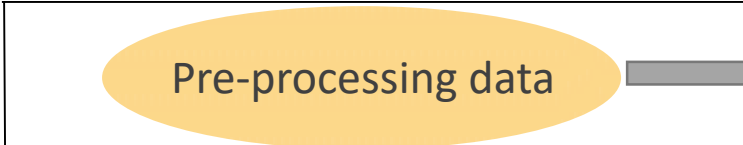


Motivation

Currently, biotechnology research is focused on applying microorganisms in the biorefinery. Such research relies heavily on computational bioinformatics analysis using genomic data. The *Clostridium beijerinckii* NRRL B-598 is a bacterium that is known as a producer of butanol. Its wider utilization can lead to a revolutionary solution for ecology. Unfortunately, we are missing its comprehensive functional annotation that would help to establish an economically viable process.

Overview of experiment



SCHWARZEROVÁ, J. Reproducible analytical pipeline for using raw RNA-Seq data from non-model organisms. In Proceedings of the 26th Conference STUDENT EEICT 2019. Brno: 2020. s. 225-228. ISBN: 978-80-214-5867-3.

Firstly, the analysis pipeline for pre-processed raw RNA-Seq was applied to data of *Clostridium beijerinckii* NRRL B-598. The whole applicable pipeline is available from Github: [JanaSchwarzerova/Analytical-pipeline-rawRNA-Seq](https://github.com/JanaSchwarzerova/Analytical-pipeline-rawRNA-Seq)

Co-expression networks

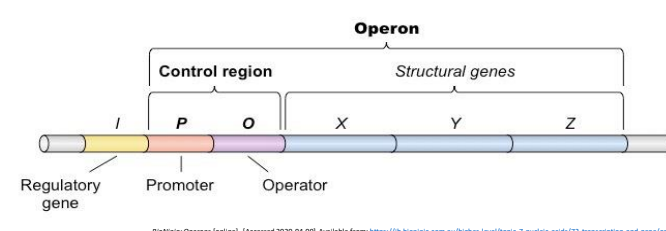
The first co-expression networks for *C. beijerinckii* NRRL B-598 were created by algorithms which are based on Pearson's correlation coefficient and mutual information.



The gene regulatory networks were created by algorithms that are based on bootstrap, random forest, and differential equation analysis. The final result was inferred using unification gene regulatory networks that were created by different approaches and penetration of gene regulatory networks which were created from different datasets using Cytoscape platform.

Operon inference

The inference of operons dataset was obtained by the use of the novel pipeline. The combination of machine learning approach which is included in the online tool Operon-mapper and expression information from the RNA-Seq data of *C. beijerinckii* NRRL B-598. 2 737 operons were predicted.



Conclusion

- ✓ The main accomplished aims of the thesis:
- ✓ Toolbox for pre-processing raw RNA-Seq
- ✓ The innovated pipeline and toolbox for infer operon dataset
- ✓ Inference of gene co-express networks
- ✓ Inference of gene regulatory networks
- ✓ Created the first gene regulation network for *C. beijerinckii* NRRL B-598 with 8787 edges (gene regulations).

