Fungal Genetics and Biology 47 (2010) 539–550

Contents lists available at ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

Identification of modules in *Aspergillus niger* by gene co-expression network analysis

Robert A. van den Berg[b,c,*,1], Machtelt Braaksma[b,1], Douwe van der Veen[a,1], Mariët J. van der Werf[b], Peter J. Punt[b], John van der Oost[a], Leo H. de Graaff[a]

[a] *Laboratory of Microbiology, Fungal Genomics Group, Wageningen University and Research Centre, Dreijenplein 10, 6703 HB Wageningen, The Netherlands*
[b] *TNO Quality of Life, P.O. Box 360, 3700 AJ Zeist, The Netherlands*
[c] *SymBioSys, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven, Belgium*

A R T I C L E I N F O

Article history:
Received 13 October 2009
Accepted 13 March 2010
Available online 27 March 2010

Keywords:
Aspergillus niger
Gene co-expression network
Co-expression
Transcriptional analysis
RNA profiling
Transcription factor binding site

A B S T R A C T

The fungus *Aspergillus niger* has been studied in considerable detail with respect to various industrial applications. Although its central metabolic pathways are established relatively well, the mechanisms that control the adaptation of its metabolism are understood rather poorly. In this study, clustering of co-expressed genes has been performed on the basis of DNA microarray data sets from two experimental approaches. In one approach, low amounts of inducer caused a relatively mild perturbation, while in the other approach the imposed environmental conditions including carbon source starvation caused severe perturbed stress. A set of conserved genes was used to construct gene co-expression networks for both the individual and combined data sets. Comparative analysis revealed the existence of modules, some of which are present in all three networks. In addition, experimental condition-specific modules were identified. Module-derived consensus expression profiles enabled the integration of all protein-coding *A. niger* genes to the co-expression analysis, including hypothetical and poorly conserved genes. Conserved sequence motifs were detected in the upstream region of genes that cluster in some modules, e.g., the binding site for the amino acid metabolism-related transcription factor CpcA as well as for the fatty acid metabolism-related transcription factors, FarA and FarB. Moreover, not previously described putative transcription factor binding sites were discovered for two modules: the motif 5′-CGACAA is overrepresented in the module containing genes encoding cytosolic ribosomal proteins, while the motif 5′-GGCCGCG is overrepresented in genes related to 'gene expression', such as RNA helicases and translation initiation factors.

1. Introduction

Genome-wide gene expression levels as generated by DNA microarray technology give insight into the behavior of individual genes at the cellular level. Such expression levels can be considered as a reflection of the physiological state of an organism and can be used to reveal details of metabolic regulation. The first fungal microarray studies were reported for the model organism *Saccharomyces cerevisiae* (DeRisi et al., 1997; Lashkari et al., 1997), followed by application of this technology for over 20 filamentous fungi including *Aspergillus niger* (Breakspear and Momany, 2007). *A. niger* is of industrial importance as it is the major production organism for citric acid world-wide (Magnuson and Lasure, 2004) and an efficient producer of both homologous and heterologous proteins (Pel et al., 2007; Punt et al., 2002). So far, *A. niger* transcriptome studies have been used to characterize polysaccharide-degrading enzyme systems (Andersen et al., 2008; Jørgensen et al., 2009; Martens-Uzunova and Schaap, 2008; van der Veen et al., 2009; Yuan et al., 2008a,b), to describe spatial colony development (Levin et al., 2007), and to study the *A. niger* response towards reductive stress (Guillemette et al., 2007) or cell wall damage (Meyer et al., 2007).

Most DNA microarray studies, including all above-mentioned *A. niger* studies, focus on differential gene expression between few experimental conditions. However, solely an observation of fold changes of expression of individual genes does not explain how biological processes work together to achieve the cell's objectives. Additional information regarding the activation and co-operation of biological processes can be obtained by comparing gene expression profiles over a range of conditions. For example, genes that encode subunits of a protein complex may have a consistently similar change of expression levels over many conditions. The similar expression of two or more genes over a range of conditions is referred to hereafter as gene co-expression.

Featherstone and Broadie deduced from a *S. cerevisiae* gene co-expression study (Featherstone and Broadie, 2002) that specific

* Corresponding author. Fax: +32 16 325 993.
E-mail address: robert.vandenberg@psy.kuleuven.be (R.A. van den Berg).
[1] These authors contributed equally to this study.

1087-1845/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2010.03.005

sets of genes interact extensively at the level of gene expression, and thus can be described in terms of an interconnected network. Such gene co-expression networks can provide a large-scale, global view of the transcriptional response of an organism.

A comparison of gene co-expression networks constructed from DNA microarray data of the evolutionary distinct organisms *S. cerevisiae*, *Homo sapiens*, *Escherichia coli*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, and *Drosophila melanogaster* indicated that these networks share common structural, or topological, properties (Bergmann et al., 2004). For example, the observed gene co-expression networks consist of co-expressed groups of genes, termed clusters or modules, which are associated with the same cellular function. However, while modules of genes involved in similar cellular functions were identified in all species analyzed (e.g., "glycolysis", "proteasome"), this study also indicated that the higher-order relations between modules differ significantly between the organisms (Bergmann et al., 2004). For example, the average gene expression profiles for genes in the "secreted protein" and "proteasome" modules correlate positively in yeast and *A. thaliana*, negatively in *D. melanogaster*, and do not appear to correlate significantly in *H. sapiens*.

Before gene co-expression networks can be generated and analyzed, two problems need to be solved. First, genomes often encode thousands of proteins. Construction of a co-expression network of this amount of genes will in most cases results in a network that is difficult to interpret due the number of genes and their many connections. Second, the physiological role of a large proportion of proteins is unknown, or at best poorly understood (Hughes et al., 2004). This lack of understanding further hampers the interpretation of any network generated. Different strategies to circumvent these problems can be employed in gene co-expression network analyses. The first strategy is to limit the analysis of co-expression networks to descriptive parameters only, such as the number of connections that a gene has with other genes (connectivity), e.g., see Jordan et al. (2008), van Noort et al. (2004). While this approach provides an understanding of the network at a higher abstraction level, such knowledge cannot be converted easily into understanding the actual underlying biological processes. The second strategy is to investigate only a subset of genes that is relevant for a certain research interest, e.g., see Bergmann et al. (2004), Lee et al. (2004), Neretti et al. (2007). For example, Bergmann and coworkers selected genes participating in eight well-defined *S. cerevisiae* biological processes, and examined co-expression of their orthologous genes in five other organisms (Bergmann et al., 2004). In a variation of this strategy, co-expression between all genes is calculated but the analysis is focused on a part of the network that is of particular biological interest, e.g., a certain oncogene (Basso et al., 2005). These approaches, however, are biased towards already known biological processes. In addition, the selected biological processes might operate distinctly in organisms other than *S. cerevisiae*. To reduce bias of using only known biological processes, yet another approach limits the analysis to genes conserved in different species, for instance, in *S. cerevisiae*, *D. melanogaster* and *C. elegans* (Daub and Sonnhammer, 2008).

In this study, we extend the latter approach to the analysis of gene co-expression networks. For the construction of a gene co-expression network of *A. niger*, a subset of genes is selected that is based on evolutionary conservation of the proteins they encode among 19 different fungal species. Even when no defined function can be assigned to such proteins, their evolutionary conservation suggests a biological role. From expression data of these conserved protein-encoding genes, a gene co-expression network is generated. Subsequently, the topology of this network is used to extend the analysis to less conserved genes excluded from the initial analysis. This approach is followed for the analysis of two *A. niger* DNA microarray data sets cultivated under distinct experimental conditions.

2. Materials and methods

2.1. Culturing

Mildly perturbed conditions: *A. niger* 872.11 (Δ*argB pyrA6 prtF28 goxC17 cspA1*) is derived from CBS 120.49. All media were based on Pontecorvo's minimal medium (pMM) (Pontecorvo et al., 1953), contained 100 mM sorbitol as carbon source and were supplemented with uridine and arginine. Glass 2.5-l fermentors (Applikon) with 2.2 l of pMM were kept at a constant temperature of 30 ± 0.5 °C while fermentor headplates were kept at 8 °C. A total of 1.0×10^6 of spores per ml were added to a fermentor. During germination, each fermentor was aerated through the headspace (50 l/h) and stirred at 300 rpm. When dissolved oxygen tension levels dropped below 60% for over 5 min, the stirrer speed was set to 750 rpm and aeration was switched to sparger inlet. In one experiment, fermentors were induced with either 0.1 mM sorbitol or D-xylose at $T = 14$ h as previously described (van der Veen et al., 2009). In a second experiment, fermentors were induced with 1 mM of various oils at $T = 14$ h and samples were taken before induction and up to 2 h after induction (Table 1).

Strongly perturbed conditions: *A. niger* N402 (*cspA1*) (van Hartingsveldt et al., 1987) is derived from CBS 120.49. All media were based on Bennett's minimal medium (bMM) (Bennett and Lasure, 1991). Both shake flask and fermentor cultures were grown in bMM medium at a constant temperature of 30 ± 0.5 °C and with differing combinations of carbon source, nitrogen source and concentration, and pH of the medium (Table 1) (Braaksma et al., 2009). Fermentor inoculum was pre-cultured in baffled 500 ml Erlenmeyer flasks containing 100 ml bMM (pH 6.5) supplemented with the carbon source and nitrogen source concentrations corresponding to fermentor conditions. These flasks were inoculated with 10^6 spores per liter and incubated in a rotary shaker at 125 rpm until approximately half of the available carbon source was consumed. Cultivations were carried out in 6.6-l fermentors (New Brunswick Scientific) with 5.0 l of bMM. The fermentors were inoculated with 4% (weight/vol) pre-culture. To prevent foaming, 1% (vol/vol) Struktol J-673 antifoam was added to the medium and additional antifoam was added during cultivation when necessary. Each fermentor was sparged with 75 l/h of air with the stirrer speed set at 400 rpm at the start of the cultivation. When dissolved oxygen tension levels dropped below 20%, the stirrer speed was automatically increased to maintain oxygen tension at 20% or until the maximum of 1000 rpm was reached. The pH was controlled by automatic addition of 8 M KOH or 1.5 M $H3PO_4$.

2.2. RNA isolation

Culture samples from mildly perturbed conditions were filtrated and biomass was snap-frozen into liquid nitrogen and stored at −80 °C. Culture samples from strongly perturbed conditions were quenched immediately in methanol at −45 °C as described previously (Pieterse et al., 2006) and centrifuged at −20 °C to remove supernatant. Biomass was frozen into liquid nitrogen and stored at −80 °C. A Trizol–chloroform extraction preceded total RNA extraction with RNeasy mini columns (Qiagen) according to the manufacturer's protocol for yeast. Concentration of total RNA was determined by spectrophotometry. RNA integrity was assessed on an Experion system (Biorad) for samples from mildly perturbed conditions by visual inspection of the electropherograms. Graphs depicting RNA integrity categories were used as visual aids (Schroeder et al., 2006). Electropherograms approximating an RNA integrity number of 8 or lower or with a 28S/18S ratio below 1.8 were discarded. For samples from strongly per-