Genomics of *Aspergillus oryzae*

Masayuki Machida

Research Center for Glycoscience
National Institute of Advanced Industrial Science and Technology (AIST)

Introduction

A. oryzae is one of the most potent secretory producers of proteins among filamentous fungi. This organism has been widely used for hundreds of years in Japanese traditional fermentation industries including oriental alcoholic beverages such as *sake* (rice wine) and *shochu* (spirits), *miso* (soybean paste) and *shoyu* (soy sauce). The long history of extensive use in the food industries placed *A. oryzae* on the list of Generally Recognized as Safe (GRAS) organisms by the Food and Drug Administration (FDA) in the United States of America (Tailor and Richardson, 1979). The safety of this organism is also supported by the World Health Organization (FAO/WHO, 1987). *A. oryzae* is also used in the commercial production of industrially valuable enzymes, amylases, proteases, lipase and so on. Since *A. oryzae* lacks a sexual generation in its life cycle and since it is difficult to obtain mutants due to the existence of multiple nuclei in conidia, traditional genetics is not applicable to an extensive analysis of *A. oryzae* genes. The lack of basic knowledge about the *A. oryzae* genes and their organization has become a significant barrier to expanding the application of *A. oryzae* to modern biotechnology in spite of its prominent potential in industrial use. It is believed that genomics will be a powerful tool for the *A. oryzae* research and development.

EST Sequencing

Large scale EST sequencing has been completed by the collaboration of O. Akita at the National Research Institute of Brewing (NRIB), Y. Kashiwagi at the National Food Research Institute (NFRI), T. Kobayashi at the Nagoya University, N. Kitamoto at the Food Research Institute of Aichi Prefectural Government, K. Kitamoto and H. Horiuchi at The University of Tokyo, M. Takeuchi at Tokyo University of Agricultural Technology and K. Gomi and K. Abe at Tohoku University and myself. The project was partly supported by private companies and societies in Japan, Amano Enzyme, Ozeki, Gekkeikan Sake, Higashimaru, Higeta, Kikkoman, Yamasa, Miso-Kyokai (the society of soy bean paste producing companies) and Tanekoji-Kumiai (the association of koji seed companies which produce *A. oryzae* conidiophores). Most of these companies have relations to traditional Japanese fermentation industries.

The *A. oryzae* strain RIB40 (ATCC-42149) was selected for the EST sequencing. In general, soy sauce companies have their own strains, selected after extensive breeding, most of which are patented. *A. oryzae* RIB40 is a wild type strain, similar to those used for sake brewing. *A. oryzae* RIB40 also has an ability to produce proteinases that are important for soy sauce fermentation. mRNA was prepared from *A. oryzae* mycelia grown in several different culture conditions including in complete medium, at high temperature and without any carbon source, etc. It was expected that the chance of finding new genes from a limited number of ESTs will be increased using these different culture conditions. cDNA was synthesized using oligo(dT) as the 1st strand synthesis primer that possessed *Not* I restriction site at its 5' end. *Eco* RI adaptors were ligated to both ends of the resulted cDNA fragments. After the digestion by *Not* I, the cDNA fragment was introduced between *Eco* RI and *Not* I sites of a plasmid vector. The 5' terminus of the cDNA was specifically

sequenced by the use of M13 universal primer (Fig. 1).

The total number of 16,808 ESTs have been sequenced. The total length analyzed by the EST sequencing reached 9.83 Mb. After clustering, the total number of the non-redundant sequences was approximately 6,000 with the total length of the contigs (non-redundant sequence) being 4.5 Mb (Table 1). The polypeptides encoded by 47% of the contigs had significant homology to those found in the public database by a BLASTX search. The contigs of the *A. oryzae* ESTs are searchable from the "Database of genomes and transcriptional regulations for filamentous fungi" on the web site of the Research Information Database (RIO-DB) at the National Institute of Advanced Industrial Science and Technology (AIST) (http://www.aist.go.jp /RIODB/ffdb/index.html). The EST contigs including those from solid state culture using steamed rice have become searchable from National Research Institute of Brewing (NRIB) (http://www.nrib.go.jp/ken/EST/db/index.html).

