Metagenome Mining of Novel Enzymes for the Bioethanol Industry

Lignocellulosic biomass is considered an attractive and immediate source for liquid biofuel. The hydrolysis of lignocellulose to soluble sugars requires the synergistic actions of several glycoside hydrolases, and thus, novel enzymes with enhanced activity are sought in order to improve the economics of the process. In the framework of this research, metagenomics and biochemical approaches were combined to identify and characterize novel enzymatic systems geared for lignocellulose degradation.

The known repertoire of glycoside hydrolases is restricted to the current sequences available in the databases. Thus, when analyzing metagenomes with conventional methods, novel glycoside hydrolase genes with low sequence similarities are usually not identified.

In the scope of this research, we developed a sequence-based screening for eliciting novel glycoside hydrolases genes from metagenomic data. We designed our search algorithm addressing three principles:

1. Glycoside hydrolases designated algorithm, rather than a full data analysis.
2. Genomic Neighborhood approach assimilation for sieving excess data.

Our search algorithm managed to reduce the analyzed data, and allowed us to focus on several metagenomic clusters, which potentially harbor novel glycoside hydrolases.

Following the scan of 75 thermal-springs metagenomes, 97 DNA scaffolds with 778 hemicellulolytic genes have been identified. Two clusters with putative cellulose degradation pathways were chosen for further evaluation and three putative novel glycoside hydrolases were further analyzed biochemically. To-date, we identified a gene product with CMCase activity in each cluster, belonging to families GH5 and GH44. The results support our working hypothesis that these clusters are involved in cellulose hydrolysis. Several of the remaining genes in the clusters are potentially new glycoside hydrolases families and are in the process of biochemical characterization.