Systematic Integration Of "Omics" Data To Improve Innovation In Beer Crafting

Barbara Dunn[1], Dan Kvitek[2], Monica Carrera[3], Xiaoyue Jiang4, Gina Tan[4], Daniel Lopez-Ferrer[4], Andreas FR Huhmer[4]
[1]Dept. of Genetics, Stanford University, Palo Alto, CA, USA.; [2]Invitae, San Francisco, CA, USA.; [3]Spanish National Research Council (CSIC), Vigo, Spain. 4Thermo Fisher Scientific, San Jose, USA.

Poster Note 64486

Overview

Purpose: This study represents a detailed analysis of individual biological components for two yeast strains: WLP001 (California Ale) and WLP300 (Hefe Weizen). The datasets are comprised of quantitative proteomics and metabolomic data collected during the fermentation process. Integration of these datasets will allow better understanding of flavor and taste differences as part of the evolutionary endeavor of both strains.

Results: This work describes the process of explaining the differences among two popular beer strains. Proteomics data was used to compare the fermentation process among both samples. Metabolomics data was then used to further explain the sensorial differences between both strains and to explain the expression differences between the two enzymatic machineries. Protein expression points out that these two ecologically similar strains are significantly different. 60 enzymes including proteolytic enzymes as well as others that are responsible for the flavor and taste of beer are statistically different. We also noticed the different profiles of molecular functions and processes for each strain when ontology enrichment was performed using Thermo Scientific™ ProteinCenter™ software. These findings were reinforced by the metabolomic data that shows very different molecular profiles. Over 2000 compounds were monitored and ~600 compounds were significantly different. This work demonstrates that complex relationships between different layers of biological information that give rise to cellular functions can only be capture by combining global measurements across these different levels.

Introduction

Recently the beer industry has experienced a boom in microbrewing. Unique organoleptic characteristics and efficient beer production is largely related to the yeast strain used. Systems biology offers the possibility to better understand the different biology among yeast strains and the combination of omics-based high throughput technologies with bioinformatic tools represents a valuable new tool and potential alternative to exploring new venues in the beer industry.

Methods

Sample Preparation

Single cell colony fermentations were carried out in triplicates of each strain in 5% malt extract growing at 25C over 100 hours. 7 different time points were collected. Cells were kept for proteomic and supernatants for metabolomic analysis. A modified Mass Spec Sample Prep Kit for Cultured Cells (Pierce, Rockford IL) was used to prepare the proteomic samples.

Liquid Chromatography and Mass Spectrometry Analyses

Peptide digests were then analyzed by LC-MS/MS analysis on Thermo Scientific™ EASY-nLC™ 1000 coupled to a Thermo Scientific™ Q Exactive™ Plus mass spectrometer over a 2-hour gradient. Cell culture supernatants were directly analyze by LC-MS/MS in positive and negative mode on an Accela LC system coupled to a Thermo Scientific™ Q Exactive Plus™ MS operating in positive and negative mode over a 15 min gradient

Data Analysis

Proteomics data was searched using SEQUEST HT. Thermo Scientific™ Sieve 2.2 software was used to process the metabolomics data. Dante RDN was used for further statistical analysis of both datasets. ProteinCenter software was used to extract biological context and set comparisons with publicly available datasets.

