Journal of Bioscience and Bioengineering
VOL. 110 No. 5, 602–607, 2010

www.elsevier.com/locate/jbiosc

Genome-wide expression analysis of *Saccharomyces pastorianus* orthologous genes using oligonucleotide microarrays

Takaaki Horinouchi,[1] Katsunori Yoshikawa,[1] Risa Kawaide,[1] Chikara Furusawa,[1] Yoshihiro Nakao,[2] Takashi Hirasawa,[1] and Hiroshi Shimizu[1,*]

Department of Bioinformatic Engineering, Graduate school of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan[1] and Suntory Holdings Limited, Research Center, 1-1-1, Wakayamadai, Shimamoto-cho, Mishima-gun, Osaka 618-8503, Japan[2]

Received 11 March 2010; accepted 20 May 2010
Available online 11 June 2010

The lager brewing yeast, *Saccharomyces pastorianus*, an allopolyploid species hybrid, contains 2 diverged sub-genomes; one derived from *Saccharomyces cerevisiae* (Sc-type) and the other from *Saccharomyces bayanus* (Sb-type). We analyzed the functional roles of these orthologous genes in determining the phenotypic features of *S. pastorianus*. We used a custom-made oligonucleotide microarray containing probes designed for both Sc-type and Sb-type ORFs for a comprehensive expression analysis of *S. pastorianus* in a pilot-scale fermentation. We showed a high degree of correlation between the expression levels and the expression changes for a majority of orthologous gene sets during the fermentation process. We screened the functional categories and metabolic pathways where Sc- or Sb-type genes have higher expression levels than the corresponding orthologous genes. Our data showed that, for example, pathways for sulfur metabolism, cellular import, and production of branched amino acids are dominated by Sb-type genes. This comprehensive expression analysis of orthologous genes can provide valuable insights on understanding the phenotype of *S. pastorianus*.

[**Key words:** *Saccharomyces pastorianus*; Lager brewing yeast; DNA microarray; Orthologous genes; Metabolic network]

The lager brewing yeast, *Saccharomyces pastorianus*, is one of the most important yeasts used in beer fermentation. *S. pastorianus* is an allopolyploid species hybrid, with 2 divergent sub-genomes: one derived from *Saccharomyces cerevisiae* (Sc-type) and the other derived from *Saccharomyces bayanus* (Sb-type) (1). This yeast is characterized by the presence of 2 divergent orthologous genes, very similar to the corresponding *S. cerevisiae* and *S. bayanus* genes, for almost every gene (2). The combination of Sc- and Sb-type genes differentiates *S. pastorianus* from yeast strains as *S. cerevisiae*. For example, *S. pastorianus* efficiently assimilates maltose and maltotriose at low temperature (e.g., 5 °C–14 °C) (3,4) and produces higher amounts of sulfite when compared with *S. cerevisiae* (5). However, due to the complexity of its genome, the underlying mechanisms of such features in *S. pastorianus* are largely unknown. Analysis of the expression levels of the Sc-type and Sb-type genes is one possible strategy to understand the phenotypic features of *S. pastorianus*.

DNA microarrays are powerful tools to comprehensively analyze the intracellular state and measure the expression levels of thousands of genes simultaneously (6). Although the expression profile of *S. cerevisiae* has been extensively studied by using DNA microarrays (7,8), a comprehensive expression profiling of the *S. pastorianus* genome are poorly understood. DNA microarrays designed for *S. cerevisiae* have been used to study the expression profile of *S. pastorianus* (9). However, using a microarray designed to analyze *S. cerevisiae* genes makes it difficult to investigate differences in the expression between Sc- and Sb-type genes (10). An *in situ* synthesized oligonucleotide microarray containing probes designed for Sc- and Sb-type genes was recently used for the simultaneous measurement of Sc- and Sb-type genes in *S. pastorianus* (11). This analysis of the differences in the expression levels of orthologous gene sets suggested that the Sc- and Sb-type genes may have different functional roles (11,12). However, the microarray used in their study did not cover entire genome of *S. pastorianus* comprehensively, due to the following 2 reasons. First, the Sc-type probes were designed based on the *S. cerevisiae* genome sequence, which is not completely homologous with the sequence of Sc-type genes in *S. pastorianus*. Second, the probes for Sb-type genes were designed based on the data from cDNA libraries, and thus some genes with low expression levels may have been missing. In contrast, recently Nakao and colleagues (13) presented the genome sequence of *S. pastorianus* and designed an oligonucleotide microarray based on this genome sequence. Since DNA identities between Sc- and Sb-type genes are high (85% in average), the multiple short oligonucleotide probes for each ORF are suitable to increase the probe specificity. Thus, Affymetrix gene-chip platform was used in which 11 oligonucleotide probes of 25 base pairs are used to quantify the expression level of each gene (14). This oligonucleotide microarray contains 6662 probe sets for Sc-type genes and 5906 probe sets for Sb-type genes, respectively. This oligonucleotide microarray is a useful tool to comprehensively and quantitatively analyze

