Analysis of Gene Expression in *Aspergillus niger* using Microarray Technology

Nadège Lory

A Thesis

In

The Department

Of

Biology

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Science at
Concordia University, Montreal, Quebec, Canada

March 2003

ABSTRACT

Analysis of Gene Expression in *Aspergillus niger* using Microarray Technology

Nadège Lory

Aspergillus niger is an important industrial fungus. This organism is extensively used to produce large quantities of desired proteins owing to its efficient post-translational modification and secretory systems. At present, the production of foreign proteins by *A. niger* uses a gene-fusion strategy that is based mainly on the glucoamylase promoter. Detailed knowledge on the global expression pattern of this organism will help us develop other expression strategies.

Here the gene transcription profile of *A. niger* cultured under different conditions was assessed using microarrays. Microarray technology permits the simultaneous investigation of expression levels of thousands of genes. Protocols for the microarray experiments were specifically designed, executed and data were analyzed. Samples of DNA corresponding to 1,700 genes of *A. niger* were printed onto coated glass slides and hybridized with cDNA probes derived from RNA extracted from cells grown under diverse conditions. The growth media were supplemented with different carbon sources: minimal medium with glucose, maltose, xylose, xylan or wheat germ and complete medium.

Two types of data were generated from these experiments: genes that are differentially expressed under different growth conditions and genes that are highly expressed under these conditions. Eighty-eight genes were found to be differentially expressed and they were grouped using hierarchical and K-means clusterings. These data

allow us to draw preliminary conclusions about the *A. niger* transcriptome under the various culture conditions. Of the 1,700 genes, 467 were considered highly expressed. By merging with the differentially expressed gene set, a total of 39 genes were shown to display both characteristics. In future experiments, the promoter of these genes could be used to increase the heterologous protein production in *A. niger*.

ACKNOWLEDGEMENTS

There is a long list of persons I would like to thank for the help and encouragement during my Master's study.

On top of this list I would like to express my special thanks to my supervisor Dr. Adrian Tsang for his scientific guidance, support, and patience. I would like to express also my gratitude to my co-supervisor Dr. Reg Storms for his helpful advices and suggestions.

My sincere thanks go also to Dr. Bill Zerges for his help and guidance as my committee member.

I am deeply grateful to all colleagues from past and present of the Biology department. In particular, this project could not have been accomplished without the precious help of Dr. Peter Ulycznyj, Shaozhen Fang and Dr. Pascale Gaudet. My special thanks extend also to Kimchi Doquang, Rosa Zito, Barbara Decelle, Yun Zheng, Dr. Natalia Semova, Alain Bataille, Simon Drouin for their constant support and technical help.

I am thankful also to the Microarray Laboratory of the Biotechnology Research Institute (BRI) for their valuable help and advices especially from Daniel Tessier and Tracey Rigby.

Besides my grateful thanks go also to Dr. Zhong-Cheng Luo from the Montreal Children Hospital for his expertise and help in my data analysis in such short notice.

Finally I want to thank my dear husband and family for all the love, the support and the encouragements.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa	amino acids
aa-dUTP	aminoallyl-2'-deoxyuridine 5'-triphosphate
BLAST	basic local alignment search tool
bp	basepair
BSA	bovine serum albumin
cDNA	complementary DNA
CFTR	cystic fibrosis transmembrane conductance regulatory
Ci	curie
CM	complete medium
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
DD	differential display
DEG	differentially expressed genes
DEPC	diethyl pyrocarbonate
dGTP	2'-deoxyguanosine 5'-triphosphate
DMSO	dimethylsulfoxide
dNTP	2'-deoxynucleotide 5'-triphosphate
DTT	dithiotreitol
dTTP	2'-deoxythymidine 5'-triphosphate
dUTP	2'-deoxyuridine 5'-triphosphate
EDTA	ethylene diamine tetraacetic acid

EST	expressed sequence tag
GRAS	generally regarded as safe
HEG	highly expressed genes
HPLC	high performance liquid chromatography
Kbp	kilo basepair
mJ	millijoule
mM	millimolar
mmole	millimole
M	molar
Mbp	mega basepair
MM	minimal medium
MOPS	3-(N-morpholino) propanesulfonic acid
NCBI	national center for biotechnology information
ng	nanogram
nm	nanometer
OD	optical density
PCR	polymerase chain reaction
PMTs	photo-multiplier tubes
RT	reverse transcription
SAGE	serial analysis of gene expression
SBH	sequencing by hybridization
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism

SOM	self organizing map
SSC	saline sodium citrate
SSPE	saline sodium phosphate-EDTA
TAE	tris-acetate-EDTA
TBE	tris-borate-EDTA
TE	tris-EDTA
tRNA	transfer RNA
U. S. F. D. A.	united states food and drug administration
μM	micromolar
WCAB	wheat chorophyll a/b

1. Introduction

1.1 Fungi

The fungi include the common mushrooms, rusts, smuts, puffballs, truffles, morels, molds, and yeasts. About 70,000 species of fungi have been recognized and described, but some mycologists estimate that the effective total number of species could reach 1.5 million (Hawksworth, 1991; Hawksworth *et al.*, 1995).

Fungi evolved from the animal group about a billion years ago and constitute since then an independent eukaryote group equal in rank to that of plants and animals (Doolittle *et al.*, 1996). The main difference between plants and fungi is the lack of photosynthesis preventing them from producing their organic food from carbon dioxide and water. On the other hand, they are closer to animals for the ability to export hydrolytic enzymes that break down biopolymers, which can be absorbed for nutrition. Fungi live in their own food supply and simply grow into new food as the local environment becomes nutrient depleted (Ingold, 1961). The relatively close relationship between fungi and animal groups is supported by the nuclear small subunit ribosomal RNA gene (18S rDNA) sequence analysis, and also by studies of elongation factor and three other proteins: alpha- and beta-tubulin and actin (Baldauf and Palmer, 1993; Bruns *et al.*, 1991; Wainright *et al.*, 1993).

According to the systematic, the kingdom of fungi is divided into four classes: Phycomycetes, Ascomycetes, Basidiomycetes and Deuteromycetes (Fungi

Imperfecti) (Ingold, 1961). Ascomycetes or sac fungi represent the largest class of fungi (75%) and are characterized by an ascus containing spores called ascospores.

1.2 Aspergillus

In the Ascomycetes class, the *Aspergillus* genus is one of the most common fungi and is widely distributed in the world mostly because of its ability to grow on a wide variety of carbon and nitrogen sources. This genus is characterized by being an asexual saprophyte fungus producing large black or brown conidia (spores) (Raper and Fennell, 1965). *Aspergillus* genus has a long history mainly for food fermentation especially in Asia for soy sauce and miso as well as for the production of organic acids (Bennett, 1985; Samson, 1994).

The first description of this genus dates from 1727 by Micheli. By 1945, because of the flourishing studies since the middle of the 19th century, huge amounts of data and information had been generated and was gathered together in a manual called "A manual of the Aspergilli" by Thom and Raper. In 1965, Raper and Fennell expanded the original manual to include newly discovered species and the growing knowledge of this genus in their guide entitled "The genus Aspergillus". To date, the genus *Aspergillus* contains more than 200 described species.

1.2.1 Aspergillus niger

1.2.1.1 Life, growth, habitat

One of the most common of the genus *Aspergillus* is *Aspergillus niger* species. Inside this group, 15 groups have been designated including *A. awamori* and *A. niger* (which are however quite distant). *Aspergillus niger* is asexual and follows a simple life cycle hypha-mycelia-conidia-hypha. The multiplication of the fungi occurs by the production of spores, the conidiospores.

Typically the vegetative mycelium of *Aspergillus* is made up of septate branching hyphae or conidophores that are colorless or brightly colored in brown in localized areas. For *A. niger* the pigmentation of the conidophore is progressive from the vesicle toward the foot, with the black aspergilline pigment absorbed from the conidia (Raper and Fennell, 1965). This particular characteristic is the explanation for the name *Aspergillus niger*, the black Aspergilli. After this period of vegetative growth, the conidophore will grow more at the terminal portion and finally enlarge into a vesicle (Raper and Fennell, 1965). At the fertile area of the vesicle grow a layer of cells called sterigmata, which will produce the chain of conidia spores. Finally these nonmotile spores are liberated and dispersed thanks to outside dissemination means such as rain, animals or wind (Ingold, 1961). If a disseminated spore finds suitable conditions (e.g. appropriate pH, temperature, and moisture), it will germinate and grow again into hypha and will form ultimately a dense mycelium (Figure 1).

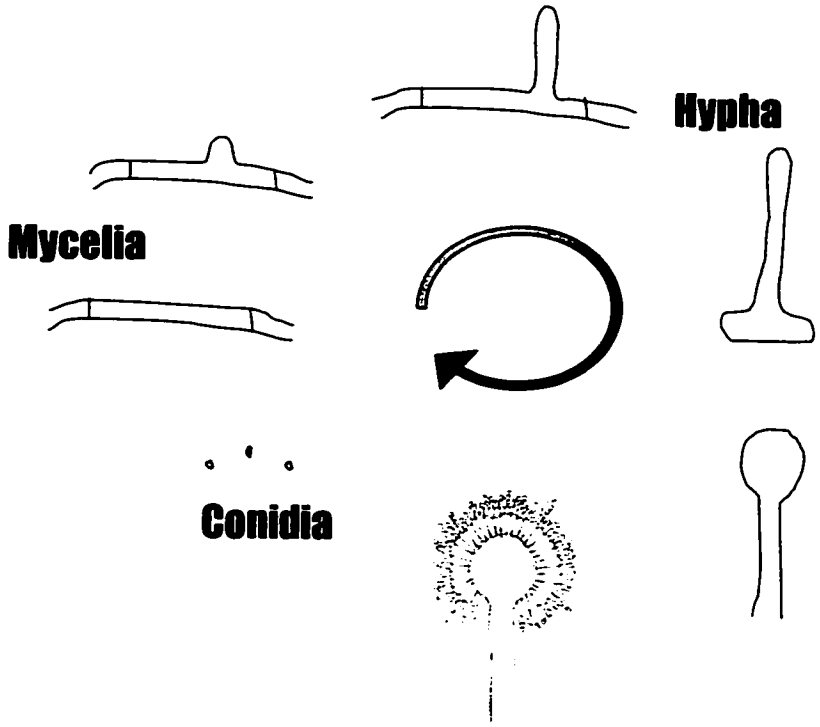


Figure 1: ***Aspergillus niger*** **life cycle (-- Hypha -- Conidia -- Hypha --)**

The mycelium is formed of hypha. After this period of vegetative growth, the terminal portion of each hypha enlarges into a vesicle. Sterigmata cells differentiate at the fertile area of the vesicle and will produce the chain of spores called conidia. These conidia are dispersed and if are found in suitable conditions, they will germinate and grow into hypha to form ultimately the mycelial mass (adapted from Thom and Raper, 1945).

Aspergilli live in a wide variety of habitats such as organic waste or plant cell material, soil, and food. Also they grow either in solid or liquid substrate and can cover a wide range of pH (from 1.5 to 9.8 for *A. niger*) (Kozakiewicz and Smith, 1994). Aspergilli are the fungi most frequently isolated from soil and they can rapidly colonize and degrade available organic matter. The abundant asexual spores are resistant to many environmental stresses including desiccation and ultra-violet light, enabling the organism to survive during inactive periods (Atlas and Bartha, 1981; Bos *et al.*, 1988). However during the mycelial stage, environmental factors such as water, temperature, pH, and gas composition (O_2/CO_2) can have a significant effect on growth. Also, added to these factors, other parameters like availability of nutrients, host resistance and interaction with other fungi can affect the production of mycotoxins (Kozakiewicz and Smith, 1994).

Aspergillus niger is not a significant human pathogen. However this organism has been involved in several cases of aspergillosis and otomycosis. *Aspergillus niger* can produce moderate to highly toxic compounds such as oxalic acid crystals, kojic acid, and malformins (Kozakiewicz and Smith, 1994).

1.2.1.2 Genetics of *Aspergillus niger*

Aspergillus niger has a compact genome of 34.5 Mbp with short distance between the genes. The *A. niger* genome encodes for about 13,000 genes (*http://www.dsm.com/newsarchive/~en/*) but only 200 genes have been reported. The genes are distributed into eight chromosomes. A genetic linkage group analysis has

been carried out to determine the chromosomal localization of a few genes including the genetic distances of specific mutations to other linked markers (Verdoes *et al.*, 1994a).

The *A. niger* sequencing project started in July 2000 by the European company DSM (*http://www5.dsm.com/en_US/html/home/dsm_home.pl*) has been successfully completed in December 2001. The complete sequence will allow a deeper understanding of gene function and gene expression, and consequently the physiology.

An important theme of research in *Aspergilli* is the analysis of the mechanisms of regulation of gene expression. Fungi, in general, have a huge enzymatic diversity and genetic and biochemical data are available for the different biological pathways. This makes them relatively attractive for the analysis of various aspects of eukaryotic gene expression (Punt and van den Hondel, 1991). Moreover molecular biological techniques are now quite developed in these organisms mostly due to the extensive analysis of the genetic model organism *Emericella nidulans* (previously referred to as *Aspergillus nidulans*) years earlier. During the last twenty years, numerous fungal genes have been isolated, characterized and further studied by genetically generating mutants. Genes encoding regulatory proteins have been isolated and confirmed by sequence structure comparison with yeast or higher eukaryotes, and also by protein/DNA binding studies. The results allow researchers to conclude the existence of a general mechanism in eukaryotes of regulation involving pathway specific and wide domain regulatory proteins (Verdoes *et al.*, 1994b). But these studies are limited at one or few genes at a time and studies using

microarray technique could give a better understanding of the global regulation and interactive networks of the genes.

1.3 Applications of *Aspergilli*

In general, fungi are widely used in industry and are an essential component for many applications in medicine, agriculture, biotechnology or basic research (Table 1).

1.3.1 Production of organic acids

As mentioned earlier, Aspergillus species have been used in the food industry and one of the eldest (used since early 20th century) and most important product of filamentous fungi is citric acid (Ingold, 1961). Large quantities of citric acid are used mainly for food, confectionery and drink industries, and also for medical purposes. The former source of citric acid was the citrus fruits but the outcome using fermentation by fungi is much more important with 500,000 tons of citric acid produced per year (Bu'Lock, 1990). A large number of fungi are able to produce citric acid in industry but *A. niger* is the species mainly used for this purpose.

Gallic acid is also one of the dominant products by Aspergillus species and is used as an antioxidant in the food industry (Miall, 1975). Finally, gluconic acid is another product of filamentous fungi that is used extensively in detergents and for other industrial applications.

Table 1: Most important fungi for main applications.

(adapted from May and Adams, 1997)

Medicine	Agriculture	Biotechnology	Basic research
Candida albicans	*Ustilago spp.*	***Aspergillus niger***	***Aspergillus nidulans***
Aspergillus fumigatus	*Magneporthe grisea*	***Aspergillus oryzae***	*Ustilago maydis*
Coccdiodes immitis	*Fusarium spp.*	***Aspergillus awamori***	*Neurospora crassa*
Histoplasma capsulatum	*Erysipe spp.*	*Penicillium chrysogenum*	*Schizophyllum commune*
Cryptococcus neoformans	***Aspergillus flavus***	*Trichoderma reesei*	*Coprinus cinereus*
Pneumocystis carnii	*Phytophthora infestans*		
	Agaricus bisporus		

1.3.2 Production of extracellular enzymes

Aspergilli have the ability to live in a variety of different habitats and to compete with other microorganisms. This is mostly due to the amazing metabolic diversity of the filamentous fungi in general (Verdoes, 1994). Indeed they can produce a lot of different enzymes, allowing them to grow on different substrates. These extracellular enzymes can be synthesized in large amounts and many of them find applications in a wide variety of areas such as in food industry (dairy, brewing, baking), in the textile and detergents fields, but also for pharmaceuticals purposes (Harvey and McNeil, 1994). Consequently, *A. niger*, *A. awamori* and *A. oryzae* have been approved by the U. S. Food and Drug Administration as GRAS (Generally Regarded As Safe) organisms. In return, this status has boosted the research to improve the different stages of the fermentation process in industry and also the development of molecular tools to genetically modify these industrial strains (Verdoes, 1994).

The first fungal enzyme used for industry was an amylase from the strain *A. oryzae* prepared by Takamine (1894) (Fogarty, 1994). Later in the 1960s the introduction of proteases in the detergents encouraged industries to develop large-scale production of fungal enzymes for other applications (Harvey and McNeil, 1994). Since then, a number of enzymes is produced massively in industry and particularly from the *Aspergilli* group (Table 2).

Table 2: Main industrial enzymes produced by *Aspergillus* species.

(adapted from Harvey and McNeil, 1994)

Enzyme	Organisms
α-amylase	*A.niger, A.oryzae*
Glucoamylase	*A.niger, A.oryzae*
Glucose oxidase	*A.niger, A.oryzae*
Lipase	*A.malleus, A.oryzae*
Neutral protease	*A.ochraceous, A.oryzae*
Catalase	*A.niger*
Hemicellulase	*A.niger*
Lactase	*A.niger, A.oryzae*
Endo-polygalacturonase	*A.niger*
Endo-polymethylgalacturonase	*A.foetidis, A.niger*
Exo-polygalacturonase	*A.niger, A.foetidis*
Oligo-1,6-glucosidase	*A.awamori*
Pectin transeliminase	*Aspergillus species*
α-Glucosyltransferase	*Aspergillus species*
β-Glucosyltransferase	*Aspergillus species*
β-1,4-xylanase	*A.niger*
Cyclohexaglucanase	*A.oryzae*
Cycloheptaglucanase	*A.oryzae*
β-1,2-glucanhydrolase	*A.fumigatus*
Anthocyanase	*A.niger*

1.3.3 Production of glucoamylase

Of all the industrial enzymes produced by the Aspergilli group, the enzyme glucoamylase (EC 3.2.1.3, 1,4-α-D-glucan glucohydrolase) produced by *A. niger* is considered one of the most commercially important. Indeed this is a major enzyme in the industry and is used for the production of glucose syrup by saccharification of starch coming from different sources like corn, barley or potato. The glucose syrup is used for the production of dextrose, fructose syrup, for the fermentation to ethanol or other food applications (Verdoes, 1994; Harvey and McNeil, 1994). This secreted enzyme is not associated with the growth of the organism and is released in the extracellular medium after growth has ceased. *A. niger* is able to produce and secrete large amounts of this enzyme. For example, the industrial microbiology company DSM has isolated a strain capable of producing 25 g/L of glucoamylase during batch culture fermentation (Finkelstein *et al.*, 1989).

Glucoamylase comprises two functional domains: a N-terminal catalytic domain (amino acid residues 1 - 440) and a C-terminal starch-binding domain (amino acid residues 509 - 616) separated by a flexible O-glycosylated linker region (Giardina *et al.*, 2001). *Aspergillus* species have two forms of the glucoamylase enzyme due to the differential splicing of an intron (Boel *et al.*, 1984). The glucoamylase form I (GAI) corresponds to the full-length protein (1 – 616 aa) and the glucoamylase form II (GAII) is missing the granular starch-binding domain (1 – 512 aa) (Figure 2). Glucoamylase hydrolyzes mainly the α-1,4-glucosidic linkages and to a lesser extent the α-1,3- and α-1,6-glucosidic linkages to release α-D-glucose from the non-

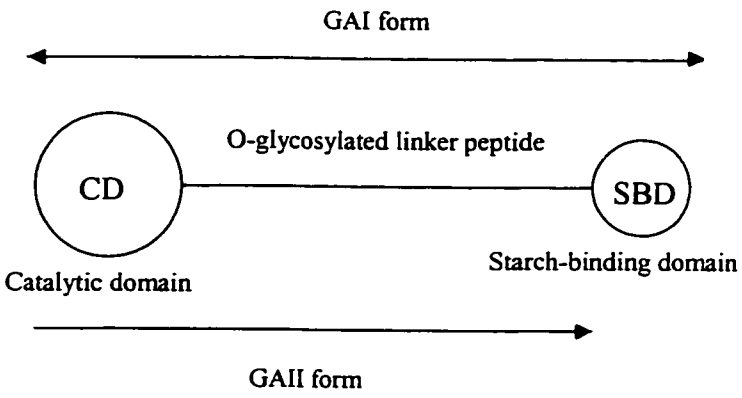


Figure 2: Schematic representation of the structure of glucoamylase protein.

The GAI form is composed of the catalytic domain and starch-binding domain linked by the O-glycosylated peptide, while the GAII form is missing the starch binding domain (Giardina et *al.*, 2001).

reducing ends of starch or other polysaccharides (Williamson *et al.*, 1997). The molecular weight of glucoamylase protein is 71,000 daltons for the GAI form and 61,000 daltons for the GAII from. The optimal pH ranges between 4.5 and 5.0 and its denaturation temperature is 70°C (Harvey and McNeil, 1994, Wallis *et al.*, 2001).

Molecular tools have been developed to overproduce the glucoamylase of *A. niger* to improve its industrial production and to increase the yield of commercially interesting homologous and heterologous proteins by a gene-fusion strategy (details in the next section).

1.3.4 Production of heterologous proteins

1.3.4.1 Advantages of *Aspergilli* as an expression system

In the past, mostly bacteria and yeasts were used for the production of proteins. *Escherichia coli* is considered as an organism of choice for the production of prokaryotic proteins and is used mostly for the commercial enzymes used in recombinant DNA technology (e.g. restriction enzymes, DNA polymerases, DNA ligases, reverse transcriptases). However for the production of eukaryotic proteins that require post-translational modifications, problems are encountered leading to the incorrect conformation of the protein. Indeed, *E. coli* lacks a significant number of enzymes involved in the post-translational processes (such as glycoproteins) and most of the extracellular eukaryotic proteins (especially the mammal proteins) synthesized in *E. coli* are inactive or partially active.

The tendency to hyperglycosylate foreign proteins by the yeast *S. cerevisiae* makes it a poor expression system for most secreted proteins because it leads to non-active or partially active proteins.

The first advantage of using *Aspergillus* species for the production of heterologous proteins is that they do not seem to exhibit problems encountered with *E. coli* and *S. cerevisiae* systems (van Brunt, 1986). Other advantages described below lead to the same conclusion and place these species as good expression systems for the production of secreted proteins (Archer and Peberdy, 1997).

First, *A. niger, A. awamori* and *A. oryzae* are considered as GRAS organisms making their use easier to obtain the regulatory acceptance for production of heterologous compounds. Secondly, these organisms have naturally the ability to produce in large amounts their own enzymes (e.g. 25 g/L of glucoamylase). Third, as mentioned earlier, *Aspergillus* species have efficient post-translational modifications enabling them to synthesize heterologous proteins in their active form. Moreover they have effective secretory systems, allowing them to produce and export the heterologous proteins in the extracellular medium. The problematic formation of inclusion bodies like in some bacteria and yeasts is quite rare and the proteins tend to be exported.

To boost the production of heterologous proteins, recombinant proteins are generated using *Aspergillus* species as hosts and the secretion is enhanced by creating gene-fusions. Basically, this gene-fusion strategy corresponds to the combination of a target gene (with industrial application) to a carrier gene (highly expressed and well processed). This leads to the large-scale production of

heterologous proteins in the native-active conformation. *Aspergillus niger* glucoamylase gene (*glaA*) is an excellent candidate to carry the target gene. Indeed the *glaA* gene is highly expressed naturally in the appropriate conditions. Also glucoamylase overproducing *A. niger* strains have been generated by introduction of multiple copies of the *glaA* encoding gene to improve the production (Verdoes *et al.*, 1993). Second, the glucoamylase protein is secreted, facilitating the recovering of the target protein. Third, the *glaA* gene is highly regulated depending of the carbon source. Highest levels of glucoamylase are produced using maltose, starch or maltodextrin. The expression level of the enzyme is decreased by about one third in presence of glucose. And finally little or no glucoamylase is synthesized in presence of glycerol or xylose (Nunberg *et al.*, 1984, Fowler *et al.*, 1990).

Therefore the *glaA* gene has been fused in front of numerous target heterologous genes, some including also a cleavage site at the junction to allow the release of the mature protein in vivo (Spencer *et al.*, 1998).

1.3.4.2 Examples of heterologous proteins expressed in *A. niger*

The main examples of successful heterologous proteins production in *Aspergillus* species are numerous (Table 3).

Table 3: Heterologous proteins produced by *Aspergillus niger*.

(adapted from Kinghorn, J.R. and Unkles, E. 1994)

		Expression signals		
Protein name	**Yield**	**Promoter**	**Terminator**	**Reference**
Calf Chymosin	>1000mg/L	Complete glaA gene		Dunn-Coleman et al (1991)
Hen egg lysosyme	1mg/L	gpdA	trpC	Archer et al (1990)
Hen egg lysosyme	12mg/L	glaA	glaA	Archer et al (1990)
Porcine pancreatic Prophospholipase A_2	10mg/L	Complete glaA gene		Roberts et al (1992)

1.3.4.2.1 Chymosin

The enzyme chymosin from mammal calf was originally synthesized in *E. coli* but the amount of protein secreted was relatively low and most of the enzyme was stored inside the organism as inclusion bodies. Later, *S. cerevisiae* was tested but results were also unsatisfactory (Emtage et *al.*, 1983; Mellor *et al.*, 1983). Finally the heterologous protein chymosin was the first successful commercial enzyme synthesized by filamentous fungi and was later recognized by the F. D. A. (U.S.A.). Basically the bovine chymosin cDNA was fused in the 3' end of the entire sequence of the *A. awamori* glucoamylase gene and was expressed in *A. awamori* (Dunn-Coleman *et al.*, 1991). The major problem encountered was the high level of proteases (mostly aspergillopepsin A) in the medium leading to the degradation of the synthesized chymosin. Berka *et al.* (1990) replaced the coding-region of the aspergillopepsin A gene (*pepA*) by gene-targeted mutagenesis, leading to chymosin yields of more than 1 g/L. More recently Tsuchiya *et al.* (1994) increased the calf chymosin production by using a fusion of the prochymosin gene with a portion of the *A. oryzae* glucoamylase gene (residues 1 - 511) lacking the starch-binding domain. Moreover the fusion with the gene precursor presented an advantage. Indeed the release of the mature chymosin from the secreted fusion protein was simply processed autocatalytically.

1.3.4.2.2 Hen Egg White Lysosyme (HEWL)

The expression of this protein was assessed first in *E. coli* and *S. cerevisiae*. The unsuccessful results lead to search another host for expression of this recombinant protein (Archer *et al.*, 1990). The HEWL cDNA was fused between the promoter and terminator of the *A. awamori* glucoamylase gene and was expressed in *A. niger*. The lysosyme secretion reached levels up to 12 mg/L, whereas yields attained 1 mg/L when the HEWL cDNA was controlled by the *A. nidulans* glyceraldehyde-3-phosphate dehydrogenase (gpdA) promoter. Later in 1993, a novel gene fusion allowed the expression to increase by almost 100-fold. The mature HEWL cDNA was fused to almost the entire *A. niger* glucoamylase gene (residues 1 to 498) including the secretion signal, the catalytic domain and most of the O-glycosylated linker region (Jeenes *et al.*, 1993). The same group showed later that the copy number of the heterologous gene had a positive effect on the secreted protein yield. The protein expression efficiency could also be affected at the transcription and post-transcription levels (Jeenes et *al.*, 1994).

1.3.4.2.3 Porcine Pancreatic Prophospholipase A_2 (PLA_2)

The production of the recombinant protein was originally developed in yeast (*S. cerevisiae*). The prophospholipase A_2 (*proPLA*$_2$) cDNA was expressed under the control of the galactose inducible GAL7 promoter, and secretion gave yields up to 8 mg/L (Bekkers *et al.*, 1991). Later the *proPLA*$_2$ gene was fused in frame with the

complete sequence of the *A. niger* glucoamylase gene allowing secreted protein yields to reach up to 10 mg/L in *A. niger* host (Roberts *et al.*, 1992). The strain was also modified to be protease-deficient (with the aspergillopepsin A gene disrupted) to prevent protease degradation of the product.

Other examples of heterologous proteins synthesized by *Aspergillus* species include the human proteins like the interferon α-2, the tissue plasminogen activator, the lactoferrin and the interleukin-6 (Broekhuijsen *et al.*, 1993, Kinghorn and Unkles, 1994).

1.3.4.3 Limitations and improvements

Emericella nidulans was the most studied filamentous fungi since the early 1940s and therefore was extensively used for the production of heterologous proteins. Later the attention was focused on three species: *A. niger*, *A. awamori* and *A. oryzae* since high levels of recombinant proteins can be obtained from them (Gouka *et al.*, 1997). The production of fungal proteins is really efficient with production levels reaching several g/L. The most highly produced protein is the cellobiohydrolase I of the improved-strain *Trichoderma reseei* with a yield of 40 g/L (Finkelstein *et al.*, 1989). The production of non-fungal proteins has always been lower and do not exceed a few ten mg/L (Jeene *et al.*, 1991, Conesa *et al.*, 2001).

Several factors at different stages can affect the foreign protein production process. First, limitations at the transcription or post-transcriptional levels can be encountered. Stability of mRNA can be influenced by the mRNA structure or by a

high content of rare codons. For example, insertion of a rare codon can lead to the brief stop of translation and subsequently increases the chance of mRNA degradation (Hentze *et al.*, 1991; Caponigro *et al.*, 1993) and/or incorrect processing of the pre-mRNA (Gouka *et al.*, 1997). Second, the translation can be problematic with inefficient translation initiation and elongation. The translocation and folding of the protein are also critical points because of important events like glycosylation, phosphorylation, subunit assembly that are necessary for the native folded form of proteins (Mellman and Simons, 1992). Finally the processing and secretion are the last decisive steps. In some cases the major limitation is at the gene expression level but most of the time, especially with *A. niger*, secretion is the major limitation. As a result the study of the secretion processes in *A. niger* and the importance of protein motifs in folding in vivo has became one of the major areas of research in the protein expression in *A. niger* (Wosten *et al.*, 1991; Gordon, *et al.*, 2000; Khalaj *et al.*, 2001). Even after proper protein secretion, the massive production of proteases by *Aspergilli* in the extracellular medium can degrade rapidly the newly made proteins (Archer, 1994).

Therefore several strategies have been developed to limit these problems but research is still in progress to improve the heterologous protein production. As mentioned earlier, the use of a carrier gene is imperative for many reasons. First, it has major stabilizing effects on the foreign gene leading to an increase in mRNA stability, and subsequently higher yields of the target protein (Archer *et al.*, 1994). The copy number of the target genes can also be increased; for example, the increase of *glaA* gene copy number from 20 up to 200-fold leads to a 10 to 20-fold

augmentation of the glucoamylase production (Verdoes *et al.*, 1993). The gene-fusion strategy has also other beneficial effects during translocation and folding. Indeed the carrier protein actually protects the foreign protein, thus the post-translational processes will occur properly and degradation will be prevented (Nyyssonen and Keranen, 1995). Also, the efficient secretion signal of the carrier protein will facilitate the secretion of the heterologous protein fused with it (Gouka *et al.*, 1997). In the case of the production of mammalian proteins, the significant difference between the fungal and mammalian glycans lead to a defective glycosylation of the foreign mammalian proteins and therefore to mostly inactive proteins. Researchers have tried to engineer fungal strains that will mimick a mammalian type of glycosylation (Maras *et al.*, 1999). Finally the use of protease-deficient strains generated by random mutagenesis can overcome the problem of degradation once the proteins are secreted in the extracellular medium (e.g. chymosin, hen egg white lysosyme, porcine pancreatic prophospholipase A_2) (van den Hombergh *et al.*, 1997; Xu *et al.*, 2000).

1.4 DNA microarray

In the past decades, different methods were developed to detect and quantitate mRNAs to lead to the global view of the gene expression profile in a particular cell type or tissue (Lockhart *et al.*, 1996; Bednar, 2000). Some of these techniques such as Northern-blot, RT-PCR and nuclease protection have the disadvantage to be inherently serial, involving a single mRNA at a time. Others like differential display

of amplified subsets of RNAs on a sequencing gel allows a broad search for expression variations but the results are not quantitative and the false positives are quite common (Liang and Pardee, 1992). The cDNAs libraries sequencing is more direct but is less sensitive when we focus on lower expressed mRNAs (Adams *et al.*, 1991; Mousses *et al.*, 1999). The Serial Analysis of Gene Expression (SAGE) method involves more complicated procedures and requires more time in analysis (Velculescu *et al.*, 1995). On the other hand, the DNA microarrays seems to be very promising by being a simple and sensitive procedure allowing the simultaneous and parallel analysis of the expression patterns of thousands of genes. Thus the architecture of the genetic regulatory networks can be discovered leading ultimately to a complete description of the transcription-control mechanisms in a cell (Schulze and Downward, 2001). Microarray technology is rapidly becoming a central platform for functional genomics.

1.4.1 The different Microarrays

Two main kinds of arrays have been developed. The first allows the spotting by a robot of cDNAs (with a size longer than 100 nucleotides) that are covalently immobilized onto a microscope slide then hybridized with a fluorescent-labelled probe (Figure 3). The first contact printing arrayer was designed and built in the laboratory of Patrick Brown at Stanford University. The second kind of array was developed first by Steve Fodor and by the firm Affymetrix (Fodor *et al.*, 1991). This method is based on a high-density spatial synthesis of oligonucleotides using

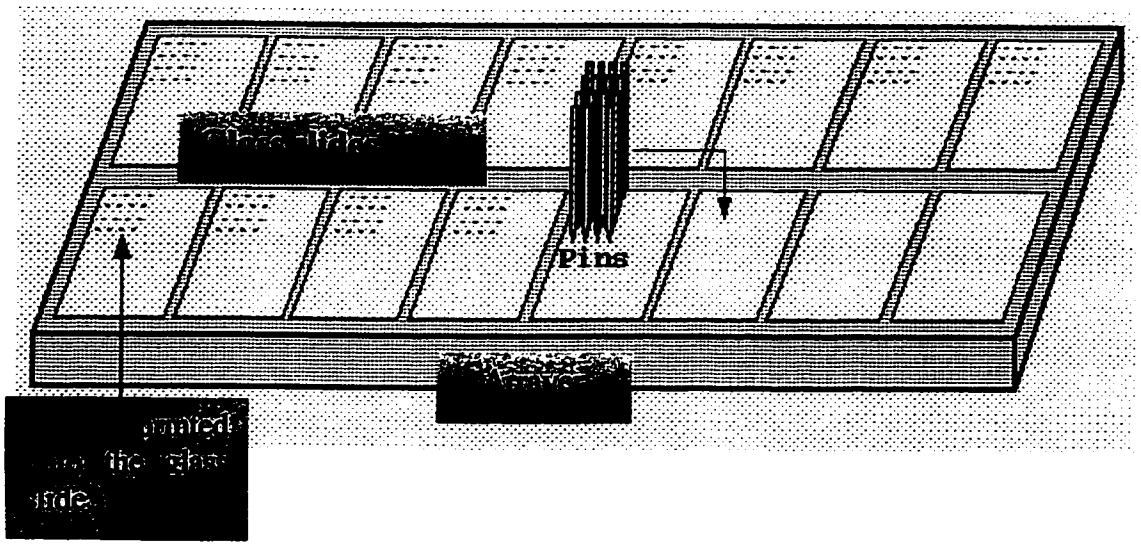


Figure 3: Schematic representation of the microarray printing of cDNAs.

The 384-well plates containing the PCR products mixed with the printing solution are loaded in the microarrayer. Typically a few nanoliters (2nl) is loaded into the spotting pins by capillary action and is deposited onto the matrix by direct physical contact between the pin and the substrate.

photolithographic masking techniques in semiconductor manufacture (McGall, 1996; Beier and Hoheisel, 2000) (Figure 4). By using a particular series of masks and light-directed chemical steps (cycle of masking-lighting-layering-incubation-washing) a defined collection of oligonucleotides (with a shorter size: up to 25 nucleotides) can

be constructed each on a 20 μm^2 space of predefined position on the array surface with each spot having a unique sequence and containing millions of identical probes (Lipshutz *et al.*, 1995; Bednar, 2000). The new generation of DNA microarrays developed by Affymetrix and Hewlett Packard called GeneChip® contains already 400,000 different oligonucleotides (Bednar, 2000). This technology can be particular attractive (easier and faster) to generate microarrays with the whole genome of sequenced organisms as well as for pathogenic organisms (Tessier *et al.*, 1997).

The cDNA arrays are more sensitive because with longer DNA sequences the hybridization is expected to have more affinity and specificity than shorter oligonucleotides (Lockhart *et al.*, 1996; Schulze and Downward, 2001). The sensitivity of cDNA microarrays is very high and a particular transcript can be detected among 500,000 others if RNA is not a limited factor (Schena *et al.*, 1996). Both strategies can be employed to determine gene expression but only oligonucleotides arrays can be used in sequence analysis and genotyping experiments (polymorphism and mutations analysis) (Khrapko *et al.*, 1989; Hacia, 1999). A survey made for the years 1999 / 2000 by the ABRF Microarray Research Group Study shows that 76% of laboratories having microarray facilities use cDNA

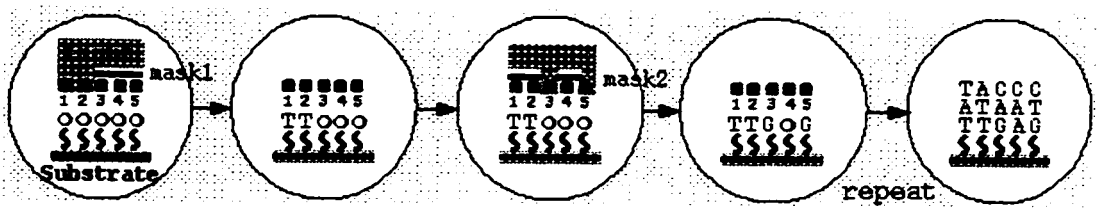


Figure 4: Principle of the oligonucleotide synthesis using photolithography.

The method combines photolithography technology from the semi-conductor industry with DNA synthetic chemistry to permit the manufacture of high-density oligonucleotide microarrays. A solid support is derivatized with a covalent linker terminated with a photolabile protecting group. A photomask is used to determine which positions react with light. Exposure to light causes dissociation of the protecting group at the exposed site. A chemical coupling reaction is then used to add a specific nucleotide to the new deprotected site, and the process is repeated using a different mask. The end result is a precision-made array, containing many oligonucleotides of known sequence.

microarrays technology, 11% use GeneChip technology (oligonucleotides chip) and 13% use both technologies (Li *et al.*, 2002).

Finally the value of microarrays for the identification of the patterns in gene expression has already been demonstrated with confidence in several model organisms: yeast, *Drosophila*, mice, and also in humans. In *Saccharomyces cerevisiae*, the gene expression patterns obtained using cDNA arrays correlate very well with the known changes in the yeast cell cycle (Spellman *et al.*, 1998), (Appendix A. Microarray Methodology).

1.4.2 Comparative expression analysis

1.4.2.1 Cellular responses to the environment

How does a cell adapt to changes in its environment? Cells can overcome different stresses and changes in their environment such as temperature, pH, nutrient availability, and the presence of toxins and ionizing radiation (Wodicka *et al.*, 1997). Usually, a change in the environment will require a modification in the gene expression profile and some genes will be induced or repressed enabling the organism to respond appropriately. In the simplest experiment, a population of cells is subjected to the stimulus and allowed to reach a steady state of transcription. Transcription levels in the altered cells can then be compared to those in a control population. The model organism yeast has been extensively studied to understand how it switches between metabolizing sugars into ethanol and ethanol, in turn, into acetic acid. This move, called diauxic shift,

involves shutting down genes for processing sugars and activating others for processing ethanol, as well as a general stress response due to the greater difficulty of deriving energy from ethanol (DeRisi *et al.*, 1997; Kuhn *et al.*, 2001). Also from these clustering studies common DNA binding motifs in the promoter regions of co-regulated genes have been identified (Cho *et al.*, 1998; Roth *et al.*, 1998).

A more complete and informative experiment is to take samples of the cell population at successive points in time (time-point experiments). In this way, the gene transcription patterns changing from the old to the new steady state can be monitored over the time. This approach confirms the up-regulation or down-regulation of the genes by showing the gradual increase or decrease in gene expression over the time (Klok *et al.*, 2002). These temporal studies can identify also the order of the changes, providing evidence about which genes are controlled by the response directly and which are only indirectly affected by it. One of the interesting applications of the time-point approach is to study the gene expression profile during the development of organisms; e.g., from the embryo to the adult (Ton *et al.*, 2002). While some of the changes may be mediated purely at the protein level, others require new transcription that can be detected by comparative hybridization.

1.4.2.2 Tissue-specific genes

Cells from different tissues are specialized for performing different functions in an organism. The role of the tissue is determined by the proteins produced, which in turn is dependent by the genes that are expressed. Generally a set of expressed genes

will be specific to a particular type of cell. Comparative hybridization experiments can reveal genes that are preferentially expressed in specific tissues.

1.4.2.3 Monitoring defects in genetic diseases

Genetic diseases are often caused by genes which are inappropriately transcribed (either too much transcribed or not enough) or which are missing altogether (Kaplan *et al.*, 2000; Tsukahara *et al.*, 2002). Such defects are especially common in cancers, which can occur when regulatory genes are deleted, inactivated, or become constitutively active (Ramaswamy and Golub, 2002). Comparative hybridization can serve two purposes in studying cancer: it can highlight the transcription differences responsible for the change from normal to cancerous cells, and it can distinguish different patterns of abnormal transcription in heterogeneous cancers. For example, microarray analysis has permitted the distinction between different subtypes of leukemia and lymphoma (Golub *et al.*, 1999). Also, cancer cell lines could be identified and classified according to their tumour origin based on their transcription profiles (Ross *et al.*, 2000). Moreover, testing drugs to verify their mechanism of action on cancers and also to find secondary drug targets and potential undesirable side effects can be predicted by microarrays (Schulze and Downward, 2000; Kurella *et al.*, 2001).

1.4.2.4 Cell cycle variations

Comparative hybridization can also be used to distinguish genes that are expressed at different times in the cell cycle. The diverse activities encountered during the cell cycle require different gene products, such as DNA polymerases for genome replication, cyclin proteins for the cell cycle check-points or microtubule spindle proteins for mitosis. Comparative hybridization analysis has been used to uncover the pathways responsible for controlling basic life processes (Spellman *et al.*, 1998).

Finally, an interesting way to exploit this comparative expression analysis is to first mutate the candidate studied gene and second overexpress it. In both cases, the effects on every single gene can be assessed by microarray analysis. This will complete the characterization of the regulatory networks that control the gene expression (DeRisi *et al.*, 1997, Brown and Botstein, 1999). This method is particularly effective in characterizing simple linear pathways but might be not so amenable for more complex combinatorial networks (Schulze and Downward, 2000).

1.4.3 The genotyping of mutations and polymorphisms

The analysis of DNA variation on a genome-wide scale is now possible using the high-density oligonucleotide arrays. The study of mutations and polymorphisms among and within species can be assessed (Chakravarti, 1999). Technically the array consists of oligonucleotides representing all known sequence variants of a gene or several genes. The genomic DNA that needs to be tested for variants or mutations is applied onto the oligonucleotide chip and the DNA will only hybridize to the complementary sequences. The technique is so sensitive that any single-nucleotide mismatch between the probe and the target can be recognized. Therefore any mutation or variation (polymorphism) in the sequence can be detected (Hacia, 1999). The first large-scale screening for mutations and variants in 1989 was to discover the different variants of the HLA-types and was blotted onto nylon membranes with detection by colorimetric assays (Saiki *et al.*, 1989). Then in 1996, the first major oligonucleotide microarrays (by photolithography) was developed to detect the mutations in the cystic fibrosis transmembrane conductance regulatory region responsible for the genetic disease (Cronin *et al.*, 1996). Recently, the technology was used for the large-scale identification and genotyping of SNPs (Single Nucleotide Polymorphisms) in the human genome (Wang *et al.*, 1998).

Another important application of oligonucleotide arrays, which is closely related to the genotyping of mutations and polymorphisms is sequencing by hybridization (SBH) (Lipshutz *et al.*, 1995; Hacia, 1999). The SBH principle is simple: all different combinations of sequences in a given length are synthesized *in*

situ onto the oligonucleotide chip. The DNA fragment that needs to be sequenced is broken in smaller pieces then labelled with a fluorophore and finally applied to the array. The smaller pieces of DNA hybridize with high specificity to the immobilized complementary sequences. Finally the sequence of the DNA fragment can be deduced from the fluorescent pattern bound to the nested sequence. Hybridization-based sequencing assays are still in their infancy and encounter difficulty like the presence into DNA of repeated and inverted sequences (Mousses *et al.*, 1999).

1.4.4 The discovery of effective antisense reagents

The microarray technology is used also to find effective antisense reagents. These are small nucleic acids sequences that have the ability to bind to their complementary target sequence of DNA or mRNA (Southern, 1997; Milner *et al.*, 1997). The nucleic acids duplexes are therefore prevented to be transcribed in mRNA or translated into the protein. Thus the targeted protein is not present in the cell and subsequently the effects can be observable *in vivo*. However obtaining effective antisense reagents is not simple and experiments using oligonucleotides arrays have been performed for this purpose. Those are made with oligonucleotides complementary to extensive regions of the mRNA target. The results have been quite encouraging because the antisense agents found via oligonucleotide array analysis were successfully verified by *in vivo* and *in vitro* tests (Sohail and Southern, 2000).

1.5 Rationale of the thesis

As discussed above, *Aspergillus niger* is an industrial fungus widely used in the production of extracellular enzymes and organic acids. At present, the strategy used to produce heterologous proteins in this organism relies mainly on the promoter of the glucoamylase gene. Genes under the control of this gene are induced when mycelia are cultured with starch or maltose as a carbon source. However, there may be situations where the presence of components present under these growth conditions interfere with applications (e.g. proteases). Transcription profiling using DNA microarray can reveal the products that are expressed under different conditions. These experiments can also be used to identify other highly expressed genes, the promoters of which can potentially be adopted to direct heterologous protein production. Moreover, exploring the gene transcription profile could lead to better knowledge about the genetic networks regulating gene activity in this organism.

The use of DNA microarrays to examine global gene expression is still an infant technology. The design, execution, and analysis of DNA microarray experiments are highly complex. Controls must be placed in different steps to ensure validity of the results. The main aim of the present work is to establish the protocols and parameters for the design, execution, and analysis of DNA microarray experiments in *Aspergillus niger*. Other goals of this project are to identify highly expressed genes and differentially expressed genes when *A. niger* is cultured under different carbon sources.

2. Materials and methods

2.1 Strains and growth conditions

2.1.1 Strains

The fungus used throughout this study was the wild type *Aspergillus niger* strain RS11.

2.1.2 Media and Solutions

2.1.2.1 Complete medium (CM)

The complete medium contains per liter: 20X salt solution, 50 ml; 2.25 M $MgSO_4$, 5 ml; D-glucose, 10 g (Bioshop); Hunter's trace elements, 1 ml; bactopeptone, 2 g; yeast extract, 1 g; casamino acids, 1 g; vitamin solution, 1 ml. The 20X salt solution used in the medium includes per liter: $NaNO_3$, 120 g; KCl, 10.4 g; KH_2PO_4, 16.3 g; K_2HPO_4, 20.9 g. The Hunter's trace elements contains per liter $ZnSO_4.7H_2O$, 22 g; H_3BO_3, 11 g; $MnCl_2.4H_2O$, 5 g; $FeSO_4.7H_2O$, 5 g; $CoCl_2.6H_2O$, 1.6 g; $CuSO_4.5H_2O$, 1.6 g; $(NH_4)_6Mo_7O_{24}.4H_2O$, 1.1 g; EDTA (Tetrasodium salt), 65 g; EDTA (Disodium salt), 7.7 g. The vitamin solution is made per liter of Pyridoxin-HCl, 0.1 g; Thiamin-HCl, 0.15 g; *p*-aminobenzoic, 0.75 g; Nicotinic acid, 2.5 g; Riboflavin, 2.5 g; Choline-HCl, 20 g; and Biotin, 0.025 g.

2.1.2.2 Minimal medium (MM)

The minimal medium contains per liter: 20X salt solution, 50 ml; 2.25 M $MgSO_4$, 5 ml; Hunter's trace elements, 1 ml. To the minimal medium is added different carbon sources per liter of minimal medium: D-glucose (1%), 10 g; D-maltose (1%), 10 g and D-xylose (1%), 10 g (all purchased from Bioshop); xylan (2%) (from Birchwood, Sigma), 20 g; wheat germ (2%), 20 g (Wheat bran Kellogg's).

2.1.2.3 Buffers and solutions

Binding buffer: 7 M Guanidine-HCl; 200 mM MES buffer, pH 5.6

Denhardts' reagent (50X): 1% (w/v) Ficoll 400; 1% (w/v) polyvinylpyrrolidone; 1% (w/v) BSA

PCR buffer (1X): 20 mM Tris (pH 9.5), 1.5 mM $MgCl_2$, 25 mM KCl, 0.05% Tween 20, 0.1 mg / ml BSA

Phosphate wash buffer: per liter, 5 ml of 1 M KPO_4, pH 8.5-8.7; 152.6 ml MilliQ water; 842.4 ml ethanol 95%

Phosphate elution buffer: per liter, 4 ml of 1 M KPO_4, pH 8.5-8.7; 996 ml MilliQ water

Saline/Tween solution: per liter, 5 g NaCl; 2 ml Tween 80

SDS (10%): per liter, 100 g of sodium dodecyl sulfate; pH 7.2

SSC (20X): per liter, 175.3 g NaCl; 88.2 g NaCitrate; pH 7.0

SSPE (20X): per liter, 175.3 g NaCl; 27.6 g $NaH_2PO_4.H_2O$; 7.4g EDTA; pH 7.4

TAE buffer (50X): per liter, 242 g Tris-base; 57.1 ml glacial acetic acid; 100 ml 0.5 M EDTA, pH 8.0

TBE buffer (5X): per liter, 54 g Tris-base; 27.5 g boric acid; 20 ml 0.5 M EDTA (pH 8.0)

TE buffer pH 8.0: 10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0

TE buffer pH 8.0 (disodium): 10 mM Tris-HCl, pH 8.0; 0.1 mM Na_2EDTA, pH 8.0

2.1.3 Maintenance of *A. niger* stock

The stock plates of *A. niger* were agar solidified medium (minimal medium + D-glucose) in which these fungi had been streaked. The plates were incubated at 30°C for 5-6 days to generate large quantities of conidia. The conidia were harvested from the stock plates by adding about 10 ml of sterile saline/Tween solution to each plate and by mixing the conidia with the solution using a glass spreader. The suspension of conidia was transferred back into the 10 ml vial, and the concentration of conidia was determined by using a haemocytometer. The suspension of conidia was kept at 4°C as a stock.

2.1.3.1 Genomic DNA extraction

In order to perform a genomic DNA extraction, 400 ml of complete medium (in a 1 L flask) was inoculated with *A. niger* conidia at a starting concentration of 1×10^6

conidia per ml. The cultures were grown at 30°C, 150 rpm overnight. The collected mycelium was filtered through Miracloth (Calbiochem), washed with sterile water (500 ml), dried between filters and weighted. The mycelium was stored at -70°C until use.

2.1.3.2 RNA extraction

Unless otherwise specified, the liquid cultures of *A. niger* were inoculated at a starting concentration of 5×10^6 conidia per 50 ml of complete medium in a 250 ml plastic flask. The flask was incubated at 30°C, 150 rpm for about 48 h with two marbles. The role of the marbles is to break the mycelium during its growth. This prevents the formation of "mycelial balls" which do not express RNA as much as the "mycelial cotton" type. After two days of incubation, the 50 ml culture was filtered through Miracloth (Calbiochem) and washed with sterile water. The mycelium was crushed using mortar and pestle and then passed through a 18G1/2 syringe (Precision Glide® Needle, Sarstedt) to homogenize the mycelial mass. The mycelium was then inoculated with the syringe in the minimal media containing different supplements: glucose (1%), xylose (1%), maltose (1%), xylan (2%) and wheat germ (2%) (in 150 ml culture volume). Another CM culture was inoculated at 5×10^6 conidia but in a 150 ml culture volume. The cultures were grown at 30°C, 150 rpm also with two marbles in 500 ml flasks for about 48 h (until the culture mass was about 100-fold that of the mycelium in the initial CM culture). The dense mycelia were then collected, filtered through Miracloth, and washed with sterile water (about

500 ml). The mycelia were squeezed to remove the residual liquid, weighted and finally stored at -70°C until use.

2.2 *Aspergillus niger* chip preparation

2.2.1 *Aspergillus niger* cDNA clones

An *Aspergillus niger* cDNA library had been constructed in the yeast expression vector pYES2 (Invitrogen Corp) using a mixed population of poly(A)+RNA. These RNAs had been extracted from mycelia cultures growing in complete medium and minimal medium. This library contained approximately 2×10^6 independent clones and 3,000 of them had been sequenced. The duplicated and overlapping ESTs were identified using FASTA program. Then over 1,700 genes were identified using the BLAST algorithm of Genbank. Approximately 3,388 analyzed clones (corresponding to the 1,700 genes) from the library were PCR amplified using the reverse and forward primers of the pYES2 plasmid. For this purpose, a small volume (5 µl) of each clone (from the plasmid mini-preparation) was diluted 10-fold by adding 45 µl of HPLC-water to obtain concentration between 10 and 50 ng/µl. Polymerase chain reaction was done for thirty-six 96-well plates in a thermocycler (Applied Biosystems GeneAmp PCR system 9700) with a final reaction volume of 50 µl per well. Each PCR reaction contained: 1X PCR buffer, 0.2 mM each dNTP; 5 units of *Taq* polymerase and 0.25 µM each of forward (pYES2) and reverse (Cyc) primers. This PCR reaction mix was added to the template. The entire mix was

denatured at 95°C for 2 min then thermacycling was performed for 25 cycles with melting, annealing, extension temperatures at 95°C for 30 sec, 50°C for 30 sec, 72°C for 1 min, respectively.

Production of the correct PCR product was verified by running 2 µl (4% of the PCR) on a 1% TBE agarose gel. The success rate of the PCR reached 93%. The 7% of the clones (about 200 clones) that could not be amplified the first time were gathered in two 96-well plates and reamplified using the same PCR master mix and cycle program. These were also run on a 1% TBE agarose gel and the final outcome of the cDNA library PCR amplification reached 96.3%.

2.2.2 *Aspergillus niger* genes from genomic DNA

To perform a genomic DNA extraction, 5 g of mycelia was grounded in liquid nitrogen (to keep the mycelia frozen and break the cells more easily) into a fine powder with a pestle and mortar. A volume of 10 ml of 10 mM Tris-HCl, 100 mM EDTA, pH 8.0 was added to the pestle, poured into a 50 ml Oakridge tube and the mortar was washed with 15 ml of this solution. To this total volume of 25 ml, 5 ml of 20% sarkosyl and 1 ml of RNAse A (10 mg/ml) were added. The tube was incubated at 65°C for 30 min, then centrifuged at 12,000 g at 15°C for 25 min. The supernatant was decanted into 1.5 ml polypropylene tubes. Extractions with equal volume of phenol and chloroform were proceeded until no protein was visible at the interface (after 3 to 4 phenol: chloroform extractions in general). The tubes were centrifuged at 12,000 g for 1 min between each extraction. The final top phase was

added to a new tube where ammonium acetate was added to obtain a final concentration of 2 M. Two volumes of isopropanol were added and the precipitation was performed overnight at -20°C. The tubes were then centrifuged at 12,000 g for 15 min at 4°C. The supernatant was discarded and the pellet was washed with 70% ethanol. The pellet was air-dried at room temperature for 5 min then resuspended in TE. Optical density at 260 nm was measured to estimate the concentration of the genomic DNA, which gave a yield of 1,074 µg corresponding to a concentration of 4.3 µg/µl. The outcome was satisfying because the expected amount is ranging between 1 to 2 mg of genomic DNA per 5 g of starting material

Published *Aspergillus niger* genes (Genbank) absent from the analyzed cDNA clones were amplified from the genomic DNA by PCR (Table 4). A total of 68 genes were found in Genbank that were missing from the analyzed cDNA clones. For each of them a pair of sequence-specific primers was designed to obtain PCR products of approximately 1 Kb (purchased from BioCorp Incorporated, Montreal). The PCR reactions were made as described earlier using the pairs of sequence-specific primers and 200 ng of genomic DNA as PCR template.

The correct PCR products were verified by gel electrophoresis and the PCR amplifications were repeated with more genomic DNA template (1 µg) if the products were missing. The same protocol as initial amplification was used for these reamplifications and the 68 genes had been successfully isolated from the genomic DNA.

