Methodology article

# Negative Subtraction Hybridization: An efficient method to isolate large numbers of condition-specific cDNAs

Anamika Ray[1], Sunita Macwana[2], Patricia Ayoubi[1], Leo T Hall[2,3], Rolf Prade[2] and Andrew J Mort*[1]

Address: [1]Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK 74078, USA, [2]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA and [3]Department of Chemistry, Eastern Oklahoma State College, Willburton, OK 74578, USA

Email: Anamika Ray - aray@biochem.okstate.edu; Sunita Macwana - macwana@okstate.edu; Patricia Ayoubi - ayoubi@okstate.edu; Leo T Hall - lhall@eosc.edu; Rolf Prade - prade@okstate.edu; Andrew J Mort* - amort@biochem.okstate.edu

* Corresponding author

## Abstract

**Background:** The construction of cDNA libraries is a useful tool to understand gene expression in organisms under different conditions, but random sequencing of unbiased cDNA collections is laborious and can give rise to redundant EST collections.

We aimed to isolate cDNAs of messages induced by switching *Aspergillus nidulans* from growth on glucose to growth on selected polysaccharides. Approximately 4,700 contigs from 12,320 ESTs were already available from a cDNA library representing transcripts isolated from glucose-grown *A. nidulans* during asexual development. Our goals were to expand the cDNA collection without repeated sequencing of previously identified ESTs and to find as many transcripts as possible that are specifically induced in complex polysaccharide metabolism.

**Results:** We have devised a Negative Subtraction Hybridization (NSH) method and tested it in *A. nidulans*. NSH entails screening a plasmid library made from cDNAs prepared from cells grown under a selected physiological condition with labeled cDNA probes prepared from another physiological condition. Plasmids with inserts that failed to hybridize to cDNA probes through two rounds of screening (i.e. negatives) indicate that they are transcripts present at low concentration in the labeled probe pool. Thus, these transcripts will be predominantly condition-specific, along with some rare transcripts.

In a screen for transcripts induced by switching the carbon source from glucose to 12 selected polysaccharides, 3,532 negatives were isolated from approximately 100,000 surveyed colonies using this method. Negative clones were end-sequenced and assembled into 2,039 contigs, of which 1,722 were not present in the previously characterized glucose-grown cDNA library. Single-channel microarray hybridization experiments confirmed that the majority of the negatives represented genes that were differentially induced by a switch from growth in glucose to one or more of the polysaccharides.

**Conclusions:** The Negative Subtraction Hybridization method described here has several practical benefits. This method can be used to screen any existing cDNA library, including full-length and pooled libraries, and does not rely on PCR or sequence information. In addition, NSH is a cost-effective method for the isolation of novel, full-length cDNAs for differentially expressed transcripts or enrichment of rare transcripts.

# Background

High throughput analysis of differentially expressed genes has been widely used to address a multitude of biological questions. For such analysis, a large collection of cDNA molecules representing the potential genes of interest is useful. A variety of techniques have been used to identify the cDNAs representing genes of interest associated with various biological processes. Some of the techniques include characterization of expressed sequence tags (EST) [1], suppressive subtractive hybridization (SSH) [2,3], and representational difference analysis (RDA) [4].

EST sequencing first became popular in 1991 [1]. The basic scheme of an EST sequencing project relies on a cDNA library constructed from a tissue of interest under a particular condition from which randomly isolated clones are sequenced until further sequencing no longer yields an acceptable frequency of identifying novel cDNAs. A drawback faced in this method is the repeated sequencing of abundant transcripts and, hence, the expense and effort of sequencing them.

The abundance of mRNAs in a typical eukaryotic cell can be divided into rare transcripts present at approximately 15 copies or less per cell, moderately abundant transcripts, and abundant transcripts present at over 1,000 copies per cell [5]. The identification of rare mRNAs from a cDNA library, based on a random selection scheme, can be difficult because of their low representation. Construction of normalized cDNA libraries based on reassociation kinetics has been used to significantly reduce the representation of abundant transcripts, thereby increasing the chances of obtaining the rare cDNAs [5]. However, normalization often results in a bias towards small inserts [6]. Widely used techniques such as SSH can be useful for identifying gene expression differences at the mRNA level, but do not adequately address the problem of redundant transcripts [7].

Another technique called representational difference analysis (RDA) has been used to clone differentially expressed cDNAs [8]. However, this technique does not solve the problem of isolating rare mRNAs from the population of abundant mRNA species [9]. There are other methods such as mRNA differential display and RNA fingerprinting by arbitrary primed PCR but both of these methods are unsuitable for experiments where the expression levels of relatively few of the genes are expected to vary [10-13]. In addition, the sensitivity of these methods depends on the primer sequences, the concentration of the template, and its potential binding sites, and sometimes both of the above methods report a high number of false positives [14].

Here we report a simple and very effective method to isolate cDNAs for transcripts induced by changes in growth conditions or differentiation. For this study we were interested in isolating cDNAs induced by switching a fungus from growth on glucose to growth on selected polysaccharides. Approximately 4,700 contigs from 12,320 ESTs were already available from a cDNA library representing transcripts isolated from glucose-grown *A. nidulans* during asexual development [15]. Our goals were to expand the cDNA collection without repeated sequencing of previously identified ESTs and to find as many cDNAs as possible representing transcripts specifically induced in complex polysaccharide metabolism.

