doi 10.18699/bgrs2024-8.4-19

Reconstruction of regulons for *Methylotuvimicrobium alcaliphilum 20Z^R* based on an analysis of a bunch of transcriptomic data

Sokolova T.S.[1]*, Kolmykov S.K.[1], Kulyashov M.A.[1], Khlebodarova T.M.[1], Kalyuzhnaya M.G.[2], Akberdin I.R.[1]

[1] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

[2] *Biology Department and Viral Information Institute, San Diego State University, San Diego, USA*

** lynxl@list.ru*

Key words: RNA-seq; transcription factor; regulon reconstruction; methanotroph; *Methylotuvimicrobium alcaliphilum 20Z^R*

Motivation and Aim: The study of methanotrophic bacteria represents a “hot” research field in microbial biotechnology, as these organisms have the potential to utilize methane as a carbon source for the production of useful target and value-added compounds. Complex interactions between transcription factors (TFs) and target genes are responsible for various molecular and phenotypic features of bacteria, including adaptation to changes in environmental and/or cultivation conditions. Despite the long-term history of methanotrophs’ investigation, both quantitative/qualitative data and knowledge of such interactions are very limited. However, the assumption that regulatory interactions between TFs and target genes are conservative among bacteria allows us to use the accumulated omics data to reconstruct the regulatory structure in closely related species. Herein, we employ the data on expression profiles for two closely related methanotrophs, *Methylotuvimicrobium alcaliphilum 20Z^R* and *Methylotuvimicrobium buryatense 5GB1*, to construct the first version of regulons in methane-oxidizing bacteria. This finding sheds light on transcriptional modules regulating molecular machinery in methanotrophs at the first time, to the best of our knowledge. Furthermore, the outcome of this study can be essential to pave the way in the fruitful development of biotechnological applications of methane-oxidizing microorganisms.

Methods and Algorithms: Transcriptomic datasets, both unpublished original and previously published [GSE51145, GSE101981, GSE253414, GSE125909, GSE221011] were used for the regulons structure. These datasets were obtained under growth conditions, when methane or methanol are used as a carbon source, and metal ions, including copper, tungsten, calcium, and lanthanum are harnessed as potential regulators of enzymes’ activities involved in the metabolism of these substrates. Gene co-expression networks (GCNs) for *M. alcaliphilum 20Z^R* and *M. buryatense 5GB1* were built using the GENIE3 algorithm [1]. The obtained sets of top gene pairs were divided into 30 clusters using the ward.D2 hierarchical clustering algorithm. Functional enrichment analysis of the clusters was carried out using homebrew R-script. The Operon-mapper web service was used to predict the operon structure of *M. alcaliphilum 20Z^R* [2]. To identify potential regulators for *M. alcaliphilum 20Z^R* we use databases with computationally predicted TFs: deepTF [3], P2TF [4] and ENTRAF DB [5]. TFBS

enrichment analysis for differentially expressed clusters was performed using the MEME suite [6].

Results: The resulting GCN, containing 30 clusters, consisted of 1,241 genes, 78 of which were putative TFs supported by at least one of the sources. To functionally annotate the resulting clusters we conducted enrichment analysis based on KEGG pathways, COG categories and GO terms. Half of the clusters have more than 1 enriched KEGG pathway and 7 clusters do not show any pathway enrichment. 24 clusters demonstrate enrichment in at least one COG functional category, 11 of which are related to only one category. Only 2 clusters show enrichment in one GO term, while 3 clusters do not show any significant enrichment of the GO terms. Thus, the remaining 25 clusters have more than one enriched GO term. We analyzed gene expression levels in the clusters for 3 conditions: CH_4 limitation, O_2 limitation and Ca-La switch. Based on the average gene expression level in the cluster, we divided the clusters into two groups: up and down regulated; the clusters that did not show significant changes in average gene expression levels were ignored. As a result, several known co-regulated groups of genes have been identified, such as the *mxaF* and *mxaI* genes encoding Ca-dependent methanol dehydrogenase, and related genes *mxaR* and *mxaJ*, which have reduced expression levels in the presence of lanthanides [7]. Several new clusters of genes controlled by one or more potential TFs have also been discovered. As an example, we observed that two co-expressed clusters (clusters 14 and 24) had significantly reduced expression under methane- and oxygen-limited growth conditions. Interestingly, these clusters could be combined into one cluster by two potential TFs, MEALZ_RS12170 (feoC) and MEALZ_RS18500 (Fig. 1).

