

PhyMet²: a comprehensive database for predicting culturing conditions of methanogens

Michał Burdukiewicz¹, Przemysław Gagat¹, Sławomir Jabłoński², Jarosław Chilimoniuk^{1*}, Michał Gaworski²,
Paweł Mackiewicz¹ and Marcin Łukaszewicz²
*jaroslaw.chilimoniuk@gmail.com

¹University of Wrocław, Department of Genomics, ²University of Wrocław, Department of Biotransformation

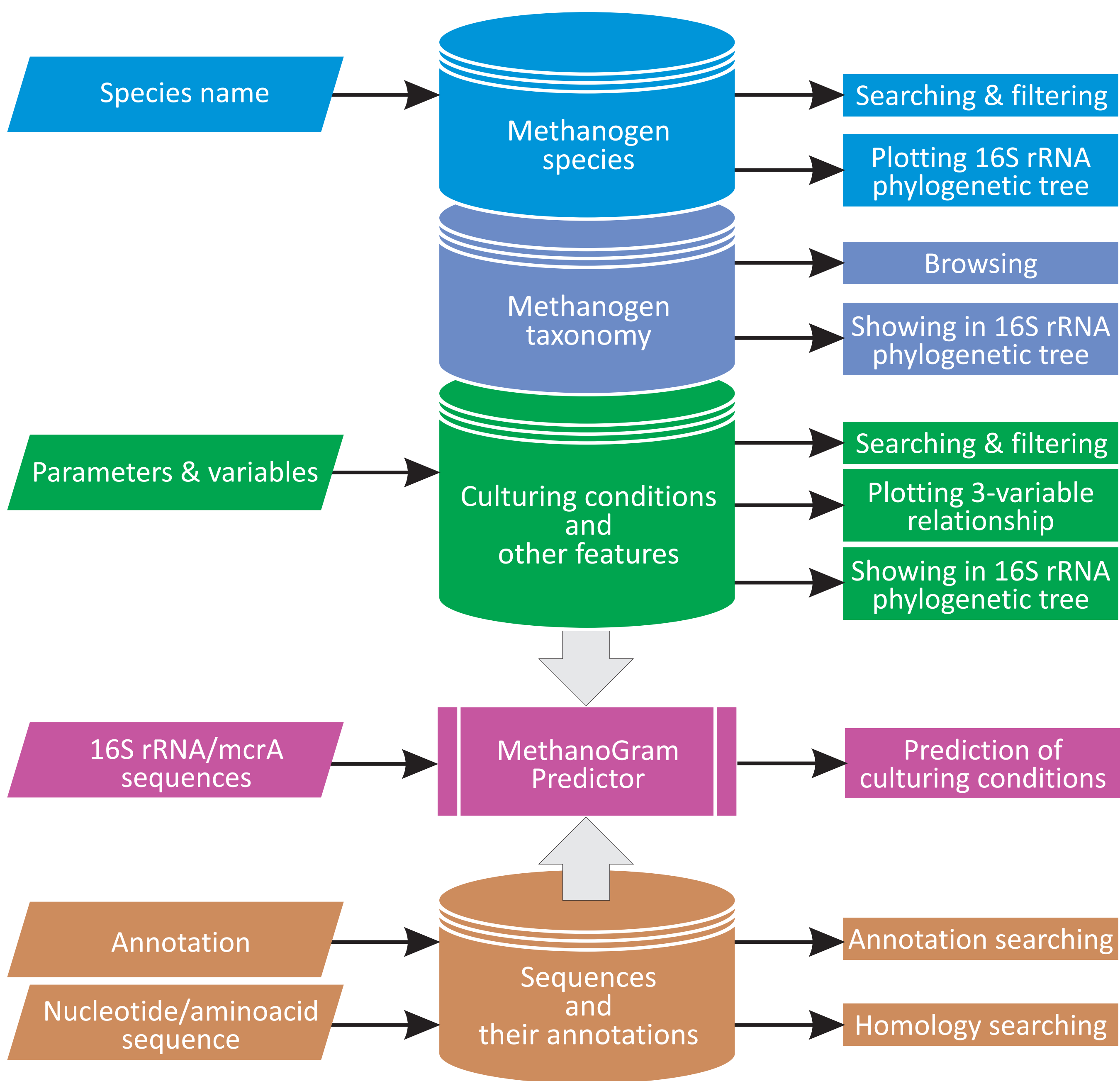
Introduction

Our rudimentary knowledge of microbes stems from difficulties concerning their isolation and culture in laboratory conditions, which is necessary for describing their phenotype, among other things, for biotechnological purposes. An important component of the understudied ecosystems are methanogens, archaea that producing methane, a potent greenhouse-effect gas methane. Therefore, we created PhyMet2, the first database that combines descriptions of methanogens and their culturing conditions with genetic information. The database contains a set of utilities that facilitate interactive data browsing, data comparison, phylogeny exploration, and searching for sequence homologues. The most unique feature of the database is the web server MethanoGram, which can be used to significantly reduce the time and cost of searching for the optimal culturing conditions of methanogens by predicting them based on 16S RNA sequences.

