Supplemental Material

Authors: Lauren Sassoubre, Kevan Yamahara, Alexandria Boehm

Title: Temporal stability of the microbial community in sewage-polluted seawater exposed to natural sunlight cycles and marine microbiota

Methods for investigating changes in OTU relative abundance between treatments using edgeR. The edgeR package facilitates statistical comparisons of the OTUs in each sample using negative binomial general linear models and user defined contrasts (1, 2). The data were filtered to remove OTUs with very low raw counts (OTUs that had a count of 2 or more in at least one sample were retained by the program). The filtered dataset contained 5338 unique OTUs. Raw OTU counts (not rarefied) were normalized to 'pseudo-counts' using a model-based method that applies a correction factor to scale all the library sizes to one 'effective library size.' Contrasts were defined as the difference between the 'pseudo-counts' in one microcosm minus the 'pseudo-counts' in another microcosm for each OTU at each time point. Four contrasts were examined (1) dark unfiltered relative to dark filtered microcosms, (2) light unfiltered relative to light filtered microcosms, (3) light filtered relative to dark filtered microcosms, (4) light unfiltered relative to dark unfiltered. To account for Type 1 errors, all p values were adjusted to False Discovery Rates (FDR) (3). Changes in OTU counts were considered statistically significant if the FDR was less than 10% (FDR<0.1).

Since we did not have biological replicates to calculate the biological coefficient of variation (BCV) (1, 2, 4-6), we based the dispersion estimates (used to calculate the

BCV) on the samples from the dark filtered microcosm at 0 h and 12 h. We made the assumption that the first two samples from the dark filtered microcosms (0 h and 12 h) could be used as replicates based on the following. Measures of observed richness and alpha diversity were similar between the dark filtered microcosms at 0 h and 12 h, with observed richness of 721 and 786, respectively, and Shannon diversity indices of 5.3 and 5.4, respectively. We also compared the dark filtered microcosm samples at 0 h and 12 h using edgeR and found no significant differences in OTU abundance between the two samples (FDR<0.1).

The edgeR manual discusses a few options for studies when biological replicates are not available and one option is to treat samples expected not to differ as replicates (in this case the dark microcosm) (6). We calculated the BCV using the dark filtered microcosm samples collected at 0 h and 12 h and found it to be equal to 0.28. This value is between 0.1 to 0.4, the range given by the edgeR manual (0.4 for human data and 0.1 for data on genetically identical model organisms) (6). The implications of assuming the dark filtered microcosm samples at 0 h and 12 h can be used as replicates to calculate the BCV are discussed in the Discussion section of the manuscript.

Methods for investigating similarities in microbial community composition using the vegan package. The vegan package was used to investigate similarities in microbial community composition between samples. Rarefied OTU counts were square root transformed and Bray-Curtis dissimilarities between samples were calculated. A taxa-time relationship was investigated using a Mantel test on the OTU dissimilarity matrices

and the temporal distance matrix. Similarities between samples were visualized with nonmetric multidimensional scaling (nMDS) plots. We used analysis of similarities (ANOSIM) to test the null hypothesis that (1) there was no significant difference in community composition at the OTU level based on sunlight exposure and (2) there was no significant difference in community composition at the OTU level based on the presence or absence of marine microbiota.

Methods for investigating changes in the abundance of sewage-derived microorganisms with time using SourceTracker. The SourceTracker package (7) was used to investigate how the sewage microbial community changed with time in each treatment. SourceTracker uses a Bayesian approach to determine the proportion of the user-identified sources and an "unknown" source in each sample (each sample is termed "sink" in the program) (7). The program rarifies OTU counts to the minimum number of sequences in a sample (4635) and thus uses a different rarified data set than we used for the rest of the analyses described earlier. The unfiltered seawater (before seeding with sieved sewage) and the filtered seawater seeded with sieved sewage at the initial time point were named as "sources". All other samples were considered "sinks". The program's "unknown" source category is used for the proportion of OTUs that are not assigned the user-identified sources.