Table 4: Sequence of the primers designed to amplify genes from the genomic DNA

Gene function	Gene name	Primers in the 5'end and 3'end
Phytase A (PhyA)	1-ANPHYAG	5'ACT ACA GCT TGG GTG CAG ATG ACC 3'
		5' GCA AAG AGA ATG GAG ATG ATG CCG 3'
Phytase B (PhyB)	2-ASNPHYB	5'TCA ACA TCA TCT CCG AGT CCG AGG 3'
		5' CAC AGC AGC CAT GTT CTT GTC ACC 3'
Xylanase	3-A19535	5'GGC TTT TGC AGG TCT TTT GG 3'
		5' ATT GAA GTC GCT ATT TCC GAA CCC 3'
Trehalose-6-phosphate synthase B (tpsB)	4-ANU63416	5'ACT CAA TCT TGT GGC CTC TTC TGC 3'
		5' AAC TTT CCG TTG ATT CTG CCC ACC 5'
26S proteasome subunit (tbpA)	5-ANU15601	5'CAA CGC CCT TGT CGA TAT CTT ACC 3'
		5' AAA ATG ATG GCC GGG GAG TTC TCC 3'
pgaC gene for polygalacturonase	6-ANPGAC	5'TAA GAA TGG TCC GTC AGC TTA TCC 3'
		5'CGA TGT TCT CGA GCA CAA AAT CCG 3'
pgaII gene for polygalacturonase	7-ANPGAII	5'TTC TGC CTC TCC TAT CGA AGC TCG 3'
		5' ACC GAA GGG AAG TTC TTG CAA GCG 3'
pgaI gene for polygalcturonase	8-ANPGA1	5'TTC GCT AAG AAG GCC TCT ACC TGC 3'
		5' GAA GGA ACG TTC TCG CAT TTA TCG 3'
Glutamine amidotransferase (trpC)	9-ANTRPCA	5'GAA GAT TGC GGT CGA TCT TGG AGC 3'
		5' AAG GCT CGA ATT TTG CCA AGA TCC 3'
Rhamno-galacturonase (rhgA)	10-ANRHGAGEN	5'CCG ACG GAG GTA ATA TGA TTC TGG 3'
		5' CCA AAG TCC GTG GTG AGA TCT GCT 3'
Endopoly-galacturonase (pgaE)	11-ANPGAE	5'TGG ATC TGA CCG ATC TGA ACG ACG 3'
		5' TGC TAA CAG CTG ATA TCG TCC GGC 3'
cAMP-dependent protein kinase catalytic subunit (pkaC)	12-ANPKAC	5'TCA TAC GCA ATC TCA CCA TGC CGG 3'
		5' CTA GAA GTC TGG GAA CAA ATG GCC 3'
Beta-xylosidase (xlnD)	13-ANZ84377	5'AAG CCG GAT ACA AAG TCA ACT TCG 3'
		5' GTA TTT TCT ACT CCT TCC CCG GCC 3'
Ferulic acid esterase A (faeA)	14-ANFAEAGEN	5'TCT TTT GGC GAC TGC AGG ACA AGC 3'
		5' AAG TAC AAG CTC CGC TCG TCA TCC 3'
Mitochondrial cobA gene for apocytochrome b	15-ASNMTCOBA	5'TAT GAT GGC TAC AGC GTT CTT AGG 3'
		5' TGG AGT TTC AAC GTG TTT TGC ACC 3'
exoB	16-A27977	5'TGC TGA TGA GGG ATA TAA CTC CGG 3'
		5' CAA ACG CCG TCT CAA TCT CAA AGC 3'
exoA	17-A27979	5'GCT GGA AGT GGA ACG AGA CAA TCG 3'
		5' CGG CAA ACT AAA GGT AAA AGT GCC 3'
Xylanolytic transcriptional activator (xlnR)	18-ANAJ1909	5'GCA TTA ATG TCA CCG AAG AGG AGC 3'
		5' TGG TAC TCC GTG TTC AAG GTC ACG 3'
Regulatory gene (areA)	19-ANAREAGEN	5'AGC ACC GTA AGC ATG TGA CGA TCG 3'
		5' ACA GAC TCA TAG TCA GCC ATT CCC 3'
(1,4)-beta-arabinoxylan arabinofuranc hydrolase (axhA)	20-ANAXHA	5'AAA TTC CTC AAA GCC AAG GGT AGC 3'
		5' CTG CTT CAA GGT AAG AAC TCC TGG 3'
Chaperone (bipA)	21-ANBIPAGEN	5'CCT CTT CAA GAA GAC CCT GAA GCC 3'
		5' AGC AGG TTG TTG TCC TTG GTC AGG 3'
Leucine zipper (cpcA)	22-ANCPCA	5'ATC GGC TCC AGT GAA TTC GAT TCC 3'
		5' GAT CAG TTA GTG CCA GTC TTG GCC 3'
Glucose oxidase (GO2)	23-ANGOD	5'CCA ACA ATC AAA CCG CGC TGA TCC 3'
		5' TTG AGC AGC TCG TGT GCC TTT TCG 3'

Table 4 (continued): Sequence of the primers designed to amplify genes from the genomic DNA

pacC gene	24-ANPACCGN	5'TCA GTT CGA CCC GAA CTC CTA TGC 3'
		5' TGA ATC ACC ATC TTC TAC CAC CGC3'
Phospho-fructokinase (pfkA)	26-ANPFKA	5'GAA TGC GCA TCG GTA TTA TTC ACG 3'
		5' ATT CTC ATA GCA ACT CCA CGT CGC 3'
tigA gene	27-ANTIGAGEN	5'CTG GTT GAA TTC TTC GCT CCC TGG 3'
		5' CTT CTC CTC ACC AAC AAA CTT GCG 3'
Rhamno galacturonase (rghB)	28-ANRHGBGEN	5'ATG GGC TCG TTT GGC TCT GAT ACC 3'
		5' CTT TAC ACG TAG CAG ACA CCC TGC 3'
Uracil phosphoribosyl transferase (furA)	29-AY017410	5'CAA CTG AAT CAA TCA TGG CCC TCC 3'
		5' ACT TCC CCA TCG CAA CAA ACA ACC 3'
1,4-beta-D-glucan cellobiohydrolase A precursor (cbhA)	30-AF156268	5'AAA CGG CGC GTT GTA CTT CAC TGC 3'
		5' GAC CGA ACT TGA TGT TTG AGA AGG 3'
1,4-beta-D-glucan cellobiohydrolase B precursor (cbhB)	31-AF156269	5'TCG CGA CCT GAA GTT CAT CAA CGG 3'
		5' AGT AGT ACG CAT TCT CAT ACG TGC 3'
nsfA gene	32-AF244544	5'AGA TGC GGG ATA ACG ACG TCA TGG 3'
		5' GCC TCA ATA GCG TTG TTG ATA ACC 3'
yptA gene	33-AF244545	5'TCT TTT CAA GTT GCT CCT CAT CGG 3'
		5' CAG TTG GCT TGT TGT TGA CAG TCG 3'
Amidase-B (gmdB)	34-AF349512	5'CCA AGT CAG CTT CCT TTA CGA AGG3'
		5' TGC GAA TAT CCC GAT CAA TGA TCC 3'
Oligosaccharyltransferase alpha subunit (ostA)	35-AY017409	5'TAC ACG ACA ACG ACC CAG AAG ACC 3'
		5' CTC ACG TCA ATA TTT CCA ATC GCC 3'
Putative DNA binding protein (facB)	36-ANU56099	5'AAT CGG AAG GAG CTG GTT TTC GCC 3'
		5' ATC TTC CAC TTC GTC TTC CAC CGC 3'
Inulinase (inuA)	37-ABO12771	5'GTA TTC TTC CAT CGC CAA TCG GGG 3'
		5' TCA AGT GAA ACA CTC CGC ACG TCC 3'
Inulinase (inuB)	38-ABO12772	5'CAG GAA GCA CCA CAT GAT ATA ACG 3'
		5' AGT GAA ACA CTC CGT ACG TCC ACC 3'
Zinc finger protein (creA)	39-ASNCREA	5'AAT TCC ACT GAT TCC ACC CCT TCC 3'
		5' CCA CTA CCG GGA TTT AAC ATA TCC 3'
Kexin (kexB)	40-ANI18127	5'CCT ACG AGA CCC ATG ATT ACT TCG 3'
		5' CCG TGG TAT GAA TTG CAT CAC TGG 3'
Citrate synthase (citA)	41-ANI243204	5'TTC ATT GAG GAG CTC ATC GAC CGC 3'
		5' AAG AGG ACG GTG TAG TAG TTG GCC 3'
Xylanase-B (xynB)	42-ASNXYNBA	5'CTC ACC AAG AAC CTT CTC CTC TGC 3'
		5' CTG AAC AGT GAT GGA CGA AGA TCC 3'
Argillopepsin A (pepA)	43-ASNPEPAB	5'GGG CTT CAC CAT CAA CCA GAT TGC 3'
		5' TTA GGT CCC TCA GAG TTG AAG ACC 3'
D-xylose reductase (xyrA)	44-AF219625	5'CCC ACA GTA AAG CTA AAC AGC GGC 3'
		5' CAA AGG ATC ATT GAA CCG CAA CCC 3'
Annexin XIV-like protein	45-AY033935	5'CAA CAA TAT CCA CAA TAT CCC CCG 3'
		5'TAC ATT CTC TCA AAC TCC CCA GCC 3'
Copper amine-oxidase	46-ANU31869	5'CGA AGA GGA GAC CAA CAT TGC TCG 3'
		5' GTT AGG CAT CTT CTT GGC CGA ACC 3'
Alternative oxidase (aox1)	47-ABO46619	5'TCG AAA TTA CTC CGG TGT CAT TGC 3'
		5' TCG ATC GCC TGG TTT AGG TTT CCC 3'
Pyruvate carboxylase (pyc)	48-ANI9972	5'CCA AAA GAT CCT TGT TGC CAA CCG 3'
		5' ACC AAC ATG GAG TCG TAG TGA GGG 3'
Endoglucanase precursor (eng1)	49-AF331518	5' TGA GTT TGG GAC CAA TAT TCC TGG 3'
		5' AGA TAT GCC TCC AGG ATA TCC AGC 3'

Table 4 (continued): Sequence of the primers designed to amplify genes from the genomic DNA

Oligosaccharyltransferase alpha subunit (ostA)	50-AY017409	5'AGC CTC CCC AAG TGT TTA AGA ACG 3' 5' ATT CCA CGC GGT TGA AGT TGT CGG 3'
Uracil phosphoribosyltransferase (furA)	51-AY017410	5' TAT CTT GGG CGT GGA AGC ATT TGC 3' 5' ACT TCC CCA TCG CAA CAA ACA ACC 3'
Pectate lyase A (plyA)	52-ANI276331	5'ACC AAC TTC AAG TGG ATC GTT GCC 3' 5' AAG TAA GAG TCT GAC CGG CAG TGC 3'
cAMP dependent protein kinase regulatory protein (pkaR)	53-ANI296317	5'GGT GGG CAC AAA AGA CGA CAA AGC 3' 5' TTC GAT TCG TAT TCG GCT CGT CGC 3'
Alpha-glucuronidase (aguA)	54-ANI290451	5'TTA ATA CTG GCT CTT CTG CTG TCC 3' 5' CAC GAC GTT CTC CTC GAA TTG ACC 3'
Secretory pathway Ca2+-ATPase (pmrA)	55-AF232827	5'CGC TGC GAT CTC CTT TTT CAT GGG 3' 5' ATT ATG GAC CTC AAA AGG CTC GGG 3'
Prolyl aminopeptidase (papA)	56-ANI315565	5'TCG GCG CTT CCA TAA TAT TCC CGG 3' 5'GCA CGT AGC ACG AGA TCT ATG TGG 3'
O-mannosyl transferase (pmtA)	57-AF396953	5'TAA ATC AGG CGT CAA GCT CTG GGG 3' 5' GAC AGA ATG TAG ACA ATG CAG GGC 3'
Aminopeptidase (aspA)	58-ANI292570	5'ACG AGA TCA AGA TCC ACA ATG CCG 3' 5' CAA CGG GGA AAC CGA CTT TCT TGG 3'
Heat shock protein (sspB)	59-ANSSPB	5'TAT CTC CAT GAT TGG TCA GTT CGG 3' 5' TCG AAC TCC TTC TCC TCC TTC TCG 3'
Beta-fructofuranosidase (fopA)	60-ABO46383	5'ACC TTA CTC TAC ACC TCC GTC TCC 3' 5' AAG AAC TTG GAG CTC GGC GAT TGC 3'
Benzoate-para-hyrdoxylase (bphA)	61-ANI347748	5'ATC TCC TTC CTT ACC TAA AGC GCG 3' 5' TAC GCA TGG TTT CCC AGA TGA CCC 3'
Alpha-L-arabinofuranoside (ABF2)	62-ANU39942	5'TGC CCT CTA CCA AAT CAA ACG AGG 3' 5' ATC CTC ATG GAA CTG CTT GGT GCC 3'
Endopolygalacturonase (pgaE)	63-ANIY18804	5'TGC CAA GCC TTT GTT CTG CCT TGC 3' 5' AGG ATG TAG ATC TCG ACA GCG TCG 3'
Endopolygalacturonase (pgaB)	64-ANIY18805	5'CTC TAG CTC CCA TAG TAC TCT TCC 3' 5' ATA CAA ACA CTC CCG GCG TGA TGC 3'
Endopolygalacturonase (pgaD)	65-ANIY18806	5'CTT TCC AAC TGT GAC TAC AAG ACC 3' 5' AAG ATA GGA TGG ACT AAC CTC GCC 3'
Beta-glucosidase	66-AF121777	5'ACG GGT AAT ACG ACT CAC TAT AGG 3' 5'AGG ATT GGA ACC CGC ATC TTC TCC 3'
Alpha-glucosidase (aglA)	67-D45356	5'TTT GGC CAG GAT ATA CAG TCT TCC 3' 5' ATC GTA CCA CAC TTC GCC ATG TCC 3'
Rhamnogalacturonan-acetyl esterase	68-ANI242854	5'CAC TAA CAA ACA ATC AGG CTG GGG 3' 5' ACT CAT CCA ACT CCA AAT CAT CCG 3'

PCR products obtained by amplification of the *A. niger* cDNA library and genomic DNA were purified through 96-well Millipore FB plates. A volume of 50 µl of Binding buffer was added into each well of the 96-well purification plates. Then the PCR product (50 µL) was added and mixed with the Binding buffer by pipetting up and down. The FB plates were placed into a Beckman AVANTI™ J-201 centrifuge with a JS-4.3 rotor and centrifuged at room temperature for 5 min at 1,000 g. The flow-through was discarded, 200 µl of 80% ethanol was added to wash the plates, then centrifuged at 1,000 g for 5 min. The wash step was repeated with centrifugation at 1,000 g for 10 min, and two subsequent spins at 1,000 g for 5 min. Purification plates were allowed to dry at room temperature for about 45-55 min, and 50 µl of disodium TE buffer was added to elute with a final centrifugation at 1,000 g, for 5 min. The purified PCR products were verified by agarose gel electrophoresis and ethidium bromide staining and the recovery rate was always satisfying.

2.2.3 Printing of purified DNA

A volume of 5 µl of the purified products corresponding to the set of 3,388 clones from the *A. niger* cDNA library and the genes amplified from the genomic DNA (68 genes) were transferred from the thirty-six 96-well plates to nine 384-well plates. Moreover negative controls have been added to the plates for the printing including no DNA, plasmid pBluescript, 5 wheat genes encoding for protoporphyrin IX magnesium chelatase, two chlorophyll a/b binding protein WCAB precursor, oxygen-evolving enhancer protein 3.1 chlorophyll precursor and photosystem type

III chlorophyll a/b binding protein. The corresponding sequences were compared by BLAST to the *Aspergillus niger* database to confirm that no homology existed between these sequences. A text file was generated to keep track of the order and location of the samples. The DNA samples were mixed with 5 µl of a printing solution of DMSO and salts (90% DMSO; 40 mM Tris-HCl, pH 6.5; 100 mM KCl). The printing solution was prepared with DMSO because it promotes denaturation and reduces the evaporation of samples to be printed. The salts were added to the DMSO as suggested by the Microarray Laboratory of the Biotechnology Research Institute (BRI personal communication) to improve the DNA printing. The plates were then centrifuged quickly (1 min at 1,000 g) in the Beckman AVANTI™ J-201 centrifuge with JS-4.3 rotor to collect the solution to the bottom of the well.

The glass slides used for microarray construction were the amino-silane coated glass slides CMT-GAPS II by Corning Incorporated Life sciences (NY, U.S.A.). The GAPS surface provides free amine groups for ionic attachment of the negatively charged phosphate groups of the DNA backbone. The PCR products were spotted in duplicate (spots side by side) onto the slides at 22°C and 45% relative humidity using the SDD2 microarrayer (Engineering Services Inc., Toronto) with Telechem Stealth SMP3 pins. Following printing the slides were allowed to dry for several hours at room temperature. The spotted DNA was bound to the slides by UV-crosslinking at 90 mJ using Stratalinker™ (Stratagene) and baked at 80°C for two hours in a Hybaid Mini Hybridization Oven. The printed slides were stored in a light-tight box at room temperature until used for hybridization.

2.3 Fluorescent-labelling probe preparation

2.3.1 Total RNA extraction

About 1 g of the mycelia stored at -70°C was grounded in liquid nitrogen into a fine powder using a mortar and pestle and transferred into a 15 ml tube. The grounded mycelium was weighted and 1 ml of TRIzol® Reagent (Invitrogen) was added per 100 mg of mycelia (TRIzol® Reagent is a mono-phasic solution of phenol and guanidine isothiocyanate). The RNA was extracted in aliquots of 1.5 ml Eppendorf tubes following the instructions of the manufacturer. The final wash of the RNA pellet was made twice with 70% ethanol. The pellet was then dried at room temperature for about 30 min and was resuspended in 20 µl of DEPC-treated water. The extracted RNA aliquots were gathered and stored at -70°C.

2.3.2 DNAse-treatment and purification of total RNA

After RNA extraction by TRIzol reagent, the RNA samples extracted from the minimal media containing glucose (1%), maltose (1%), xylose (1%) and wheat germ (2%) and the complete medium were treated with DNA-free™ kit (Ambion) using the protocol provided in the kit. The RNAse-free DNAse enzyme is especially designed to remove any trace amounts of DNA from RNA preparations. The last step of centrifugation allowed to block the enzyme and divalent cations in the

DNAse Inactivation Reagent pellet where as the RNA in the aqueous phase was transferred to a fresh tube.

The RNA sample extracted from MM+xylan (2%) was passed through silica column from the RNeasy kit (Qiagen) following the protocol. This step allowed the purification of RNA from all contaminants (including DNA) and also for desalting. This RNA sample was treated differently compared to the other RNA samples. The next step of reverse transcription was performed correctly only when the sample was purified by silica column. Unfortunately the yield of the purified RNA was considerably diminished from 50% to 70% of loss.

2.3.3 RNA quality control assays

The quality of the RNA is very important because it can have significant adverse effects on the results and hence the degradation state must be evaluated (Loriod *et al.*, 2001).

The concentration and purity of RNA were determined by measuring the absorbance at 260 nm and 280 nm in a spectrophotometer. Absorbance readings at 260 nm measured the RNA concentration and should be greater than 0.15 to ensure accuracy. An absorbance of 1 unit at 260 nm corresponds to 40 μg of RNA per ml. The ratio between the readings taken at 260 nm and 280 nm provides an estimate of the purity of RNA and pure RNA is supposed to have a ratio comprised between 1.8 and 2.1. To detect any RNA degradation characterized by a lower smear (below the 28S and 18S ribosomal RNAs bands), an 1.4% agarose RNA-gel electrophoresis

(MOPS/EDTA/Formaldehyde) prepared according to the "Molecular cloning" protocol (Sambrook and Russell, 2001) was run. Also the resulting signal intensity of the 28S band must be twice that of the 18S band to assure good quality RNA.

Finally the RNA quality was also assessed by using the Agilent RNA 6000 chip with the Agilent 2100 Bioanalyzer software for analysis (Agilent Technologies, CA). The microcapillary chip had its micro-channels filled with a sieving polymer and fluorescence dye. The RNA samples were loaded in the sample-wells of the chip. Then the pin-electrodes created electrokinetic forces capable of driving fluids or molecules within fluids, and the samples were electrophoretically separated. The samples were detected by their fluorescence and translated into gel-like images (bands) and electropherograms (peaks). The advantage of using this system is that this miniaturized fluid pathway reduces the time of determination (about 7-8 min per chip as compared to 5 hours per gel), and the strong electrokinetic driving forces improves the resolution.

2.3.4 Aminoallyl incorporation during reverse transcription

The aminoallyl incorporation and next steps were performed following the instructions provided in The Institute for Genomic Research TIGR protocol (*http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml*). The RNA was reverse-transcribed into cDNA with incorporation of a modified nucleotide aa-dUTP (aminoallyl-dUTP). To 25 μg of total RNA, 2 μl of oligo-dT(18) (2 μg/μl) was added and the final volume was brought to 18.5 μl with DEPC-treated water. The

RNA and oligo-dT were mixed and incubated at 70°C for 10 min. After snap-frozen the tubes in dry-ice-ethanol bath for 30 sec, the samples were briefly centrifuged and the following reagents were added: 6 µl of 5X First Strand buffer, 3 µl of 0.1 M DTT, 0.6 µl of 50X aminoallyl-dNTP mix (prepared with 5 µl of 100 mM dATP, 5 µl of 100 mM dCTP, 5 µl of 100 mM dGTP, 3 µl of 100 mM dTTP, 2 µl of 100 mM aa-dUTP to a final volume of 20 µl), 2 µl of Superscript™ II RT (200 U/µl; Life Technologies).

The samples were mixed briefly and then incubated at 42°C for approximately 3-4 hours. The RNA was hydrolysed by adding 1 µl of RNAse A (4.53 mg of protein/ml) and 1 µl of RNAse H (5 U/µl) and incubated at 37°C for 15-30 min. The hydrolysis using RNAses was suggested by the BRI to improve the probe quality. The samples were purified to remove unincorporated aa-dUTP and free amines by using the PCR purification kit (Qiagen) following the manufacturer's protocol but with the wash and elution buffers prepared as described in the TIGR protocol. The product of reverse transcription was verified by running 10% of the elutant on a 2% TAE agarose gel. Bright smears comprised between 400 bp and 2,000 bp corresponding to the synthesized cDNAs were observable.

2.3.5 The coupling of aa-cDNA to cy dye ester

The aminoallyl-labelled cDNAs were resuspended in 4.5 µl of 0.1 M Na_2CO_3, pH 9.0 and 4.5 µl of the appropriate NHS-ester Cy dye. The reactions were incubated for at least 1 h in the dark at room temperature. Two dyes were used: first

the FluoroLink™ Cy3 monofunctional (Amersham Biosciences), which is an orange fluorescing cyanine that can be excited at wavelength 532 nm, and the FluoroLink™ Cy5 monofunctional (Amersham Biosciences) producing an intense signal in the far-red of the spectrum (wavelength: 635 nm). Both of them can bind to compounds that have free amino groups (aa-dUTP). Before use, these dyes were resuspended in 73 µl of DMSO and aliquoted in 4.5 µl for storage at -70°C.

2.3.6 Quenching and cleaning

Before using the Cy3 and Cy5 labelled samples for hybridization, the reactions were quenched to prevent cross-coupling following the protocol of the Ontario Cancer Institute (*http://www.carleton.ca/~kbstorey/oci.pdf*). For that, 4.5 µl of 4 M hydroxylamine solution was added to the labelled probes and incubated at room temperature in the dark for 15 min. A volume of 35 µl of 100 mM NaOAc, pH 5.2 was added to the probes. Then the labelled aa-cDNAs were cleaned to remove any uncoupled dye by the PCR purification kit (Qiagen) following the manufacturer's protocol. After elution, the probes were mixed together and then dried by speed vacuum (Savant Speedvac System) at 50°C for about 45 min.

2.4 Hybridization and Analysis

These final steps were done using the instructions of the OCI protocol.

2.4.1 Hybridization

The prehybridization solution was prepared as follows: per 10 ml of DIG Easy Hybridization Solution (Roche Molecular Biochemicals), 500 µl of 10 mg/ml yeast tRNA and 500 µl of 10 mg/ml salmon sperm DNA were added. The prehybridization solution was filtered through 0.2 µm-syringe filter (Sarstedt) and 0.11 g of BSA (Sigma, Fraction V) was added and mixed by vortexing. Before use the solution and the hybridization containers (Coplin jars, VWR Canlab) were pre-heated at 42°C for 30 min. About 60 µl of the prehybridization solution was transferred onto the surface of the microarray slide and covered with a 22 mm × 60 mm polyethylene hydrophobic coverslip (PGC Scientific). The slide was placed in the Coplin jar in a 42°C Hybaid mini-oven for at least 30 min. Some DIG Easy Hybridization solution was also added at the bottom of the Coplin jar to prevent drying.

After the prehybridization, the slides were cleaned very carefully. Two Coplin jars were filled with MilliQ water and another one with isopropanol. The slides were rinsed five times in each Coplin jar beginning with MilliQ water. The slides were dried at room temperature and then the water/water/isopropanol wash cycle was repeated a second time. The slides were kept longer in the last isopropanol bath (~ 2 min).

The dried probe was resuspended in 60 µl of hybridization solution, which was prepared the same way as the prehybridization solution except that the BSA was not included. The hybridization solution was heated at 95°C for 3 min (as recommended

by the BRI) and transferred to a 42°C water-bath until use. The probe in the hybridization solution was added at the surface of the slide and covered by a 22 mm × 60 mm coverslip. The slides were transferred into a Boekel InSlide Out Hybridization Oven set at 42°C for 12-16 h.

2.4.2 Washing and Scanning

Following hybridization, the slides were washed in a staining glass container with a stir bar agitating the solution. The first wash was with a low-stringency buffer (1X SSC; 0.2% SDS) prewarmed at 42°C for 10 min. The coverslip was gently removed from the slide surface during this first wash. The wash solution was discarded and a second wash with pre-warmed low-stringency buffer was repeated. The slide was then washed 10 min in a high-stringency buffer (0.1X SSC; 0.2% SDS) pre-warmed at 37°C followed by a wash at room temperature with 0.1X SSC for 10 min. The last wash was in MilliQ water at room temperature for 10 min. The slide was finally dipped into a Coplin jar filled with MilliQ water and air-dried before scanning. The times of washes were doubled to 10 min compared to the original OCI protocol as the background diminished.

After hybridization, the DNA microarray was scanned to monitor the fluorescence of each probe that was successfully hybridized to the target. The scanning of both channels was simultaneous allowing the acquirement of both images at the same time and was proceeded as follow. First the microarray was inserted into the scanner with wavelengths 635 and 532 nm (Axon 4000b, Axon

Instruments, CA) face down and a "preview scan" was run with a low scan resolution (40 μm). This step finds the area on the array to be scanned at high resolution and also allows a quick view of the array quality. The next step was to delimit a small portion of area and to scan at high resolution (10 μm) in order to optimize the scanner settings. Two important factors optimize the scanner settings. First, maximize the dynamic range and having pixels distributed at all intensities from low intensity (1) to saturation (65535). This step was made by increasing the PMTs of both channels in order to cover as much as possible the full range of intensities. Second, balance the two channels by adjusting the PMTs to obtain the same amount of red and green signal resulting in a pixel ratio of approximately 1.0. To maximize the signal-to-noise ratio quality, the final scan was proceeded by averaging multiple lines. The scanner scanned several times the same line and averaged the value.

2.4.3 Normalization and filtering

All microarray slides were scanned then normalized using the total normalization method (Appendix A). Briefly this method assumes that genes spotted on the slides are not differentially expressed in the different conditions tested and therefore the mean of the ratio of the median intensities of all features in one slide should be equal to 1. If the value was different than 1, a normalization factor was calculated by the software GenePix™ Pro 4.0 and applied to the whole slide to rescale it to 1.

After the normalization step, a preset quality control test generated by the GenePix™ Pro 4.0 software was applied to each slide. This quality control test was based on the following set of criteria. Basically at least 55% of the pixels of each spot must be above the background added to one standard deviation for each channel ([% > B635+1SD] > 55 and [% > B532+1SD] > 55). The regression ratio of the red channel (635 nm) divided by the green channel (532 nm) should be above the numerical value of 0.3 ([Rgn R^2 (635/532)] > 0.3). All spots that are called as "not found" (no gene expression), "absent" (no gene spotted) or "bad" (flagged after visual verification) were also discarded ([Flags] <> [Bad]; [Flags] <> [Absent]; [Flags] <> [Not Found]). Also the number of spots reaching the saturation should count for less than 3% in the entire slide ([F635 % Sat.] < 3; [F532 % Sat.] < 3). Finally the sum of the median intensities for the red channel and green channel in each spot should be at least above 500 ([Sum of Medians] > 500).

2.4.4 Clustering

Two different kinds of clustering were used to analyze the microarray data (Appendix B). First the hierarchical clustering was performed as previously described (Eisen *et al.*, 1998) using the program Cluster. The resulting cluster was visualized using the program TreeView (both programs were developed at Stanford University and are available at *http://rana.stanford.edu/software/*). Briefly the hierarchical clustering algorithm regrouped together genes with similar expression profile and the data were graphically displayed using a color-scale. The gene

transcription abundance was therefore represented by shades of red for an increase and shades of green for a decrease (black color standing for missing data). A dendogram was simultaneously built during the clustering and the branch lengths indicated the degree of similarity between genes. The dendogram branch pointed in the middle of the cluster as the average-linkage method was adopted in this hierarchical clustering (Gasch *et al.*, 2000).

Second, the K-means clustering was performed using MaxdView software (developed by the Bioinformatics group of the University of Manchester and available at *http://bioinf.man.ac.uk/microarray/maxd/index.html*). In this clustering, the number of clusters is set by the user and the K-means algorithm allows the genes to move from one cluster to the other to regroup together genes with the best similarity.

2.5 RNA-blot hybridization analysis

The formaldehyde RNA gels were prepared following the "Molecular cloning" protocol (Sambrook and Russell, 2001) and 1 µl of ethidium bromide (200 µg/ml) was included to the samples. The running time was about 4-5 h at a speed of 5 V/cm and the gels were checked under a UV illuminator before the transfer to ensure even loading.

The nucleic acids were transferred onto nylon membranes Nytran SuperCharge (Schleicher & Schuell Biosciences) following the protocol provided in "Molecular cloning" (Sambrook and Russell, 2001) using 20X SSC as the transfer buffer. After

4 h of transfer, the membranes were rinsed in the electrophoretic buffer then UV cross-linked at 120 mJ using Stratalinker™ (Stratagene) and stored at 4°C until use.

The starting amount of probe used for the random priming was 100 ng of DNA and this step was done following the instructions provided in the Random Prime Labelling System *redi*prime™II kit (Amersham Biosciences). The radioactively-labelled probes were then purified through MicroSpin™ G-25 Column (Amersham Biosciences) following the recommendations of the manufacturer. The incorporation rate of the radioisotope [α-^{32}P] dCTP (3000 Ci/mmole; Amersham Biosciences) was checked by scintillation counter.

The prehybridization was performed at 68°C for at least 1.5 h in hybridization bottles in the Hybaid mini-oven with 20 ml of prehybridization solution (6X SSPE; 5X Denhardts' reagent; 0.5% SDS; 100 µg/ml sonicated and boiled salmon sperm DNA). The radioactively-labelled probe was directly added to the prehybridization solution and the hybridization was conducted overnight at 68°C. After the hybridization the membranes were rinsed 3 times with a low stringency solution (1X SSC; 0.5% SDS) at room temperature. Then the following 3 washes were carried out with a high stringency solution (0.5X SSC; 0.5% SDS) at 58°C for 25 min with agitation in a waterbath. Finally the membrane was wrapped in Saran Wrap (Reynolds® film 906RF) and subjected to exposure to X-ray films (Super RX, Fujifilm) with intensifying screens. The resulting bands were quantified using Syngene Digital Imaging System.

2.6 RT-PCR

The reverse transcriptions were performed with a starting amount of 5 μg of total RNA. The RNA was mixed with 0.25 μl of oligo-dT (2 μg/μl) and heated at 70°C for 5 min. The tubes were chilled on ice for 5 min and the following reagents were added: 4 μl of 5X First Strand buffer, 2 μl of 10 mM dNTPs mix, 2 μl of DEPC-treated water. The reaction mixture was incubated at 37°C for 5 min and then 200 units (1 μl) of the reverse transcriptase (Superscript™ II RT; Life Technologies) was included and the tubes transferred at 42°C for 60 min. After the incubation time, the reverse transcriptase was inactivated by heating at 70°C for 10 min.

The PCR was directly done after the inactivation of the reverse transcriptase. The PCR reactions were carried out as previously using GlaT5' and GlaT336 as primers (GlaT5' sequence: 5' AGA ATG AAT TCT TTG GCC GGC CCG ACC GCG ACG GTG ACT GAC 3'; GlaT336 sequence: 5' CGC GAA TAT TGA GCT CAT GGC TAT GCA TTG 3'). The PCR template corresponded to 1 μl of the reverse transcription product and the extension time of the PCR cycle was increased to 3 min. The RT-PCR products were then verified by gel electrophoresis on a 1% TBE agarose gel and staining with ethidium bromide.

3. Results

3.1 Quality control

3.1.1 DNA and RNA quality

As the use of DNA microarrays in transcription profiling is very sensitive, all the components involved must be of the highest quality. First, all the DNA samples used to spot on the microarray slides were checked by agarose-gel electrophoresis to confirm the presence of the expected size products (Figure 7). The PCR products were then purified by Millipore FB plates and selected examples are shown in Figure 8. The results of PCR amplification and purification are summarized in Table 5.

Another important factor in ensuring reproducibility in microarray experiments is the quality of the RNA samples used to generate probes. Degradation of the RNA samples before the reverse-transcription step can affect the final results, as certain transcripts would be missing. Contamination with any cellular components (cellular protein, lipid, carbohydrate and DNA) can hinder the hybridization performance as these contaminants can bind non-specifically to the fluorescent-labelled probe, resulting in a significant and uneven background and affecting the reproducibility of the experiments (Li *et al.*, 2002). The yield and concentrations of RNA were verified spectrophotometrically (Table 6) while the integrity and purity were examined by capillary electrophoresis on Agilent RNA 6000 chip (Figures 9 & 10). Most RNA samples showed satisfactory yield and quality.

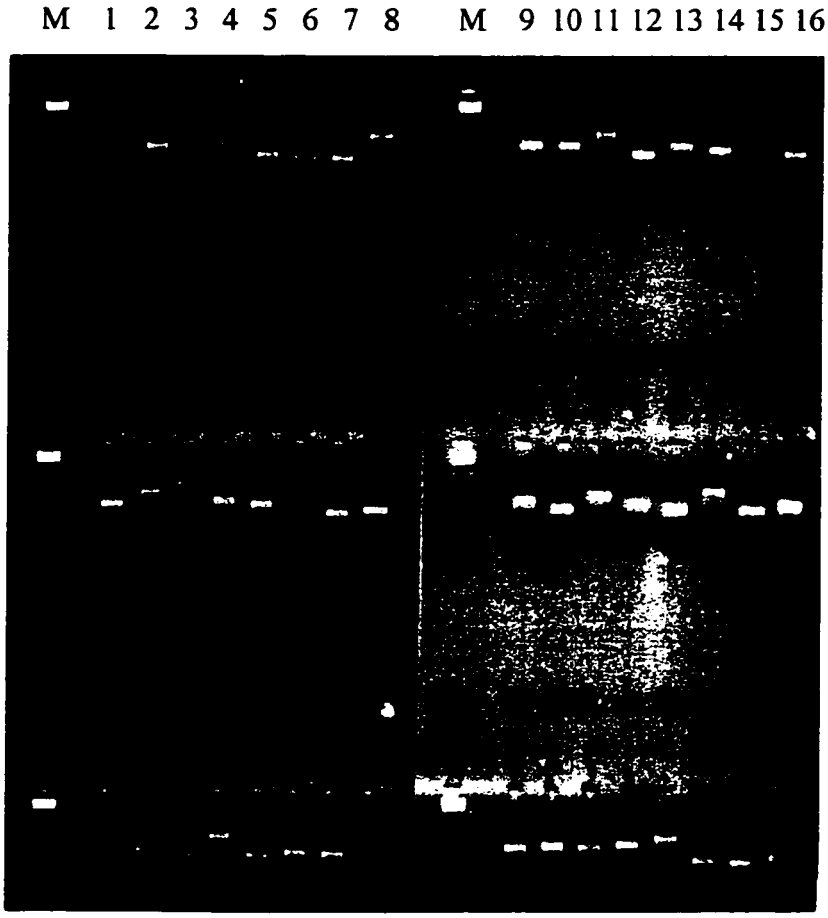


Figure 7: Agarose-gel electrophoresis of 48 PCR products from the analyzed cDNA clones.

A total of 72 agarose gels, each containing 48 PCR products (6 rows X 8 wells), were examined. The molecular size marker, lambda DNA digested with *Hind*III (0.5 μg) was loaded in the first well of each row (M). The gel shown here reveals the products of the PCR amplification. This representative subset corresponds to the clones an_1740 to an_1789.

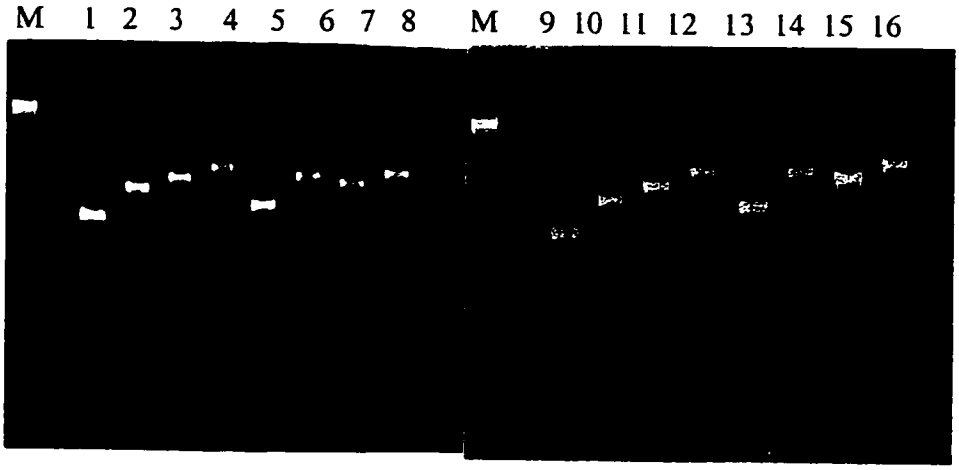


Figure 8: Agarose gel electrophoresis of 8 PCR-amplified products before and after purification.

Lane M corresponds to lambda DNA digested with *Hind*III.

a) The lanes 1 to 8 represent the PCR products of the clones an_1907, an_1908, an_1909, an_1910, an_1911, an_1912, an_1913, an_1914 before purification.

b) The lanes 9 to 16 represent the PCR products of the same set of clones (same order) after purification.

On average, 80% of the PCR products were recovered after purification.

Table 5: Summary of PCR amplification and purification yields.

Starting number of clones	Number of clones successfully PCR amplified	% of clones successfully PCR amplified	Average concentration of PCR products	Average concentration after purification
3,456	3,331	96.38%	183.75 ng/μl	145.85 ng/μl

Table 6: Amounts of mycelia and RNA isolated from different growth conditions.

Conditions	Mycelium weight (g)	RNA concentration (µg/µl)	TotalRNA (µg)	%RNA / mycelia weight
MM + Glucose	2.33	7.47	1494	0.097
	3.43	4.92	885.6	0.077
MM + Maltose	2.04	10.7	1070	0.13
	1.92	11.7	1170	0.21
MM + Xylose	1.56	7.45	1043	0.14
	1.98	7.52	1203.2	0.14
MM + Wheat germ	4.64	1.31	200	0.01
	5	2.08	457.6	0.024
MM + Xylan	2.43	4.36	872	0.06
	2.55	5.46	764.4	0.14
Complete medium (CM)	2.6	2.29	503.8	0.04
	3.45	0.677	94.78	0.012

Figure 9: RNA-gel representation using the Agilent RNA 6000 chip

Lane L: RNA 6000 ladder

Lane 1: RNA from MM+maltose (1%) (1^{st} biological replicate)

Lane 2: RNA from MM+maltose (1%) (2^{d} biological replicate)

Lane 3: RNA from MM+xylose (1%) (1^{st} biological replicate)

Lane 4: RNA from MM+xylose (1%) (2^{d} biological replicate)

Lane 5: RNA from MM+wheat germ (2%) (1^{st} biological replicate)

Lane 6: RNA from MM+wheat germ (2%) (2^{d} biological replicate)

Lane 7: RNA from MM+xylan (2%) (1^{st} biological replicate)

Lane 8: RNA from MM+xylan (2%) (2^{d} biological replicate)

Lane 9: RNA from CM (1^{st} biological replicate)

Lane 10: RNA from CM (2^{d} biological replicate)

Lane 11: RNA from MM+glucose (1%) (1^{st} biological replicate)

Lane 12: RNA from MM+glucose (1%) (2^{d} biological replicate)

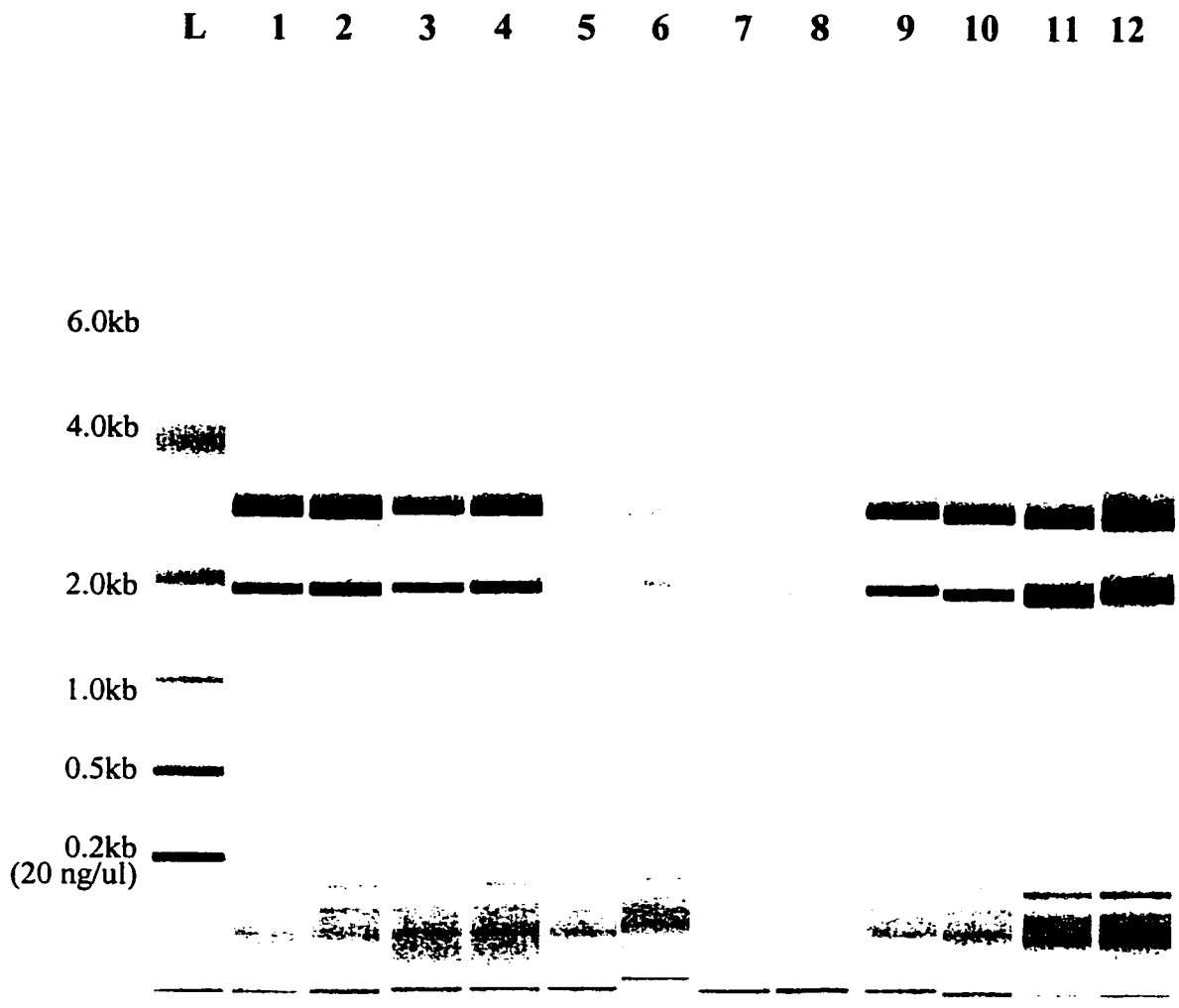

L 1 2 3 4 5 6 7 8 9 10 11 12
6.0kb
4.0kb
2.0kb
1.0kb
0.5kb
0.2kb
(20 ng/ul)

Figure 10: RNA quality control using Agilent Bioanalyzer 2100 chip

The four graphs show the scans of the RNA gels. The first small and sharp peak corresponded to the marker followed by the wide peak of tRNA, the two next sharp and long peaks represent the 18S and 28S ribosomal RNAs. The sharp peaks corresponding to 18S and 28S suggest that the RNA samples are of high quality with little or no degradation.

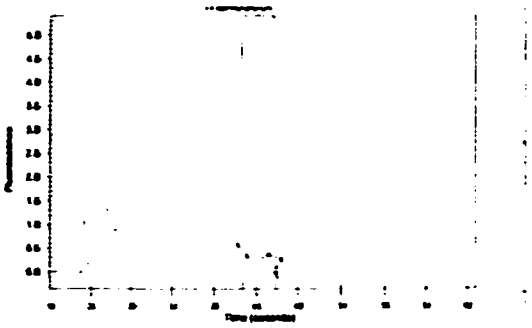
Graph 1: Total RNA from MM+glucose (1^{st} biological replicate)

Graph 2: Total RNA from MM+maltose (1^{st} biological replicate)

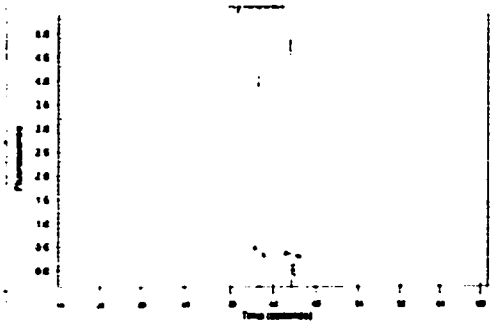
Graph 3: Total RNA from CM (1^{st} biological replicate)

Graph 4: Total RNA from MM+xylose (1^{st} biological replicate)

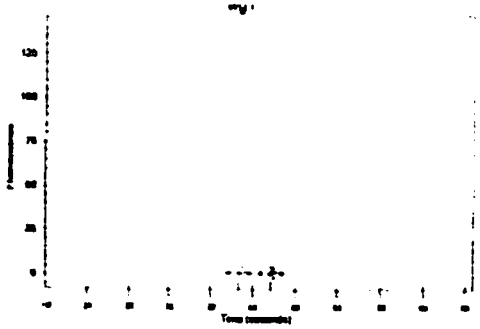
a.



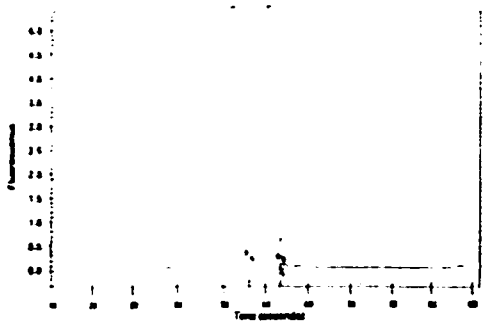
b.



c.



d.



Nevertheless, RNA extracted from MM+wheat germ was less abundant and slightly degraded (Figure 9, lanes 5 & 6). In 10 independent extractions, all the samples from MM+wheat germ resulted consistently to lower yields and degradation in comparison to the other RNA samples. Moreover RNA samples extracted from the MM+xylan condition revealed also a lower yield and slight degradation (Figure 9, lanes 7 & 8). These latter RNA samples showed an unusual brown colour, which was present in all independent extractions. The subsequent reverse transcription step was performed correctly when these RNA samples were purified by silica columns, thus removing this brown contaminant. Unfortunately this purification reduced dramatically the yield (from 50 to 70%) and lead to minor degradation.

3.1.2 Generation of microarray data

After checking the quality of DNA products and RNA samples, the PCR products were spotted on the microarray slides and the RNA samples were fluorescently labelled and hybridized to the microarray. Two cDNA labelled probes were applied onto the microarrays, which contained about 1,700 genes of *A. niger*. These probes corresponded to the RNA isolated from *A. niger* grown in MM+glucose and labelled with Cy5 (control condition) plus RNA from one of the other growth conditions and labelled with Cy3 (MM+maltose, MM+xylose, MM+xylan, MM+wheat germ or Complete Medium). Thus, in theory, 5 printed arrays must be used to test all desired conditions. Moreover each microarray slide contained the set of 1,700 genes in duplicate (2X 1,700 genes) constituting the two

technical replicates. Besides the hybridization was repeated twice with RNA extracted from replicate cultures. Thus two slides were used per condition tested and constituted the two biological replicates. Therefore, a total of 10 microarray slides were hybridized to assess the different conditions in replicate.

The hybridization is based on the sequence complementarity between the cDNA probes and the DNA spotted on the array. If a cDNA probe has its sequence complementary to the sequence DNA on a given spot, that cDNA will hybridize to the spot under the proper conditions, and it will be detectable by its fluorescence. Therefore, every spot on an array is an independent assay for the presence of a particular cDNA. Moreover there is enough DNA deposited per spot to allow the hybridization of both probes at the same time without interference (Appendix A).

After hybridization, the microarray slides were scanned to generate two-color images representing the hybridization of the different labelled probes (Figure 11). The relative abundance of each transcript was estimated by measuring the fluorescence of the two dyes bound to each spot and expressed as a ratio. As one sample was always taken as the control biological condition, the ratio generated gave a direct indication of responses relative to that control (Gracey and Cossins, 2003). The ratios were generated for each spot by dividing the median intensities of the green channel at 532 nm (probe labelled with Cy3) by the median intensities of the red channel at 635 nm (probe of control condition labelled with Cy5). Moreover the ratios were converted to logarithms in base 2 since this is the standard way to measure microarray data and the most convenient way to evaluate the relative

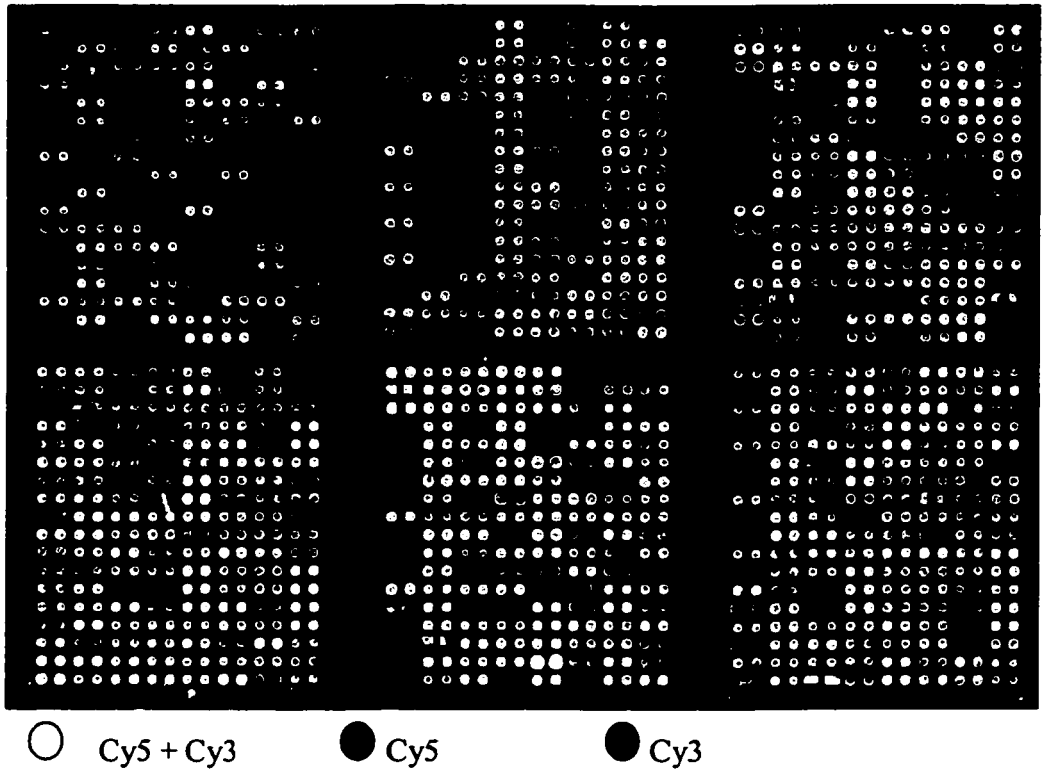


Figure 11: Two-color image of the microarray section with MM+wheat germ probe (Cy3) hybridized with MM+glucose probe (Cy5).

This portion of microarray represents 864 clones spotted in duplicate side by side. Basically a green spot signifies that mainly the Cy3 probe hybridized, the gene is expressed at higher levels under MM+wheat germ only. A red spot means that hybridization is mainly with the Cy5 probe, the gene is expressed under MM+glucose. A yellow spot corresponds to the hybridization of both probes (Cy3+Cy5), the gene is expressed at about the same level under both conditions. A black spot is no hybridization, the gene is expressed at levels below detection under both conditions of MM+wheat germ and MM+glucose.

expression in the range of several folds (represented by $\log_2$ratio (Cy3/Cy5)) (Quackenbush, 2001).

3.1.3 Data filtering

Given the scale of the microarray data sets, filtering steps are often used in the analysis process to reduce the data set to a more manageable size. The first filtering was to remove all improper spots. The data were filtered using the quality control set criteria of the GenePix™ Pro 4.0 software (Material and methods, section 2.4.3) which removes improper spots based mainly on the shape and intensity of the spot. An average of 46% of the clones were retained (Table 7) and this was considered a little bit low but still reasonable for a microarray data (personal communication, Dr P. Fawcett). One set of data was completely rejected (Table 7), as the hybridization of the second MM+xylan biological replicate was problematic. This array hybridized well to the RNA probe labelled with Cy5 but a fluorescent signal was completely absent for the RNA probe labelled with Cy3. As the reverse-transcription of the RNA into cDNA appeared successful based on agarose gel analysis (result not shown), we suspect that subsequent labelling and/or purification of the probe failed. This set of data was therefore ignored for the rest of the analysis and the number of replicates for MM+glucose / MM+xylan was reduced to two.

Furthermore the negative controls (Material and methods, section 2.2.4) showed a very low intensity as expected and were mostly discarded during this filtering step.

Table 7: Results of the microarray quality control set of criteria applied to each slide.

	Slide number	Fluorescently labelled probes applied	Spots after filtering (total 15,552 spots)	% spots after filtering
Data Set 1	12991652	Glucose/Maltose	5,301	34%
	12991633	Glucose/Complete medium	4,250	27%
	12991651	Glucose/Xylose	7,722	50%
	12991634	Glucose/Xylan	7,213	46%
	12991649	Glucose/Wheat germ	8,771	56%
Data Set 2	12991630	Glucose/Maltose	9,334	60%
	12991629	Glucose/Complete medium	6,427	41%
	12991653	Glucose/Xylose	9,244	60%
	12991655	Glucose/Xylan	not determined	not determined
	12291631	Glucose/Wheat germ	5,583	36%
				Average = 46%

3.1.4 Quality control graphs

After collecting the microarray data and removing the bad spots, the data sets were plotted in several graphs. These graphs allowed us to visualize the distribution and repartition of the data, and therefore, were indicative of the data quality before starting any further analysis. First the sets of data in each slide were plotted one (red data at 635 nm) versus the other (green data at 532 nm) to verify the balance between the two channels (Figure 12). This type of graph is called a correlation graph and is important in image processing and data acquisition. The $\log_2$(Cy3) intensity was plotted versus the $\log_2$(Cy5) intensity for all conditions, but only the correlation graph of MM+maltose versus MM+glucose is shown (Figure 12). The correlation factor calculated from the plot must be as close as possible to 1, which is the perfect equilibrium between the two sets of data. In the example of the Figure 12, the correlation factor of the MM+maltose plotted against MM+glucose is 0.956, showing that the sets of data were well balanced and could therefore be retained for further analysis. The histogram and the graph of Dutoit presented in Figure 13 shows that the clone distribution in both channels and the data are centered around the value 0. Most of the genes were regrouped around the $\log_2$ratio (Cy3/Cy5) equal to zero, meaning that most of the genes were not differentially expressed between one condition and the other as expected. The data of the other conditions gave similar encouraging results with correlation factors all close to 1 (0.924 for MM+glucose / MM+xylose; 0.850 for MM+glucose / MM+xylan; 0.969 for MM+glucose /

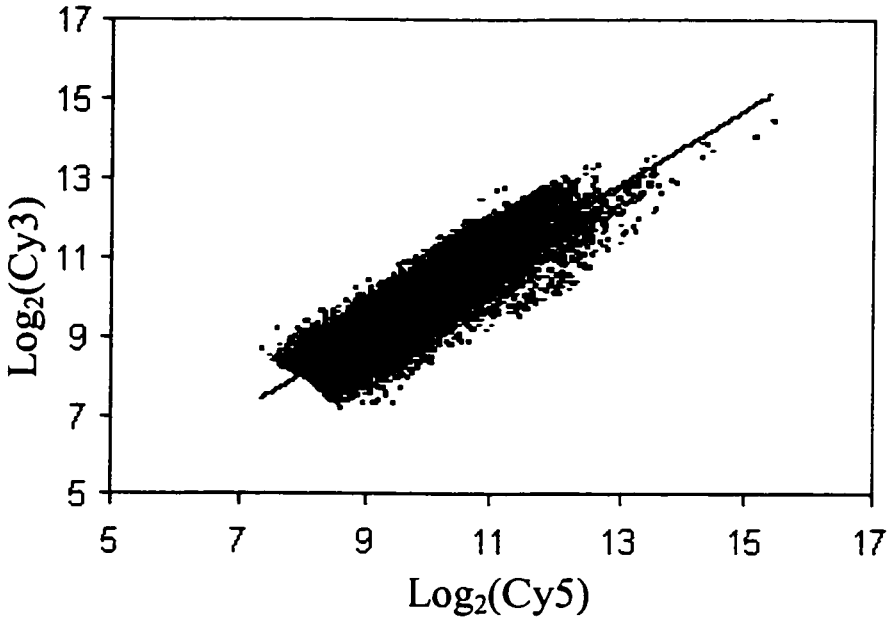


Figure 12: Correlation graph of the MM+maltose (Log_2Cy3) versus MM+glucose (Log_2Cy5).

