Identifying the role of genome methylation in the bacterium *Bacillus subtilis*

Bacillus subtilis is a Gram-positive soil bacterium that is known to have genome methylation, however the location of these modifications and their biological role remains unknown. The goal of this research is to identify the genome-wide sites for N6 methyladenosine (N6mA) and N4 methylcytosine (N4mC). Current work in the Simmons lab has identified a N6mA methylation pattern in the sequence motif 5`GACG**A**G 3` and a N4mC methylation pattern in the sequence 5`**C**TCGARB 3` in the genome of *Bacillus subtilis.* Using homology searches we identified two genes encoding putative methyltransferases (MTases), *yabB* and *yeeA* [1]. I hypothesize that YabB and YeeA are methyltransferases responsible for the observed methylation in the sequences of interest and that these modifications of genomic DNA function as either part of a restriction modification system or play a role in regulation of transcription [2][3].

The first objective will be to identify and characterize the MTases responsible for the observed methylation patterns. To test the suspected MTases (YabB and YeeA), I have created strains with clean deletions of *yabB* and *yeeA*. Diagnostic PCR was used to confirm loss of *yabB* and *yeeA.* Next Generation Sequencing using the Pacific Biosciences (PacBio) platform will be used to observe any differences in genomic methylation between wild type, *ΔyabB*, and *ΔyeeA* genomic DNA. To confirm differences in genome-wide methylation found using PacBio sequencing, I will treat genomic DNA with methylation-sensitive restriction endonucleases followed by Southern blot analysis with a radiolabeled probe, with the goal of finding different DNA lengths in the respective digests. In *ΔyabB* and *ΔyeeA* cells, I will clone yabB and yeeA for ectopic expression to restore methylation and confirm this restoration with methylation-sensitive restriction endonuclease treatment. I will also use an *in vitro* assay to show that YabB and YeeA are both necessary and sufficient for methylation of these sequence motifs.

In order to elucidate the effects of methylation, I will begin by using computational methods. Using the *B. subtilis* reference genome I will search for regions where the sequences of interest appear. Once I have generated a list of the regions where the sequence appears I will perform factor analysis to find properties of the genome that are correlated with methylation of genomic DNA. A correlation between a sequence of interest and a particular region of the genome may indicate a mechanism for regulation, whereas a random distribution of the sequence may be more indicative of a role in a restriction modification system.

To analyze the effect of the methylated bases on cell physiology I will use the results from our MTase characterization and computational methods to refine my analysis. I will begin by generating growth curves for Δ*yabB*, Δ*yeeA*, and wild type strains to determine if lack of methylation affects growth rate. I will also determine if lack of genome methylation has any effect on mutation rate. The results from the first portion of this proposal will allow us to narrow down potential assays/screens to determine the function of genome methylation in *B. subtilis*.

Overall, our main goals will be to determine and characterize the MTases responsible for methylation of the previously described sequence motifs and to determine the effect of this methylation. Elucidating the function of DNA methylation in *B. subtilis* will provide a foundation for further study of base

modification in related Gram-positive pathogens, which will serve in better understanding mechanisms of gene regulation. *B. subtilis* will serve as a model organism to provide further insight to the consequences of non-sequence alterations of DNA in bacterial genomes.

[1] Turner, Stephen et al. “Direct detection of DNA methylation during single-molecule, real-time sequencing.” *Nature Methods.* (2010): 461-467. Web.

[2] Kozdon, Jennifer et al. “Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle.” *PNAS.* (2013): E4658-E4667. Web.

[3] Wilson, Geoffrey. “Organization of restriction-modification systems.” *Nucleic Acids Research.* (1991): 2539-66. Web.