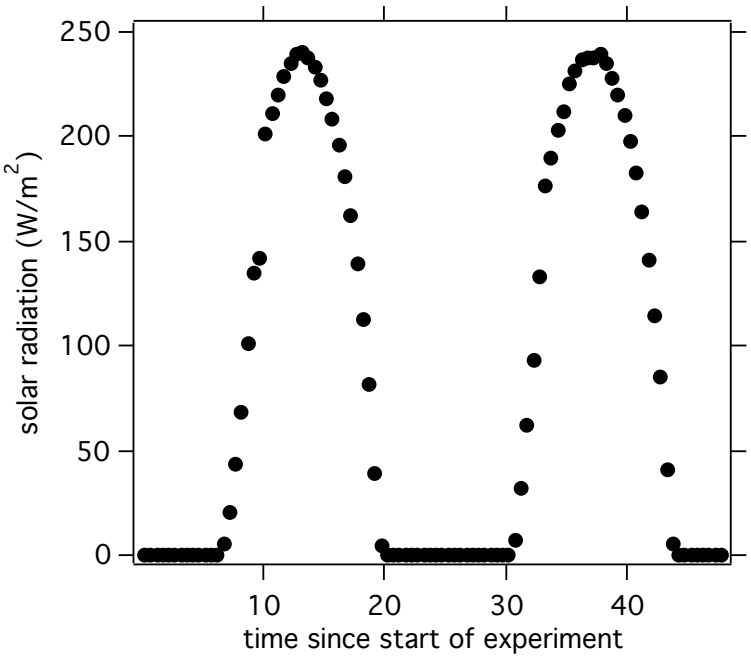
Methods for quantification of bacterial 16S rRNA by Quantitative Polymerase Chain Reaction (qPCR). In each sample used for microbial community analysis, the number of copies of a region of the16S rRNA gene specific to bacteria was quantified using qPCR

on an Applied Biosystems StepOnePlus. Twenty microliter reactions consisted of 1X Platinum Quantitative PCR SuperMix-UDG with ROX (includes 3 mM MgCl) (Invitrogen, Life Technologies, Carlsbad, CA), 0.2 μM forward primer (BACT1369F: CGGTGAATACGTTCYCGG) (64), 0.2 μM reverse primer (PROK1492R:GGWTACCTTGTTACGACTT) (64), 0.1 μM probe (TM1389F:CTTGTACACACCGCCCGTC) (64), and 2 μl template. Samples were run at 1:100 dilutions. While the assay targeted bacteria broadly, standard curves were constructed using genomic DNA from one bacterium with a known number of rRNA operons (4), *Enterococcus faecalis* (ATCC 29212). Standard curves, no template controls (NTCs) and samples were run in triplicate on each plate. The following cycling parameters were used: 50 °C for 2 minutes, 94 °C for 10 minutes followed by 40 cycles of 94 °C for 15 seconds and 56 °C for 30 seconds (64).

