**THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**

# "Systems Biology of Glucose Sensing and Repression in *Aspergillus niger*"

**Margarita Salazar Peña**

Systems Biology

Department of Chemical and Biological Engineering

**CHALMERS UNIVERSITY OF TECHNOLOGY**

Göteborg, Sweden 2010

**LIST OF PUBLICATIONS**

This thesis recapitulates the work contained in the following publications:

1) Andersen MR, **Salazar M**, et al. (2010). Sequencing of *A. niger* ATCC 1015 and comparison to *A. niger* CBS 513.88. *Submitted to Genome Res.*

2) **Salazar M***, Vongsangnak W*, Panagiotou G, Andersen MR, Nielsen J. (2009) Uncovering transcriptional regulation of glycerol metabolism in Aspergilli through genome-wide gene expression data analysis. Mol Genet Genomics. 282(6): 571-586. (* equal contribution).

3) Vongsangnak W*, **Salazar M***, Hansen K, Nielsen J. (2009). Genome-wide analysis of maltose utilization and regulation in aspergilli. Microbiology. 155: 3893-3902. (* equal contribution).

4) **Salazar M**, Calzada-Funes J, Bruno K.S., Culley D.E., Baker S.E., Nielsen J. (2010). Transcriptome profiling of *Aspergillus niger* AdrA, FacB and CreA mutant genotypes during growth on glucose or glycerol as carbon sources. *Submitted to Microbiology*.

5) **Salazar M**, Bruno K.S., Culley D.E., Baker S.E., Nielsen J. (2010). Deletion of a fungal regulatory gene of the GATA family: *areB* in *Aspergillus niger*. *Submitted to Fungal Genet. Biol.*

My contribution to each of these publications was:

***Publication 1***:

MS carried out the batch fermentations, enzymatic assays and microarray experiments with two different sequenced strains of *A. niger*, ATCC1015 and CBS 513.88 using glucose as carbon source, supported on the microarray data analysis and assisted in manuscript preparation.

***Publication 2***:

MS conducted the batch fermentations and microarray experiments with *A. niger* strain BO1 on two different carbon sources, glucose and glycerol, performed microarray data analysis and wrote the manuscript.

***Publication 3***:

MS carried out the batch fermentations and microarray experiments with *A. niger* strain BO1 on two different carbon sources, maltose and glucose, assisted on analysis of the data and preparation of manuscript.

***Publication 4***:

MS constructed the *A. niger* transcription factor deletion strains, conducted the batch fermentations and microarray experiments with the three different genetically modified strains on two different carbon sources, glucose and glycerol, performed microarray data analysis and wrote the manuscript.

***Publication 5***:

MS constructed the *A. niger areB* deletion strain, carried out the physiological characterization of the deletion mutant with different nitrogen sources, conducted batch fermentations and microarray experiments using two different carbon sources, glycerol and glucose, performed microarray data analysis and wrote the manuscript.

**OTHER PUBLICATIONS RELATED TO MY PROJECT**

1) Andersen MR, Vongsangnak W, Panagiotou G, **Salazar M**, Lehmann L, Nielsen J. (2008) A tri-species *Aspergillus* micro array – advancing comparative transcriptomics. Proc Natl Acad Sci. 105:4387-4392.

2) Papini M*, **Salazar M*,** Nielsen J. (2010). Industrial systems biology of yeast and fungi. "Biosystems Engineering". Advances in Biochemical Engineering and Biotechnology. (* equal contribution). Feb 12. [Epub ahead of print]. *In press*.

3) Vongsangnak W, Nookaew I, **Salazar M**, Nielsen J. (2010). Analysis of Genome-Wide Coexpression and Coevolution of *Aspergillus oryzae* and *Aspergillus niger*. OMICS: A J Integrative Biology. 14(2): 165-175.

My contribution to these manuscripts was:

***Publication 1***:

MS conducted the batch fermentations and microarray experiments with the *A. niger* strain BO1 using xylose as sole carbon source and energy and assisted in manuscript preparation.

***Publication 2***:

This book chapter focused on Industrial systems biology of yeast and fungi. It was divided in six main parts where I contributed to three of them, mainly to the sections 3, 5 and 6. Section 3 explained the Systems Biology Toolbox that included: Genomics, Transcriptomics, Proteomics, Metabolomics, Fluxomics and Metabolic Modeling. Section 5 was devoted to fungal cell factories and included a throughout description of the characteristics and applications of *Aspergillus* species as cell factories, strain improvement of the industrial production hosts *Aspergillus oryzae* and *Aspergillus niger*, *Aspergillus* species as platform for production of chemicals and heterologous protein production. Finally, section 6 elaborated on systems Biology as a driver for Industrial Biotechnology where I focused on three case

studies where *Aspergillus* species were exploited. I also assisted in reviewing and formatting the full manuscript.

***Publication 3***:

MS assisted on manuscript preparation, data analysis for answering reviewer's comments and resubmission of manuscript.

**OTHER PUBLICATIONS NOT RELATED TO MY PROJECT**

1) Mouse paper: Collaboration with Food Science.

MS performed cDNA synthesis of 18 samples and its preparation for microarray hybridization on GeneChip® Mouse Gene 1.0 ST Array.

2) Human paper: Collaboration with Gothenburg University.

MS performed cDNA synthesis of 36 samples and further processing for microarray hybridization on GeneChip® Human Gene 1.0 ST Array.

**ABSTRACT**

## 1. Introduction

### 1.1 Characteristics and applications of *Aspergillus niger* as a cell factory

*Aspergillus* comprises a genus of multicellular eukaryotic microorganisms containing more than 185 different species described on a taxonomically basis, i.e. *Aspergillus nidulans*, *A. fumigatus*, *A. flavus*, *A. oryzae*, *A. awamori*, *A. terreus*, etc (http://www.aspergillus.man.ac.uk) (YU *et al.* 2005). These ascomycetes are of great economical and medical importance as some of them are used for metabolite production i.e. *A. terreus* which produces lovastatin (ASKENAZI *et al.* 2003), and others are pathogens such as *A. fumigatus*. Among them, *A. niger* is one of the preferred hosts for chemical production. Fungi have been used in food production since ancient times, and today the diversified metabolism of these organisms is exploited for the production of pharmaceuticals, flavours, enzymes, and bulk and fine chemicals (HOFMANN *et al.* 2003).

*Aspergillus niger* is an excellent producer of many metabolites that are valuable commodities, e.g. citric acid, itaconic acid and gluconic acid. Apart from metabolites, *Aspergillus niger* is able to secrete a wide range of enzymes, representing a rich source for the production of enzymes that can be used in detergents, food processing and material handling (see Table 1).

For example, the market for industrial enzymes is estimated to be worth nearly US$ 5 billion by 2009, of which filamentous fungi account for roughly half of the production (LUBERTOZZI and KEASLING 2009) and in 2008, the global citric acid production reached 1.4 million tonnes, increasing annually at 3.5 to 4.0% in demand and consumption (ANASTASSIADIS *et al.* 2008).

Of particular interest is that *A. niger*, like other *Aspergillus* spp., has the capability to degrade a wide range of natural organic substrates including plant materials (BAKER and BENNETT 2008).

*A. niger* is one of the most important species of the genus *Aspergillus* which possesses asexual reproduction producing black conidia. An important feature of *A. niger*, is the GRAS (Generally Recognized as Safe) status already granted by the Food and Drug Administration of the US government (PERRONE *et al.* 2007). This is a highly desired classification in the food industry and moreover, as suggested by Baker and Bennett (BAKER and BENNETT 2008), because of the long history in the use of *A. niger*, the process for production of new products from this fungus can therefore be easily approved.

**Table 1**. Commercial enzyme preparations produced by *Aspergillus niger* for food, feed and technology processing, i.e., detergents industry [adapted from AMFEP (Association of Manufacturers and Formulators of Enzyme Products, Brussels) (AMFEP, 2007; Varga et al., 2008)].

| Enzyme | Host organism | Donor organism | IUB number | Application | | |
|---|---|---|---|---|---|---|
| | | | | Food | Feed | Technology |
| Aminopeptidase | *Aspergillus niger* | None | 3.4.11.x | Y | N | N |
| Amylase (alpha) | *Aspergillus niger* | None | 3.2.1.1 | Y | N | N |
| Arabinanase | *Aspergillus niger* | None | 3.2.1.99 | Y | Y | N |
| Arabinofuranosidase | *Aspergillus niger* | *Aspergillus* sp. | 3.2.1.55 | Y | N | N |
| Asparaginase | *Aspergillus niger* | *Aspergillus* sp. | 3.5.1.1 | Y | N | N |
| Carboxypeptidase (serine-type) | *Aspergillus niger* | *Aspergillus* sp. | 3.4.16.x | Y | N | N |
| Catalase | *Aspergillus niger* | None | 1.11.1.6 | Y | N | Y |
| Cellulase | *Aspergillus niger* | None | 3.2.1.4 | Y | Y | N |
| Galactosidase (alpha) | *Aspergillus niger* | None | 3.2.1.22 | Y | Y | N |
| Glucanase (beta) | *Aspergillus niger* | none | 3.2.1.6 | Y | Y | N |
| Glucoamylase or Amyloglucosidase | *Aspergillus niger* | none | 3.2.1.3 | Y | N | Y |
| Glucose oxidase | *Aspergillus niger* | none | 1.1.3.4 | Y | N | Y |
| Glucosidase (alpha) | *Aspergillus niger* | none | 3.2.1.20 | Y | N | N |
| Hemicellulase | *Aspergillus niger* | none | - | Y | Y | N |
| Inulase | *Aspergillus niger* | none | 3.2.1.7 | Y | N | N |
| Lipase triacylglycerol | *Aspergillus niger* | none | 3.1.1.3 | Y | N | N |
| Mannanase (endo-1.4-beta) | *Aspergillus niger* | none | 3.2.1.78 | Y | Y | N |
| Pectin lyase | *Aspergillus niger* | none | 4.2.2.10 | Y | Y | N |
| Pectin methylesterase or Pectinesterase | *Aspergillus niger* | *Aspergillus* sp. | 3.1.1.11 | Y | Y | N |
| Phosphatase | *Aspergillus niger* | none | 3.1.3.2 | Y | N | N |
| Phospholipase A | *Aspergillus niger* | *Aspergillus* sp. | 3.1.1.4 | Y | N | N |
| Phospholipase B | *Aspergillus niger* | none | 3.1.1.5 | Y | N | N |
| Phytase | *Aspergillus niger* | none | 3.1.3.8 | Y | N | N |
| Polygalacturonase or Pectinase | *Aspergillus niger* | none | 3.2.1.15 | Y | Y | Y |
| Protease (incl. milkclotting enzymes) | *Aspergillus niger* | none | 3.4.2x.x | Y | Y | N |
| Tannase | *Aspergillus niger* | none | 3.1.1.20 | Y | Y | N |
| Transglucosidase | *Aspergillus niger* | none | 2.4.1.24 | N | N | Y |
| Xylanase | *Aspergillus niger* | none | 3.2.1.8 | Y | Y | N |

Those enzymes which do not have an IUB (International Union of Biochemistry) number (IUPAC-IUB, 1971) are enzyme complexes, where the listed activity is the result of the sum of many single active enzyme proteins. There is no general IUB number for aminopeptidases. Nevertheless, all these enzymes fall under the 3.4.11.x category according to the IUPAC-IUB Commission on Biochemical Nomenclature (CBN).

Similarly, there is no general IUB number for proteases. But all these enzymes fall under the 3.4.2x category IUPAC-IUB Commission on Biochemical Nomenclature (CBN).

The main concern of using *Aspergillus* as a production host, as with many filamentous fungi, is the potential production of exometabolites with toxic properties. Consistently, the safety of *A. niger* as production organism has been widely documented (BARBESGAARD *et al.* 1992; SCHUSTER *et al.* 2002) and metabolites profiles produced by several isolates of this strain used in enzyme production have been published (BLUMENTHAL 2004; VAN DIJCK *et al.* 2003).
Even though *Aspergillus* is a genus widely exploited for production of organic acids and enzymes, little is known about i.e. the regulation of its metabolism; therefore *Aspergillus* research is a fertile ground for quantitative and modelling studies of complex cellular processes such as signalling and metabolic networks.

### 1.2 Strain improvement of the industrial production host *Aspergillus niger*

So far *Aspergillus* cell factories have been mainly exploited for the production of endogenous metabolites and enzymes, nevertheless, they are also attractive for non endogenous metabolite production. For this and other reasons, the fungal community is developing robust technologies for genetic manipulation. Since the genome sequencing of the genetic model organism *A. nidulans* (GALAGAN *et al.* 2005), the pathogens *A. fumigatus* (NIERMAN *et al.* 2005) and *A. flavus* (YU *et al.* 2005), the production host *A. oryzae* (MACHIDA *et al.* 2005), and later *A. niger* (PEL *et al.* 2007), and more which will be published soon (Scott Baker, personal communication), the genetics and genomics fields have had a tremendous development.
While classical strain improvement has been mainly done through chemical mutagenesis or natural mutagenesis caused by selective pressure, it is now feasible to conduct direct gene targeting at a specific locus. However, targeted gene modification is usually hampered by the low transformation efficiencies naturally achieved in filamentous fungi (MEYER 2008). Furthermore, the lack of a sexual cycle in *A. niger* has hindered the use of genetics for strain improvement (MACCABE *et al.* 1998). Targeting and replacement of gene loci in filamentous fungi, as in other organisms, are supported by the cellular machinery that accomplishes recombination and DNA repair (BIRD and BRADSHAW 1997). Especially the rate of homologous recombination (HR) in a given host determines in the efficiencies in knockout approaches using genetic markers that are flanked by homologous sequences of the gene locus

to be replaced. In filamentous fungi, these homologous sequences are generally larger (several hundreds of base pairs) compared to other organisms (KRAPPMANN *et al.* 2006; MEYER 2008), i.e., bacteria and yeast where a minimal length of 30 to 50 bp is sufficient to ensure a high yield of HR (HUA *et al.* 1997). There are two main mechanisms of DNA repair, by homologous recombination and by non-homologous end joining (NHEJ). In filamentous fungi repair of DNA damages seems to occur primarily by using the NHEJ machinery, and a DNA fragment that is desired to be integrated at a certain position in the genome is therefore often integrated ectopically impeding the achievement of the desired phenotype. Recently, disruption of some genes involved in the NHEJ pathway, namely *ku70* and *ku80*, enhanced the gene targeting efficiency in the yeast *Kluyveromyces lactis* (KOOISTRA *et al.* 2004) and in the filamentous fungus *Neurospora crassa* (NINOMIYA *et al.* 2004). This finding was immediately applied to industrially relevant filamentous fungi, for instance, *A. oryzae* and *A. sojae* (TAKAHASHI *et al.* 2006) and later to *A. niger* (MEYER *et al.* 2007). More recently, deletion of DNA ligase IV (LigD), another factor involved in the NHEJ pathway, resulted in a targeting efficiency as high as 100% in *N. crassa* (ISHIBASHI *et al.* 2006). It has already been applied to *A. oryzae* (MIZUTANI *et al.* 2008) and it will hopefully be transferred to *A. niger* soon. This technology will allow the engineering of these species by allowing the deletion of industrially undesirable traits to further improve productivity and safety. The development of recombinant DNA technologies has given the possibility of introducing targeted mutations (i.e. over-expressing genes or deletion of undesired ones) instead of the random generation and further screening for the desired phenotypes which is very costly and time consuming. One of the most used transformation methods for filamentous fungi is the protoplast mediated transformation (PMT) method, developed earlier for *S. cerevisiae* and adapted for filamentous fungi. Nevertheless, this method preferentially produces multicopy integration events (DE GROOT *et al.* 1998). Alternative methods for fungal transformation such as electroporation, biolistic transformation and *Agrobacterium* mediated transformation (AMT) have been developed (MICHIELSE *et al.* 2005; RUIZ-DIEZ 2002). These methods are valuable for fungal strains that do not form sufficient amount of protoplasts or for strains where their protoplasts do not regenerate. Overall, as discussed by Meyer (MEYER 2008) individual species have to be considered independently and the most appropriate method identified and optimized for each strain. A schematic representation of the chemical transformation process of *Aspergillus* is shown on Figure 1.

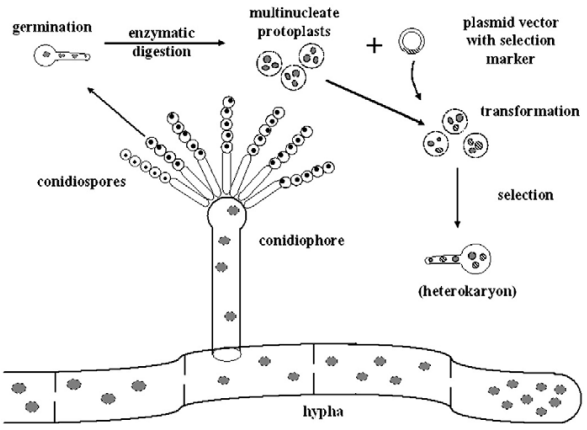



**Figure 1**. Chemical transformation of *Aspergillus*. Conidiospores are harvested and germinated in a nutrient broth, then subjected to an enzymatic treatment to lyse the cell wall, liberating protoplasts, which are incubated with transforming DNA in a medium containing $CaCl_2$ and polyethylene glycol (PEG). Plating on selective medium allows regrowth of transformants. Since the conidia are uninucleate in most *Aspergillus* species, a homokaryotic strain is readily obtained by reselection of transformants. Electroporation transformation is a similar process, using hydrated instead of fully germinated conidia, while biolistic and *Agrobacterium* methods make use of intact hyphae. Figure taken from (LUBERTOZZI and KEASLING 2009).

### 1.3 Comparative genomics of *Aspergillus niger* with other close Aspergilli

Genomics is the process of revealing the entire genetic contents of an organism by high throughput sequencing of the DNA and bioinformatics identification of all of the genes (YU *et al.* 2005). Since the publication in 1977 of the first complete genome sequence, that of the Bacteriophage fX174, a viral genome with only 5,368 base pairs (bps) (SANGER *et al.* 1977), the field of genomics has been of growing importance to biological studies. Thanks to the up to date sequencing technologies and powerful bioinformatics techniques, this field is growing rapidly. Genome sequencing of filamentous fungi is generally considered to have started as late as February 2001 (GALAGAN *et al.* 2003) with the publication of a draft version of the genome sequence of *Neurospora crassa*. Nearly three decades have passed since the invention of electrophoretic methods for DNA sequencing, often referred to as Sanger sequencing, and

its cost-effectiveness has mainly been driven down following the introduction of automation and the numerous refinements of this technology (SHENDURE *et al.* 2004).

The availability of genome sequence data facilitates research on basic biology, genetic regulation and evolution of close related species such as *Aspergillus* through comparative genomic studies. The completion of a ~8x coverage genome sequence of *A. niger* CBS 513.88, a strain used in industrial enzyme production, was published in 2007 (PEL *et al.* 2007) (see Table 2). This strain was derived from *A. niger* NRRL 3122, a classically improved strain selected for increased glucoamylase A production (VAN LANEN and SMITH 1968). Although, this *A. niger* genome sequencing initiated a number of new genome based investigations (CULLEN 2007; MARTENS-UZUNOVA and SCHAAP 2009; SUN *et al.* 2007; YUAN *et al.* 2008a; YUAN *et al.* 2008b), there remains a need for uncovering fundamental differences between strains used for different purposes, i.e., citric acid production and enzyme production (CULLEN 2007).

**Table 2**. General genome statistics for the *A. niger* citric acid producer strain ATCC 1015, the *A. niger* enzyme producer strain CBS 513.88 and for comparison with other *Aspergillus* spp., the *A. oryzae* wild type strain RIB40. Except genome sizes and the number of gene models, all values are averages.

| | ***A. niger* ATCC 1015**[a] | ***A. niger* CBS 513.88**[b] | ***A.nidulans FGSC A4***[c] | ***A. oryzae RIB40 ATCC 42149***[d] |
|---|---|---|---|---|
| Gene models | 11,200 | 14,165 | 9,541 | 14,063/12074[e] |
| Genome size (Mbp) | 34.85 | 33.93 | 30.07 | 37.05 |
| Gene length (bp) | 1,696.1 | 1,572.8 | 1,868 | 1,414 |
| Protein length (aa) | 484.3 | 439.9 | | 448.8 |
| Exons per gene | 3.1 | 3.6 | 3.6 | 2.9 |
| Exon length (bp) | 480.8 | 370.0 | | 542.8 |
| Intron length (bp) | 93.8 | 97.2 | | 231.4 |

[a] The genome assembly published by Andersen et al. (ANDERSEN *et al.* 2009b).

[b] The genome assembly published by Pel et al. (PEL *et al.* 2007).

[c] The genome assembly published by Galagan et al. (GALAGAN *et al.* 2005).

[d] The genome assembly published by Machida et al. (MACHIDA *et al.* 2005).

[e] The genomics of *A. oryzae* published by Kobayashi et al. (KOBAYASHI *et al.* 2007).

In new comparative genomic studies, gene profiling using microarrays provides a powerful tool to detect and profile whole sets of genes transcribed under specific conditions and several examples of this approach are being generated at a fast pace, e.g., a recent tri-species comparative transcriptomics study conducted by Andersen and coworkers. (ANDERSEN *et al.* 2008c).

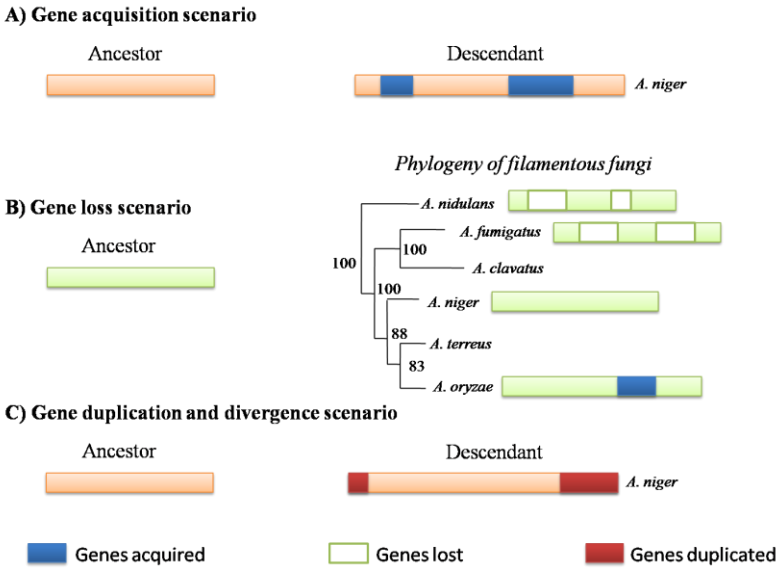



**Figure 2**. Mechanisms of gene and genome size expansion of *A. niger*. A) The common ancestor of the four *Aspergillus* species, namely *A. niger*, *A. oryzae*, *A. nidulans* and *A. fumigatus*, is assumed to have smaller genome size than *A. niger* and *A. oryzae* and similar to *A. nidulans* and *A. fumigatus*. *Aspergillus niger* might have acquired genes by horizontal gene transfer during evolution. B) The common ancestor is assumed to have the genome size similar to *A. niger*. The other species might have lost genetic material during evolution. The phylogenetic relationship of the species is taken from Pel et al. (PEL *et al.* 2007). Maximum likelihood tree based on concatenation of twenty orthologous proteins. Numbers at nodes are bootstrap values. C) The *A. niger* genome might have expanded by gene duplication followed by divergence of one of the duplicated genes as it might also have occurred to *A. oryzae*. The figure was adapted from ((MACHIDA *et al.* 2008).

Breakthroughs in *A. niger* genomics may lead to improvement of production processes and its wider exploitation as a cell factory. The availability of *Aspergillus* genomic data marks a new era in research for fungal biology. A comparison table of the genome statistics of the

acidogenic and the enzyme producer *A. niger* strains compared to its close relatives *A. nidulans* and *A. oryzae* is shown below (Table 2).

We can appreciate the differences in genome size which might be due to several reasons such as gene loss, gene acquisition, or gene duplication if we consider that all these three *Aspergillus* species diverted from a common ancestor (see Figure 2). According to Galagan et al. (GALAGAN *et al.* 2005), phylogenetic analysis of Aspergilli using the whole genome data showed that *A. nidulans* branched off earlier than *A. oryzae* and *A. fumigatus*.

### 1.4 The phenomenon of glucose repression in *Aspergillus* species

Glucose repression is the mechanism by which the presence of glucose represses transcription of genes involved in the utilization of other less favored carbon sources, genes encoding gluconeogenic and glyoxylate cycle enzymes and genes involved in secondary metabolism. In yeast, the main effect of glucose takes place at the transcriptional level (GANCEDO 1998). Nevertheless, for a subset of genes regulated by glucose, control is operating on mRNA stability instead of (or in addition to) on transcription rates. For example, in yeast, the *CYC1* mRNA half life was shown to decrease from 12 min in derepressed cells to about 2 min when glucose was present (ZITOMER *et al.* 1979). Another example is the case of *MAL62* mRNA half life, where the decrease was from 25 to 6 min (FEDEROFF *et al.* 1983). In *Aspergillus* species, glucose has been shown to repress expression of several genes. Table 3 shows a classification of carbon sources according to their level or repression over other pathways or genes. In *Aspergillus* species such as *A. nidulans*, a number of catabolic pathways are affected by glucose repression. Examples include utilization pathways for ethanol (FELENBOK and KELLY 1996), proline (CUBERO and SCAZZOCCHIO 1994), acetate (KELLY and HYNES 1977), arabinan (RUIJTER *et al.* 1997), xylan (DE VRIES *et al.* 1999; TAMAYO *et al.* 2008), pectin (DE VRIES *et al.* 2002a; SOLIS-PEREIRA *et al.* 1993) and cellulose (LOCKINGTON *et al.* 2002). Enzymes from central carbon metabolism, such as acetyl-CoA synthase (*facA*) and isocitrate dehydrogenase, and glyoxylate cycle enzymes, such as isocitrate lyase (*acuD*) and malate synthase (*acuE*) (SZEWCZYK *et al.* 2001), are also affected.

Perhaps the most studied model system is the ethanol utilization pathway, where glucose represses transcription of the specific regulatory gene, *alcR*, the alcohol dehydrogenase I encoding gene, *alcA*, the aldehyde dehydrogenase encoding gene, *aldA*, and the genes *alcM* and *alcX* (FELENBOK and KELLY 1996).

In *S. cerevisiae*, the main components of the glucose signaling and repression cascade have been elucidated since long time, being the transcription factor Mig1, one of the key players. In *Aspergillus* spp., the orthologue of Mig1 is known as CreA. CreA is a negatively acting regulator of carbon catabolite repression (BAILEY and ARST 1975). CreA in *Aspergillus niger* contains 427 amino acids. Alignment with the *Aspergillus nidulans* CreA showed 90% sequence similarity (82% identity) at the amino acid level. It contains two zinc-finger DNA-binding motifs of the $Cys_2$-$His_2$ class similar to the transcription factor Mig1 in *Saccharomyces cerevisiae* which recognizes a DNA sequence that is GC-rich, an alanine rich region consisting of 8 residues and an acidic region (DRYSDALE *et al.* 1993). The DNA binding motifs have a consensus hexanucleotide 5'-CPyGGG-3'. Several CreA mutants have been constructed in *A. nidulans*, e.g., *creA1*, *creA30*, *creA204*, *creA218*, *creA220*, *creA221* and *creA225* (SHROFF *et al.* 1996), and the mutations fall mainly into two main classes: missense mutations within the zinc-finger region, or frameshift or nonsense mutations somewhat after this region. Furthermore, transcriptome analysis has been conducted in an *A. nidulans* CreA mutant using a high-density oligo array with probes for 3,278 selected genes using the Febit Geniom® One array system (MOGENSEN *et al.* 2006b).

**Table 3**. Carbon sources classification. Adapted from (RUIJTER and VISSER 1997).

| Strong repressors | Intermediate repressors | Non or de-repressing C sources |
|---|---|---|
| **D-glucose, D-xylose, sucrose and acetate** | **D-mannose, maltose, D-fructose, D-mannitol and D-galactose** | **Glycerol, melibiose, lactose, L-arabinose and ethanol** |

The mechanism(s) regulating mRNA turnover in response to the carbon source remains to be worked out, but it is clearly established that CreA does not require the presence of exogenous glucose or any glucose-dependent activation process to be active (MATHIEU and FELENBOK 1994). For instance, CreA exerts a permanent repression of the *alc* genes under derepressed conditions in *A. nidulans*. Overall it seems that, although some of the general regulatory genes involved in carbon catabolite repression in yeasts are conserved in Aspergilli, the detailed mechanisms are somewhat different.

### 1.5 Carbon utilization in *Aspergillus* spp.

*Aspergillus niger* is a saprophytic fungus, mainly present in the soil, feeding preferably on organic matter such as plant cell wall polysaccharides (cellulose, hemi-cellulose and pectin),

and on plant storage polysaccharides (starch and inulin). *A. niger* is able to degrade a wide range of sugars such as the monosaccharides glucose, fructose and xylose, disaccharides such as sucrose and maltose, trisaccharides such as raffinose and maltotriose and a broad range of polysaccharides as it is shown on Figure 3.