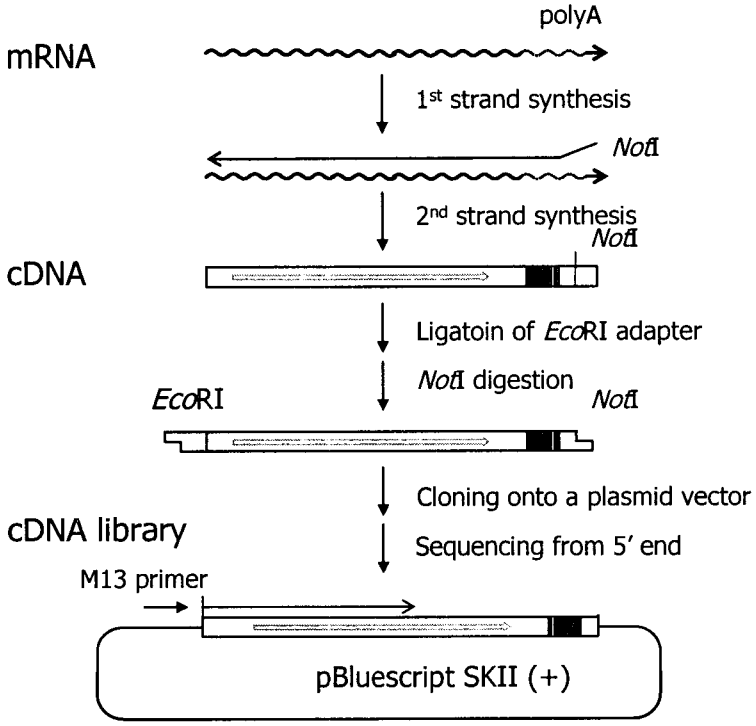

Fig. 1. Construction of the libraries for 5' EST sequencing.

Genome Sequencing

For the comprehensive analysis of the nucleotide sequence of *A. oryzae* genes and their organization on the chromosomes, *A. oryzae* genome sequencing was promoted since 1998. The *A. oryzae* genome consists of 8 chromosomes raging from 2.8 Mb to 7 Mb in length, and was estimated to have a total genome size of 35 Mb (Kitamoto et al., 1994). The shortest band at 2.8 Mb is thought to be derived from two chromosomes, VII and VIII. The second shortest band (VI) is expected to have rDNA which repeats in variable numbers as is observed in chromosome XII of S. cerevisiae (Rustchenko et al., 1993).

A whole genome sequencing project for *A. oryzae* was launched in 2001 at the National Institute of Technology and Evaluation (NITE) by the cooperation of The Consortium for *A. oryzae* Genomics consisting of the National Institute of the Advanced Industrial Science and Technology (AIST), the National Research Institute of Brewing (NRIB), the National Food Research Institute (NFRI), The University of Tokyo, Tokyo

University of Agricultural Technology, Tohoku University, Nagoya University, Axiohelix, Amano Enzyme, Gekkeikan Sake, Higeta, Intec Web and Genome Informatics, Kikkoman, Kyowa-Hakko Kogyo, Ozeki and the Brewing Society of Japan. The DNA libraries were prepared mainly by AIST and the large scale sequencing is being done by NITE (Fig. 2). The members of the consortium will focus on the analysis of gene function and the utilization of data derived from the *A. oryzae* genome.

Table 1. ***Aspergillus oryzae*** **5' ESTs analyzed to date.**

Liquid rich medium (glucose)	2,478
Liquid rich medium (long insert)	215
Liquid rich medium (high temp.)[1]	2,072
Liquid medium starved	1,790
Liquid medium starved (long insert)	163
Liquid rich medium (maltose)[2]	932
Liquid alkaline condition[3]	751
Solid phase culture[4]	6,309
Solid phase culture (low temp.)	1,049
Germination[5]	1,049
Total	16,808

[1]Kashiwagi et al. National Food Research Institute,
[2]Kobayashi et al. Nagoya University,
[3]Kitamoto et al. University of Tokyo,
[4]Akita et al. National Research Institute of Brewing,
[5]Takeuchi et al. Tokyo University of Agricultural Technology

The sequencing has been done by a whole genome shotgun sequencing approach in combination with some chromosome specific shotgun sequencing. A rough draft of the *A. oryzae* genome was completed in January 2002 by accumulating sequences of greater than 6X depth of coverage including the sequences of cosmid's ends. Tentatively, the total genome size of *A. oryzae* was estimated to be 37 Mb.

Transcription Regulation

Comprehensive analysis of gene expression using a DNA microarray is one of the most powerful methods for both investigating gene function and for discovering useful genes. A 1st generation *A. oryzae* microarray was prepared by Maeda *et al.* at Tohoku University consisting of approximately 2,000 cDNAs amplified from EST clones. Fig. 3 shows that most of the genes except fbpA, which encodes fructosebisphosphatase, were induced by glucose by the DNA microarray analysis, which is completely consistent with the results by Northern hybridization.

