \log(\text{Sr})$	6	0.5896
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{Br}^-) + \log(\text{NO}_3^-) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	7	0.5851
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{Br}^-) + \log(\text{NO}_3^-) + \log(\text{Na}) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	8	0.5919
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{Br}^-) + \log(\text{SO}_4^{2-}) + \log(\text{Na}) + \log(\text{K}) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	9	0.5879
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{Br}^-) + \log(\text{NO}_3^-) + \log(\text{SO}_4^{2-}) + \log(\text{Na}) + \log(\text{K}) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	10	0.5811
$\log(\text{F}^-) + \log(\text{Cl}^-) + \log(\text{Br}^-) + \log(\text{NO}_3^-) + \log(\text{SO}_4^{2-}) + \log(\text{Mg}) + \log(\text{Na}) + \log(\text{K}) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	11	0.5541

Table S9: Environmental parameter models and their correlations with the indicator community, as determined by BIOENV.

Model	Model Size	Correlation
$\log(\text{Cl}^-)$	1	0.5965
$\log(\text{Cl}^-) + \log(\text{Mn})$	2	0.6240
$\log(\text{Cl}^-) + \log(\text{Mg}) + \log(\text{Si})$	3	0.6673
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Mg}) + \log(\text{Si})$	4	0.6804
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Mg}) + \log(\text{Si}) + \log(\text{Mn})$	5	0.6615
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Mg}) + \log(\text{Na}) + \log(\text{Si}) + \log(\text{Mn})$	6	0.6413
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Br}^-) + \log(\text{Mg}) + \log(\text{TDP}) + \log(\text{Si}) + \log(\text{Mn})$	7	0.6496
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Br}^-) + \log(\text{Mg}) + \log(\text{Na}) + \log(\text{TDP}) + \log(\text{Si}) + \log(\text{Mn})$	8	0.6378
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Br}^-) + \log(\text{Mg}) + \log(\text{Na}) + \log(\text{TDP}) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	9	0.6227
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Br}^-) + \log(\text{Mg}) + \log(\text{Na}) + \log(\text{K}) + \log(\text{TDP}) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	10	0.6160
$\log(\text{Cl}^-) + \log(\text{NO}_2\text{--N}) + \log(\text{Br}^-) + \log(\text{Ca}) + \log(\text{Mg}) + \log(\text{Na}) + \log(\text{K}) + \log(\text{TDP}) + \log(\text{Si}) + \log(\text{Mn}) + \log(\text{Sr})$	11	0.6072

### 3. Supplemental Figures

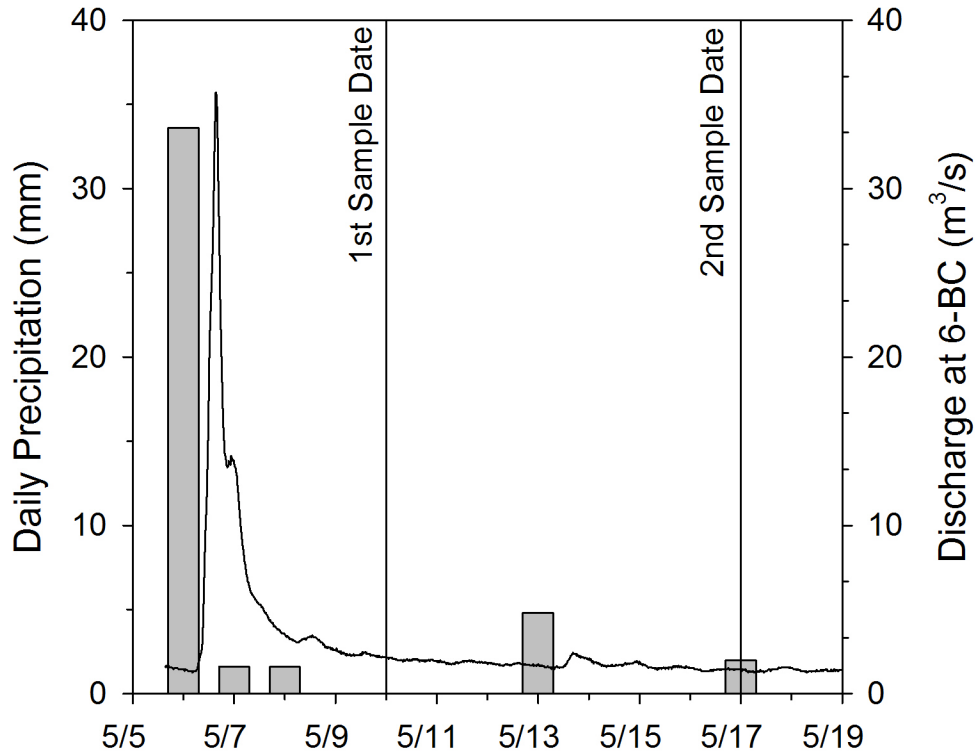


Figure S1: Daily Precipitation (left axis, bars) and continuous discharge (right axis, plotted curve) at 6-BC (USGS station number 01473900) for the study period. There was 3.5 cm of precipitation on May 6<sup>th</sup> and a trace amount of rainfall (1.6 mm) on May 7<sup>th</sup> and 8<sup>th</sup>. There were trace amounts of rainfall before the May 17<sup>th</sup> sampling event, but the total amount of precipitation before the second sampling date was only 4.8 mm. Source data available at <https://waterdata.usgs.gov/usa/nwis/uv?01473900>.

