One man's trash is another's treasure: Understanding microbial diversity and community structure via metagenomic analysis in sediments near a sewage outfall

Elizabeth Makrides[3,*], Christopher Graves[1,*], Victor Schmidt[1,2,*], Anne Giblin[2], Zoe Cardon[2], David Rand[1]

1. Ecology and Evolutionary Biology, Brown University, Providence, RI; 2. Marine Biological Laboratories, Woods Hole, MA;
3. Division of Applied Mathematics, Brown University, Providence, RI; * Contributed equally to this project

Introduction

Just as different human communities may grow up depending on the type of jobs available (think Silicon Valley), microbial community structures depend on the resources available. One question researchers have asked is whether the response to resource availability is primarily reflected in shifting species, or if there is a "lottery" system, in which any species that satisfies a certain functional niche can enter and remain. The lottery hypothesis suggests that different environments will primarily change the functional composition of the communities (i.e. what traits are present) rather than the taxonomic composition (i.e. what species are present). Our project examines the functional and taxonomic structure of microbial communities along Greenwood Creek in the Plum Island Estuary Long Term Ecological Research (LTER) site. Due to the presence of a sewage outfall at the top of the creek, a strong gradient of nitrogen and other compounds has existed in Greenwood Creek for many years which could have substantial effects on community structure. Since this nitrogen gradient co-occurs with a strong salinity gradient, we also took sediment samples from a nearby creek with a similar salinity gradient to put our findings in context. Following sampling, we extracted DNA from the soil and used "next-generation" sequencing technology to sequence millions of random DNA fragments from the community of microbes in the sediment. We will compare these sequences to DNA sequences in existing databases to determine the presence and abundance of different functional genes and taxonomic groups along the two creeks. This combination of field ecology, environmental biology, high-throughput sequencing and bioinformatics allows us to get a full picture of the otherwise invisible structure and function of microbial communities and address basic hypotheses regarding microbial ecology.

Sampling procedure

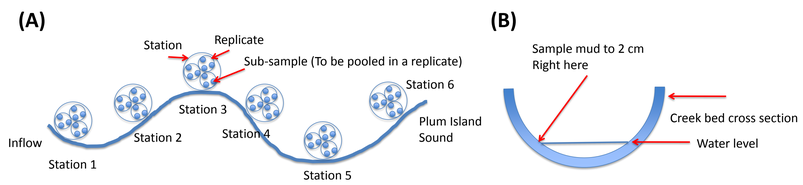


Figure 1. (A) Overall schematic of sampling. **(B)** Sample location relative to water level.

Sediments were sampled at 6 sites along two creeks in the Plum Island Estuary LTER. The focal creek, Greenwood creek, has a 'freshwater' input of treated sewage effluent, whereas the reference creek, Egypt River, is fed by a drinking water reservoir. The reference was chosen based on similar size and tidal salinity cycle to that of Greenwood creek. Samples of the sediments were taken within 0.3 meters of the water level during low tide on October 8, 2011 by taking 2 cm cores. Three biological replicates were sampled within a range of 5 m at each site along the creek by mixing 3 of the 2 cm cores (taken within 0.5 meters of one another) into a sterile sample cup. The 3 cores in each biological replicate were homogenized thoroughly and aliquots of 1.5 ml of mixed sediment were flash frozen in Liquid N2. GPS coordinates, water and sediment temperature, and salinity were measured at each sampling site. Additional sediment cores were also taken at each biological replicate for bulk density analysis of soil composition, moisture content, CHN quantification, and N isotopes.

Sample Sites

Figure 2. Left: Representative sample site (Egypt River – reference station 4). Right: scouting trip on a salt marsh in the Plum Island LTER.

Figure 3. Above: Greenwood sample sites. Right: reference (Egypt) sample sites. Note reference 5 and 6 are reversed relative to the mouth of the river. The two creeks are separated by approximately 5km.

