Computational analysis of the conservation of the $\sigma^E$-ChrR regulon in $\alpha$- and $\gamma$-proteobacteria

Yann Dufour & Timothy Donohue

It has been determined in *Rhodobacter sphaeroides*, an $\alpha$-proteobacterium, that the $\sigma$/anti-$\sigma$ transcription factor pair, $\sigma^E$ and ChrR, mediates the transcriptional response to stress induced by singlet oxygen. It has also been shown that the transcriptional response involves a second level of transcriptional regulation since $rpoH_{II}$, which encodes a second $\sigma^{32}$-like factor in this bacterium, contains a $\sigma^E$-dependent promoter. To characterize and discriminate target genes potentially activated by each of the two alternative sigma factors, we have chosen a computational approach to take advantage of available genomic sequences and microarray gene expression data. A clustering analysis of RNA level patterns was performed on a dataset composed of 67 microarrays. It revealed a cluster of genes from 6 operons with RNA abundance patterns that showed strong correlation with $rpoE(\sigma^E)$ and $rpoH_{II}$ RNA levels and ~40 potential operons forming a second cluster showing RNA level patterns similar to the first cluster. The gene promoter sequences from each cluster were analyzed with sequence motif discovery algorithms to identify putative transcription factor binding sites. A conserved bi-partite motif typical of those used for DNA binding by $\sigma$ factors was found in all the promoter sequences of the first cluster. A position-weighted matrix (PWM) was constructed from the conserved sequences and used to score all promoters to identify additional $\sigma^E$ target genes that may contain multiple promoters. All candidate promoters were experimentally tested to obtain a set of 9 validated $\sigma^E$-dependent promoters. A putative DNA binding motif for RpoH II was also identified in the larger cluster of genes and is currently being experimentally validated. Subsequently, we used the information gathered for the $\sigma^E$ regulon to investigate its conservation across the bacterial kingdom. A group of 75 sequenced bacterial species, composed of both $\alpha$- and $\gamma$-proteobacteria, was found to contain $rpoE$ and $chrR$ homologs. Orthologous groups of genes were constructed across all species using the OrthoMCL program. Promoter sequences of genes in each orthologous group were scored against the $\sigma^E$ binding site PWM to obtain a group specific motif score. This analysis revealed that groups with the highest scores correspond to the $\sigma^E$ target genes that were validated in *R. sphaeroides*. This result indicates that both the binding sequence for $\sigma^E$ and the biological functions represented in its regulon are conserved across groups of $\alpha$- and $\gamma$-proteobacteria, thus, these bacteria may be exposed to stresses of the same nature.