

## **Identification and classification of *Clostridium difficile***

Pavel Shelyakin

*IITP RAS, f.serval@gmail.com*

We classified promoter motifs *Clostridium difficile* 630 based on experimentally determined positions of the transcription start sites (TSS) and data on gene expression changes in bacteria with knocked out sigma factor genes.

Positional weight matrices (PWM) were constructed using experimental data about gene regulated by specific sigma factor, or compilation of promoters from *Bacillus subtilis*. PWM scores for each TSS were then converted to quantiles and the most likely sigma factor was assigned to each TSS. The same procedure was applied to related *Clostridium* genomes, and conserved promoters were used to update the PWMs.

As a strict criterion, a sigma factor is assigned to a promoter if it had the highest quantile score and at least two orthologs had candidate promoters scoring above the threshold set to lowest observed score of experimentally validated promoters. At that, 25, 14, 25 and 11 promoters were assigned to sigma factors E, F, G and K, respectively. Among them 17, 5, 5 and 9 genes were experimentally validated.