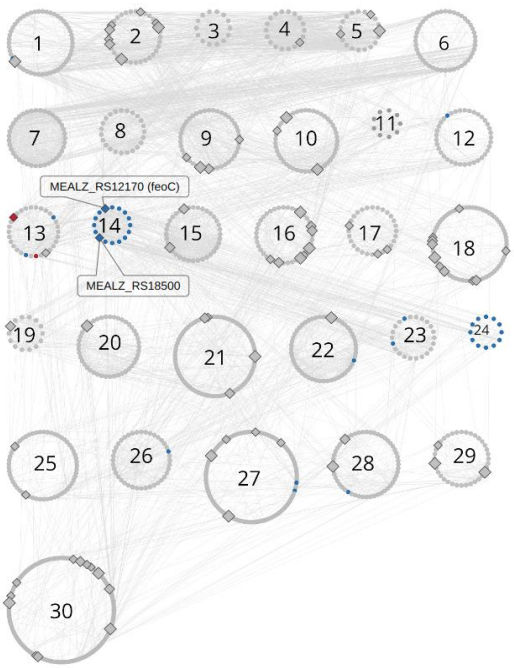

Fig. 1. Down-expressed clusters 14 and 24 in low CH_4 growth condition. Blue color of both circles and diamonds in a cluster means down-expressed genes and potential TFs (MEALZ_RS12170 (feoC) and MEALZ_RS18500) correspondingly. Genes in the cluster are associated with the iron metabolism

We hypothesize that these genes encode true TFs whose expression and genes in their operons are associated with the iron metabolism or reduction in response to methane and/or oxygen limitation. In addition to the clusters identified and described above, a cluster of genes mainly related to nitrogen metabolism transport and regulation has been shown to have reduced expression levels under hypoxia conditions. Their only co-expressed potential TF, CRP, is a regulator of carbon source utilization gene expression in the absence of glucose. It also regulates the expression of GlnG, a known TF mainly regulating nitrogen transport and metabolism in the bacterial cell. In order to confirm the potential regulatory role of revealed transcription factors we performed a search of regulatory signals (motifs) in the upstream of promoters for genes highly co-expressed with the potential TF in a bunch of analyzed transcriptomic profiles. Furthermore, we alternatively explore the regulatory motifs via an interrogation of orthologous genes encoding those genes in closely related methanotrophic strains.

The source code and all data for the whole analysis are available on a GitHub repository https://github.com/Lynxlazy/20ZR-Regulons.

Funding: The study was financially supported by the Russian Science Foundation (project No. 23-24-00606, https://rscf.ru/en/project/23-24-00606/).

References

1. Huynh-Thu V.A. et al. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. 2010;5(9):e12776. doi 10.1371/journal.pone.0012776
2. Taboada B. et al. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*. 2018;34(23):4118-4120. doi 10.1093/bioinformatics/bty496
3. Bao X.R. et al. DeepTF: Accurate prediction of transcription factor binding sites by combining multi-scale convolution and long short-term memory neural network. In: Intelligence Science and Big Data Engineering. Big Data and Machine Learning. IScIDE 2019. Lecture Notes in Computer Science. Vol. 11936. Springer, 2019;126-138. doi 10.1007/978-3-030-36204-1_10
4. Ortet P. et al. P2TF: a comprehensive resource for analysis of prokaryotic transcription factors. *BMC Genomics.* 2012;13:628. doi 10.1186/1471-2164-13-628
5. Ledesma L. et al. Prediction of DNA-binding transcription factors in bacteria and archaea genomes. In: Prokaryotic Gene Regulation. Methods in Molecular Biology. Vol. 2516. New York: Humana, 2022;103-112. doi 10.1007/978-1-0716-2413-5_7
6. Bailey T.L., Johnson J., Grant C.E., Noble W.S. The MEME Suite. *Nucleic Acids Res*. 2015;43(W1):W39-W49. doi 10.1093/nar/gkv416
7. Akberdin I.R. Rare earth elements alter redox balance in *Methylomicrobium alcaliphilum 20Z^{R}*. *Front Microbiol*. 2018;9:2735