For each condition, the logarithm in base 2 of the median intensities of each clone was calculated then plotted onto correlation graphs. This data series corresponding to the MM+glucose versus MM+maltose conditions had a correlation coefficient of 0.956 showing that both channels are well balanced.

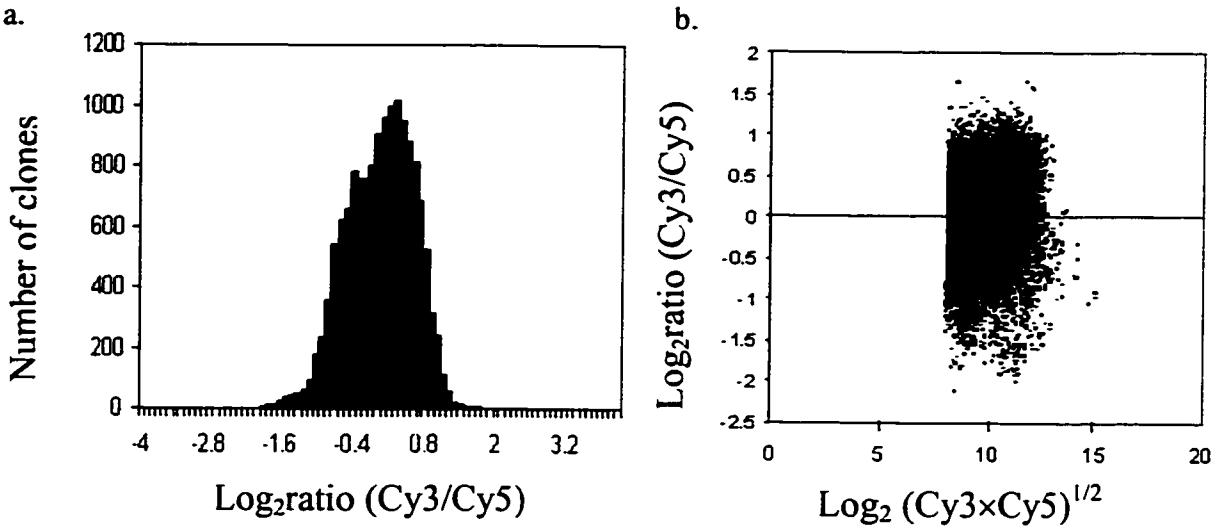


Figure 13: Quality control graphs of the data sets of MM+maltose versus the control condition MM+glucose.

a. Histogram representing the number of clones versus the $\log_2$ratio (Cy3/Cy5)

b. Graph of Dutoit representing the $\log_2$ratio (Cy3/Cy5) versus $\log_2(Cy3 \times Cy5)^{1/2}$

Both representations show that the distribution and repartition of the data are centered around zero.

MM+wheat germ; and 0.981 for MM+glucose / Complete Medium) and data mainly centered around zero.

3.2 Microarray analysis

The different quality controls and filtering steps led to data with acceptable quality and therefore the analysis of the microarrays data was performed. As mentioned earlier, the microarray experiments were repeated with two technical replicates (*A. niger* arrays hybridized with the same probe coming from the same culture) and two biological replicates (*A. niger* arrays hybridized with probes coming from replicate cultures). Therefore four replicates were available for each set of conditions (MM+glucose, MM+maltose, MM+xylose, Complete Medium and MM+wheat germ) except for MM+xylan, where only two replicates were available. The software GenePix™ Pro 4.0 displayed three types of data corresponding to three important values calculated for each clone: the median intensity (subtracted from the local background of the spot) in both channels (532 nm and 635 nm), and the $\log_2$ratio (median intensity at 532 nm / median intensity at 635 nm). The median intensities for both channels were used to establish the list of the highly expressed genes while the $\log_2$ratio (Cy3/Cy5) was utilized to determine whether the genes were differentially expressed.

3.2.1 Highly expressed genes (HEG)

The most highly expressed genes in one or several conditions were assessed by considering the median intensities in the two channels separately. The signal intensities of all four replicates (or two for MM+xylan) were merged in the same table (data not shown). Basically we generated 6 tables for the 6 different conditions with 4 columns for the 4 replicates (except MM+xylan with 2 columns). First we calculated the median value of each set of intensities per column (a total of 4 median values for the 4 columns or 2 for MM+xylan condition). Then each intensity value corresponding to each gene was divided by the median value calculated for the column, creating therefore a ratio for each gene. This ratio was then converted to logarithm in base 2 (internal $\log_2$ratio) and the direct comparison of the internal $\log_2$ratio for each gene in the four replicates allowed us to verify again the quality of the data. The value should be very similar between replicates and the result of this extra quality control test was very satisfactory (data not shown). Therefore the mean value of the internal $\log_2$ratio between replicates was calculated and data were sorted.

The highly expressed genes corresponded to those with the highest internal $\log_2$ratio. Arbitrary threshold values were set to find the most highly expressed genes. When the threshold value was set at two (internal $\log_2$ratio set at 2), in total 467 genes out of 1,700 genes (27.4%) were highly expressed. 41.3 % of these highly expressed genes have a known function (Appendix C). If the threshold value was increased to three, 3.8% were considered highly expressed (65 genes).

From these results, we found that the majority of the HEG (with internal $\log_2$ratio > 2) had an unknown function (58.7%). Nevertheless the rest of the HEG could be sorted into functional classes and most of them were found to belong to the ribosomal proteins (8.56%) group, to the energy/TCA cycle (4.06%), protein turnover (3.64%) and transport (3.42%) categories (Figure 14). Those groups correspond to genes normally highly expressed in favorable growing conditions (Wodicka *et al.*, 1997). The glucoamylase gene was found highly expressed in all conditions (with internal $\log_2$ratio values equal to 4.17 for MM+maltose, 3.8 for MM+glucose, 3.7 for MM+wheat germ, 2.65 for Complete Medium) except in MM+xylan (0.39) and MM+xylose (1.25) (Appendix C).

3.2.2 Comparison of gene expression

The transcription profile was obtained using the $\log_2$ratio (Cy3/Cy5) based on the ratio of medians (median intensity at 532 nm / median intensity at 635 nm) directly calculated by the GenePix™ Pro 4.0 software for each slide. At this stage, further filtering was performed with a program specifically created for our set of data using the SAS software (SAS Institute Incorporated, Cary, U.S.A.). This filtering was necessary to remove the inconsistent data between replicates and also the non-interesting data (i.e. genes not differentially expressed) to allow subsequent clustering analysis.

First, the four $\log_2$ratio (Cy3/Cy5) replicates generated in each condition were gathered in the same table (a total of 5 tables) (data not shown). The genes with two,

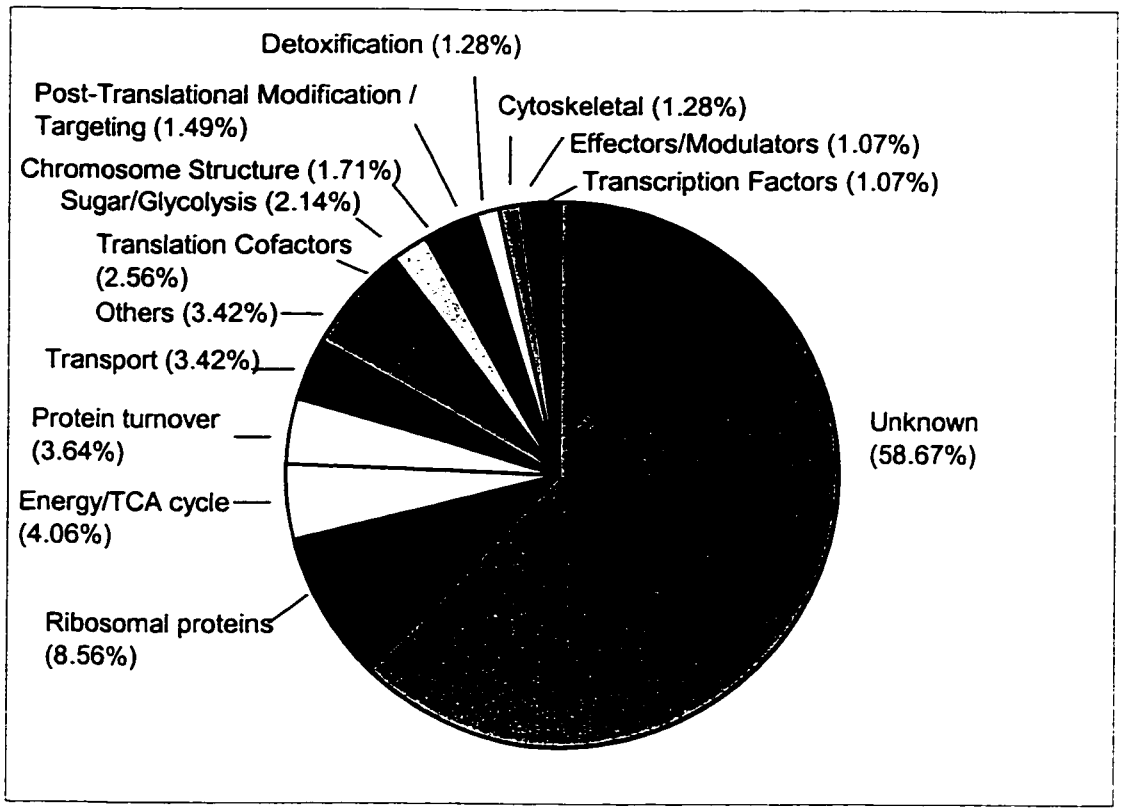


Figure 14: Pie chart grouping highly expressed genes in functional classes.

The 467 highly expressed genes were sorted by functional classes. Most of the HEG have no function (58.67%), then 8.56% are encoding ribosomal proteins followed by energy/TCA cycle genes (4.06%) and protein turnover genes (3.64%).

three and four values for the replicates were retained whereas the ones with only one value (i.e. no replicate data) were discarded. The median value and the standard deviation (SD) were calculated for each gene. Confidence intervals were calculated for each gene based on the Student's *t* distribution. When using a Student's *t* distribution, two, three and four standard variations are equivalent to 90%, 95% and 97% confidence intervals for as little as only a few replicates (Finkelstein *et al.*, 2002). Therefore, all genes with more than 2 SD from the population mean of the $\log_2$ratio must be consistent with 90% of confidence interval between the replicates and those data were therefore retained. The rest was discarded.

We decided to use the cutoff value of 3-fold ($\log_2$ratio = +1.5 or -1.5) to consider the genes as differentially expressed as many studies use cutoff values from 2-fold to 3-fold (2-fold for Schena *et al.*, 1996; 3-fold for Wodicka *et al.*, 1997; 2-fold for Hegde *et al.*, 2000; 2.5-fold for Schenk *et al.*, 2000 and 3-fold personal communication, Dr. P. Fawcett). Genes with a $\log_2$ratio between -1.5 and +1.5 were therefore filtered, as they were not considered differentially expressed (Table 8). Furthermore, the $\log_2$ratio (Cy3/Cy5) values of the differentially expressed genes were indicated in all tested conditions (not only in the particular condition(s) where the genes were differentially expressed). However a majority of these data was removed when proceeding to the 2 SD filtering step. This is explained by the fact that the SD values can be very low when the $\log_2$ratio (Cy3/Cy5) are closed to zero and thus most of the genes which are not differentially expressed were discarded at this step.

Table 8: Differentially expressed genes (3-fold change).

Log$_2$ratio (Cy3/Cy5)					Gene name	Description
Xylose	**CM**	**Maltose**	**Wg**	**Xylan**		
		0.66	-1.79	-0.76	an_3665	26S proteasome regulatory particle chain RPN12
0.47	-1.52	0.72	-1.19		an_2646	40S ribosomal protein S10
	-1.78	0.70	-1.18	-0.56	an_2255	40s ribosomal protein s3
	-1.27	0.30	-1.86	-0.51	an_2360	40S ribosomal protein S3A
	-1.61			-0.75	an_1030	40s ribosomal protein s5
	-1.42		-1.75	-0.67	an_3183	60S ribosomal protein L16 (L11)
-0.08	-2.08		-1.80	-0.78	an_2586	60S ribosomal protein L17
	-2.41		-1.07	-0.88	an_2344	60S ribosomal protein L19B
	-0.91	0.24	-1.69		an_1324	60S ribosomal protein L24
0.46	-1.54	0.62	-1.27		an_771	60S ribosomal protein L29 (L27A)
	-0.8		-1.51		an_2100	60S ribosomal protein L9
	-1.57	0.83			an_835	Acidic ribosomal protein P1
-0.41	-1.4		-1.88	-0.76	an_2369	ADP,ATP Carrier protein
0.73	-0.87	0.63	-1.78		an_2430	ATP synthase 9
		0.41	-2.26	-2.07	ANI347748	Benzoate-para-hydroxylase
0.62			2.19		an_3367	Carbamoyl phosphate synthetase small subunit
				1.68	an_2347	Carboxyvinyl-carboxyphosphonate phosphorylmutase
	-1.13	0.74	-1.54		an_1040	Cobalamin-independent methionine synthase
-2.11		-1.01		-1.53	AF219625	D-xylose reductase
0.03		0.33	-1.61	0.84	an_1230	Elongation factor 1 alpha
			-2.00	-0.87	an_2763	Elongation factor 2
		0.61	-1.92		an_3281	F-actin capping protein beta subunit
-2.94				-2.96	an_2467	Glucoamylase
	-3.4		-1.67	-0.99	an_2588	Glutamine synthetase
			-2.27	-1.11	an_636	Glyceraldehyde 3-phosphate dehydrogenase
-0.82			-2.59	-1.43	an_2904	Glyceraldehyde 3-phosphate dehydrogenase
	-1.69	0.53	-1.78	-0.92	an_661	Glycine-rich protein
	-0.78		-1.54		an_3404	Heat shock 70 kDa protein
	-0.93		-2.00	-0.74	an_3438	Heat shock protein
	1.63	-0.68		0.78	an_2522	Hemolysin
0.67		0.26	1.50	0.7	an_1986	Homeodomain DNA-binding transcription factor
0.4	-1.65	0.67			an_1008	Hypothetical protein
	1.6			0.93	an_1084	Hypothetical protein
	-1.62			-0.26	an_2776	Malate dehydrogenase
			-1.57	-0.71	an_1810	Mold-specific MS8 protein
-0.21	1.53		-1.68	0.01	an_668	Peroxisomal hydratase-dehydrogenase-epimerase
			-1.94	-1.06	an_3131	Plasma membrane H+-ATPase
0.32			1.54		an_2672	Plasma membrane H+-ATPase
			-1.6		an_1995	Polyubiquitin
	-0.51		-1.67		an_2374	Probable Peroxisomal Membrane Protein PMP20
			-1.97	-1.12	an_3118	Probable phosphate transport protein MIR1
-1.86		-0.46	-1.87	-1.33	ANI9972	Pyruvate carboxylase
			-1.9	-0.73	an_924	Pyruvate decarboxylase
	-1.11		-2.04	-0.81	an_2760	Ribosomal protein L10
0.39	-1.93		-1.65	-0.81	an_2744	Ribosomal protein L18a
	1.82			1.14	an_595	Similar to serine/threonine kinase
	-0.53		-2.22		an_3554	Thioredoxin
0.67		0.09	1.93		an_2485	Thioredoxin, mitochondrial precursor
-0.62		-1.51	0.3		an_2096	Verprolin related protein
-0.77			-1.94	-1.57	ASNXYNBA	Xylanase B

Table 8 (continued): Differentially expressed genes (3-fold change).

Log$_2$ratio (Cy3/Cy5)					Gene name	Description
Xylose	**CM**	**Maltose**	**Wg**	**Xylan**		
			-1.97	-0.92	an_963	Function unknown
		-0.64	1.8		an_742	Function unknown
-0.64	-0.76		-1.73	-1.01	an_734	Function unknown
			1.5		an_575	Function unknown
				1.53	an_3658	Function unknown
	1.63			0.51	an_3635	Function unknown
	1.61			1.36	an_3634	Function unknown
	-0.61		1.93		an_3591	Function unknown
	-1.08	-0.31	-1.59	-0.53	an_3577	Function unknown
-0.25			-1.75	-1.1	an_3573	Function unknown
	1.55			0.56	an_3420	Function unknown
	1.52				an_3386	Function unknown
	1.58			0.77	an_3326	Function unknown
	-1.55		-1.24	-0.61	an_3319	Function unknown
	1.53			0.47	an_3186	Function unknown
	0.86	-1.6			an_2987	Function unknown
	-0.97		-1.82		an_2970	Function unknown
	1.59			0.62	an_2930	Function unknown
-1.4			-2.33	-1.33	an_2927	Function unknown
0.14	-1.56		-1.26	-0.37	an_2836	Function unknown
	2.17	-1.04			an_2826	Function unknown
-0.4			-2.42	-1.31	an_2803	Function unknown
	-0.56		-1.8		an_2670	Function unknown
	1.57			0.64	an_2666	Function unknown
0.25			-1.56		an_2642	Function unknown
	-2.99		-4.16		an_2625	Function unknown
				1.57	an_2453	Function unknown
			-1.5	0.42	an_2378	Function unknown
	-1.69		-1.86		an_2302	Function unknown
-0.69		0.39	-1.52	-0.34	an_2201	Function unknown
0.74	2.03		2.1	0.3	an_2197	Function unknown
			-1.58		an_2185	Function unknown
-0.31	-0.81		-1.5	-0.63	an_1664	Function unknown
			-1.58		an_1613	Function unknown
		-1.59			an_1494	Function unknown
	1.67				an_1390	Function unknown
			-2		an_1198	Function unknown
0.56	1.67	0.28	1.72		an_1191	Function unknown
-0.4		-1.64			an_1150	Function unknown
		-0.57	-1.59		an_1137	Function unknown

Finally the differentially expressed genes (DEG) in at least one of the conditions represented about 5.1% of the genes printed on the microarray (a total of 88 genes) with more than half of these (55.5%) having a known function. These results could also be sorted by functional classes and the DEG corresponded mostly to ribosomal protein genes (14.44%), and genes involved in metabolism processes: sugar/glycolysis (6.66%), energy/TCA cycle (5.55%), transport (4.44%) (Figure 15 & Table 8).

3.2.3 Clustering analysis

A natural basis for organizing gene expression data is to group together genes with similar patterns of expression. Moreover the rapid visualization is simplified by using a color scale instead of tables of numbers. Therefore the 88 DEG found earlier were clustered using both hierarchical and K-means clustering methods (Appendix B). Those complementary methods showed 10 potential clusters that can be recognizable in Figure 16 with the representation of the hierarchical clustering.

These clusters regrouped 24 induced genes (3 genes in MM+xylan, 14 genes in CM, 2 genes in both CM and MM+wheat germ conditions and 5 genes in MM+wheat germ only) and 64 repressed genes (4 genes in MM+maltose, 2 genes in both MM+xylan and MM+xylose conditions, 25 genes in both MM+wheat germ and CM conditions, 14 genes in both MM+wheat germ and MM+xylan conditions, 2 genes in the three conditions: MM+wheat germ and MM+xylan and MM+xylose, 13

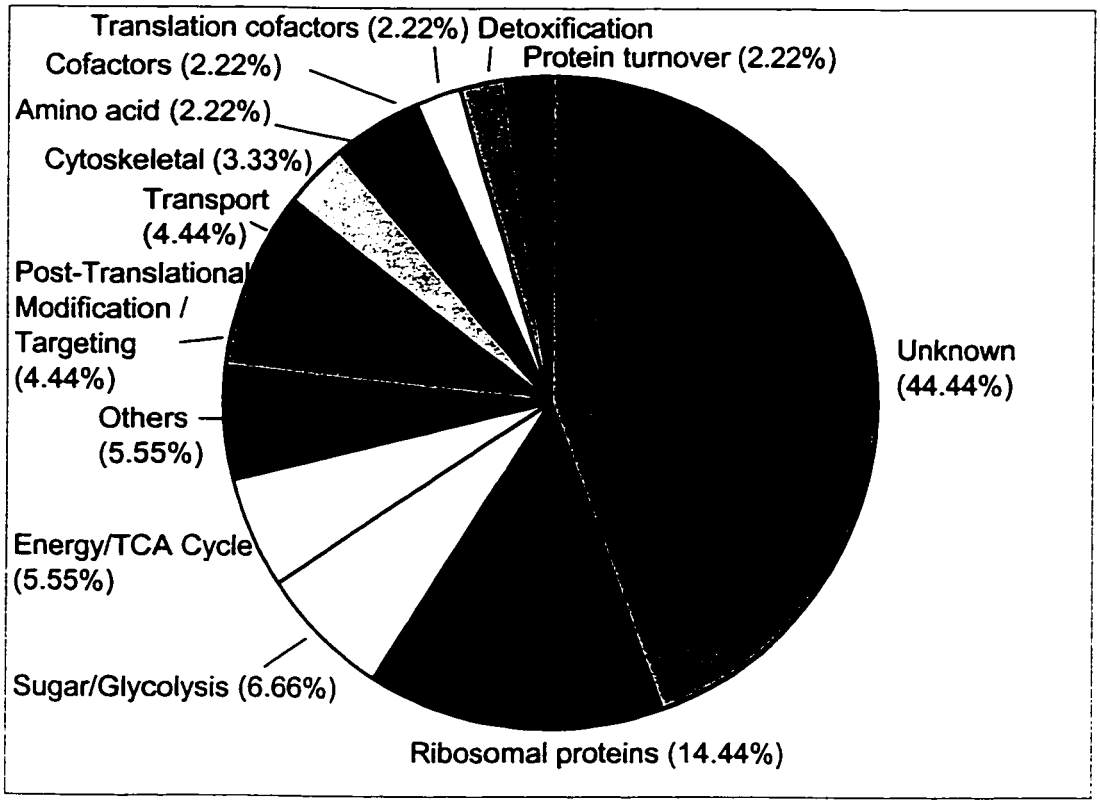
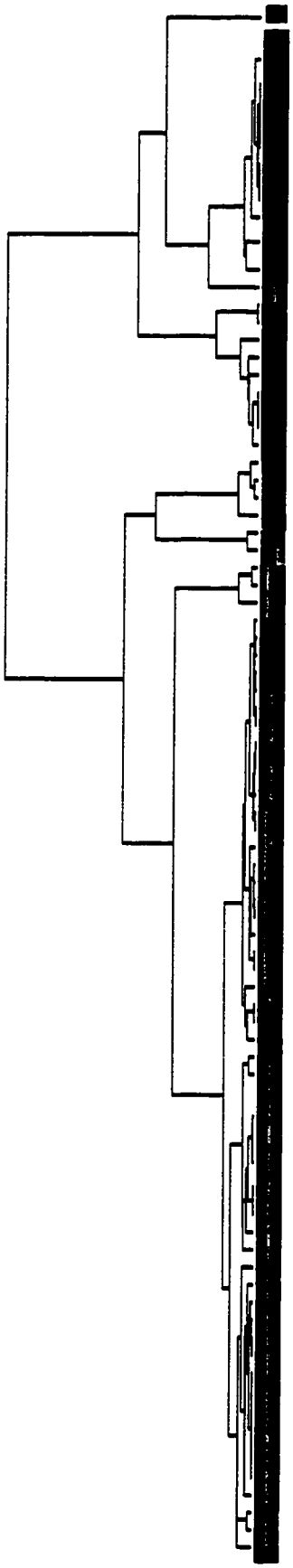


Figure 15: Pie chart grouping responsive (more than 3-fold change) genes in functional classes.

The 88 differentially expressed genes were sorted by functional classes. About half of the DEG have no known function (44.44%), then 14.44% are encoding ribosomal proteins followed by sugar/glycolysis genes (6.66%) and energy/TCA cycle genes (5.55%).

Figure 16: Hierarchical cluster analysis of combined *A. niger* data sets in MM+maltose, CM, MM+wheat germ, MM+xylan and MM+xylose.

The 88 DEG were clustered and a color scale was applied to facilitate the visualization. The red color indicates genes that are induced and green indicates genes that are repressed in comparison to the control MM+glucose condition.



	Role	Subrole	Description
an_2347	Metabolism	Sugar/Glycolysis	Carboxyvinyl-carboxyphosphonate phosphorylmutase
an_2453			
an_3658			
an_3634			
an_3186			
an_3635			
an_3420			
an_2666			
an_2930			Hypothetical protein
an_1084	Protein Synthesis	Post-Translational Modification/Targeting	Similar to serine/threonine kinase
an_595			
an_3326	Cell/Organism Defense	Stress Response	Hemolysin
an_2522			
an_3386			
an_1390			
an_2826			
an_668	Cell/Organism Defense	Detoxification	Peroxisomal hydratase-dehydrogenase-epimerase
an_2197			
an_1191			
an_1986	RNA Synthesis	Transcription Factors	Homeodomain DNA-binding transcription factor
an_742			
an_3591			
an_2485	Metabolism	Cofactors	Thioredoxin, mitochondrial precursor
an_2672	Metabolism	Transport	Plasma membrane H+-ATPase
an_3367	Metabolism	Amino Acid	Carbamoyl phosphate synthetase small subunit
an_575			
an_1494			
an_1150			
an_2096	Cell Structure/Cytoskeleton	Cytoskeletal	Verprolin related protein
an_2987			
AF219625	Metabolism	Sugar/Glycolysis	D-xylose reductase
an_2467	Metabolism	Sugar/Glycolysis	Glucoamylase
an_2776	Metabolism	Energy/TCA Cycle	Malate dehydrogenase
an_1030	Protein Synthesis	Ribosomal Proteins	40s ribosomal protein s5
an_1008			Hypothetical protein
an_3577			
an_2369	Metabolism	Transport	ADP,ATP Carrier protein
an_1664			
an_3183	Protein Synthesis	Ribosomal Proteins	60S ribosomal protein L16 (L11)
an_2360	Protein Synthesis	Ribosomal Proteins	40S ribosomal protein S3A
an_2760	Protein Synthesis	Ribosomal Proteins	Ribosomal protein L10
an_3438	Protein Synthesis	Post-Translational Modification/Targeting	Heat shock protein
an_3404	Protein Synthesis	Post-Translational Modification/Targeting	Heat shock 70 kDa protein
an_2970			
an_2100	Protein Synthesis	Ribosomal Proteins	60S ribosomal protein L9
an_1324	Protein Synthesis	Ribosomal Proteins	60S ribosomal protein L24
an_2625			
an_2302			
an_2255	Protein Synthesis	Ribosomal Proteins	40s ribosomal protein s3
an_2744	Protein Synthesis	Ribosomal Proteins	Ribosomal protein L18a
an_2836			
an_2586	Protein Synthesis	Ribosomal Proteins	60S ribosomal protein L17
an_3319			
an_661			Glycine-rich protein
an_2588	Metabolism	Energy/TCA Cycle	Glutamine synthetase
an_2344	Protein Synthesis	Ribosomal Proteins	60S ribosomal protein L19B
an_2430	Metabolism	Energy/TCA Cycle	ATP synthase 9
an_771	Protein Synthesis	Ribosomal Proteins	60S ribosomal protein L29 (L27A)
an_2646	Protein Synthesis	Ribosomal Proteins	40S ribosomal protein S10
an_1040	Protein Synthesis	Post-Translational Modification/Targeting	Cobalamin-independent methionine synthase
an_2378			
an_1230	Protein Synthesis	Translation Cofactors	Elongation factor 1 alpha
an_1613			
an_2185			
an_1198			
an_1995	Protein Synthesis	Protein Turnover	Polyubiquitin
an_2642			
an_2670			
an_2374	Cell/Organism Defense	Detoxification	Probable Peroxisomal Membrane Protein PMP20
an_3554	Metabolism	Cofactors	Thioredoxin
an_3281	Cell Structure/Cytoskeleton	Cytoskeletal	F-actin capping protein beta subunit
an_1137			
an_734			
ANI347748	Metabolism	Amino Acid	Benzoate-para-hydroxylase
ASNXYNBA	Metabolism	Amino Acid	Xylanase B
an_2904	Metabolism	Amino Acid	Glyceraldehyde 3-phosphate dehydrogenase
an_2803			
an_3573			
an_924	Metabolism	Energy/TCA Cycle	Pyruvate decarboxylase
an_636	Metabolism	Sugar/Glycolysis	Glyceraldehyde 3-phosphate dehydrogenase
an_2763	Protein Synthesis	Translation Cofactors	Elongation factor 2
an_963			
an_1810	Cell Structure/Cytoskeleton	Cytoskeletal	Mold-specific MS8 protein
an_3118	Metabolism	Transport	Probable phosphate transport protein MIR1
an_3131	Metabolism	Transport	Plasma membrane H+-ATPase
an_3665	Protein Synthesis	Protein Turnover	26S proteasome regulatory particle chain RPN12
ANI9972	Metabolism	Energy/TCA Cycle	Pyruvate carboxylase
an_2927			
an_2201			

genes in MM+wheat germ and 4 genes in CM). The general feature of these analyses showed that most of the genes (54 genes) were repressed under MM+wheat germ condition. Also in the CM condition, the analysis revealed several DEG with 16 genes induced while 29 genes were repressed. For MM+xylan condition: 3 genes were induced while 18 genes were repressed. No significant gene expression variation was observable for MM+glucose, MM+xylose and MM+maltose conditions.