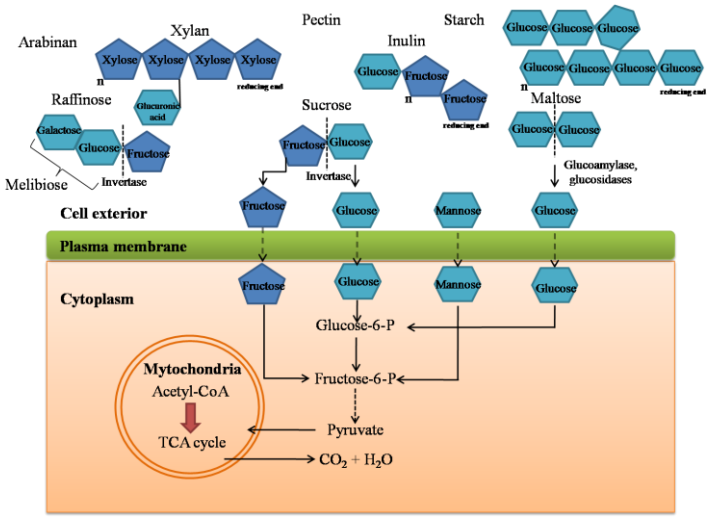



**Figure 3**. Assimilation of carbon sources in *Aspergillus niger*. Hexagons represent six carbon sugars and pentagons represent five carbon sugars.

There are several known transporters capable of transporting the monosaccharides such as glucose, xylose and mannose into the cell for subsequent phosphorylation and conversion into mainly biomass and $CO_2$. The characterized high-affinity transporters MstA and MstF, and the low affinity transporter MstC (JORGENSEN *et al.* 2007), as well as a number of putative transporters present in the genome of *A. niger* are responsible of the transport across the cell membrane. In contrast, polysaccharides are cleaved extracellularly by means of a broad range of extracellular enzymes such as amylases (CARLSEN and NIELSEN 2001a), glucoamylase (PEDERSEN *et al.* 2000b), galactosidases (SHANKAR and MULIMANI 2007), glucosidases (YUAN *et al.* 2008b), fructofuranosidases (FERNANDEZ *et al.* 2007), among others.

### 1.6 Glycolysis and gluconeogenesis

When a substrate is available as the sole source of carbon and energy, all the cellular components must be synthesized from this compound via appropriate metabolic pathways. Thus, this requires the organism to have the ability to rearrange the expression of gene-

encoding enzymes catalyzing the appropriate steps in the pathways according to the substrates available. The organism, therefore, induces enzymes specific for the breakdown of the particular compound. This section emphasizes on carbon metabolism and the utilization of carbon sources metabolized via the TCA cycle and requiring gluconeogenesis. *Aspergillus* species are mainly saprophytes growing in the soil on decaying plant material and capable of growing on an extremely diverse variety of carbon sources. *A. niger* is not an exception to this generalization and it is one of the Aspergilli most capable of growing on a wider range of carbon sources (DE VRIES 2008). Gluconeogenic substrates are ethanol, acetate and fatty acids, all of which result in the production of acetyl-CoA. Ethanol is converted to acetate by alcohol dehydrogenase and acetaldehyde dehydrogenase. Acetate generates acetyl-CoA by means of the enzyme acetyl-CoA synthetase (Acs1, EC 6.2.1.1). Likewise, fatty acids are converted to acetyl-CoA by β-oxidation in peroxisomes (HYNES *et al.* 2008). The glyoxylate cycle, comprising the enzymes, isocitrate lyase (ICL, EC 4.1.3.1) and malate synthase (MAS, EC 4.1.3.2), is necessary for the net conversion of acetyl-CoA via malate to oxaloacetate, which is then used in gluconeogenesis. An overview of the enzymes mentioned and its activity for degradation of the carbon sources described is presented on Figure 4. In *S. cerevisiae*, the genes for metabolism of acetyl-CoA by means of acetylCoA synthetase, the glyoxylate bypass and gluconeogenesis is controlled by the Zn(II)2Cys6 transcriptional regulators Cat8 and Sip4 (GANCEDO 1998). Furthermore, growth on ethanol or acetate is dependent on Cat8, Sip4 and Adr1 transcriptional activators and on the protein kinase Snf1 (sucrose non-fermenting 1) (GANCEDO 1998; YOUNG *et al.* 2003; YOUNG *et al.* 2002). Genes regulated by Cat8 or Sip4 contain one or more cis-acting elements termed carbon source responsive elements (CSRE) in their 5' regions where these regulatory proteins bind (ROTH *et al.* 2004). Snf1 in *S. cerevisiae* has diverse regulatory functions; however, in particular it is required for growth on fermentable carbon sources. Conserved orthologues are found in filamentous fungi, i.e. *A. nidulans* AN7695.3 (HYNES 2008), and by blastP comparison in *A. oryzae* and *A. niger*, AO090701000767 and JGI54719, respectively. Deletion of Snf1 orthologues in other fungi, such as *Fusarium oxysporum*, has been found to affect the utilization of some carbon sources (OSPINA-GIRALDO *et al.* 2003). A potential phosphorylation site was found in orthologues of CreA for some fungi (FLIPPHI and FELENBOK 2004). Nevertheless, there is no clear effect described on carbon source requiring gluconeogenesis. Consistent with the data from *Cochliobolus carbonum* (TONUKARI *et al.* 2000), mutations in the Snf1 orthologue in *F. oxysporum* diminished the transcription of

genes encoding cell wall-degrading enzymes and reduce its virulence on *Arabidopsis thaliana* and *Brassica oleracea* (OSPINA-GIRALDO *et al.* 2003).

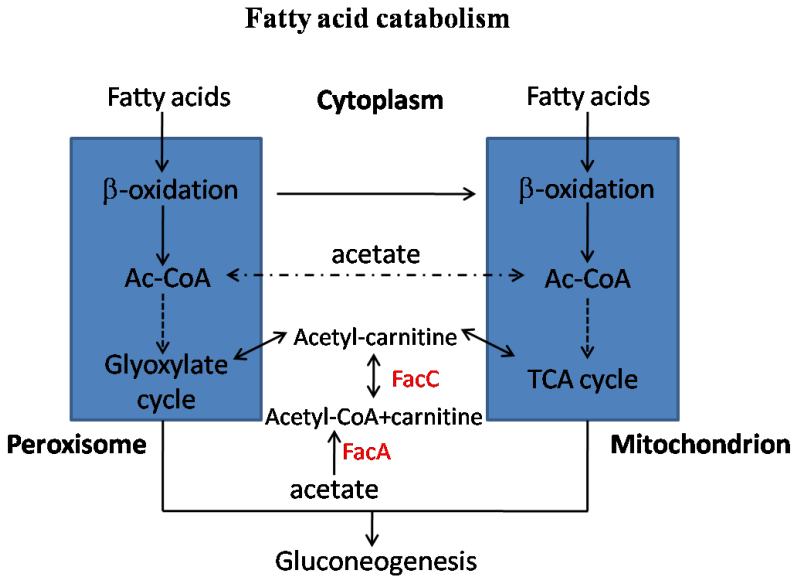



**Figure 4**. Fatty acids catabolism. β-oxidation of fatty acids takes place in the mitochondria and in peroxisomes. Fatty acids are metabolized in peroxisomes to yield acetyl-CoA which, via the glyoxylate cycle, yields malate for gluconeogenesis. Acetyl-CoA is also transferred to the mitochondria via acetyl-carnitine for metabolism via the TCA cycle. Short-chain fatty acids can directly enter the mitochondria for β-oxidation. Acetate is converted in the cytoplasm to acetyl-CoA and subsequently enters the mitochondria and the peroxisomes for metabolism via the glyoxylate cycle and the TCA cycle. Figure adapted from Hynes in (HYNES 2008).

The most important enzymes for gluconeogenesis are phosphoenolpyruvate carboxykinase (PEPCK, EC 4.1.1.32), which converts oxaloacetate to phosphoenolpyruvate and fructose-1,6-biphosphatase (FBP, EC 3.1.3.11), which hydrolizes fructose-1,6-biphosphate into fructose-6-phosphate and phosphate.

In contrast, in *A. niger*, FacB which is the orthologue of Cat8 has not been characterized and there are no reports of Sip4 orthologues. A comparison of the diverse transcriptional regulatory circuits controlling pathways for growth on gluconeogenic carbon sources of *S. cerevisisae* and *A. niger* is shown on Figure 5. Some of the genes involved on fatty acids catabolism, glyoxylate bypass and gluconeogenesis are also regulated by Adr1 in *S. cerevisiae*

(YOUNG *et al.* 2003). Nevertheless, an Adr1 orthologue has not been characterized before in *Aspergillus* species (SALAZAR *et al.* 2009).

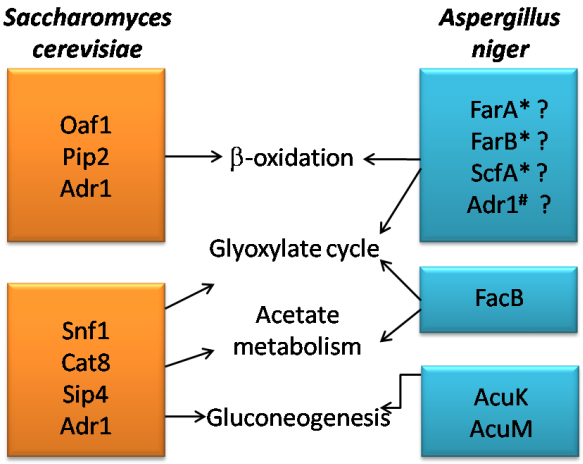



**Figure 5**. Comparison of transcriptional regulatory circuits controlling pathways for growth on gluconeogenic carbon sources of *S. cerevisisae* and *A. niger*. *The transcriptional regulators FarA, FarB and ScfA have been identified and mutants lacking these regulatory proteins haven been characterized in *A. nidulans*. #Orthologues to the *S. cerevisiae* transcriptional regulator Adr1 have been identified in *A. nidulans*, *A. oryzae* and *A. niger*. Figure adapted from Hynes in (HYNES 2008).

### 1.7 Nitrogen regulation

The *Aspergillus niger* transcriptional activator AreA is a key regulator of nitrogen metabolism. It was cloned and characterized since the end of the 90's and it shows an overall identity with its orthologues from other fungal species which varies between 32 and 72 % (MACCABE *et al.* 1998). Northern analysis indicated the synthesis of multiple transcripts, similarly to the *A. nidulans* orthologue, where the major species lie between 2.95 kb and 3.1 kb (MACCABE *et al.* 1998). In *A. nidulans*, AreA transcription is highly regulated in response to nitrogen nutrient quality or availability, differential *areA* mRNA stability and interaction with other co-repressors and co-activators such as NmrA and TamA (WONG *et al.* 2009a).

AreA contains a C-terminal GATA zinc finger DNA binding domain, whereas AreB, which is another regulator of nitrogen metabolism identified in *A. nidulans* and *P. chrysogenum*, contains a N-terminal GATA zinc finger DNA binding domain as well as a C-terminal leucine zipper domain. AreB possibly competes with AreA for DNA binding. In order to investigate

the role of AreB in nitrogen regulation in *A. niger*, we analyzed the *areB* deletion phenotype in different cultivation conditions and conducted transcriptome analysis. In *A. nidulans*, AreB antagonizes AreA activation and it has been reported to function in growth, asexual development and conidial germination, but not in sexual development (WONG *et al.* 2009a). Over-expression of AreB prevents AreA-dependent gene expression and confers severe growth inhibition. Functional characterization of the AreB domains has shown that both, the GATA zinc finger DNA-binding domain and the leucine zipper domain are required for its function. Preliminary studies on *A. nidulans* on different nitrogen sources have shown that AreB has a wide domain of action including but not limited to nitrogen regulation (WONG *et al.* 2009a).

**1.8 Transcriptional regulation**

Processes such as fungal development, metabolism, stress responses and other responses to diverse signals are regulated by a number of proteins known as transcription factors. Filamentous fungi, in particular *Aspergillus nidulans* and *Neurospora crassa*, have been used as model genetic systems to understand numerous processes and mechanisms underlying gene regulation (CADDICK 2004; FELENBOK and KELLY 1996; PENALVA and ARST 2004). Putative functional links between genes and proteins can be pictured on the basis of observed coordinated expression. Thus, this provides a valuable way to dissect the components of specific biological processes, therefore making transcriptomics a powerful tool to understand fundamental aspects of gene regulation.

The DNA sequence defines the binding sites for regulatory proteins, the relative proximity and arrangement of which combined with the DNA's physical properties, determine the function of a given motif (CADDICK and DOBSON 2008). Genome analysis of regulatory processes through comparative genomics searches for sequence elements and features involved in gene expressions, which are conserved, where there is functional conservation.

Regulatory proteins interact at specific promoters to either activate or repress transcription. Approximately 45 transcription factors have been characterized mainly in *A. nidulans*, and for many of them, the respective DNA binding motif has been also identified (CADDICK and DOBSON 2008; NAKAJIMA *et al.* 2000; TODD and ANDRIANOPOULOS 1997). The availability of the genome sequences has allowed the identification of a large number of genes that putatively encode DNA-binding proteins. These can be classified according to distinct classes with regards of the PFAM (protein family) domain, resulting in at least 86 different classes

(CADDICK and DOBSON 2008), where the relative distribution is generally well conserved among Aspergilli. The Zn(II)2Cys6 (Zn binuclear cluster) motif is regarded as being fungal specific (TODD and ANDRIANOPOULOS 1997) and is proposed to have evolved after the fungi diverged from other eukaryotes (CADDICK and DOBSON 2008). It is one of the most important classes of transcription factors including regulators of a wide range of biological processes such as primary and secondary metabolism, development and drug resistance.

In general, Zn(II)2Cys6 transcription factors possess a well conserved N-terminal localized DNA binding motif (CX2CX6CX5–16CX2CX6–8), followed by a less well defined conserved domain known as a fungal specific transcription factor domain. The C-terminal part of the Zn(II)2Cys6 transcription factor normally contains the activation domain (YUAN *et al.* 2008a). A model for zinc cluster protein DNA recognition proposed by MacPherson et al. (MACPHERSON *et al.* 2006) is shown in Figure 6.

In the genome of filamentous fungi several putative transcription factors are present. In the recently released genome sequence of *A. niger* CBS 513.88, 296 ORFs were identified showing the Zn(II)2Cys6 motif (PEL *et al.* 2007), and similar numbers are predicted from the genomes of *A. niger* ATCC 1015 and in other *Aspergillus* species such as *A. oryzae* and *A. nidulans* (PEL *et al.* 2007)). The role of only a few of the transcription factors in the filamentous fungus *A. niger* has been studied in detail (see Table 4). At least three genes which effect control across a broad spectrum of metabolic activities (wide domain) have been identified. In *A. niger*, these wide domain regulatory genes have been cloned and characterized: *creA* (DRYSDALE *et al.* 1993), the negatively-acting regulator of carbon catabolite repression, *pacC* (MACCABE *et al.* 1996; SARKAR *et al.* 1995) which regulates gene expression in response to external pH and *areA* (MACCABE *et al.* 1998), a positive regulator of nitrogen metabolite repression, where the presence of preferred nitrogen sources such as ammonium and L-glutamine, leads to repression of activities involved in the utilization of other less- favoured nitrogen sources. Recently, a global regulator of secondary metabolism, *laeA*, was identified (BOK and KELLER 2004). The fact that the predicted *laeA* gene seems intact in *A.oryzae/A.niger* may suggest the existence of a global regulation mechanism of secondary metabolism by other transcriptional regulatory factors (MACHIDA *et al.* 2008).

*A. niger* possesses by far, a larger number of putative Zn(II)2Cys6 transcription factors than its close relatives *A. nidulans*, *A. oryzae*, *A. fumigatus*, and the more distant ascomycetes species *N. crassa* and *S. cerevisiae*. A distribution of putative DNA-binding domains in the six fungal species is presented by Caddick and Dobson (CADDICK and DOBSON 2008).

**Table 4**. Known regulatory proteins in *Aspergillus niger*

| Factor | Function | Binding domain | Recognition motif | References |
|---|---|---|---|---|
| CreA | Carbon catabolite repression | C2H2 zinc finger | SYGGRG | (DRYSDALE *et al.* 1993) |
| PacC | Regulator in response to external pH | C2H2 zinc finger | GCCARG | (MACCABE *et al.* 1996; SARKAR *et al.* 1995) |
| CpcA ? | Regulator of amino acid biosynthesis | bZip | TTGASTCWG | Not confirmed in *A. niger* |
| AreA | Positive-acting regulator of nitrogen metabolite repression | GATA zinc finger | n.d. | (MACCABE *et al.* 1998) |
| FacB | Regulator of acetate and acetamide metabolism | Zn(II)2Cys6 | n.d. | (TODD and ANDRIANOPOULOS 1997; TODD *et al.* 1997) |
| RlmA | Regulator of genes induced in response to cell wall stress | MADS-box | $CTA(T/A)_4TAG$ | (DAMVELD *et al.* 2005a) |
| AmyR | Regulator of starch degrading enzymes | Zn(II)2Cys6 | CGGN8CGG | (PETERSEN *et al.* 1999; YUAN *et al.* 2008b) |
| InuR | Regulator of inulinolytic genes | Zn(II)2Cys6 | CGGN8CGG | (YUAN *et al.* 2008a) |
| XlnR | Regulator of xylanolytic enzymes required for xylan degradation and endoglucanase gene expression | Zn(II)2Cys6 | GGCTAAA | (VAN PEIJ *et al.* 1998) |
| AraR | Regulator of arabinose degrading enzymes* | Zn(II)2Cys6 | n.d. | de Vries et al., FGC 2009. |
| PrtT | Regulator of extracellular proteases | Zn(II)2Cys6 | n.d. | (PUNT *et al.* 2008) |

Note: DNA-binding regulatory proteins formally characterized in *A. niger*, their regulatory role, class of DNA-binding domain, recognition motif if known are listed. Ambiguous bases in the consensus sequences are given as R=A or G, Y=C or T, H=A, C or T, K=G or T, M=A or C, S=G or C, and W=A or T, N=any base. n.d.: not determined. Zn(II)2Cys6 is also known as Zinc binuclear cluster according to (CADDICK and DOBSON 2008). * ΔAraR mutants were unable to grow on L-arabinose and L-arabitol (de Vries et al., Fungal Genetics Conference, Asilomar 2009). AraR might have generated from gene duplication in Aspergilli and is only present in Aspergilli and not in other filamentous fungi. According to the authors, AraR is very similar to XlnR.

Many of these proteins are involved in regulating metabolism, i.e., XlnR or InuR, being responsible for monitoring the presence of specific metabolites and the regulation of the corresponding clusters of genes (YUAN *et al.* 2008a).

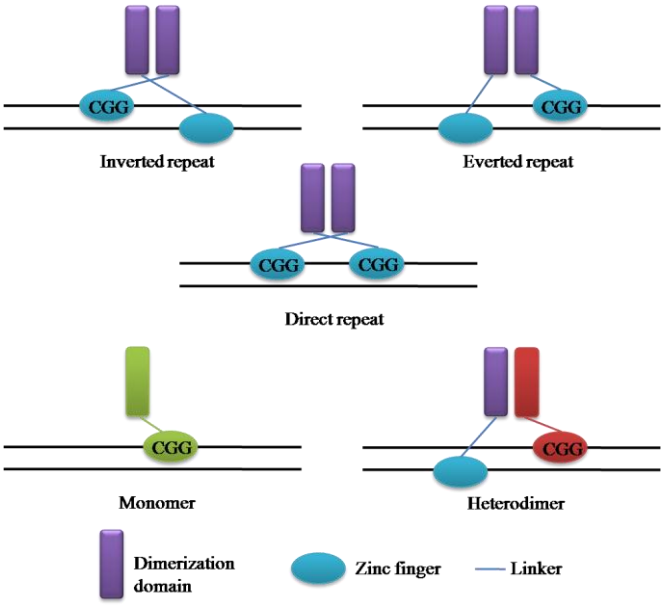



**Figure 6**. Model for zinc cluster protein DNA recognition. Zinc cluster proteins preferentially bind to CGG triplets and they can be oriented in three different configurations: inverted, everted and direct repeats. The orientation of CGG triplets and the nucleotide spacing between the triplets are the two major determinants of DNA-binding specificity (MACPHERSON *et al.* 2006). Zinc cluster proteins can bind as monomers (in green), as homodimers (two molecules in purple) and as heterodimers (one molecule in purple and one in red). Picture modified from (MACPHERSON *et al.* 2006).

A primary goal is the identification of key regulatory elements within the genome allowing for prediction of gene expression. This includes the identification of specific DNA motifs as well as the characterization of additional features within the sequence that influence their function. Processes established before divergence of species are likely to be conserved retaining the regulatory components, e.g., the DNA-binding domains of regulatory proteins, such as the GATA factors regulating nitrogen metabolism (SCAZZOCCHIO 2000) and the PacC orthologues, which are responsible for regulation of the response to ambient pH and are conserved across ascomycetes (PENALVA and ARST 2004). In contrast, newly evolved

regulatory systems will be specific to another group of species, therefore defining novel functions whereas other regulatory proteins might have been lost during evolution. An example of this, it's the existence of some orthologous proteins conserved in ascomycetes, but not present in *S. cerevisiae* such as LaeA (BOK and KELLER 2004).

A major source of underutilized information lies in the promoters and other regulatory elements, as they define which parts of the genome are transcribed, its level of expression and regulation (CADDICK and DOBSON 2008). There are well-established in vitro and in vivo methods for determination of specific binding of specific transcription factors to a given DNA sequence. Nevertheless, there are various examples where motifs found in vitro appear not to be functionally significant when analyzed in vivo (Punt et al. 1995; Mathieu et al. 2000; Gomez et al. 2003). Recently, the development of chromatin immunoprecipitation assays (ChIP) has been successfully developed for *A. nidulans* (CADDICK and DOBSON 2008). The combination with intergenic arrays to undertake ChIP on CHIP experiments (Horak et al. 2002) would give a global picture of transcription factor function in Aspergilli.

However, conventional analysis of transcription factors combined with genome data to assign putative functions has led to the identification of genes subject to transcription factor regulation and the establishment of putative consensus sequences. An example of this is the FarA/FarB approach followed by Hynes et al. (Hynes et al. 2006) for identification of fatty acid catabolism regulated genes or the CpcA element (Gcn4 orthologue in *S. cerevisiae*), which was identified as a conserved sequence upstream of a number of genes involved in amino acid transport and metabolism in *A. nidulans*, *A. fumigatus* and *A. oryzae* (GALAGAN *et al.* 2005), and which was consistent with the known function of CpcA (Hoffmann et al. 2001).

A fundamental aspect of gene regulation is that different signals are coordinated via specific regulatory elements, where the competition or cooperative binding at a given sequence results in the appropriate regulatory response. An example is the regulation of ethanol catabolism by AlcR in *A. nidulans* (Mathieu et al., 2000).

The identification of regulatory motifs in silico is an important component of functional genomics. There is a wide range of tools available that apply different bioinformatic algorithms for identification of consensus motifs such as RSAT (van Helden, 1998), Cosmo (Bembom et al. 2007), MEME (http://meme.sdsc.edu/meme/meme.html), Asap (Marstrand et al. 2008), among others and which run as web-server applications or are downloadable as separate applications. The choice of method and interpretation of the results is responsibility of the user, but it is generally recommended to use more than one method to confirm results.

## 2. Materials and Methods

### 2.1 Strains

The strains used in this project were the *A. niger* strains BO1, ATCC 1015, ATCC 11414 and CBS 513.88. The former is a glucoamylase producer strain obtained from Novozymes (PEDERSEN *et al.* 2000b), the second is the *A. niger* citric acid producer strain sequenced by the Joint Genome Institute (JGI) and ATCC 11414 is an improved citric acid producer strain derived from ATCC 1015, both obtained from the IBT collection (Biosys, DTU). The later is the *A. niger* strain sequenced by DSM (PEL *et al.* 2007), ancestor of the industrial protein production strains. All organisms were maintained as frozen spore suspensions at -80 °C in 20% glycerol.

### 2.2 Cultivation media

The composition of the batch cultivation medium was the following: 7.3 g $L^{-1}$ $(NH_4)_2SO_4$, 1.5 g $L^{-1}$ $KH_2PO_4$, 1.0 g $L^{-1}$ $MgSO_4.7H_2O$, 1 g $L^{-1}$ NaCl, 0.1 g $L^{-1}$ $CaCl_2.2H_2O$, 0.05 ml/L antifoam 204 (Sigma) and 1 ml $L^{-1}$ of trace elements solution. Trace elements solution composition (g $L^{-1}$): 7.2 g $ZnSO_4.7H_2O$, 1.3 g $CuSO_4.5H_2O$, 0.3 g $NiCl_2.6H_2O$, 3.5 g $MnCl_2.4H_2O$ and 6.9 g $FeSO_4.7H_2O$. In the studies where batch cultivations were carried out, glucose monohydrate (20 g $L^{-1}$), xylose (20 g $L^{-1}$), glycerol (20 g $L^{-1}$) and maltose monohydrate (20 g $L^{-1}$) were used (ANDERSEN *et al.* 2008c).

The composition of the continuous cultivations medium was the following: 2.5 g $L^{-1}$ $(NH_4)_2SO_4$, 0.75 g $L^{-1}$ $KH_2PO_4$, 1.0 g $L^{-1}$ $MgSO_4.7H_2O$, 1 g $L^{-1}$ NaCl, 0.1 g $L^{-1}$ $CaCl_2.2H_2O$, 0.05 ml $L^{-1}$ antifoam 204 (Sigma) and 1 ml $L^{-1}$ of trace elements solution. The concentration and carbon sources used varied depending on the study carried out.
In the study where continuous cultivations were conducted, the concentrations of the carbon sources used were: glucose monohydrate (4 g $L^{-1}$), xylose (3.64 g $L^{-1}$) and glycerol (3.72 g $L^{-1}$) giving the same amount of C-mol fed to the bioreactors, 0.121 Cmol $L^{-1}$.

Complex media composition for *A. niger* spores propagation: 10 g $L^{-1}$ glucose monohydrate, 2 g $L^{-1}$ yeast extract, 3 g $L^{-1}$ tryptone, 0.52 g $L^{-1}$ KCl, 0.52 g $L^{-1}$ $MgSO_4.7H_2O$, 1.52 g $L^{-1}$ $KH_2PO_4$, 20 g $L^{-1}$ agar and 1 ml $L^{-1}$ of trace elements solution. The trace elements solution

used in this case contained (g $L^{-1}$): 0.4 g $CuSO_4.5H_2O$, 0.04 g $Na_2B_4O_7.10H_2O$, 0.8 g $FeSO_4.7H_2O$, 0.8 g $MnSO_4.H_2O$, 0.8 g $Na_2MoO_4.2H_2O$ and 8 g $ZnSO_4.7H_2O$.

Minimal media composition for *A. niger* cultivation: 10 g $L^{-1}$ glucose monohydrate, 6 g $L^{-1}$ $NaNO_3$, 0.52 g $L^{-1}$ KCl, 0.52 g $L^{-1}$ $MgSO_4.7H_2O$, 1.52 g $L^{-1}$ $KH_2PO_4$, 15 g $L^{-1}$ agar, 1 ml $L^{-1}$ vitamins solution and 1 ml $L^{-1}$ trace elements solution. The pH was adjusted to 6.35 with 1M NaOH previously to sterilization. Vitamins solution composition (g $L^{-1}$): 1 g d-biotin, 1 g pyridoxine, 1 g thiamine, 1 g riboflavin, 1 g p-aminobenzoic acid (PABA), 1 g nicotinic acid and 20 drops of chloroform as preservative. The trace elements solution used in this case contained (g $L^{-1}$): 22 g $ZnSO_4 \cdot 7\ H_2O$, 11 g $H_3BO_3$, 5 g $MnCl_2 \cdot 4\ H_2O$, 5 g $FeSO_4 \cdot 7\ H_2O$, 1.7 g $CoCl_2 \cdot 6\ H_2O$, 1.6 g $CuSO_4 \cdot 5\ H_2O$, 1.5 g $Na_2MoO_4 \cdot 2\ H_2O$ and 50 g $Na_4EDTA$, pH was adjusted to 6.5 with KOH .

Transformation recovery media composition for *A. niger*: 10 g $L^{-1}$ glucose monohydrate, 182.17g $L^{-1}$ sorbitol, 6 g $L^{-1}$ $NaNO_3$, 0.52 g $L^{-1}$ KCl, 0.52 g $L^{-1}$ $MgSO_4.7H_2O$, 1.52 g $L^{-1}$ $KH_2PO_4$, 8 g $L^{-1}$ agar, 1 ml $L^{-1}$ vitamins solution and 1 ml $L^{-1}$ trace elements solution (same as described in minimal media composition for *A. niger* cultivation). If the transformations were conducted to disrupt the *pyrG* gene, the media was supplemented with 1 g $L^{-1}$ uridine and 1 g $L^{-1}$ uracil. The pH was adjusted to 6.35 with 1M NaOH previous sterilization.

### 2.3 Preparation of inoculum

*A. niger* BO1, ATCC 1015 and CBS 513.88 fermentations were inoculated with spores propagated on complex media plates, incubated for 6-8 days at 30 °C. The same stock of spores was used to inoculate all plates. In all cases, *A. niger* spores were harvested by adding 10 ml of Tween 80 0.01%, subsequently washed with 0.9% NaCl solution and filtered through sterile Miracloth (Calbiochem, San Diego, CA, USA). Spores were resuspended in 100 ml Tween 80 0.01% and counted. *A. niger* cultivations were inoculated with a spore suspension to obtain a final concentration of $5.7X10^9$ spores $L^{-1}$.

