

Metabolomics-based prediction models of yeast strains for screening of metabolites contributing to ethanol stress tolerance

Z Hashim¹, E Fukusaki²

¹Department of Bioprocess Engineering, Faculty of Chemical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

²Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

E-mail: ¹zanariahhashim@utm.my, ²fukusaki@bio.eng.osaka-u.ac.jp

Abstract. The increased demand for clean, sustainable and renewable energy resources has driven the development of various microbial systems to produce biofuels. One of such systems is the ethanol-producing yeast. Although yeast produces ethanol naturally using its native pathways, production yield is low and requires improvement for commercial biofuel production. Moreover, ethanol is toxic to yeast and thus ethanol tolerance should be improved to further enhance ethanol production. In this study, we employed metabolomics-based strategy using 30 single-gene deleted yeast strains to construct multivariate models for ethanol tolerance and screen metabolites that relate to ethanol sensitivity/tolerance. The information obtained from this study can be used as an input for strain improvement via metabolic engineering.

1. Introduction

Increasing energy usage, high cost and environmental concerns require for the development of sustainable and renewable fuels and chemicals. Fuels and chemicals have long been produced from plant and animal-based resources, but this has raised issues such as competition with food, high prices and unethical land-use practices. A sustainable alternative is to produce directly from abundant and cost-effective renewable resources by microbial conversions.

To solve the problems faced by current biofuel microbial productions, several strategies have been implemented. Mainly there are three approaches; strain development using metabolic engineering, raw material development which includes searching for new and cost effective resources, and process optimization to obtain high yield and low cost. Metabolic engineering of microbial strain has been considered to be a powerful approach to improve biofuel manufacturing. In particular, the potential of metabolic pathways to produce biofuel molecules from microbes has been described [1]. Currently ethanol dominates the biofuel market. Besides ethanol, other candidate compounds for biofuels such as higher alcohols or fatty acid-derived compounds [2] have also been identified. Although yeast has a native pathway to produce ethanol, production yield is low for commercial biofuel production. In

^{2 2} To whom any correspondence should be addressed.



addition, ethanol is toxic to yeast, and thus one strategy to improve ethanol production is via increased ethanol tolerance.

Metabolomics, the exhaustive profiling of metabolites [3,4] contained in a cell or an organism is a powerful tool to study phenotypic changes. Metabolomics offers an unbiased view on cellular functions and links between genotype and phenotype [3]. Previously, Yoshida *et al.* [5] demonstrated that metabolite profile under a standard growth condition is able to predict the lifespan of yeast mutant strains. Using the same concept, we now apply metabolite profiling to construct multivariate models to predict ethanol tolerance and screen for responsible metabolites. The candidate metabolites obtained from this study can be linked back to their putative pathways and considered for metabolic engineering, for example using gene deletion or overexpression strategies, to improve ethanol tolerance in yeast. It is known that the screening of candidate genes/pathways to be modified in metabolic engineering has been difficult due to the complexity of metabolic landscapes [6,7], and the candidates are often picked from known pathways. Thus, a direct measurement of metabolites may reveal better information about metabolite alteration and provide a semi-rational approach for metabolic engineering.

2. Materials and method

2.1. Yeast strains

Yeast BY4742 (*MATa leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1*) and its isogenic derivatives, *yfgΔ* (*MATa leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1 yfg::kanMX*) were purchased from Open Biosystems (Huntsville, AL, USA). The 30 yeast strains used were:

std1Δ, *ace2Δ*, *sut1Δ*, *hcm1Δ*, *mig1Δ*, *tec1Δ*, *fkp2Δ*, *skn7Δ*, *kar4Δ*, *mss11Δ*, *bas1Δ*, *swi5Δ*, *pho2Δ*, *rgt1Δ*, *crz1Δ*, *ino4Δ*, *mgalΔ*, *cup2Δ*, *rtg1Δ*, *rlm1Δ*, *tos4Δ*, *mbp1Δ*, *sok2Δ*, *hms2Δ*, *hot1Δ*, *stp2Δ*, *mot3Δ*, *rpn4Δ*, *srb2Δ*, *aft1Δ*

The gene function based on gene ontology (GO) annotation can be obtained from SGD (Saccharomyces Genome Database). These strains were chosen randomly to represent a wide variety of ethanol tolerance values.

2.2. Yeast culture and metabolite extraction

Frozen stocked (-80 °C) yeast strains were plated onto YPD agar plate (yeast extract 10 g/L, peptone 20 g/L, dextrose 20 g/L, agarose 20 g/L) with added geneticin G418 200 µg/mL (except for wild-type strain) and grown at 30 °C for two days. The cultivation and sampling were performed as described previously [8]. Briefly, pre-cultured yeast cells were grown in YPD medium at 30 °C in a rotary shaker (200 rpm) to OD₆₀₀ = 1.0 (optical density at 600 nm, equivalent to 1 mg cells-dry weight) and harvested using a rapid filtration system. After washing with 5 mL of water, the filter membrane containing yeast cells was immediately folded, inserted into a 2.0 mL Eppendorf tube and plunged into a liquid nitrogen bath to quench metabolism. Time from start pouring culture broth until quenching was kept within 30 seconds. Samples were then stored at -80 °C until extraction.

