

## Gene prospecting for industrial applications through comparative genomic analysis

Victor G. Capone\*, José Madeira, Juliana José, Andreas K. Gombert

### Abstract

Second generation ethanol is extracted from sources other than the sugar cane juice such as leaves and bagasse, some of the challenges of this process are the high temperatures and anaerobic environment inside the vats where the fermentation occurs, usually by the action of the yeast *S. cerevisiae* which can grow in full anaerobic conditions, furthermore, this yeast is not thermotolerant and can't grow in temperatures above 32°C, in this work we compare 6 different species of yeasts to try and determine the genetic mechanisms behind anaerobic growth and thermotolerance.

### Key words:

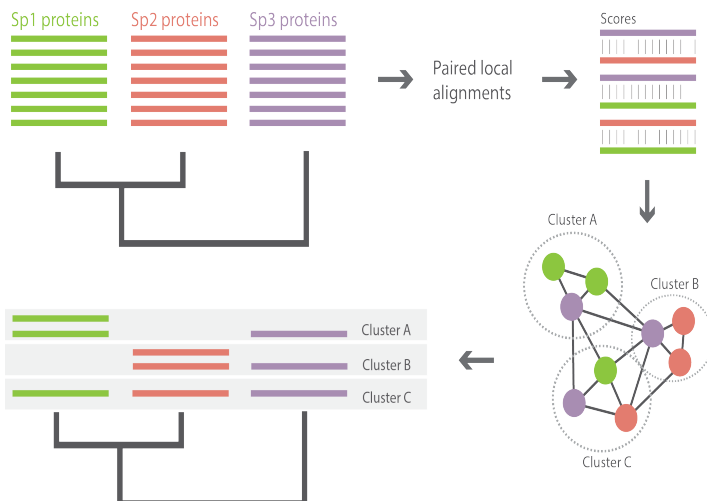
*bioinformatics, comparative genomics, ethanol*

### Introduction

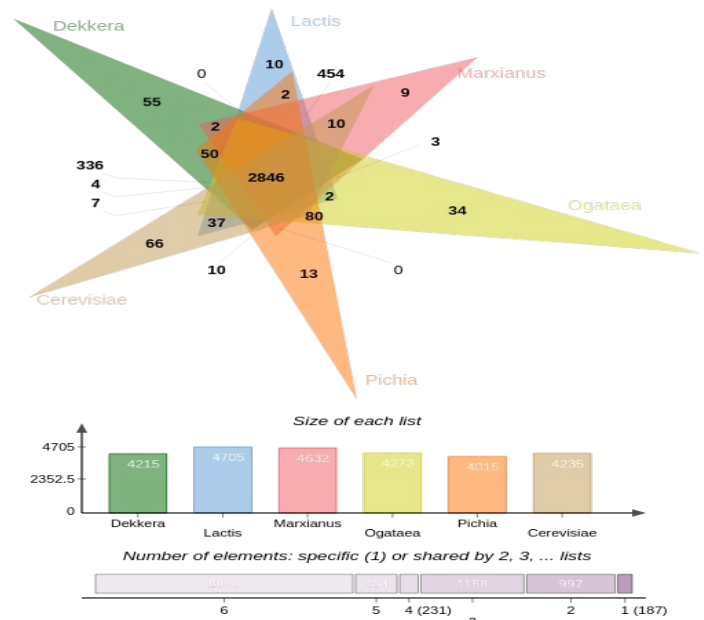
Some chemicals, such as ethanol, pyruvate, acetone and lactate, which are used to create complex compounds, are only produced by fermentation, which require an anaerobic environment to happen, also, the high temperatures reached by the vats are usually non optimal for the growth of the microbial strains used in the process and the cooling costs correspond up to about 20% of the total. Using comparative genomics over 6 different yeast species, this work prospects for genes related to thermotolerance and ability to grow in anaerobiosis, this is done by looking at gene families discovered using the Markov clustering algorithm over BLAST's similarity queries.

### Results and Discussion

Using the OrthoMCL framework, the protein sequences from 6 yeast species, *S. cerevisiae*, *K. lactis*, *K. marxianus*, *O. Polymorpha*, *D. bruxellensis* and *P. pastoris*, were filtered to remove noise aminoacids in the intergenic space, the resulting sequences were used to generate a BLAST database, from which good candidates matches for the MCL algorithm were extracted, the MCL algorithm was then applied to cluster the gene families from the similarity matrix.



**Image 1.** Clustering Process Overview



**Image 2.** Venn Diagram of the gene families

### Conclusions

After comparing the data with genes related to thermotolerance and anaerobic growth described in literature, experiments will be performed to determine if any genes found are related to those characteristics.

### Acknowledgement

We acknowledge Fundação de Apoio a Pesquisa do Estado de São Paulo (FAPESP) for its financial aid in this work.

Dongen, Stijn van, *Graph clusterization by flow simulation*, 2000

Snoek, I.S. Ishtar, Steensma, H. Yde *Why does Kluyveromyces lactis not grow under anaerobic conditions? Comparison of essential anaerobic genes of Saccharomyces cerevisiae with the Kluyveromyces lactis genome*, 2005