### 2.4 *Aspergillus niger* cultivations

#### 2.4.1 Batch cultivations

*A. niger* batch cultivations were carried out in either 5 L reactors with a working volume of 4.5 L (custom-designed) or 2.7 L with a working volume of 2 L (DASGIP AG, Jülich

Germany). The bioreactors were equipped with two Rushton four-blade disc turbines, pH and temperature control. Inlet air was controlled with a mass flowmeter. The concentrations of oxygen and carbon dioxide in the exhaust gas from the 5 L bioreactors were monitored with a gas analyzer (1311 Fast response Triple gas, Innova combined with multiplexer controller for Gas Analysis MUX100, B. Braun Biotech International). The concentrations of oxygen and carbon dioxide in the exhaust gas from the 2.2 L bioreactors were monitored with the DASGIP automated system. The temperature was maintained at 30°C and the pH was controlled by automatic addition of 2 N NaOH. The pH was initially set to 3.0 to prevent spore aggregation; when spores started to germinate, the pH was gradually increased to 4.5. Similarly, the stirring speed was initially set to 200 rpm and the aeration rate to 0.05 vvm (volume of gas per volume of liquid per minute) to prevent loss of hydrophobic spores from the medium to the head-space of the reactor. After germination, these parameters were increased to 600 rpm and 0.89 vvm and kept steady throughout all the rest of the fermentation.

### 2.4.2 Carbon-limited continuous cultivations

To determine the physiological characteristics of *A. niger* ATCC 1015 and to collect samples for transcriptome analysis, batch and continuous cultivations were carried out. The inlet feeding rate was started in the late exponential phase of the batch cultivations. The dilution rate was controlled either at 0.10 $h^{-1}$ or 0.05 $h^{-1}$. All cultivations were mass controlled. Carbon-limited conditions were assured by measuring the concentration of the liquid effluent which was 0.0 g $L^{-1}$ in all cases. The cultivations were assumed to be in a physiological steady-state when no significant variation in biomass dry weight and $CO_2$ production was observed after 3 retention times. To prevent wall growth in the upper part of the bioreactor, its top was cooled down to 4 °C.

#### 2.4.2.1 Batch phase of the continuous cultivations

These fermentations were performed in 2 L Braun bioreactors with a working volume of 1.6 L. Reactors were equipped with three Rushton four-blade disc turbines, pH and temperature control. The temperature was maintained at 30 °C and the pH was controlled by automatic addition of 2 N NaOH. The pH was initially set to 3.0 to prevent spores aggregation and only when spores started to germinate, the pH was increased to 4.5 and kept constant through the cultivation. Likewise, the stirring speed was initially set to 100 rpm and the aeration rate to

0.06 vvm (volume of gas per volume of liquid per minute) to prevent pellet formation and loss of hydrophobic spores from the medium to the headspace of the reactor. After germination, these parameters were increased to 600 rpm and 0.89 vvm and kept steady throughout all the rest of the fermentation. The cultivation conditions allowed dispersed filamentous growth of the fungus. The concentrations of oxygen and carbon dioxide in the exhaust gas were monitored with a gas analyzer.

**2.5 Sampling**

For quantification of cell mass and extracellular metabolites, a known volume of cell broth was withdrawn from the reactor, filtered and washed. The culture filtrates were frozen at -20 °C for subsequently sugar and extracellular metabolite quantification. Cell dry weight was determined using nitrocellulose filters (pore size 0.45 μm, Gelman Sciences). The filters were pre-dried in an oven at 100 °C for 24 h, cooled in a desiccator and subsequently weighed. A known volume of cell culture was filtered, washed with distilled water and dried on the filter for 24 h in an oven at 100 °C. The filter was cooled in a desiccator and weighed for cell mass concentration determination. For gene expression analysis, mycelium was harvested in the mid exponential phase or in steady state after three retention times. The cultures were filtered through sterile Miracloth and washed with a suitable amount of 0.9% NaCl solution. The mycelium was quickly dried by squeezing and subsequently frozen in liquid nitrogen. Samples were stored at -80 °C until RNA extraction.

**2.6 *Aspergillus niger* protoplastation method**

A liquid culture of 100 ml of complete media or PDB (potato dextrose broth) was initiated by inoculating a suitable amount of *A. niger* ATCC 1015 spores to achieve a final concentration of $10^6$ conidia $ml^{-1}$. The culture was grown overnight for ~18 h by shaking at 150 rpm and $30^oC$. Mycelia was harvested by filtering the culture through sterile Miracloth and subsequently rinsed with sterile water. A protoplastation solution (20 ml) was prepared by dissolving the cell wall digesting enzyme Vinoflow FCE (Novozymes, Denmark) in protoplasting buffer (0.6M $(NH_4)_2SO_4$ and 50 mM maleic acid, pH adjusted to 5.5) to achieve a final concentration of 60 mg $ml^{-1}$. Mycelia were added to the protoplastation solution and digested by shaking at $30^oC$ and 70 rpm for 3 to 4 hours. The progression of the digestion was checked at various intervals by taking small samples and looking at them under the microscope (400x magnification). The protoplasts are large round cells that are very sensitive

(larger size than conidia). When most or all the mycelia were digested, the culture was filtered through sterile Miracloth. The filtrate was centrifuged at 4$^{o}$C and 800 g for 10 minutes in order to pellet the cells and then resuspended in 25 ml of STC solution (1 M sorbitol, 100 mM Tris-HCl pH 7.5, 10 mM $CaCl_2.2H_2O$). The protoplasts were centrifuged again and resuspended in less than 1 ml STC in order to have a concentrated solution. The protoplasts were counted to get approximately $10^8$ protoplasts ml$^{-1}$ and diluted to a concentration of at least 1.2x10$^7$ protoplasts ml$^{-1}$. Subsequently, a 40% PEG solution (PEG 4000 dissolved in STC) was added to the protoplasts solution and mixed gently to a final concentration of 20% v/v. Then, dimethyl sulfoxide (DMSO) was added to the protoplasts-PEG solution and mixed thoroughly to get a final concentration of 7% v/v. Protoplasts were stored at -80 °C in aliquots of 200 µl per tube and 100 µl were used per transformation.

**2.7 *Aspergillus niger* transformation**

*A. niger* transformations were conducted with the polyethylene glycol (PEG) transformation method using 100 µl protoplasts thawed on ice per transformation each time. The protoplasts were aliquoted into 15 ml falcon screwcap conical tubes. Subsequently, 1 to 10 µg of DNA dissolved in TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA pH 8.0) were added (10 µl volume at a maximum). The mixture was incubated on ice for 15 min and 1 ml 40% PEG 4000 dissolved in STC solution was added for subsequent incubation for 15 min at room temperature. Then, 8 ml of liquid transformation recovery media were added. Tubes were incubated at 70-80 rpm and 30 °C for 1 h. After incubation, the tubes were centrifuged at 800 g for 8 min at room temperature using a swinging bucket rotor, supernatant was poured off and 12 ml of 48 °C transformation recovery media containing 0.8% agar was added. The selection compound was added in this step. In case of using hygromycin, a final concentration of 100 µg ml$^{-1}$ was used. In case of using 5-fluoroorotic acid (5-FOA), the final concentration was 1 or 2 mg ml$^{-1}$. Then, the transformation mixture was mixed gently by inversion and poured into pre-labelled plates swirl to spread. Plates were left in the sterile hood until the agar solidified and subsequently wrapped with microporous tape for incubation at 30 °C until colonies were visible (~3 to 8 days depending on the selection compound used). Plates were overlaid the next day with 8 ml transformation media (media containing 1 M sorbitol, 0.8% agar and the required selection compound). The plates were incubated until visible colonies grew through the overlay; usually the ones that managed to go through the surface were

transformed. If there were none or significantly fewer colonies on the no DNA control plate, 12 to 14 well isolated colonies were picked and transferred to slants containing the selection compound.

Subsequently, spores from each of the transformation slants were collected and a liquid culture was started for genomic DNA extraction and later screening by polymerase chain reaction (PCR). The remaining spore suspension from each of the colonies was stored.

### 2.8 *Aspergillus niger* genomic DNA extraction

Genomic DNA extraction was conducted using the CTAB method where 3 ml of stationary cultures were initiated in 15 ml falcon tubes and incubated for app. 24 h at 30 °C. Mycelium was hooked from the media surface and the excess incubation media was dried using paper towels. Mycelium was transferred to 1.5 ml vials and frozen in liquid nitrogen for subsequent vacuum drying for several hours to overnight. Dried material was transferred to new 2 ml screwcap tubes and two 2 mm glass beads were added for subsequent bead beating during 1 min at full speed. Immediately, 750 µl CTAB buffer were added (2% CTAB, 100 mM Tris-HCl pH 8.0, 20 mM EDTA and 1.4 M NaCl) and tubes were incubated at 57 °C for 15 min to 1 h. A phenol/chloroform (P/C) extraction was conducted by adding 750 µl of P/C. The mixture was centrifuged at 14,000 g for 5 min and the supernatant was transferred to new 1.5 ml vials. One volume of 2-propanol was added following centrifugation at 6,000 g for 10 min (if centrifugation speed was too low, no pellet was formed). The supernatant was removed and the pellet was rinsed with a generous amount of 70% ethanol. Tubes containing the pelleted DNA were air-dried and the pellet was resuspended by adding 50 to 100 µl TE buffer.

### 2.9 Gene deletions

The transformation method described previously was used to conduct several gene deletions in the *A. niger* sequenced strain ATCC 1015. The targeted genes were the ones coding for the orthologue proteins in yeast Adr1, Cat8 (FacB *Aspergillus* orthologue), and the *Aspergillus nidulans* orthologue proteins CreA, AreB, and PyrG. The gene (*hph*) that confers resistance to the antibiotic hygromycin was used as a selectable marker. It was amplified from the plasmid PCB1003 (CARROLL *et al.* 1994), obtained from the Fungal Genetics Stock Center (http://www.fgsc.net/). The *pyrG* gene from *A. fumigatus*, amplified from plasmid pFNO3, was used to test the correct disruption of the *pyrG* gene in the *A. niger* strain by

complementation assays. The primers used for construction of the gene constructs are shown in Table 5.

**Table 5**. Primers used to conduct the genetic manipulations.

| Primer name | Sequence |
|---|---|
| **AdrA** | |
| AdrA-F1 | CACCGAAGGAAAGGTGGACG |
| AdrA-F2 | GTCAGAATGGGCGACCAGCA |
| AdrA-R3 | TGACCTCCACTAGCTCCAGCTCCTAGTAGAGCAGCGACACG |
| AdrA-HphF | CGTGTCGCTGCTCTACTAGGAGCTGGAGCTAGTGGAGGTCA |
| AdrA-HphR | GCCCGCGTAAGGAATGAATGCGGTCGGCATCTACTCTATT |
| AdrA-F4 | AATAGAGTAGATGCCGACCGCATTCATTCCTTACGCGGGC |
| AdrA-R5 | GCAACAAGTCACCTGCTCTG |
| AdrA-R6 | ACGGGCATCACTGCTTGCTG |
| **CreA** | |
| CreA-F1 | TTAACACACCGTGCGTGGCC |
| CreA-F2 | CTGATGGAGACACCCGTTTGC |
| CreA-R3 | TGACCTCCACTAGCTCCAGC AGCTTGTCCCAAGACCGA |
| CreA-HphF | TCGGTCTTGGGACAAGCTTCGCTGGAGCTAGTGGAGGTCA |
| CreA-HphR | GCACAAGTCTATTCGGTCGTAGCCGGTCGGCATCTACTCTATT |
| CreA-F4 | AATAGAGTAGATGCCGACCGGCTACGACCGAATAGACTTGTGC |
| CreA-R5 | GTAAGTCCCCATGACTTGCGG |
| CreA-R6 | CGGAAGTTCGGCATGAGAAGTC |
| **AreB** | |
| AreB-F1 | CGTCGTATACCACTCCCGGA |
| AreB-F2 | CAGTTCGGTCATTTGTGGCC |
| AreB-R3 | TGACCTCCACTAGCTCCAGCACAGTAGGATCACGCGAGGA |
| AreB-HphF | TCCTCGCGTGATCCTACTGTGCTGGAGCTAGTGGAGGTCA |
| AreB-HphR | GTCCATCGTCCGAACAAAGCCGGTCGGCATCTACTCTATT |
| AreB-F4 | AATAGAGTAGATGCCGACCGGCTTTGTTCGGACGATGGAC |
| AreB-R5 | GCTTATTGATTCTCCGCCTCG |
| AreB-R6 | AGCAGCAGCAAGGCAGTAAG |
| **FacB** | |
| FacB-F1 | CAGCGAGGAGAATGATGCCG |
| FacB-F2 | TCCGATGCAATACTCCGCCT |
| FacB-R3 | TGACCTCCACTAGCTCCAGCGTTGGCTGGATGCTTTGGCG |
| FacB-HphF | CGCCAAAGCATCCAGCCAACGCTGGAGCTAGTGGAGGTCA |
| FacB-HphR | TAGCTCAGACAGCCAGTCGTCGGTCGGCATCTACTCTATT |
| FacB-F4 | AATAGAGTAGATGCCGACCGACGACTGGCTGTCTGAGCTA |
| FacB-R5 | TATCGATCACGCATCGCAGC |
| FacB-R6 | TCTAGTCTAGCTTTCGGAGC |

| **PyrG** | |
|---|---|
| PG-F1 | GCAGGGAAAAATACGAGCTCCAATG |
| PG-F2 | AACCTGGGTGTGGCAACTTCAATGG |
| PG-R3B | TATGGGCTC ACTTATCTAGAATTGCTTCTGGACAGTGTTGCCAAT |
| PG-F3B | AGAAGCAATTCTAGATA AGTGAGCCCATATCATCAACTGCAGCA |
| PG-R4 | CACCCGTCGCCATTTGCTCTACGCA |
| PG-R5 | AAGCTTATCACCGTCCCTTATCAGC |
| ***hph* screen primer** | TGACCTCCACTAGCTCCAGC |

Nucleotides colored in red indicate sequences present in the hygromycin resistance gene and used for fusing the fragments.

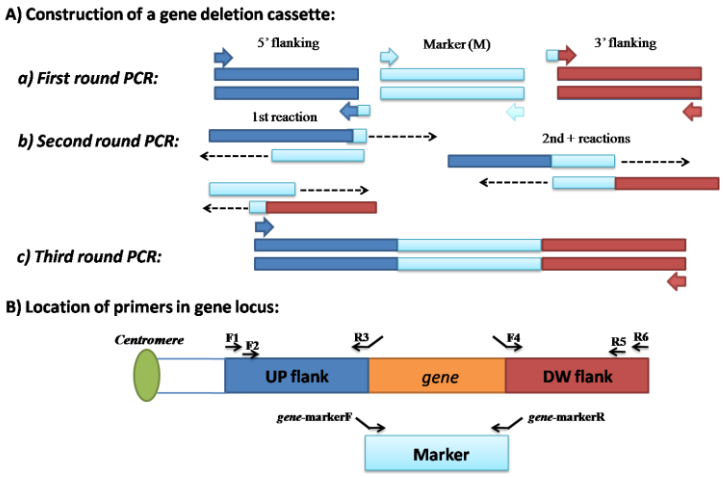



Figure 7. Schematic representation of gene deletion cassette construction. A) A typical reaction fuses DNA fragments of a 5' flanking sequence, a 3' flanking sequence and a marker (M). *a) First round PCR*: amplification of the components using the specific and chimeric primers. *b) Second round PCR*: the assembly reaction is carried out without using any specific primers, as the overhanging chimeric extensions act as primers. The first two cycles are shown in detail. *c) Third round PCR*: amplification of the final product using nested primers (primers *gene*-F2 and *gene*-R5). B) Location of primers in each gene locus targeted. The arrows numbered from 1 to 6 represent the primers used in the forward or reverse direction to amplify the up and down flanking regions of the manipulated gene. Primers 3 and 4 carried 20-25 bases of homologous sequence overlapping with the ends of the selectable marker of choice. Primers named *gene*-markerF and *gene*-markerR were used to amplify the marker used; they contain a ~20 bases tail which sequence is homologous to the sequence of the gene to be deleted, which make a total of 40 bases homologous between marker and gene.

The fragment construction strategy was the one designed earlier by Yu et al. (YU *et al.* 2004) and based on construction of the deletion cassette by PCR. Figure 7 shows the schematic representation of the gene deletion cassette construction steps where the upstream and downstream flanking sequences of the gene to be deleted were amplified by PCR. The location of the primers in each targeted gene locus is depicted in Figure 7, panel B). The arrows numbered from 1 to 6 represent the primers used in the forward or reverse direction to amplify the upstream and downstream flanking regions of the manipulated gene. Primers 3 and 4 carried 20-25 bases of homologous sequence overlapping the ends of the selectable marker of choice. Primers named *gene*-markerF and *gene*-markerR were used to amplify the marker; they contain a 20 bases tail which sequence is homologous to the sequence of the gene to be deleted. This accounts for a total of 40 bases homologous between the marker and the gene, i.e., *hph* and *A. niger ADR1* orthologue.

**2.10 PCR screening of transformants**

The genomic DNA extracted from each of the transformants was used to conduct PCR screening using the Takara Ex Taq kit according to the manufacturer's protocol (Takara Bio Inc., Shiga, Japan). Primers named *gene*-F1, i.e., Adr1-F1 and *hph* screen inner primer were used (Table 5). The *gene*-F1was located outside the manipulated gene locus for each of the targeted genes and the *hph* screen primer was located inside the *hph* gene sequence, as shown in Figure 7.

**2.11 Southern Blot**

Southern blot was conducted to check the correct disruption of the genes. It relies on the genomic DNA extraction from each of the pre-screened positive transformants. Those who gave positive in the PCR screening, followed restriction enzyme digestion and subsequent transfer of DNA to a membrane for hybridization using the designed biotin-labeled probe. Biotin-labeled DNA probes were prepared using the North2South Biotin Random Prime labeling Kit from Pierce according to the manufacturer's protocol (Pierce, Rockford, IL, USA). The genomic DNA was extracted using the CTAB method previously described and RNAse treated with the RNAse cocktail from Ambion (Ambion, Inc., TX, USA) to degrade RNA from the samples. For each strain, approximately 1 μg of genomic DNA was restriction enzyme digested using the appropriate restriction enzyme at 37 °C for 2 h (Fermentas International Inc., Canada). A gel electrophoresis was run including 10 ng BstEIIλ ladder on

lane 1, samples (digested DNA from transformants) and controls (digested DNA from wild type strain ATCC 1015). Alkaline transfer of DNA to the nylon membrane was done using the Turboblotter system developed by Whatman according to their protocol (Whatman Schleicher & Schuell, NJ, USA). DNA was fixed to the membrane by drying it in an oven at 80 °C for 2 h. Blocking of membrane was done by adding 5-10 ml of church buffer (0.5 M $NaHPO_4$ pH 7.2, 7% SDS, 2 ml 0.5 M EDTA pH 8 and 10g BSA) and incubated for 1 h at 60 °C. Lambda probe (5-10 ng $ml^{-1}$) and southern probe (30 ng probe $ml^{-1}$) were added to 1 ml church buffer and denatured for 5-10 min in a boiling bath. Subsequently, denatured probes were added to pre-blocked membrane. Southern hybridization with biotin-labeled probe was carried out at 60 °C over night and subsequently stringent washed. Probe detection was done using the chemiluminescent nucleic acid detection module (Pierce, Rockford, IL, USA) and a gel doc imaging system (Bio-Rad Laboratories Inc., CA, USA).

**2.12 Sugars and extracellular metabolites quantification**

The concentration of sugars and extracellular metabolites in the filtrates were determined using high pressure liquid chromatography (HPLC) on an Aminex HPX-87H ion-exclusion column (BioRad, Hercules, CA). The column was eluted at 60 °C with 5 mM $H_2SO_4$ at a flow rate of 0.6 mL $min^{-1}$. Metabolites were detected with a refractive index detector and an UV detector.

**2.13 Total RNA extraction**

*A. niger* total RNA was isolated using the Qiagen RNeasy Mini Kit (QIAGEN Nordic, Ballerup, Denmark), according to the protocol for isolation of total RNA from plant and fungi. For this purpose, approximately 100 mg of frozen mycelium were placed in a 2 ml tube, pre-cooled in liquid nitrogen, containing three RNase-treated steel balls (two balls with a diameter of 2 mm and one ball with a diameter of 5 mm). The tubes were subsequently shaken in a Mixer Mill, at 5 °C for 10 minutes, until the mycelium was ground to powder and thus ready for extraction of total RNA. All samples were inspected for good quality of total RNA extracted with a BioAnalyzer (2100 BioAnalyzer, Agilent Technologies Inc., Santa Clara, CA, USA). RNA quantification was performed in a spectrophotometer (Amersham Pharmacia Biotech, GE Healthcare Bio-Sciences AB, Uppsala, Sweden) and total RNA was stored at -80 $^{o}$C until further processing.

**2.14 Microarray manufacturing and design**

Affymetrix arrays were used for the analysis of the transcriptome data of *A. nidulans*, *A. oryzae and A. niger* (Affymetrix company, Santa Clara, CA, USA). The arrays were packaged in an Affymetrix® GeneChip cartridge (49 format), and were processed with GeneChip reagents in the GeneChip® Instrument System. The design and selection of probes for interrogating gene expression levels based on the genomes of *A. nidulans* FGSC A4 (BROAD INSTITUTE DATABASE), *A. oryzae RIB40* (DOGAN DATABASE) *and A. niger ATCC 1015* (JGI DATABASE) was performed by Andersen et al. (ANDERSEN *et al.* 2008c). The arrays contain a maximum of 11 non-overlapping perfect match (PM) probes of 25 oligomers length per gene. 11,122 probe sets were represented in the microarray *for A. niger*, 12,039 probe sets plus an EST collection (courtesy of Novozymes) for *A. oryzae* and 10,656 probe sets for *A. nidulans*.

**2.15 Preparation of biotin-labeled cRNA and microarray processing**

Biotin-labeled cRNA was prepared from either 1 μg or 5 μg of total RNA, according to the protocol described in the Affymetrix GeneChip® Expression Analysis Technical Manual (AFFYMETRIX and GENECHIP 2007). All samples were prepared in the same manner depending on the study. The cRNA was cleaned before fragmentation using the Qiagen RNeasy Mini Kit (protocol for RNA Cleanup), in order to guarantee good-quality cRNA samples for subsequent processing. Biotin-labeled cRNA was quantified in a spectrophotometer (Amersham Pharmacia Biotech, GE Healthcare Bio-Sciences AB, Uppsala, Sweden) and 20 μg were fragmented following the manufacturer recommendations. Approximately 15 μg of fragmented cRNA was hybridized to the 3AspergDTU Affymetrix GeneChip (ANDERSEN *et al.* 2008c) following the Affymetrix GeneChip® Expression Analysis protocol. Arrays were washed and stained using a GeneChip® Fluidics Station FS-400, and scanned on an Agilent GeneArray® Scanner 3000. The scanned probe array images (.DAT files) were converted into .CEL files using the Affymetrix GeneChip Operating Software.

**2.16 Transcriptome analysis**

Affymetrix CEL-data files were preprocessed using the statistical language R version 2.7.1 (R DEVELOPMENT CORE TEAM 2007) and Bioconductor version 2.2 (GENTLEMAN *et al.* 2004). The probe intensities were normalized for background by using the robust multiarray average

method with perfect match (PM) probes only (IRIZARRY *et al.* 2003). Subsequent normalization was performed using the qspline algorithm (WORKMAN C *et al.* 2002). Gene expression indexes were calculated from the PM probes with the median polish summary method (IRIZARRY *et al.* 2003). All statistical preprocessing methods were implemented in affy package (GENTLEMAN *et al.* 2004) using R scripts (DUDOIT *et al.* 2003). Statistical analysis was applied to identify differential gene expression levels. Moderated Student's t-tests between the different carbon sources for *A. niger* or the corresponding *Aspergillus* spp. was conducted using limma package (SMYTH 2004). Empirical Bayesian statistics were used to moderate the standard errors within each gene and Benjamini-Hochberg's method to adjust for multiple testing (BENJAMINI and HOCHBERG 1995). Unless otherwise stated, a cut-off of adjusted p value < 0.05 was used to assess for statistical significance. The Limma package was also used to conduct ANalysis Of VAriance (ANOVA) for the comparison of each *A. niger* knock-out strain grown on glucose or glycerol as carbon sources versus the wild type strain ATCC 1015 grown under the same conditions, i.e., a typical dataset would be: gene deletion mutant_glucose, gene deletion mutant_glycerol, ATCC 1015_glucose and ATCC 1015_glycerol.

**2.17 Protein sequence comparisons**

A cross comparison between the amino acid sequences of the predicted ORFs from each of the three *Aspergillus* genomes, based on DOGAN (DOGAN DATABASE), JGI (JGI DATABASE) and BROAD Institute databases (BROAD INSTITUTE DATABASE) using the BLASTP algorithm, was applied (ALTSCHUL *et al.* 1990b). The *A. oryzae* genome sequence (NBRC 100959), *A. nidulans* FGSC A4 version 3.0 and *A. niger* ATCC 1015 version 1.0 were used. An estimated expectation value cut-off of 1E-30 was set to assess for statistical significance. The best hit, based on the score, was selected for the case in which the protein query produced more than one hit. Bi-directional best hits were found by comparing the lists of best hits for two species against each other (i.e. *Niger_Oryzae, Oryzae_Niger*) and selecting those genes where the best hit in the other organism was the same best hit, thus giving a conservative set of 1:1 homologues for all three pair-wise comparisons. Tri-directional best hits were found by comparing the three lists of bi-directional hits (*Niger_Oryzae, Nidulans_Oryzae, Niger_Nidulans*) and selecting the genes that had a 1:1:1-relationship in all comparisons between all three species. The full subset of tri-directional homologues is given in Appendix 6: Supplementary Table 1.

### 2.18 Detection of conserved regulatory elements

For the publication where conserved motifs were searched (manuscript 2), several bioinformatics tools were applied for the detection of conserved regulatory elements. As a first step, pattern recognition was conducted in RSAT (Regulatory Sequence Analysis Tools) (VAN HELDEN *et al.* 1998) using the option of oligo-analysis. The method is based on the detection of over-represented oligonucleotides. The statistical significance of a site was assessed based on pre-computed tables of oligonucleotide frequencies observed in all non-coding sequences from *A. oryzae* and *A. nidulans* genomes, respectively, as these two organisms are supported by the application. In the case of *A. niger*, our own frequency table was calculated based on the intergenic regions from scaffold 1 of the *A. niger* ATCC 1015 genome sequence for 6, 7 and 8 base pairs (bps) oligonucleotides. Intergenic regions containing unknown bases (N's) were removed from the training set leaving 1214 sequences. The motif recognition was computed by running the analysis with a 1000 bps upstream region counted from the start codon of each gene or predicted transcription start site in the case of *A. niger*. A subset of 243 promoters, 3 times 81 promoters for each of the species, was analyzed. Statistical analysis was conducted to find consensus motifs in the subsets of 81 up-regulated conserved genes as well as in the 5 down-regulated conserved genes in the three Aspergilli species investigated, *A. nidulans*, *A. niger* and *A. oryzae*. The analysis was done considering a different length of consensus patterns, ranging from 6 to 8 bps for each *Aspergillus*. After having a number of probable consensus conserved motifs; these were further inspected using R 2.7.1 and Cosmo package (BEMBOM *et al.* 2007). Default settings were used and the program was run for different patterns length. A background Markov model was computed using the intergenic regions from scaffold 1 of the *A. niger* ATCC 1015 genome sequence as previously reported (ANDERSEN *et al.* 2008c). The two component mixture (TCM) model was used to search for a conserved motif where the maximum number of sites was increased to include all 174 binding sites. Finally, a more refined search for potential transcription factor binding sites in the subset sequences was done with the pattern search program Patch using TRANSFAC 6.0 public sites (http://www.gene-regulation.de/).

### 2.19 GO-term enrichment analysis

GO-term enrichment analysis was conducted with the *A. niger* ATCC 1015 conserved up-regulated genes list from manuscript 2 (81 genes) and with the significantly differentially

expressed genes list (p value < 0.05) using R 2.7.1 (R DEVELOPMENT CORE TEAM 2007) with BioConductor (GENTLEMAN *et al.* 2004) and the topGO-package v. 1.2.1 with the elim algorithm to remove local dependencies between GO terms (ALEXA *et al.* 2006). GO-term assignments were based on automatic annotation of the *A. niger* ATCC 1015 version 1.0 gene models, a cut-off of elim p value < 0.05 was used to assess significance.

**2.20 Clustering analysis**

A consensus clustering algorithm was used to identify similar expression profiles of the genes which changed expression level. The algorithm was implemented in the MATLAB toolbox ClusterLustre (GROTKJAER *et al.* 2006) and it is available at www.sysbio.se (CVIJOVIC *et al.* 2010). Pearson correlation coefficient was used as a similarity metrics for data processing. A partitioning clustering method by k-means (POLLARD and VAN DER LAAN 2005) was used to account for average expression of the biological replicates under each condition.

## 3. Results and Discussion

### 3.1 Genome sequencing of *Aspergillus niger*

It is well known that the filamentous fungus *Aspergillus niger* exhibits a great diversity in its phenotype. It is found all over the globe, both as marine and terrestrial strains, produces both organic acids and bio-material degrading enzymes in high amounts and has been known to exhibit pathogenicity. For these reasons and more, it is of great interest to a number of communities ranging from basic research to applied sciences such as biotechnology and medicine.

Although the industrial enzyme-producing *A. niger* CBS 513.88 was already sequenced and published in 2007 (PEL *et al.* 2007), the diversity of this species still allowed for additional exploration. Therefore, a whole genome sequencing project of the acidogenic *A. niger* wild type strain ATCC 1015 was initiated by the Joint Genome Institute (JGI) as an initiative of the Department of Energy of the United States and it produced a sequence of very high quality. The sequence is available in the *Aspergillus niger* JGI Genome Portal (http://genome.jgi-psf.org/Aspni5/Aspni5.home.html). The assembly release version 1.0 of whole genome shotgun reads was constructed with the JGI assembler, Jazz, using paired end sequencing reads at coverage of 8.9X. The Genome Portal includes an overview of automatic and manual annotation, where more than 2,100 gene models have been subjected to manual annotation, task conducted by our former fungal research group at DTU.