Prior to extraction, the filter-bound cells were disrupted with MM 301 mixer mills (Retsch, Haan, Germany) at 20 Hz for 1 min. 500 µL of extraction solvent (acetonitrile / methanol / water = 2 / 2 / 1 with added 1.2 µg/mL of 1,4-piperazinediethanesulfonic acid (Dojindo, Kumamoto, Japan) as an internal standard) precooled to -30 °C was added, and extraction was carried out on ice for 15 min. After centrifugation at 15000 rcf for 5 min, the supernatant was transferred to a new tube. A second round of extraction was done with another 500 µL of extraction solvent. Supernatants from both extractions were combined, and 500 µL of chloroform was added. After vortex and centrifugation, the upper polar phase was transferred to a new tube (about 450 µL). 400 µL water was added to the remaining mixture (containing chloroform), vortexed and centrifuged. Another 450 µL of upper polar phase was collected and combined with the previous polar extract. Finally, after centrifugation, 700 µL of polar extract was collected and filtered (0.2 µm PTFE, Millipore, MA, USA), and applied to a

centrifugal concentrator VC-96R (Taitec, Osaka, Japan) for 2 h. The remaining extract was then transferred to a glass vial and ready for LC analysis.

2.3. Ethanol tolerance measurement

Frozen cells were plated onto YPD agar plates and let to grow for two days. A single colony from each strain was transferred into 1 mL of YPD medium placed in a deep 96-well plate. This plate was set on a plate shaker (1050 rpm) at 30 °C, and cells were let to grow overnight (about 16-18 h). The overnight pre-culture broth from each well was diluted 100 times in fresh YPD medium and distributed to a 96-well microplate, 100 μ L per well. The plate was incubated in a 30 °C oven with no shaking for 4 h. After 4 h, 25 μ L of 35% ethanol (for 7% ethanol stress condition) or sterilized water (for stress-free condition) was added, and incubation was continued for 8 h. OD₆₀₀ was measured every 1 h. Growth rate for mutant and parental strains were compared in stress and stress-free conditions.

2.4. Metabolite profiling

The LC-MS platform consists of ACQUITY Ultra Performance LC and LCT Premier XE (Waters, Milford, MA, USA). Separation was achieved on Acquity UPLC HSS T3 column (2.1 x 150 mm, 1.8 μ m particle size, Waters), using reversed phase chromatography with ion pairing agent tributylamine in the aqueous phase. Solvent A is 97/3 water/methanol with 10 mM tributylamine and 15 mM acetic acid; solvent B is methanol. Solvent gradient profile is as follows; 0 min, 0% B; 2.5 min, 0% B; 5 min, 20% B; 7.5 min, 20% B; 13 min, 55% B; 15.5 min, 95% B; 19 min, 0% B; 25 min, 0% B. The solvent flow rate is kept at 200 μ L/min. Other LC parameters are autosampler temperature 4 °C, injection volume 10 μ L (full loop), and column temperature 25 °C. Ionization method was ESI in negative mode. The mass spectrometer was tuned and calibrated before use, and final analysis parameters are: capillary voltage 2200 V, sample cone voltage 25 V, desolvation temperature 300 °C, source temperature 120 °C, cone gas flow 50 L/h, desolvation gas flow 500 L/h. Leucine encephalin 2.0 ng/ μ L was used as a reference solution, with 5.0 μ L/min flow rate. The scan speed was 0.1 s (10 scans/s) and mass range was 70-900 Da.

2.5. Data processing

Raw data from LC-MS was read using Masslynx (proprietary MS software from Waters). The mass window was set at 0.05 Da. Based on the retention time values extracted, peak identification method was done using Quanlynx (Waters) and verified by visual inspection. The peak list table was constructed, for each semi-quantified metabolite, composing of metabolite type and abundance in each yeast mutant strain. Identification is based on 48 validated metabolites, and 63 putative metabolites from literature [9]. Peak intensities were expressed as the peak area after alignment and baseline correction where necessary, and normalized to internal standard and total raw signal.

2.6. Multivariate data analysis and data visualization

Multivariate data analysis was performed using SIMCA-P+ ver. 12 software (Umetrics, Umeå, Sweden). Peak list table in Excel format was imported to SIMCA-P+ and multivariate plots (PCA, OPLS and OPLS-DA) were computed automatically. Data was scaled to unit variance.

3. Results and discussion

3.1. Yeast growth

The gross specific growth rate μ (h^{-1}) was determined by the relation between cell mass concentration and time, based on the kinetic model of cell growth. The result is shown in Table 1. The specific growth rate is a key control parameter in industrial application of yeast and serves as a “fitness” indicator of yeast strains [10], therefore monitoring it is of prime importance. Out of the 30 strains,

only *srb2Δ* demonstrated slow growth (approximately 12% lower gross specific growth rate compared to wild-type BY4742), whereas all other strains exhibited comparable growth rates.

