Integrating Eukaryotic and Prokaryotic Plankton Community Transcriptomics into an Ecological Network Analysis of the Neuse River Estuary

Christian Stackhouse[1], Jamie Browne[1], Nathan Hall[2], Dave Schruth[1], Hans Paerl[2], ModMon Team[2] and Adrian Marchetti[1]

University of North Carolina-Chapel Hill Marine Sciences Department[1], UNC-CH Institute of Marine Sciences[2]

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Sea Grant North Carolina www.ncseagrant.org

Abstract

Eutrophication of estuarine and coastal systems can often result in harmful algal blooms (HABs), hypoxia and fish kills. Non-point sources of pollution in the Neuse River Estuary (NRE) watershed have been increasing as the area has experienced steady growth in agriculture, industry and urbanization. Efforts to reduce nutrient inputs of phosphorus have had positive results; however, similar efforts for nitrogen input reduction have not been as successful. Necessary to the effective management of this ecosystem are the identification of the abiotic and biotic components of the system and subsequent understanding of the relationships of those components.

The aims of this project are to:

a) provide insight into bloom dynamics, causes, and effects, and to provide new molecular tools that may aid in forecasting HABs.

b) build ecological relationships between the plankton communities incorporating transcriptomic (analysis of mRNA sequences) data from eukaryotic phytoplankton, cyanobacteria, and heterotrophic bacteria. These data will be contextualized with environmental data routinely collected by the ModMon monitoring program.

c) elucidate important ecological networks that can then be used to guide best management practices to promote healthy ecosystems that may be expandable to other coastal and estuarine systems.

Data Collection and Processing

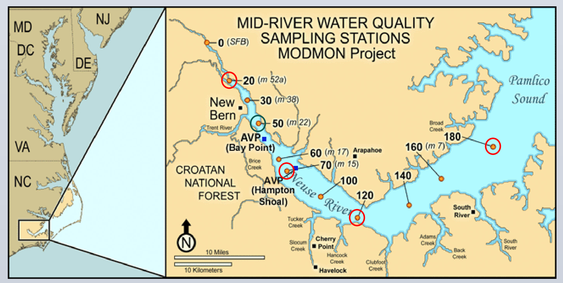

Fig. 1: ModMon Sampling Stations Map
Red circles indicate stations sampled in this study. The green circle marks the area where a large bloom of the dinoflagellate *Gyrodinium* sp. was sampled.

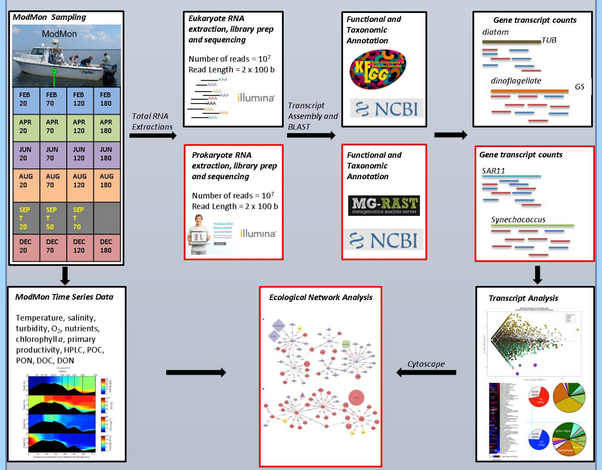

Fig. 2: Data processing and assimilation pipeline
This graphic illustrates how biotic (transcriptomic) and abiotic (environmental data) will come together to enable novel ecological network analysis.