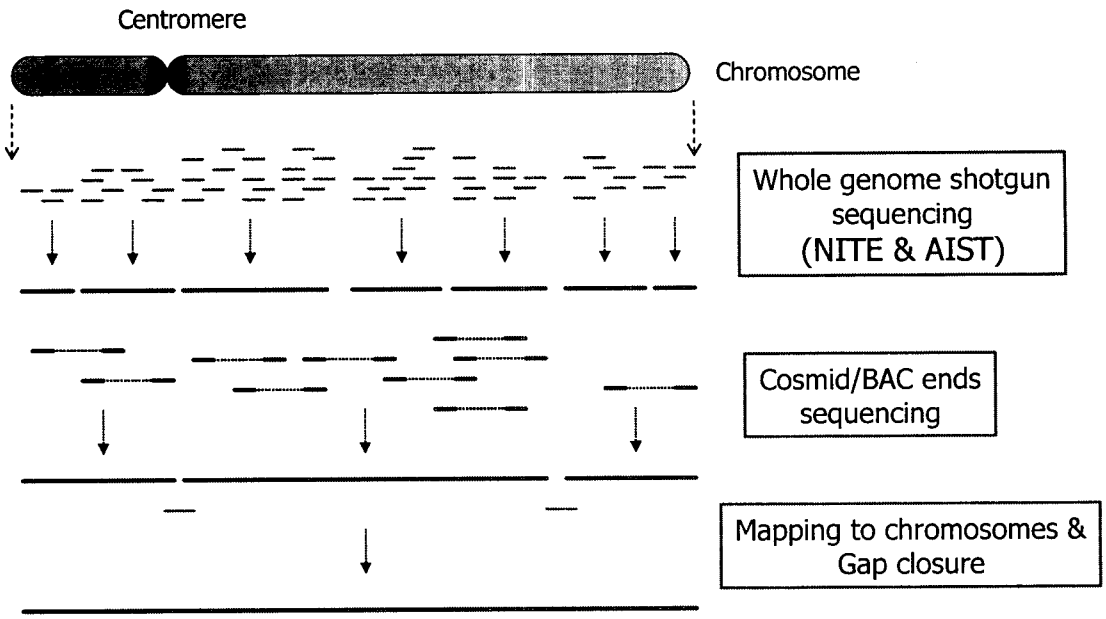

Fig. 2. Strategy for the sequencing of *A. oryzae* genome.

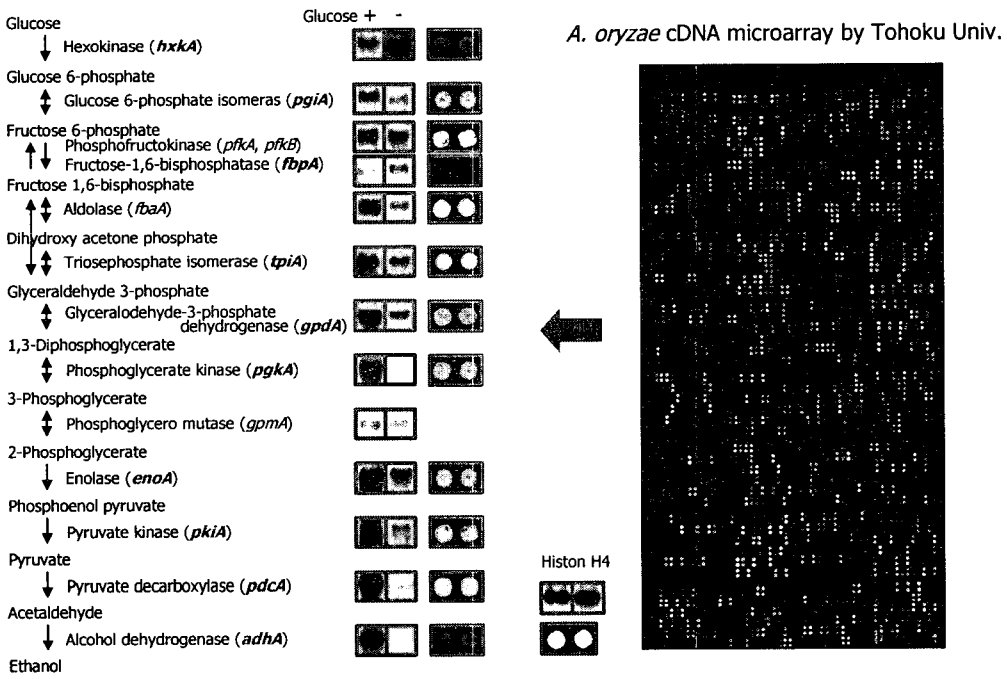

Fig. 3. Expression profiling by DNA microarray.

Analyses of the regulatory elements and the transcription factor of a particular gene are important for utilizing its function by predicting expression in the production condition. *In vitro* analysis using an electrophoretic mobility shift assay (EMSA) generates useful information about transcription regulatory elements. Scanning of the element(s) based on the sequence-specific binding of cellular factors(s) by EMSA with highly sensitive fluorescence detection has remarkable potential for the rapid determination of these elements (Sano *et al.*, 2001) (Fig. 4). The combination of the DNA microarray and analyses based on

DNA-protein interaction may be a useful way to generate information about transcription of industrially important organisms for which in vivo analysis techniques have not yet been well established.

