Reconstruction of transcription control networks in Mollicutes by high-throughput identification of promoters

Gleb Fisunov, Irina Garanina, Daria Evsyutina, Tatiana Semashko, Vadim Govorn

Research and Clinical Centre of Physical-Chemical Medicine, Malaya Pirogovskaya 1a, Moscow 119992, Moscow, Russia, herr.romanoff@gmail.com

Bacteria of class Mollicutes feature extremely reduced genomes, yet are capable of self-replication on an artificial medium. Mollicutes and in particular mycoplasmas are extensively studied as objects for systems biology including the synthesis of artificial genome by Venter et al [1]. Mycoplasmas as the most reduced group of Mollicutes are believed to have extremely reduced gene expression control systems along with the general genome reduction. Here we used a combination of omics and computational methods to reveal transcription control network in a set of Mollicutes representatives: *Acholeplasma laidlawii, Spiroplasma melliferum* and *Mycoplasma gallisepticum*. In this set A. laidlawii represent a species with the richest repertoire of 52 transcription factors (TF's), M. gallisepticum is the most reduced species with 10 TF's and S. melliferum occupies middle position with 23 TF's. In this study we were focused predominantly on *M. gallisepticum* as the most reduced species and used other for comparative studies.

First we obtained a single-nucleotide quantitative map of transcription start sites for three species using 5'-Enriched RNA Sequencing [2]. The obtained data allows to identify both: promoter position with single-nucleotide resolution and its activity in quantitative fashion. Then we constructed a pangenome of conserved transcription factors for 50 species of Mollicutes. We used a combination of experimental data on promoters of reference species with the conservation analysis of respective promoters and TF's across the Mollicutes to identify binding sites of respective TF's. As a result we identified 34 putative binding sites for *A. laidlawii*, 12 for *S. melliferum* and 5 for *M. gallisepticum* [3]. The functional analysis of these TF's and respective regulons demonstrates that they predominantly regulate various parts of metabolic network. The reduction of metabolic capacities from Acholeplasmas to Mycoplasmas corroborates well with the reduction of respective regulators. The activity of 3 TF's in *M. gallisepticum* was proved experimentally: HrcA [4], MraZ [5], HsdC [6].

Second we used quantitative data on promoters obtained by 5'-ERS method to build a quantitative model that accurately predicts promoters' power on the basis of its sequence. This data on promoter power (theoretical prediction) in a combination with the data on promoters' activity (experimental data) was used to identify promoters which activity is deviant from their power. Those promoters were considered to be exposed to a regulatory impact from some kind of transcription control systems. We used heat stress model (5'-ERS data) of *A. laidlawii* to demonstrate that the activity of promoters of heat stress regulon (the sole regulon with completely known function) becomes equal to their power in heat stress but demonstrates repressed state under the normal growth. Further we compared our prediction of TF's binding sites with the activity and power of respective promoters and identified that almost all of them repress respective promoters. Their repressor function corroborates with the position of respective binding sites within promoters: they overlap with transcription start site or locate within spacer between -10 and -35 elements, e.g. they impede RNA-polymerase binding if bound to promoter. Our model predicts about 60 promoters for *M. gallisepticum* that are affected by repressors, which is drastically higher than 10 TF identified by conservation. Since that the almost all of the respective promoters lack common conserved motifs except promoter per se we propose that there is a significant amount of unknown TS's even in such a reduced bacterium as *M. gallisepticum*.

References

1. C.A. Hutchison et al (2016) Design and synthesis of a minimal bacterial genome, *Science,* **351**(6280)**:**aad6253.

2. P.V. Mazin et al. (2014) Transcriptome analysis reveals novel regulatory mechanisms in a genome-reduced bacterium, *NAR,* **42:**13254–68.

3. G.Y. Fisunov (2016) Reconstruction of Transcription Control Networks in Mollicutes by High-Throughput Identification of Promoters, *Front. Microbiol.,* **7:**1977.

4. L. Chang (2008) Mycoplasmas regulate the expression of heat-shock protein genes through CIRCE-HrcA interactions, *Biochem. Biophys. Res. Commun.,* **367:**213–8

5. G.Y. Fisunov (2016) Binding site of MraZ transcription factor in Mollicutes, *Biochimie*, **125:**59–65.

6. G.Y. Fisunov (2017) Binding site of restriction-modification system controller protein in Mollicutes, *BMC Microbiol.*, **17:**26.