Environmental Data

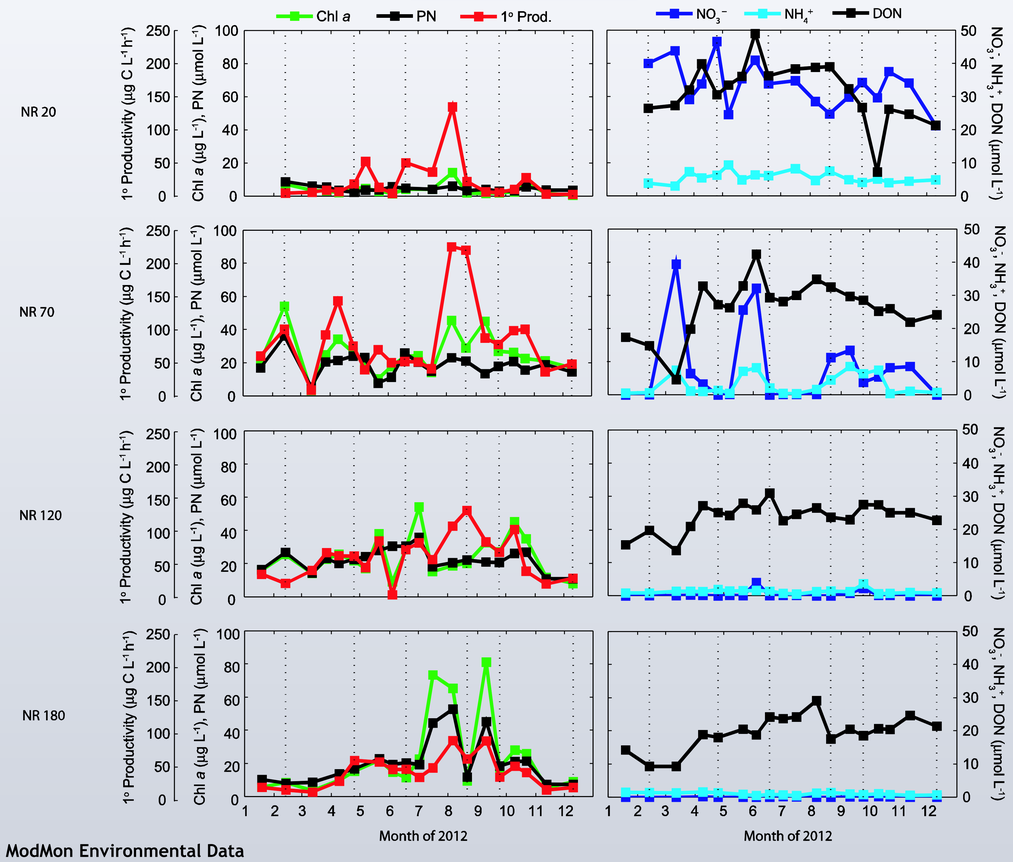

Fig. 3: ModMon Environmental Data
Chlorophyll a (Chl a), particulate nitrogen (PN), primary productivity (1˚ Prod), dissolved nutrient concentrations of nitrate (NO_3), ammonium (NH_4) and dissolved organic nitrogen (DON) at select ModMon in 2012. Dashed lines indicate times where samples for transcriptomic sequencing were collected as part of this study.

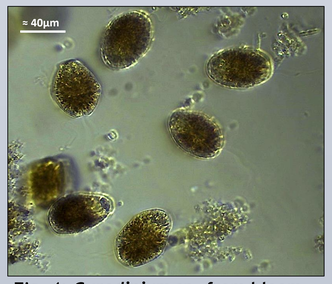

Fig. 4: *Gyrodinium* sp. from bloom

Fig. 5: Filtered cells to be extracted

Table 1: Raw sequencing and BLASTx statistics.

					KEGG		*nR*	
Date	**Station**	**Approximate Size (bp)**	**Number Raw Reads**	**Number of contigs**	**Number hits**	**% Hit**	**Number hits**	**% Hit**
2/13/2012	***20***	7,654,698,200	76,546,982	1,024,616	534,217	52	599,808	59
	70	6,811,459,000	68,114,590	868,114	417,406	48	467,449	54
	120	9,816,328,000	98,163,280	1,146,114	573,897	50	646,617	56
	180	8,454,741,200	84,547,412	1,086,447	514,656	47	553,508	51
4/25/2012	***20***	8,015,339,800	80,153,398	577,328	343,785	60	384,462	67
	70	7,719,591,600	77,195,916	649,712	391,290	60	444,511	68
	120	10,846,515,200	108,465,152	891,591	501,088	56	576,959	65
	180	5,348,416,400	53,484,164	427,353	245,412	57	280,832	66
6/18/2012	***20***	6,749,777,600	67,497,776	629,947	378,639	60	422,709	67
	70	8,868,293,400	88,682,934	759,464	411,182	54	446,179	59
	120	8,313,215,800	83,132,158	739,718	358,325	48	396,550	54
	180	6,207,331,800	62,073,318	116,876	68,346	58	77,843	67
8/21/2012	***20***	10,136,731,800	101,367,318	778,232	462,779	59	514,035	66
	70	8,581,611,200	85,816,112	898,048	452,992	50	545,989	61
	120	12,306,494,600	123,064,946	985,983	526,112	53	575,244	58
	180	7,849,497,800	78,494,978	613,750	315,632	51	351,782	57
9/24/2012	***20***	18,200,401,800	182,004,018	1,624,338	739,161	46	599,026	37
	****Bloom***	11,187,443,000	111,874,430	1,050,048	406,319	39	452,360	43
	70	12,183,958,800	121,839,588	1,168,651	564,192	48	611,538	52
12/10/2012	***20***	6,585,302,800	65,853,028	731,683	371,366	51	411,422	56
	70	9,874,267,800	98,742,678	953,157	431,647	45	488,392	51
	120	12,877,809,400	128,778,094	862,182	413,336	48	480,755	56
	180	6,340,245,800	63,402,458	531,392	261,379	49	298,539	56
	Totals:	**210,929,472,800**	**2,109,294,728**	**19,114,744**	**9,683,158**		**10,626,509**	