# Results

### *Isolation, sequencing, and classification of NSH-derived ESTs*

A non-normalized, non-amplified cDNA plasmid library was constructed from pooled RNAs extracted from *A. nidulans* grown on different polysaccharides (see materials and methods). If a fungal culture is given a mixture of complex polysaccharides the cells will most likely utilize one class of polysaccharide preferentially over another (Mort et al., unpublished results). Keeping this in mind, we grew fungal cultures individually in twelve different polysaccharides. RNA was extracted from each culture and pooled to make a composite cDNA library. We hoped these twelve different polysaccharides would represent all of the sugars and sugar linkages present in plant cell walls and would thereby induce a wide variety of enzymes necessary for the degradation and metabolism of the complex carbohydrates.

About 100,000 colonies from the cDNA library were screened with DIG-labeled cDNAs reverse transcribed from RNAs extracted from glucose-grown fungus. Colonies showing no or very faint hybridization signals (negatives), indicating potentially condition-specific transcripts, were selected and transferred to 96-well plates, and then subjected to secondary hybridization using the same probe. Out of a total of approximately 100,000 colonies plated, 8,000 negatives were counted and then finally 3532 (3.5% of those screened) well isolated colonies were selected as negatives for further analysis. Plasmid DNAs were isolated, the plasmid inserts end-sequenced and the sequence data were processed using PipeOnline (POL) [16,17], which generated a database containing 2,039 unique contigs assembled from overlapping sequences. All NSH derived EST sequences were deposited at NCBI's dbEST database [18] with accession numbers ranging from CK445320-CK449149 and CK468506-CK468532.

Fig. 1 shows the distribution of predicted functions of the proteins represented by the ESTs as classified by POL
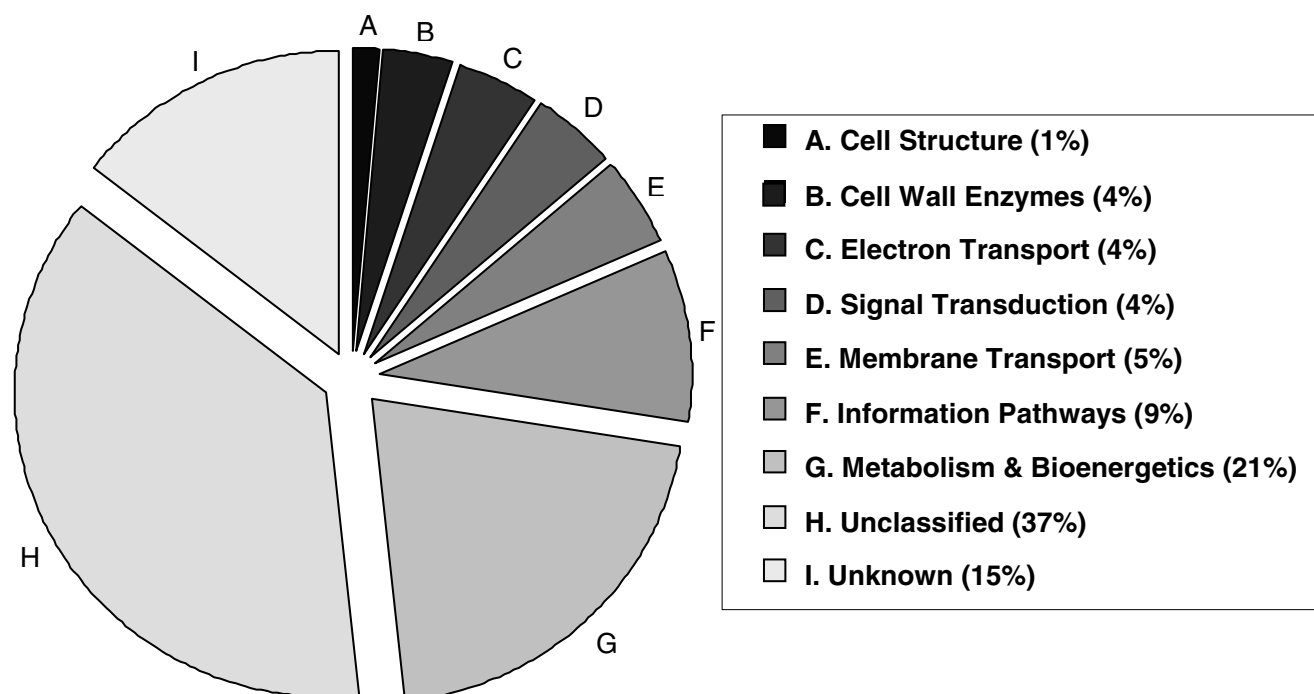
**A. Cell Structure (1%)**
**B. Cell Wall Enzymes (4%)**
**C. Electron Transport (4%)**
**D. Signal Transduction (4%)**
**E. Membrane Transport (5%)**
**F. Information Pathways (9%)**
**G. Metabolism & Bioenergetics (21%)**
**H. Unclassified (37%)**
**I. Unknown (15%)**

