

Bioinformatics

P342

PHENO-METABOLOMICS CHARACTERIZATION OF A *SACCHAROMYCES CEREVISIAE* STRAIN COLLECTION BY INTEGRATIVE DATA ANALYSIS APPROACHES

Ricardo Franco-Duarte¹; Lan Umek²; Inês Mendes¹; Cristiana C. Castro³; João Silva³; Rosa Martins⁴; António C. Silva Ferreira⁴; Blaz Zupan²; Dorit Schuller¹

¹CBMA – Molecular and Environmental Research Centre, Department of Biology, University of Minho, Braga, Portugal

²Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

³IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Braga, Portugal

⁴Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Porto, Portugal

The objective of the present work was to characterize the pheno-metabolomic diversity of a *Saccharomyces cerevisiae* wine strain collection, by combining genetic, phenotypic and metabolomic data using bioinformatic approaches.

A strain collection was constituted comprising 172 strains with different geographical origins and technological uses. Phenotypic screening (30 physiological tests) and genetic characterization (11 microsatellite loci) was previously performed [1], and revealed strain variability. Must fermentations were performed with each strain and from the combined data (fiber optics spectroscopy, physiological and molecular characterization) a sub-group of 24 strains were chosen for further approaches.

HPLC analysis revealed variable results, with glucose, fructose and acetic acid contributing most for inter-strain variability. Relevant metabolites related to aromatic profiles were determined by GC-MS. PCA showed significant variance between strains with a clear separation between the amounts of alcohols and esters produced, according to the technological group. Hexyl acetate, ethyl octanoate and decanoic acid showed the highest weight in strain variability, according to the PCA plot. Partial least squares discriminant analysis (PLS-DA) was then used in a pairwise approach to predict the metabolic profiles of the strains, using phenotypic and genetic data. For 15 (70%) of the 24 metabolites, the PLS model found associations (coefficient of correlation above 0.80) with phenotypic or genetic data.

A more holistic matrix factorization approach [2] was then used to project data onto a common system of coordinates, in which the most related variables are weighted together and placed apart from the axis origin. The search for multi-dimensional modules is currently underway, to obtain a group of most-correlated features.

This work was financially supported by the program PTDC (AGR-ALI/103392/2008).

[1] Mendes, Franco-Duarte, *et al* (2013) Computational models for prediction of yeast strain potential for winemaking from phenotypic profiles Computational models for prediction of yeast strain potential for winemaking from phenotypic profiles. *Plos One* 8(7):e66523

[2] Zhang, *et al* (2012) Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Research* 40(19):79-91