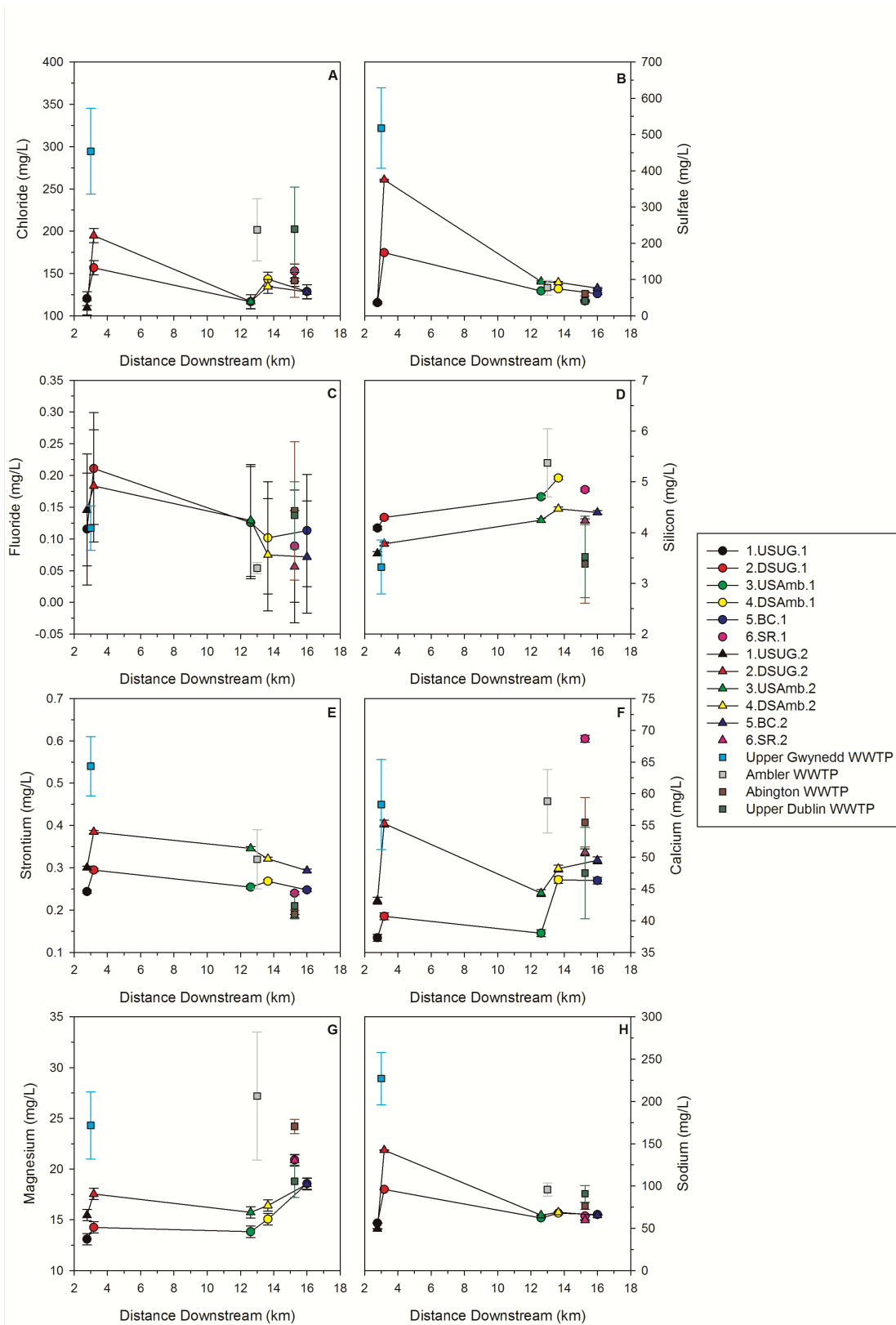


Figure S2:  $\text{Cl}^-$  (A),  $\text{SO}_4^{2-}$  (B),  $\text{F}^-$  (C), Si (D), Sr (E), Ca (F), Mg (G), and Na (H) concentrations with distance along the streams. Error bars indicate analytical error. Average WWTP concentrations from three different samples are shown in squares. Error bars indicate the standard deviation among the three samples.