* Corresponding author. Tel./fax: +81 6 6879 7446.
E-mail address: shimizu@ist.osaka-u.ac.jp (H. Shimizu).

1389-1723/$ - see front matter © 2010, The Society for Biotechnology, Japan. All rights reserved.
doi:10.1016/j.jbiosc.2010.05.011

Sc- and Sb-type gene expression and characterize the phenotype of *S. pastorianus*.

In this study, we performed a comprehensive expression profiling of *S. pastorianus* in a pilot-scale lager fermentation process. We explored correlations in the expression levels of Sc- and Sb-type genes. In addition, we performed a genome-wide screening of genes exhibiting differences in expression levels between the Sc- and Sb-type genes. We also looked at functional categories and metabolic pathways, which are significantly different in the Sc-and Sb-type genes. We believe that our analyses will help in understanding the mechanisms underlying the differences between *S. pastorianus* and other yeast strains.

MATERIALS AND METHODS

Strain, culture conditions, and DNA microarray analysis The lager brewing strain *S. pastorianus* Weihenstephan 34/70 used in this study was provided by Fachhochschule Weihenstephan, Freising, Germany. The laboratory strains *S. cerevisiae* FY834 (MATα his3Δ200 ura3-52 leu2Δ1 lys2Δ202 trp1Δ63) was used for DNA microarray analysis. Fermenting tests were performed in wort medium and yeast was cultivated in 2 l E.B.C. (European Brewery Convention) tube at 15 °C. The initial dissolved oxygen concentration was set to 11 ppm, and the initial cell concentration was set to 7.5 g/l-wet cell weight. Apparent extract concentration, which is defined as the concentration of apparent total dry matter in the wort media, was determined empirically by measuring the specific gravity (Density Meter DMA46; Anton Paar Co., Graz, Austria) of the wort. Cells were sampled at different time points (0, 21, 45, 69, 98, 189, and 239 h) for DNA microarray analysis. Total RNA extraction, cDNA synthesis, fragmentation, labeling, washing, and scanning of the gene-chip oligonucleotide microarray were performed according to the manufacturer's instructions (Affymetrix, Inc.). The microarray used in this study was designed in CustomExpress® Array System in 49-7875 format with 11-μm spot size, which contain about 23,000 probe sets in total. The probes were designed based on genome sequence of *S. pastorianus* Weihenstephan 34/70 (13).

Microarray data analysis Expression analysis was performed using the Affymetrix Microarray Suite (MAS5.0) to make presence/absence calls for each gene (15). We used a modification of the model-based expression indexes (MBEI) method (16,17) to obtain the expression levels of each gene from the microarray raw data and evaluate the expression levels of the genes. After obtaining the expression levels of the genes, data from multiple arrays were normalized using the quantile normalization method (18).

Analysis of orthologous gene products based on functional category and metabolic network Functions of gene products were classified using the Functional Catalogue (FunCat) described in the Munich Information Center for Protein Sequence (MIPS) database (19). In order to screen functional categories where orthologous gene sets with different expression levels are significantly overrepresented, we used a hypergeometric distribution with the following formula:

$$P(X = x|N, M, n) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

where N is the total number of orthologous gene sets inspected, M is the number of gene sets related to a functional category (referred as "A" in the following part of the paragraph) in the total genes, and n is the number of orthologous gene sets with significantly different expression levels. This probability function describes the probability that we found x orthologous gene sets related to category "A" when we sampled n orthologous gene sets with different expression levels. By using this probability function, we obtain the probability p that we found more than k orthologous gene sets related to category "A" in the orthologous gene sets with different expression levels, as follows:

$$p = \sum_{i=k}^{n} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

Thus, when we found this p-value is enough small in real data, this indicates that the orthologous gene sets related to the functional category "A" is significantly overrepresented in the orthologous gene sets showing different expression levels.