Table 1. Growth kinetics under YPD medium for 30 yeast mutant strains used in this study. (RSD: relative standard deviation, measured for $n = 3$)

Strain no.	Strain name	Gross specific growth rate, μ (h^{-1})	Relative standard deviation, RSD % ($n = 3$)	Doubling time, t_d (h)
1	<i>std1Δ</i>	0.46	0.77	1.50
2	<i>ace2Δ</i>	0.48	0.79	1.43
3	<i>sut1Δ</i>	0.47	0.51	1.48
4	<i>hcm1Δ</i>	0.44	0.59	1.57
5	<i>mig1Δ</i>	0.46	0.72	1.51
6	<i>tec1Δ</i>	0.46	0.47	1.50
7	<i>fkh2Δ</i>	0.44	1.33	1.57
8	<i>skn7Δ</i>	0.47	0.78	1.48
9	<i>kar4Δ</i>	0.46	0.55	1.51
10	<i>mss11Δ</i>	0.46	0.45	1.50
11	<i>bas1Δ</i>	0.41	0.43	1.71
12	<i>swi5Δ</i>	0.44	0.28	1.57
13	<i>pho2Δ</i>	0.44	0.19	1.56
14	<i>rgt1Δ</i>	0.46	0.29	1.50
15	<i>crz1Δ</i>	0.45	0.50	1.55
16	<i>ino4Δ</i>	0.47	0.92	1.48
17	<i>mgal1Δ</i>	0.47	0.88	1.48
18	<i>cup2Δ</i>	0.47	0.46	1.49
19	<i>rtg1Δ</i>	0.47	0.24	1.47
20	<i>rlm1Δ</i>	0.45	0.79	1.55
21	<i>tos4Δ</i>	0.46	0.50	1.50
22	<i>mbp1Δ</i>	0.45	1.19	1.55
23	<i>sok2Δ</i>	0.42	0.53	1.67
24	<i>hms2Δ</i>	0.46	0.09	1.50
25	<i>hot1Δ</i>	0.47	0.68	1.49
26	<i>stp2Δ</i>	0.46	0.28	1.52
27	<i>mot3Δ</i>	0.42	0.82	1.66
28	<i>rpn4Δ</i>	0.42	0.24	1.67
29	<i>srb2Δ</i>	0.36	2.46	1.91
30	<i>aft1Δ</i>	0.46	0.54	1.50
Wild-type	BY4742	0.41	0.59	1.67

3.2. Ethanol stress tolerance

Ethanol stress tolerance for 30 yeast mutant strains and wild-type strain was measured by comparing the gross specific growth rates under stress condition (7% ethanol) with non-stress condition. The result is shown in Table 2. Based on the stress tolerance (Y_2/Y_1) values, there were five strains classified as ethanol sensitive which are *swi5Δ*, *mbp1Δ*, *stp2Δ*, *mot3Δ*, and *srb2Δ* (indicated in **bold** in Table 2). These strains have 10% or more reduction of tolerance value compared to wild-type strain ($p > 0.1$).

To compare current finding with previous works, we searched for the five genes (*SWI5*, *MBP1*, *STP2*, *MOT3*, *SRB2*) in Saccharomyces Genome Database (SGD, <http://www.yeastgenome.org/>). Table 3 summarizes gene description, competitive fitness and ethanol resistance under null mutation. Indeed, the loss of these genes resulted in decreased competitive fitness (except for *SWI5* which has been reported as both increased and decreased competitive fitness). Moreover, while there was no data for *SWI5* and *MOT3* regarding ethanol sensitivity, *MBP1*, *STP2* and *SRB2* all showed decreased ethanol resistance (Table 3). Thus, our result is consistent with previous reports.

Table 2. Growth rate of yeast strains under stress-free (normal growth), Y1 and stressed (7% ethanol), Y2 conditions and the relative stress tolerance Y2/Y1. **Bold** indicates ethanol sensitive strains, defined as those having 10% or more reduced ethanol tolerance compared to BY4742 (RSD: relative standard deviation, measured for n = 3)

Strain no.	Strain name	Stress-free (no ethanol)		7% ethanol		Stress tolerance
		Y1 (h ⁻¹)	RSD (%)	Y2 (h ⁻¹)	RSD (%)	Y2/Y1 (%)
1	<i>std1Δ</i>	0.46	0.77	0.26	6.32	55.7
2	<i>ace2Δ</i>	0.48	0.79	0.25	8.64	51.4
3	<i>sut1Δ</i>	0.47	0.51	0.28	7.99	59.1
4	<i>hcm1Δ</i>	0.44	0.59	0.24	8.12	54.8
5	<i>mig1Δ</i>	0.46	0.72	0.26	4.46	57.1
6	<i>tec1Δ</i>	0.46	0.47	0.26	5.89	56.4
7	<i>fkp2Δ</i>	0.44	1.33	0.25	14.16	57.4
8	<i>skn7Δ</i>	0.47	0.78	0.30	8.81	63.7
9	<i>kar4Δ</i>	0.46	0.55	0.24	3.02	52.3
10	<i>mss11Δ</i>	0.46	0.45	0.24	0.83	50.9
11	<i>bas1Δ</i>	0.41	0.43	0.28	4.90	68.9
12	<i>swi5Δ</i>	0.44	0.28	0.18	4.34	41.0
13	<i>pho2Δ</i>	0.44	0.19	0.25	1.21	56.6
14	<i>rgt1Δ</i>	0.46	0.29	0.25	2.04	55.1
15	<i>crz1Δ</i>	0.45	0.50	0.26	6.43	58.5
16	<i>ino4Δ</i>	0.47	0.92	0.29	2.56	62.4
17	<i>mgalΔ</i>	0.47	0.88	0.24	5.55	50.9
18	<i>cup2Δ</i>	0.47	0.46	0.24	4.14	50.6
19	<i>rtg1Δ</i>	0.47	0.24	0.25	9.47	53.7
20	<i>rlm1Δ</i>	0.45	0.79	0.23	7.05	51.0
21	<i>tos4Δ</i>	0.46	0.50	0.21	6.55	46.4
22	<i>mbp1Δ</i>	0.45	1.19	0.17	5.59	38.7
23	<i>sok2Δ</i>	0.42	0.53	0.25	11.64	59.1
24	<i>hms2Δ</i>	0.46	0.09	0.25	5.97	55.1
25	<i>hot1Δ</i>	0.47	0.68	0.23	6.99	48.6
26	<i>stp2Δ</i>	0.46	0.28	0.18	3.21	40.1
27	<i>mot3Δ</i>	0.42	0.82	0.17	2.53	41.4
28	<i>rpn4Δ</i>	0.42	0.24	0.22	7.16	52.0
29	<i>srb2Δ</i>	0.36	2.46	0.13	4.68	37.1
30	<i>aft1Δ</i>	0.46	0.54	0.23	0.09	50.1
Wild-type	BY4742	0.41	0.59	0.21	3.12	49.9

