

MED4 downregulated cluster motif search: Methods and Results

May 4, 2006

John Aach

Supplemental material for

Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability

Andrew C. Tolonen, John Aach, Debbie Lindell, Zackary I. Johnson, Trent Rector, George M. Church, Sallie W. Chisholm

This document describes the methods used to search for motifs in the upstream sequences of downregulated clusters 7 and 9 for MED4, and the results obtained from this effort. (See main text for information on the clustering.)

The principal purpose of this search was to see if the NtcA motif could be found in upstream regions of repressed clusters, since NtcA has been reported to be a repressor of some genes (see manuscript). However, since the NtcA motif was found upstream of *upregulated* clusters in MED4 only and not MIT9313 (see Supplemental Information document NtcA_motif_reconstructions.pdf), this search in repressed clusters was limited to MED4.

METHODS

Methods for this search were identical to those for the motif search in upregulated clusters as documented in NtcA_motif_reconstructions.pdf, except that

- the upstream regions came from MED4 clusters 7 and 9, and not 1 and 2. Note that clusters 7 and 9 (and not 8 and 9) were the most downregulated clusters at the 12 hour time point (see Figure 3 in the main text).
- *only* MED4 downregulated clusters were analyzed (not MIT9313, as noted above).
- AlignACE was run using columns = 8 and 10, but not 12.

Motif statistics spreadsheet: Motif MAP, specificity, and NtcA ComparACE scores were compiled in MED4.downregclus.motifdata.xls. Specificity, MAP, and CompareACE scores were also ranked from best to worst (best = 1), where rankings were specific to each AlignACE run (so that, e.g., there are two rank 1 specificity scores for MED4, one for each of the AlignACE column 8 and 10 runs).

Motif logos ppt file: Selected motif logs (see Results) were placed in MED4_downregulated_cluster_motif_sequence_logos.ppt.

RESULTS

- AlignACE generated 17 motifs with columns=8, and 12 with columns=10. The best MAP score for any motif was 8.69716 (columns=8). The best specificity score for columns==8 was 4.36e-8 (motif 8), and for columns=10 was 1.97e-9 (motif 6).
- ComparACE scores with the known NtcA site "training set" from Su et.al. (2005) NtcA motif were all poor (maximum 0.338805 for motif 5, columns=8). As NtcA was detected in upregulated genes for MED4, AlignACE should have been able to detect it if it is present upstream of downregulated genes. This suggests that NtcA does not play a major role in directly repressing the most repressed genes under N starvation.
- Sequence logos were generated for the 10 best motifs from the columns=8 AlignACE run.

REFERENCES

- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E., 2004, WebLogo: A sequence logo generator, *Genome Research* 14:1188-90
- Hughes, J.D., Estep, P.W., Tavazoie S., and Church, G.M., 2000, Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*, *J Mol Biol* 296(5):1205-14.
- Roth, F.R., Hughes, J.D., Estep, P.E., and Church, G.M., 1998, Finding DNA Regulatory Motifs within Unaligned Non-Coding Sequences Clustered by Whole-Genome mRNA Quantitation, *Nat Biotechnol* 16(10):939-45.
- Schneider T.D. and Stephens, R.M., 1990, Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res* 18:6097-100
- Su, Z. Oلمان, V., Fenglou, M., Xu, Y., 2005, Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis. *Nucleic Acids Res* 33(16): 5156-71