Thermo SCIENTIFIC

Study Design

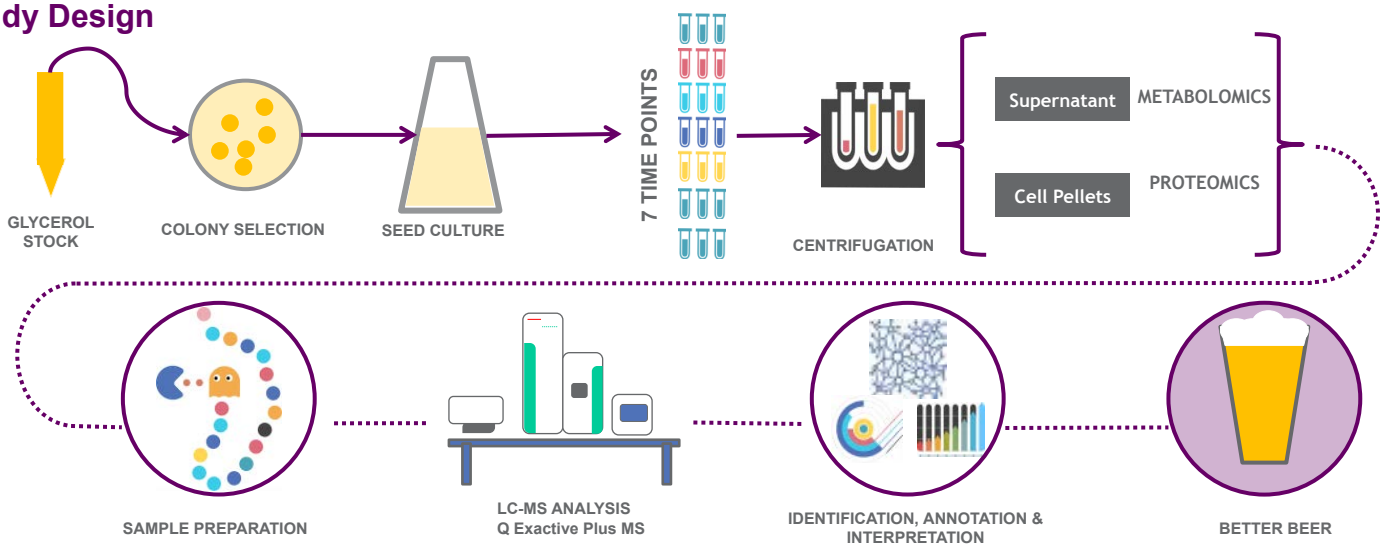

FIGURE 1. Workflow summarizing the analytical approach. Samples were obtained from single cell colony fermentations carried out in triplicates. Collected samples were centrifuged, cell pellets were submitted for proteomic analysis and supernatants were kept for metabolomic analysis. For proteomics, proteins from the cell lysate were precipitated and digested. After digestion, the samples were acidified, dried down and analyzed by LC-MS. Finally, peptide and protein identification was performed, as well as, label free quantitation. Supernatants were directly analyzed by LCMS and features/components were extracted. Subsequently, statistics and data mining including gene ontology enrichment was performed.

Results

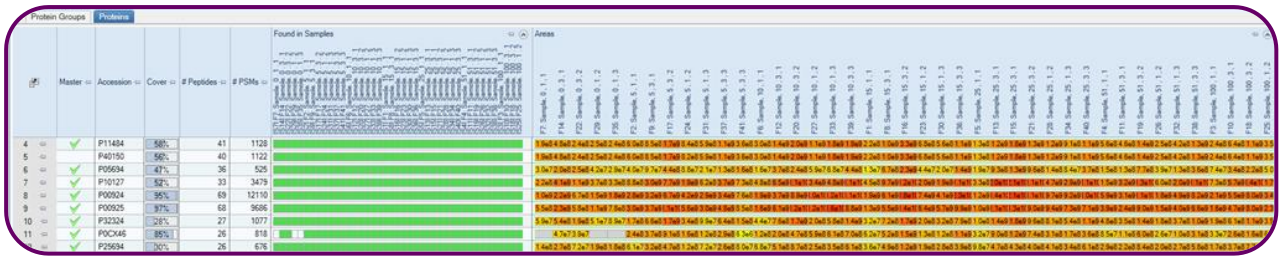

FIGURE 2 shows a screenshot of the Proteome Discoverer software results. After database search, areas from the identified peptides were extracted using the Precursor Ions Area Detector plug-in. The "Areas" table shows the average of the peak areas of the top 3 identified peptides for any given protein in each sample. The table can be directly exported as a text file to easily perform further statistical analysis