Other genes seemed to be specific for one condition and indicated an increase with a particular carbon source. This was the case with the carboxyvinyl-carboxyphosphonate phosphoryl mutase gene involving in sugar/glycolysis process and two other genes (an_3658 and an_2453) that revealed a 3-fold increase in MM+xylan condition. In total 14 genes were induced only in CM condition with at least a 3-fold increase including hemolysin, peroxisomal hydratase-dehydrogenase-epimerase, genes similar to serine/threonine kinase and several genes with unknown function (clones an_1084, an_3635, an_3634, an_3420, an_3386, an_3326, an_3186, an_2930, an_2826, an_2666, an_1390). Five genes were induced only in MM+wheat germ condition including the carbamoyl phosphate synthetase small subunit, homeodomain DNA-binding transcription factor, plasma membrane H+-ATPase and thioredoxin genes plus three genes with unknown function (clones an_742, an_575, and an_3591). Also two genes (clones an_2197 and an_1191) were induced in only MM+wheat germ and CM.

Finally as expected, the glucoamylase gene was found to be repressed under MM+xylan and MM+xylose conditions compared to the control condition MM+glucose.

3.2.4 Carrier gene candidates

As mentioned before, the purpose of this project was to find carrier genes to enhance protein production in *Aspergillus* species. Those potential candidates should share several characteristics with the well-characterized carrier gene glucoamylase and possess two important traits: be highly expressed and differentially regulated. By merging the data obtained for the highly expressed genes (467 HEG) and the differentially expressed genes (88 DEG), a total of 39 genes were found to be potential candidates. These genes were also screened for potential signal-peptide sequences in their peptide using the SignalP V1.1 prediction server (Nielsen *et al.*, 1997). In total 8 genes showed a secretion signal including also the control glucoamylase gene (Table 9).

3.3 RNA-blot hybridization analysis

Traditional approaches were performed to verify the validity of our microarray results including RNA-blots and RT-PCR for the control glucoamylase gene.

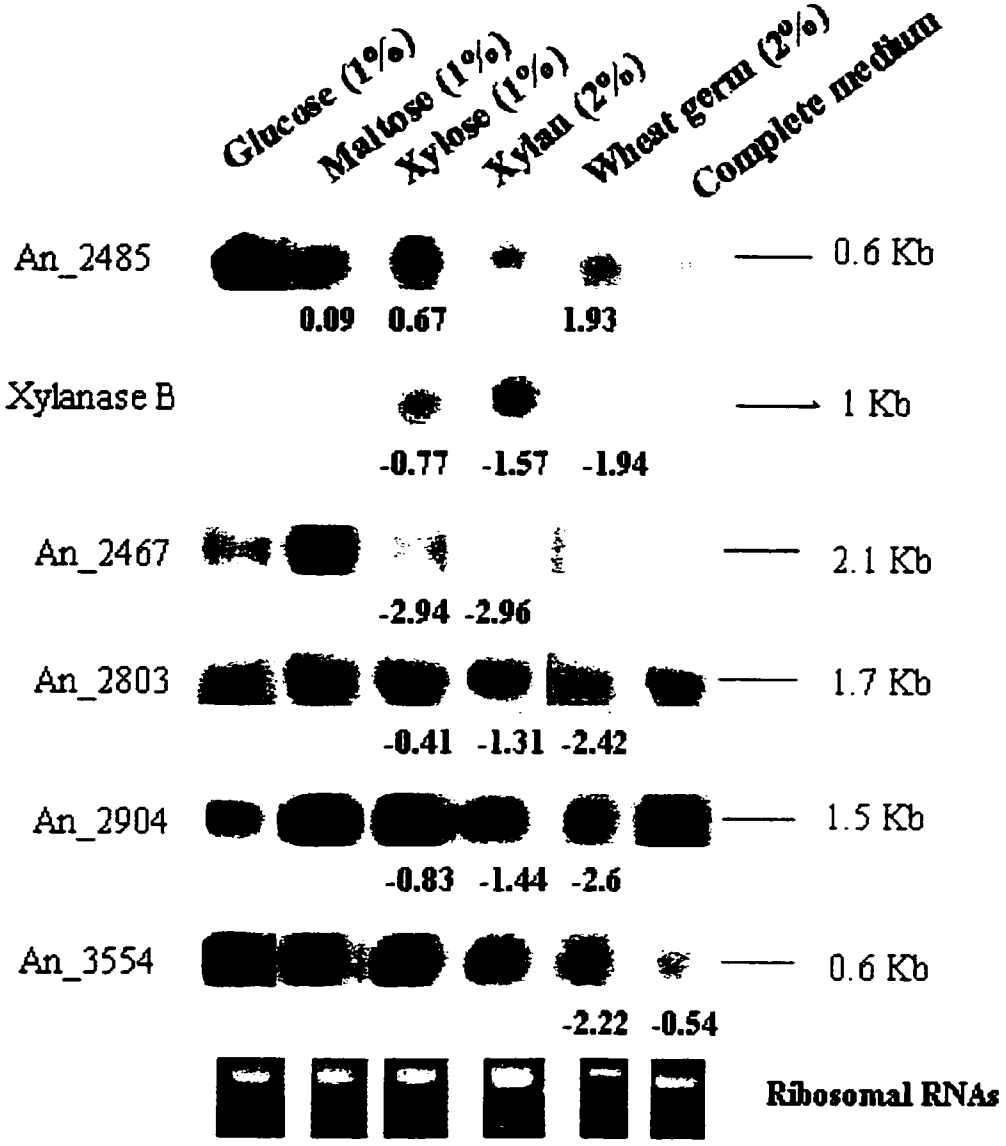
Four genes were assessed in 6 different culture conditions using RNA-blot analysis (Figure 17). The glucoamylase gene was first tested by labelling the clone

Table 9: List of the candidate carrier genes differentially and highly expressed.

Log_2ratio (Cy3/Cy5)					Secretory signal	Gene name	Description
Xylose	**CM**	**Maltose**	**Wg**	**Xylan**			
	1.83			1.15	YES	an_595	Similar to serine/threonine kinase
	-1.11		-2.04	-0.82	NO	an_2760	Ribosomal protein L10
-1.86		-0.47	-1.88	-1.34	NO	ANI9972	Pyruvate carboxylase
			-1.97	-1.13	NO	an_3118	Probable phosphate transport protein MIR1
			-1.61		NO	an_1995	Polyubiquitin
			-1.94	-1.07	YES	an_3131	Plasma membrane H+-ATPase
-0.21	1.54		-1.69	0.02	NO	an_668	Peroxisomal hydratase-dehydrogenase-epimerase
0.41	-1.66	0.67			YES	an_1008	Hypothetical protein
	-0.93		-2.00	-0.75	NO	an_3438	Heat shock protein
	-1.69	0.53	-1.79	-0.93	YES	an_661	Glycine-rich protein
-0.83			-2.60	-1.44	NO	an_2904	Glyceraldehyde 3-phosphate dehydrogenase
-2.94				-2.96	YES	an_2467	Glucoamylase
		0.62	-1.93		NO	an_3281	F-actin capping protein beta subunit
0.03		0.33	-1.61	0.85	NO	an_1230	Elongation factor 1 alpha
	-1.13	0.74	-1.54		NO	an_1040	Cobalamin-independent methionine synthase
-0.42	-1.40		-1.89	-0.77	YES	an_2369	ADP,ATP Carrier protein
0.46	-1.54	0.62	-1.27		NO	an_771	60S ribosomal protein L29 (L27A)
	-1.28	0.30	-1.87	-0.51	NO	an_2360	40S ribosomal protein S3A
		0.67	-1.80	-0.76	NO	an_3665	26S proteasome regulatory particle chain RPN12
				1.54	NO	an_3658	Unknown function
	1.64			0.51	NO	an_3635	Unknown function
	1.62			1.37	NO	an_3634	Unknown function
-0.25			-1.75	-1.11	NO	an_3573	Unknown function
	1.56			0.56	NO	an_3420	Unknown function
	1.52				NO	an_3386	Unknown function
	1.54			0.48	NO	an_3186	Unknown function
	1.60			0.62	NO	an_2930	Unknown function
-1.41			-2.34	-1.33	YES	an_2927	Unknown function
	1.57			0.65	NO	an_2666	Unknown function
				1.57	NO	an_2453	Unknown function
			-1.50	0.42	NO	an_2378	Unknown function
-0.69		0.39	-1.52	-0.34	NO	an_2201	Unknown function
-0.31	-0.82		-1.51	-0.64	NO	an_1664	Unknown function
			-1.98	-0.93	NO	an_963	Unknown function
	-0.61		1.94		YES	an_3591	Unknown function
-0.41			-2.42	-1.31	NO	an_2803	Unknown function
0.26			-1.56		NO	an_2642	Unknown function
	-1.69		-1.87		NO	an_2302	Unknown function
	1.68				NO	an_1390	Unknown function

Figure 17: RNA-blots of six genes under 6 different culture conditions

Aspergillus niger was grown for 2 days in CM, then transferred for 2 more days to MM with the different supplements. In parallel *A. niger* was grown also for two days in CM. Total RNA was extracted, resolved by denaturing gel electrophoresis, and blotted onto membranes. Autoradiographs obtained from probing the membranes for six genes (an_2485: thioredoxin, mitochondrial precursor; Xylanase B; an_2467: glucoamylase; an_2803: no known function; an_2904: glyceraldehyde-3-P-dehydrogenase; an_3554: thioredoxin) are shown. Those genes are also represented with the corresponding $\log_2$ratio (Cy3/Cy5) values generated from the microarray analysis. The RNA gels (before transfer) revealing the ribosomal RNAs were shown as indicative of the RNA quantity loaded.


Glucose (1%)
Maltose (1%)
Xylose (1%)
Xylan (2%)
Wheat germ (2%)
Complete medium
An_2485
0.6 Kb
0.09
0.67
1.93
Xylanase B
1 Kb
-0.77
-1.57
-1.94
An_2467
2.1 Kb
-2.94
-2.96
An_2803
1.7 Kb
-0.41
-1.31
-2.42
An_2904
1.5 Kb
-0.83
-1.44
-2.6
An_3554
0.6 Kb
-2.22
-0.54
Ribosomal RNAs

an_2467 with [α-^{32}P] dCTP and probing RNA-blots. One major band appeared at the expected size of 2.1 Kbp under the MM+maltose condition. With the glucose as the carbon source, the enzyme has been reported to be expressed one third of the level observed with maltose or starch (Nunberg *et al.*, 1984; Fowler *et al.*, 1990). Clearly a 2.1 Kbp lighter band appeared in the MM+glucose condition, confirming the expected pattern of glucoamylase expression previously characterized. In MM+xylan and MM+xylose, glucoamylase enzyme is not synthesized and as expected no band was present on the blot. For MM+wheat germ and CM, no band was present on the blot.

Five other genes were also tested including glyceraldehyde-3-P-dehydrogenase gene (clone an_2904), the clone an_2803 (no known function) and the thioredoxin gene (clone an_3554), the mitochondrial precursor of the thioredoxin gene (an_2485) and xylanase B gene. The two first clones showed a high expression level in all conditions except for the MM+wheat germ where there was a clear decrease. The thioredoxin gene (clone an_3554) was also showing a high expression in all conditions except in MM+wheat germ and CM where a decrease was noticed. The second thioredoxin gene (an_2485) showed mainly an increase in MM+glucose while the xylanase B gene was present under MM+xylose and MM+xylan conditions. Moreover the ribosomal RNAs appearing in the RNA gels were indicative of the total RNA quantity loaded in the gel. The total RNA was loaded evenly in all conditions except for the RNA from the MM+wheat germ where the quantity seemed slightly lower.

3.4 RT-PCR

Reverse transcription was performed with RNA from different conditions (Figure 18). The PCR reaction was set up using primers (GlaT5' and GlaT336) designed to amplify a small portion of the terminal region of the glucoamylase gene (336 bp). A PCR product with expected size was observable in MM+maltose, MM+glucose but not in MM+xylose and MM+xylan, as expected. A 336 bp PCR product was also present in MM+wheat germ and CM.

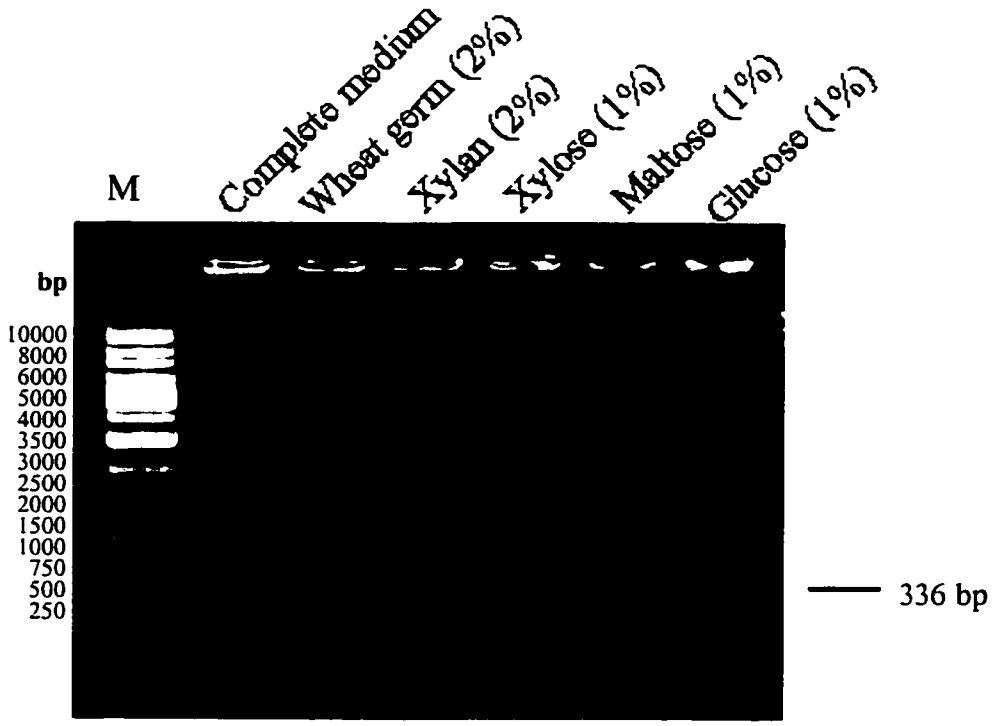


Figure 18: RT-PCR results using glucoamylase terminator region primers.

Cultures were grown 2 days in CM, then 2 days in MM with the 5 different supplements while the CM culture was grown for two days. *Aspergillus niger* total RNA was extracted, reverse transcribed, then PCR amplified using glucoamylase primers to obtain a 336 bp portion of the transcript. The 1 kb marker (M) was loaded in the first lane followed by the RT-PCR products in the next lanes.

4. Discussion

Aspergillus niger is an important industrial organism as host for protein production. Various improvements were applied to increase the secreted protein yield (Verdoes *et al.*, 1993; Maras *et al.*, 1999; Xu *et al.*, 2000). Nevertheless the complete knowledge of the gene expression pattern, or transcriptome has not been completely elucidated. Useful information such as the expression level of genes and their regulation could be obtained. Such information will be helpful in the rational design of expression systems for increasing protein production.

In recent years, several high-throughput techniques have been developed to analyze gene expression. These methods include the comparative expressed sequence tag (EST) sequencing, the PCR-based subtractive cloning, the serial analysis of gene expression (SAGE), differential display (DD) and microarray technology (Carulli *et al.*, 1998). The EST sequencing method presents some disadvantages: it requires automated DNA sequencing capabilities, it is not sensitive for lowly expressed mRNAs and the throughput is low compared to other methods (Adams *et al.*, 1991). The PCR-based subtractive cloning method also has poor sensitivity and low throughput. In contrast, the SAGE technology can achieve very high throughput. However, the procedure to generate good SAGE libraries and the analysis of the data are quite complex (Velculescu *et al.*, 1995). The DD technique has become popular in the past ten years but the approach generates a lot of false-positives (made during the PCR or the cloning processes) and the results are not quantitative. In conclusion, the characterization of highly expressed and

regulated genes under different conditions requires the quantitative expression comparison of thousands of genes and the most appropriate and suitable method to fulfill these requirements is microarray technology (Hegde *et al.*, 2000).

4.1 Cluster analysis

4.1.1 Results

The comparison of the gene expression profile using clustering analysis allows the preliminary characterizations of the gene regulatory network. The microarray experiments reported here were based on 1,700 genes tested in 6 different conditions: MM+glucose, MM+maltose, MM+xylose, MM+wheat germ, MM+xylan and complete medium. All conditions were referred to the control condition of MM+glucose to allow the comparison of gene expression for all conditions. Genes that showed at least a 3-fold increase or decrease were considered differentially expressed, and they corresponded to 88 genes (Table 8) that could be grouped into 10 clusters.

Results from the clustering analysis reveal that most of the differentially expressed genes (54 genes) are repressed when grown under the MM+wheat germ condition. These genes are involved in metabolism and protein synthesis, suggesting that wheat germ provides poor growing condition for the organism (Gasch *et al.*, 2000). This result was also reflected by the poor yield and instability of the RNA extracted in wheat germ compared to the other conditions (Table 6)

and also by the finding that housekeeping genes like glyceraldehyde-3-phosphate dehydrogenases were not expressed under this stressful condition (details in 4.1.2). Whereas wheat germ is a favourite for the growth of other fungi for protein production, the growth and gene expression were limited for *A. niger*.

Five genes seemed to be induced only in MM+wheat germ. These include a DNA-binding transcription factor gene, a gene involved in amino acid synthesis and the thioredoxin gene (an_3554). The DNA-binding transcription factor gene relates to the MM+wheat germ poor growing condition as this gene, in general, is induced during starvation (Tao *et al.*, 1999). Another gene identified as being a thioredoxin gene (but different from the former one according to sequence comparison, clone an_2485) had the opposite expression profile and was repressed in MM+wheat germ. The divergent expression of these two thioredoxin genes were verified by RNA-blot analysis. While the former thioredoxin (an_3554) showed an increase in the MM+wheat germ condition, similarly to the microarray analysis, the latter (an_2485) was not confirmed and was actually overexpressed in MM+glucose (Figure 17).

For CM, genes for 6 ribosomal proteins (clones an_2646, an_2255, an_1030, an_2586, an_2344 and an_771) were repressed. This is unexpected as usually rich growing condition promotes the expression of genes for ribosomal proteins (Wodicka *et al.*, 1997). Nevertheless, an excess of ribosomal proteins synthesized in rich condition can cause a negative autoregulation at the transcription level by destabilizing the ribosomal transcripts. This is done by direct binding of certain ribosomal proteins to the mRNAs coding for ribosomal proteins and, therefore,

decreases the level of transcripts (Tao *et al.*, 1999). As expected, the malate dehydrogenase gene involved in the energy/TCA cycle was repressed under complete medium condition, with 3-fold decrease compared to minimal conditions (Tao *et al.*, 1999). On the other hand, 16 genes were induced in complete medium, with at least a 3-fold increase. These include hemolysin, peroxisomal hydratase-dehydrogenase-epimerase, genes similar to serine/threonine kinase and several genes with unknown function (clones an_1084, an_3635, an_3634, an_3420, an_3386, an_3326, an_3186, an_2930, an_2826, an_2666, an_2197, an_1390 and an_1191).

The expression of some genes was not as expected; for example, the xylanase B gene and the D-xylose reductase gene. Both genes express products involved in D-xylose catabolism and while xylanase B degrades the complex heterogenous polymer xylan into D-xylose, D-xylose reductase reduces D-xylose to xylitol. Therefore we expected the genes to be expressed under MM+xylose and MM+xylan (van Peij *et al.*, 1998) and the induction of xylanase B under these conditions. This was confirmed by RNA-blot analysis. But according to the microarray results, xylanase B showed a 3-fold repression under MM+xylan ($\log_2$ratio(Cy3/Cy5) = -1.57) and no significant expression under MM+xylose (-0.77). D-xylose reductase revealed a repression under both conditions MM+xylan (-1.53) and MM+xylose (-2.11).

The results can be explained by a certain variability observed on microarray experiments. Indeed many comparisons between microarrays and more traditional techniques like the RNA-blot and RT-PCR were performed (Taniguchi *et al.*, 2001;

Seki et *al.*, 2001; Kucharski *et al.*, 2002) These studies demonstrated that even though the majority of the genes were confirmed by RNA-blots or RT-PCRs, certain variations were also revealed, especially for low level transcripts (Brown *et al.*, 2001). Microarray experiments were found to be slightly less sensitive than the RNA-blot technique and some differential expression reaching levels of 1000-fold using RNA-blot technique was almost not detectable by microarrays (Holland, 2002).

4.1.2 Data validation

The data were validated using RNA blot analysis and/or RT-PCR. First the well-characterized glucoamylase gene was used to verify the results. The glucoamylase enzyme is produced at higher level in the presence of maltose than in glucose, and with little or no synthesis in the presence of xylan or xylose (Nunberg *et al.*, 1984; Fowler *et al.*, 1990).

The RNA-blot confirmed that the glucoamylase gene was expressed in the minimal medium supplemented with maltose. With glucose as carbon source, the expression is decreased by about 3-fold. The RT-PCR experiment showed glucoamylase expression under MM+maltose, MM+glucose but also complete medium and MM+wheat germ. The glucoamylase expression in complete medium condition, also shown when using microarrays, was probably caused by the glucose (1%) contained in the medium. Glucoamylase was also expressed under MM+wheat germ due to the residual glucose induction, as the mycelium was first

grown for two days in complete medium and then transferred to MM+wheat germ medium.

In conclusion, the expected transcription profile of the well-characterized glucoamylase gene was successfully revealed by the microarray technique and confirmed by RNA-blot and RT-PCR techniques.

In addition to the glucoamylase gene, we verified a number of other genes by RNA-blot analysis to confirm the microarray results. The first one was the glyceraldehyde-3-phosphate dehydrogenase gene (clone an_2904), which is a housekeeping gene and therefore expressed under all growing conditions. The microarray and RNA-blot results showed an even expression in all conditions except in MM+wheat germ, where a decrease of about 6-fold for the microarray and close to 4-fold for the RNA-blot was detected. A second glyceraldehyde-3-phosphate dehydrogenase gene (clone an_636), different from the first one by sequence comparison, showed the exact same behavior and confirmed this result. These genes were clustered with the clone an_2803, which was also repressed under MM+wheat germ and showed a decrease of about 5-fold (in both microarray and RNA-blot results). The thioredoxin gene (clone an_3554) was verified and showed almost a 5-fold decrease in the microarray and 3.5-fold decrease in the RNA-blot analysis.

As several genes picked randomly had the same expression profile in the different conditions for both the microarray and RNA-blot techniques, the microarray data were validated and could be used for further analysis.

4.1.3 Potential carrier genes

Using the glucoamylase gene of *Aspergillus* species as a carrier gene was shown to be very promising for the expression of many heterologous proteins. Numerous examples with mammal proteins are available: chymosin in *A. awamori* with a yield of 1 g/L, hen egg white lysosyme in *A. niger* reached 1.2 g/L and the pancreatic prophospholipase A expressed in *A. niger* led to levels of 10 mg/L (Kinghorn and Unkles, 1994). Nevertheless other carrier genes could improve the expression level and be better adapted for the transport of proteins into the extracellular medium.

The search for potential carrier genes was initiated by first looking through the large-scale microarray experiment (containing 1,700 genes) for genes with both characteristics of being highly expressed at least in one condition and being regulated. These could permit the expression in large quantities of the desired protein and to control its expression by simply growing the organism in a different medium.

In total, 39 genes showed both characteristics (Table 9). Nevertheless, these genes can be used as carrier genes using a gene-fusion strategy with interesting enzymes. These carrier genes were not only induced in maltose like the glucoamylase gene but also in other conditions. Basically 23 genes were induced in maltose, 21 genes in glucose, 16 genes in xylose, 15 genes in complete medium, 7 genes in xylan and one gene in wheat germ.

These 39 candidate genes were also tested for secretory signal in their peptide sequence. According to the prediction server SignalP V1.1, 8 of these genes possessed a secretory signal. The gene-fusion strategy between the carrier genes (including their secretory signal) and the target genes (which are not normally secreted) will allow production and secretion to the extracellular medium of the desired proteins. From the 8 genes, 5 could be identified, including the well-characterized glucoamylase gene. The other genes included two transmembrane proteins (ATPase and ADP, ATP carrier protein), a serine/threonine kinase-like protein and a glycine-rich protein. The rest of the candidate genes whose peptide do not possess a secretory signal (31 genes) could still be fused with target genes whose products are secreted by including the secretory signal in the fusion.

4.2 Future improvements

The microarray technique was developed and optimized to assess the gene expression of *A. niger* clones analyzed in different growing conditions. These preliminary results were satisfactory but improvements need to be made.

First, the microarray chip carried only a small portion of the *A. niger* genome (1,700 genes out of 13,000 genes) and the analysis and determination of the cross-connected regulations, gene interactions and eventual pathways was difficult as the number of genes was limited. Moreover, the microarray experiments could be repeated several times (by increasing the number of biological replicates to 3 or 4) to further confirm the gene expression. As the technical replicates are not as

important as the biological replicates, the 2 technical replicates we performed should be enough. Besides further confirming the gene expression, some time-point experiments could be performed by taking RNA samples at different times during the *A. niger* growth. The transcription level varying through the time would confirm with certainty the expression of the genes (Klok *et al.*, 2002).

Before proceeding with microarray analysis, better methods than using the categorical set of criteria of the GenePix software could be employed to eliminate the improper spots. Tran *et al.* (2002) have argued that the data retained, after applying a mean and median correlation method, could be more accurate and numerous than the traditional GenePix method (in average 80% of the data retained).

Finally the quantitative methods like RNA-blot and RT-PCR could be applied to a bigger set of genes of the microarray in order to validate the microarrays results.

4.3 Conclusion

In conclusion, the design, execution and analysis of microarray data were successfully performed. The analysis of the 1,700 genes of *A. niger* in various conditions added with different supplements leaded to preliminary conclusions. First, the MM+wheat germ medium seems to not be a favorable growing condition for *A. niger*. Secondly, no apparent product seems to interfere with the glucoamylase gene production (no protease genes). Thirdly, we found 39 potential

carrier genes with a high and regulated expression. Eight of them also possess a secretory signal in their peptide sequence. Therefore other promoters and signal peptides other than the glucoamylase promoter could be employed to produce heterologous proteins in *A. niger*. Nevertheless, further microarray analysis and protein expression studies should be performed to verify the validity of these genes as potential carrier genes.