**Figure 1**
**Functional classification of ESTs isolated by NSH** NSH-ESTs were classified into the major functional categories according to PipeOnline [17]. The percentages indicate distribution of predicted EST functions in broad functional categories. All of the broad categories defined by PipeOnline are represented.

using the best BLASTX hit with an expectation value $E \leq$ 1e-3. The ESTs were grouped as follows: 1% into cell structure, 4% into cell wall enzymes, 4% into electron transport, 4% into signal transduction, 5% into membrane transport, 9% into information pathways, 21% into metabolism and bioenergetics, 37% were unclassified and 15% were unknown. Unclassified ESTs are those sequences that had a BLASTX hit with an expectation value $E \leq$ 1e-3, but no "known" function attributed to them. ESTs with no significant homology to known peptides were designated "unknown". From Fig. 1 it is evident that genes of unclassified and unknown function gave rise to half of the ESTs found in the NSH screen.

Fig. 2 shows the percentage of contigs containing a given number of NSH-ESTs plotted against the number of ESTs per contig. The majority (75%) of contigs contain one NSH-EST, which indicates low sequence redundancy in the NSH collection.

***Validation of the negatives using slide-based hybridization***
If the NSH method worked as predicted, one would expect most clones selected to contain cDNA inserts from genes not expressed in glucose-grown fungus but induced by forcing the fungus to utilize a complex carbohydrate as its carbon source or de-repressed after removal of glucose. To test this, clones representing about half of the contigs from our inventory of negatives which were also not present in the previous EST collection were PCR amplified and microarrayed on glass slides. A subset of the polysaccharides used for library construction namely pectin, arabinan, carboxymethyl cellulose, locust bean gum, and gum arabic were used separately as substrates for fungal cultures, and from those individual cultures labeled cDNAs were prepared. Cultures grown on glucose and in the absence of glucose (starved) were also used to prepare cDNAs. The cDNAs prepared from each condition were used in single channel non-competitive slide-based hybridizations.

For each carbohydrate tested, background-subtracted hybridization signals were normalized between slides
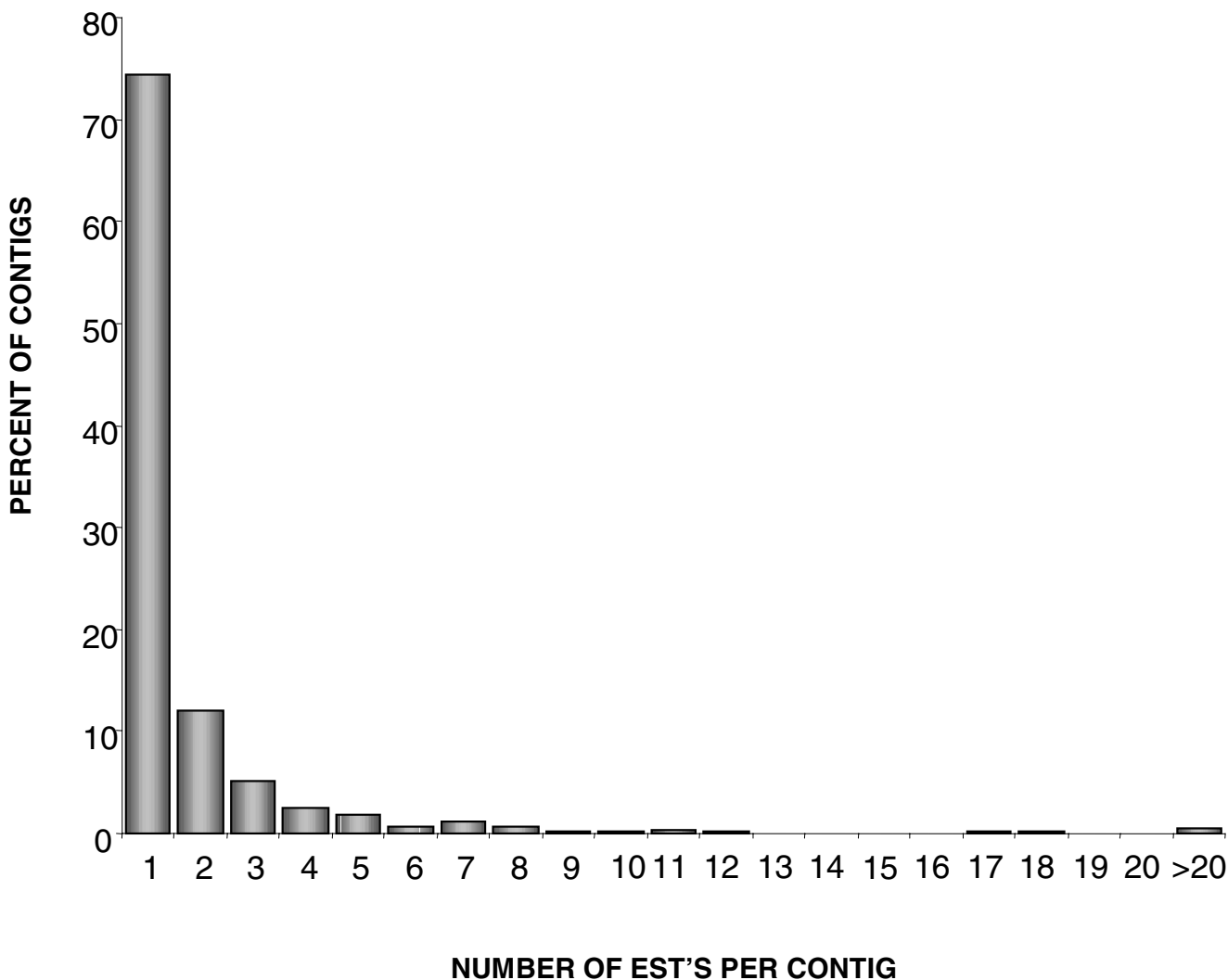
**Figure 2**
**Bar graph showing the redundancy rates of NSH-ESTs** The NSH-ESTs were assembled into contigs and the number of ESTs per contig was determined, against which was plotted the percentage of contigs containing that number of NSH-ESTs. Seventy-five percent of the contigs are composed of one NSH-EST.