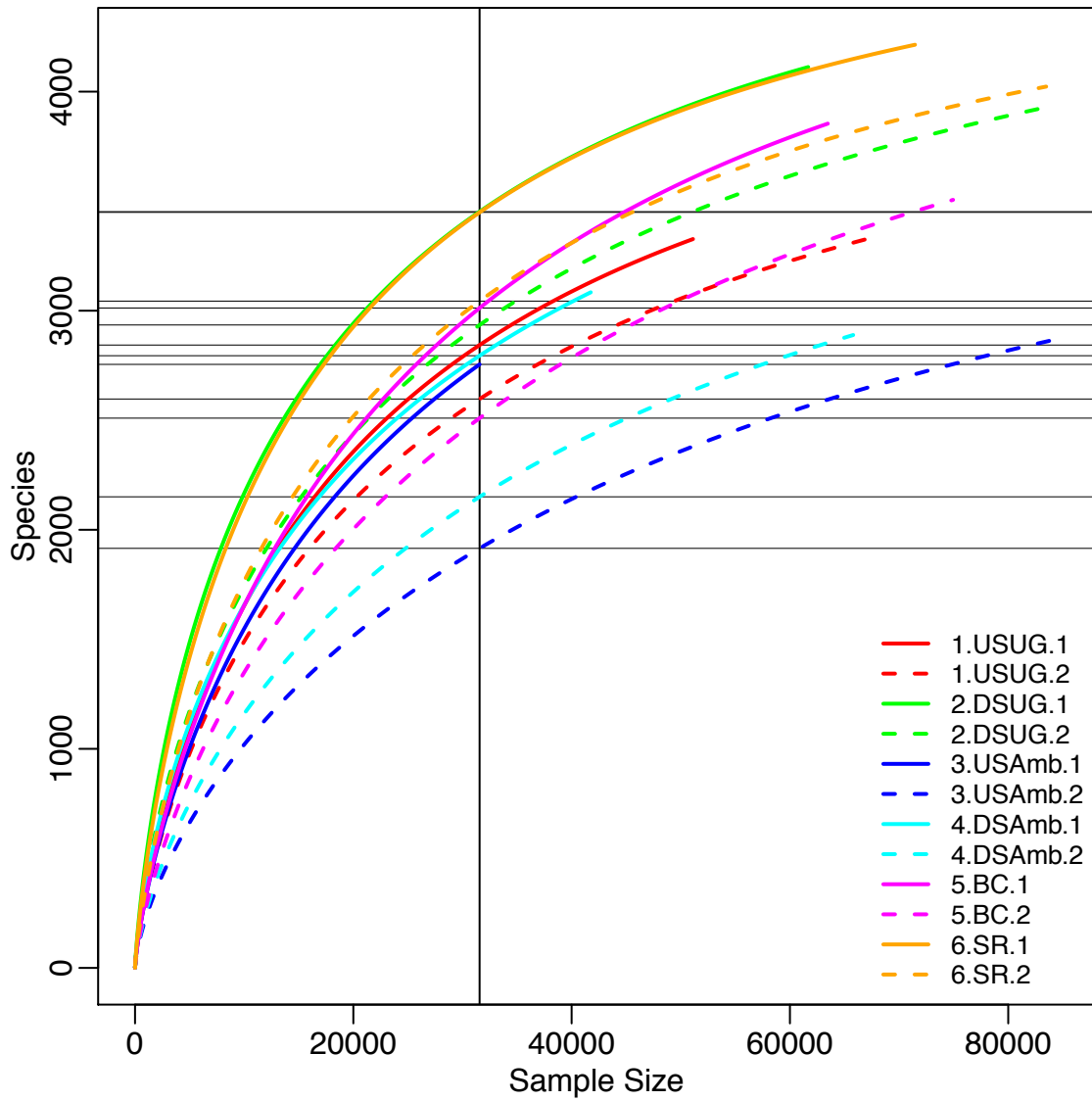


Figure S3: Plot of the rarefaction curves indicate that the amplicon libraries have been sufficiently sampled.

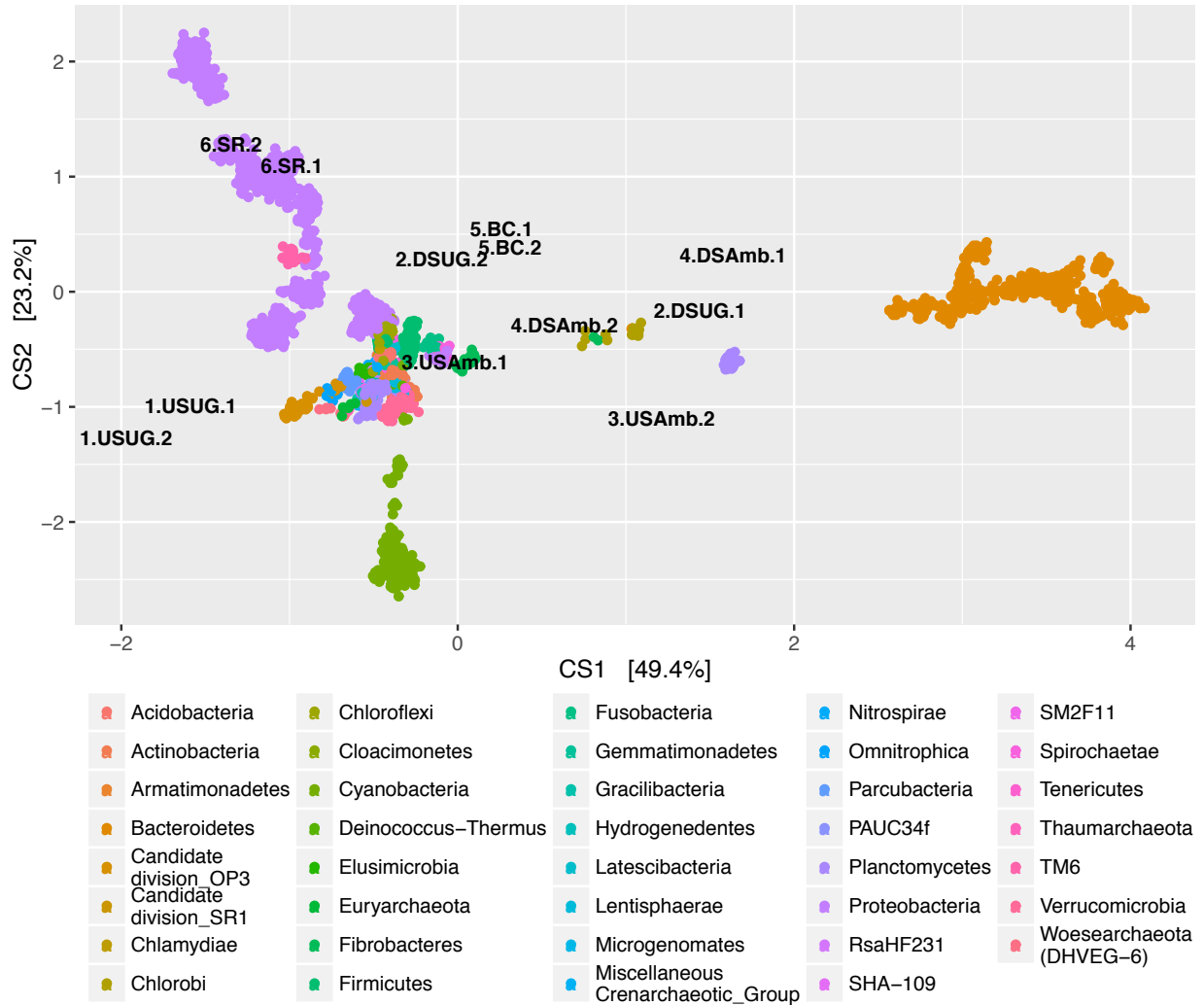


Figure S4: Plot of DPCoA results for the entire microbial community. Sample positions have been inflated 10x to facilitate visual distinction between them. Colors designate each taxa's phylum membership. The three most abundant phyla distinguish themselves, with Bacteroidetes dominating the first axis and Proteobacteria and Cyanobacteria segmenting on the positive and negative values of the second axis respectively.



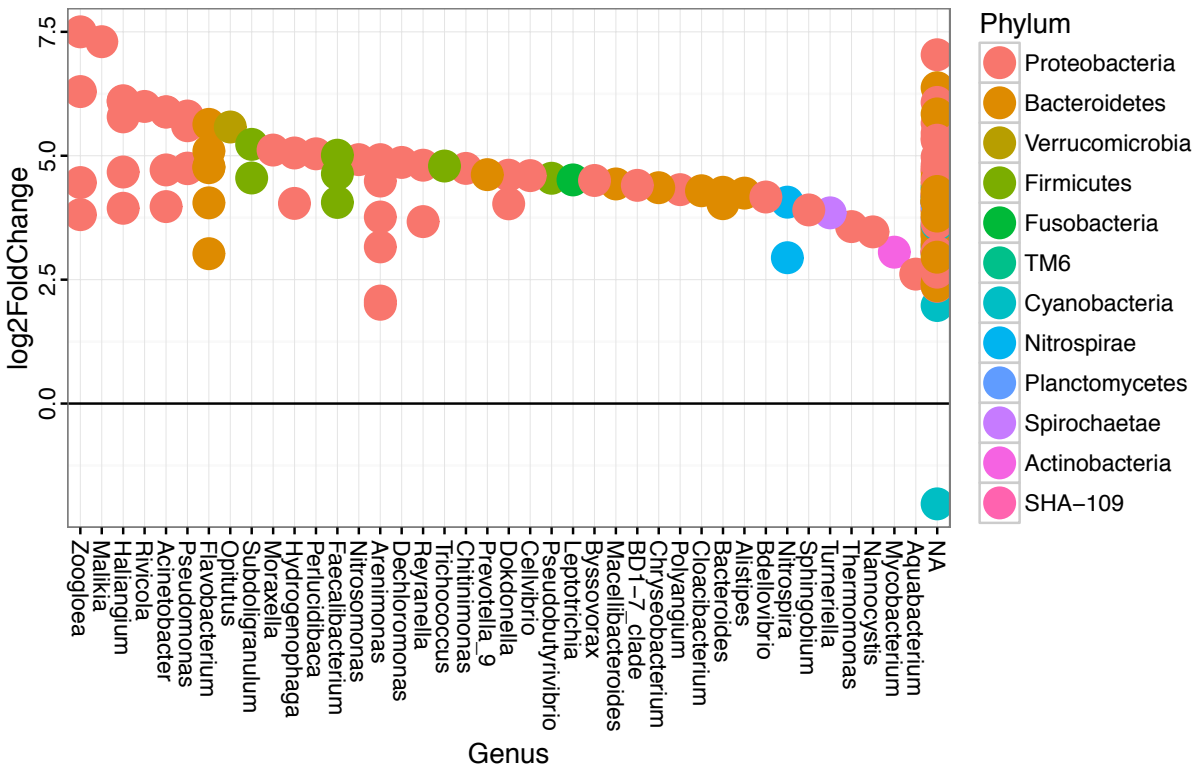


Figure S5: Log2-fold changes of taxa found to be differentially abundant between sites immediately up and downstream of effluent sources. Positive values indicate an increase in abundance at downstream sites. Multiple circles falling within a Genus indicate cases where multiple taxa within that Genus were found to be differentially abundant. Differentially abundant taxa not annotated at the Genus level are located above “NA” on the ordinal axis (far right).