Sediment characteristics

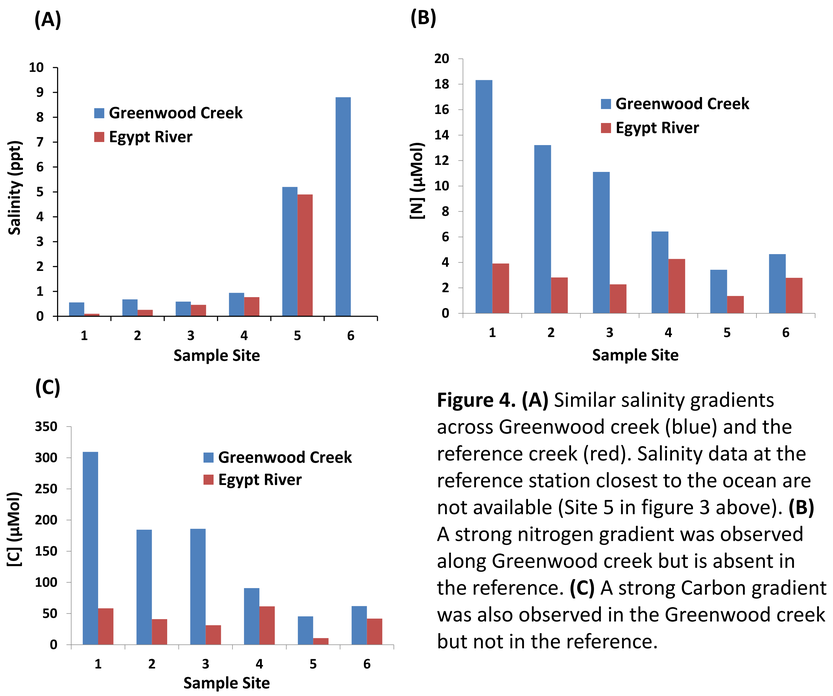


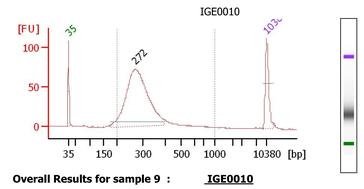
Figure 4. (A) Similar salinity gradients across Greenwood creek (blue) and the reference creek (red). Salinity data at the reference station closest to the ocean are not available (Site 5 in figure 3 above). **(B)** A strong nitrogen gradient was observed along Greenwood creek but is absent in the reference. **(C)** A strong Carbon gradient was also observed in the Greenwood creek but not in the reference.

Next Generation Sequencing

(A)

Peak table for sample 9 : IGE0010

Peak		Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	◀	35	125.00	5,411.3	Lower Marker
2		272	840.87	4,691.4	
3	▶	10,380	75.00	10.9	Upper Marker



Overall Results for sample 9 : IGE0010

Number of peaks found:	1
Noise:	0.2
Corr. Area 1:	1,102.6

Region table for sample 9 : IGE0010

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
200	1,000	1,102.6	95	307	28.5	897.86	4,730.5	

(B)

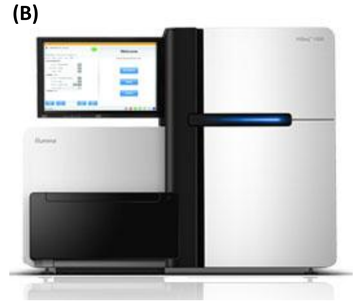


Figure 5. (A) Agilent Biolanalyzer characterization of the quality and size distribution of one of the next generation sequencing libraries. **(B)** Samples are now being sequenced on the Illumina HiSeq 2000, which sequences millions of DNA fragments

Following sample collection, we extracted genomic DNA from the sediment samples and sheared the DNA into random fragments with a median size of ~175 bps. We then add specific DNA sequences called "adapters" onto each end of the fragments so that they bind onto the sequencing machine. The fragments are then amplified and the next generation DNA sequencing libraries are complete. The Illumina HiSeq 2000 sequences 100 base pairs from each end of our DNA fragments for millions of fragments.

Data analysis

Our next generation sequencing data will consist of sequences from millions of DNA fragments representing the "metagenomes" of the sediment samples. We will combine the 100 bp sequences from each end of the fragment to produce longer sequences for analysis. The community metagenomic data will provide two kinds of information: (1) 16S and 18S rDNA fragments will allow us to identify some of the microbial community members present, and (2) other DNA fragments will provide information about the functional capacity of the community as a whole. The shotgun metagenomic data derived from the Illumina run will be run through MG-RAST (a publicly available pipeline http://metagenomics.anl.gov/), which queries a number of on-line databases for sequence matches to those sequences detected in our samples, and reports detected functional capabilities in defined "subsystems" (e.g. anaerobic and aerobic respiratory metabolisms, nitrogen and sulfur cycling pathways, etc.) We will use the raw data and the subsystem data to explore whether long-term exposure to treated sewage effluent has altered microbial community structure and functionality.

Acknowledgements

Christopher Graves, Elizabeth Makrides, and Victor Schmidt are PhD students supported by the Brown University IGERT Training Grant in Reverse Ecology. We would like to thank the numerous faculty members and staff at Brown University and the MBL for their support and assistance on the project. Also, we would like to thank Christoph Schorl and Hilary Hartlaub at the Brown Genomics Core Facility for their assistance.