GENE ONTOLOGY ANNOTATION OF CLOSTRIDIUM BEIJERINCKII NRRL B-598 GENOME

Sedlar K.¹, Gruber M.², Branska B.³, Csaba G.², Zimmer R.², Patakova P.³, Provaznik I.¹

¹Department of Biomedical Engineering, Brno University of Technology, Brno, Czechia
²Institut für Informatik, Ludwig-Maximilians-Universität München, Munich, Germany
³Department of Biotechnology, University of Chemistry and Technology Prague, Prague, Czechia

Modern research of microbial organisms relies heavily on publicly available genomic data. Although public databases contain currently more than 13,000 of complete bacterial genomes, the majority of them is not well studied regarding the potential functional capabilities of their genomes. Unfortunately, this applies also to solventogenic bacteria from the genus Clostridium, capable of biofuels production, and complicates their full understanding. Here, we present an augmented genome sequence of the strain C. beijerinckii NRRL B-598, a promising butanol producer. By genome resequencing using highly accurate Illumina platform, we were able to introduce 136 changes into the genome assembly and improve the annotation of the genome. The current genome sequence consist of 5,128 protein-coding genes, 166 pseudogenes, and 148 genes for RNA. Additionally, we decided to study potential function of the discovered genomic elements by Gene Ontology (GO) annotation. We searched UniProt, InterPro, Gene Ontology Consortium (GOC), and RNAcentral databases and found 22,013 annotations assigned to 3,917 distinct genomic elements. We removed duplicated GO term assignments and obsolete terms. We took remaining genomic elements without any assigned GO term and used them for sequence-based annotation in InterPro and GO databases. Finally, we obtained 1,702 distinct GO terms assigned to 4,455 genomic elements in 18,020 unique assignments. In combination with RNA-Seq data, we performed GO enrichment analysis to reveal biological processes, molecular functions, and cellular component that are statistically significant in the life cycle of C. beijerinckii NRRL B-598. Not only can this analysis help to understand the full functional potential of our strain, the methodology can be easily used for analyses of other solvent producing clostridia and help to understand the process of bacterial biorefinery in general.

This work has been supported by grant project GACR 17-00551S.