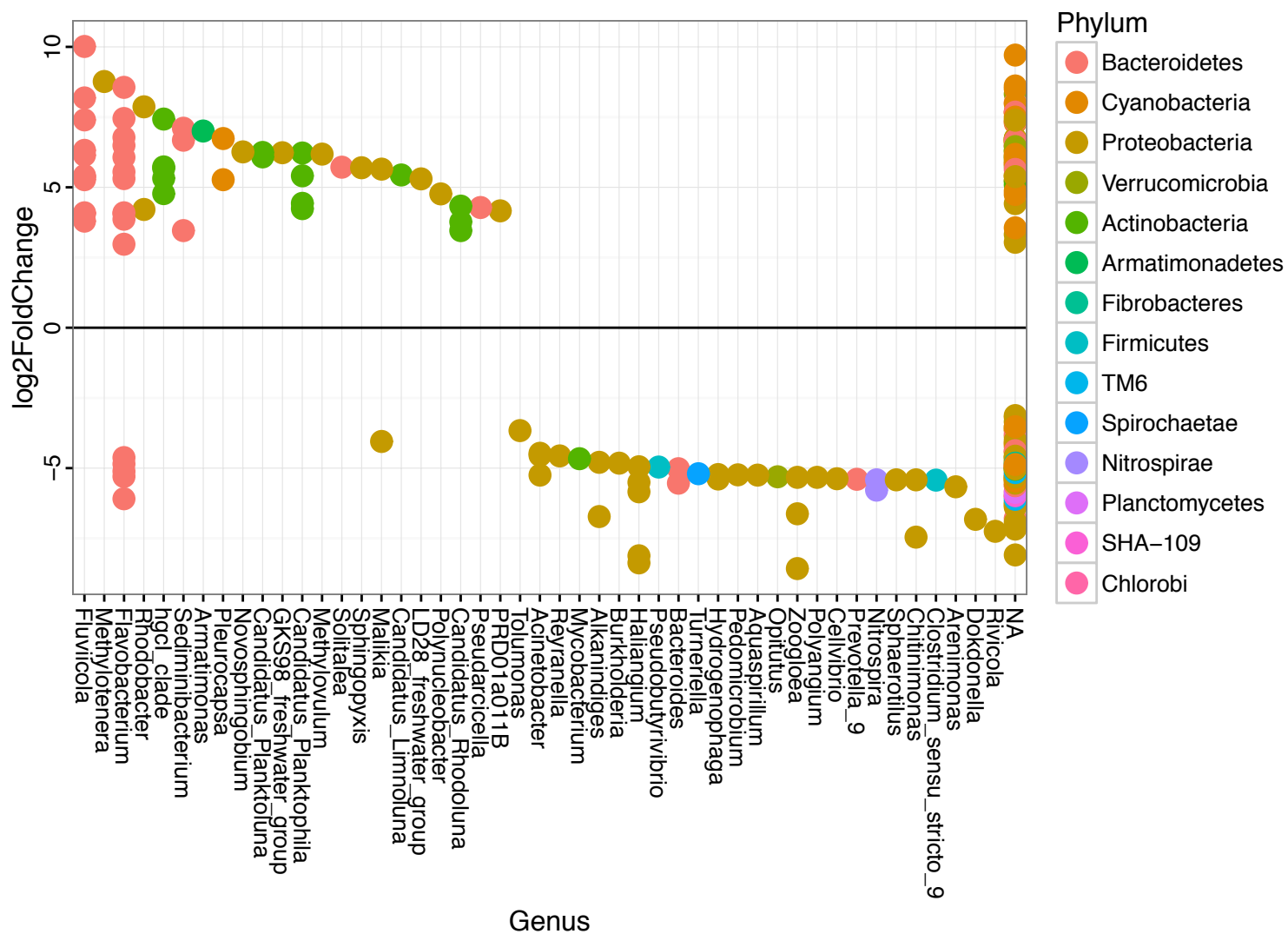


Figure S6: Log2-fold changes of taxa found to be differentially abundant between Sites 2 and 3. Positive values indicate an increase in abundance at Site 3, negative value indicate a decrease in abundance at Site 3. Multiple circles falling within a genus indicate cases where multiple taxa within that genus were found to be differentially abundant. Differentially abundant taxa not annotated at the genus level are located above “NA” on the ordinal axis (far right).

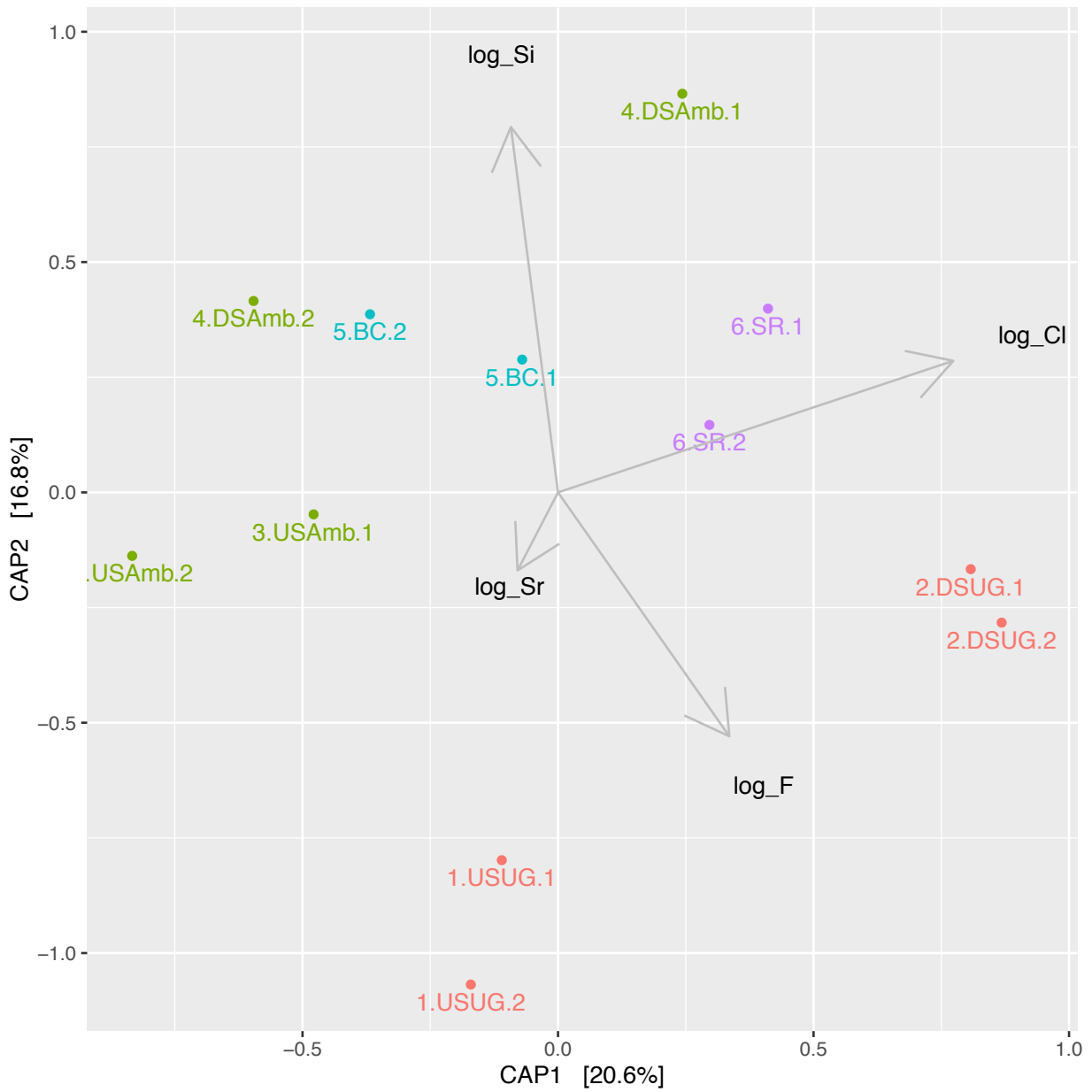


Figure S7: db-RDA of sampling sites and VST OTU counts of the full microbial community. CAP refers to constrained analysis of principal components.