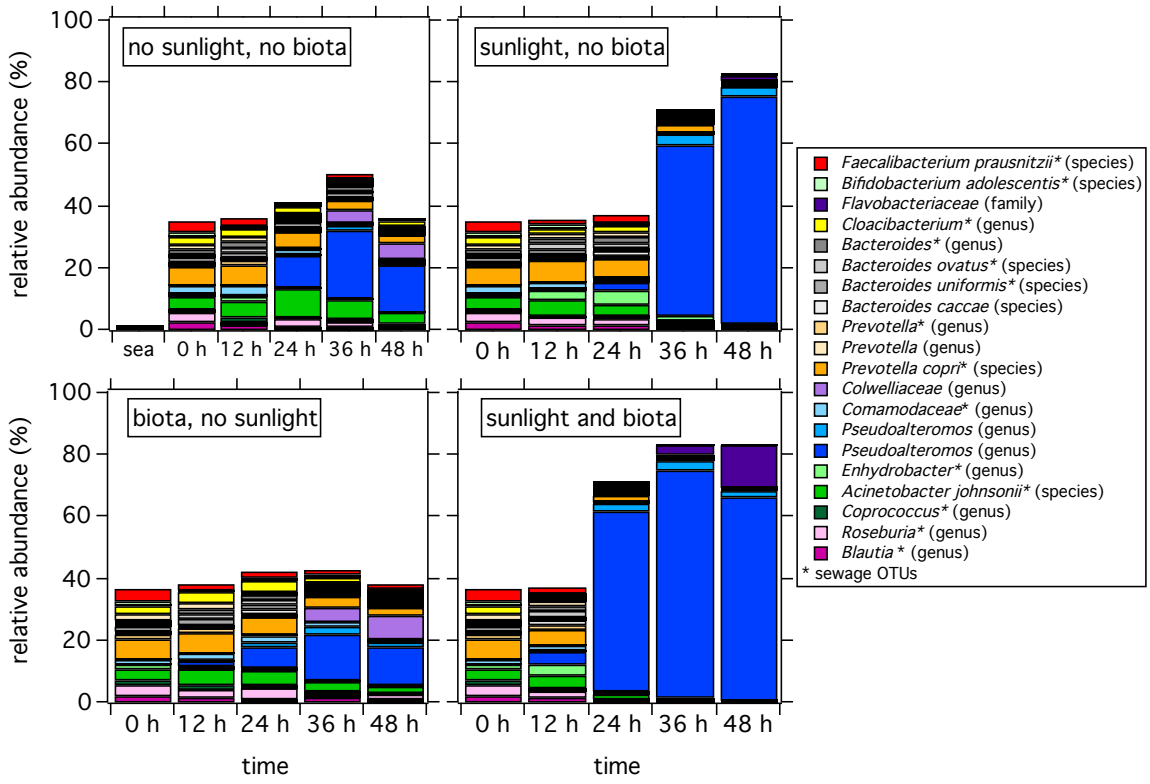
Methods for enumeration of culturable enterococci. Standard methods were used to enumerate enterococci on mEI media in each dialysis bag (4). The natural log of the enterococci concentrations was used to calculate the inactivation rates k and T_{90} values assuming first order decay.

Results for quantification of the 16S rRNA gene by QPCR and enterococci by culture-based methods. The total number of bacterial 16S rRNA gene copies remained relatively stable over time (Figure S3). Concentrations of culturable enterococci (from sewage and/or seawater) decreased to below the detection limit (1 CFU/100 mL) by the end of the first day of sunlight exposure but did not change in the microcosms kept in the dark