Table 3. Gene description, competitive fitness and ethanol resistance for five genes related to reduced ethanol tolerance, retrieved from Saccharomyces Genome Database (SGD).

Gene	Description	Competitive fitness (null mutation)	Ethanol resistance (null mutation)
<i>SWI5</i>	Transcription factor that recruits Mediator and Swi/Snf complexes; activates transcription of genes expressed at the M/G1 phase boundary and in G1 phase; required for expression of the <i>HO</i> gene controlling mating type switching; localization to nucleus occurs during G1 and appears to be regulated by phosphorylation by Cdc28p kinase; <i>SWI5</i> has a paralog, <i>ACE2</i> , that arose from the whole genome duplication	Decreased/ Increased	No data
<i>MBP1</i>	Transcription factor; involved in regulation of cell cycle progression from G1 to S phase, forms a complex with Swi6p that binds to MluI cell cycle box regulatory element in promoters of DNA synthesis genes	Decreased	Decreased
<i>STP2</i>	Transcription factor; activated by proteolytic processing in response to signals from the SPS sensor system for external amino acids; activates transcription of amino acid permease genes; <i>STP2</i> has a paralog, <i>STP1</i> , that arose from the whole genome duplication	Decreased	Decreased
<i>MOT3</i>	Transcriptional repressor and activator with two C2-H2 zinc fingers; involved in repression of a subset of hypoxic genes by Rox1p, repression of several DAN/TIR genes during aerobic growth and ergosterol biosynthetic genes in response to hyperosmotic stress; contributes to recruitment of Tup1p-Cyc8p general repressor to promoters; involved in positive transcriptional regulation of <i>CWP2</i> and other genes; relocates to the cytosol in response to hypoxia; can form [MOT3+] prion	Decreased	No data
<i>SRB2</i>	Subunit of the RNA polymerase II mediator complex; associates with core polymerase subunits to form the RNA polymerase II holoenzyme; general transcription factor involved in telomere maintenance	Decreased	Decreased

3.3. Metabolites detected and peak list

A total of 111 metabolites were detected in yeast extract, as shown in Table 4. Metabolites indicated in **bold** are those validated in this study (48 metabolites), while the rest is putative metabolites based on the retention time and *m/z* values from Lu *et al.*, 2010 [9] (63 metabolites). We selected liquid chromatograph with mass spectrometer (LCMS) for metabolite quantification platform as this method has been reported to cover a broad range of metabolites. Additionally, the use of an ion pairing reagent tributylamine in the mobile phase improved the retention of polar metabolites such as sugar phosphates and nucleotides that are central in primary metabolism. A peak list table was constructed, listing the relative abundance of these 111 metabolites, for each yeast mutant sample. Here, the metabolite intensities were calculated as relative to internal standard and were normalized to total raw signal. This dataset (sample vs. metabolite intensities) was then subjected to multivariate data analysis such as PCA and OPLS-DA as described in the following sections.

Table 4. List of metabolites detected from yeast extracts. (* indicates structural isomers that are non-separable with current method)