According to the ATCC 1015 genome sequence analysis conducted, only 15 gaps were present in the sequence, half of the telomeric regions were elucidated and 11,200 protein-coding genes were predicted. Genome statistics are summarized in Table 6. The finished contigs (24 in total, spanning 34.85 Mb) are available from NCBI (Acc.nos. XX).

**Comment [p1]:** Check accession numbers

The contiguous ATCC 1015 sequence was used to close 186 contig gaps between adjacent contigs found in the CBS 513.88 sequence detected by PCR followed by sequencing. This led to a number of improved gene models resulting from mergers of terminal truncated genes and the inclusion of a number of new genes (see results section in manuscript 1 for details). An updated version of the *A. niger* CBS 513.88 genome sequence can be accessed through NCBI (Acc.nos. XX-XX). However, a difference of 0.8 megabase (Mb) still remained, accounting for approximately 2.5% of the full genome.

**Comment [p2]:** Check accession numbers

The genome sequence analysis as well as transcriptome analysis of the citric acid-producing *A. niger* wild type strain ATCC 1015 and of the enzyme producing strain CBS 513.88 is the foundation of manuscript 1 submitted to the journal *Genome Res.*

**Comment [p3]:** Check the Journal, where it was finally sent

**Table 6**. General genome statistics for *A. niger* ATCC 1015 and *A. niger* CBS 513.88. Except for genome sizes and the number of gene models, all values are reported as averages.

| | ***A. niger* ATCC 1015** | ***A. niger* CBS 513.88[a]** | ***A. niger* CBS 513.88[b]** |
|---|---|---|---|
| Gene models | 11,200 | 14,165 | 14,082 |
| Genome size (Mbp) | 34.85 | 33.93 | 34.02 |
| Gene length (bp) | 1696.1 | 1572.8 | 1589.0 |
| Transcript length (nt)[c] | 1501.3 | 1322.5 | 1330.4 |
| Protein length (aa) | 484.3 | 439.9 | 442.5 |
| Exons per gene | 3.1 | 3.6 | 3.6 |
| Exon length (bp) | 480.8 | 370.0 | 371.6 |
| Intron length (bp) | 93.8 | 97.2 | 96.9 |

[a]The genome assembly published by Pel et al. (2007) (PEL *et al.* 2007).

[b]Genome assembly of *A. niger* CBS 513.88 after gap closure using sequence information from ATCC 1015.

[c]The high difference in "Gene length (nt)" and "Transcript length (nt)" between *the A. niger* ATCC 1015 and the two versions of the *A. niger* CBS 513.88 is most likely due to differing definitions of both terms. Since for the *A. niger* CBS 513.88 genome versions no predictions of promotor or terminator sequences are available, the 'Gene length' is calculated from START codon to STOP codon. The 'Transcript length' is calculated the same way, but without intron sequences.

### 3.1.1 Comparative genomics of *A. niger* CBS 513.88 and ATCC 1015

Based on pair-wise alignments of the corresponding chromosomal arms of the two sequenced *A. niger* strains, CBS 513.88 and ATCC 1015, we can tell that the two strains are largely syntenic. Sequence analysis revealed an average of 7.84 single nucleotide polymorphisms per kilobase (SNPs/kb) with levels as high as 160 SNPs/kb in hyper variable regions. Comparative genomics uncovered several genome rearrangements, a clear case of strain specific horizontal gene transfer (HGT) and identified 0.8 megabase of novel sequence.

In an un-matched region identified in the left arm of chromosome III, in comparison with ATCC 1015, the CBS 513.88 genome harbors two additional and identical α-amylase encoding genes which are identical to the α-amylase encoding genes AO090023000944 and AO090120000196 from *A. oryzae* RIB40. These findings strongly suggested that the strain CBS 513.88 recently acquired these duplicate An12g06930/An05g02100 α-amylase genes

through horizontal gene transfer (HGT) from most likely an *A. oryzae* strain. The suggested directionality of the HGT recombination event was inferred from the fact that the particular α-amylase gene and immediate five downstream genes are present in the genomes of *A. oryzae* RIB40 and of strain CBS 513.88, but they are not present in the ATCC 1015 genome. Supporting our findings, it has been reported that in other black Aspergilli, α-amylase encoding genes are > 99 % identical at DNA level to the *A. oryzae* RIB40 and the *A. niger* CBS 513.88 α-amylases (SHIBUYA ET AL., 1992; KORMAN ET AL., 1990) leaving room for alternative HGT scenarios.

**Comment [p4]:** Insert reference in correct way through EndNote

All this genome sequence analysis forms the foundation of manuscript 1, where we present the genome sequence of the citric acid-producing *A. niger* wild type strain ATCC 1015, and an updated genome sequence for the enzyme producing strain CBS 513.88. In this manuscript, the study of the diversity of the *A. niger* species based on exometabolite profiling conducted by our partners in DTU, and new genomic sequences from seven isolates confirmed wide variation within the species. A detailed list of alleles was generated, and differences in genotype were observed to accumulate in metabolic pathways essential to acid production as well as protein synthesis. In this study we present a number of genome-scale comparative data-sets which will serve as the foundations for new hypotheses useful in studying and optimizing either type of producer-strain. This multi-disciplinary comparative analysis identified a number of factors on multiple levels that are indicative of specialization as an acidogenic or an enzyme-producing strain.

### 3.1.2 Characterization and transcription analysis of batch cultures with the two sequenced *A. niger* strains CBS 513.88 and ATCC 1015 show distinct differences in the phenotype and transcriptome profile

In order to evaluate the effect of the differences in genome sequence on the physiology of the two *A. niger* strains, CBS 513.88 and ATCC 1015, a comparative transcriptome analysis was performed. The two strains were grown exactly under the same conditions in batch cultures in a glucose based minimal medium designed for enzyme production and dispersed filamentous growth as reported in Materials and Methods. Samples were harvested for transcriptome and metabolite analysis at the mid exponential phase of growth. The transcriptome profiles were analyzed using an Affymetrix array based on the genome sequence of ATCC 1015 (ANDERSEN *et al.* 2008c). Several characteristics of the cultures physiology were measured

and are summarized in Table 7. Glucoamylase A activity was measured and resulted to be more than six times higher in cultivations of the glucoamylase producing strain *A. niger* CBS 513.88, whereas the citric acid concentration in the culture filtrates was fairly similar.

**Table 7**. Statistics for batch cultivations of *A. niger* ATCC 1015 and *A. niger* CBS 513.88. Fermentations were performed in three biological replicates for each strain. Values are presented as average±standard deviation. $\mu_{max}$ and $Y_{sx}$ are general statistics for the fermentations, while the remaining values are specific for the time of sampling for transcription analysis. GlaA refers to glucoamylase A.

| | **ATCC 1015** | **CBS 513.88** |
|---|---|---|
| mRNA sample (h) | 24.5±1.2 | 40.2±4.2 |
| Biomass (g $L^{-1}$) | 5.0±0.1 | 4.0±0.5 |
| $\mu_{max}$ ($h^{-1}$) | 0.17±0.01 | 0.15±0.01 |
| Glucose (g $L^{-1}$) | 10.0±0.6 | 9.5±0.4 |
| Glycerol (g $L^{-1}$) | 0.09±0.02 | 0.27±0.03 |
| $Y_{sx}$ (Cmol/Cmol)[a] | 0.67±0.03 | 0.55±0.03 |
| GlaA (U $mL^{-1}$)[b] | 9.6±3.2 | 62.8±9.3 |
| Citric acid (g $L^{-1}$) | 0.10±0.12 | 0.14±0.03 |

[a] Biomass was converted to Cmol by means of the factor 24.9 g biomass $Cmol^{-1}$ (NIELSEN *et al.* 2003) .

[b] One unit of glucoamylase can be assumed to correspond to 25µg of protein (PESL protein assay; Boehringer Mannheim, Mannheim, Germany).

Statistical analysis showed 4,784 significantly differentially expressed genes under the cut-off imposed ($p < 0.05$), where 2,431 genes had a higher expression index in *A. niger* CBS 513.88 and 2,353 a higher index in *A. niger* ATCC 1015. To further explore the differences in metabolism of the two strains, the significantly differentially expressed genes were examined in the context of the metabolic network. An almost equal number of genes were up-regulated in each of the two strains. A substantial subset of the metabolic genes was up-regulated in CBS 513.88, including central pathways such as glycolysis and the tricarboxylic acid cycle. This was rather surprising, since the specific maximum growth rate of ATCC 1015 was higher than for CBS 513.88, therefore expecting higher expression levels of genes involved in the

production of precursors for biomass and replication machinery for ATCC 1015 and not for CBS 513.88. Apparently, the strain with higher glucoamylase productivity at a lower growth rate demanded a higher metabolic activity in comparison to the acidogenic strain ATCC 1015.

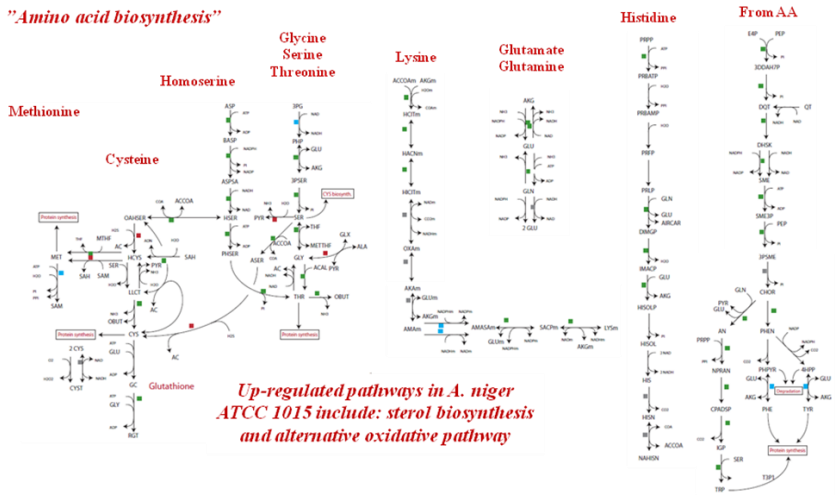



**Figure 8**. Significantly regulated genes mapped into the reconstructed *A. niger* metabolic network. Green boxes denote reactions where the corresponding genes are up-regulated in *A. niger* CBS 513.88. Red boxes show up-regulation of genes in ATCC 1015 strain. Blue boxes demonstrate that putative iso-enzymes exist that are up-regulated in each of the two strains. White boxes/No box indicate no significant change between the strains. The plot is an adaptation of a figure from Andersen et al. (ANDERSEN *et al.* 2008a). Complete metabolic map is shown in supplementary Figure 5 of Manuscript 1.

When examining Figure 8 for entire up-regulated pathways in either strain, the biosynthetic pathways of threonine, serine, lysine and tryptophan were found up-regulated in the CBS 513.88 strain. A closer look using a codon usage analysis of the glucoamylase A gene (*glaA*) (details in manuscript 1), revealed that *glaA* is uncommon in that it has a higher percentage of tryptophan codons than 90% of all the predicted genes of *A. niger* CBS 513.88, and twice as high a tryptophan content as the average in the biomass composition measured by Christias et al. (CHRISTIAS *et al.* 1975). Many SNPs were found in the biosynthetic pathway of tryptophan, which was an unforeseen coincidence. Additionally, threonine and serine codons also have high levels in *glaA*. The combination of these results suggested that high-yield production of enzymes is highly dependent on corresponding increased production of amino acids that are over-represented in the product.

Other up-regulated pathways in CBS 513.88 strain included purine and pyrimidine biosynthesis and chitin biosynthesis. Very few metabolic pathways were generally up-regulated in ATCC 1015. The few that were up-regulated included most of the sterol biosynthesis pathway and the alternative oxidative pathway.

In order to identify significant trends in the transcriptome profile of the two strains, a GO term over-representation analysis was conducted with the up-regulated genes from each strain. The analysis showed that up-regulated genes in CBS 513.88 had a number of significant traits relevant to high protein production yield, specifically amino acid biosynthesis and tRNA aminoacylation activities (see Table 8).

**Table 8**. Summary of GO term over-representation analysis of the 1000 most up-regulated significantly differentially expressed genes found in batch cultivations of *A. niger* CBS 513.88. The top ten biological processes GO terms are shown.

| # | GO.ID | Term | Annotated | Significant | Expected |
|---|---|---|---|---|---|
| 1 | GO:0006418 | tRNA aminoacylation for protein translation | 50 | 21 | 5.38 |
| 2 | GO:0006526 | Arginine biosynthetic process | 8 | 6 | 0.86 |
| 3 | GO:0009082 | Branched chain family amino acid anabolism | 12 | 7 | 1.29 |
| 4 | GO:0007049 | Cell cycle | 34 | 12 | 3.66 |
| 5 | GO:0000059 | Protein import into nucleus, docking | 10 | 6 | 1.08 |
| 6 | GO:0009073 | Aromatic amino acid family biosynthetic process | 19 | 8 | 2.04 |
| 7 | GO:0016043 | Cell organization and biogenesis | 220 | 43 | 23.67 |
| 8 | GO:0009067 | Aspartate family amino acid biosynthetic process | 9 | 5 | 0.97 |
| 9 | GO:0050658 | RNA transport | 3 | 3 | 0.32 |
| 10 | GO:0015031 | Protein transport | 107 | 26 | 11.51 |

The same GO term over-representation study in ATCC 1015 strain did not propose terms of a similar trend. However, electron transport was the most significantly over-represented biological process GO term (see Table 9). Individual examination of the regulated genes showed the presence of an alternative oxidase and the mitochondrial proton gradient driven ATPase. Up-regulation of genes annotated to have functions in carbohydrate transport and transport of organic acids was also seen.

An examination of the individual regulated genes to find single genes of special interest, showed that regulation of glucoamylase A was significantly up-regulated in CBS 513.88, but the fold change was ~3 compared to the more than 6-fold higher glucoamylase activity measured in the enzymatic assays of these cultivations.

**Table 9**. Summary of GO term over-representation analysis including all the statistically significant up-regulated genes found in batch cultivations of *A. niger* ATCC 1015. The top ten biological processes GO terms are reported.

| # | GO.ID | Term | Annotated | Significant | Expected |
|---|---|---|---|---|---|
| 1 | GO:0006118 | Electron transport | 599 | 153 | 115.07 |
| 2 | GO:0006564 | L-serine biosynthetic process | 18 | 10 | 3.46 |
| 3 | GO:0051341 | Regulation of oxidoreductase activity | 34 | 15 | 6.53 |
| 4 | GO:0008643 | Carbohydrate transport | 101 | 30 | 19.4 |
| 5 | GO:0009410 | Response to xenobiotic stimulus | 22 | 9 | 4.23 |
| 6 | GO:0042221 | Response to chemical stimulus | 99 | 27 | 19.02 |
| 7 | GO:0006865 | Amino acid transport | 87 | 24 | 16.71 |
| 8 | GO:0015837 | Amine transport | 87 | 24 | 16.71 |
| 9 | GO:0015849 | Organic acid transport | 87 | 24 | 16.71 |
| 10 | GO:0046942 | Carboxylic acid transport | 87 | 24 | 16.71 |

Differences were seen not only at the transcriptome level, but also at the phenotypic level. Morphological differences were observed when growing the two *A. niger* strains on solid media plates. *A. niger* ATCC 1015 produced dark black conidiophores with long strings of connected spores, whereas *A. niger* CBS 513.88 produced conidiophores with less brown spores with a sectorial morphology as seen in Figure 9.

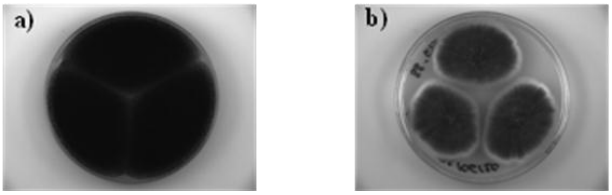



**Figure 9**. Morphological differences of *A. niger* strains a) ATCC 1015 and b) CBS 513.88 when grown on solid rich media. Plates were incubated for 6 days at 30 °C.

### 3.1.3 Gene expression mapping into the *A. niger* genome identifies secondary metabolite cluster activities and reveals a whole-arm inversion in chromosome VI

Differences in gene expression relative to chromosome positions between the two *A. niger* strains compared, ATCC 1015 and CBS 513.88, were also examined. The log2-ratios of the gene expression indices from the transcriptome analysis were mapped to the synteny maps constructed (details in results section of Manuscript 1). This allowed the identification of chromosome regions with a uniformly higher expression in one of the strains.

The parts of the genome assembly of ATCC 1015 that did not find hits in the CBS 513.88 sequence were of special interest, since uniformly higher expression in the ATCC 1015 strain of all genes in the entire region suggested that this region was not present in the CBS 513.88 strain. This was the case for six large regions, including the largest area in the genome not found in CBS 513.88 as explained in details in manuscript 1.

Two putative secondary metabolite clusters were found using the log2-ratios from the transcriptome analysis. The cluster on chromosome VIII (including a putative polyketide synthase gene, JGI211885), which is nearly identical to an *A. fumigatus* gene-AFUA_1G17740) appeared to be unique for the ATCC 1015 strain since this region was not found in the genome of the CBS 513.88 strain. The second cluster was present in both genomes and located close to the telomeric region on the right arm of chromosome I. This cluster was significantly up-regulated in the CBS 513.88 *A. niger* strain and contained a putative non-ribosomal peptide synthase (NRPS) (JGI43555) and a putative transcription factor (JGI188323).

The transcriptome mapping successfully confirmed the inversion of the entire right arm of chromosome VI (details in Manuscript 1). The telomeric position effect has been described earlier in *S. cerevisiae* by Gottschling et al. (GOTTSCHLING *et al.* 1990), and the same

philosophy was applied to find out the orientation of chromosome arms in *A. niger* CBS 513.88. Thus, if an arm has been inverted, reduced expression should be found at opposite ends in ATCC 1015 and CBS 513.88, thereby influencing the log-2 ratios. This was seen on the right arm of chromosome VI with the numerical log2-ratios increasing towards the two ends.

### 3.2 Insights into glycerol metabolism and its regulation in *Aspergillus*

The findings reported in this section are the foundation of manuscript 2: Uncovering transcriptional regulation of glycerol metabolism in Aspergilli through genome-wide gene expression data analysis. The driving force for conducting this study was mainly due to the fact that glycerol has become of considerable importance in industrial fermentation processes as being a major by-product from biodiesel production. Thereby, it represents a cheap carbon source for bio-based production of chemicals. Glycerol is a non-fermentable carbon source that can be used by many yeast species, including *S. cerevisiae* and filamentous fungi such as *A. nidulans*, *A. oryzae*, and *A. niger*. For this reason, we considered that it would be valuable to identify regulatory nodes that control glycerol consumption in industrial relevant Aspergilli in order to convert this by-product into other chemicals or proteins of higher added value.

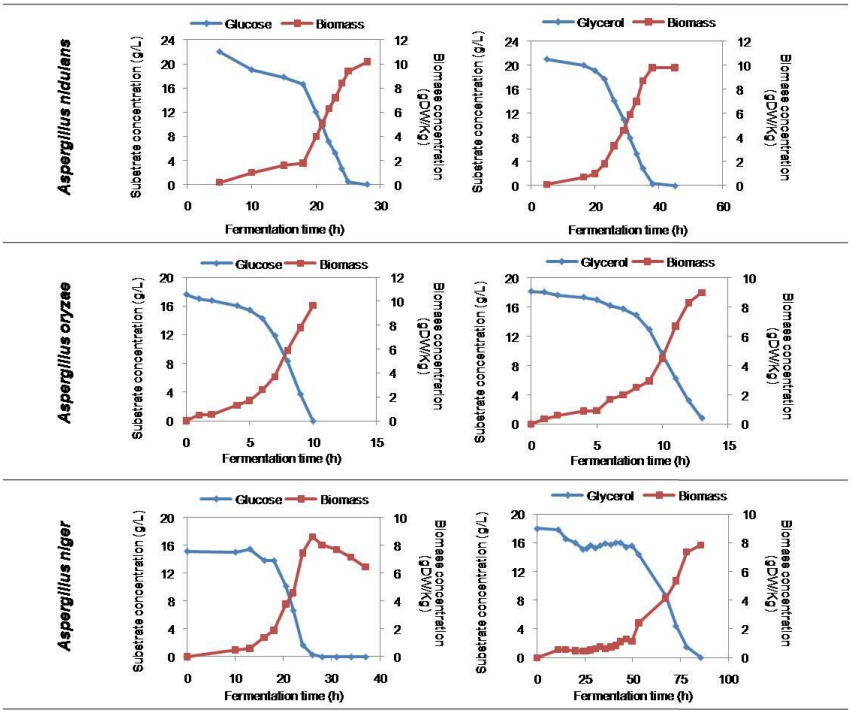
In the yeast *S. cerevisiae* and Aspergilli, glycerol degradation occurs via a two steps glycerol phosphorylative pathway. In the first step, glycerol is converted to glycerol-3-phosphate by glycerol kinase (EC 2.7.1.30), product of the gene *GUT1* in *S. cerevisiae*. Then, glycerol-3-phosphate crosses the outer mitochondrial membrane, where it is oxidized to glycerone phosphate by the inner mitochondrial membrane enzyme, $FAD^+$-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5), which is encoded by the *S. cerevisiae* gene *GUT2* (DAVID *et al.* 2006; RONNOW and KIELLAND-BRANDT 1993). Finally, glycerone phosphate enters the cytosol, where it is used either in the glycolytic or in the gluconeogenic pathways.

To identify the transcriptional regulation of glycerol metabolism in *Aspergillus*, we analyzed data from triplicate batch fermentations of three different Aspergilli (*Aspergillus nidulans*, *Aspergillus oryzae* and *Aspergillus niger*) with glucose and glycerol as carbon sources and the main facts are reported and discussed in the following subsections.

#### 3.2.1 Fermentation results using three different *Aspergillus* species with glucose or glycerol as carbon source

To have a complete dataset of fermentation and transcriptome data in all three *Aspergillus* species, namely, *A. oryzae*, *A. niger* and *A. nidulans*; we collected fermentation data from our previous work and conducted new fermentation experiments (details in manuscript 2). Fermentations on glycerol with *A. oryzae* and *A. niger* were conducted specifically for the study. Each *Aspergillus* species had its own specific cultivation medium and all fermentations were run in three biological replicates. A summary of all fermentation results are shown in Figure 10, where panel a) Shows the substrate and biomass concentration profiles for each Aspergilli, and panel b) Shows the statistics of the physiological characterization data. In the case of *A. nidulans* and *A. niger*, the growth rates differences on the two carbon sources were more prominent. For *A. nidulans*, the maximum specific growth rate on glucose was double that on glycerol; and for *A. niger*, growth on glycerol was four times slower when compared to glucose (see Figure 10 b). The difference on growth rates in the three *Aspergillus* spp. might be due to several reasons, but we hypothesize that it is very likely that glycerol is a favorite carbon source for the *A. oryzae* strain A1560 used in this study (VONGSANGNAK *et al.* 2008). Besides the cultivation conditions, have been optimized for growth on glycerol (Novozymes's fermentation conditions). Culture samples were harvested and processed further for genome-wide gene expression analysis.

**a)**



**b)**

| Strain | Carbon source | $\mu_{max}$ ($h^{-1}$) | $Y_{sx}$ (g DW/g substrate) | Time of sampling (h) | Biomass concentration (g DW/Kg) |
|---|---|---|---|---|---|
| ***A. nidulans*** | Glucose | 0.23±0.020 | 0.47±nd | ~ 22 | 6.33±0.40 |
| | Glycerol | 0.11±0.010 | 0.42±nd | ns | 6.50±0.50 |
| ***A. oryzae*** | Glucose | 0.38±0.004 | 0.54±0.013 | ~6 | 2.50±0.09 |
| | Glycerol | 0.30±0.004 | 0.52±0.008 | ~8 | 2.44±0.05 |
| ***A. niger*** | Glucose | 0.22±0.015 | 0.57±0.053 | ~21 | 3.74±0.06 |
| | Glycerol | 0.05±0.007 | 0.40±0.022 | ~36 | 0.88±0.29 |

**Figure 10**. Summary of batch fermentation parameters of *A. nidulans*, *A. oryzae* and *A. niger* grown on glucose or glycerol as sole carbon source. a) Fermentation profiles of a representative replicate. (♦): Substrate concentration (g/L). (●): Biomass concentration (g DW/Kg). All fermentations were performed in three biological replicates. b) Summary of batch cultivations statistics. For all cultivations, maximum specific growth rate ($\mu_{max}$), biomass yield ($Y_{sx}$), time of sampling for transcriptome analysis (TA), and biomass concentration at the time of sampling for transcriptome analysis (TA) are given. ns: not specified.

### 3.2.2 Protein comparison and transcriptome analysis

Protein comparisons and cross analysis with gene expression data of all three species resulted in the identification of 88 genes having a conserved response across *A. nidulans*, *A. oryzae* and *A. niger*. In order to reach this result, first, genes having orthologues in the three species were identified using a BLASTP based comparison (ALTSCHUL *et al.* 1990a). Genome-wide protein sequences from the three Aspergilli were compared among each other in order to obtain tri-directional orthologues. By defining a threshold of E-value of 1E-30, 5,190 orthologues were found to be conserved in all three *Aspergillus* species producing the list of 1:1:1 orthologues shown in Supplementary Table 1 (Appendix). This set of conserved genes (1:1:1 orthologues) was used for further analysis.

Then, a t-test pair-wise comparison for each *Aspergillus* spp. on glycerol versus glucose identified 904, 1,145 and 3,058 genes as significantly differentially expressed for *A. nidulans*, *A. oryzae* and *A. niger*, respectively.

Subsequently, these three subsets of significant genes in all three species were cross compared to the list of 5,190 conserved genes in the three Aspergilli as well as with each other. This

resulted in the identification of 88 conserved genes that were differentially expressed in all three species (Figure 11). Among them, 81 genes were up-regulated during growth on glycerol, 5 genes were down-regulated and 2 genes did not show a clear trend (details in manuscript 2).

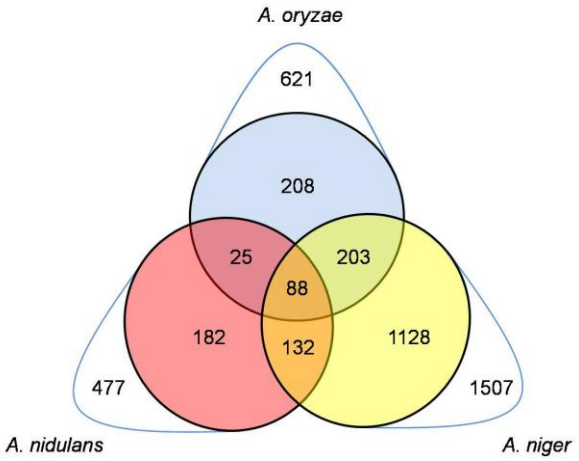



**Figure 11**. Venn diagram of significantly differentially expressed genes from glycerol versus glucose by pair-wise comparison for each *Aspergillus* species. The colored overlapping middle area contains the genes that are significantly differentially expressed and conserved in all three *Aspergillus* species. The numbers on a white background represent the non conserved genes in all three Aspergilli, but still differentially expressed in a single species. Adjusted p value cut-off < 0.05.

### 3.2.3 Detection of conserved motifs

One or more conserved transcriptional regulators were suspected to be up-regulating the subset of 81 genes or down-regulating the subset of 5 down-regulated genes within the group of 88 genes having a conserved transcriptional response. Therefore, statistical promoter analysis was conducted for all three data sets of 81 up-regulated genes on glycerol medium. By inspecting the upstream sequences of each *Aspergillus* up-regulated orthologues dataset, accounting for 243 promoters in total (3 X 81 promoters), we found the motif "TGCGGGGA" as the most over-represented pattern. The corresponding logo plot is shown in Figure 12. The same analysis was conducted with the subset of down-regulated genes, but no consensus *cis*-acting regulatory element was found. Based on a literature search, we proposed that TGCGGGGA is the consensus binding sequence of the transcriptional activator Adr1, which

regulates several pathways in *S. cerevisiae* (YOUNG *et al.* 2003) and in humans (DAS and BAEZ 2008).

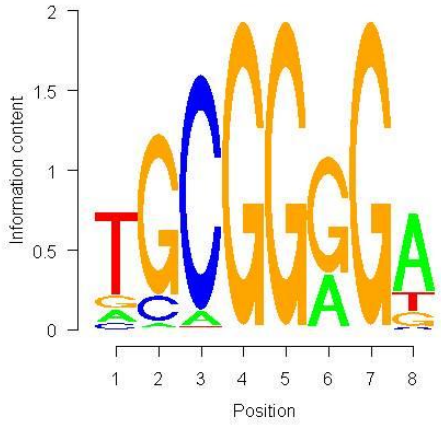



**Figure 12**. Logo plot of the over-represented motif from the 81 promoter regions of *A. nidulans*, *A. oryzae* and *A. niger* genes significantly up-regulated on glycerol medium. The nucleotides representing the sequence are stacked on top of each other for each position in the aligned sequences. The height of each nucleotide is made proportional to its frequency, and the nucleotides are sorted so that the most common is on top. The height of the entire stack is then adjusted to signify the information content of the sequences at that position (SCHNEIDER and STEPHENS 1990). The "x axis" indicates the position of the corresponding nucleotide (A, T, C or G). "y axis" represents the information content of the corresponding nucleotide at each position in a bits scale, where 2 is the maximum value.