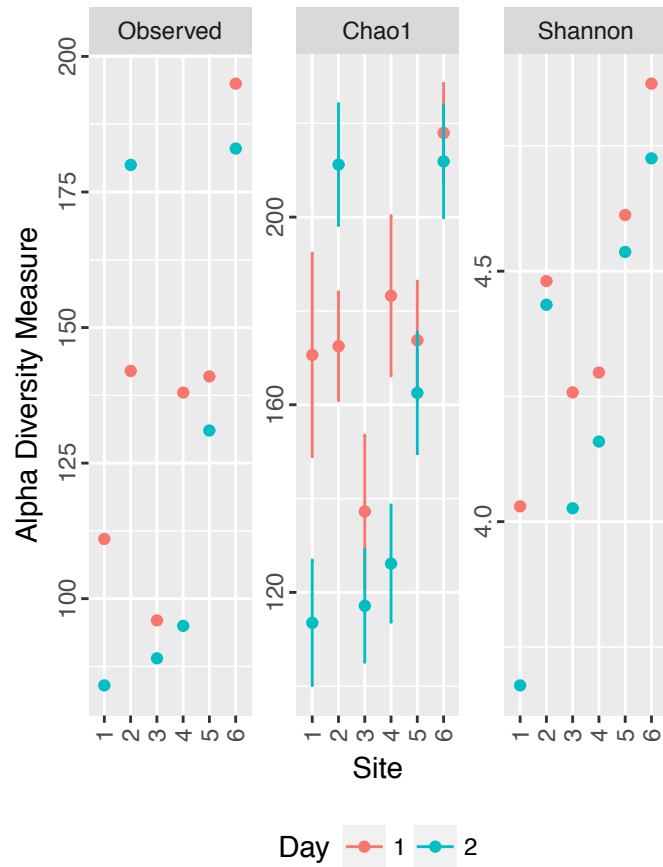


Figure S8: Alpha diversity measures for the indicator subset community. Error bars for Chao1 represent standard error (SE).

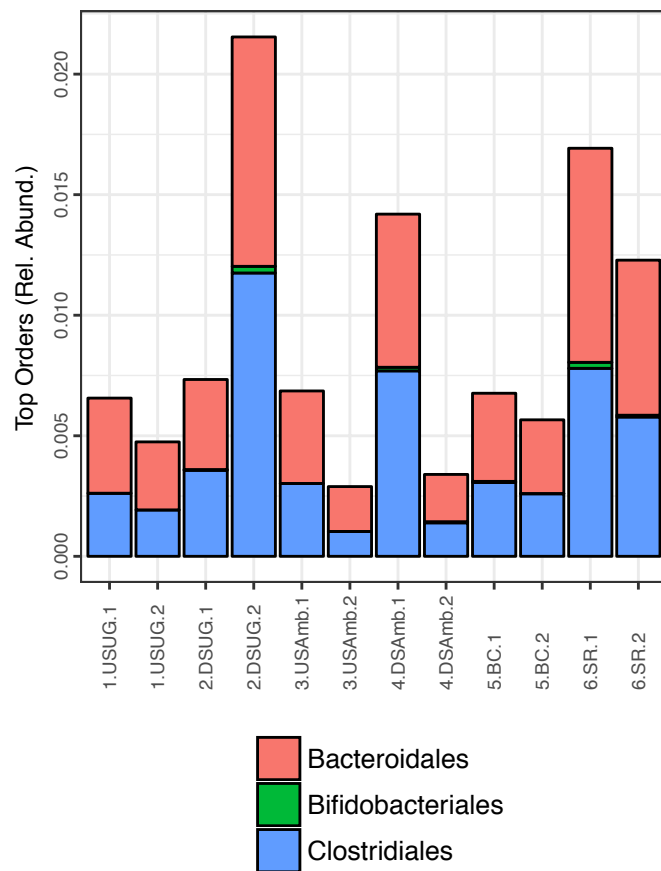


Figure S9 Barcharts representing the relative abundance of the three orders in the indicator subset.

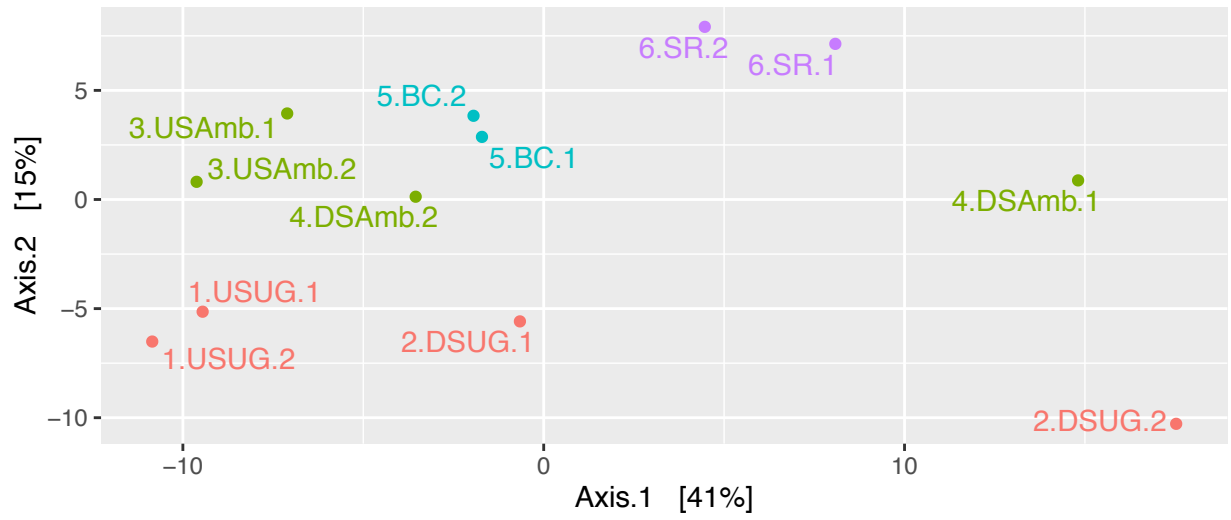


Figure S10: A plot of the PCA results for the indicator community subset.