REFERENCES

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M.,Polymeropoulos, M.H., Xiao, H. Complementary DNA sequencing : expressed sequence tags and human genome project. Science 252, 1651-1656 (1991)

Archer, D.B., Jeenes, D.J., MacKenzie, D.A., Brightwell, G., Lambert, N., Lorne, G., Radford, S., and Dobson, C.M. Hen egg lysosyme expressed and secreted from *Aspergillus niger* is correctly processed and folded. Bio/technology 8, 741-745 (1990)

Archer, D.B. Enzyme production by recombinant Aspergillus. Bioprocess. Technol. 19, 373-393 (1994)

Archer, D.B., Jeenes, D.J., Mackenzie, D.A. Strategies for improving heterologous protein production from filamentous fungi. Antonie Van Leeuwenhoek 65, 245-250 (1994)

Archer, D.B., Peberdy, J.F. The molecular biology of secreted enzyme production by fungi. Crit. Rev. Biotechnol. 17, 273-306 (1997)

Atlas, R. and Bartha, R. Microbial ecology: Fundamentals and applications. Addison Wesley Publishing, Company, Inc., Phillipines. 3, 216-231 (1981)

Baldauf, S. L., and Palmer, J. D. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc. Natl. Acad. Sci. (USA) 90, 11558-11562 (1993)

Bednar, M. DNA microarray technology and applications. Med. Sci. Monit. 6, 796-800 (2000)

Beier, M. and Hoheisel, J.D. Production by quantitative photolithographic synthesis of individually quality checked DNA microarrays. Nuc. Ac. Res. 28, i-vi (2000)

Bekkers, A.C., Franken, P.A., Van den Bergh, C.J., Verbakel, J.M., Verheij, H.M., De Haas, G.H. The use of genetic engineering to obtain efficient production of porcine pancreatic phospholipase A2 by *Saccharomyces cerevisiae*. Biochim. Biophys. Acta. 1089, 345-351(1991)

Bennett, J.W. Molds, manufacturing and molecular genetics. In W.E. Timberlake, (ed.), Molecular genetics of filamentous fungi. Alan R. Liss, Inc., NY. (1985)

Berka, R.M., Ward, M., Wilson, L.J., Hayenga, K.J., Kodama, K.H., Carlomagno, L.P., and Thompson, S.A. Molecular cloning and deletion of the gene encoding aspergillopepsin A from *Aspergillus awamori*. Gene 86, 153-162 (1990)

Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, K.E. and Fill, N.P. Glucoamylase G1 and G2 from *Aspergillus niger* are synthesized from two different but closely related mRNAs. EMBO J. 3, 1097-1102 (1984)

Bos, C.J., Stam, P. and van der Veen, J.H. Interpretation of UV-survival curves of Aspergillus conidiospores. Mut. Res. Elsevier sci. pub. B.V. 197, 67-75 (1988)

Brazma, A. Gene expression data mining and analysis. DNA microarrays : gene expression applications, chapter 6: 106-126 (2001)

Broekhuijsen, M.P., Mattern, I.E., Contreras, R., Kinghorn, J.R., van den Hondel C.A. Secretion of heterologous proteins by *Aspergillus niger*: production of active human interleukin-6 in a protease-deficient mutant by KEX2-like processing of a glucoamylase-hIL6 fusion protein. J. Biotechnol. 31, 135-145 (1993)

Brown, A.J.P., Planta, R.J., Bailey, D.A., Butler, P.R., Cadahia, J.L., Cerdan, M.E., de Jonge, M., Gardner, D.C.J., Gent, M.E., Hayes, A., Kolen, C., Lombardia, L.J., Murad, A., Oliver, R.A., Sefton, M., Thevelein, J.M., Tournu, H., Delft, Y., Verbart, D.J., Winderickx, J. and Oliver, S.G. Transcript analysis of 1003 novel yeast genes using high-throughput northern hybridizations. J. E.M.B.O. 20, 3177-3186 (2001)

Brown, P. and Botstein, D. Exploring the new world of the genome with DNA microarrays. Nat. Gen. 21, 33-37 (1999)

Bruns, T.D., White, T.J., and Taylor, J.W. Fungal molecular systematics. An. Rev. Eco. Sys. 22, 525-564 (1991)

Bu'Lock, J.D. Swings and roundabouts for citric acid producers. Biotechnol. Insight 84: 5-6 (1990)

Caponigro, G., Muhlrad, D. and Parker, R. A small segment of the MATα1 transcript promotes mRNA decay in S. cerevisiae: a stimulatory role for rare codons. Mol. Cell Biol. 13, 514-518 (1993)

Carulli, J.P., Artinger, M., Swain, P.M., Root, C.D., Chee, L., Tulig, C., Guerin, J., Osborne, M., Stein, G., Lian, J., and Lomedico, P.T. High throughput analysis of differential gene expression. J. Cell. Biochem. 30/31, 286-296 (1998)

Chakravarti, A. Population genetics-making sense out of sequence. Nat. Gen. 21, 56-60 (1999)

Chen, Y. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Opt. 2, 364-374 (1997)

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W. A genome-wide transcriptional analysis of the mitotic cell-cycle. Mol. Cell. 2, 65-73 (1998)

Conesa, A., Punt, P.J., van Luijk, N., van den Hondel, C.A. The secretion pathway in filamentous fungi: a biotechnological view. Fungal Genet. Biol. 33, 155-171 (2001)

Cronin, M.T. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. Hum. Mut. 7, 244-255 (1996)

DeRisi, J.L., Iyer, V.R., and Brown, P.O. Exploring the Metabolic and Genetic Control of Gene expression on a genomic scale. Science 278, 680-686 (1997)

Doolittle, R.F., Feng, D.F., Tsang, G., Cho, G. and Little, E. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271, 470-477 (1996)

Dunn-Coleman, N.S., Bloebaum, P., Berka, R.M., Bodie, E., Robinson, N., Armstrong, G., Ward, M., Przetak, M., Carter, G.L., LaCost, R., Wilson, L.J., Kodama, K.H., Baliu, E.F., Bower, B.,Lamsa, M. and Heinsohn, H. Commercials level of chymosin production by *Aspergillus*. Bio/technology 9, 976-982 (1991)

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95, 14863-14868 (1998)

Eisen, M.B. and Brown, P.O. DNA arrays for analysis of gene expression. Methods Enzymol 303,179-205 (1999)

Emtage, J.S., Angal, S., Doel, M.T., Harris, T.J., Jenkins, B., Lilley, G., and Lowe, P.A. Synthesis of calf prochymosin (prorennin) in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 80, 3671-3675 (1983)

Finkelstein, D.B., Rambosek, J., Crawford, M.S., Soliday, C.L., Mc Ada, P.C., and Leach, J. Genetics and Molecular biology of industrial microorganisms. Am. Soc. Microbiol. 295-300 (1989)

Finkelstein, D., Ewing, R., Gollub, J., Sterky, F., Cherry, M., and Somerville, S. Microarray data quality analysis: lessons from the AFGC project. Plant. Mol. Biol. 48, 119-131 (2002)

Fodor, S.P.A. Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. Light-directed, spatially addressable parallel chemical synthesis. Science 251, 767-773 (1991)

Fowler, T., Berka, R.M., Ward, M. Regulation of the glaA gene of *Aspergillus niger*. Curr. Genet. 18, 537-545 (1990)

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell. Vol.11, 4241-4257 (2000)

Gasch, A.P. and Eisen, M.B. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. Gen. Biol. 3, 0059.1-0059.22 (2002)

Giardina, T., Gunning, A.P., Juge, N., Faulds, C.B., Furniss, C.S.M., Svensson, B., Morris, V.J. and Williamson, G. Both binding sites of the starch-binding domain of Aspergillus niger glucoamylase are essential for inducing a conformation change in amylose. J. Mol. Biol. 313, 1149-1159 (2001)

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531-537 (1999)

Gordon, C.L., Archer, D.B., Jeenes, D.J., Doonan, J.H., Wells, B., Trinci, A.P., Robson, G.D. A glucoamylase::GFP gene fusion to study protein secretion by individual hyphae of *Aspergillus niger*. J. Microbiol. Methods, 42, 39-48 (2000)

Gouka, R.J., Punt, P.J., van den Hondel, C.A.M.J.J. Efficient production of secreted protein by *Aspergillus* : progress, limitations and prospects. Appl. Microbiol. Biotechnol. 47, 1-11 (1997)

Grimmond, S. and Greenfield, A. Expression profiling with cDNA microarrays : a user's perspective and guide. DNA microarrays : Gene expression applications, 15-33 (2001)

Hacia, J.G. Resequencing and mutational analysis using oligonucleotide microarrays. Nat. Gen. 21, 42-47 (1999)

Hawksworth, D. L. The fungal dimension of biodiversity: magnitude, significance, and conservation. Mycol. Res. 95, 641-655 (1991)

Hawksworth, D. L., Kirk, P. M., Sutton, B. C., and Pegler, D. N. Ainsworth and Bisby's Dictionary of the Fungi (8th Ed.) 616 (1995)

Hedenfalk, I. Gene-expression profiles in hereditary breast cancer. N. Engl. J. Med. 344, 539-548 (2001)

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N. and Quackenbush, J. A concise guide to cDNA microarray analysis. Biotech. 29, 548-562 (2000)

Hentze, M.W. Determinants and regulation of cytoplasmic mRNA stability in eukaryotic cells. Biochim. Biophys. Acta. 1090, 281-292 (1991)

Herrero, J., Valencia, A. and Dopazo, J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. Bioinfo. 17, 126-136 (2001)

Holland, M.J. Transcript abundance in yeast varies over six orders of magnitude. J. Biol. Chem. 277, 14363-14366 (2002)

Ingold, C.T. The biology of fungi. Hutchinson Educational Ltd. London, UK (1961)

Jeenes, D.J., Marczinke, B., MacKenzie, D.A., Archer, D.B. A truncated glucoamylase gene fusion for heterologous protein secretion from *Aspergillus niger*. FEMS Microbiol. Let. 107, 267-271(1993)

Jeenes, D.J., Mackenzie, D.A., Archer, D.B. Transcriptional and post-transcriptional events affect the production of secreted hen egg white lysozyme by *Aspergillus niger*. Transgen. Res. 3, 297-303(1994)

Kaplan, J.M., Kim, S.H., North, K.N., Rennke, H., Correia, L.A., Tong, Q.H., Mathis, B.J., Rodriguez-Perez, J.C., Allen, P.G., Beggs, A.H., and Pollak, M.R. Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. Nat. Genet. 24, 251-256 (2000)

Khalaj, V., Brookman, J.L., Robson, G.D. A study of the protein secretory pathway of *Aspergillus niger* using a glucoamylase-GFP fusion protein. Fungal Genet. Biol. 32, 55-65 (2001)

Kinghorn, J.R. and Unkles, E. Molecular genetics and foreign proteins in the genus Aspergillus. Aspergillus. Plenum Press Pub. Corp., 65-100 (1994)

Klok, E.J., Wilson, I.W., Wilson, D., Chapman, S.C., Ewing, R.M., Somerville, S.C., Peacock, W.J., Dolferus, R., and Dennis, E.S. Expression Profile Analysis of the Low-Oxygen Response in Arabidopsis Root Cultures. Plant Cell 14, 2481-2494 (2002)

Kohonen, T. Self-Organizing Maps. (Springer, Berlin, 1997)

Kozakiewicz, Z and Smith, D. Physiology of Aspergillus. Aspergillus. Plenum Press Pub. Corp., 23-40 (1994)

Khrapko, K.R., Lysov, Y.P., Khorlyn, A.A., Shick, V.V., Florentiev, V.L., and Mirzabekov, A.D. An oligonucleotide hybridization approach to DNA sequencing. FEBS letter. 256, 118-122 (1989)

Kucharski, R. and Maleszka, R. Evaluation of differential gene expression during behavioral development in the honeybee using microarrays and northern blots. Gen. Biol. 0007.1-0007.9 (2002)

Kuhn, K.M., DeRisi, J.L., Brown, P.O., and Sarnow, P. Global and specific translational regulation in the genomic response of Saccharomyces cerevisiae to a rapid transfer from a fermentable to a nonfermentable carbon source. Mol. Cell. Biol. 21, 916-927 (2001)

Kurella, M., Hsiao, L.L., Yoshida, T., Randall, J.D., Chow, G., Sarang, S.S., Jensen, R.V., and Gullans S.R. DNA microarray analysis of complex biological processes. J. Am. Soc. Nephrol. 12, 1072-1078 (2001)

Lee, J.K. Analysis issues for gene expression array data. Clinic. Chem. 47, 1350-1352 (2001)

Li, X., Gu, W., Mohan, S., and Baylink, D.J. DNA microarrays: their use and misuse. Nat. Pub. Gr. 9, 13-22 (2002)

Liang, P. and Pardee, A.B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257, 967-971 (1992)

Lipshutz, R.J. Morris, D., Chee, M., Hubbell, E., Kozal, M.J., Shah, N.,Shen, N., Yang, R., Fodor, S.P.A. Using oligonucleotides probe arrays to access genetic diversity. Bio. Tech. 19, 442-447 (1995)

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.C., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. Expression monitoring by hybridization to high density oligonucleotide arrays. Nat. Biotech. 14, 1675-1680 (1996)

Loriod, B. cDNA macroarrays and microarrays on nylon membranes with radioactive detection. DNA microarrays: gene expression applications, 57-84 (2001)

Maras, M., van Die, I., Contreras, R., and van den Hondel, C.A.M.J.J. Filamentous fungi as production organisms for glycoproteins of bio-medical interest. Glycoconjugate J. 16, 99-107 (1999)

May, G.S. and Adams, T.H. The importance of fungi to man. Gen. Res. Cold Spring harbor Lab. Press 7, 1041-1044 (1997)

McGall, G., Labadie, J., Brock, P. Light-directed synthesis of high-density oligonucleotide arrays using semi-conductor photoresistance. Proc. Natl. Acad. Sci. USA 93, 13555-13560 (1996)

Mellman, I. and Simons, K. The Golgi complex: In vitro veritas? Cell 68, 829-840 (1992)

Mellor, J., Dobson, M.J., Roberts, N.A., Tuite, M.F., Emtage, J.S., White, S., Lowe, P.A., Patel, T., Kingsman, A.J., and Kingsman, S.M. Efficient synthesis of enzymatically active calf chymosin in *Saccharomyces cerevisiae*. Gene 24, 1-14 (1983)

Miall, L.M. Historical development of the fungal fermentation industry. Vol.1, Industrial mycology. London 104-107 (1975)

Milner, N., Mir, K.U. and Southern, E.M. Selecting effective antisense reagents on combinatorial. Nat. Biotech. 15, 537-541 (1997)

Mittal, V. DNA array technology. Molecular cloning. Sambrook, J. and Russell, D.W. Vol.3 Appendix A10.1-A10.19 (2001)

Mousses, S., Bittner, M.L., Chen, Y., Dougherty, E.R., Baxevanis, A., Meltzer, P.S. and Trent, J.M. Gene expression analysis by cDNA microarrays. Chapter6, 113-136 (1999).

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot. Engin., 10, 1-6 (1997).

Nunberg, J.H., Meade, J.H., Cole, G., Lawyer, F.C., McCabe, P., Schweickart, V., Tal, R., Wittman, V.P., Flatgaard, J.E., Innis, M.A. Molecular cloning and characterization of the glucoamylase gene of *Aspergillus awamori*. Mol. Cell. Biol. 4, 2306-2315 (1984)

Nyyssonen, E., and Keranen, S. Multiple roles of the cellulase CBHI in enhancing production of fusion antibodies by the filamentous fungus *Trichoderma reesei*. Curr. Genet. 28, 71-79 (1995)

Punt, P.J. and van den Hondel, C.A.M.J.J. Analysis of transcription control sequences in filamentous fungi. Molecular biology of filamentous fungi. Berlin, 24-29, 177-187 (1991)

Quackenbush, J. Computational analysis of microarray data. Nature 2, 418-427 (2001)

Ramaswamy, S., and Golub, T.R. DNA microarrays in clinical oncology. J. Clin. Oncol. 20, 1932-1941 (2002)

Raper, K.B. and Fennel, D.I. The genus Aspergillus. Williams and wilkins Co., Baltimore, MD (1965)

Roberts, I.N., Jeenes, D.J., MacKenzie, D.A., Wilkinson, A.P., Sumner, I.G., and Archer, D.B. Heterologous gene expression in *Aspergillus niger*: a glucoamylase-porcine pancreatic prophospholipase A_2 fusion protein is secreted and processed to yield mature enzyme. Gene 122, 155-61 (1992)

Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J.C., Lashkari, D., Shalon, D., Myers, T.G., Weinstein, J.N., Botstein, D. and Brown, P.O. Systematic variation in gene expression patterns in human cancer cell lines. Nat. Genet. 24, 227-235 (2000)

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. Finding DNA regulatory motifs within aligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat. Biotech. 16, 939-945 (1998)

Saiki, R.K., Walsh, P.S., Levenson, C.H. and Erlich, H.A. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci. USA 86, 6230-6234 (1989)

Sambrook, J. and Russell, D.W. Molecular cloning. Cold Spring Harbour Lab., Protocol 6, 7.31 & Protocol 7, 7.35 (2001)

Samson, R.A. Taxonomy-Current concepts of Aspergillus systematics. Aspergillus. Plenum Press Publ. Corp., 1-22 (1994)

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. USA 93, 10614-10619 (1996)

Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., Somerville, S.C. and Manners, J.M. Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc. Natl. Acad. Sci. USA 97, 11655-11660 (2000)

Schulze, A. and Downward, J. Analysis of gene expression by microarrays: cell biologist's gold mine or minefield? J. Cell Sc. 113, 4151-4156 (2000)

Schulze, A. and Downward, J. Navigating gene expression using microarrays-a technology review. Nat. Cell Biol. 3, E190-E195 (2001)

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y. and Shinozaki, K. Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13, 61-72 (2001)

Sohail, M., and Southern, E.M. Hybridization of antisense reagents to RNA. Curr. Opin. Mol. Ther. 2, 264-271 (2000)

Southern, E.M., Milner, N., and Mir, K.U. Discovering antisense reagents by hybridization of RNA to oligonucleotide arrays. Ciba Found. Symp. 209, 38-44; discussion 44-46 (1997)

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Bolstein, D., and Futcher, B. Comprehensive identification of cell-cycle regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol. Biol. Cell 9, 3273-3297 (1998)

Spencer, J.A., Jeenes, D.J., MacKenzie, D.A., Haynie, D.T., and Archer, D.B. Determinants of the fidelity of processing glucoamylase-lysozyme fusions by *Aspergillus niger*. Eur. J. Biochem. 258, 107-112 (1998)

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub, T.R. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA. 96, 2907-2912 (1999)

Taniguchi, M., Miura, K., Iwao, H. and Yamanaka, S. Quantitative assessment of DNA microarrays - Comparison with Northern Blot analyses. Genomics 71, 34-39 (2001)

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. Functional genomics: expression analysis of Escherichia coli growing on minimal and rich media. J. Bacterio. 181, 6425-6440 (1999)

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. and Church, G.M. Systematic determination of genetic network architecture. Nat. Gen. 22, 281-285 (1999)

Tessier, D.C., Thomas, D.Y. and Brousseau, R. A DNA microarray Fabrication strategy for research laboratories. BRI website (http://www.bri.nrc.ca) (1997)

Thom, C. and Raper, K.B. A manual of the Aspergilli. Williams & Wilkins. Baltimore, MD (1945)

Tibshirani, R., Hastie, T., Eisen, M., Ross, D., Botstein, D. and Brown, P.O. Clustering methods for the analysis of DNA microarray data. (1999)

Ton, C., Stamatiou, D., Dzau, and V., Liew, C. Construction of a zebrafish cDNA microarray: gene expression profiling of the zebrafish during development. Biochem. Biophys. Res. Commun. 296, 1134-1139 (2002)

Toronen, P., Kolehmainen, M., Wong, G., and Castren, E. Analysis of gene expression data using self-organizing maps. FEBS letter 451, 142-146 (1999)

Tran, P.H., Pfeiffer, D.A., Shin, Y., Meek, L.M., Brody, J.P., and Cho, K.W.Y. Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. Nuc. Ac. Res. 30, No 12 (2002)

Tsuchiya, K., Nagashima, T., Yamamoto, Y., Gomi, K., Kitamoto, K., Kumagai, C., and Tamura, G. High level secretion of calf chymosin using a glucoamylase-prochymosin fusion gene in Aspergillus. Biosci. Biotechnol. Biochem. 58, 895-899 (1994)

Tsukahara, T., Tsujino, S., and Arahata, K. cDNA microarray analysis of gene expression in fibroblasts of patients with X-linked Emery-Dreifuss muscular dystrophy. Mus. Ner. 25, 898-901 (2002)

van Brunt, J. Fungi, the perfect hosts? Bio. Techn. 4, 1057-1062 (1986)

van den Hombergh, J.P., van de Vondervoort, P.J., Fraissinet-Tachet, L. and Visser, J. Aspergillus as a host for heterologous protein production : the problem of proteases. Trends Biotechnol. 15, 256-263 (1997)

van Peij, N., Gielkens, M., de Vries, R.P., Visser, J. and de Graaff, L.H. The transcriptional activator XlnR regulates both xylanotic and endoglucanase gene expression in *Aspergillus niger*. App. Env. Microbiol. 64, 3615-3619 (1998)

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. Serial analysis of gene expression. Science 270, 484-487 (1995)

Verdoes, J.C., Punt, P.J., Schrickx, J.M., van Verseveld, H.W., Stouthamer, H., and van den Hondel, C.A.M.J.J. Glucoamylase overexpression in Aspergillus niger: molecular genetic analysis of strains containing multiple copies of the glaA gene. Transgen. Res. 2, 84-92 (1993)

Verdoes, J.C. Molecular genetic studies of the overproduction of glucoamylase in *Aspergillus niger*. Chapter 1: Strain improvement strategies for the overproduction of fungal proteins by filamentous fungi, 7-29 (1994)

Verdoes, J.C., Calil, M.R., Punt, P.J., Debets, F., Swart, K., Stouthamer, A.H. and van den Hondel, C.A.M.J.J. The complete karyotype of *Aspergillus niger*: the use of introduced electrophoretic mobility variation of chromosomes for gene assignment studies. Mol. Gen. Genet. 244, 75-80 (1994a)

Verdoes, J.C., Punt, P.J., Stouthamer, A.H. and van den Hondel, C.A.M.J.J. The effect of multiple copies of the upstream region of the *Aspergillus niger* glucoamylase-encoding gene on expression. Gene 145, 179-187 (1994b)

Wainright, P. O., Hinkle G., Sogin M. L., and Stickel, S. K. Monophyletic origins of the Metazoa: an evolutionary link with fungi. Science 260, 340-342 (1993)

Wallis, G.L., Swift, R.J., Atterbury, R., Trappe, S., Rinas, U., Hemming, F.W., Wiebe, M.G., Trinci, A.P., Peberdy, J.F. The effect of pH on glucoamylase production, glycosylation and chemostat evolution of *Aspergillus niger*. Biochim. Biophys. Acta. 1527, 112-122 (2001)

Wang, D.G. Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. Science 280, 1077-1082 (1998)

Williamson, M.P. Function of conserved tryptophans in the Aspergillus niger glucoamylase 1 Starch Binding Domain. Biochem. 36: 7535-7539 (1997)

Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., and Lockhart, D.J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat. Biotech. 15, 1359-1367 (1997)

Wosten, H.A., Moukha, S.M., Sietsma, J.H., Wessels, J.G. Localization of growth and secretion of proteins in *Aspergillus niger*. J. Gen. Microbiol. 137, 2017-2023 (1991)

Xu, J., Wang, L., Ridgway, D., Gu, T., Moo-Young, M. Increased heterologous protein production in *Aspergillus niger* fermentation through extracellular proteases inhibition by pelleted growth. Biotechnol. Prog. 16, 222-227 (2000)

APPENDICES

Appendix A. Microarray methodology

There are five basic and important steps in performing a DNA microarray experiment:

- Processing and printing of the cDNA clones to generate the chip
- Sample RNA isolation
- Preparation of the probe (e.g., reverse transcription and labelling)
- Hybridization of the labelled probe DNA to the DNA spotted on the substrate
- Image acquisition of hybridization results and image analysis

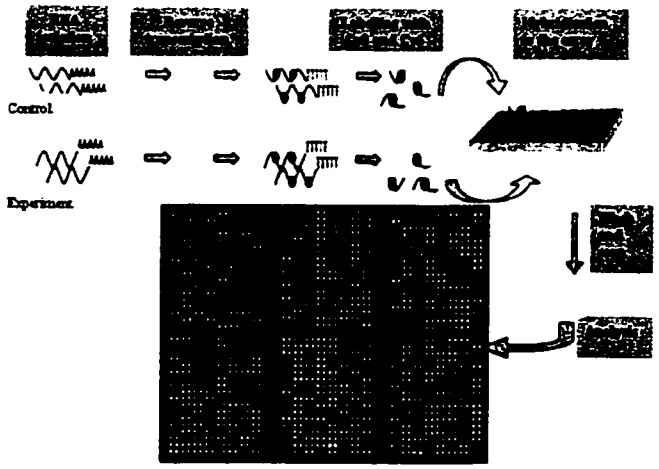


Figure 5: Representation of the Microarray experiment main steps.

1.1 Immobilization of DNA onto glass slides

The genes that are chosen to be printed onto the microarray depends on the availability of the genes and clones for a particular organism, on the size-capacity of the slide but overall it depends on the purpose of the microarray. In our experiment, we printed genes with a known function and others that we ignore the role. This array could allow the characterization of the gene expression pattern of known and unknown genes and, consequently, the eventual discovery of new genes (Mittal, 2001).

Genes can be robotically spotted onto the modified microscopes slides and constitute cDNA arrays. In our experiment these genes are PCR products that have been amplified then purified through a column to remove unwanted salts, detergents, PCR primers and proteins that could affect the spotting and the following hybridization step. Typically a few nanoliters (2 nl) corresponding to a concentration of 100 - 500 μg/ml is loaded into a spotting pin by capillary action and is deposited onto the matrix by direct physical contact between the pin and the substrate (DeRisi *et al.*, 1997). Usually each DNA sample is printed in duplicate in two adjacent spots to test the reproducibility of the printing but also the hybridization process (Saiki *et al.*, 1989). The slides are let to air-dry then UV-cross linked to form covalents bonds between the thymidine residues in the DNA and the positively charged amine groups on the silane slides (Cheung *et al.*, 1999). To achieve maximal hybridization and denaturation of the spotted DNA, the cDNA arrays are also baked.

1.2 RNA Extraction

Genes coding for a protein are transcribed into messenger RNAs (mRNAs) in the cell nucleus. Messengers RNAs account for 1% of total RNA in *Aspergillus niger*. These mRNAs are translated into proteins by ribosomes in the cytoplasm. The transcription level of a gene is taken to be the amount of its corresponding mRNA present in the cell. Comparative hybridization experiments compare the amounts of many different mRNAs in two cell populations. This comparative study is done in parallel and give two kinds of information: a **static information** about the expression of a particular gene (expressed or not in specific conditions) and a **dynamic information** about the expression pattern of this gene compared to the others or compared at different time-points (DeRisi *et al.*, 1997). Total RNA was extracted from the *Aspergillus niger* mycelia and the purity of the RNA is critical for the hybridization performance as any cellular components (cellular protein, lipid and carbohydrate) can bind non-specifically to the fluorescent-labelled probe.

1.3 Reverse transcription into cDNAs and fluorescent labelling

The purified RNA undergone reverse transcription into cDNAs by using oligo(dT) for primer, which limit in theory the presence of ribosomal RNAs in the hybridization solution (BRI, personal communication). Using oligo(dT) as primer permits also to label the probe at the 3'end of the gene directly complementary to the targeted immobilized EST sequence on the array (DeRisi *et al.*, 1997). Besides the

reverse transcription into more stable form (DNA) prevents the experimental samples from degradation, as the environment is full of RNA-digesting enzymes. Moreover during this reverse transcription, modified nucleotides (aa-dUTP) were added into the reaction mixture, which are incorporated far more efficiently into the cDNA than the traditional Cy-dye-labelled dNTPS (direct labelling). The fluorophores dyes were then bound with their ester group to the free amino groups present on the modified nucleotides incorporated in the cDNA strand (Grimmond and Greenfield, 2001).

In order to detect cDNAs attached to the microarray, those must be labelled with a reporter molecule identifying their presence. The reporters currently used in microarrays experiments are the fluorescent dyes Cy3 and Cy5. These two dyes are frequently paired because of the high incorporation efficiency, the good photostability and yield, and the wide separation of their excitation and emission spectra, allowing highly discriminating optical filtration (DeRisi *et al.*, 1997). We called the labelled cDNA samples probes because they are used to probe the collection of spots on the array (but the cDNA immobilized can also be referred as probes as they would be actually the probes applied to the blot in a typical RNA-blot experiment). The colors are not directly observed on the array, and the slides need to be scanned to observe a green (Cy3), red (Cy5) or yellow (Cy3+Cy5) color coming from the light emitted by the dye and capted by a detector measuring the fluorescence. The number of fluorophore molecules attached to the cDNA depends on its length and possibly its sequence composition. The incorporation of the dye is in average about 1 fluorophore every 50 bp for the indirect labelling (i.e. reverse

transcription followed by labelling separately). As said earlier, the direct labelling (reverse transcription with labelling simultaneously) is less efficient and leads to the incorporation of 1 fluorophore every 200 bp (TIGR protocol available at *http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml*).

1.4 Hybridization to the cDNA microarray

The two cDNA probes are tested by hybridization onto the cDNA microarray. The array can hold until thousands of spots with each of them containing a different DNA sequence. If the probe contains a cDNA whose sequence is complementary to the DNA on a given spot, that cDNA will hybridize to the spot, where it will be detectable by its fluorescence. In this way, every spot on an array is an independent assay for the presence of a particular cDNA. There is enough DNA on each spot that both probes can hybridize to it at once without interference. The use of differentially labelled mixtures avoids most of the complications of hybridization kinetics (by using ratio) (Brown and Botstein, 1999). The sealed humidifier chamber prevents the microarray slide to dry and the hybridization solution to evaporate ruining then the experiment (Grimmond and Greenfield, 2001).

1.5 Scanning the Hybridized Array

The relation between the amount of RNA for a given gene and the intensity of the spot signal on the array is quite complex. This depends on a multitude of factors

such as the labelling method, hybridization conditions, target features and the gene sequence. Therefore the microarray technique is better used to assay the **relative representation** of RNA in two or more samples than considering one RNA species and its **absolute amount**. Thus the relative representation of a gene in the two samples is measured by taking the ratio of the fluorescence intensities of the two dyes (Eisen and Brown, 1999).

A good signal corresponds to the best signal-to-noise ratio. Efficient removal of unincorporated products, avoiding evaporation of hybridization solution while on the microarray and cleanliness of reagents therefore must be a priority (Grimmond and Greenfield, 2001). Also very important factors include the amount of cDNA immobilized on support, the amount of labelled RNA, the specific activity, the hybridization kinetics and finally the exposure time (Loriod *et al.*, 2001).