Data collection

PhyMet2 contains 153 manually curated and up-to-date high quality records of methanogenic species. Sequence data was collected from the NCBI (www.ncbi.nlm.nih.gov) and Silva (www.arb-silva.de) databases, and additional information was obtained by thorough manual search of literature.

PhyMet² as multifunctional platform



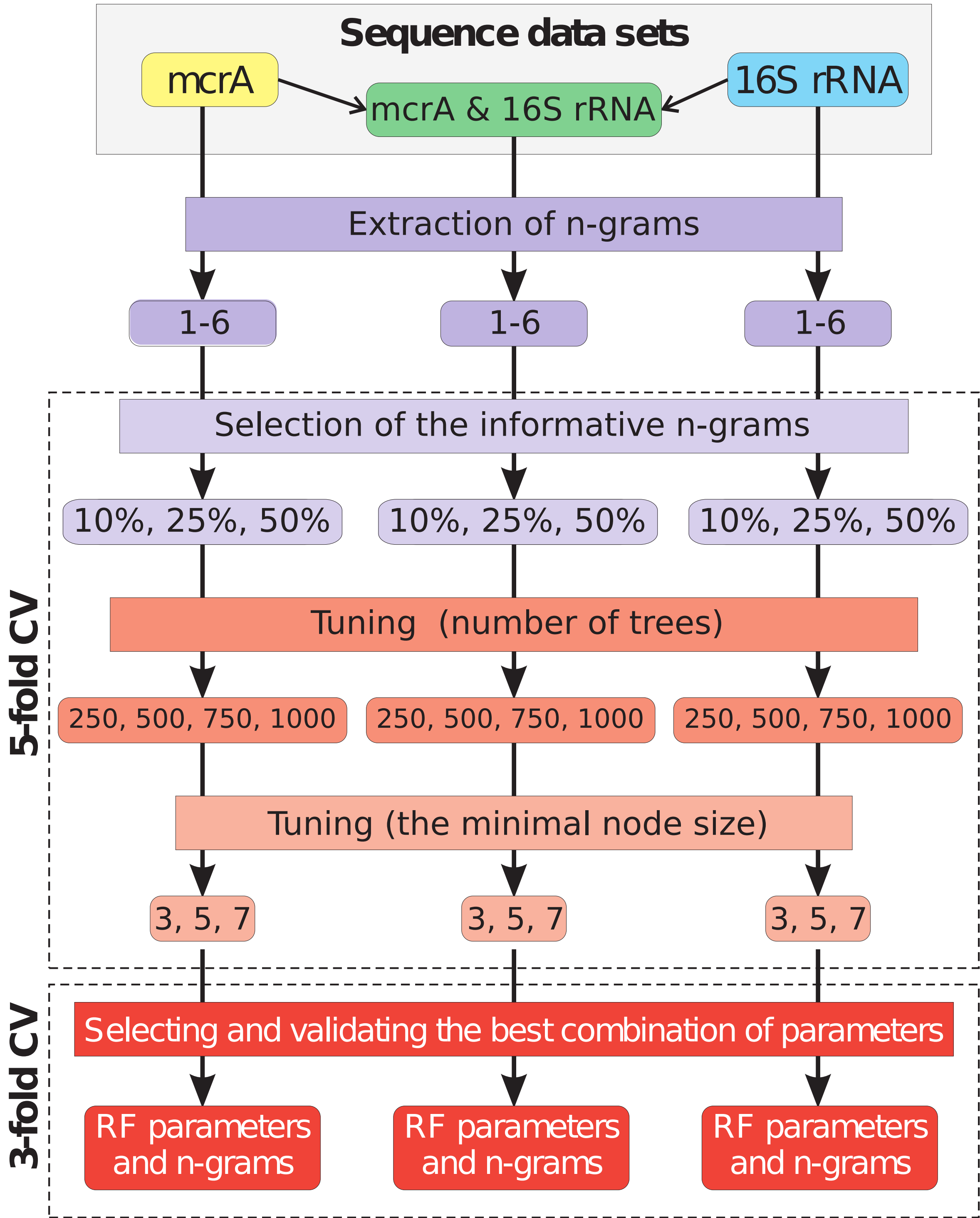
PhyMet² (Phylogeny and Metabolism of Methanogens) is the largest data analysis platform that provides information on culturing conditions and sequence data for methanogenic archaea with a user-friendly interface. The analyses include advanced data browsing, exploring phylogeny, plotting selected features, searching for potential sequence homologues and predicting key culturing conditions for newly discovered methanogens based on 16S rRNA sequences.

MethanoGram

The unique feature of PhyMet2 is a web server, MethanoGram, that predicts conditions for the optimal growth of methanogens: temperature, pH, NaCl concentration, and growth doubling time. MethanoGram is one of the first approaches aiming at predicting the phenotype of microorganisms based on molecular markers, and hopefully will boost further research in the field. We would also like to apply our algorithm to predict culturing conditions for other microorganisms.