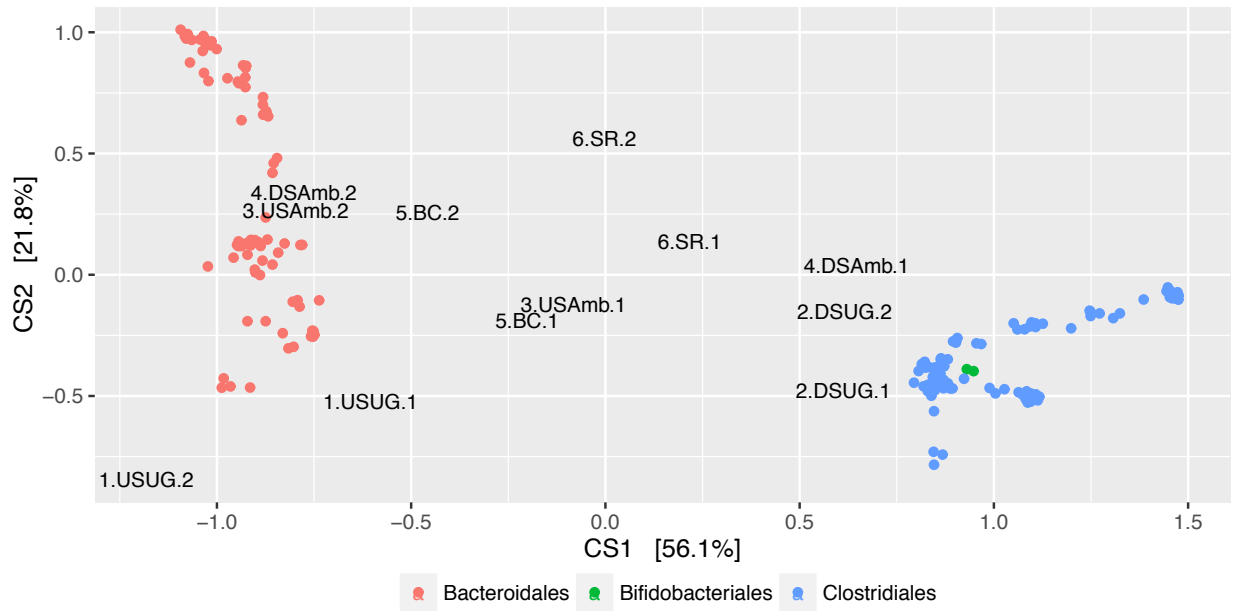


Figure S11: Plot of DPCoA results for the indicator subset of the microbial community. Sample positions have been inflated 5x to facilitate visual distinction between them.

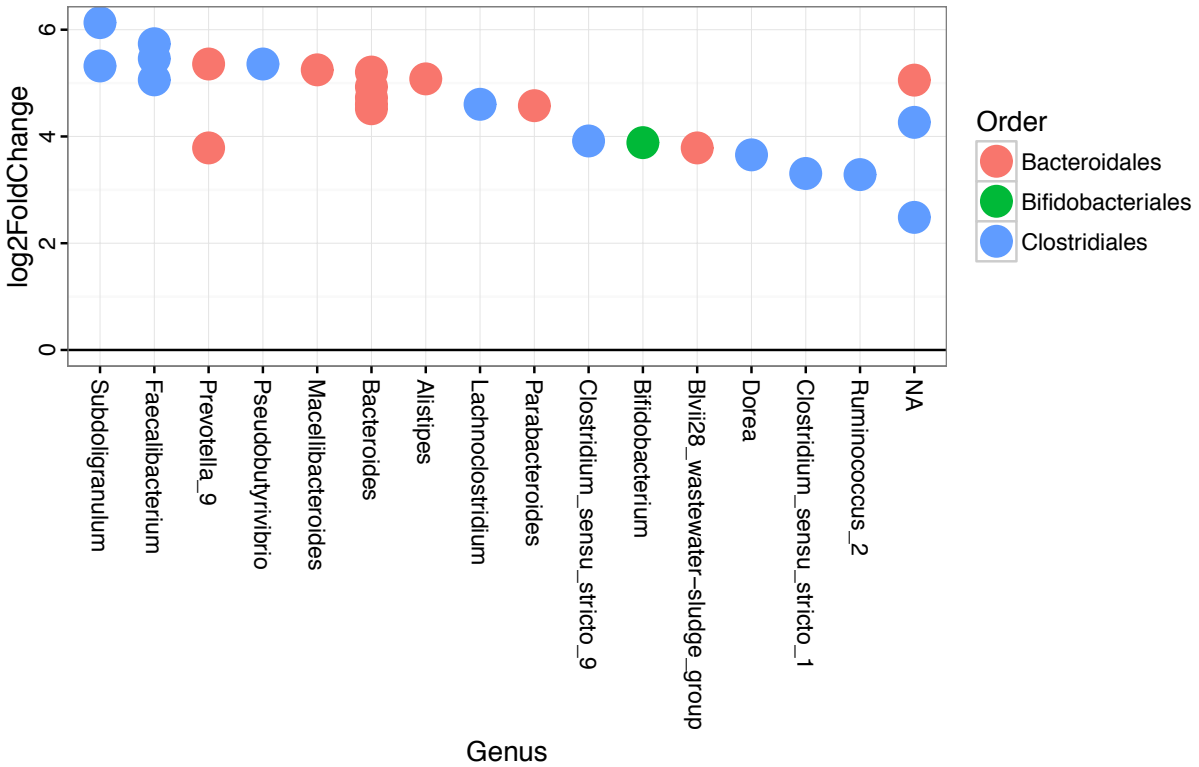


Figure S12: Log<sub>2</sub>-fold changes of taxa found to be differentially abundant in the indicator community between sites immediately upstream and downstream of effluent sources. Positive values indicate an increase in abundance at downstream sites. Multiple circles falling within a genus indicate cases where multiple taxa within that genus were found to be differentially abundant. Differentially abundant taxa not annotated at the genus level are located above “NA” on the ordinal axis (far right).