Once the cDNA probes have been hybridized to the array and any loose probe has been washed off, the array must be scanned to determine how much of each probe is bound to each spot. The probes tagged with fluorescent reporter molecules emit detectable light when stimulated by a laser. The emitted light is captured by a detector, either a charge-coupled device (CCD) or a confocal microscope, which records its intensity. The device used in these experiments is the laser scanning confocal microscope from Axon Instruments, Inc. The scanner produced a light with a wavelength appropriate for each dye (532nm for Cy3 and 635nm for Cy5). This light passed through a regular microscope objective and then excited a small area on the array. The array emitted fluorescence with a wavelength different from the excitation wavelength. The emission light is separated from all other unwanted lights

by using a series of mirrors, filters and lenses, and finally converted to an electrical signal by the photomultiplier tubes (PMTs) (Schena *et al.*, 1998; Eisen and Brown, 1999). Spots with more bound probe will have more reporters and will therefore fluoresce more intensely. It is possible to adjust the PMTs in order to balance the two channels and obtain the same amount of red and green signal resulting in a pixel ratio of approximately 1.0.

1.6 Image analysis and normalization

After scanning, the computerized images made of thousands of spots have different "color behaviours". Most of the yellow spots correspond to the genes that are expressed in both experimental conditions, and have therefore the same amount of labelled probe bound to them. Spots whose mRNAs are present at a higher level in one or the other tested condition show up as predominantly red or green. The ratio of fluorescent intensities for a spot is interpreted as the ratio of concentrations for its corresponding mRNA in the two samples. And it was demonstrated that changes of as little as a factor of two can be detected by the microarray technique and be confirmed later by alternate form of RNA blotting (Schena *et al.*, 1996).

The first step before any analysis is the subtraction of the background to reduce the effect of non-specific fluorescence like the autofluorescence of the glass slide or the non-specific binding of both samples (control and tested experiments). Basically two methods are available: first for the local background subtraction, a different background value is computed for each spot. This value corresponds to the median

value of the pixels comprised in a circle with a 3-times diameter compared to the actual spot diameter. Also the spot pixels and a two-pixel ring around the spot are excluded.

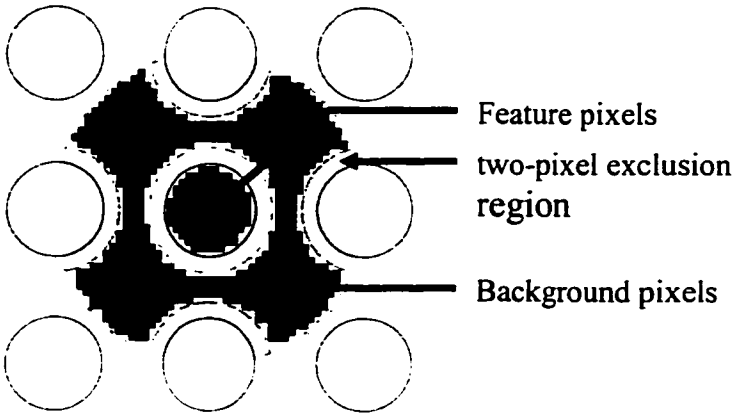


Figure 6: Schematic representation of the background subtraction.

The second method called the global background subtraction use only one value to subtract from the all slide. This value is the average of all local median background values of all spots of the array. This method assumes that the background is uniform through the entire slide, which is rarely the case. Therefore we used for our experiments the local background subtraction method.

After this step, we needed to normalize the relative fluorescence intensities in each of the two channels. Indeed differences at each step in the experiment can cause variations and shift in the average ratio of the two dyes Cy5 and Cy3. These error sources are the variations in the labelling and detection efficiencies, chemical properties of different dyes, pin tips, slides batches as well as the quantity of the

starting RNA (Quackenbush, 2001). The type of normalization must be chosen carefully to correct most of these variations. Also by normalizing each array, this will allow the comparison across many arrays.

Three major normalization methods can be used depending on the characteristics of the experiment, on the known and expected sources of error. First, the total intensity normalization is based on the assumption that most of the genes are not differentially expressed from one condition to the other, and that some genes will be up-regulated and others down-regulated. Consequently, the total integrated intensity computed for all genes should be the same in both channels and the mean of the ratio of the median intensities of all features is equal to 1.0. If the number is different than 1.0, a normalization factor is then generated and applied to the whole slide to rescale it to 1.0. This total intensity normalization is suitable for arrays with a large number of genes or moderate number of genes but randomly selected (Li *et al.*, 2002).

Secondly, the normalization using control genes assume that a group of genes have a consistent behaviour in the conditions tested (like housekeeping genes) and therefore must have the mean of the control features equal to 1.0. Assuming that the slides are uniform across the array, a normalization factor can be calculated from these features and then applied to the whole array (Chen *et al.*, 1997). This normalization is adequate for arrays with a small number of genes or for nylon membrane-based arrays (Li *et al.*, 2002).

Finally, the normalization using regression techniques is based on the scatter plot of Cy5 versus Cy3 intensities (or their logarithm) in which the genes are aligned

with a slope equal to 1.0 if the labelling and the detection efficiencies are the same for both samples. In the normalization process, the intensities will be adjusted using regression techniques to have a slope of 1.0 (Hedenfalk *et al.*, 2001).

Appendix B. Clustering methods

"Clustering is the most popular method currently used in the first step of gene expression matrix analysis. The goal of clustering is to put together objects with similar properties" (Brazma, 2001). Indeed the clustering process is quite efficient to group thousands of gene expression values based on their associations and in identifying patterns of their co-expression (Lee, 2001). However this task is not simple and a perfect and single correct classification does not exist yet. But several unsupervised approaches might be more or less appropriate for different data sets, and actually combining them together can lead to a significant result. The different clustering approaches for expression profiles are:

-hierarchical (complete linkage, average linkage and single linkage)

-partitioning based clustering algorithms K-means

-self-organizing maps (SOMs)

3.1 The hierarchical clustering

The hierarchical clustering is the most widely used clustering method (Sokal and Mitchener, 1958; Eisen and Brown, 1999). This agglomerative technique arranges genes according to their similarity in expression profiles across all of the array experiments such that genes with similar expression are clustered together. Technically a pair-wise distance matrix is calculated for all genes that need to be clustered, and then the closest genes are put together to form a cluster. Finally

clusters are compared between them and merged according to the similarity. Typically the Pearson correlation coefficient or Euclidean distances are used to quantify the similarity between genes and clusters of genes. The graphical visualization of the hierarchical algorithm is illustrated in two-dimensional space by dendograms where each merger is represented by a binary tree, and the length of each branch is indicative of the distance between two samples. There are several hierarchical clustering algorithms differing by the used to measure the distance between clusters as they are constructed (Sokal and Sneath, 1973). Mainly, there is first the simple-linkage clustering (called also the minimum, or nearest-neighbour method) which represents the smallest distance between a gene of the first cluster and a gene of the second cluster when two clusters are compared. Secondly, the complete-linkage clustering (called also the maximum or furthest-neighbour method) which is the opposite of simple-linkage clustering, calculates the distance as the greatest between members of two clusters. Finally, the average-linkage clustering is the most widely used approach for the hierarchical clustering and corresponds to the average distance between clusters.

3.2 The K-means clustering

In contrary to the hierarchical method, the k-means clustering (or Lloyd's algorithm) is a divisive approach to partition the genes or experiments into groups that have similar expression patterns (Quackenbush, 2001). This method can be a good alternative to the hierarchical clustering if the approximate number of clusters

is known (Tavazoie *et al.*, 1999). Indeed the genes are partitioned into a fixed number (k) of clusters and an average expression vector is calculated for each cluster and is used to compute the distances between the clusters. Then the genes will be allowed to move to one cluster to the other if they have more affinity with the second one. This maximizes the expression similarity of the genes in each cluster and this is repeated until no genes move. Basically the main purposes of the k-mean clustering are to minimise variability within clusters, to maximize variability between them and to lead ultimately to non-overlapping clusters. Recently a more flexible method called fuzzy k-means clustering has been developed and allows the genes to belong to more than one group. Indeed individual genes can be regulated by multiple condition-specific systems resulting in the co-expression of genes under different situations. Therefore genes should be allowed to be present in more than one cluster at a time to reveal more complex correlations between gene-expression patterns. This method has already shown to be efficient in yeast for the identification of putative transcription factor binding sites in the promoters of the genes (Gasch and Eisen, 2002).

3.3 The self organizing maps (SOMs)

The SOM algorithm organizes the data into a multi-dimensional matrix by an iterative process based on the relative similarity of the expression pattern of genes. This method is a divisive method close to the k-means clustering except that the expression vectors are compared to reference vectors defined for each partition

(Kohonen, 1995). For that purpose, the genes are represented as the points in the multi-dimensional space and their coordinates corresponds to the expression levels at various time points (Toronen *et al.*, 1999). The expression vectors are first initialized randomly then iteratively adjusted and the process continues for 20,000 to 50,000 iterations in which each iteration will move the nodes of the expression vectors into the direction of favorized points (Tamayo *et al.*, 1999). This unsupervised neural network is supposed to be a more robust and accurate approach for the clustering of large amounts of noisy data (Herrero *et al.*, 2001).

In conclusion, the K-means clustering produces tighter clusters than hierarchical clustering. But hierarchical clustering can produce a great number of small clusters, which is more convenient for discovery. Unlike K-means, hierarchical also produces an ordering of the objects interesting for the data display. SOM clustering is less used as it is usually recommended to check the results against K-means clustering (Tibshirani *et al.*, 1999).

Appendix C. Highly expressed genes (with internal log_2ratio > 2)

G: MM+glucose

M: MM+maltose

Xose: MM+xylose

Xan: MM+xylan

Wg: MM+wheat germ

CM: complete medium

internal log_2ratio > 3

internal log_2ratio > 2

internal log_2ratio < 2

Name	Description	G	M	Xose	Xan	Wg	CM
an_2467	Glucoamylase			1.25	0.39		2.65
an_933	Translation elongation factor 1 (tef gene)			2.87	2.63	2.59	2.12
an_595	Similar to serine/threonine kinase						
an_2965	H(+)-ATPase, plasma membrane				2.25	0.99	1.34
an_3409	Ubiquitin UB1				2.64		2.66
an_3118	Probable membrane protein		2.46	1.92	1.64	1.33	2.46
an_3281	F-actin capping protein beta-subunit	2.99		2.89	2.12	1.98	2.74
ANI9972	Pyruvate carboxylase	2.99	2.62	1.16	1.15	1.4	
an_601	Translational elongation factor 1-alpha	2.98			2.73	2.74	2.51
an_3300	Translational inhibitor protein p14.5	2.91					
an_3665	26S proteasome regulatory particle chain RPN12	2.91	2.99		1.44	2.19	2.39
an_1995	Rub1 mRNA for polyubiquitin	2.9	2.54	1.26	1.64	1.06	1.45
an_2085	ADP,ATP carrier protein	2.89		2.13	1.28	1.07	1.35
an_1604	Probable lysophospholipase L2 protein	2.8	2.26	2.89	2.66		1.61
an_3034	Antifungal protein precursor	2.78	2.66	2.05	2.92	1.41	2.81
an_1593	Translational elongation factor 1-alpha	2.73	2.84	2.83	2.05	2.16	1.84
an_907	Farnesyltransferase beta subunit	2.66	2.57	2.78	2.73	2.51	1.32
an_3248	Phosducin-like protein, trimeric G protein modulator	2.65			1.7	1.86	1.31
an_2369	ADP,ATP carrier protein	2.63	2.89	2.82	1.44	1.42	1.21
an_2904	Glyceraldehyde-3-phosphate dehydrogenase	2.52	2.12	1.62	1.93	0.41	1.06
an_3432	Hypothetical protein	2.5	2.28	1.91	2.78	2.79	2.15
an_2360	40s ribosomal protein s3ae (S1)	2.48	2.52	2.78	1.25	1.3	0.97
an_3185	Aldehyde dehydrogenase (NAD+)	2.46	2.4	1.53		1.75	2.57
an_1318	Cobalamin-independent methionine synthase (metH/D)	2.44	2.71	2.39	2.01	2.08	1.8
an_1348	60S ribosomal protein L32	2.43	2.58	2.18	2.57	2.32	2
an_3377	Maltose O-acetyltransferase	2.41	2.46		1.85	1.65	1.27
an_3285	Carbon catabolite repressor,zinc finger	2.39	2.17	1.72	1.11	1.92	2.42
an_3172	60S ribosomal protein L6-B	2.34	2.7	2.49	2.28	2.67	1.97
an_2854	Elongation factor 1 alpha	2.33	2.5	2.08	1.22	1.16	1.51
an_577	Mitochondrial F(1)-ATPase beta-subunit precursor	2.32	2.76		1.49	1.62	1.33
an_2543	Aspartic protease (Aspergillopepsin A precursor)	2.32	2.16	1.93	1.4	2.28	1.6
an_800	Putative protein	2.32	2.18	1.42	2.06	2.02	2.34
an_607	Elongation factor 2	2.26	2.41	2.64	1.72	1.79	1.22
an_1587	Histone H3	2.25	2.23	2.23	1.85	2.15	2.07
an_571	Probable zinc-binding dehydrogenase	2.25	2.09	2.56	2.06	2.43	1.66
an_1933	Transcritption factor Btf3	2.24	2.24	1.63	0.21	1.3	1.98
an_957	Cytochrome c oxidase holoenzyme assembly or stability	2.18	2.02	1.31	1.87	1.9	1.49
an_1766	40s ribosomal protein S31	2.18	1.99	1.71	2.36	1.73	1.97

an_2510	60S ribosomal protein L43	2.17	2.44	2.25	2.29	2.91	1.52
an_1790	Histone H2A	2.16	2.43	2.71	2.04	2.29	2.13
an_3599	40S ribosomal protein S30	2.14	2.04	1.8	2.19	2.24	2.02
an_955	Probable phosphate transport protein, mitochondrial	2.14	2.27	1.85	2.32	2.24	2.44
an_1414	ATP adenosine-5'-phosphosulfate 3'-phosphotransferase	2.12	1.81	1.47	2.63	2.12	2.10
an_3504	Ubiquitin conjugating enzyme	2.12	2.23	1.55	1.92	2.32	2.29
an_1846	Cross-pathway control protein A - cpcA (leucine zipper)	2.11	1.31	1.37	1.27	1.9	2.13
an_3153	Putative dual specificity protein kinase Pom1p	2.11	1.92	2.12	1.71	1.7	2.07
an_1292	Polyprotein	2.1	1.63	1.55	2.24	2.16	2.19
an_668	Peroxisomal hydratase-dehydrogenase-epimerase	2.09	2.11	1.88	1.33	0.71	1.29
an_1621	Ribosomal protein L13	2.08	2.18	2.11	1.1	1.78	1.06
an_2554	40S ribosomal protein S13	2.07	1.97	1.54	1.73	1.88	1.68
an_1279	Putative transcriptional regulatory protein	2.06	1.5	1.3	1.94	1.99	1.87
an_1278	Cyclophilin-like peptidyl prolyl cis-transisomerase (cypA)	2.06	1.99	2.02	1.71	2.36	1.54
an_974	Putative carboxymethylenebutenolidase	2.05	1.56	1.63	2.48	2.31	2.03
an_1649	60S ribosomal protein L19B	2.05	2.02	1.52	1.73	1.29	1.78
an_3391	Hexokinase-like protein	2.04	2.25	2.32	1.72	2.78	2.22
an_1693	NADH-ubiquinone oxidoreductase	2.03	2.17	1.79	2.21	2.75	2.02
an_1795	Replication factor C like protein	2.03	1.83	1.67	2.28	2.27	2.15
an_1307	60s ribosomal protein l39	2.03	1.88	1.76	1.57	1.46	1.37
an_2687	40S ribosomal protein S8	2.02	2.3	2.74	1.18	2.28	1.36
an_3275	Cofilin	2.02	1.88	2.5	1.08	1.46	1.51
an_2409	60S ribosomal protein L28	2.02	2.6	2.62	1.74	2.72	1.97
an_1641	Putative carboxy-phosphoenolpyruvate mutase	2.02	2.09	2.2	2.6		2.44
an_3477	Plasma membrane H+-ATPase	2.02	2.53	2.34	1.47	0.66	1.2
an_3074	60S ribosomal protein L10	2	1.91	1.64	1.55	1.66	1.37
an_1791	histone H2A	2	1.85	1.95	1.58	2.04	1.87
an_2524	60S acidic ribosomal protein P2	1.99	2.19	2.74	1.6	2.24	0.87
an_2380	ATP synthase subunit 4	1.99	2.16	1.76	2.1	1.83	1.44
an_3170	Histone H3	1.99	2.37	1.32	1.15	2.33	1.6
an_1535	40s ribosomal protein S2	1.98	1.72	1.59	1.78	2.43	1.79
an_3642	Elongation factor 1-alpha	1.98	2.25	1.95	1.66	1.11	0.92
an_587	Plasma membrane H+-ATPase (PMA1)	1.97	2.68	2.55	1.94	1.54	1.27
an_3412	Pepstatin insensitive protease, lysosome	1.97	2.33	2.39	1.26	1.42	0.69
an_694	60S ribosomal protein L18	1.94	2.39	1.87	1.56	1.49	1.53
an_3508	Mold-specific MS8 protein (hypae formation)	1.93	2.21	2.05	1.44	1.89	2.28
an_3506	Multifunctional beta oxidation protein, peroxisome	1.93	1.49		2.34	2.42	
an_1598	Glutamine synthetase (glnA)	1.93	2.01	2.29	1.48	1.56	1.44
an_1892	Probable phosphate transport protein MIR1	1.92	1.84	1.82	1.99	2.16	1.57
an_1214	Ankyrin repeat gene family protein	1.91	1.73	1.34	2.15	2.18	2.02
an_1674	Conserved hypothetical protein; PHD finger	1.91	1.5	1.51	1.97	2.24	1.76
an_1383	Phenol hydroxylase related protein	1.91	1.96	2.12	1.38	2.55	1.21
an_1464	Hypothetical protein	1.9	1.26	1.24	1.89		1.85
an_1377	Wee1-like protein	1.9	2.17	2.48	1.92	2.35	1.8
an_1584	Succinate dehydrogenase (ubiquinone)	1.89	2.03	1.26	1.5	2.44	1.87
an_3175	60S ribosomal protein L17	1.88	2.17	1.68	2.08	2.64	1.54
an_3133	40S ribosomal protein S5	1.87	2.07	0.82	1.32	1.5	0.8
an_1319	Antigenic cell wall galactomannoprotein MP	1.86	1.51	1.46	1.86	2.12	2.46
an_3174	Hypothetical protein	1.86	1.97	1.66	1.72	2.09	1.59
an_1819	14-3-3-like protein (regulator)	1.85	1.83	1.34	1.85	2.15	1.71
an_1309	Essential for assembly of a functional F1-ATPase	1.84	1.34	1.21	2.05	2.48	1.96
an_1408	3-oxoacyl-(acyl carrier protein) reductase	1.84	1.43	1.68	2.11	2.85	1.9
an_1699	Mucin	1.83	1.5	1.36	2.34	2.18	2.05
an_2899	Similar to keratin associated protein 9.4	1.83	1.77	1.82	1.93	2.32	2.35

an_1418	40s ribosomal protein S25	1.82	1.69	1.8	2.56	1.95	1.83
an_2659	6-phosphogluconate dehydrogenase	1.82	1.91	2.55	1.34	2.26	2.07
an_3437	Thioredoxin, mitochondrial precursor	1.81	2.24	2.18	1.46	2.23	1.87
an_2956	60S ribosomal protein L31	1.81	2.13	1.56	1.99	2.09	1.24
an_3099	60S ribosomal protein L33	1.8	1.43	0.07	2.49	1.05	1.62
an_2949	Putative protein	1.8	2.43	2.46	2.23	2.95	1.98
an_786	HEX1 (hex1) protein	1.8	2.13	1.18	1.08	0.14	0.63
an_2470	E6-associated protein (related to ubiquitination)	1.79	2.21	2.06	1.58	2.54	1.31
an_1558	Hypothetical protein	1.78	1.56	1.5	2.26	2.51	1.57
ANAJ1909	Xylanotic transcriptional factor (xlnR)	1.78	1.92	1.9	2.16	1.74	1.92
an_2193	40S ribosomal protein S18	1.78	1.89	2.16	1.43	1.44	1.11
an_2425	Putative ribosomal protein L35	1.77	2.29	2.52	1.76	2.02	1.15
an_2748	Required for degradation of cell cycle control proteins	1.77	1.2	1.55	2.22	2.89	2.16
an_1280	26S Proteasome regulatory subunit	1.76	1.55	1.62	2.01	2.54	1.92
an_1446	Mitochondrial ribosomal protein MRP17	1.76	1.31	1.32	1.94	2.56	1.67
an_2751	Histone H4	1.76	1.76	1.77	1.89	1.99	1.55
an_3407	Oligopeptide ABC transporter (permease)	1.75	2.07	2.17	2.03	2.91	1.27
an_1286	40S ribosomal protein S27	1.75	1.6	1.39	2.05	1.57	1.88
an_3405	RNA-binding protein	1.74	2.42	1.7	1.9	2.56	1.32
an_682	Related to the component Tra1 of the SAGA complex	1.74	1.69	1.37	1.86	2.48	2.08
an_2497	60S ribosomal protein L36	1.74	1.76	2.13	1.03	1.86	0.88
an_2434	Phenol-2-monooxygenase	1.73	1.32	1.46	2.15	2.91	1.8
an_1477	Electron transfer protein	1.73	0.96	1.51	1.44	1.86	2.12
an_1659	Y20 protein	1.71	1.19	1.16	1.81	2.65	1.42
an_1930	Acidic ribosomal protein P2	1.71	1.97	1.73	1.92	2.23	1.6
an_3214	Ribosomal protein L18	1.7	2.19	1.93	2.11	2.43	1.53
an_3001	60S ribosomal protein L44	1.7	1.61	0.65	1.61	2.19	1.14
an_1814	Similarity to phosphomannomutases	1.69	1.53	2.18	1.6	1.89	1.18
an_1652	1,4-benzoquinone reductase	1.68	1.33	1.55	2.12	2.36	1.69
an_1015	Voltage-dependent Na+ channel beta-1 subunit	1.68	0.83	-0.3	2.86	0.9	2.03
an_2884	60S ribosomal protein L5	1.68	2.09	1.49	2.22	2.07	1.97
an_1870	Translation initiation factor 3 (eIF3)	1.66	1.32	1.01	1.34	1.57	2.18
an_2513	ATP synthase 9	1.66	2.54	1.08	1.27	-0.8	-0.49
an_1970	1,3-beta-D-glucan synthase catalytic subunit	1.65	2.19	1.61	1.13	1.01	0.81
an_2723	Initiation factor 3	1.65	2.09	1.75	1.96	2.23	2.24
an_666	Initiation factor 5A	1.65	2.01	2.7	1.8	1.91	2.04
an_2569	Initiation factor 3	1.64	1.95	1.54	2.19	2.53	2.08
an_1753	Chitin synthase	1.62	1.37	1.44	2.33	2.18	1.72
an_2799	Prefoldin subunit	1.59	1.46	1.21	1.64	2.39	1.36
an_2132	RepA2 replication protein	1.59	1.48	1.6	1.9	2.41	1.44
an_2020	40S ribosomal protein S4	1.58	2.26	1.98	0.54	0.79	0.12
an_1869	Methylmalonate-semialdehyde dehydrogenase	1.57	1.1	1.24	1.99	2.78	1.94
an_1605	Related to ATP synthase epsilon chain (mitochondrial)	1.56	1.59	1.33	1.36	2.03	1.68
an_1934	UL37 immediate-early glycoprotein	1.55	1.62	1.44	1.92	2.04	2.66
an_771	60S ribosomal protein L27A	1.54	1.5	2.11	1.06	1.21	0.74
an_897	Similar to serine/threonine kinase	1.54	1.67	1	2.57	1.83	2.74
an_1367	Related to 6-phosphofructo-2-kinase	1.54	1.56	1.78	1.19	2.32	1.36
an_2749	40S ribosomal protein S24	1.52	1.32	2.19	1.37	0.87	0.38
an_1916	Similar to serine/threonine kinase	1.51	0.93	0.46	2.38	1.57	2.45
an_3277	Metallothionein (proliferation)	1.51	1.25	0.96	2.33	1.98	2.25
an_1432	Probable autophagy protein AUT7	1.5	0.97	1.24	2.26	1.57	1.67
an_2152	40S ribosomal protein S9 (S7)	1.48	2.31	0.6	2.21	2.41	1.69
an_960	Hemolysin	1.47	1.29	1.24	1.67	1.41	2.61
an_597	Serine rich protein	1.47	1.81	2.38	1.24	1.3	1.5

an_687	Probable integral membrane protein	1.47	0.93	1.02	2.36	1.8	1.77
an_1237	Hypothetical protein	1.47	0.97	0.69	2.22	1.5	1.25
an_3200	Ubiquitin carboxyl-terminal hydrolase	1.42	0.66	0.84	1.69	2.61	1.75
an_2787	Ribonucleoprotein, U6 snRNP	1.41	1.34	1.06	2.22	1.99	1.61
an_1490	Related to signal sequence receptor alpha chain	1.41	0.73	0.51	1.77	2.08	1.24
an_3122	NADP-dependent glutamate dehydrogenase	1.41	1.98	1.69	1.39	0.84	0.17
an_2715	Hypothetical protein	1.41	2.08	1.81	1.4	1.97	1.21
an_2644	Acyl carrier protein, mitochondrial precursor	1.41	1.97	1.55	1.4	1.94	1.06
an_3216	Ribosomal protein S14	1.38	2.07	1.54	1.62	1.68	1.23
an_1775	Putative Oligosaccharyltransferase gamma subunit	1.38	1.42	1.16	1.59	1.93	1.44
an_1991	Conserved hypothetical protein	1.37	1.47	0.84	2.12	2.62	1.34
an_2121	SCF complex protein cul-1 homolog	1.37	1.34	1.42	1.87	2.53	1.64
an_1787	Polyketide synthase	1.36	0.64	0.92	2.05	1.41	1.4
an_2077	60S ribosomal protein L29	1.36	1.33	1.16	1.41	2.16	1.32
an_797	Nonhistone protein 6	1.35	1.07	1.98	1.17	2.05	1.12
an_787	40S ribosomal protein S11	1.35	1.82	2.31	0.7	1.19	-0.01
an_3451	Glutaredoxin	1.35	1.18	2.05	1.4	2.34	1.38
an_1619	Translation initiation factor 3	1.33	1.6	1.67	0.63	2.48	1.27
an_3414	G-beta like protein	1.33	2.19	1.83	1.4	1.07	0.53
an_895	Arginase	1.33	1.96	1.52	1.82	2.27	1.79
an_2149	Endosomal protein	1.32	1.13	1.46	1.87	2.49	1.57
an_1782	Conserved hypothetical protein	1.32	0.93	1.13	0.88	2.29	1.21
an_1423	40S ribosomal protein S21	1.28	2.21	1.82	1.39	1.83	1.47
an_1774	Glycinamide ribonucleotide transformylase	1.27	0.92	1.37	1.67	2.16	0.89
an_1889	Calcium P-type ATPase	1.27	1.7	1.43	2.16	1.85	2.13
an_3610	NADH ubiquinone oxidoreductase	1.25	1.76	2.03	1.21	0.11	0.77
an_1557	Conserved hypothetical protein	1.19	1.16	1.44	1.05	2.33	1.1
an_2722	60S ribosomal protein L21	1.17	1.34	2.03	0.81	0.76	-0.14
an_1657	60S ribosomal protein L15	1.09	1.3	1.23	2.16	1.23	2.8
an_2313	Peroxisomal organisation and biogenesis protein	1.05	1.24	1	2.22	1.74	1.5
an_697	CipC protein (cellulosome scaffoldin)	1.05	1.28	1.88	0.94	1.28	1.86
an_3289	Mutanase-alpha-1,3-glucanase	1.02	0.92	1.36	0.4	2.21	0.85
an_3469	60S ribosomal protein L19	1	1.58	2.07	0.83	1.06	0.97
an_1008	Hypothetical protein	0.92	1.24	2.04	0.55	0.69	-0.03
an_890	Myo-inositol transport protein ITR1	0.87	0.38	0.6	1.4	2.41	1.45
an_3556	Cytochrome P450 monooxygenase	0.83	0.44	1.32	1.51	2.05	2.44
AY017409	Oligosaccharyl transferase alpha subunit	0.76	0.33	2.18	-1.19	-1	-0.34
A27977	ExoB-arabinan-degrading enzymes	0.66	0.54	-0.3	1.16	2.12	1.07
an_1029	Histone H3	0.47	0.28	-0.5	[redacted]	0.75	1.38
an_2747	Carboxyphosphonoenolpyruvate mutase	0.41	0.23	0.39	2.42	-0	0.71