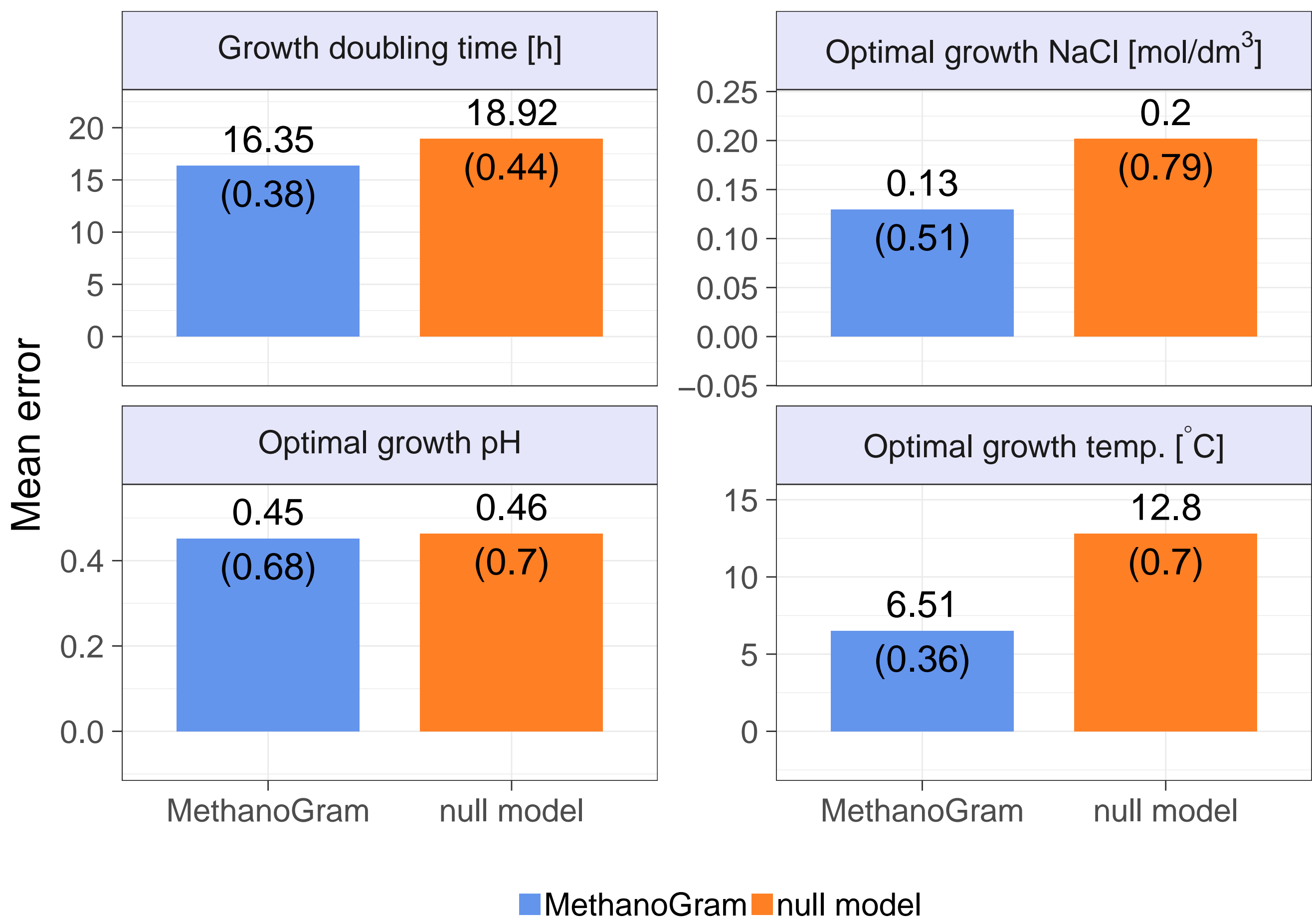
Tuning of MethanoGram

In order to train MethanoGram, we used n-grams, i.e. subsequences of the length n that were extracted from 16S rRNA and mcrA. We chose only those species that have known 16S rRNA, mcrA sequences and all important culturing conditions.



To estimate the culturing conditions we implemented random forests algorithm and performed nested cross-validation.

Mean error



Null model does not incorporate any sequence-based information. Values above columns represent mean error of algorithm, below show mean errors of the best predictors found in the nested cross-validation.

Funding and aviability

PhyMet² is available at: <http://metanogen.biotech.uni.wroc.pl/>.
MethanoGram is available as a web-server: <http://www.smorfland.uni.wroc.pl/shiny/mgp/>.
This work was supported by the Leading National Research Center (KNOW) and the National Science Centre grant no. 2015/17/N/NZ2/01845 and 2017/24/T/NZ2/00003.

Bibliography

To appear in Environmental Microbiology Reports:
Burdukiewicz, M., Gagat, P., Jabłoński, S., Chilimoniuk, J., Gaworski, M., Mackiewicz, P., and Łukaszewicz, M.. PhyMet2: a database and toolkit for phylogenetic and metabolic analyses of methanogens.