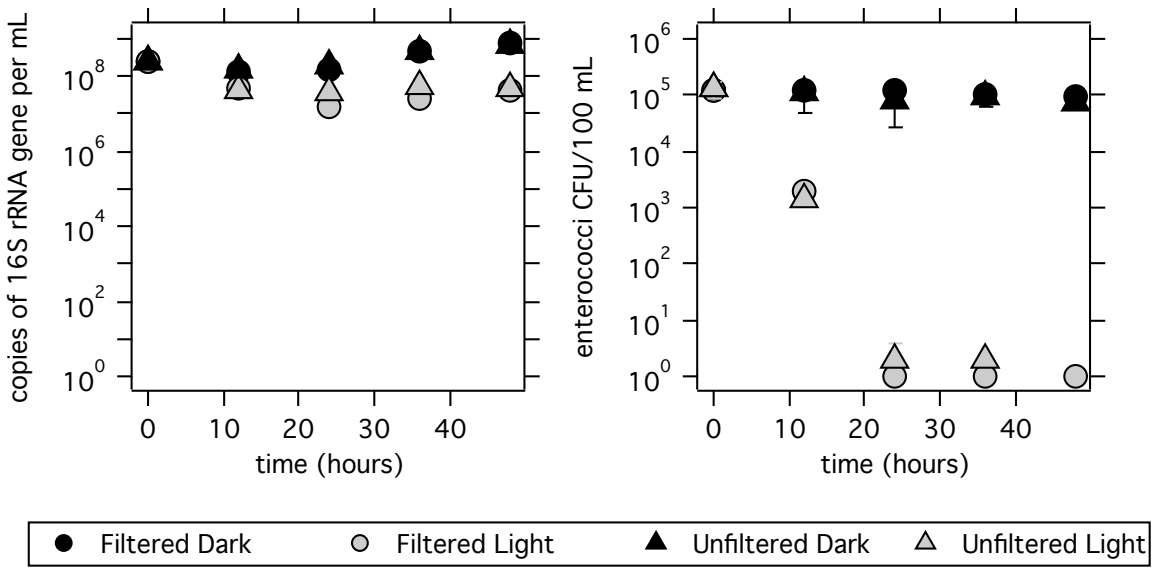
(~10^5 colony forming units (CFU)/100 mL) (Figure S3). The enterococci inactivation rates (k) for the sunlit microcosms with and without marine microbiota were calculated using the first three time points and we found to be -0.49 $\pm$ 0.085 hr^{-1} and -0.46 $\pm$ 0.048 hr^{-1}, corresponding to T_{90}s of 4.26 and 4.47 hr, respectively. While culturable enterococci concentrations as high as 10^5 CFU/100 ml were measured, only one OTU was assigned to the genus *Enterococcus* in the NGS analysis, and the relative abundance of this OTU was below 0.1% in all the microcosms it was measured in. This OTU assigned to *Enterococcus* was identified in the dark filtered microcosms at 0 h, 12 h, 24 h, the light filtered microcosms at 0 h, 12 h, 24 h, 36 h, 48 h, the dark unfiltered microcosms at 24 h and 36 h, and the light unfiltered microcosms at 12 h. *Enterococcus* likely represents a very low proportion of the bacteria present in sewage and seawater, making it too rare a target to be detected using NGS at the sequencing depth used. This is consistent with previous studies investigating *Enterococcus* in beach sands using culture-based methods and NSG (66, 67). Decay of the NGS signal for sewage microorganisms followed a similar trend to enterococci bacteria but only in the sunlit microcosms where culturable enterococci concentrations decayed significantly within 24 hours. In the dark microcosms, enterococci levels remained elevated well after the sewage signal from NGS had decayed.



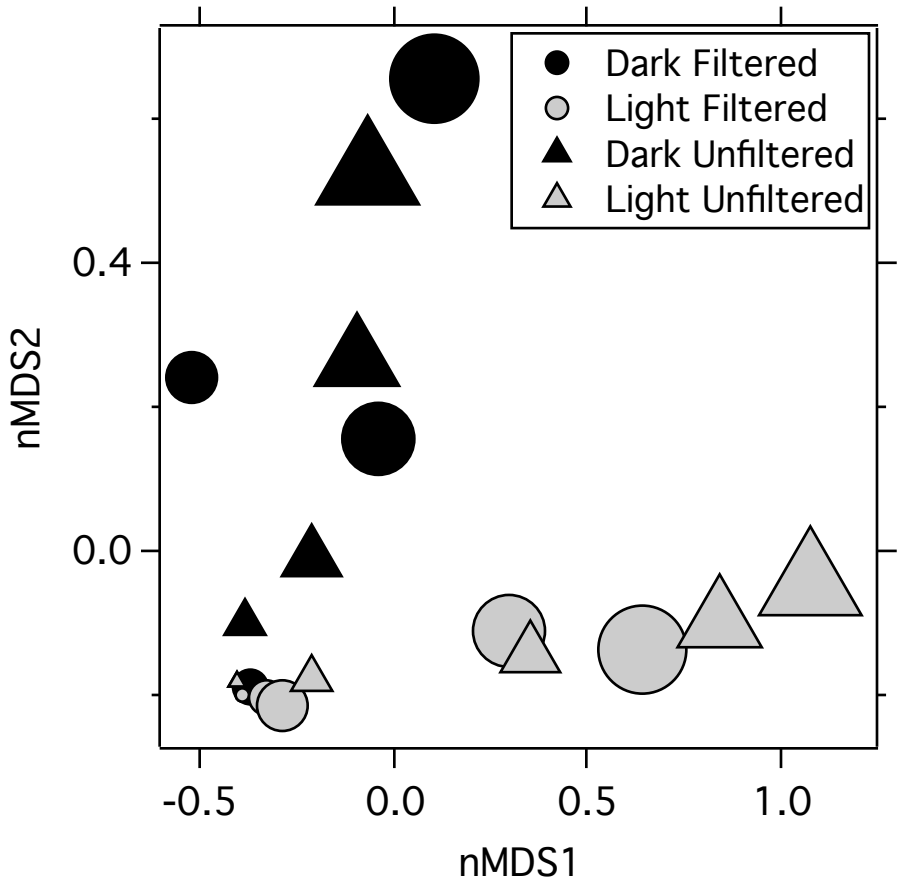
Supplemental Material Figure S1. Solar irradiance collected every 30 minutes during the experiment. Irradiance includes energy between 400 and 2700 nm.