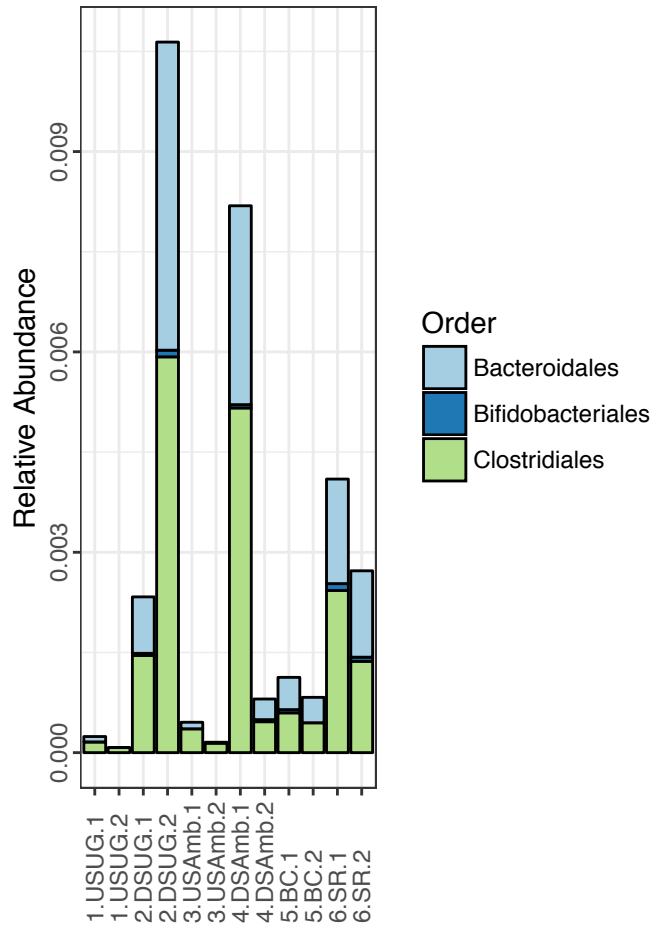


Figure S13: Relative abundance of differentially abundant taxa, observed during the analysis of the indicator subset. Samples from sampling sites 5.BC and 6.SR are included only for comparison and were not included in the differential abundance testing.

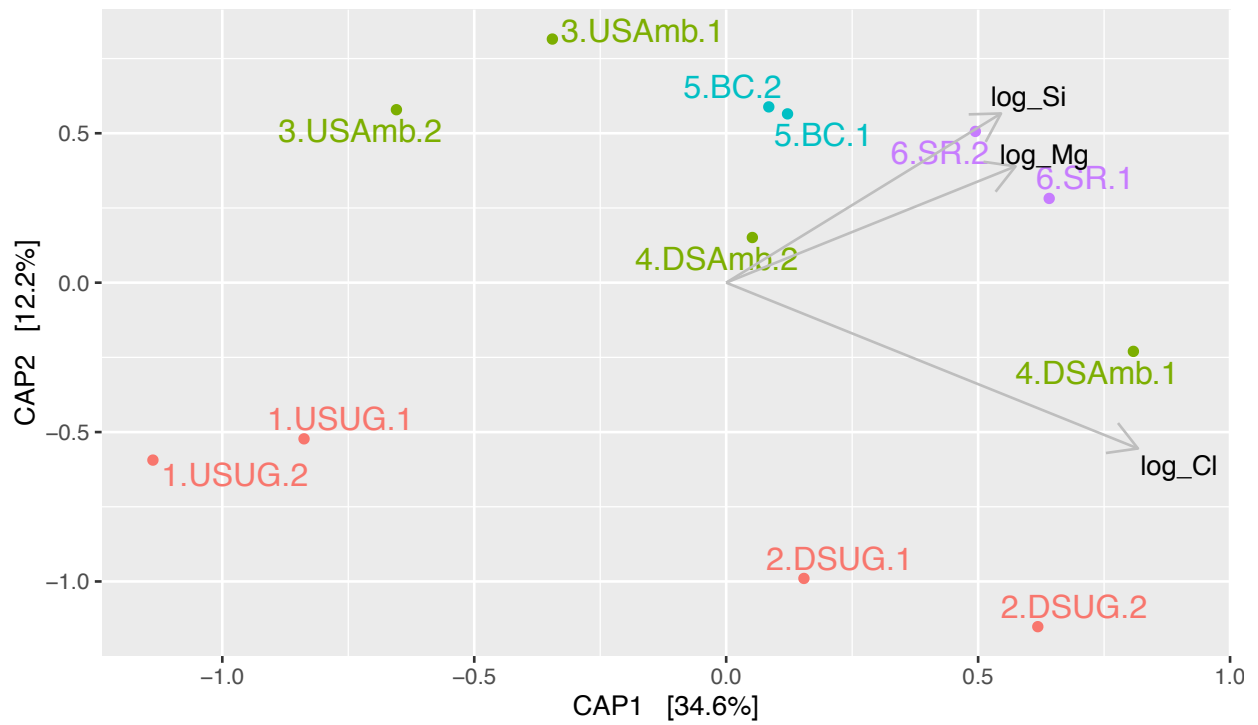


Figure S14: db-RDA of sampling sites and VST OTU counts of the indicator community.

## **Supplemental References**

Homer C, Dewitz J, Yang LM, Jin S, Danielson P, Xian G, et al. Completion of the 2011 National Land Cover Database for the Conterminous United States - Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering and Remote Sensing* 2015; 81: 345-354.