Analysis of Transcriptional Regulation of *Bradyrhizobium japonicum* Based on dRNA-seq Data

J. Chuklina[1, 2]

chuklina.jelena@gmail.com

N. Lyubimov[3], E. Evguenieva-Hackenberg[4], M.S. Gelfand[2, 3]

[1]*Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, Russia*

[2]*Institute for Information Transmission Problems RAS, 19 Bolshoy Karetny per., Moscow, Russia,*

[3]*Moscow State University, GSP-1, Leninskie Gory, Moscow, Russia*

[4]*Justus-Liebig-University of Giessen, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany*

The analysis of transcriptome can provide important insights on gene functions and regulatory network of the organism. In addition to the comparison of expression in varying environment, dRNA-seq allows one to map transcription start sites (TSSes) in bacteria, using terminal exonuclease (TEX) to degrade transcripts with processed 5'-ends (as described in [1]). Knowing TSSes also allows for detection of promoter motifs, one of the key elements in controlling differential expression. Here we present the transcriptome analysis of nitrogen fixing α-protebacterium, soy symbiont *Bradyrhizobium japonicum USDA 110* with RNA taken from soy root nodules and free-living state (see [2] for experimental details).

To analyze the dRNA-seq data, we have developed a software package "Transcription Start Site Finder" (TSSF) that detects TSSes and promoters and is complemented with publicly available algorithms for the identification of terminators [3-5]. This allows for a combination of an automatic search and expert analysis using a convenient interface.

To detect putative TSSes, we have computed a salience function (convolution with the Haar wavelet) that indicates where a sharp jump in read coverage occurs. To be scored as a TSS, a peaks should have the same coordinate in TEX-treated and non-treated library, be enriched in the former one, have expression above the noise level and satisfy several other empirical characteristics. These features were taken as input data for a machine-learning classification by SVM, which used a subset of manually reviewed TSSes as a training set. To date, 10071

putative TSSes were detected in 9Mb genome of *B.japonicum*, 6875 TSSes corresponded to putative non-coding RNAs, and only 2744 of 8371 protein-coding genes possessed at least one predicted TSS.

To analyze transcriptional regulation, we developed a tool for *de novo* promoter prediction. It detects positionally overrepresented pairs of oligonucleotides, as sigma-factors σ^{70} and σ^{54}, predominantly used under experimental conditions tested, are known to recognize a motif comprised of two boxes. Having determined the most likely positions of the motif, we constructed positional-weight matrices (PWM) that were then used to score candidate promoters and select the associated sigma-factor.

The dRNA-seq data analysis demonstrates an important role of small regulatory RNAs, both *trans*- and *cis*-encoded (such as 3'-anti-sense RNAs abundant in *Bradyrhizobia*-related species). It also is a powerful method to detect transcripts enriched in nodules thus possibly involved in symbiotic nitrogen fixation.

References

[1] Sharma C. M., Hoffmann S., Darfeuille F., Reignier J., Findeiss S., Sittka A., Chabas S., et al. (2010). The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature*, *464*(7286), 250–5.

[2] Madhugiri R., Pessi G., Voss B., Hahn J., Sharma C.M, Reinhardt R., Vogel J., Hess W.R., Fischer H.-M., Evguenieva-Hackenberg E. (2012) Small RNAs of the Bradyrhizobium/Rhodopseudomonas lineage and their analysis. *RNA Biology* 9 (1): 47-58.

[3] Mitra A., Kesarwani, Anil K., Pal D., Nagaraja V. (2010) WebGeSTer DB--a transcription terminator database. *Nucleic acids research*.. 39 (Database issue), D129–35.

[4] Naville M., Ghuillot-Gaudeffroy A., Marchais A., Gautheret D. (2011) ARNold: A web tool for the prediction of Rho-independent transcription terminators. *RNA Biology*.. 8(1), 11–13.

[5] Gardner P.P., Barquist L., Bateman A., Nawrocki E. P., Weinberg Z. (2011) RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic acids research*. 39(14), 5845–5852.