Taxonomic Analysis

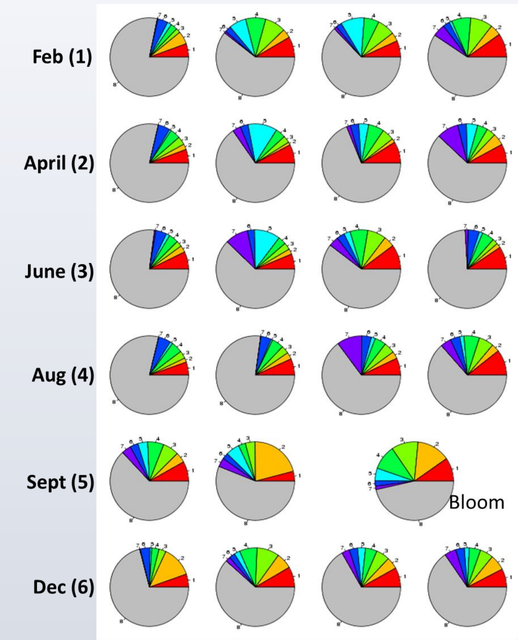

Figure 6: Taxonomic distribution of best hits to Marine Ref Database
Best taxonomic hits (Blastx, e-value $<10^{-3}$) of environmental sequences from each sampling time point (month) and station to the Moran Lab's Marine Ref database. The *Gyrodinium sp.* bloom sample is indicated. *Alexandrium tamarense* (red; dinoflagellate), *Mesodinium pulex* (orange; ciliate), *Scrippsiella trochoidea* (yellow green; dinoflagellate), *Lingulodinium polyedrum* (green; dinoflagellate), *Skeletonema costatum* (cyan; centric diatom), *Tiarina fusus* (blue; ciliate), *Chaetoceros cf. neogracile* (purple; centric diatom), and other (grey).

Conclusions and Future Directions

All sampling has been completed for this project. The analysis of these data has required the development of a novel high throughput analysis pipeline which can be utilized in future metatranscriptomic studies. Nearly 211 gigabases have been sequenced.

Next steps include:

a) A thorough assessment of taxonomic distributions obtained by molecular sequences in comparison to other methods to evaluate phytoplankton community composition (e.g. HPLC pigments)

b) Differential expression (DE) analysis of genes, modules and pathways will be performed for each sample.

c) The DE data, taxonomic data, and environmental data will be joined for each station/time point. We hope that the bloom data will reveal novel correlations between the data to characterize the bloom event.

d) Identification of key genes that can be used as molecular indicators to forecast the presence of HAB species and potential blooms of the species in the NRE.

e) Incorporation of prokaryotic gene sequences in order to perform an ecological network analysis between eukaryotic and prokaryotic plankton in conjunction with environmental data.

Acknowledgements

We are especially grateful to UNC Research Computing for providing all cluster time necessary to perform our sequence analyses.

Contact Information

Marchetti Lab Mailing Address:
3202 Venable Hall
123 South Road
Chapel Hill, NC 27599

phone: (919) 883-9388
email: ctstackh@email.unc.edu
email: amarchet@email.unc.edu