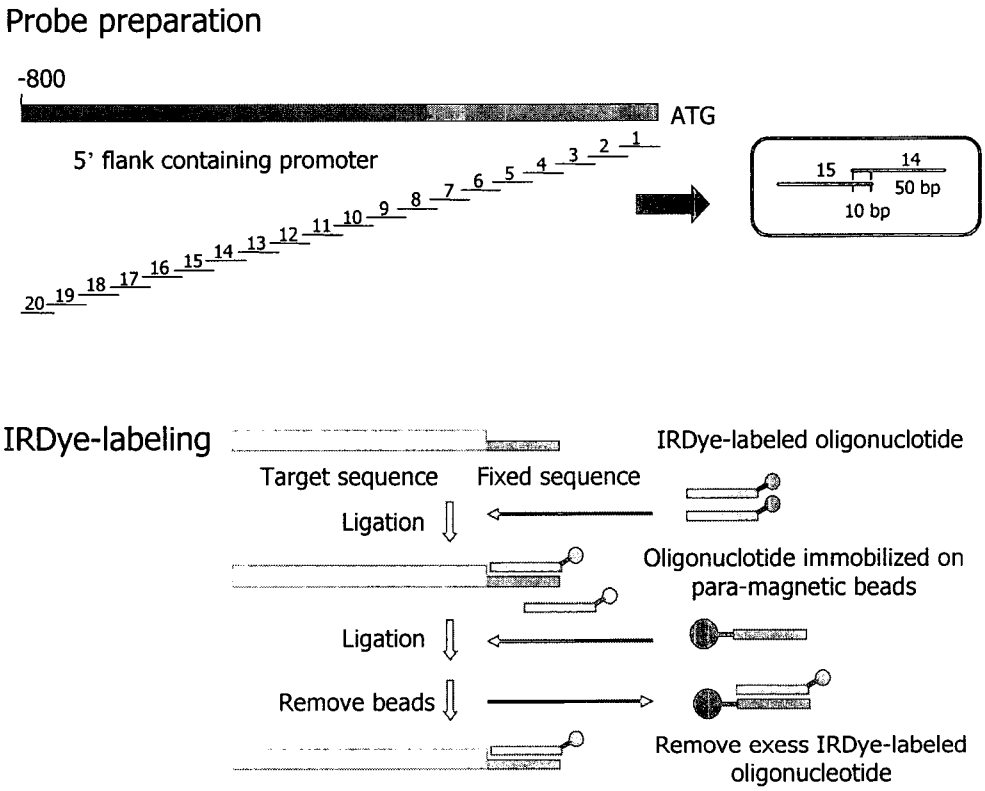

Fig. 4. Fluorescence EMSA scanning

Conclusions

Fermentation results from variety of biological functions of microorganisms, including metabolism, biosynthesis, reactions by the secreted enzymes and so forth. Huge efforts have been made to improve the fermentation efficiency, by both the breeding of *A. oryzae* and modifying the fermentation conditions. Comprehensive analyses of genes, proteins, gene expression and protein interactions will establish an indispensable knowledge for research on *A. oryzae*. The comprehensive genomic analyses will enable us to reconstruct biological systems including transcription networks and metabolic pathways *in silico*. Very recently, whole genome sequencing of several *Aspergillus* species including *A. nidulans*, *A. fumigatus* and *A. niger* as well as *A. oryzae* has been announced to be nearly completed. Comparative genomics among the *Aspergillus* species should dramatically accelerate the research of *Aspergilli*.

References

1. Kitamoto, K., Kimura, K., Gomi, K. and Kumagai, C. (1994) Electrophoretic karyotype and gene assignment to chromosomes of *Aspergillus oryzae*. *Biosci. Biotechnol. Biochem.* 58, 1467-1470.

2. Rustchenko, E.P., Curran, T.M. and Sherman, F. (1993) Variations in the number of ribosomal DNA units in morphological mutants and normal strains of *Candida albicans* and in normal strains of *Saccharomyces cerevisiae*. *J. Bacteriol.* 175, 7189-7199.
3. Sano, M., Ohyama, A., Takase, K., Yamamoto, M. and Machida, M. (2001) Electrophoretic mobility shift scanning using an automated infrared DNA sequencer. *BioTechniques* 31, 1056-1058.
4. Tailor, M.J. and Richardson, T. (1979) Applications of Microbial Enzymes in Food Systems and in Biotechnology. *Adv. Appl. Microbiol.* 25, 7-35.