Supplementary figures for “Genome wide analysis of protein production load in Trichoderma reesei”

Contents
1. Lipase production .................................................................................................................. 2
2. Ethanol production .............................................................................................................. 2
3. Intracellular free amino acid concentrations ...................................................................... 3
4. Gene expression clusters .................................................................................................... 13
5. Correlations of genes and cultivation parameters .............................................................. 15
6. Overlap of gene expression clusters and lists of significantly changing genes .................. 17
7. CAZY and related genes known to be highly produced based on proteomics ...................... 18
8. FIRE analysis for detecting shared promoter motifs in gene expression clusters ............... 20
9. Comparison of gene’s correlation to specific protein production rate in this publication and in Arvas 2011 “Correlation of gene expression and protein production rate - a system wide study” .............. 22
10. Flux clusters ..................................................................................................................... 23
11. Glycerol identification ....................................................................................................... 24
12. Cellotriose identification ................................................................................................... 25
13. Fits of heteroscedastic Gaussian processes ...................................................................... 26
   A. Biomass i.e. CDW (g/l) ..................................................................................................... 26
   B. Extracellular protein (g/l) ............................................................................................... 27
   C. Cellobiose (g/l) .............................................................................................................. 27
   D. MUL (nkat/l) ............................................................................................................... 28
   E. Cellotriose (g/l) ............................................................................................................. 28
   F. Glucose (g/l) ................................................................................................................. 29
   G. Glycerol (g/l) ................................................................................................................. 29
1. Lipase production

2. Ethanol production
3. Intracellular free amino acid concentrations

- Ser (umol/g)
- Thr (umol/g)

Graphs showing the concentration changes over time (h) for various conditions: Cel4d, Cel4dCt, CutCBHd, CutCBHdCt, LipPr4d, and LipPr4dCt.
4. Gene expression clusters
5. Correlations of genes and cultivation parameters

20 genes with lowest correlation above the FDR cut-off of 0.00005. Y-axes cultivation parameter, X-axes gene expression (rlog2). Title, the un-adjusted p-value of correlation and the correlation.
6. Overlap of gene expression clusters and lists of significantly changing genes
7. CAZY and related genes known to be highly produced based on proteomics

Gene expression as rlog2 value.
8. FIRE analysis for detecting shared promoter motifs in gene expression clusters

Motif 1:

Motif 2:

Motif 3:
9. Comparison of gene’s correlation to specific protein production rate in this publication and in Arvas 2011 “Correlation of gene expression and protein production rate - a system wide study”

On y-axes the gene expression’s correlation to specific protein production rate in this publication. On x-axes gene expression’s correlation to specific protein production rate in Arvas 2011. Each dot is a single gene.
10. **Flux clusters**
11. Glycerol identification

The glycerol determination was confirmed by gas Chromatography–mass spectrometry (GC/MS). The dried growth media samples (A) and glycerol reference compound (B) were derivatized with MSTFA containing 1% of TMCS. Retention time of the peak in the sample had matching retention time and mass spectra with glycerol reference compound NIST MS library search produced a good match (923) with glycerol reference spectra (C).
12. Cellotriose identification

Extracted ion chromatograms, m/z 503.15 showing a growth media sample (A) and cellotriose reference compound (B). Mass spectra of the peaks 1 and 2 at retention times 3.57 and 3.70 min respectively, in the sample chromatogram were similar as the spectra of cellotriose reference compound. Based on accurate mass of [M-H]⁻ and [M+Cl]⁻ the compounds 1 and 2 were similar trisaccharides but the precise isomer cannot be assign based on mass spectra. The samples were analysed using ultra high performance-hydrophilic interaction liquid chromatography-mass spectrometry (UHP-HILIC-MS). The instrument was Waters Acquity Ultra Performance LC™ (UPLC) combined with Synapt G2-S mass spectrometer and Acquity UPLC™ BEH Amide (2.1 x 100 mm with 1.7 μm particles) column was used.
Fits of heteroscedastic Gaussian processes

In the figures the dots present actual measurement data, the solid red line is the distribution mean, coloured region indicates the 95% interval, while the dashed line includes also the observational noise model.

A. Biomass i.e. CDW (g/l)
B. Extracellular protein (g/l)

C. Cellobiose (g/l)
D. MUL (nkat/l)

E. Cellotriose (g/l)
F. Glucose (g/l)

G. Glycerol (g/l)