The consensus binding sequence of Adr1 in humans is "GCGGGGA", and regulates the transcription of *psen1* (gene encoding presenilin 1) (DAS and BAEZ 2008), a transmembrane protein that functions as part of the gamma-secretase protease complex. In *S. cerevisiae*, Adr1 is known to regulate several pathways including glycerol metabolism and fatty acid metabolism (YOUNG *et al.* 2003). The consensus binding sequence in *S. cerevisiae* is TTGG(A/G)GA, and according to Cheng et al. (CHENG *et al.* 1994), only four base pairs are essential: GG(A/G)G. From the subset of 81 up-regulated genes, 24 genes had the motif TGCGGGGA in all three *Aspergillus* species and 30 of the total of 72 genes showed it more than once in their promoters. In general, the motif was located at an average position of 650 bps upstream from the start codon. From this subset of 24 up-regulated genes, 5 of them had orthologues in *S. cerevisiae*; *ADH2* (YMR303C/*alcB* in *Aspergillus)*, *ALD5* (YER073W/*aldA*

in *Aspergillus)*, *ACS2* (YLR153C), *CCC1* (YLR220W) and *PUT2* (YHR037W). The location of the Adr1 promoter binding sites are summarized in Supplementary Table 5 of manuscript 2. This conserved up-regulatory response in all three *Aspergillus* spp. suggested that these genes could be activated by a common cross species conserved transcription factor, which could be the regulatory protein Adr1 as it occurs in *S. cerevisiae*, responsible for regulation of *ADH2* and *ALD5* (YOUNG *et al.* 2003) (details in Table 1 of manuscript 2).

The transcriptome analysis indicated that genes involved in ethanol, glycerol, fatty acid, amino acids and formate utilization were putatively regulated by Adr1 in Aspergilli as they are in *S. cerevisiae* and this transcription factor, therefore is likely to be cross species conserved among *Saccharomyces* and distant *Ascomycetes*. Transcriptome data was further used to evaluate the high osmolarity glycerol (HOG) pathway. All the components of this pathway present in yeast have orthologues in the three Aspergilli studied and its gene expression response suggested that this pathway functions as in *S. cerevisiae* (details in manuscript 2).

### 3.2.4 Glycerol utilization in *A. nidulans*, *A. oryzae* and *A. niger*

A closer look at the transcriptome results showed differences on the preference of glycerol utilization pathways in each Aspergilli studied. In naturally glycerol utilizing fungi, glycerol (GL) can be phosphorylated either into glycerol 3-phosphate (GL3P) and further oxidized by the $FAD^+$-dependent glycerol-3-phosphate dehydrogenase into glycerone phosphate (T3P2), which then enters glycolysis. In the other pathway, glycerol can be converted through $NAD^+$/$NADP^+$ glycerol dehydrogenases into glycerone (GLYN) and further phosphorylated by glycerone kinase into T3P2. A simplified scheme of the metabolic pathways leading to or from glycerol is illustrated in Figure 13.

It is likely that both pathways leading to the glycolytic intermediate T3P2 are involved in glycerol utilization in *Aspergillus* species. Nevertheless, according to our transcriptome data, the most active pathway in *A. oryzae* is probably the one using glycerol dehydrogenase and glycerone kinase to produce glycerone phosphate. In contrast, in *A. niger* and *A. nidulans*, the pathway using glycerol kinase and the $FAD^+$ dependent glycerol-3-phosphate dehydrogenase is most likely to be the dominant (see Figure 13).

Another study in *A. niger* supporting our findings has shown that glycerol accumulated in a glycerol kinase mutant (WITTEVEEN and VISSER 1995), which was able to synthesize glycerol,

but not able to catabolize it, suggesting that the activity of this pathway is important for glycerol catabolism.

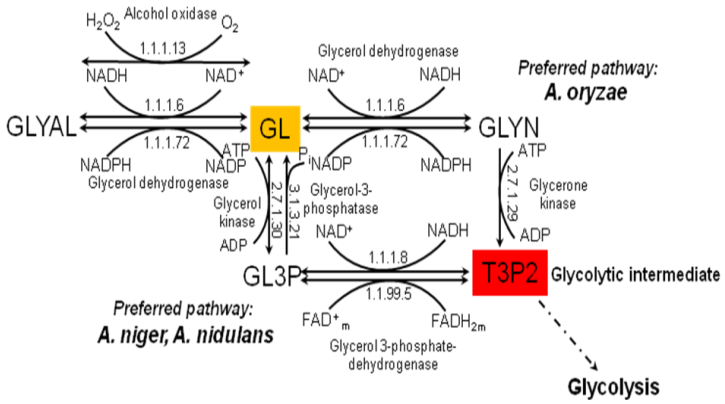



**Figure 13**. Glycerol utilization pathways in *Aspergillus* species leading to the production of the glycolytic intermediate glycerone phosphate. The abbreviation of metabolites is described as follows. GL, glycerol; GLYAL, D-glyceraldehyde; GLYN, glycerone; GL3P, sn-glycerol 3-phosphate; T3P2, glycerone phosphate.

### 3.3 Transcriptional regulation of maltose uptake in *A. oryzae* and *A. niger*

The findings reported in this section are the base of manuscript 3: Genome-wide analysis of maltose utilization and regulation in aspergilli. In this case, the motivation for conducting the study was mainly because maltose is one of the most effective inducers used for enzyme production in aspergilli such as for α-amylase production by *A. oryzae* (CARLSEN and NIELSEN 2001b; CARLSEN *et al.* 1996), but also for glucoamylase production in *A. niger* and *Aspergillus nidulans* (KATO *et al.* 2002; LARRY *et al.* 1972). Compared to the yeast *Saccharomyces cerevisiae* little is known about maltose utilization, transport and regulation at the molecular level in aspergilli. Thereby, we propose that digging into the mechanisms and its effectors can help on boosting enzyme production at higher yields.

#### 3.3.1 Maltose utilization

Maltose utilization in *S. cerevisiae* is under control of three general regulatory mechanisms: induction, glucose repression and glucose inactivation (NOVAK *et al.* 2004). The presence of maltose in the environment is necessary for induction of synthesis of maltase and maltose

transporter. The metabolism and regulation of maltose requires the presence of *MAL* regulon. There are several *MAL* regulons identified in different strains of *S. cerevisiae*, but the *MAL6* locus is the most well studied (KLEIN *et al.* 1996). The gene structure of the *MAL6* locus is composed of a cluster of three genes: *MAL61* (*MALT*) encoding maltose permease, *MAL62* (*MALS*) encoding maltase (EC: 3.2.1.20) and *MAL63* (*MALR*), encoding a transcriptional activator specifically activating expression of the *MALT* and *MALS* genes (NEEDLEMAN *et al.* 1984). As many other processes, expression of both *MALT* and *MALS* is carbon catabolite repressed by glucose through the transcription factor Mig1 and coordinately induced by maltose (KLEIN *et al.* 1996).

### 3.3.2 Identification of MAL regulon in Aspergilli

Using the gene structure of *S. cerevisiae MAL*6 locus as a model, we aimed at identification of the *MAL* gene cluster in ten different sequenced *Aspergillus* genomes by using BLASTP (ALTSCHUL *et al.* 1990a) (see Methods from manuscript 3 for details). Five different *Aspergillus* species, specifically, *A. oryzae*, two strains of *A. fumigatus*, *A. flavus*, *A. clavatus*, and *A. fischeri* showed at least one *MAL* gene cluster in their genomes as illustrated in Figure 14.

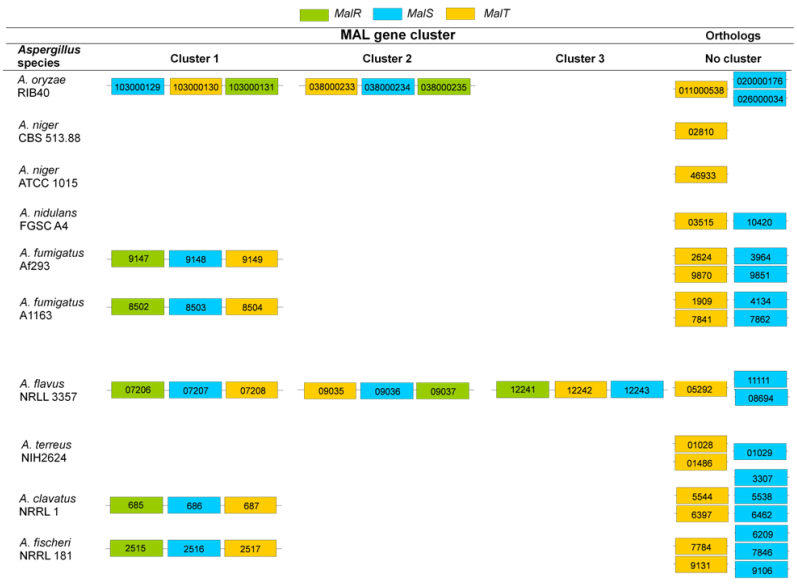



**Figure 14**. The diagram shows the comparative sequence analysis of *MAL* gene cluster between *S. cerevisiae* and 10 different *Aspergillus* species. Values in each rectangle represent the shortened ORF name. For individual full name, the shorten ORF is prefixed by "AO090"

for *A. oryzae* RIB40, "An02g" for *A. niger* CBS 513.88, "JGI" for *A. niger* ATCC 1015, "ANID_" for *A. nidulans* FGSC A4, "ORF" for *A. fumigatus* Af293 and A1163, *A. clavatus* NRRL 1 as well as *A. fischeri* NRRL 181, "AFL2T_" for *A. flavus* NRRL 3357, "ATET_" for *A. terreus* NIH2624. Complete details are shown in Supplementary File 1 of manuscript 3.

*A. oryzae* and *A. flavus* contained at least two *MAL* gene clusters. Phylogenetic analysis suggested that events of gene duplication and horizontal gene transfer may have occurred in these strains (see Supplementary file 1 from manuscript 3). In contrast, no *MAL* cluster was found in *A. nidulans*, *A. terreus* and two strains of *A. niger* under the statistical constraints imposed. Therefore, it is most likely that these four *Aspergillus* strains do not have the *MAL* regulon for maltose utilization. In all the sequenced *Aspergillus* genomes, we were able to identify multiple orthologues encoding maltase or α-glucosidase enzymes and maltose transporters as shown in Figure 14, but not found in a gene cluster.

### 3.3.3 Physiological characterization of *A. oryzae* and *A. niger* in batch fermentations

The strains used in these experiments were *A. oryzae* wild type strain A1560 (an ancestor of strains used for commercial α-amylase production) and *A. niger* wild type strain BO1 (an ancestor of strains used for commercial glucoamylase production). Both strains were obtained from Novozymes (CARLSEN and NIELSEN 2001b; PEDERSEN *et al.* 2000a).

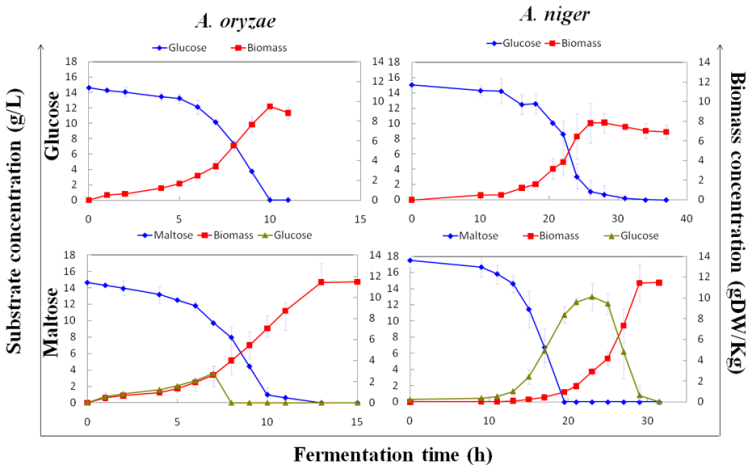



**Figure 15**. Biomass and substrate concentration profiles of *A. oryzae* and *A. niger* with glucose or maltose as carbon sources in batch fermentations.

To evaluate their physiology and harvest samples for global regulatory response analysis when adjusting metabolism to the use of glucose or maltose, we grew the two *Aspergillus* species in well-controlled bioreactors to perform reproducible batch fermentations. The cultivations were carried out in three biological replicates on defined minimal medium (ANDERSEN *et al.* 2008b; PEDERSEN *et al.* 2000a) (see Methods from manuscript 3 for details). Biomass growth and substrate concentration profiles are shown in Figure 15. In comparison to *A. niger*, *A. oryzae* grew faster than *A. niger* in the two carbon sources used. Glucose was exhausted in 10 h and maltose in 12 h, at rates of 3.09±0.02 g $L^{-1}.h^{-1}$ and 2.46±0.02 g $L^{-1}.h^{-1}$, respectively. A summary of typical fermentation characterization parameters are shown in Table 10.

**Table 10**. Physiological characterization data. For all batch cultivations, maximum specific growth rate ($\mu_{max}$), biomass yield ($Y_{sx}$), sampling time for transcriptome analysis (TA), and biomass concentration at the time of sampling for transcriptome analysis (TA), are given. Average values and standard deviations are reported.

| Strain | Carbon source | $\mu_{max}$ ($h^{-1}$) | $Y_{sx}$ (g DW/g Csource) | Sampling time for TA (h) | Biomass for TA (g DW/Kg) |
|---|---|---|---|---|---|
| ***A. oryzae*** | Glucose | 0.38±0.01 | 0.51±0.01 | 6±0 | 2.50±0.09 |
| | Maltose | 0.32±0.05 | 0.49±0.05* | 7±0 | 2.27±0.09 |
| ***A. niger*** | Glucose | 0.22±0.01 | 0.57±0.05 | 21±1 | 3.74±0.06 |
| | Maltose | 0.31±0.02 | 0.62±0.02* | 24±0 | 3.55±0.51 |

*Biomass yield was calculated based on glucose (g DW/g glucose)

In *A. oryzae*, as shown in Figure 15, glucose modestly accumulated in the media during growth on maltose. In contrast, with *A. niger*, glucose accumulated on maltose cultivations, probably due to a very high extracellular glucosidase activity expressed by *A. niger*, which allowed the fungus to grow very fast on this carbon source. Besides growth rates and biomass yields of *A. oryzae* and *A. niger*, sampling times and biomass yields when sampling biomass for gene expression analysis were reported (see Table 10).

### 3.3.4 Transcriptome analysis of maltose metabolism

We validated the presence or absence of the *MAL* gene cluster in *A. oryzae* and *A. niger* by using our custom designed Affymetrix GeneChip for genome-wide expression analysis (ANDERSEN *et al.* 2008b). To do this, we performed Student's t-test pair-wise transcriptome

comparisons of glucose and maltose to examine expression of putative *MAL* gene clusters in *A. oryzae* and in *A. niger* (for complete details see manuscript 3). Solely 16 genes were significantly differentially expressed in *A. oryzae*. While, in contrast, for *A. niger*, no genes were statistically differentially expressed. Among the significant up-regulated genes, there were genes encoding maltase (AO090103000129 and AO090038000234) and maltose permease (AO090103000130 and AO090038000233) in *A. oryzae*, which are the functionally related orthologous genes of *MALS* and *MALT*, respectively in *S. cerevisiae* (see Figure 14 from section 3.3.2). The two *A. oryzae* genes orthologous to the *S. cerevisiae MALR* transcription factor, AO090103000131 and AO090038000235, were also up-regulated, but not statistically significant. Notably, we found a CreA binding site in the upstream region of *MALR* in *A. oryzae* (see details in results section of manuscript 3). We therefore proposed that the mechanism behind the *MAL* regulon in *A. oryzae* is similar to *MALR* function in *S. cerevisiae*, where it is activated by maltose and repressed by glucose (KLEIN *et al.* 1996).

*A. oryzae* has two *MAL* regulons and each regulon contains one *MALR* transcriptional activator (e.g. AO090103000131 or AO090038000235). *MALR* regulator induces maltose permeases (*MALT*) to transport extracellular maltose into the cell. *MALR* also induces maltase (*MALS*) that hydrolyzes intracellular maltose into glucose which is then channeled through glycolysis. Figure 16 a) illustrates the proposed mechanisms for regulation of maltose utilization in *A. oryzae*. In contrast, no *MAL* gene cluster was identified in *A. niger*, at least no closely homologous to the one existing in *S. cerevisiae* (see Figure 14). Maltose utilization in *A. niger* do not involve a *MAL* regulon, but occurs through another regulatory system, e.g., via AmyR regulator for glucoamylase (*glaA*) and/or induction of other glucosidases. Further supporting our findings, Yuan and coworkers (YUAN *et al.* 2008c) using the wild type strain N402 (ATCC 9029) did not identify up-regulated components of the *MAL* regulon when comparing gene expression data on xylose versus maltose cultivations. Based on findings in the literature (YUAN *et al.* 2008c) and our findings, we suggest that *A. niger* utilize maltose by means of extracellular hydrolysis by glucoside hydrolases such as glucoamylase followed by glucose uptake and metabolism. Figure 16 b) summarizes the mechanism for regulation of maltose utilization in *A. niger*.

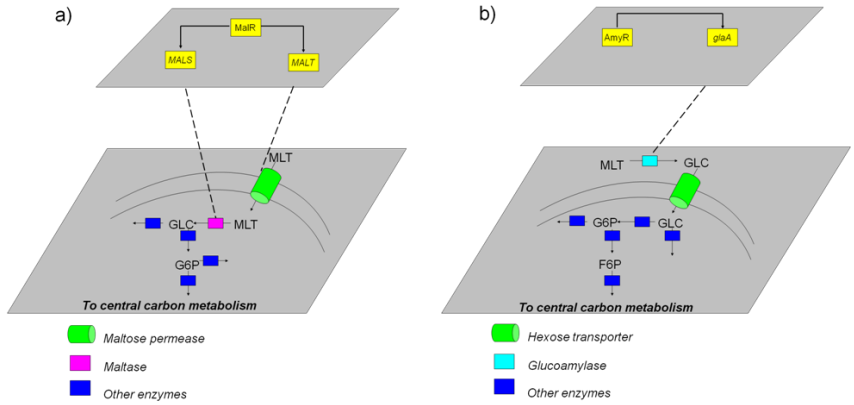



**Figure 16**. Mechanism of maltose utilization and regulation in a) *A. oryzae* and b) *A. niger*.

### 3.4 Strain construction

In order to investigate the role of several transcription factors involved in different regulatory pathways, gene knock-out strains were constructed. The *A. niger* ATCC 1015 strain was chosen as genetic background to conduct the genetic manipulations. The constructed gene deletion strains were further physiologically characterized for growth on solid media plates using different carbon sources as well as in submerged batch cultivations. Samples for transcriptome analysis were taken in the mid-exponential phase of growth and analyzed accordingly to the methods described. The gene deletion strains were used for conducting the studies reported in manuscripts 4 and 5.

#### 3.4.1 AdrA deletion

In an attempt to identify the existence of Adr1, AdrA in *Aspergillus* species, i.e., in *A. niger* and to understand its regulatory effect, an *A. niger adrA* deletion strain was constructed. According to our results from transcriptome studies (manuscript 2), Adr1/AdrA regulates pathways such as glycerol and fatty acid metabolism, amino acid metabolism, ribosome biogenesis and peroxisomal biogenesis. Therefore, deletion of this putative transcription factor would result in down-regulation of genes involved in those pathways where it functions as an activator. The method used for deleting the *ADR1 S. cerevisiae* orthologue in *A. niger*, JGI210333, is described in Materials and Methods.

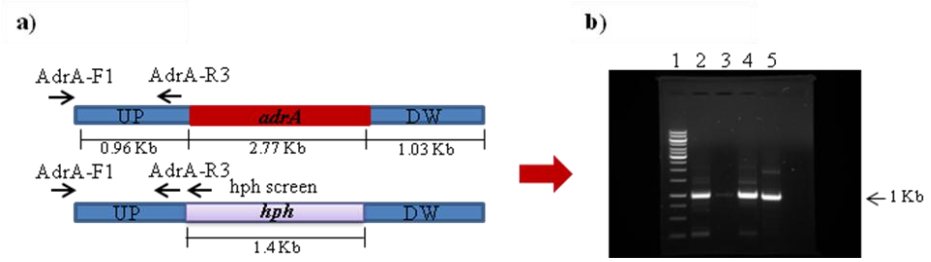



**Figure 17**. Schematic representation of the deletion of *A. niger* AdrA orthologue. On the left hand side, panel a) Genomic region of *adrA* locus (JGI210333), panel b) Gel electrophoresis. Lanes are as follows: 1 Kb DNA ladder (lane 1), PCR product from genomic DNA of wild type strain ATCC 1015 amplified with primers AdrA-F1 and AdrA-R3 (lane 2), PCR product from genomic DNA of wild type strain ATCC 1015 amplified with primers AdrA-F1 and *hph* screen primer (lane 3), PCR product from genomic DNA of *adrA* gene deletion mutant amplified with primers AdrA-F1 and AdrA-R3 (lane 4) and PCR product from genomic DNA of *adrA* gene deletion mutant amplified with primers AdrA-F1 and *hph* screen primer (lane 5). 1 kb DNA GeneRuler (Fermentas AB, Sweden) was used. Gel electrophoresis was carried out with 1% agarose gel run for 1 h at 70 V. Fragment sizes are included in Kb.

The transformation efficiency in *A. niger* is fairly low (MEYER *et al.* 2007); therefore, a vast number of transformants were screened by PCR using the primers AdrA-F1 and *hph* screen, and AdrA-F1 and AdrA-R3 as control reaction (for primers sequences see Table 5). An example of PCR screening results is shown in Figure 17. According to the expected band sizes, the transformant obtained is a gene deletion mutant, where the gene deletion cassette containing the gene conferring hygromycin resistance (*hph*) has been integrated in the targeted locus as shown in lane 5 of Figure 17.

### 3.4.2 CreA deletion

CreA is a transcription factor and one of the main components if not, the most important protein, involved in the glucose repression signaling cascade. It is the orthologue of Mig1, known to be regulating carbon catabolite repression in *S. cerevisiae* (GANCEDO 1998).

Previous results from physiological studies as well as transcriptome studies have shown that CreA regulates a number of pathways. According to Mogensen et al. (MOGENSEN *et al.* 2006a), the regulated genes can be classified in three main groups, namely: group 1: genes used for the metabolism of less favourable carbon sources, group 2: genes encoding

gluconeogenic and glyoxylate cycle enzymes and group 3: genes involved in secondary metabolism. Deletion of this key transcription factor would result in up-regulation of genes involved in those regulated pathways where it functions as a repressor. In order to study the effects of deleting this transcription factor over glucose repression orthologue components, and to understand its regulatory effect in a genome-wide way, an *A. niger* CreA/JGI206681, knock-out strain was constructed. As expected in filamentous fungi, a large number of transformants was screened by PCR in order to find potential transformants (see Table 5 for primer sequences). Results from PCR screening of transformant 3.2 are shown in Figure 18. This strain showed the expected PCR products band sizes after integration of the deletion cassette in the targeted locus (see lane 5 from Figure 18) and therefore, was used in further experiments.

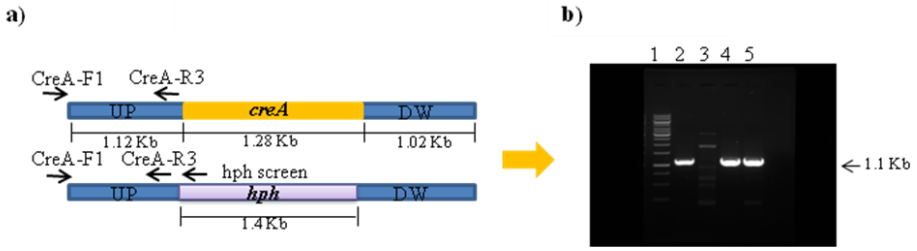



**Figure 18**. Schematic representation of the deletion of CreA in *A. niger* ATCC 1015. Panel a) Genomic region of *creA* locus (JGI206681). Panel b) Gel electrophoresis. Lanes are as follows: 1 Kb DNA ladder (lane 1), PCR product from genomic DNA of wild type strain ATCC 1015 amplified with primers CreA-F1 and CreA-R3 (lane 2), PCR product from genomic DNA of wild type strain ATCC 1015 amplified with primers CreA-F1 and *hph* screen primer (lane 3), PCR product from genomic DNA of *creA* gene deletion mutant amplified with primers CreA-F1 and CreA-R3 (lane 4) and PCR product from genomic DNA of *creA* gene deletion mutant amplified with primers CreA-F1 and *hph* screen primer (lane 5). Gel electrophoresis was conducted under the same conditions as above (section 3.4.1).

### 3.4.3 FacB deletion

In *S. cerevisiae*, Cat8 is a transcription factor involved in the regulation of several genes and activated during the diauxic shift, which is the transition between fermentative and non-fermentative metabolism. Cat8 is the orthologue of FacB, known to regulate acetamide and acetate metabolism in *A. nidulans* (TODD *et al.* 1997). Genomic studies have shown that least 30 genes, encoding proteins involved in gluconeogenesis, ethanol utilization and the

glyoxylate cycle are regulated by Cat8 (TACHIBANA *et al.* 2005). In addition, Cat8 together with the transcription factor Adr1 co-regulate a number of genes in *S. cerevisiae* (YOUNG *et al.* 2003).

In an attempt to understand the regulatory role of FacB in *A. niger*, besides regulation of acetate metabolism, an *A. niger facB* deletion strain was constructed. Transformant 1.2 was a successful gene deletion strain, as demonstrated by PCR screening results shown in Figure 19, and was used in subsequent experiments. Growth assays confirmed that *facB* is required for growth on acetate as a sole carbon source as reported in the following sections and in manuscript 4.

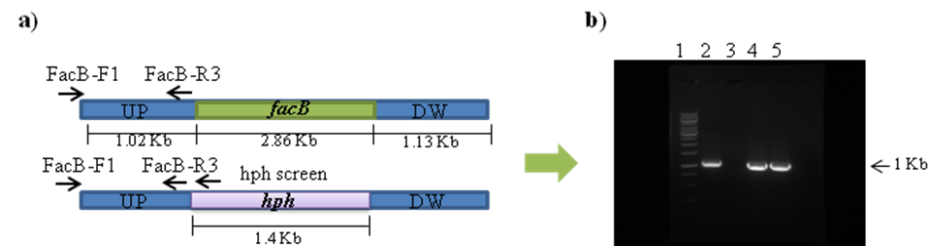



**Figure 19**. Schematic representation of the deletion of FacB in *A. niger* ATCC 1015. Panel a) Genomic region of *facB* locus (JGI139020). Panel b) Gel electrophoresis. Similar gel electrophoresis was run for *facB* gene deletion strain using the corresponding primer pair combination. Lane 1 corresponds to the 1 Kb DNA ladder used and lanes 2 and 3 for PCR products amplified with genomic DNA from ATCC 1015 wild type strain. Lanes 4 and 5 corresponds to PCR products amplified with genomic DNA of *facB* gene deletion mutant. Gel electrophoresis was run under the same conditions as above.

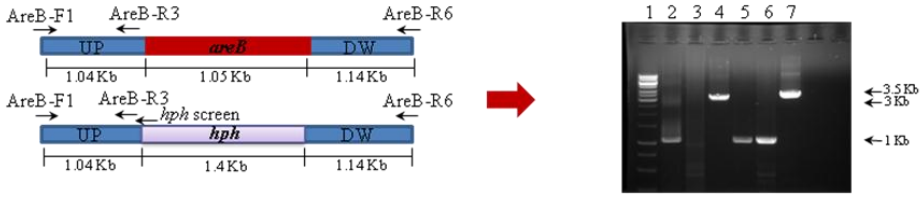
### 3.4.4 AreB deletion

In order to study the regulatory role of AreB in *A. niger*, an *areB* deletion mutant was constructed. In *A. nidulans*, genetic analysis and physiological studies have shown that AreB is a GATA type transcriptional regulator involved in regulation of nitrogen metabolism in concert with AreA (CONLON *et al.* 2001). Similarly, in *S. cerevisiae*, nitrogen metabolism is regulated by GATA transcription factors (COFFMAN *et al.* 1997). In this case, there are two negative-acting GATA factors, Dal80 and Gzf3, which function requires a repressive carbon source. However, mutations in those two genes showed only minor effects at the level of growth phenotype (CONLON *et al.* 2001).

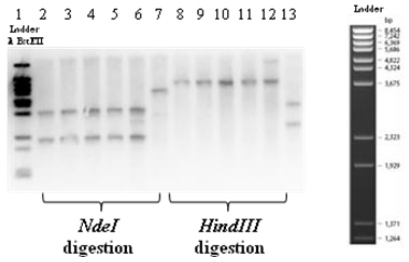
We suspected that deletion of this transcription factor in *A. niger* would result in up-regulation of genes involved in those pathways where it functions as a repressor or show a growth defect.

Transformants were first screened by PCR (see Table 5 for primers sequences) and when candidates were found (Figure 20 a), Southern blot hybridization was carried out to confirm the correct deletion of the gene (Figure 20 b). As a typical PCR screening example, the results from colony 12.1 are shown in Figure 20 a).

**a)**




**b)**




| Restriction enzyme | Wild type (band size) | Knock-out (band size) |
|---|---|---|
| NdeI | 5.63 Kb | 3.73 Kb |
| HindIII | 4 Kb, 2.7 Kb | 6.68 Kb |

**Figure 20**. Schematic representation of *areB* gene deletion in *A. niger* ATCC 1015. a) Targeted *areB* (JGI36739) locus and PCR screening results. Gel electrophoresis lanes are as follow: *Lane 1*: 1 Kb DNA ladder; *Lane 2*: PCR product from genomic DNA of wild type strain ATCC 1015 amplified with primers AreB-F1 and AreB-R3; *Lane 3*: PCR product from genomic DNA of wild type strain ATCC 1015 amplified with primers AreB-F1 and *hph* screen primer; *Lane 4*: PCR product from genomic DNA of wild type strain ATCC 1015 amplified with primers AreB-F1 and AreB-R6; *Lane 5*:PCR product from genomic DNA of *areB* gene deletion mutant amplified with primers AreB-F1 and AreB-R3; *Lane 6*: PCR product from genomic DNA of *areB* gene deletion mutant amplified with primers AreB-F1 and *hph* screen primer and *Lane 7*: PCR product from genomic DNA of *areB* gene deletion mutant amplified with primers AreB-F1 and AreB-R6. Gel electrophoresis was carried out as above. b) Southern blot hybridization. Gel electrophoresis lanes are described as follow: *Lane 1*: λ DNA-BstEII digest ladder; *Lanes 2 to 6*: genomic DNA from transformants 10.1, 10.2,

10.3, 12.1 and 12.2, respectively after NdeI restriction enzyme (RE) digestion; *Lane 7*: genomic DNA from *A. niger* ATCC 1015 strain after NdeI RE digestion. *Lanes 8 to 12*: genomic DNA from transformant 10.1, 10.2, 10.3, 12.1 and 12.2, respectively after HindIII RE digestion; *Lane 13*: genomic DNA from *A. niger* ATCC 1015 strain after HindIII RE digestion. The band sizes of the ladder used and expected fragment sizes after genomic DNA RE digestion using the enzymes NdeI and HindIII are included.

