CFFT19-05	AI and Big Data projects

Analysis of fermentation data from ethanol strain 7442 production in Baltimore plant

Peeter Meos[1], Ljubov Käes[1], Tanel Peet[1], Ott Kekišev[1], Jean Blouin, Michel Dextraze, Zhigen Zhang and Ildar Nisamedtinov

[1] AS Proekspert, Tallinn, Estonia

Economic Justification

The quality of yeast may depend on various fermentation parameters, both uncontrollable (e.g. molasses) and controllable (pH, DO, T, substrate feeding rates etc). The modern cultivation software allow to collect and record in short time intervals a relatively wide range of digital data, which can further be processed by statistical analysis for identification of the root causes leading to quality deviations and perhaps even online prediction of the yeast quality. The avoided yeast quality problems due to better process analysis will allow to save hundreds of thousands dollars every year. In this pilot project we aim to study feasibility of such Big Fermentation Data analysis. The case study is LBDS' TransFerm ethanol yeast (7442), which from time to time has poor performance (glucose tailing and lower ethanol concentration). This defect is related to genetic mutation of *IRA2* gene for still unknown environmental factors.

Executive summary

The purpose of this study is to investigate and provide possible insight into the causes of undesirable mutations in gene *IRA2* in ethanol yeast strain 7442, leading to poor fermentation performance. Lallemand has established a hypothesis that the mutations could be caused by an unknown and unfavorable combination of parameters of the production processes. This analysis, while limited in scope, takes a first look at the collected data. The study looks at the quality and usability of the data, performs feature reduction and takes first steps in modelling and predicting the mutation. Additionally, in the end we provide recommendations how to make the data collection, processing and analysis more automated, holistic and less error prone as opposed to current partly manual data collection. As a result, we have concluded that process step F2 has some irregularities that could be interesting for further research. Nevertheless, the data is too sparse to support robust claims that the differences in F2 between \rejected" and \released" batches are indeed statistically significant. However, since some patterns do exist, F2 step should be studied further.