Supplemental Material Figure S2. The 20 most abundant OTUs labeled with their lowest taxonomic designation assigned by QIIME workflow (noted in parentheses in the legend). An asterisk next to the taxonomic designation indicates a 'sewage OTU.' Top left panel. The microcosms kept in the dark with no biota (dark filtered microcosms). Added to this plot are the relative abundances for the unfiltered seawater sample labeled "sea". Top right panel. The microcosms exposed to sunlight with no biota (light filtered microcosms). Bottom left panel. The microcosms kept in the dark with biota (dark unfiltered microcosms). Bottom right panel. The microcosms exposed to sunlight with biota (light unfiltered microcosms). The y-axes in all plots represent the percent relative abundance. The x-axis in all plots is time. Note that *Pseudoalteromonas* (genus) and *Prevotella* (genus) are listed twice in the legend because two unique OTUs received this as their lowest taxonomic assignment.



Supplemental Material Figure S3. Changes in the bacterial 16S rRNA signal measured by qPCR (left panel) and changes in culturable enterococci concentrations (right panel). Dark circles represent dark filtered microcosms. Light circles represent light filtered microcosms. Dark triangles represent dark unfiltered microcosms. Light triangles represent light filtered microcosms. Time since the start of the experiment is shown on the x-axes. Error bars on the left panel represent the standard deviation of triplicate qPCRs and are difficult to see because they are small. Error bars on the right panel represent the standard deviation of replicate plate counts.



Supplemental Material Figure S4. Nonmetric multidimensional scaling (nMDS) based on Bray-Curtis dissimilarities (stress=0.066). Black circles and triangles represent dark filtered and unfiltered microcosms, respectively. Grey circles and triangles represent light filtered and unfiltered microcosms, respectively. The size of the marker corresponds to the time the sample was collected; the smallest markers represent 0 h and the largest represent 48 h.

References

1. **McCarthy DJ, Chen Y, Smyth GK.** 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. **40:**4288-4297.
2. **Robinson MD, Smyth GK.** 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics **9:**321-332.
3. **Benjamini Y, Hochberg Y.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B **57:**289-300.
4. **Robinson MD, McCarthy DJ, Smyth GK.** 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26:**139-140.
5. **Pal B, Bouras T, Shi W, Vaillant F, Sheridan JM, Fu N, Breslin K, Jiang K, Ritchie ME, Young M, Lindeman GJ, Smyth GK, Visvader JE.** 2013. Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. Cell reports **3:**411-426.
6. **Robinson M, McCarthy D, Chen Y, Smyth G.** 2013. edgeR: differential expression analysis of digital gene expression data User's Guide.
7. **Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST.** 2011. Bayesian community-wide culture-independent microbial source tracking. Nat. Methods **8:**761-763.