Southern blot results with genomic DNA from five different transformants digested with the restriction enzymes NdeI, HindIII or SacII are shown (Figure 20 b). All of them, transformants 10.1, 10.2, 10.3, 12.1 and 12.2, showed the expected band sizes after RE digestion using as probe the full knock-out construct when correct integration of the deletion cassette at the targeted locus has occurred.

### 3.4.5 PyrG deletion

The fact that many transformation experiments were carried out in order to get the knock-out strains, and only a few genes were successfully deleted in the targeted locus, indicates the difficulty of conducting gene targeting with *A. niger* due to many reasons, especially the existence of the non-homologous end joining pathway (MEYER 2008). Efforts towards engineering this pathway have been made in other *Aspergillus* species, i.e., *A. oryzae* and *A. sojae* (TAKAHASHI *et al.* 2006) as well as in other *A. niger* strains such as the *A. niger* N4O2 strain (MEYER *et al.* 2007). We aimed at blocking this pathway by disrupting the *kusA* coding gene in the *A. niger* ATCC 1015 strain. In order to do this, we first constructed a *pyrG* disrupted strain. The transformant was unable to grow without the addition of the growth requirements uridine and uracil. Confirmation of the *pyrG* gene disruption was done by amplifying the targeted *pyrG* locus using the primers PG-F1 and PG-R5 producing a 2 kb fragment. The PCR product was purified and digested with the restriction enzymes XbaI or KpnI, respectively. KpnI cuts inside the amplified fragment from the wild type strain and XbaI cuts only if the *pyrG* gene has been successfully disrupted due to the design of the primers, where an XbaI restriction site was included (details in Figure 21).

The transformant was checked for correct genetic behavior by doing a complementation assay. The plasmids pFNO3 (*A. fumigatus pyrG*), pRF281 (*Neurospora crassa pyr4)*, and pLH1 (*Trichoderma reesei pyrG*) were used for amplification of the orotidine 5-phosphate

decarboxylase encoding gene required for the transformation and recovery of a functional pyrimidine biosynthetic pathway (LACROUTE 1968).

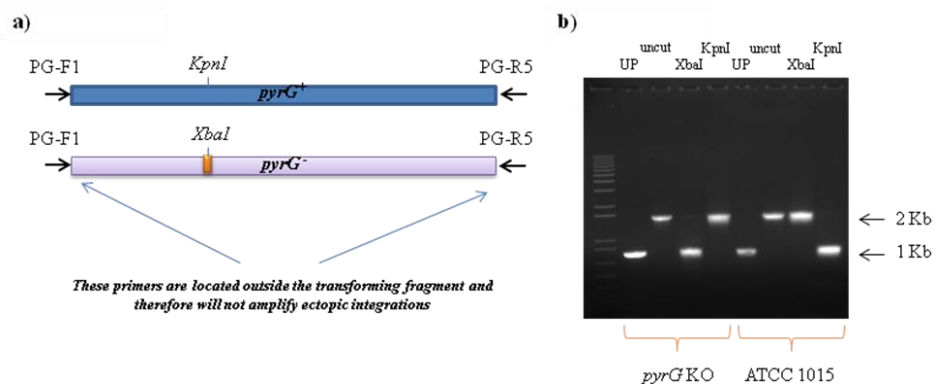



**Figure 21**. Schematic representation of *pyrG* deletion. a) Genomic region of the targeted *pyrG* locus. *pyrG*$^{+}$ corresponds to the ATCC 1015 *pyrG* wild type locus, *pyrG*$^{-}$ corresponds to the ATCC 1015 *pyrG* disrupted locus. b) PCR fragments digested with the RE XbaI or KpnI. The gel electrophoresis lanes are described as follows. Ladder: λ DNA-BstEII digest ladder. UP: amplified upstream flanking region of *pyrG* with PG-F1 and PG-R3B. Uncut: amplified *pyrG* fragment with PG-F1 and PG-R5. *XbaI*: amplified *pyrG* fragment with PG-F1 and PG-R5 and digested with the RE XbaI. *KpnI*: amplified *pyrG* fragment with PG-F1 and PG-R5 and digested with the RE *KpnI*.

Protoplasts and transformation procedures were carried out as described earlier. Transformants were able to grow on minimal media plates without the addition of uracil and uridine, therefore, confirming the introduction of the *pyrG* gene in the genome of the *A. niger* ATCC 1015 strain and the correct recovery of pyrimidine biosynthesis (LACROUTE 1968).
Subsequently, this strain was used for the deletion of the *kusA* gene. Fragments were constructed in order to have a transient KusA disruption mutant as described previously for *A. nidulans* (NIELSEN *et al.* 2008). This work was continued by Pacific Northwest National Laboratory (PNNL) staff in the United States.

**Comment [p5]:** Should I leave this comment?

### 3.5 Analysis of transcription factor functions with transcriptomics

The findings reported in this section are the foundation of manuscript 4: Transcriptome profiling of *Aspergillus niger* AdrA, FacB and CreA mutant genotypes during growth on glucose or glycerol as carbon sources, and manuscript 5: Deletion of a fungal regulatory gene

of the GATA family: *areB* in *Aspergillus niger*. The knowledge gap existing in terms of regulatory mechanisms was the driving force for conducting this kind of studies. When trying to improve production yields, there are bottlenecks which cannot be overcome by simply over-expressing or deleting pathways, mainly because the cell's metabolism is a highly buffered system regulated by a number of transcription factors. Therefore, a more thorough knowledge of their regulatory roles is required. Even though *A. niger* is a widely exploited species for production of organic acids and enzymes, relatively little is known about the regulation of its metabolism. Thus, we made use of four transcription factor knock-out strains and gene expression profiling in order to reveal novel regulatory functions of the transcription factors investigated.

### 3.5.1 Transcriptome profiling of *Aspergillus niger adrA, facB* and *creA* gene deletion strains during growth on repressing and derepressing conditions

Glucose repression is a widely known mechanism by which the presence of glucose represses transcription of genes involved in the utilization of other less favored carbon sources. A full mechanistic explanation of this phenomenon and the other key players involved, besides CreA, has not yet been elucidated in Aspergilli.

Using genome-wide transcription analysis we identified genes either affected directly or indirectly by AdrA, FacB and CreA transcription factors during growth on glucose or glycerol as carbon sources. In order to do this, we conducted well controlled batch fermentations with the corresponding deletion strains (strain construction details given in previous section) and collected mycelium used for further transcriptome profiling. Clustering of genes defined for each gene deletion strain allowed the identification of genes presumably regulated by each transcription factor or, alternatively, indirectly regulated.

#### 3.5.1.1 Physiology of *adrA, facB* and *creA* deletion strains

Physiology of *adrA, facB* and *creA* transcription factor deletion strains was evaluated on solid media plates (see Figure 22) and in liquid batch cultivations (see Table 11).

In summary, colony shape and conidiation on glucose was affected in all three deletion strains compared to the wild type strain ATCC 1015. As expected, *facB* deletion mutant could not grow on acetate as sole carbon source, but growth was normal on the other carbon sources evaluated. Growth of *adrA* deletion strain was slightly affected on lactose and on L-proline

when compared to ATCC 1015 strain, forming fewer conidia, but not affected at all on glycerol. In contrast, *creA* knock-out morphology was affected on all carbon sources tested.
In liquid cultivations, morphology of all cultures was filamentous. See manuscript 4 for complete results.

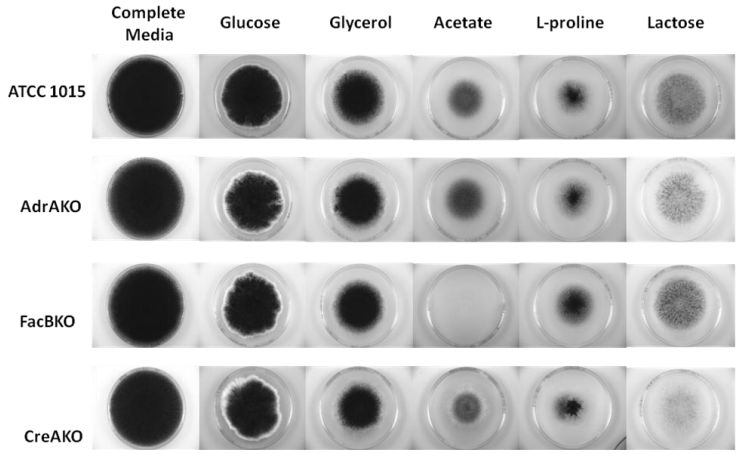



**Figure 22**. Growth assay with *adrA*, *facB* and *creA A. niger* deletion strains constructed. Strains were grown on different carbon sources and compared to wild type strain ATCC 1015. The following carbon sources were added to minimal medium containing 2% agar: glucose (10 g $L^{-1}$), glycerol (10 g $L^{-1}$), acetate and L-proline (50 mM), and lactose (5 g $L^{-1}$). Complete media contains (10 g $L^{-1}$ glucose monohydrate, 2 g $L^{-1}$ yeast extract and 3 g $L^{-1}$ tryptone). The pH was adjusted to 6.35. Plates were incubated for 9 days at 30 °C.

**Table 11**. Kinetic parameters and yield coefficients of each strain grown on glucose or glycerol. TA: Time of biomass sampling for transcriptome analysis.

| Strain | Carbon Source | $\mu_{max}$ ($h^{-1}$) | $Y_{s/x}$ (g DW/g substrate) | Consumption rate (g substrate/Lh) | Time at TA (h) | Biomass concentration (g DW/Kg) |
|---|---|---|---|---|---|---|
| ATCC 1015 | Glucose | 0.202 ± 0.032 | 0.450 ± 0.068 | 1.590 ± 0.018 | 25 | 5.384 ± 0.802 |
| (WT) | Glycerol | 0.056 ± 0.004 | 0.375 ± 0.009 | 0.114 ± 0.019 | 60 | 1.660 ± 0.248 |
| Adr1KO | Glucose | 0.194 ± 0.020 | 0.545 ± 0.092 | 1.472 ± 0.118 | 25 | 4.340 ± 0.537 |
| | Glycerol | 0.069 ± 0.002 | 0.592 ± 0.034 | 0.081 ± 0.001 | 60 | 1.295 ± 0.094 |
| FacBKO | Glucose | 0.209 ± 0.002 | 0.507 ± 0.032 | 1.975 ± 0.095 | 25 | 1.300 ± 0.233 |
| | Glycerol | 0.072 ± 0.003 | 0.707 ±0.021 | 0.111 ± 0.002 | 60 | 1.857 ± 0.554 |
| CreAKO | Glucose | 0.203 ± 0.002 | 0.512 ±0.075 | 1.704 ± 0.056 | 25 | 2.133 ± 0.692 |
| | Glycerol | 0.062 ± 0.004 | 0.653 ±0.020 | 0.089 ± 0.012 | 60 | 1.376 ± 0.355 |

The maximum specific growth rate on glucose for all mutants was similar to ATCC 1015 strain. Whereas on glycerol, $\mu_{max}$ and yields were considerably higher in all three gene deletion mutants than in ATCC 1015 (see Table X). Even though, biomass concentration at the time of sampling for transcriptome analysis (TA) was different for each strain, samples for transcriptome analysis were harvested at 25 h for glucose fermentations and at 60 h for glycerol fermentations.

#### 3.5.1.2 Transcriptional profiling and identification of patterns by gene clustering

Gene expression changes were assessed by using moderated Student's t-test and ANOVA statistics.

T-test pair-wise comparison of *adrA* deletion mutant compared to ATCC 1015 detected 980 significant gene expression changes on glycerol, while none on glucose (for complete results see manuscript 4). Among those, 386 genes were significantly up-regulated in the *adrA* deletion mutant where we found lipids and acetyl-CoA metabolic processes as well as histidine biosynthetic process and ribosome biogenesis over-represented biological process GO-terms.

Three ANOVA analyses, one corresponding to each deletion mutant compared to the wild type strain, were run. In each case, we compared the gene expression profiles of each gene deletion mutant grown on glucose or glycerol to the transcriptome profiles of the ATCC 1015 strain under the same conditions. Three main effects were evaluated, one for the genotype effect (mutant/wild type), one for the carbon source effect (glucose/glycerol) and one for the combined effect or interaction (genotype/carbon source). The complete datasets are given in Supplementary Tables S1.1., S1.2., and S1.3 of manuscript 4. The major effect was due to the change of carbon source for all three mutants, whereas the genetic effect was the least important. Clustering analysis was performed considering the non-overlapping totally differentially expressed genes obtained from ANOVA analysis for *adrA* and *creA* deletion mutants; hence, 3,507 genes and 3,962 genes, respectively.

ANOVA analysis showed 131 genes affected in response to deletion of *adrA* and the same over-represented biological processes GO-terms, as in case of the t-test pair-wise comparison on glycerol, were found.

Deletion of *facB* and transcriptome profiling of the strain under repressing (glucose) and derepressing (glycerol) conditions showing no significant changes in gene expression on other pathways was major evidence supporting the fact that FacB is a regulator specific to acetate

and acetamide metabolism, even though the genes relevant to these pathways were not affected under this specific comparison (see details in manuscript 4).

ANOVA analysis of *creA* deletion mutant identified 1,041 genes as changing expression level, and 319 genes were specific to the genetic perturbation (details in results section of manuscript 4). Among those 319 differentially expressed genes, we identified several glycoside hydrolases and putative transcriptional regulators. Similarly, moderated student's t-test pair-wise comparisons of *creA* deletion mutant versus ATCC 1015 strain were performed on glucose and on glycerol. On glucose, 121 genes were identified as having statistically significant expression level changes, where 112 genes were up-regulated in the CreA knock-out strain including methyl isocitrate lyase (JGI196237), alpha amylase (JGI140567), and feruloyl esterase (*faeB*/JGI51478) (DE VRIES *et al.* 2002b). On glycerol, 1,224 gene expression changes were identified. We hypothesize that these facts could be an indication of the strong repression caused by glucose by means of other transcriptional regulators besides CreA, where deletion of CreA did not alleviate repression of genes caused by glucose.

Clustering confirmed the complex regulatory effect of AdrA and CreA at the transcriptome level, where eight clusters were chosen as displaying interesting patterns (Figure 23). Cluster 1 (582 genes) from Figure 23 a) showed up-regulated genes in the *adrA* deletion mutant with respect to ATCC 1015 when grown on glycerol, but exhibited similar levels of expression on glucose. Over-represented biological processes included amino acid and $NAD^+$ biosynthetic and metabolic processes (see Figure 23 a).

In contrast, for example, cluster 6 (253 genes) grouped genes showing higher transcript levels on glucose in the *adrA* deletion strain compared to ATCC 1015 indicating loss of repression, while having higher expression level in ATCC 1015 compared to *adrA* deletion strain on glycerol indicating loss of activation (see Figure 23 a). Biological processes related to regulation of transcription and signal transduction as well as DNA and RNA processing were over-represented (details in results section of manuscript 4). This cluster contained numerous transcription factors, i.e., TorA (JGI53581); an HLH transcription factor, Hpa3 (JGI181931); a putative GATA transcription factor, SreP (JGI52040); a negative nitrogen transcriptional regulator, AreB (JGI36739); a forkhead family transcription factor (JGI214129); a homeobox transcription factor (JGI126405) and a histone transcription regulator 1 (JGI175742). A summary of the pathways regulated by AdrA, as indicated by cluster analysis, is shown in Figure 24.

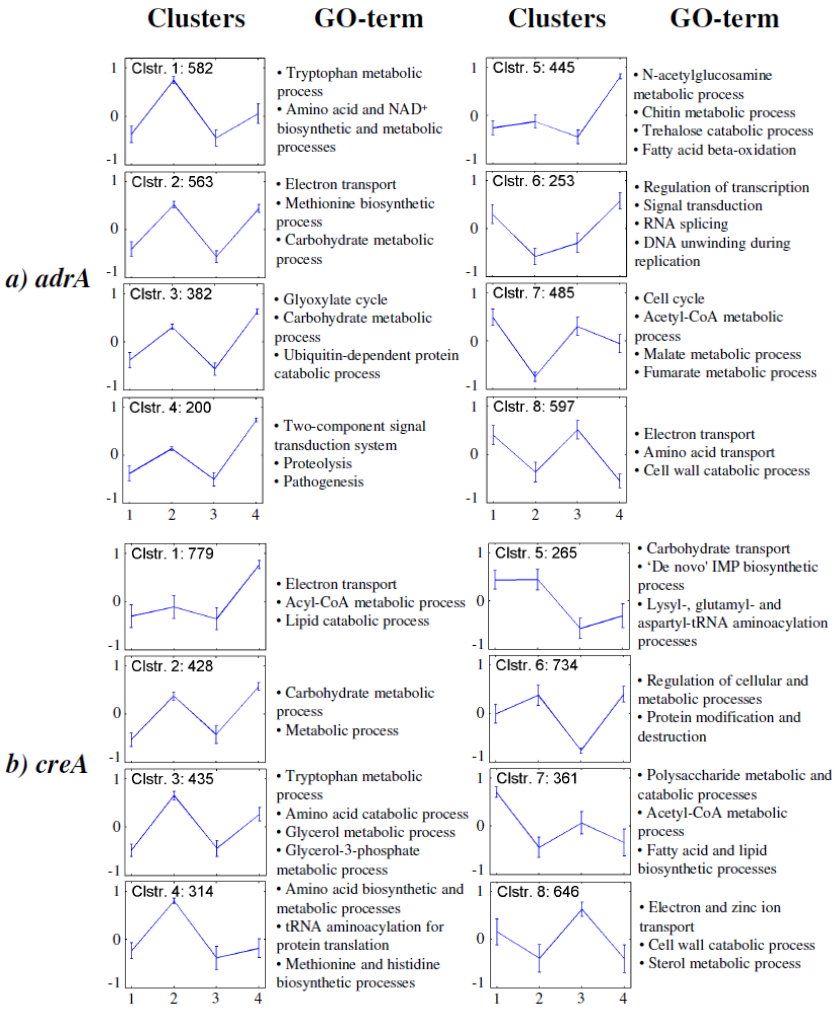



**Figure 23**. Cluster patterns and over-represented biological processes GO-terms using the significantly differentially expressed genes obtained by ANOVA analysis. a) *adrA* deletion mutant clustering patterns; b) *creA* deletion mutant clustering patterns. The "X" axis represents the four different conditions investigated: 1) Deletion mutant_Glucose; 2) Deletion mutant_Glycerol, 3) Wild type ATCC 1015_Glucose; 4) Wild type ATCC 1015_Glycerol; the "Y" axis represents normalized gene expression intensities.

The effect of *creA* deletion when the strain was grown under "non-repressing conditions" was fairly evident indicating probable repression when glycerol was used as a carbon source or relief of repression over other transcriptional activators via CreA as shown in cluster 4 (Figure

23 b). Here, higher transcript levels of *creA* deletion mutant compared to ATCC 1015 strain were obtained. Up-regulation of genes including those involved in amino acid related processes, i.e., amino acid biosynthetic process were identified (Figure Xb), but also several transcription factors, i.e., the acid regulatory protein PacC (JGI47049) (ANDERSEN *et al.* 2009a; PENALVA and ARST 2004) and the putative NF-X1 finger transcription factor (JGI191797).

Cluster 5 also grouped genes that were clearly derepressed by the deletion of *creA* (Figure 23 b) and included numerous genes already known to be glucose repressed through CreA, i.e., alpha-glucosidase B (JGI119858), alpha-galactosidase C (JGI212736), glycosyl hydrolase (JGI173507), aldehyde dehydrogenase (JGI196874), the transcriptional activator XlnR (JGI48811), but also others likely affected, such as a putative Zn2Cys6 transcription factor (JGI188323) and oxaloacetate acetylhydrolase (JGI57241).

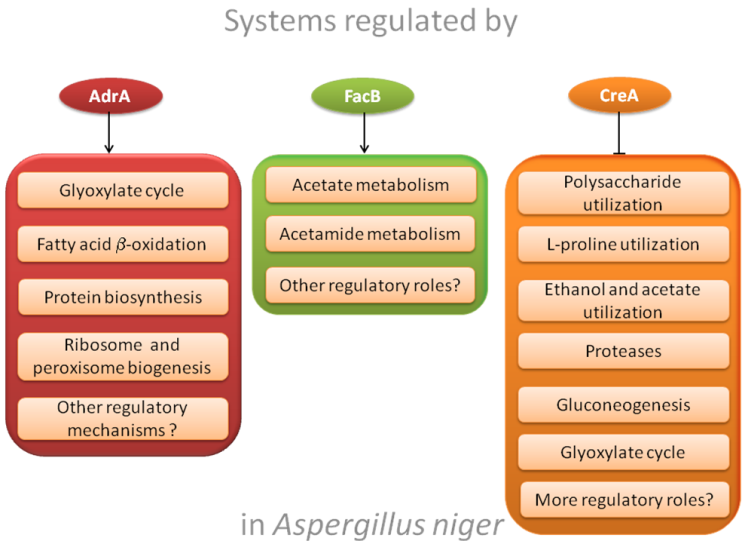



**Figure 24**. Systems regulated by AdrA, FacB and CreA transcription factors in *A. niger*.

Similarly, cluster 6 grouped derepressed genes on glucose, where processes such as regulation of cellular and metabolic processes were the most enriched. Among those genes we found at least 16 putative or annotated transcription factors including TorA (JGI53581), the pH-response regulator protein PalI (JGI52449) (ARST *et al.* 1994), the putative PrnA transcription factor (JGI208882), which was earlier established not to be either self-regulated or significantly affected by carbon and/or nitrogen metabolite repression in *A. nidulans*

(CAZELLE *et al.* 1998), the acetate metabolism regulator FacB (JGI139020), the MADS-box transcription factor RlmA (JGI51606) (DAMVELD *et al.* 2005b) and the nitrogen regulatory protein AreA (JGI53926). Similarly to AdrA, a summary of the pathways regulated by CreA are shown in Figure 24.

The expression profiles obtained confirmed that regulation of metabolism, i.e., carbon catabolite repression is not a simple on/off process dependent on the presence of only a few key transcription factors, such as CreA or AdrA, but a more complex regulatory system. Overall, these results suggest that regulation of glycoside hydrolases is not solely dependent on CreA transcriptional regulation, because we would expect up-regulation of these enzymes in the *creA* knock-out strain, as suggested by de Vries et al. (DE VRIES *et al.* 1999). In contrast, we still found down-regulated glycoside hydrolases on glucose in the *creA* deletion mutant suggesting that CreA is not the only regulator involved, but a more complex and tight regulatory system, where other transcription factors such as the transcriptional activators XlnR (TAMAYO *et al.* 2008), AraR, recently characterized by de Vries et al. (DE VRIES 2009), and others act in concert with CreA to regulate several pathways. In contrast, FacB is a substantially less broad regulator which regulates specific pathways such as aldehyde and acetamide metabolism and does not seem to regulate any other pathways.

One important feature found in this study was the large number of putative and annotated regulatory proteins identified as changing expression level in response to the deletion of the transcription factors AdrA or CreA. To our knowledge, it has not been reported that TorA is up-regulated during growth on glycerol compared to glucose in *A. niger*, while the opposite occurring in the *adrA* and *creA* deletion strains, and in the later case, no transcript level changes were reasonably detected on glycerol (see results section in manuscript 4). Our results could indicate crosstalk between different regulatory pathways in *A. niger* as occurs in other organisms such as *S. cerevisiae*. On the other hand, it seems that the TOR pathway plays only a minor role in regulation of nitrogen metabolism in *Aspergillus* spp. (FITZGIBBON *et al.* 2005), where AreA is the key regulator, giving room for a role on other regulatory processes.

### 3.5.2 An insight into the role of AreB in carbon and nitrogen assimilation and metabolism

In the presence of preferred nitrogen sources, such as ammonium or glutamine, AreA mediated regulation leads to repression of activities involved in the utilization of other less-favored nitrogen sources (MACCABE *et al.* 1998). AreB, is another GATA type regulator of nitrogen metabolism identified in *A. nidulans* (CONLON *et al.* 2001) and in *Penicillium chrysogenum* (HAAS *et al.* 1997). Studies on *A. nidulans* demonstrated that the transcription factor AreB has a wide action domain including, but not restricted to, nitrogen regulation.

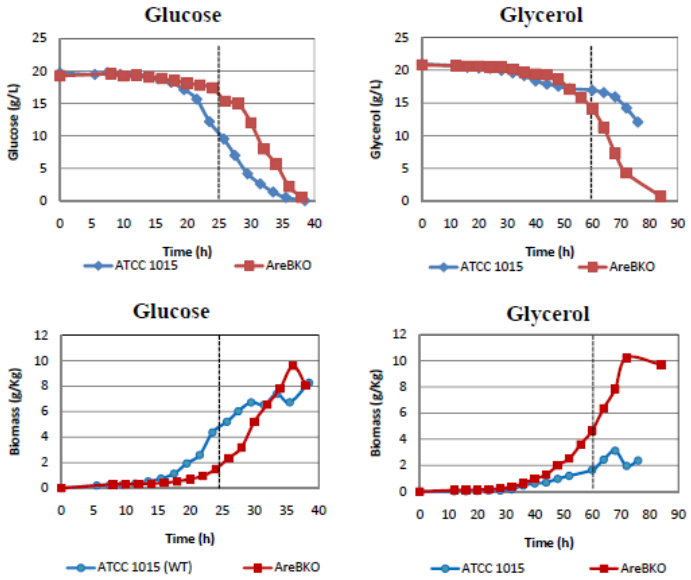
Possibly, AreB competes with AreA for DNA binding (CONLON *et al.* 2001). In order to investigate the role of AreB in nitrogen and possibly in carbon regulation in *A. niger*, we analyzed the *areB* deletion phenotype in different cultivation conditions and conducted transcriptome analysis to further expand on our knowledge about regulation of metabolism.

#### 3.5.2.1 Deletion of the *Aspergillus niger* nitrogen regulatoy gene *areB* reveals pleiotrophic phenotypes

Growth assays on agar plates and in shake flask cultivations using the *areB* deletion mutant and ATCC 1015 wild type strain demonstrated that nitrogen and carbon source utilization was affected in the deletion mutant (see results section in manuscript 5). Nitrogen source utilization was tested using minimal medium and several nitrogen sources, including: alanine, ammonium, arginine, glutamine, L-proline and nitrate with either glucose or glycerol as carbon source. A replicate with L-proline as both carbon and nitrogen source was also included. There were a number of clear growth differences between the two strains. In general, *areB* gene deletion mutant accumulated more biomass with ammonium, nitrate, and glutamine as nitrogen sources with glucose or glycerol as carbon sources (see complete details in results section of manuscript 5). In contrast, ATCC 1015 *A. niger* strain had increased growth on alanine and L-proline with glucose when compared to glycerol. Whereas, on glycerol plus alanine, the growth of *areB* deletion strain was better than in ATCC 1015 strain.

Fermentation profiles based on average values are shown in Figure 25. Glucose was consumed at similar specific consumption rates at the time of sample harvest for transcriptome analysis, while on glycerol, the specific consumption rate was considerably higher in the *areB* deletion mutant. As shown in Figure 25, *areB* deletion mutant $\mu_{max}$ on glycerol was approximately 57% higher compared to ATCC 1015 $\mu_{max.}$ Overall, deletion of AreB caused a substantial growth effect. In shake flask cultivations and confirmed in batch

fermentations, *areB* deletion strain was able to accumulate more biomass on both glucose and glycerol when using ammonia as a nitrogen source.




| Strain | Carbon Source | $\mu_{max}$ ($h^{-1}$) | $Y_{s/x}$ (g biomass g substrate$^{-1}$) | Consumption rate[a] (g substrate g biomass$^{-1}$ $h^{-1}$) |
|---|---|---|---|---|
| ATCC 1015 (WT) | Glucose | 0.202 ± 0.032 | 0.450 ± 0.068 | 1.478 ± 0.487 |
| | Glycerol | 0.056 ± 0.004 | 0.375 ± 0.009 | 0.066 ± 0.056 |
| AreBKO | Glucose | 0.180 ± 0.002 | 0.467 ± 0.071 | 1.263 ± 1.161 |
| | Glycerol | 0.088 ± 0.001 | 0.567 ± 0.022 | 0.444 ± 0.118 |

**Figure 25**. Physiological characterization in batch fermentations. Substrate consumption and biomass growth profiles for *A. niger areB* gene deletion mutant and the wild type strain ATCC 1015 are shown. Dashed lines indicate the time, where samples for transcriptome profiling were harvested. Kinetic parameters and yield coefficients of each strain grown on glucose or glycerol are reported. [a] Due to the time-wise dependency of specific substrate consumption rates, these were calculated using the biomass and substrate concentrations measured at the time of sample harvest for transcriptome analysis and the closest previous values, where $q_s=(\Delta S/\Delta x)/\Delta t$; S=substrate, x=biomass and t=time.

Based on this finding, we suspected that AreB could act as a repressor of growth related genes, such as those involved in amino acid metabolism and other biosynthetic genes. To investigate the possibility, the transcriptomes of both strains, AreBKO and ATCC 1015 strain, were compared following growth on glucose or glycerol as carbon source and ammonium as nitrogen source.

### 3.5.2.2 Transcriptional profiling of *areB* deletion mutant and identification of patterns by gene clustering

In our analysis, we first did data quality assessment and two microarrays of doubtful quality were discarded (details in results section in manuscript 5). Two statistical analyses were run, ANOVA analysis (Figure 26 a) and moderated Student's t-test pair-wise comparison on each carbon source.

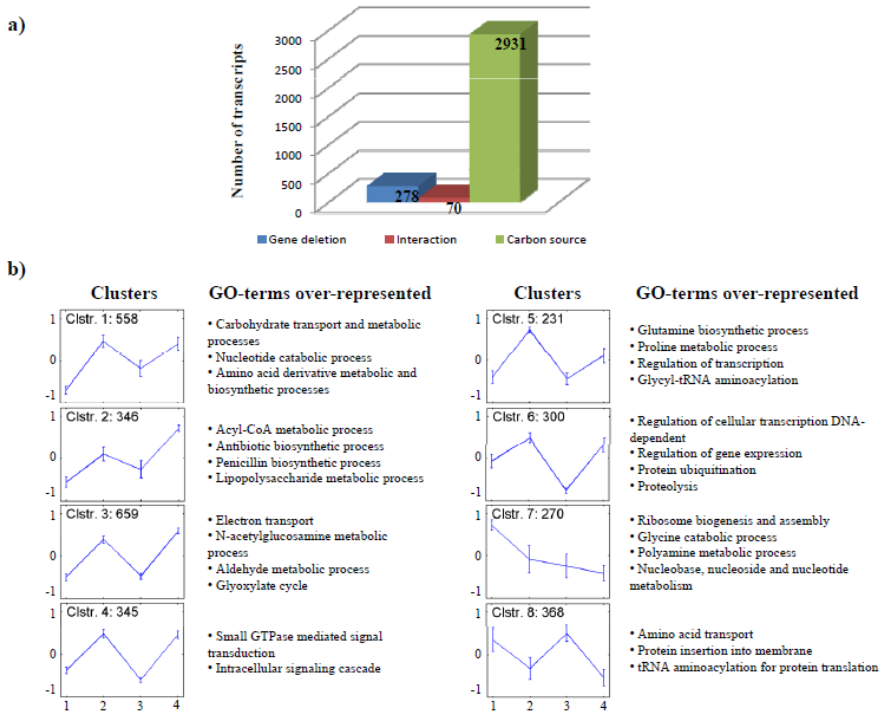



**Figure 26**. a) Statistically significant differentially expressed genes identified by ANOVA analysis for the *areB* gene deletion strain compared to the ATCC 1015 wild type strain grown on glycerol or glucose. Three main effects were evaluated: 1) gene deletion effect (mutant/wild type), 2) interaction effect (genotype/carbon source) and 3) carbon source effect (glucose/glycerol). The "X" axis represents the affected genes under each effect. The "Y" axis represents the number of transcripts identified as changing expression level under a statistical cut-off of adjusted p value < 0.05. b) Clustering patterns using all non-overlapping differentially expressed genes captured by ANOVA analysis. The "X" axis represents the four different conditions investigated: 1) *areB* deletion strain_Glucose; 2) *areB* deletion strain_Glycerol, 3) ATCC 1015_Glucose; 4) ATCC 1015_Glycerol; the "Y" axis represents

normalized gene expression intensities. Over-represented biological process GO-terms are included for each cluster.

ANOVA identified 278 genes changing expression level due to *areB* deletion. Complete differentially expressed genes list is shown in Supplementary File 3 from manuscript 5. Genes included glycosyl hydrolases, glycosyl transferases, acetyl transferases and proline oxidases. GO-term enrichment analysis showed as top biological processes GO-terms: polysaccharide catabolic process and glutamine family amino acid catabolic process, GO-term associated to the chemical reactions and pathways resulting in the breakdown of amino acids of the glutamine family, comprising arginine, glutamate, glutamine and proline.
Further clustering analysis considering all the non-overlapping 3,077 genes changing expression level and captured by the ANOVA analysis showed that AreA, TorA, and PalI were affected in their expression levels (found in cluster 6, Figure 26 b).
For example, clusters 5 and 6 from Figure 26 b) grouped genes that showed a certain level of derepressed phenotype. They contained up-regulated genes in the *areB* gene deletion strain compared to ATCC 1015 strain when grown on glycerol (cluster 5) or on glucose (cluster 6). Over-represented biological processes GO-terms with genes from cluster 5 included glutamine biosynthetic process, proline metabolic process, glycyl-tRNA aminoacylation and several other regulation GO-terms (see Figure 26 b). Therefore, clustered genes included regulatory proteins such as XlnR/JGI48811 (VAN PEIJ *et al.* 1998) and PacC/JGI47049 (PENALVA and ARST 2004). Similarly, cluster 6 grouped several regulatory proteins including AreA/JGI53926, PalI/Rim9/JGI52449, Hpa3/JGI181931, TorA/JGI53581, RlmA/JGI51606 (DAMVELD *et al.* 2005b) and putative C6 transcription factors (JGI184609 and JGI131636). Not surprisingly, biological processes related to regulation of transcription were over-represented, i.e., regulation of gene expression (Figure 26 b).

Moderated Student's t-test statistics of *areB* gene deletion mutant compared to ATCC 1015 strain on glycerol, detected 569 significant gene expression changes using an adjusted p value $< 0.05$ as a cut-off to assess significance. Among them, 316 genes were up-regulated and included proline (JGI178560) and GABA permeases (JGI197679), neutral amino acid permease (JGI191223), epoxide hydrolase (JGI51646), ubiquinol oxidase (JGI47967) and a wide range of genes encoding putative dehydrogenases, hydrolases, oxidases and permeases.

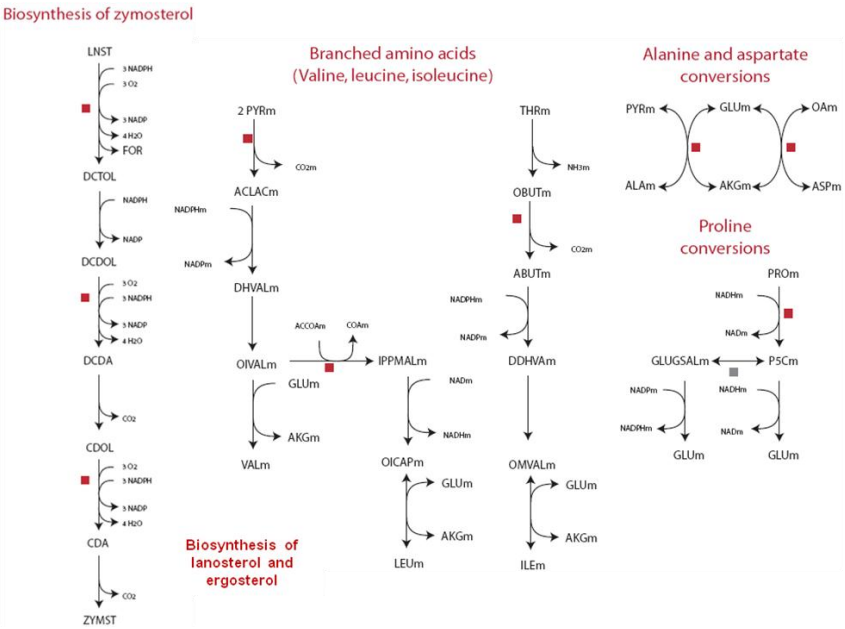



**Figure 27**. Up-regulated metabolic pathways identified after mapping the 569 significant differentially expressed genes obtained by t-test pair-wise comparison of *areB* deletion mutant and ATCC 1015 strain on glycerol into the *A. niger* metabolic map. Red boxes refer to statistically significant up-regulated genes involved in the enzymatic reactions depicted.

Figure 27 shows a relevant subset of up-regulated metabolic pathways found after mapping the 569 significant genes into the *A. niger* metabolic map (ANDERSEN *et al.* 2008a). The branched amino acids biosysnthesis (valine, leucine and isoleucine); alanine, aspartate and proline conversions as well as biosynthesis of lanosterol, zymosterol and ergosterol pathways were up-regulated. The complete metabolic map is shown in Supplementary File 7 of manuscript 5.

Transcriptome analysis provides a valuable tool for determining components of specific biological processes, making this a powerful tool for understanding fundamental aspects of gene regulation (CADDICK and DOBSON 2008). Building on previous work in *A. nidulans* and other fungal species, we believe that our global transcriptional analysis pinpointed new regulatory roles for AreB in addition to its role in nitrogen utilization. By analogy with the action of *S. cerevisiae* Dal80 and Nil2, AreB has been proposed to function negatively by competing with AreA for binding to the promoters of AreA-regulated genes (CONLON *et al.* 2001) such as nitrate and formamide utilization genes (FRASER *et al.* 2001), or even to the

*gabA* gene, encoding the major GABA permease (DAVIS *et al.* 1996). Consistent with this suggestion, we found several up-regulated GABA permeases in the *A. niger areB* gene deletion mutant, e.g., JGI197679, JGI178560 and JGI190162 (see results section in manuscript 5). Similar to *A. nidulans*, *areB* is not an essential gene in *A. niger* (WONG *et al.* 2009b), as shown by the viability of these fungi following complete deletion of the *areB* gene. In contrast, loss of function *nreB* mutants, *areB* homologue in *P. chrysogenum*, could not be isolated (HAAS *et al.* 1997). A role of *A. niger* AreB in conidiation is also suspected as supported by sparser conidia formation observed on nitrate with both glucose and glycerol as carbon source in the *areB* gene deletion mutant compared to the reference strain (solid media growth test) and by the down-regulation of a gene (JGI46001) coding for a protein similar to the conidiation-specific protein Con-10 (ROBERTS *et al.* 1988) detected in the transcriptome analysis.

To investigate several scenarios, the availability of the *A. niger areB* deletion mutant will facilitate the molecular analysis of nitrogen metabolism in this organism despite its lesser genetic amenability compared to other filamentous fungi.

## 4. Conclusions

It is widely known that processes like fungal development, metabolism, stress responses and other responses to diverse signals are regulated by a number of transcription factors in all living organisms. Within this area, this thesis recapitulates the work carried out with filamentous fungi, and specifically, with *Aspergillus niger* in this search for novel regulatory mechanisms and the transcription factors associated to them. To do this, we applied a functional genomics approach, where we made use of three very basic steps: construction of gene deletion strains, physiological characterization and transcriptome profiling. Filamentous fungi, in particular *Aspergillus nidulans* and *Neurospora crassa*, have been traditionally used as model genetic systems to understand numerous processes and mechanisms underlying gene regulation and it is well established that functional links between sets of genes or proteins can be postulated on the basis of observed coordinated gene expression. Therefore, it has been proposed that transcriptomics provides a valuable way to dissect components of specific biological processes, making this a powerful tool for understanding fundamental aspects of gene regulation.

Currrently, approximately 45 transcription factors have been already characterized in *A. nidulans*, however, as discussed earlier through the thesis, this number is substantially lesser in *A. niger*. In *A. niger*, at least three genes exert control across a broad spectrum of metabolic activities, known as wide domain regulatory genes, and they have been cloned and characterized since the early 90's: *creA*, the negatively-acting regulator of carbon catabolite repression; *pacC*, regulator in response to external pH; and *areA*, a positive regulator of nitrogen metabolite repression. However, as shown in the functional annotation of the recently published genome sequence of *A. niger* CBS 513.88 and in the *A. niger* ATCC 1015 JGI genome portal site, there is a vast number of genes encoding putative DNA-binding proteins, identified through their automatic gene modeling pipelines and conserved domain analysis, but a detailed characterization of their functions is still awaiting. The studies reported in this thesis illustrate the complexity of the regulatory circuits regulating cellular processes.

Using genome-wide transcriptome profiling of several *A. niger* strains, from wild type strains, protein producers to gene deletion mutants, we could dig more into the reasons, for example: that are indicative of specialization as an acidogenic or an enzyme-producing strain, where the most plausible explanation for the observed differences was evolution of the strains when exposed to different environments (**manuscript 1**) or, like in manuscripts 2 and 3, the reasons

why glycerol metabolism and maltose metabolism are different in so closely related species, where, i.e., through comparative genomics approaches, we showed that *A. niger* and *A. nidulans* have the same preferred glycerol consumption metabolic pathway, while *A. oryzae* prefers another route for glycerol conversion into the glycolytic intermediate glycerone phosphate (**manuscript 2**). We also demonstrated that *A. niger* and *A. oryzae* does not have the same maltose uptake, metabolism and regulatory mechanisms (**manuscript 3**). In contrast, **manuscripts 4** and **5** were not as focused as the previous manuscripts, but more exploratory in terms of looking for novel regulatory roles played by the transcription factors studied: AdrA, FacB and CreA (**manuscript 4**) and AreB (**manuscript 5**).

As shown in **manuscript 1**, the transcriptome comparison of the two *A. niger* sequenced strains, ATCC 1015 and CBS 513.88, revealed up-regulation of the electron transport chain components, specifically the alternative oxidative pathway in ATCC 1015, while the enzyme producer CBS 513.88 showed significant up-regulation of amino acids biosynthesis genes with over-representation in glucoamylase A, tRNA-synthases and protein transporters. As demonstrated by the gene expression profiling the genetic diversity between both *A. niger* strains ATCC 1015 Vs CBS 513.88 compared was observed to accumulate in metabolic pathways essential to acid production as well as protein synthesis. This multi-disciplinary comparative analysis, where we contributed with the physiological characterization and transcriptome profiling of the strains, identified a number of factors on multiple levels that are indicative of specialization as an acidogenic or an enzyme-producing strain.

Specialization and evolution of strains is not only seen within a single species, but also in different species, such as in different Aspergilli. For example, when studying carbon metabolism with a simple substrate like glycerol, which is catabolized by a wide range of microorganisms including *Aspergillus* species, pathway preferences were also seen (**manuscript 2**). To identify the transcriptional regulation of glycerol metabolism in *Aspergillus*, we analyzed data from triplicate batch fermentations of three different Aspergilli (*A. nidulans*, *A. oryzae* and *A. niger*) with glucose and glycerol as carbon sources. Protein comparisons and cross analysis with gene expression data helped us to identify 88 genes with a conserved transcription response across the three species. Through promoter analysis with the up-regulated genes we detected over-representation of the motif 5'-TGCGGGGA-3' in their upstream regions. This putatively conserved binding site found is similar to the binding site of Adr1 in the yeast *S. cerevisiae* and it also exists in humans. However, as discussed within the thesis, their regulatory role is quite different. Our transcriptome analysis indicated

that genes involved in ethanol, glycerol, fatty acid, amino acids and formate utilization were likely to be regulated by Adr1 in Aspergilli, which could be a sign that this transcription factor is cross species conserved among *Saccharomyces* and more distant *Ascomycetes*, such as *Aspergillus* spp.

The maltose utilization and regulation study (**manuscript 3**), is one more example of using our group previously designed *Aspergillus* GeneChip for validation of the presence of *MAL* gene cluster in the *A. oryzae* genome and the absence of *MAL* gene cluster in the *A. niger* genome, where we identified two *MAL* clusters in *A. oryzae*, but no *MAL* cluster in *A. niger*. In order to utilize maltose, *A. niger* requires a different regulatory system that involves the AmyR regulator for glucoamylase (*glaA*) induction. Through comparative genomics analysis, we also found at least one *MAL* cluster in other Aspergilli, e.g., as in the case of *A. clavatus*, *A. fumigatus* and *A. fischeri* as well as three *MAL* clusters in *A. flavus*. We believe that, although the amount of knowledge on maltose transport and metabolism is far from being complete in *Aspergillus* spp., our study helps to understand the sugar preference in industrial fermentation processes as discussed thoroughly in **manuscript 3**.

Combining the construction of an *areB* deletion mutant in *A. niger* ATCC 1015 with a global transcriptome comparison of the mutant and wild type ATCC 1015 strain, we searched for novel regulatory functions of AreB in the overall metabolism of *A. niger* (**manuscript 5**).

Until now, it was not clear whether *areB* had an essential function, if it had a clear role in nitrogen metabolism in *Aspergillus* or whether this function was conserved across filamentous fungi. To answer these questions, we deleted the *areB* gene in *A. niger* and investigated nitrogen-regulated and carbon-regulated gene expression in this strain. Comparison to the recently obtained results of a deletion mutant in *A. nidulans* were extremely helpful, where it was demonstrated that AreB is a negative regulator of nitrogen metabolic genes in nitrogen limitation and starvation conditions rather than under nitrogen-repressing conditions. In contrast to *A. nidulans*, *A. niger* growth in liquid cultures was enhanced in the *areB* deletion mutant under N-sufficient conditions (ammonium) when both glucose or glycerol were used as carbon source and also in the presence of N-limited conditions (alanine) on glycerol, but not with glucose; and with nitrate on glucose, but not with glycerol. Therefore, it seems that in *A. niger*, AreB negatively affects AreA activation under both nitrogen-repressing and nitrogen-limiting conditions, and that there is a crosstalk of nitrogen and carbon metabolism pathways which is not completely understood. From transcriptome data using ammonium as nitrogen source, elevated *areA* gene expression was observed in the *areB* deletion strain,

indicating that AreB in *A. niger* has a role in nitrogen regulation to negatively modulate AreA activity, as it is proposed to occur in *A. nidulans*. In *S. cerevisiae*, a similar situation occurs, producing a highly responsive, but equally highly buffered control circuit.

By gene clustering of the gene-expression profiles together with GO-term enrichment analysis and metabolic pathway mapping, we could confirm the effect of AreB on nitrogen metabolic pathways such as those related to amino acid biosynthetic pathways, but also on other pathways such as ergosterol, zymosterol and lanosterol biosynthesis and on nitrogen catabolic genes such as the GABA shunt and its probable influence on other transcription factors such as CreC, TorA, PacC, PalI, RlmA and Hpa3. Our results indicate that *A. niger* AreB is a key component in control of carbon and nitrogen metabolism. Improved understanding of the metabolic regulation by AreB and other transcription regulators will lead to improvement of production processes and wider exploitation of *A. niger* as a cell factory.

## 5. References


AFFYMETRIX, and GENECHIP, 2007 Affymetrix Genechip Expression Analysis Technical Manual. P/N 702232, Affymetrix, Santa Clara, CA, Revision 2.

ALEXA, A., J. RAHNENFUHRER and T. LENGAUER, 2006 Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics **22:** 1600-1607.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990a Basic local alignment search tool. J Mol Biol **215:** 403-410.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990b Basic Local Alignment Search Tool. Journal of Molecular Biology **215:** 403-410.

ANASTASSIADIS, S., I. G. MORGUNOV, S. V. KAMZOLOVA and T. V. FINOGENOVA, 2008 Citric Acid Production Patent Review. Recent Patents on Biotechnology **2:** 107-123.

ANDERSEN, M. R., L. LEHMANN and J. NIELSEN, 2009a Systemic analysis of the response of Aspergillus niger to ambient pH. Genome Biology **10:** R47.

ANDERSEN, M. R., M. L. NIELSEN and J. NIELSEN, 2008a Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger. Molecular Systems Biology **4:** 13.

ANDERSEN, M. R., M. SALAZAR, J. NIELSEN and S. E. BAKER, 2009b A Snapshot of Evolution: *Aspergillus niger* ATCC 1015 versus CBS 513.88. Submitted.

ANDERSEN, M. R., W. VONGSANGNAK, G. PANAGIOTOU, P. S. MARGARITA, L. LEHMANN *et al.*, 2008b A trispecies *Aspergillus* microarray - advancing comparative transcriptomics. PNAS **105:** 4387-4392.

ANDERSEN, M. R., W. VONGSANGNAK, G. PANAGIOTOU, M. P. SALAZAR, L. LEHMANN *et al.*, 2008c A trispecies *Aspergillus* microarray: Comparative transcriptomics of three *Aspergillus* species. Proceedings of the National Academy of Sciences of the United States of America **105:** 4387-4392.

ARST, H. N. J., E. BIGNELL and J. TILBURN, 1994 Two new genes involved in signalling ambient pH in *Aspergillus nidulans*. Mol Gen Genet **245:** 787-790.

ASKENAZI, M., E. M. DRIGGERS, D. A. HOLTZMAN, T. C. NORMAN, S. IVERSON *et al.*, 2003 Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. Nature Biotechnology **21:** 150-156.

BAILEY, C., and H. N. ARST, 1975 Carbon catabolite repression in *Aspergillus nidulans*. Eur J Biochem **51:** 573-577.

BAKER, S. E., and J. W. BENNETT, 2008 An Overview of the Genus Aspergillus, pp. 3-13 in *The Aspergilli: Genomics, Medical Aspects, Biotechnology, and Research Methods*, edited by G. H. GOLDMAN and S. A. OSMANI. CRC Press. Taylor & Francis Group, Boca Raton, Florida.

BARBESGAARD, P., H. P. HELDTHANSEN and B. DIDERICHSEN, 1992 ON THE SAFETY OF ASPERGILLUS-ORYZAE - A REVIEW. Applied Microbiology and Biotechnology **36:** 569-572.

BEMBOM, O., S. KELES and M. J. VAN DER LAAN, 2007 Supervised detection of conserved motifs in DNA sequences with cosmo. Statistical Applications in Genetics and Molecular Biology **6**.

BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological **57:** 289-300.

BIRD, D., and R. BRADSHAW, 1997 Gene targeting is locus dependent in the filamentous fungus Aspergillus nidulans. Mol Gen Genet **255:** 219-225.

BLUMENTHAL, C. Z., 2004 Production of toxic metabolites in Aspergillus niger, Aspergillus oryzae, and Trichoderma reesei: justification of mycotoxin testing in food grade enzyme preparations derived from the three fungi. Reg Toxicol Pharmacol **39:** 214-228.

BOK, J. W., and N. P. KELLER, 2004 LaeA, a regulator of secondary metabolism in Aspergillus spp. Eukaryotic Cell **3:** 527-535.

BROAD INSTITUTE DATABASE, ***Aspergillus nidulans*** **genome database**.

CADDICK, M. X., 2004 Nitrogen regulation in mycelia fungi in *The Mycota: III: Biochemistry and Molecular Biology*, edited by G. MARZLUF and R. BRAMBL. Springer Verlag, Berlin.

CADDICK, M. X., and C. DOBSON, 2008 Gene Regulation in *The Aspergilli: Genomics, Medical Aspects, Biotechnology, and Research Methods*, edited by G. H. GOLDMAN and S. A. OSMANI. CRC Press. Taylor & Francis Group, Boca Raton, Florida.

CARLSEN, M., and J. NIELSEN, 2001a Influence of carbon source on alpha-amylase production by Aspergillus oryzae. Applied Microbiology and Biotechnology **57:** 346-349.

CARLSEN, M., and J. NIELSEN, 2001b Influence of carbon source on alpha-amylase production by *Aspergillus oryzae*. Appl. Microbiol. Biotechnol. **57:** 346-349.

CARLSEN, M., J. NIELSEN and J. VILLADSEN, 1996 Growth and alpha-amylase production by Aspergillus oryzae during continuous cultivations. Journal of Biotechnology **45:** 81-93.

CARROLL, A. M., J. A. SWEIGARD and B. VALENT, 1994 Improved vectors for selecting resistance to hygromycin. Fungal Genet Newsl **41:** 22.

CAZELLE, B., A. POKORSKA, E. HULL, P. M. GREEN, G. STANWAY *et al.*, 1998 Sequence, exon-intron organization, transcription and mutational analysis of *prnA*, the gene encoding the transcriptional activator of the *prn* gene cluster in *Aspergillus nidulans*. Mol Microbiol **28:** 355-370.

CHENG, C., N. KACHEROVSKY, K. M. DOMBEK, S. CAMIER, S. K. THUKRAL *et al.*, 1994 Identification of potential target genes for Adr1p through characterization of essential nucleotides in UAS1. Mol Cell Biol **14:** 3842-3852.

CHRISTIAS, C., C. COUVARAKI, S. G. GEORGOPOULOS, B. MACRIS and V. VOMVOYANNI, 1975 PROTEIN-CONTENT AND AMINO-ACID COMPOSITION OF CERTAIN FUNGI EVALUATED FOR MICROBIAL PROTEIN PRODUCTION. Applied Microbiology **29:** 250-254.

COFFMAN, J. A., R. RAI, D. M. LOPRETE, T. CUNNINGHAM, V. SVETLOV *et al.*, 1997 Cross regulation of four GATA factors that control nitrogen catabolic gene expression in Saccharomyces cerevisiae. Journal of Bacteriology **179:** 3416-3429.

CONLON, H., I. ZADRA, H. HAAS, H. N. ARST, M. G. JONES *et al.*, 2001 The Aspergillus nidulans GATA transcription factor gene areB encodes at least three proteins and features three classes of mutation. Molecular Microbiology **40:** 361-375.

CUBERO, B., and C. SCAZZOCCHIO, 1994 2 different, adjacent and divergent Zinc-finger binding-sites are necessary for CreA-mediated carbon catabolite repression in the proline gene-cluster of *Aspergilllus nidulans*. EMBO Journal **13:** 407-415.

CULLEN, D., 2007 The genome of an industrial workhorse. Nature Biotechnology **25:** 189-190.

CVIJOVIC, M., R. OLIVARES-HERNÁNDEZ, R. AGREN, N. DAHAR, W. VONGSANGNAK *et al.*, 2010 BioMet Toolbox: genome-wide analysis of metabolism. Submitted.

DAMVELD, R. A., M. ARENTSHORST, A. FRANKEN, P. A. VANKUYK, F. M. KLIS *et al.*, 2005a The *Aspergillus niger* MADS-box transcription factor RlmA is required for cell wall einforcement in response to cell wall stress. Mol Microbiol **58:** 305-319.

DAMVELD, R. A., M. ARENTSHORST, A. FRANKEN, P. A. VANKUYK, F. M. KLIS *et al.*, 2005b The *Aspergillus niger* MADS-box transcription factor RlmA is required for cell wall reinforcement in response to cell wall stress. Mol Microbiol **58:** 305-319.

DAS, H. K., and M. L. BAEZ, 2008 ADR1 interacts with a down-stream positive element to activate PS1 transcription. Front Biosci **13:** 3439-3447.

DAVID, H., G. HOFMANN, A. OLIVEIRA, H. JARMER and J. NIELSEN, 2006 Metabolic network driven analysis of genome-wide transcription data from *Aspergillus nidulans*. Genome Biol **7**.

DAVIS, M. A., A. J. SMALL, S. KOURAMBAS and M. J. HYNES, 1996 The *tamA* gene of *Aspergillus nidulans* contains a putative zinc cluster motif which is not required for gene function. J Bacteriol **178:** 3406-3409.

DE GROOT, M. J. A., P. BUNDOCK, P. J. J. HOOYKAAS and A. G. M. BEIJERSBERGEN, 1998 Agrobacterium tumefaciens-mediated transformation of filamentous fungi. Nature Biotechnology **16:** 839-842.

DE VRIES, R. P., 2008 A genomic look at physiology and extracellular enzymes of Aspergillus in relation to utilization of plant matter, pp. 87-101 in *Aspergillus in the genomic era*, edited by J. VARGA and R. A. SAMSON. Wageningen Academic Publishers, The Netherlands.

DE VRIES, R. P., 2009 AraR mutants are unable to grow on L-arabinose and L-arabitol. Fungal Genetics Conference Abstract, Asilomar.

DE VRIES, R. P., J. JANSEN, G. AGUILAR, L. PARENICOVA, V. JOOSTEN *et al.*, 2002a Expression profiling of pectinolytic genes from *Aspergillus niger*. FEBS Lett **530:** 41-47.

DE VRIES, R. P., P. A. VANKUYK, H. C. M. KESTER and J. VISSER, 2002b The *Aspergillus niger faeB* gene encodes a second feruloyl esterase involved in pectin and xylan degradation and is specifically induced in the presence of aromatic compounds. Biochem. J. **363:** 377-386.

DE VRIES, R. P., J. VISSER and L. H. DE GRAAFF, 1999 CreA modulates the XlnR-induced expression on xylose of *Aspergillus niger* genes involved in xylan degradation. Res Microbiol **150:** 281-285.

DOGAN DATABASE, ***Aspergillus oryzae*** **genome database**.

DRYSDALE, M. R., S. E. KOLZE and J. M. KELLY, 1993 The *Aspergillus niger* carbon catabolite repression encoding gene, CREA. Gene **130:** 241-245.

DUDOIT, S., R. C. GENDEMAN and J. QUACKENBUSH, 2003 Open source software for the analysis of microarray data. Biotechniques**:** 45-51.

FEDEROFF, H. J., T. R. ECCLESHALL and J. MARMUR, 1983 Carbon catabolite repression of maltase synthesis in *Saccharomyces carlsbergensis*. Journal of Bacteriology **156:** 301-307.

FELENBOK, B., and J. M. KELLY, 1996 Regulation of carbon metabolism in mycelia fungi, pp. 369–380 in *The Mycota: III: Biochemistry and Molecular Biology* edited by G. MARZLUF and R. BRAMBL. Springer Verlag, Berlin, pp 369-380.

FERNANDEZ, R. C., C. A. OTTONI, E. S. DA SILVA, R. M. S. MATSUBARA, J. M. CARTER *et al.*, 2007 Screening of beta-fructofuranosidase-producing microorganisms and effect of pH and temperature on enzymatic rate. Applied Microbiology and Biotechnology **75:** 87-93.

FITZGIBBON, G. J., I. Y. MOROZOV, M. G. JONES and M. X. CADDICK, 2005 Genetic analysis of the TOR pathway in *Aspergillus nidulans*. Eukaryotic Cell **4:** 1595-1598.

FLIPPHI, M., and B. FELENBOK, 2004 The onset of carbon catabolic repression and interplay between specific induction and carbon catabolite repression in *Aspergillus nidulans* in *The Mycota: III: Biochemistry and Molecular Biology*, edited by G. MARZLUF and R. BRAMBL. Springer Verlag, Berlin.

FRASER, J. A., M. A. DAVIS and M. J. HYNES, 2001 The formamidase gene of *Aspergillus nidulans*: Regulation by nitrogen metabolite repression and transcriptional interference by an overlapping upstream gene. Genetics **157:** 119-131.

GALAGAN, J. E., S. E. CALVO, K. A. BORKOVICH, E. U. SELKER, N. D. READ *et al.*, 2003 The genome sequence of the filamentous fungus Neurospora crassa. Nature **422:** 859-868.

GALAGAN, J. E., S. E. CALVO, C. CUOMO, L. J. MA, J. R. WORTMAN *et al.*, 2005 Sequencing of Aspergillus nidulans and comparative analysis with A-fumigatus and A-oryzae. Nature **438:** 1105-1115.

GANCEDO, J. M., 1998 Yeast carbon catabolite repression. Microbiology and Molecular Biology Reviews **62:** 334-+.

GENTLEMAN, R. C., V. J. CAREY, D. M. BATES, B. BOLSTAD, M. DETTLING *et al.*, 2004 Bioconductor: open software development for computational biology and bioinformatics. Genome Biology **5**.

GOTTSCHLING, D. E., O. M. APARICIO, B. L. BILLINGTON and V. A. ZAKIAN, 1990 POSITION EFFECT AT SACCHAROMYCES-CEREVISIAE TELOMERES - REVERSIBLE REPRESSION OF POL-II TRANSCRIPTION. Cell **63:** 751-762.

GROTKJAER, T., O. WINTHER, B. REGENBERG, J. NIELSEN and L. K. HANSEN, 2006 Robust multi-scale clustering of large DNA microarray datasets with the consensus algorithm. Bioinformatics **22:** 58-67.

HAAS, H., K. ANGERMAYR, I. ZADRA and G. STOFFLER, 1997 Overexpression of *nreB*, a new GATA factor-encoding gene of *Penicillium chrysogenum*, leads to repression of the nitrate assimilatory gene cluster. J Biol Chem **272:** 22576-22582.

HOFMANN, G., M. MCINTYRE and J. NIELSEN, 2003 Fungal genomics beyond Saccharomyces cerevisiae? Current Opinion in Biotechnology **14:** 226-231.

HUA, S. B., M. S. QIU, E. CHAN, L. ZHU and Y. LUO, 1997 Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. Plasmid **38:** 91-96.

HYNES, M. J., 2008 Gluconeogenic Carbon Metabolism in *The Aspergilli: Genomics, Medical Aspects, Biotechnology, and Research Methods*, edited by G. H. GOLDMAN and S. A. OSMANI. CRC Press. Taylor & Francis Group, Boca Raton, Florida.

HYNES, M. J., S. L. MURRAY, G. S. KHEW and M. A. DAVIS, 2008 Genetic analysis of the role of peroxisomes in the utilization of acetate and fatty acids in *Aspergillus nidulans*. Genetics **178:** 1355-1369.

IRIZARRY, R. A., B. HOBBS, F. COLLIN, Y. D. BEAZER-BARCLAY, K. J. ANTONELLIS *et al.*, 2003 Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:** 249-264.

ISHIBASHI, K., K. SUZUKI, Y. ANDO, C. TAKAKURA and H. INOUE, 2006 Nonhomologous chromosomal integration of foreign DNA is completely dependent on MUS-53 (human Lig4 homolog) in Neurospora. Proceedings of the National Academy of Sciences of the United States of America **103:** 14871-14876.

JGI DATABASE, ***Aspergillus niger* genome database**

JORGENSEN, T. R., P. A. VANKUYK, B. R. POULSEN, G. J. G. RUIJTER, J. VISSER *et al.*, 2007 Glucose uptake and growth of glucose-limited chemostat cultures of Aspergillus niger and a disruptant lacking MstA, a high-affinity glucose transporter. Microbiology-Sgm **153:** 1963-1973.

KATO, N., Y. MURAKOSHI, M. KATO, T. KOBAYASHI and N. TSUKAGOSHI, 2002 Isomaltose formed by alpha-glucosidases triggers amylase induction in Aspergillus nidulans. Current Genetics **42:** 43-50.

KELLY, J. M., and M. J. HYNES, 1977 Increased and decreased sensitivity to carbon catabolite repression of enzymes of acetate metabolism in mutants of *Aspergillus nidulans*. Mol Gen Genet **156:** 87-92.

KLEIN, C. J. L., L. OLSSON, B. RONNOW, J. D. MIKKELSEN and J. NIELSEN, 1996 Alleviation of glucose repression of maltose metabolism by MIG1 disruption in *Saccharomyces cerevisiae*. Applied and Environmental Microbiology **62:** 4441-4449.

KOBAYASHI, T., K. ABE, K. ASAI, K. GOMI, P. R. JUVVADI *et al.*, 2007 Genomics of Aspergillus oryzae. Biosci Biotechnol Biochem **71:** 646-670.

KOOISTRA, R., P. J. J. HOOYKAAS and H. Y. STEENSMA, 2004 Efficient gene targeting in Kluyveromyces lactis. Yeast **21:** 781-792.

KRAPPMANN, S., C. SASSE and G. H. BRAUS, 2006 Gene targeting in Aspergillus fumigatus by homologous recombination is facilitated in a nonhomologous end-joining-deficient genetic background. Euk Cell **5:** 212-215.

LACROUTE, F., 1968 REGULATION OF PYRIMIDINE BIOSYNTHESIS IN SACCHAROMYCES CEREVISIAE. Journal of Bacteriology **95:** 824-&.

LARRY, L., I. BARTON, E. CARL, GEORGI and R. L. DAVID, 1972 Effect of maltose on glucoamylase formation by *Aspergillus niger*. Journal of Bacteriology **111:** 771-777.

LOCKINGTON, R. A., L. RODBOURN, S. BARNETT, C. J. CARTER and J. A. KELLY, 2002 Regulation by carbon and nitrogen sources of a family of cellulases in *Aspergillus nidulans*. Fungal Genetics and Biology **37:** 190-196.

LUBERTOZZI, D., and J. D. KEASLING, 2009 Developing Aspergillus as a host for heterologous expression. Biotechnology Advances **27:** 53-75.

MACCABE, A. P., J. VANDENHOMBERGH, J. TILBURN, H. N. ARST and J. VISSER, 1996 Identification, cloning and analysis of the Aspergillus niger gene pacC, a wide domain regulatory gene responsive to ambient pH. Molecular and General Genetics **250:** 367-374.

MACCABE, A. P., S. VANHANEN, M. D. S. GELPKE, P. J. I. VAN DE VONDERVOORT, H. N. ARST *et al.*, 1998 Identification, cloning and sequence of the Aspergillus niger areA wide domain regulatory gene controlling nitrogen utilisation. Biochimica Et Biophysica Acta-Gene Structure and Expression **1396:** 163-168.

MACHIDA, M., K. ASAI, M. SANO, T. TANAKA, T. KUMAGAI *et al.*, 2005 Genome sequencing and analysis of Aspergillus oryzae. Nature **438:** 1157-1161.

MACHIDA, M., O. YAMADA and K. GOMI, 2008 Genomics of Aspergillus oryzae: Learning from the History of Koji Mold and Exploration of Its Future. DNA Res **15:** 173-183.

MACPHERSON, S., M. LAROCHELLE and B. TURCOTTE, 2006 A fungal family of transcriptional regulators: The zinc cluster proteins. Microbiology and Molecular Biology Reviews **70:** 583-+.

MARTENS-UZUNOVA, E. S., and P. J. SCHAAP, 2009 Assessment of the pectin degrading enzyme network of Aspergillus niger by functional genomics. Fungal Genetics and Biology **46:** S170-S179.

MATHIEU, M., and B. FELENBOK, 1994 The *Aspergillus nidulans* CreA protein mediates glucose repression of the ethanol regulon at various levels through competition with the AlcR specific transactivator. Embo Journal **13:** 4022-4027.

MEYER, V., 2008 Genetic engineering of filamentous fungi - Progress, obstacles and future trends. Biotechnol Adv **26:** 177-185.

MEYER, V., M. ARENTSHORST, A. EL-GHEZAL, A. C. DREWS, R. KOOISTRA *et al.*, 2007 Highly efficient gene targeting in the Aspergillus niger kusA mutant. J Biotechnol **128:** 770-775.

MICHIELSE, C. B., M. ARENTSHORST, A. F. J. RAM and C. VAN DEN HONDEL, 2005 Agrobacterium-mediated transformation leads to improved gene replacement efficiency in Aspergillus awamori. Fungal Genetics and Biology **42:** 9-19.

MIZUTANI, O., Y. KUDO, A. SAITO, T. MATSUURA, H. INOUE *et al.*, 2008 A defect of ligD (human lig4 homolog) for nonhomologous end joining significantly improves efficiency of gene-targeting in Aspergillus oryzae. Fungal Genetics and Biology **45:** 878-889.

MOGENSEN, J., H. NIELSEN, G. HOFMANN and J. NIELSEN, 2006a Transcription analysis using high-density microarrays of Aspergillus nidulans wild type and creA mutant during growth on glucose or ethanol. Fungal Genet Biol **43:** 593-603.

MOGENSEN, J., H. B. NIELSEN, G. HOFMANN and J. NIELSEN, 2006b Transcription analysis using high-density micro-arrays of Aspergillus nidulans wild-type and creA mutant during growth on glucose or ethanol. Fungal Genetics and Biology **43:** 593-603.

NAKAJIMA, K., M. SANO and M. MACHIDA, 2000 Current Progress in the Analysis of Transcriptional Regulation in the Industrially Valuable Microorganism *Aspergillus oryzae*. Biotechnol. Bioprocess Eng. **5:** 253-262.

NEEDLEMAN, R. B., D. B. KABACK, R. A. DUBIN, E. L. PERKINS, N. G. ROSENBERG *et al.*, 1984 MAL6 of *Saccharomyces*: A complex genetic locus containing three genes required for maltose fermentation. Proc. Natl. Acad. Sci. USA **81:** 2811-2815.

NIELSEN, J., J. VILLADSEN and G. LIDÉN, 2003 *Bioreaction Engineering Principles*. Kluwer Academic/Plenum Publishers, New York, N.Y.

NIELSEN, J. B., M. L. NIELSEN and U. H. MORTENSEN, 2008 Transient disruption of non-homologous end-joining facilitates targeted genome manipulations in the filamentous fungus Aspergillus nidulans. Fungal Genetics and Biology **45:** 165-170.

NIERMAN, W. C., A. PAIN, M. J. ANDERSON, J. R. WORTMAN, H. S. KIM *et al.*, 2005 Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature **438:** 1151-1156.

NINOMIYA, Y., K. SUZUKI, C. ISHII and H. INOUE, 2004 Highly efficient gene replacements in Neurospora strains deficient for nonhomologous end-joining. Proceedings of the National Academy of Sciences of the United States of America **101:** 12248-12253.

NOVAK, S., V. ZECHNER-KRPAN and V. MARIC, 2004 Regulation of maltose transport and metabolism in Saccharomyces cerevisiae. Food Technology and Biotechnology **42:** 213-218.

OSPINA-GIRALDO, M. D., E. MULLINS and S. KANG, 2003 Loss of function of the Fusarium oxysporum SNF1 gene reduces virulence on cabbage and Arabidopsis. Current Genetics **44:** 49-57.

PEDERSEN, H., M. BEYER and J. NIELSEN, 2000a Glucoamylase production in batch, chemostat and fed-batch cultivations by an industrial strain of *Aspergillus niger*. Appl. Microbiol. Biotechnol. **53:** 272-277.

PEDERSEN, H., M. BEYER and J. NIELSEN, 2000b Glucoamylase production in batch, chemostat and fed-batch cultivations by an industrial strain of *Aspergillus niger*. Applied Microbiology and Biotechnology **53:** 272-277.

PEL, H. J., J. H. DE WINDE, D. B. ARCHER, P. S. DYER, G. HOFMANN *et al.*, 2007 Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88. Nature Biotechnology **25:** 221-231.

PENALVA, M. A., and H. N. ARST, 2004 Recent advances in the characterization of ambient pH regulation of gene expression in filamentour fungi and yeasts. Annual Review of Microbiology **58:** 425-451.

PERRONE, G., A. SUSCA, G. COZZI, K. EHRLICH, J. VARGA *et al.*, 2007 Biodiversity of Aspergillus species in some important agricultural products. Stud Mycol**:** 53-66.

PETERSEN, K. L., J. LEHMBECK and T. CHRISTENSEN, 1999 A new transcriptional activator for amylase genes in Aspergillus. Molecular and General Genetics **262:** 668-676.

POLLARD, K. S., and M. J. VAN DER LAAN, 2005 Cluster Analysis of Genomic Data, pp. 209-228 in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, edited by R. GENTLEMAN, V. J. CAREY, W. HUBER, R. A. IRIZARRY and S. DUDOIT. Springer Science+Business Media, LLC, New York, NY, USA.

PUNT, P. J., F. H. J. SCHUREN, J. LEHMBECK, T. CHRISTENSEN, C. HJORT *et al.*, 2008 Characterization of the *Aspergillus niger* prtT, a unique regulator of extracellular protease encoding genes. Fungal Genet Biol **45:** 1591-1599.

R DEVELOPMENT CORE TEAM, 2007 **A Language and Environment for Statistical Computing**

ROBERTS, A. N., V. BERLIN, K. M. HAGER and C. YANOFSKY, 1988 Molecular analysis of a *Neurospora crassa* gene expressed during conidiation. Mol Cell Biol **8:** 2411-2418.

RONNOW, B., and M. C. KIELLAND-BRANDT, 1993 *GUT2*, a gene for mitochondrial glycerol 3-phosphate dehydrogenase of *Saccharomyces cerevisiae*. Yeast **9:** 1121-1130.

ROTH, S., J. KUMME and H. J. SCHULLER, 2004 Transcriptional activators Cat8 and Sip4 discriminate between sequence variants of the carbon source-responsive promoter element in the yeast Saccharomyces cerevisiae. Current Genetics **45:** 121-128.

RUIJTER, G. J. G., S. A. VANHANEN, M. M. C. GIELKENS, P. J. I. VANDEVONDERVOORT and J. VISSER, 1997 Isolation of Aspergillus niger creA mutants and effects of the mutations on expression of arabinases and L-arabinose catabolic enzymes. Microbiology-Uk **143:** 2991-2998.

RUIJTER, G. J. G., and J. VISSER, 1997 Carbon repression in Aspergilli. Fems Microbiology Letters **151:** 103-114.

RUIZ-DIEZ, B., 2002 Strategies for the transformation of filamentous fungi. J Appl Microbiol **92:** 189-195.

SALAZAR, M., W. VONGSANGNAK, G. PANAGIOTOU, M. R. ANDERSEN and J. NIELSEN, 2009 Uncovering transcriptional regulation of glycerol metabolism in Aspergilli through genome-wide gene expression data analysis. Mol Genet Genomics **282:** 571-586.

SANGER, F., G. M. AIR, B. G. BARRELL, N. L. BROWN, A. R. COULSON *et al.*, 1977 NUCLEOTIDE-SEQUENCE OF BACTERIOPHAGE PHICHI174 DNA. Nature **265:** 687-695.

SARKAR, S., M. X. CADDICK, E. BIGNELL, J. TILBURN and H. N. ARST, 1995 Regulation of gene expression by ambient pH in Aspergillus: Genes expressed at acid pH, pp. 360-363 in *Biochemical-Society 657th Meeting Aberdeen*, Aberdeen, Scotland.

SCAZZOCCHIO, C., 2000 The fungal GATA factors. Current Opinion in Microbiology **3:** 126-131.

SCHNEIDER, T. D., and R. M. STEPHENS, 1990 Sequence Logos - a New Way to Display Consensus Sequences. Nucleic Acids Research **18:** 6097-6100.

SCHUSTER, E., N. DUNN-COLEMAN, J. C. FRISVAD and P. W. M. VAN DIJCK, 2002 On the safety of Aspergillus niger - a review. Applied Microbiology and Biotechnology **59:** 426-435.

SHANKAR, S. K., and V. H. MULIMANI, 2007 Alpha-galactosidase production by Aspergillus oryzae in solid-state fermentation. Bioresour Technol **98:** 958-961.

SHENDURE, J., R. D. MITRA, C. VARMA and G. M. CHURCH, 2004 Advanced sequencing technologies: Methods and goals. Nature Reviews Genetics **5:** 335-344.

SHROFF, R. A., R. A. LOCKINGTON and J. M. KELLY, 1996 Analysis of mutations in the creA gene involved in carbon catabolite repression in Aspergillus nidulans. Canadian Journal of Microbiology **42:** 950-959.

SMYTH, G., 2004 Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3**.

SOLIS-PEREIRA, S., E. FAVELA-TORRES, G. VINIEGRA-GONZALEZ and M. GUTIERREZ-ROJAS, 1993 Effects of different carbon sources on the synthesis of pectinase by *Aspergillus niger* in submerged and solid state fermentations. Appl. Microbiol. Biotechnol. **39:** 36-41.

SUN, J. B., X. LU, U. RINAS and A. P. ZENG, 2007 Metabolic peculiarities of Aspergillus niger disclosed by comparative metabolic genomics. Genome Biol **8**.

SZEWCZYK, E., A. ANDRIANOPOULOS, M. A. DAVIS and M. J. HYNES, 2001 A Single Gene Produces Mitochondrial, Cytoplasmic, and Peroxisomal NADP-dependent Isocitrate Dehydrogenase in *Aspergillus nidulans*. Journal of Biological Chemistry **276:** 37722-37729.

TACHIBANA, C., J. Y. YOO, J. B. TAGNE, N. KACHEROVSKY, T. I. LEE *et al.*, 2005 Combined global localization analysis and transcriptome data identify genes that are directly coregulated by Adr1 and Cat8. Molecular and Cellular Biology **25:** 2138-2146.

TAKAHASHI, T., T. MASUDA and Y. KOYAMA, 2006 Enhanced gene targeting frequency in ku70 and ku80 disruption mutants of Aspergillus sojae and Aspergillus oryzae. Mol Genet Genomics **275**.

TAMAYO, E. N., A. VILLANUEVA, A. A. HASPER, L. H. DE GRAAFF, D. RAMON *et al.*, 2008 CreA mediates repression of the regulatory gene xlnR which controls the production of xylanolytic enzymes in *Aspergillus nidulans*. Fungal Genetics and Biology **45:** 984-993.

TODD, R. B., and A. ANDRIANOPOULOS, 1997 Evolution of a fungal regulatory gene family: The Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genetics and Biology **21:** 388-405.

TODD, R. B., R. L. MURPHY, H. M. MARTIN, J. A. SHARP, M. A. DAVIS *et al.*, 1997 The acetate regulatory gene facB of Aspergillus nidulans encodes a Zn(II)2Cys6 transcriptional activator. Molecular & General Genetics **254:** 495-504.

TONUKARI, N. J., J. S. SCOTT-CRAIG and J. D. WALTON, 2000 The Cochliobolus carbonum SNF1 gene is required for cell wall-degrading enzyme expression and virulence on maize. Plant Cell **12:** 237-247.

VAN DIJCK, P. W. M., G. C. M. SELTEN and R. A. HEMPENIUS, 2003 On the safety of a new generation of DSM Aspergillus niger enzyme production strains. Reg Toxicol Pharmacol **38:** 27-35.

VAN HELDEN, J., B. ANDRE and J. COLLADO-VIDES, 1998 Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. Journal of Molecular Biology **281:** 827-842.

VAN LANEN, J. M., and M. B. SMITH, 1968 Process of producing glucoamylase and an alcohol product. US patent 3,418,211.Hiram Walker & Sons, Inc., pp., Peoria, IL.

VAN PEIJ, N., J. VISSER and L. H. DE GRAAFF, 1998 Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in Aspergillus niger. Molecular Microbiology **27:** 131-142.

WITTEVEEN, C. F. B., and J. VISSER, 1995 Polyols pools in *Aspergillus niger*. FEMS Microbiol Lett **134:** 57-62.

WONG, K. H., M. A. DAVIS, M. J. HYNES and R. B. TODD, 2009a Deletion of the *Aspergillus nidulans* nitrogen regulatory gene *areB* reveals pleiotropic phenotypes, pp. in *The Sixth International Aspergillus Meeting, Asperfest*, Asilomar Conference Center.

WONG, K. H., M. J. HYNES, R. B. TODD and M. A. DAVIS, 2009b Deletion and overexpression of the *Aspergillus nidulans* GATA factor AreB reveals unexpected pleiotropy. Microbiology **155:** 3868-3880.

VONGSANGNAK, W., I. NOOKAEW, M. SALAZAR and J. NIELSEN, 2010 Analysis of genome-wide co-expression and co-evolution of *Aspergillus oryzae* and *Aspergillus niger*. OMICS: A Journal of Integrative Biology **14:** 165-175.

VONGSANGNAK, W., P. OLSEN, K. HANSEN, S. KROGSGAARD and J. NIELSEN, 2008 Improved annotation through genome-scale metabolic modeling of Aspergillus oryzae. Bmc Genomics **9:** 14.

WORKMAN C, JENSEN L, JARMER H, BERKA R, GAUTIER L *et al.*, 2002 A new non-linear normalization method for reducing variability in DNA microarray experiments. Gen Biol **3**.

YOUNG, E. T., K. M. DOMBEK, C. TACHIBANA and T. IDEKER, 2003 Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. Journal of Biological Chemistry **278:** 26146-26158.

YOUNG, E. T., N. KACHEROVSKY and K. VAN RIPER, 2002 Snf1 protein kinase regulates Adr1 binding to chromatin but not transcription activation. Journal of Biological Chemistry **277:** 38095-38103.

YU, J., T. CLEVELAND, W. NIERMAN and J. BENNETT, 2005 Aspergillus flavus genomics: gateway to human and animal health, food safety, and crop resistance to diseases. Rev. Iberoam. Micol. **22:** 194-202.

YU, J. H., Z. HAMARI, K. H. HAN, J. A. SEO, Y. REYES-DOMINGUEZ *et al.*, 2004 Double-joint PCR: a PCR-based molecular tool for gene manipulations in filamentous fungi. Fungal Genetics and Biology **41:** 973-981.

YUAN, X. L., J. A. ROUBOS, C. VAN DEN HONDEL and A. F. J. RAM, 2008a Identification of InuR, a new Zn(II)2Cys6 transcriptional activator involved in the regulation of inulinolytic genes in Aspergillus niger. Molecular Genetics and Genomics **279:** 11-26.

YUAN, X. L., R. M. VAN DER KAAIJ, C. VAN DEN HONDEL, P. J. PUNT, M. VAN DER MAAREL *et al.*, 2008b Aspergillus niger genome-wide analysis reveals a large number of novel alpha-glucan acting enzymes with unexpected expression profiles. Molecular Genetics and Genomics **279:** 545-561.

YUAN, X. L., R. M. VAN DER KAAIJ, C. VAN DEN HONDEL, P. J. PUNT, M. VAN DER MAAREL *et al.*, 2008c *Aspergillus niger* genome-wide analysis reveals a large number of novel alpha-glucan acting enzymes with unexpected expression profiles. Mol Genet Genomics **279:** 545-561.

ZITOMER, R. S., D. L. MONTGOMERY, D. L. NICHOLS and B. D. HALL, 1979 Transcriptional regulation of the yeast cytochrome C gene. Proceedings of the National Academy of Sciences of the United States of America **76:** 3627-3631.

Hoffmann B, et al. (2001). Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene cpcA of filamentous fungus Aspergillus nidulans. Mol. Biol. Cell. 12:2846.

van Helden J, Andre B, Collado-Vides J (1998) Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. J Mol Biol 281:827-842

Bembom O, Keles S, van der Laan MJ (2007) Supervised detection of conserved motifs in DNA sequences with cosmo. Stat Appl Genet Mol Biol 6

Marstrand TT, Frellsen J, Moltk I, Thiim M, Valen E, Retelska D, Krogh A. (2008). Asap: A Framework for over-Representation Statistics for Transcription Factor Binding Sites. PLoS ONE 3(2):e1623

Shibuya I, Tamura G, Ishikawa T, Hara S (1992) Cloning the alpha-amylase cDNA of Aspergillus shirousamii and its expression in Saccharomyces cerevisiae. BioSci Biotech Biochem 56:174-179.

Korman DR, Bayliss FT, Barnett CC, Carmona CL, Kodama KH, Royer TJ, Thompson SA, Ward M, Wilson LJ, Berka RM (1990) Cloning, characterization, and expression of two alpha-amylase genes from *Aspergillus niger* var. *awamori*. Curr Genet 17: 203-12.

## 6. Appendix

**Supplementary Table 1**: Full subset of tri-directional homologues of *A. niger*, *A. nidulans* and *A. oryzae*.

## 7. Acknowledgements

The last four years have been fascinating; they have developed me both as a scientist and as a person and have given me the greatest job opportunity I could ever imagine. I had the occasion to conduct my PhD studies in two places, the Center for Microbial Biotechnology (CMB) at the Technical University of Denmark (DTU) and the Department of Chemical and Biological Engineering in Chalmers University of Technology, in an extremely friendly atmosphere and with the most skilled people from all over the world. Therefore many thanks to the staff at CMB and in Chalmers with whom I have been so fortunate to work during the last years.

I would like to give my highest appreciation to my Supervisor, Dr. Jens Nielsen, who has been there for me throughout the project and has always supported me, even in times when I could not see the end. In this context, I would like to address special thanks to the rest of the members of our old (in DTU) and new (in Chalmers) research group, for fruitful discussions and relevant critique:
In DTU: Kiran R. Patil, Ana Paula Oliveira, Michael Jewett, Roberta Mustacchi, Kanchana,
To others at CMB, but not part of our research group: Dr. Jens C. Frisvad and Dr. Jette Thykaer
In Chalmers: Marta Papini, Gionata Scalcinati, Siavash Partow, Dr. Verena Siewers, Roberto Olivares-Hernandez, Jie Zhang,

Special thanks to Dina Petranovic for being my cosupervisor who gave me guidance in the most difficult times of my PhD, which was towards the end.

I would like to give special thanks to Birgitte Karsbol, Pernille Winther and Trine Bro, secretariat staff at DTU, and to Tina Johansen, Pia Friis, Peter Meincke, Martin Nielsen and Elisabeth Kroger for their invaluable assistance with the experimental work.

I would also give special thanks to Marie Nordqvist, Johanna Zandén and Malin Nordvall at Chalmers University of Technology for their kind assistance with the experimental work and lastly and no least, to Erica Dahlin and Martina Butorac for taking care of all kind of practicalities regarding my stay in Chalmers and in general in Sweden.

Special thanks to Dr. Jakob B. Nielsen and Dr. Michael L. Nielsen at Technical University of Denmark who helped me and encouraged me when I was struggling in the strain construction with the really “unfriendly bug” *Aspergillus niger* and who taught me to run my first PCR reaction even when I had no idea why I had to add the primers and for all the very first hands in experience in Molecular Biology. Special thanks also to Dr. Mikael R. Andersen for his support, help and provision of all kind of tools for data analysis and for being a wonderful and long lasting collaborator.

My greatest appreciation to Dr. Gerald Hofmann, now at DSM, for his friendship and support, for sharing knowledge with me and for teaching me to run my first blastP comparison when I had no idea of what I was trying to do. Also special thanks to Dr. Audrey Diano, now at Novozymes, for being my supportive office mate

in the beginning of my PhD, for becoming my close friend and for teaching me to run my first fermentation ever, for being kind, humble and extremely open with me, I simply have no words to define it. To Dr. Susan Meijer for sharing her experience and giving me her friendship and guidance at all times while we were together in DTU. To Dr. Irina Borodina for being my friend, for her guidance and for her objectiveness at all times.

To my dearest collaborators Dr. Mikael R. Andersen, Dr. Wanwipa Vongsangnak and Dr. Gianni Panagiotou for their excellent input at all times.

I would like to thank my French Master student Camille Clement who worked really hard and always showed positive attitude even when things were not working. And also special thanks to Spanish visiting PhD student Javier Calzada-Funes, who also worked really hard and helped me to complete the very last experiments reported in this thesis.

My greatest gratitude to Dr. Scott E. Baker, Dr. Kenneth S. Bruno, Dr. David E. Culley and Dr. Jon K. Magnuson for tutoring me in Pacific Northwest National Laboratory (PNNL) in Washington, United States in a research stay for three months, for their support, friendship, for sharing time with me after working hours and for organizing extra activities to keep me busy. To Dr. Sue (Ken's wife) for sending me pictures of their kids and to Dr.Shuang and Dr. Dai at PNNL for fruitful discussions.

Special double thanks to my new boss Dr. Lisbeth Olsson for her support and patience and to my new research mates, for embracing me in their group and for their friendship: Hampus Sunner, Christian Thörn, Ali Kazemi, Rakesh Koppram, Gupta Udatha, Ximena Rozo, Dr. Valeria Mapelli, Dr. Yun Shen, Dr. Eva Albers and Dr. Carl Johan Franzen, and lastly and no less, to my dearest and supportive office mate Magnus Ask.

This work could not have been possible without the financial support provided by the Chalmers Foundation and Knut and Alice Wallenberg Foundation and the National Council of Research Conacyt-Mexico financing my PhD stipend.

To my family and friends for their love and support at all times, to the ones still here and to the ones who have already left. Specially to my mother who should get all the credits because without her I would just simply have abandoned this enterprise.

To the love of my life…

**Margarita Salazar Peña**

**Gothenburg, Sweden; june 2010**.