Amino acids (35)	
Alanine	Amino adipic acid
4-aminobutyrate	Aconitate
Serine	Arginine
Proline	Tyrosine
Valine	3-phosphoserine
Threonine/Homoserine*	<i>N</i> -acetyl-glutamine
Hydroxyproline	Acetyllysine
<i>N</i> -acetyl-L-alanine	<i>N</i> -acetyl-glutamate
Leucine/ Isoleucine*	Tryptophan
Asparagine	Cystathione
Ornithine	Glutathione
Aspartate	<i>S</i> -adenosyl-L-homocystein
Homocystein	Glutathione disulfide
<i>p</i> -aminobenzoate	CDP-ethanolamine
Glutamate	UDP- <i>N</i> -acetyl glucosamine
Methionine	D-glucosamine-1/6- phosphate*
Histidine	<i>N</i> -acetyl-D-glucosamine-1/6-phosphate*
Orotate	
Organic acids (28)	
Glyoxylate	Pantothenate
Glycolate	2-keto-isovalerate
Pyruvate	Citraconate
Lactate	Hydroxycaproate
Glycerate	Phenylpyruvate
Fumarate	Phenyllactate
Succinate	Phosphoenolpyruvate
Nicotinate	2-isopropylmalate
Malate	Xanthurenate
2-Oxoglutarate	2,3-diphosphoglycerate
β-Glycerate	2-dehydro-D-gluconate
Shikimate	D-gluconate
3-Phosphoglycerate	6-phospho-D-gluconate
Citrate /Isocitrate*	2-hydroxy-2-methylbutanedioate
Sugars/ Sugar phosphates (17)	
Erythrose-4-phosphate	ADP-ribose
Ribose-5-phosphate	UDP-glucose
Ribulose-5-phosphate	ADP-glucose
Fructose-6-phosphate	Ribulose-1,5- bisphosphate
Glucose-1-phosphate	5-phosphoribosyl-1-pyrophosphate
Glucose-6-phosphate	Trehalose-6-phosphate
Sedoheptulose-7-phosphate	Trehalose/Sucrose/Cellobiose*
Fructose-1,6-bisphosphate	DHAP/GAP*
	<i>sn</i> -glycerol-3-phosphate
Nucleotides / nucleotide derivatives (27)	
TMP CMP UMP AMP GMP	Cytidine
CDP UDP ADP GDP	Uridine
dTMP dTDP dCTP	Inosine
CTP UTP ATP GTP	3',5'-cyclic AMP
NAD NADH	Xanthosine-5'-phosphate
NADP NADPH	
FMN FAD	
Others (4)	
Indole	Acetyl-coenzyme A
Thiamine	Coenzyme A

3.4. Principal component analysis (PCA)

As the first step in multivariate data analysis, principal component analysis (PCA) was performed. PCA provides an unbiased overview of the dataset without any prior information. In PCA, each observation is plotted in a new dimensional space so that the dataset will be fitted to new axes known as principal components or PC, and so that the data is represented in low-dimensional planes.

Here, it was found that ethanol sensitive mutants, indicated by red triangles in Figure 1, showed a tendency to cluster at the lower right quadrant of the PCA score plot. When the ethanol sensitive strains were picked up, together with wild-type strain, and plotted in a new PCA plot (Figure 2), some clustering was observed. Although the strains were not completely arranged by their stress response values, *srb2Δ* which has a significantly lower growth rate under 7% ethanol (0.135 h^{-1}) was clearly discriminated from other strains (Figure 2).

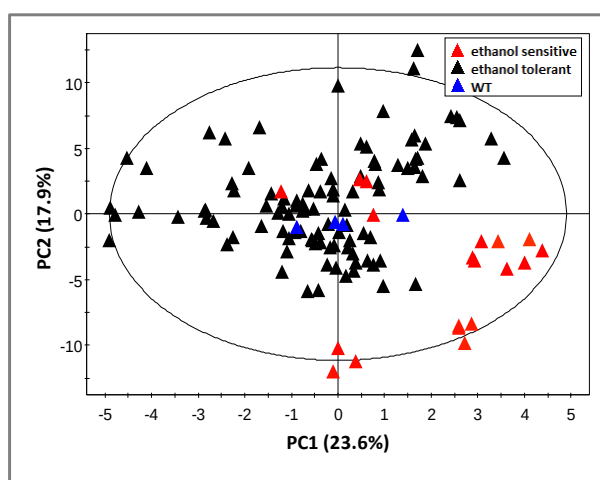


Figure 1. PCA score plot of metabolite profile of 30 yeast mutant strains and wild-type strain WT (Scaling: unit variance).

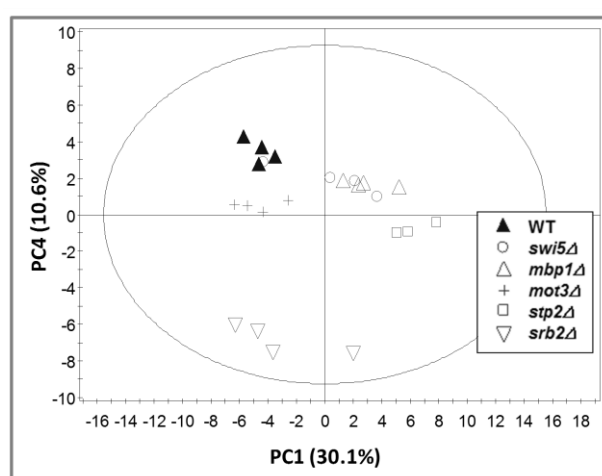


Figure 2. PCA score plot of ethanol sensitive strains and wild-type strain. Strain *srb2Δ* (indicated by reverse open triangles) was clearly separated from other strains.

