**Potential titles:**

*Emiliania huxleyi* species complex physiological response and strain variation under changing N and P environments

Pan genome of *Emiliania huxleyi* maintains species complex in response to nutrient pulses in oligotrophic ocean

Strain variation facilitates success of cosmopolitan phytoplankton in changing nutrient environments

Evidence of strain heterogeneity and modulation in field populations of cosmopolitan phytoplankton in changing nutrient environments

**Potential authors:**

Harriet Alexander1,2, Mónica Rouco3, Sheean T. Haley3, and Sonya T. Dyhrman3,\*

1 MIT-WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Cambridge, MA 02139, USA

2 Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

3 Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA

*Do we need to add DK and STW?*

*Word limit: 5000 words including M/M*

**Written for submission to ISME Journal**

**Organizing questions:**

What is the role of strain diversity in the success of *Ehux* in changing nutrient environments?

**Abstract:**

**Introduction:**

Central to the global carbon cycle, marine phytoplankton are estimated to constitute nearly half of global primary productivity (Field, Behrenfeld, Randerson, & Falkowski, 1998). Cocolithophores, a group of marine phytoplankton, play a significant role in the carbon cycle. Beyond their significant contributions to primary production (1-10% of total marine carbon fixation), cocolithophores are an important source of calcite (CaCO3­­)to oligotrophic systems, constituting about 50% of calcite deposition to sediments (Poulton, Adey, Balch, & Holligan, 2007). *Emiliania huxleyi* is the most abundant cocolithophore species in the modern ocean, known for its cosmopolitan distribution in the surface ocean and ability to form large blooms in both eutrophic coastal and oligotrophic open ocean regions (Paasche, 2001). The paradigm of bloom dynamics in biological oceanography has diatoms blooming first after a nutrient pulse (Margalef, 1978). Following the termination of the diatom bloom, nutrients such as nitrogen (N), phosphorus (P), and silica (Si) are low, and cocolithophores, namely *E. huxleyi*, thrives and bloom (Litchman, Klausmeier, Miller, Schofield, & Falkowski, 2006). *E. huxleyi* is known to be well adapted to such low-nutrient environments with its ability to scavenge nutrients from organic compounds (Bruhn, LaRoche, & Richardson, 2010; Dyhrman & Palenik, 2003; Palenik & Henson, 1997).

The broad distribution of *E. huxleyi* across many ecological gradients in combination with many laboratory studies, indicate a high degree of physiological variability within this species complex (Iglesias-Rodriguez et al., 2008; Riebesell et al., 2000). Genomic surveys of cultured isolates of *E. huxleyi* have shown a high level of variability amongst the genomes (Read et al., 2013), which mirrored the physiological variability observed in the field and laboratory. Though historically believed to be a single species, the genome of *E. huxleyi*, termed a pan genome, demonstrates great variability across isolated strains. The pan genome was found to consist of a set of core genes, common to all strains surveyed, as well as a variable set of genes distributed amongst strains. Such genomic variability has been described in other cosmopolitan species (Kashtan et al., 2014) and may be central to the success of the species in diverse environmental conditions (Biller, Berube, Lindell, & Chisholm, 2014).

Metabolic plasticity in response to environmental change is a current cornerstone to the study of phytoplankton physiology, with much effort being put towards characterizing transcript- and protein-level shifts following perturbation (Bertrand et al., 2012; Dyhrman et al., 2006, 2012; Frischkorn, Harke, Gobler, & Dyhrman, 2014; Jones et al., 2013; Wurch, Bertrand, Saito, Van Mooy, & Dyhrman, 2011). Different species within a functional group (e.g. diatoms) are known to mount unique metabolic responses to the same environmental conditions (Alexander, Jenkins, Rynearson, & Dyhrman, 2015; Bender, Durkin, Berthiaume, Morales, & Armbrust, 2014). An alternative (or compliment) to such metabolic plasticity may be genomic variability, which enables stable niche partitioning within cosmopolitan populations. Unique transcriptional responses and physiological abilities have also seen at the strain-level in various *Prochloroccous* ecotypes (Thompson, Huang, Saito, & Chisholm, 2011).

Here, we examine the role of metabolic plasticity and strain variability in the stability and success of natural populations of *E. huxleyi* in the North Pacific Subtropical Gyre.

The molecular mechanisms underlying the functional group difference are beginning to be better understood, with recent work in the oligotrophic North Pacific Subtropical Gyre (NPSG) demonstrating diverging transcriptional response between diatoms and haptophytes, which were dominated by *E. huxleyi* (Alexander et al. 2015\*\*).

**Results/Discussion:**

*Emiliania huxleyi* is known to exist as many strains (or ecotypes), which exhibit clear differences cell physiology, morphology, reproduction, and ecology (Paacsh). Such fundamental differences among strains make direct, global comparisons between culture or laboratory studies and field predictions, models, or observations. The recently published pan genome of *E. huxleyi* described the genetic basis for such variability, where there exists a core set of genes across all strains and a variable set not common to all strains (citation). Such gene variability is hypothesized to define niches for different strains (e.g. success in low N, high P, or low light), and consequently may be at the heart of its cosmopolitan nature (citation).

Though relatively well described in culture, the dynamics of these strains have not yet been assessed in the field, particularly in relation to environmental perturbation. Here we employ a novel metatranscriptomic approach to query the presence/absence and success of different strains of *E. huxleyi* in the environment under changing nutrient environments. We hypothesize that rather than a single strain exhibiting great metabolic plasticity in response to perturbation of the nutrient environment, different strains of *E. huxleyi* will come to occupy the newly created niche.

With focus on the large phytoplankton assemblage (> 5.0 um), the total mRNA from the surface mixed layer at Station ALOHA was deeply sequenced six times during the summer of 2012, following a eularian sampling scheme. To perturb the nutrient environment of the community, two microcosm experiments were conducted with natural populations ~2 weeks apart. These experiments were designed to skew the nitrogen (N) : phosphorus (P) ratio of the microcosm, consequently disrupting the nutrient environment. Sequence reads from the *in situ* and experimental treatments were conservatively mapped to a custom database comprised of all publicly available transcriptomes (MMETSP) as well as to a curated set of available *E. huxleyi* transcriptomes and genomes.

*Functional group species level responses to nutrient input different.*

Previous work from Alexander et al. (2015) demonstrated a clear delineation between the diatom and haptophyte functional group response to deep seawater (DSW) addition, reflective of previously defined r- and K-selected lifestyles, respectively. Additionally, it was shown that following DSW addition the rank-order of the species shifted drastically for diatoms, but relatively little for haptophytes, with *E. huxleyi* consistently being the most highly represented taxa. This pattern holds true, not only for DSW amendment, but also for

Figure 1. Diatoms

*Shifts between strain-specific and core metabolism indicate a shift towards a new strain in the field?*

Figure 2. Strain diversity

*Doesn’t really matter where things come from as long as they are changing.*

Figure 3. Differential expression analysis

Figure 4.

Pangenome: everything is everywhere but the environment selects. Shift in diversity following perturbation?

What is making theses species successful? What selection is driving the selection?

N is a particulary strong driver.

Hypothesis: strains should modulate abundance with changing environment?

**Materials and Methods:**

*Sample collection and shipboard nutrient incubation experiments*

Seawater for the *in situ* eukaryote community mRNA analysis was collected at the HOT, Station ALOHA (22˚45’ N, 158˚00’ W) from a depth of 25 m at 1400 hrs (local time)on six occasions during August and September 2012 using a Eulerian sampling scheme as part of the HOE-DYLAN research expedition as per Alexander et al 2015. Water was collected in acid-washed 20-L carboys and approximately 60 L of seawater was prescreened through 200 µm mesh and then filtered onto polycarbonate filters (5.0 µm pore size, 47 mm, Whatman) by way of peristaltic pump. Filters were changed every 20 minutes or when flow rate decreased. Filters were placed in cryovials and stored in liquid nitrogen until mRNA extraction. The total length of filtration time did not exceed 3 hours.

In conjunction with these field-based surveys, two factorial nutrient amendment incubation experiments focused on the macronutrients N and P were performed with natural communities (T0=S1 and S4) (STable 1). Incubations were modeled after a simulated 10% deep seawater (DSW) upwelling as described in Alexander 2015 and designed to tease apart the potential nutritional components of DSW upwelling. Using historical nutrient profile data collected from HOT-DOGS, macronutrient amendments were based off of the 10 year average nutrient concentration at 700 m (below the nutricline). The concentration of iron was modeled after (Marchetti et al., 2012) and vitamin B12 was modeled after (Bertrand et al., 2007). Triplicate 20-L carboys of each treatment were incubated at 30% surface light-levels using on-deck incubators for 7 days and processed as described above, on the final day at 1400 hrs (local time). Chlorophyll *a* was measured on whole water samples collected onto GF/F filters (25 mm, Whatman) using a 90% acetone extraction and assayed by fluorescence using the AquaFluor Turner TD700 (Parsons, Yoshiaki, & Lalli, 1984). Nutrient analyses… blahblah

*RNA Extraction and Sequencing*

RNA was extracted from individual filters with the RNeasy Mini Kit (Qiagen), following a modified version of the yeast protocol. Briefly, lysis buffer and RNA-clean zirconia/silica beads was added to the filter and samples were vortexed for 1 minute, placed on ice for 30 seconds, and then vortexed again for 1 minute. Samples were then processed following the yeast protocol. The resulting RNA was eluted in water and then treated for possible DNA contamination using TURBO DNA-free Kit (Ambion) following the Rigorous DNase protocol. RNA from individual filters was then pooled by sample, using the RNA Cleanup Protocol from the RNeasy Mini Kit (Qiagen). The resulting RNA sample thus represented approximately 60 L of total seawater for the *in situ* sample. Filters were pooled across like triplicate bottles by treatment, totaling 60 L from each of the incubation treatments. The total RNA sample was then enriched for eukaryotic mRNA through a poly-A pull down. The resulting enriched mRNA sample then went through library preparation with the Illumina TruSeq mRNA Prep Kit (Illumina). Libraries were sequenced with the Illumina HiSeq2000 at Columbia Genome Center (New York, NY). Each sample was sequenced to produce a targeted 60 million, 100 base pair, paired end reads. Raw sequence data quality was visualized using FastQC and then cleaned and trimmed using Trimmomatic v 0.27 (paired end mode; 6-base pair wide sliding window for quality below 20; minimum length 25 base pair).

*Database Creation and Mapping*

Transcriptome sequences and annotations generated through the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) that were made public as of 17 March 2014 were collected, representing 401 transcriptomes across 209 species or cultured isolates. Transcriptomes from like species (regardless of strain or condition) and cultured isolates were pooled and clustered using CD-HIT-EST (98% identity; word size of 9). The resulting clustered set of transcripts was considered to be the representative transcriptome for the species or cultured isolate. Transcriptomes were annotated with KEGG Orthology annotations using the bi-directional best hit (BBH) method through the KEGG Automatic Annotation Server (KAAS) (Moriya, Itoh, Okuda, Yoshizawa, & Kanehisa, 2007). The 209 transcriptomes created in this manner were concatenated to form a comprehensive species-level transcriptome database from the MMETSP library. Due to the large size of the resulting MMETSP database, trimmed reads were mapped to the MMETSP using the Burrows-Wheeler Aligner (Li & Durbin, 2010) (BWA-mem, parameters: -k 10 -aM) and then counted using the HTSeq 0.6.1 package (Anders et al., 2014).

Ehux strains from experimetns (provide numbers) were used. Orthologus clustering using orthoMCL (values)

*Differential expression analysis*

Counts obtained from HTSeq were pooled for like-KEGG orthologs across all species in a functional group. The quantitative metabolic fingerprint (QMF) was assessed by normalizing global patterns of expression at the module level to the total mapped reads. PCA and confidence ellipses of the QMF signals by functional group and sample type (*in situ*, no addition control, and DSW addition) were calculated and visualized using FactorMineR package in R (Figure 2, Supplemental Figure 7). No significant difference was seen between the *in situ* and no addition control samples (Supplemental Figure 7). For each functional group, the pooled KEGG counts from the *in situ* samples (S1-S3) were compared to those from the corresponding DSW amendment (E1-E3) using Analysis of Sequence Counts (ASC), an empirical Bayes method(Wu et al., 2010). Genes were considered to be differentially abundant between treatments if for a fold change of 2.0 the posterior probability (post-*p*) was greater than 0.95 (2012). Patterns of differential abundance were visualized using Circos (Krzywinski et al., 2009). Global shifts in the expression of genes independent of functional group were assessed with TMM normalization using the Microbial Assemblage Normalized Transcript Analysis package (MANTA, v. 1.12.0)(Marchetti et al., 2012).

**Tables:**

**STable 1. Nutrient incubation experimental design.**

**Figures:**

Figure 1: Bargraphs (alternative to 3D). Something that shows the consistency in mapping over time?

Figure 2: +N/+P

Figure 3: