**Promoter library analysis and sequence prediction with Jupyter-notebook assisted random forest regression**

Ulf W. Liebal1†, Sebastian Köbbing1, Lars M. Blank1

1 Institute of Applied Microbiology-iAMB, Aachen Biology and Biotechnology-ABBt, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany

KEYWORDS: Biotechnology, Synthetic Biology, Promoters, Pseudomonas, Machine Learning, Random Forest, Python

**ABSTRACT**

In metabolic engineering heterologous enzymes are expressed to introduce new metabolic pathways to micro-organisms. Tight control of gene expression is beneficial to fine-tune enzyme activity, possibly across multiple hosts to either identify the optimal host or to separate cloning and production host. Among the main determinants of heterologous gene expression is the promoter region. However, detailed knowledge of the translation of a promoter sequence to an expression strength is missing. Here we present a Jupyter notebook workflow that uses a promoter library from *Pseudomonas putida* and trains a random forest regressor. The random forest regressor allows for the expression prediction of thousands of novel promoter sequences that cover a broad range of expression profiles while minimizing genetic adjustments to a reference sequence. The associated statistical analyses identify the sequence exploration space of the promoter library within which reasonable predictions are possible. The random forest regression extracts a sequence-logo like results representing the position-nucleotide impact on the prediction and reproduces the outstanding effects of -35 and -10 promoter domains. The workflow identifies novel promoter sequences with defined activities thereby enlarging the promoter toolbox for controlled gene expression. Moreover, the notebook is flexible to also analyze multiple promoter libraries across multiple species to predict defined cross-host activity. It can be adapted to investigate any sequence library to identify sequence factors that determine targeted quantitative outcomes.