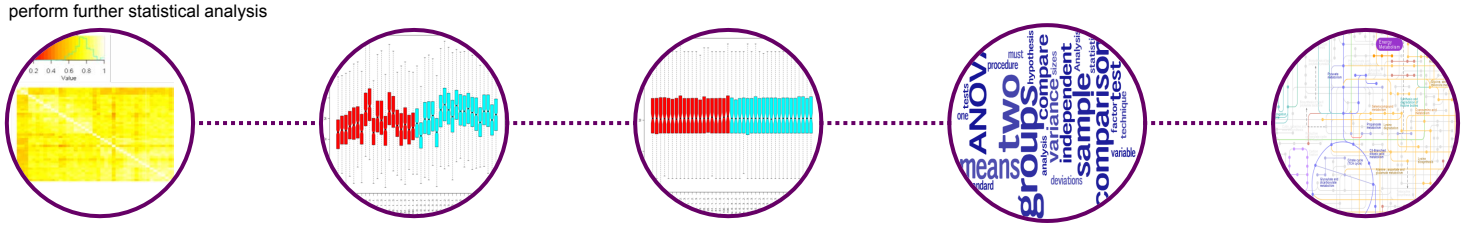

FIGURE 3. Statistical Analysis Workflow. Reproducibility was first evaluated using a correlation plot to discard outliers. The average correlation was 0.78. Then missing value imputation was performed, raw intensities were $\log_2$ transformed and normalized using quantile normalization. For the metabolomics data a similar workflow was applied. However, the intensities were normalized using LOWESS. Finally, an ANOVA test was performed to classify the samples and discover those proteins/compounds that changed in abundance. A Welch-approximation was used to calculate the variance. P-values were based on permutations by randomly grouping the samples 1000 times. Bonferroni correction was applied to adjusted the proportion of false positives not exceed 0.01 significant features. Significant features were further analyzed with Protein Center software (proteomic data) or KEEG, ChemSpider and mzCloud (metabolomic data).

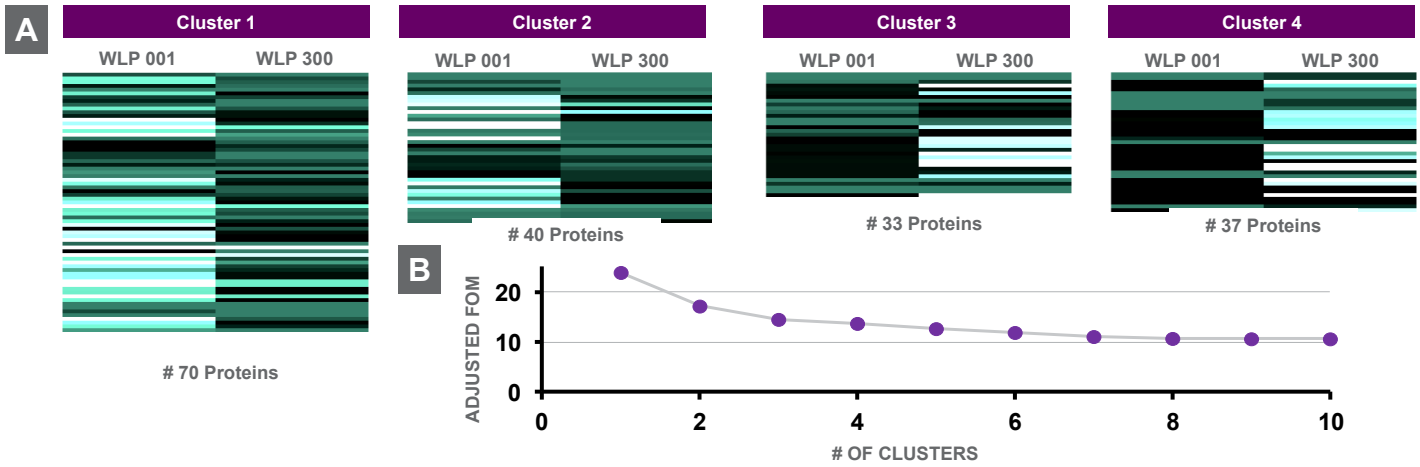

Figure 6. Differential protein expression during fermentation for both beer strains. (A) Differentially, expressed proteins (q-value <0.01) from more than 1000 proteins accurately quantified were clustered using k-means method. **(B)** We used multiple cluster validation metrics based on Figures Of Merit (FOM) to select the best performing clustering algorithm as well as the most adequate number of clusters .