based on the intensity of the controls and ranked from 0–3 depending on the normalized signal intensity. Fig. 3 summarizes all of the results by representing the average relative fluorescence intensity of each gene in each treatment by one of three shades of red or by black if it was deemed non-detectable. The order of the features appearing in the figure is relative to the intensity of signal obtained from glucose-grown labeled fungal cDNA (right-hand column) ranked from the highest to the lowest and secondarily by the sum of relative fluorescence intensity

levels in all the other hybridizations. The features can be grouped into four categories: 1) Those that appear to be expressed at high levels in glucose and at high or lower levels in starvation or complex carbohydrates 2) Those that are expressed to a low extent in glucose-grown cultures but are expressed to higher levels when grown in other conditions tested 3) Those that are non-detectable in the glucose-grown cultures but are expressed in one or more of the other conditions 4) Those that are apparently not expressed under any condition tested.
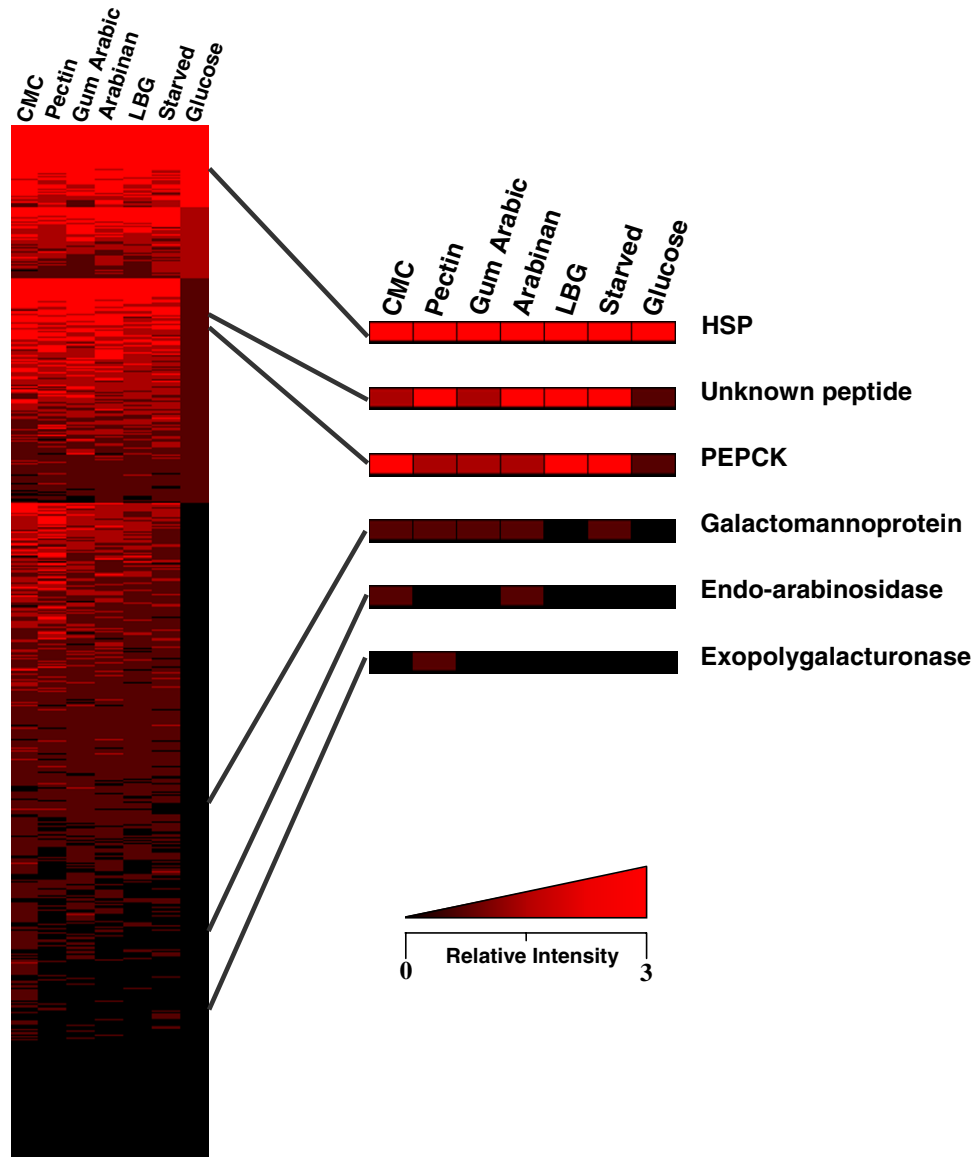
**Figure 3**
**Heatmap representation of normalized hybridization signals** For each carbohydrate tested, background-subtracted hybridization signals were normalized between slides to the intensity of the controls and then assigned a number 0–3 depending on the normalized intensity. Signal intensities of 0–4999 were assigned a number of 0 and were colored black in the heatmap; those 5000–19,999 were assigned a number of 1 and were colored light red; those 20,000–39,999 were assigned a number of 2 and were colored medium red; and any intensities above 40,000 were assigned a number of 3 and were colored bright red to assist in visualization of the hybridization signals. Signal intensities of the selected cDNAs coding for heat shock protein (HSP), unknown peptide, phosphoenolpyruvate carboxykinase (PEPCK), galactomannoprotein, endo-arabinosidase and exopolygalacturonase are shown on the right hand side of the diagram.