3.5. Stress tolerance prediction model

To find which metabolites that have a strong contribution to the ethanol stress response, analysis using OPLS-DA (orthogonal projection to latent structure-discriminant analysis) was performed. The yeast strains were each pre-assigned to two classes, class 1 which is ethanol tolerant, and class 2 which is ethanol sensitive, based on the growth rate under 7% ethanol, relative to wild-type strain (Figure 3). From this graph, the corresponding S-plot was obtained (Figure 4), which indicates the contribution of metabolites towards the OPLS-DA model. It was found that citrate/isocitrate and uridine 5'-monophosphate (UMP) were important in ethanol sensitivity and ethanol tolerance, respectively. Tryptophan, which is well known to have effects on ethanol stress in yeast, was also identified.

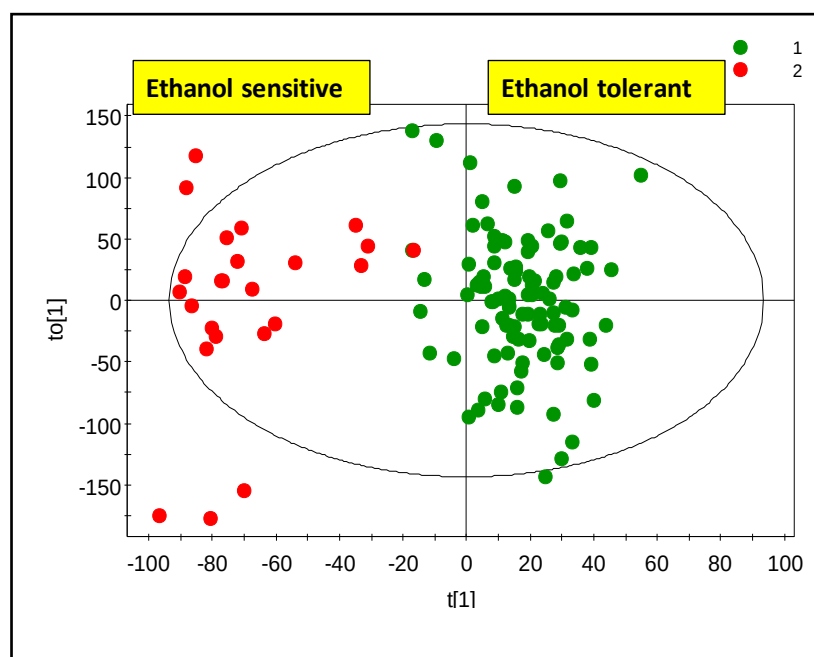


Figure 3. OPLS-DA plot based on the pre-assigned classes of ethanol tolerant (class 1) and ethanol sensitive (class 2) of 30 yeast mutant strains.

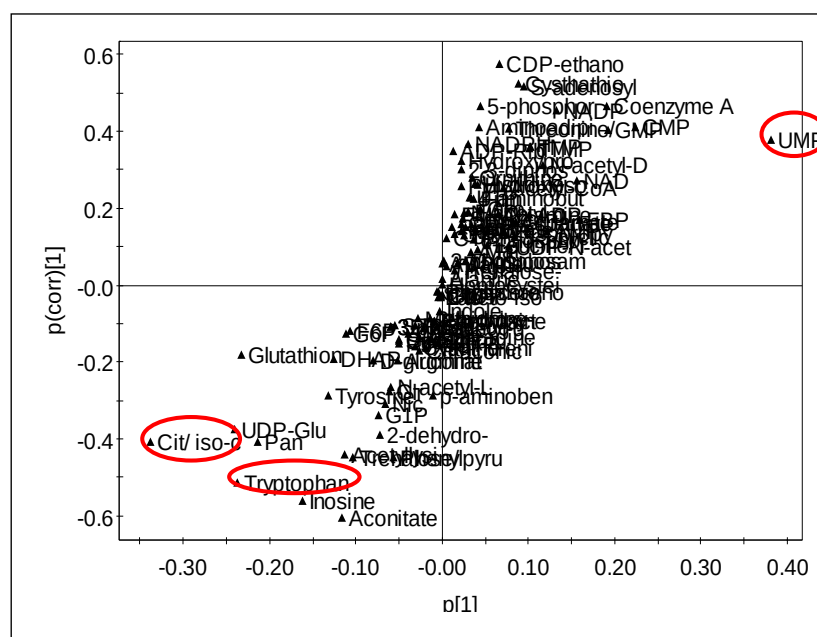


Figure 4. S-plot corresponding to the OPLS-DA model, showing metabolites contribution towards the pre-assigned classes. Citrate/isocitrate and tryptophan contribute to ethanol sensitivity while uridine 5'-monophosphate (UMP) contributes to ethanol tolerance.

Next, prediction model using OPLS (orthogonal projection to latent structures) was performed to yield a linear prediction profile. The horizontal axis shows the predicted ethanol stress from metabolome

data, and the vertical axis shows the actual measured stress response (from Section 3.2). Here, a model with good linearity and prediction ability, indicated by R^2 and Q^2 values respectively, was obtained (Figure 5). $Q^2 > 0.8$ is considered as an excellent model, and the difference between R^2 and Q^2 must be within 0.2 - 0.3 (SIMCA-P+ User guide).

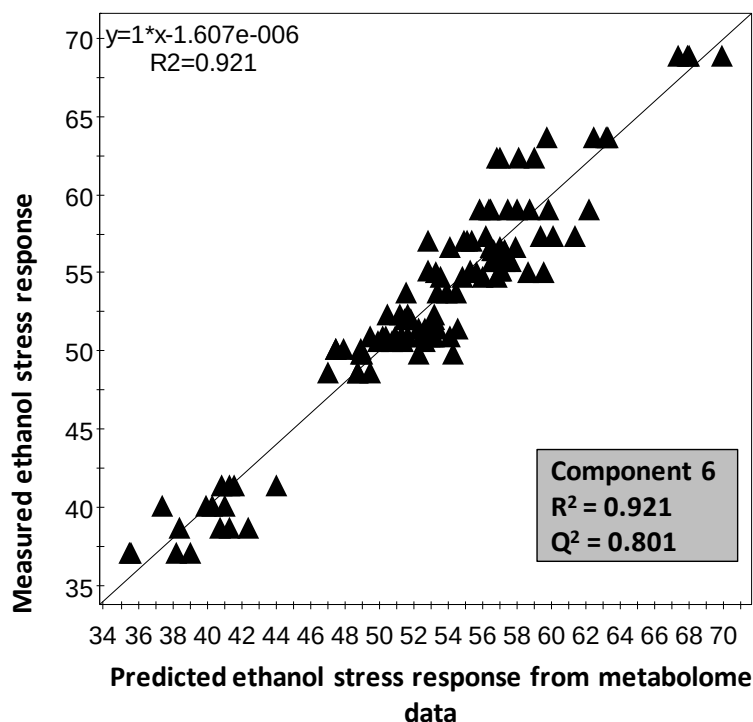


Figure 5. OPLS prediction model of ethanol stress response (ratio of growth rate under 7% ethanol to non-stress growth) based on metabolome data.

This result demonstrates that the metabolite profile of yeast strains under normal growth condition has the potential to be used in phenotype prediction. When the variables influence on OPLS modeling was closely examined using VIP (variables influence on projection) values, we could see that amino acids and many central metabolism compounds contribute greatly to the model. Figure 6 shows the metabolite list with VIP values larger than 1, which are considered important to the model. Amino adipic acid and hydroxyproline topped the list. Although there is no report about the relation of these two metabolites with ethanol stress response in yeast, increased levels of amino adipic acid in stressed rats and increased levels of hydroxyproline in stressed humans were previously reported. Amino adipic acid is an intermediate in the metabolism of lysine and saccharopine. Decker *et al.* [11] investigated the relation of amino acid levels with stress in rats. Levels of amino adipic acid in the pancreas were significantly increased (more than 2 times) in stressed rats. Similarly, hydroxyproline has been demonstrated to increase its level in urine under stressful condition among humans [12].

Tryptophan and trehalose, two known metabolites that have influence on ethanol stress response in yeast, were also identified, and this result is consistent with previous findings by other researchers. For example, intracellular trehalose accumulation was found to lead to higher ethanol tolerance [13,14] while strains lacking 'metabolism of tryptophan' genes were sensitive to ethanol [15]. Our results demonstrated that VIP analysis gives us a hint of which metabolites that may be influential or has a tendency towards ethanol stress response, and validation experiments under the specific stress condition will be needed for confirmation of these metabolites importance.

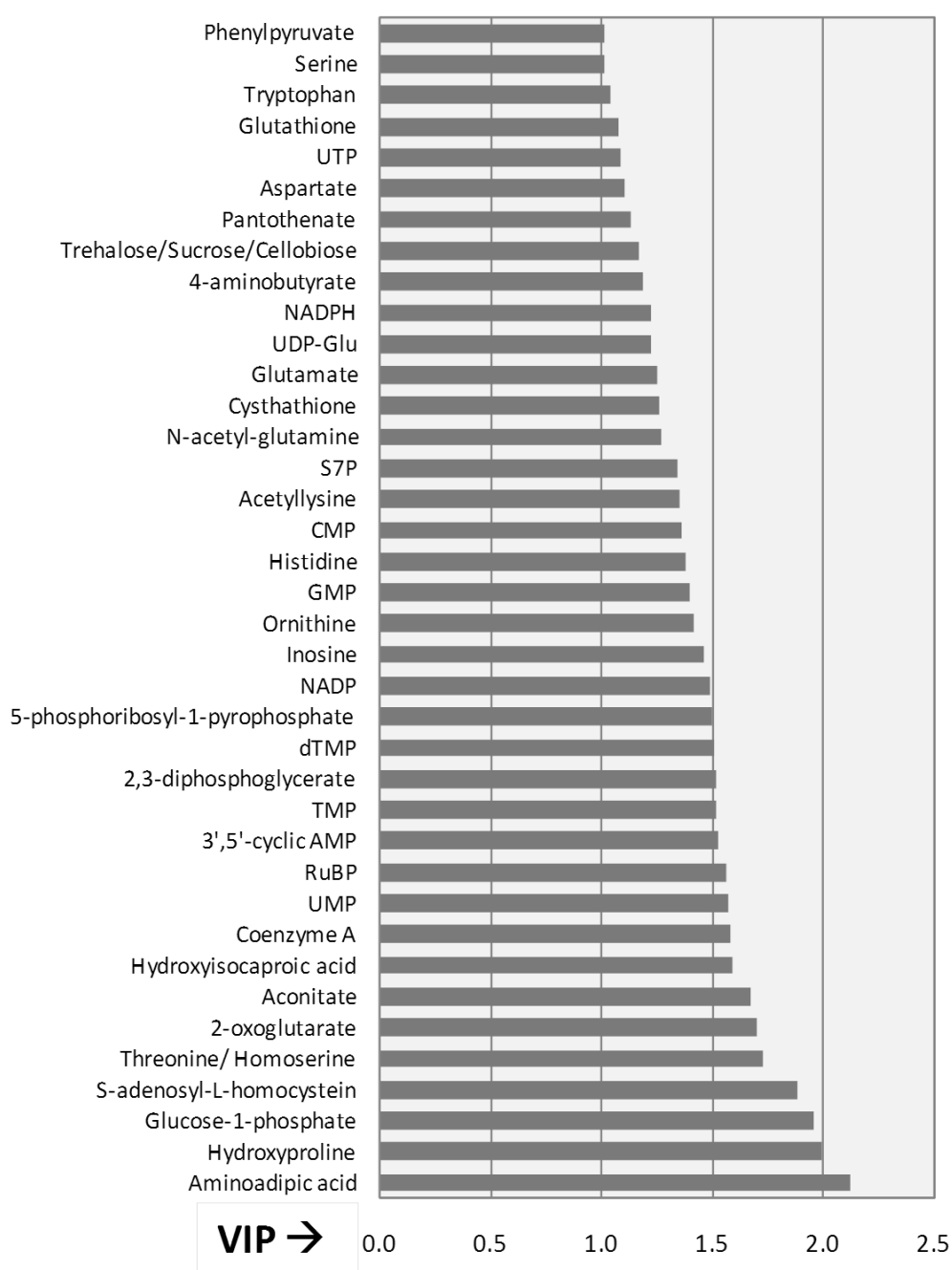


Figure 6. List of metabolites with VIP (variables influence on projection) value > 1.0, considered as influential in the OPLS prediction model of ethanol stress response.

4. Conclusion

In the current study, metabolite profiling of 30 yeast strains each lacking a single gene was performed. Using the metabolite profile, various models for ethanol stress tolerance were built. The prediction models provide information regarding which metabolites that contribute to such a behavior. From OPLS-DA model, citrate/isocitrate and tryptophan contribute to ethanol sensitivity while uridine 5'-monophosphate (UMP) contributes to ethanol tolerance. From OPLS model, aminoadipic acid and

hydroxyproline were found to contribute to the tolerance in yeast, while trehalose and tryptophan, which have been reported previously, were also found but with less importance. Taken together, these metabolites can be considered as candidates for metabolic engineering. This study serves as a preliminary screening for semi-rational selection of metabolites and/or metabolic pathways. Moreover, the metabolite profile studied was under a standard growth condition and thus may also be extended for prediction of other phenotypes.

5. References

- [1] Fortman J L, Chhabra S, Mukhopadhyay A, Chou H, Lee T S, Steen E J and Keasling J D 2008 *Trends Biotechnol.* **26** 375
- [2] Steen E J, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, Del Cardayre S B and Keasling J D 2010 *Nature* **463** 559
- [3] Fiehn O 2002 *Plant Mol. Biol.* **48** 155
- [4] Fukusaki E and Kobayashi A 2005 *J. Biosci. Bioeng.* **100** 347
- [5] Yoshida R, Tamura T, Takaoka C, Harada K, Kobayashi A, Mukai Y and Fukusaki E 2010 *Aging Cell* **9** 616
- [6] Alper H, Moxley J, Nevoigt E, Fink G R and Stephanopoulos G 2006 *Science* **314** 1565
- [7] Alper H and Stephanopoulos G 2007 *Metab. Eng.* **9** 258
- [8] Hashim Z, Teoh S T, Bamba T and Fukusaki E 2014 *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **966** 83
- [9] Lu W, Clasquin M F, Melamud E, Amador-Noguez D, a Caudy A and Rabinowitz J D 2010 *Anal. Chem.* **82** 3212
- [10] Van Hoek P, Van Dijken J P and Pronk J T 1998 *Appl. Envir. Microbiol.* **64** 4226
- [11] Decker R, Stuckrad-Barre S v, Milakofsky L, Hofford J M, Harris N and Vogel W V 1995 *Life Sci.* **57** 1781
- [12] Lee H M, Kim Y M, Lee C H, Shin J H, Kim M K and Choi B Y 2011 *J. Prev. Med. Public Health* **44** 74
- [13] Kim J, Alizadeh P, Harding T, Hefner-Gravink A and Klionsky D 1996 *Appl. Envir. Microbiol.* **62** 1563
- [14] Jung Y J and Park H D 2005 *Biotechnol. Lett.* **27** 1855
- [15] Yoshikawa K, Tanaka T, Furusawa C, Nagahisa K, Hirasawa T and Shimizu H 2009 *FEMS Yeast Res.* **9** 32

Acknowledgements

The authors would like to thank Fukusaki Lab, Osaka University Japan for the strains and technical support. Z.H. acknowledges the Ministry of Higher Education Malaysia/ University Teknologi Malaysia for financial support (Q.J130000.2744.01K94).