Hypergeometric distribution is a probability distribution that describes the number of successes in a sequence of draws from a finite population without replacement, while binomial distribution describes the number of successes for draws with replacement. If a population size is enough large, the above two distributions are almost identical. However, in the case of small population size, the difference between with and without replacement is not negligible. In our study, we used hypergeometric distribution to avoid "double counting" of genes since the number of genes in a functional category is often small. For the same reason, hypergeometric test is widely used in microarray analysis, for example, to describe an significant overrepresentation of up/down-regulated genes in a specific functional category (20,21).

We looked for metabolic pathways where Sc-type or Sb-type genes have higher expression levels than the corresponding orthologous genes in order to analyze the roles of these orthologous gene products more comprehensively. For this analysis, we

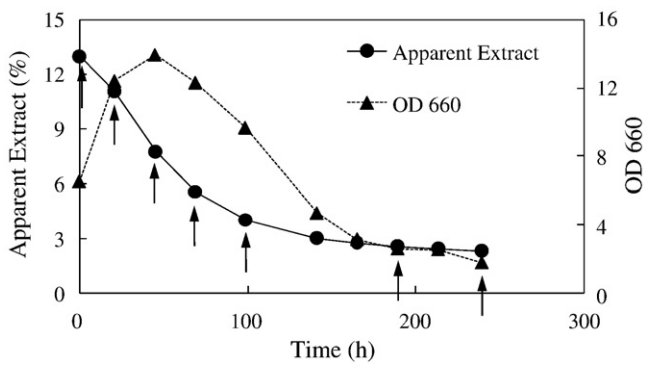

FIG. 1. Time courses of the cell growth (triangle) and the apparent extract (circle) in the laboratory-scale fermentation test of *Saccharomyces pastorianus* at 15 °C. The arrows indicate 7 sampling time points for the microarray analysis.

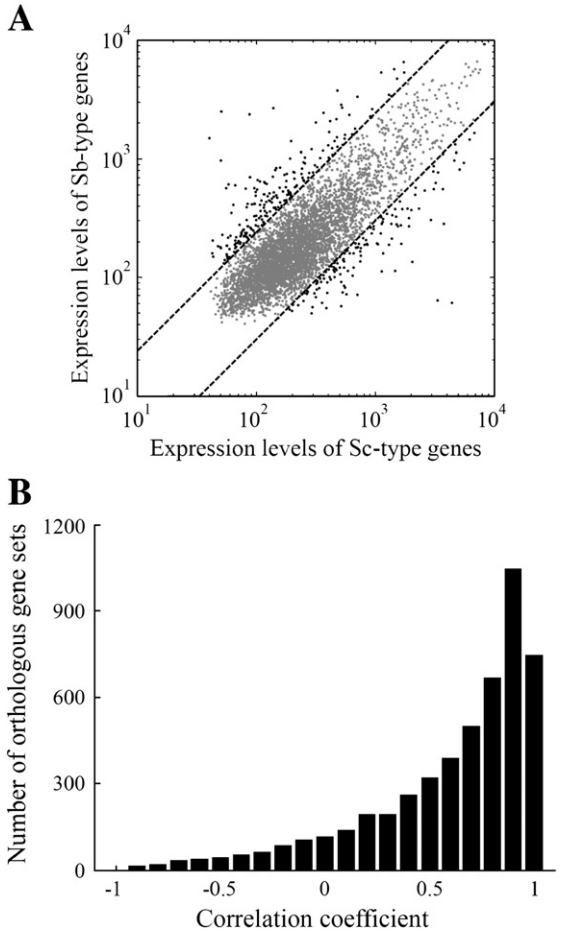

FIG. 2. The correlation between the expression levels of orthologous gene sets. (A) Scatter plot of expression levels of orthologous gene set. The gray data points close to diagonal line represent similar expression levels between the corresponding orthologous gene sets. The black points represent orthologous gene sets having top 5% and bottom 5% expression ratio. (B) The correlation coefficient between time-series expression data of orthologous genes.