These results also revealed substrate-specific expression of genes. For instance, cDNA encoding exopolygalacturonase hybridized only to labeled cDNA made from pectin-grown fungal mycelium. Similarly, cDNA encoding endo-arabinanase hybridized only to labeled cDNA made from arabinan and arabinogalactan protein-grown fungal cultures but not to labeled cDNA extracted from fungal cultures grown in glucose (Fig. 3).

The intensity of hybridization to several cDNAs for which there was a relatively high redundancy in our NSH collection were investigated to determine whether they were actually highly expressed in fungal cultures grown in any of the conditions other than glucose. A cDNA encoding phosphoenol pyruvate carboxykinase (PEPCK), involved in gluconeogenesis, (picked 9 times during the NSH screen) hybridized at very high levels to labeled cDNA extracted from starved cultures but only weakly to labeled cDNA prepared from glucose-grown fungal cultures.

Twenty-one clones selected during the NSH procedure contained inserts with sequences matching that of a 30 kDa heat shock protein. A corresponding cDNA hybridized very strongly to labeled cDNAs from fungus grown on all carbon sources, including glucose. Another cDNA, sequenced 14 times and with an unknown function, hybridized to a lesser extent to labeled cDNA from glucose grown fungus compared to that of fungus grown on all of the complex polysaccharides or no carbon source. The cDNA encoding cell wall galactomannoprotein (sequenced 6 times in the NSH collection) gave no signal from glucose-grown fungus, but a low signal in starvation and three of the complex carbohydrates. There were also 80 features on the microarray which did not hybridize to any of the labeled cDNAs tested.

### Northern evaluation of the negative subtraction technique and microarray data

Several of the cDNAs reported above were labeled and used to perform northern analysis using RNA extracted from fungal cultures grown on all of the different carbon sources used for the library construction, glucose, and no carbon source, i.e. starvation to mimic growth on a complex carbon source the fungi could not digest. Fig. 4 shows these results. All of the results indicated that clones selected as negatives in the NSH and examined further by northern hybridization were indeed either not expressed or expressed only to a low extent in glucose-grown fungus, but induced to considerably higher levels in one or more of the complex carbohydrates or starved fungal cultures.

## Discussion

Our goal was to isolate cDNAs induced in *A. nidulans* grown on complex carbohydrates rather than glucose and to greatly expand the number of ESTs available for *A. nid-*

*ulans* without re-sequencing those generated from a glucose-grown conidiating culture [15]. In the previous EST project randomly picked clones were end-sequenced and 12,320 ESTs were assembled into 4,595 contigs. Some of the clones (especially some coding for heat shock proteins) were sequenced hundreds of times reflecting their high transcript abundance. Approximately 100,000 colonies from the cDNA library made from fungus growing on complex polysaccharides were screened with probes made from glucose-grown fungus. This allowed us to discriminate against colonies harboring cDNAs representing transcripts present in glucose-grown fungus. The screen eliminated ∼92% of the colonies from further analysis. This suggests that over 90% of the messages in the complex carbohydrate-grown fungus are common to fungus grown on glucose. We expect that a large fraction of the messages are from abundant transcripts. About half of the colonies which were negative in the initial screen were not picked because of lack of separation, small size, or failure to test negative in the secondary screen. From the original 100,000 colonies plated, 3,532 were finally picked for sequencing. After processing the 3,532 resulting ESTs, 2,039 unique contigs were obtained. Of these 1,772 had not been found previously within the glucose-grown conidiating library. Thus, the NSH method was very efficient in selecting for the desired cDNAs.