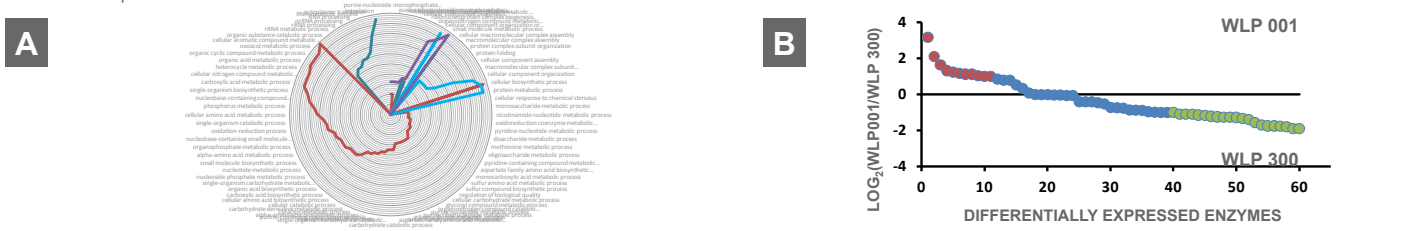

Figure 6. Ontology enrichment of the differentially expressed proteome. The four clusters from the differentially, expressed proteins (p-value<0.01) were analyzed using ProteinCenter software. GO terms, pathways and enzymes significantly over-represented when compared towards the *Saccharomyces cerevisae* proteome were selected. And submitted for further statistical analysis. **Figure 6A** shows a radar map using the ontology terms and their significance values to highlight the differences within the molecular processes between both strains (Red=cluster1, Dark green=cluster 2, Purple=cluster 3 and Blue=cluster 4). Furthermore, over 60 enzymes were differentially expressed, which may explain the different organoleptic properties.**(B)**

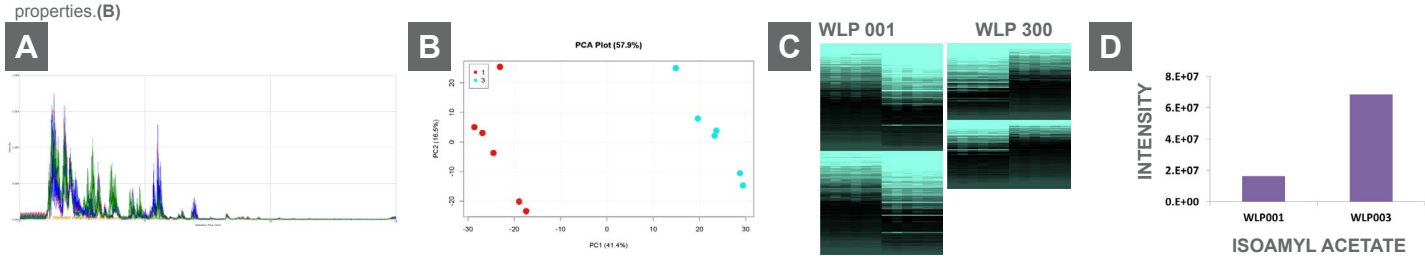

Figure 7. Metabolomics dataset. (A) Screenshot of the alignment for all the metabolomic datasets. **(B)** PCA plot highlighting the differences between both strains. **(C)** Compounds with significantly different abundances (q-value<0.01) were clustered. 23% of the compounds were significantly different. Most of them belong to yeast specific metabolic pathways, although degradation of aromatic compounds as well as biosynthesis of secondary metabolites pathways were highly represented. As expected isoamyl acetate is higher abundant in the Hefe Weizen strain compared to the California Ale **(D).** The metabolomic dataset found that isoamyl acetate in its sodiated adduct was significantly higher in abundance in the Hefe Weizen strain. This compound is responsible for the strong banana nose with a slight hint of clove.

Conclusion

- Simple, but yet powerful proteomic and metabolomics workflow, based on label free quantitation, single LC runs on a bench top mass spectrometer and data analysis by Proteome Discoverer software, Protein Center software and Sieve software (2.2).
- Integration of omics data using systems biology approaches and mass spectrometry hold a great potential to better understand molecular differences between evolutionary distant, but ecologically very similar strains and to design new strains that could enhance the organoleptic properties of beer.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic) 400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 45 453 9100
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

PN64486-EN 0616S

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand