Title: Genome-wide mapping of cis-regulatory elements and regulation of nitrate assimilation in *Phaeodactylum tricornutum*

Authors: Sarah R. Smith,[1,2] Hong Zheng,[2] Pratap Venepally,[2] Leo Baumgart,[3] Ronan O'Malley,[3] and **Andrew E. Allen[2,4]*** (aallen@ucsd.edu)

Institutions: [1]Moss Landing Marine Laboratories, San José State University, Moss Landing, CA; [2]J. Craig Venter Institute, La Jolla, CA; [3]Joint Genome Institute, [4]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA; and [4]University of State, City

Website URL:

Project Goals: Short statement of goals. (Limit to 1,000 characters)

Abstract Text: Despite the importance of diatoms in the marine environment, and their relevance for biofuels development, little is known about how diatoms sense and respond to shifts in environmental conditions. Transcription factors (TFs) regulate gene expression programs by binding DNA and promoting (activating) or blocking (repressing) recruitment of RNA polymerase to initiate transcription, and undoubtedly have a role in the response of diatoms to their environment. We used DNA affinity purification sequencing (DAP-Seq), a high-throughput *in vitro* method, to characterize transcription factor binding sites genome-wide. Using this method, we have successfully mapped the transcription factor binding sites for 58 TFs from the *P. tricornutum* genome. Several TF classes failed *in vitro* to generate significant signals, which could be due to incorrect gene models, the lack of true DNA binding domains, or insufficient maturation (lack of heteromers or posttranslational modifications). The most successful classes of TFs in the DAP-seq pipeline were the bZIPs, heat shock factors (HSFs), Myb and MybSHAQKY TFs. Our results corroborate findings from previous investigators for the few transcription factors that have been functional characterized, and greatly expand the catalog of these key molecular components of signal transduction cascades in diatoms. Genome-wide, several bZIPs (Aureochrome1a, bZIP10, bZIP11, bZIP13, bZIP15) were associated with patterns of gene expression driven by diel signatures, pointing to a role for these TFs in regulating the shift between illuminated and dark cell physiology. Fewer TFs were associated with overall macronutrient status. Careful dissection of the architecture of transcription factor binding sites (TFBS) in the promoters of a tightly coordinated set of nitrate assimilation genes that are highly sensitive to nitrate availability (highly nitrate sensitive, or HNS regulon) revealed the existence of TFBS hotspots within different promoters. Analysis of these hotspots suggests that HNS genes are likely regulated by a combination of activation and repression from various TFs, including light and cAMP-sensitive bZIPs, low N induced HSFs, and a homolog of a fungal regulator of the nitrate regulon. This mode of regulation shares little similarity with what is known from cyanobacteria, plants, other algae (including *Chlamydomonas*), and filamentous fungi, and gives insight into the functional significance of the HSFs in diatoms, which have undergone massive evolutionary radiation relative to HSFs in other eukaryotes. Precise

knowledge of mechanisms regulation transcription from the HNS promoters will further our understanding of how diatoms integrate the assimilation of nitrate with other cellular energetic demands during growth. This detailed knowledge is also essential in synthetic biology to most effectively design transgenic lines that use the nitrate reductase (NR) promoter as an inducible promoter.

Funding Statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC00006719 and DE-SC0008593.