It had been estimated that *A. nidulans* codes for approximately 8,000 genes [19] although recently 9,500 ORFs have been predicted from the whole genome sequence [20]. The combination of these two EST libraries accounts for almost 70% of the predicted transcripts.

This technique was also successful in avoiding redundancy. For example, the cDNA encoding a particular 30 kDa heat shock protein was isolated 411 times during the random screening of the cDNA library made from glucose-grown conidiating fungus whereas it was found only three times during the NSH screen. It was most likely abundant in the original cDNA library but was removed through the NSH screening procedure.

Since our library was constructed from pooled RNAs of fungal cultures grown in twelve different plant polysaccharides, the isolated cDNAs could be from one or more conditions. To test if the isolated cDNAs are expressed in the manner expected (i.e. are induced by one or more of these polysaccharides but absent when grown on glucose) we selected unique cDNAs representing 728 contigs and hybridized them by conducting single-channel, slide-based microarray hybridization to labeled cDNAs made from fungal cultures grown on individual polysaccharides. Approximately 65% of the ESTs probed did not hybridize to labeled cDNAs from glucose-grown fungal cultures. Of these, some of the cDNAs were detected very
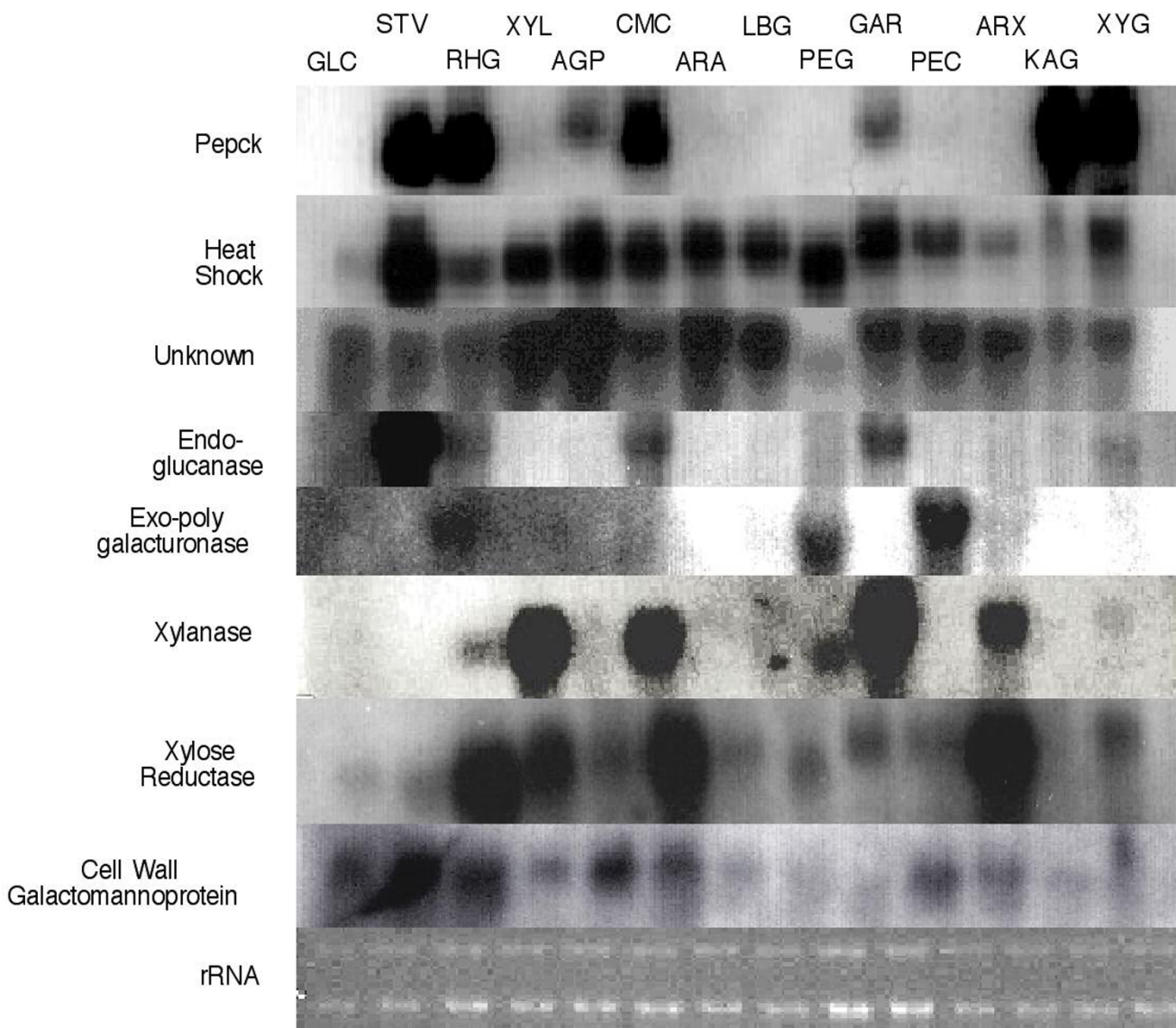
**Figure 4**
**Northern analysis** Northern blot expression pattern of *A. nidulans* mRNAs corresponding to the cDNAs encoding (from top to bottom) phosphoenolpyruvate carboxykinase (PEPCK), heat shock protein, protein with an unknown function, endogluca-nase, exopolygalacturonase, xylanase, xylose reductase, and cell wall galactomannoprotein. Total RNA was isolated from fungal mycelia and ~10 μg of RNA was separated on a 1% agarose/formaldehyde gel, blotted on nylon membranes, UV cross crosslinked and hybridized to cDNA probes as indicated. The order of total RNA in the gel from left to right are: glucose plus (GLC), glucose minus or starved (STV), rhamnogalacturonan (RHG), xylan (XYL), arabinogalactan protein (AGP), carboxy methyl cellulose (CMC), arabinan (ARA), locust bean gum (LBG), pectic galactan (PEG), gum arabic (GAR), pectin (PEC), arab-inoxylan (ARX), karaya gum (KAG), xyloglucan (XYG).

specifically in hybridization using labeled cDNAs pre-pared from cultures grown on single or a few different car-bon sources, but many were detected in hybridization using labeled cDNAs from a variety of cultures. This perhaps indicates a generalized shift in metabolism caused by the switch from glucose as carbon source to less readily metabolized polysaccharides with concomitant release from carbon catabolite repression which invites further study.

Although we initially designed the NSH system to find de-repressed and induced transcripts, it is clear from previous work [21] that transcripts which are of low abundance, including those in glucose-grown fungus, could be detected as negatives in this type of screen because the corresponding labeled cDNA used to screen the library would be at too low a concentration to produce a detectable signal. Nelson and coworkers screened macroarrays of randomly picked colonies of a cDNA library from human prostate tissue with labeled cDNA from the same tissue and selected the non-hybridizing colonies as rare transcripts [21]. It is therefore likely that some of the 80 "selected negatives" which did not hybridize in our single channel microarray studies also represent rare transcripts. It could also be possible that some of the 80 non-hybridizing arrayed negatives may be transcripts specific to one of the six carbon sources used in preparing the cDNA library but not tested in the microarray analysis.

Despite our efforts to eliminate transcripts found in glucose-grown fungus, approximately 35% of the ESTs from our collection of NSH-ESTs arrayed on the glass slides hybridized to labeled cDNAs extracted from fungus grown in glucose. Of these, one-quarter hybridized strongly under all nutritional conditions whereas three-quarters hybridized more strongly in some of the complex carbon sources compared to glucose. Selection of "false-negatives" can be attributed to experimental limitations common to hybridization screening. For example, colonies may have been in a region of the membrane that was not uniformly exposed to the labeled cDNA probe or mistakes could have been made in selecting negative colonies from over-crowded plates. Repressed growth of individual colonies or incomplete transfer of colonies during membrane lifts would also result in a reduction or complete loss of hybridization signal. These types of mistakes could account for the occasional selection of cDNA clones which gave a high degree of hybridization to glucose-grown fungal cDNA in microarray experiments.

A few of the clones showing intense hybridization to glucose-grown labeled fungal cDNAs used in the microarray experiments, were picked multiple times using NSH, thus making it unlikely that they were the result of repeated double selection during the screening. In addition, most of the false negatives were not present in the previously characterized EST collection made from a glucose-grown conidiating library. Some of the selected negative cDNA clones may have produced a weakly positive signal in northern or microarray hybridization experiments due to fairly small stretches of sequence similarity to transcripts present in glucose-grown fungus resulting in cross-hybridization. For example, in the combined EST libraries, we observed five different cDNA sequence contigs presumed to code for 30 kDa heat shock proteins, but the EST found

by NSH (derived from 21 cDNA sequences) contains an approximately 100 nucleotide stretch with high homology to the other four heat shock proteins. We believe this would be sufficient to allow cross-hybridization with different heat shock protein transcripts from glucose-grown fungal cDNA and yield a positive signal in microarray and northern hybridizations.

The question remains as to why it behaved as a negative during the NSH experiment yet was positive in northern hybridization. The cDNA probe concentration used in the various hybridization experiments could provide another factor in the selection of false negatives in NSH. Considering that the probes were applied in the hybridization methods at the same total DNA concentration, the effective concentration of a particular cDNA probe species was much lower in the NSH hybridization as it consisted of heterogeneous mixtures containing thousands of different sequences, compared to that used in northern blots which consisted of single, homogeneous cDNA probes. Several apparent discrepancies between NSH, northern blots and microarray hybridizations may also be related to basic technical differences and limitations of these methods. Further, while the NSH method worked well in this study, the potential for cross-hybridization within gene families is likely if applied to organisms with highly complex genomes.

## Conclusions

Some of the advantages of NSH are listed: This technique should be applicable to any preferably non-normalized cDNA library as long as it contains a low percentage of empty vectors which would, of course, appear as negatives. The absence of a PCR step in the construction of the library reduces the risk of disproportionate amplification of some sequences. There are no restriction digestion steps involved during the construction of the cDNA library, which lead to small insert size. Since the library used in the NSH method is made from potentially full-length cDNAs, each negative selected has a good chance of containing a full-length cDNA. It should be straightforward to automate most of the steps in the NSH procedure. Thus, NSH is an efficient method for isolation of cDNAs for differentially expressed and, very likely, rarely expressed transcripts.

## Methods

### *Isolation of RNA and construction of cDNA library*

The *A. nidulans* strain FGSC C26 (genotype: *biA1 veA1*) was inoculated at $10^6$ spores/ml of minimal medium containing 1% glucose and grown at 37°C for 18 hr with constant shaking at 300 rpm. Minimal medium with the appropriate supplements was prepared as described by Pontecorvo [22]. Fungal mycelia present in 100 ml of the culture were collected, washed with water and added to

different 250 ml flasks containing minimal medium supplemented with 1% of a single complex carbon source representing those found in plant polysaccharides. The fungal cultures were then grown at 37°C, pH 6.5 at 200 rpm for an additional 8 hr in minimal medium supplemented with one of the following carbon sources: carboxy methyl cellulose (Sigma), xyloglucan (Megazyme), rhamnogalacturonan (Megazyme), pectin (Sigma), karaya gum (Sigma), gum arabic (Sigma), locust bean gum (Sigma), arabinogalactan protein (provided by Dr. Jinhua An, Pharmagenesis), arabinoxylan (Megazyme), pectic galactan (Megazyme), xylan (Sigma) or arabinan (Megazyme). Total RNA from lyophilized mycelia of *A. nidulans* was isolated using Tri reagent (Life Technologies) based on the guanidinium thiocyanate-phenol-chloroform extraction method [23]. Equal amounts of the total RNA from cultures grown in each medium were combined to create a single pool of total RNA. A non-normalized cDNA library from the pooled RNA from *A. nidulans* was constructed by Life Technologies, Inc. using the vector, pCMVSport 6.0. The titer of the non-normalized library was $5.06 \times 10^6$ cfu/ml, and 98% of the clones from the library had inserts with the average insert size being 1.49 kb. The plasmid library was plated on 150 mm petridishes containing LB-Amp (100 µg/ml) such that each 150 mm plate had approximately 1,000 colonies.

### Screening of cDNA library and negative subtraction hybridization

Probes used for screening the cDNA library were made from cDNA reverse transcribed from total RNA of *A. nidulans* grown in Minimal Medium containing 1% glucose at 37°C for 18 hr, collected, washed and transferred to fresh medium of the same composition and grown for an additional 8 hr. The cDNAs were PCR amplified using Advantage® cDNA PCR Kit according to the manufacturer's instructions and labeled with the Digoxygenin (DIG) labeling and detection system from Roche Molecular Biochemicals using the random prime labeling method according to the manufacturer's instructions. DIG-labeled cDNAs used as probes were adjusted to a final concentration of 20 ng of probe/ml of hybridization buffer. The labeled probes were used to screen the cDNA library of *A. nidulans* by colony hybridization [24].

Stringency washes following hybridization were performed according to the manufacturer's instructions supplied by Roche Molecular Biochemicals. The membranes were exposed to the chemiluminescent substrate CDP-Star between two plastic sheets, allowed to incubate for 5 min, sealed in plastic bags and then exposed to Lumi-film (Roche Molecular Biochemicals) for detection. Exposure times of 5 min, 10 min and 20 min were taken. The X-Ray films were aligned to the nylon membranes attached in the autoradiography cassettes, then aligned to the LB-Amp

plates containing the colonies and the positively hybridized colonies were flagged. Those colonies which did not hybridize to the probe (cDNAs from glucose-grown fungus) were selected and grown in 200 µl of LB-Amp (100 µg/ml) in a 96-well plate at 37°C for 14–16 hr.

Cultures from the 96-well plates were gridded on nylon membranes (Amersham Pharmacia Biotech) [25] for secondary hybridization using the same probe as used in the primary hybridization. Glycerol stocks of the clones selected at the primary hybridization stage were also maintained.

The clones which were negative after the second round of hybridization were selected and 100 µl of their glycerol stocks were added to 1000 µl of Terrific Broth (Amp, 100 µg/ml) in 96-well culture blocks and grown at 37°C for 16 hr with shaking at 200 rpm. Plasmid DNAs were prepared using the 96-well alkaline lysis miniprep kit from Edge Biosystems according to the manufacturer's instructions. The DNAs were PCR amplified in 96-well thin-walled V-bottom microtitre plates (USA Scientific) using DyeDeoxy "Terminator PRISM" mix using the following program: 96°C for 30 s, 45°C for 15 s, 60°C for 4 min for 49 cycles [26]. The PCR products were purified to remove unincorporated dyes and primers using the 96-well gel filtration and purification kit from Edge Biosystems and sequenced on an ABI 3700 sequencer.

The raw sequences (ABI chromatograms) were processed using PipeOnline [16,17] for functional annotation. PipeOnline (POL) is a fully automated EST processing program designed to take raw sequence trace files as input, call bases, remove vector sequences, assemble contigs and annotate function to them wherever possible.

### Testing of the negatives using microarray experiments

Inserts from the clones from our inventory of negatives were PCR amplified using T7 and SP6 primers. Thermal cycling conditions consisted of an initial denaturation of 96°C for 3 min, followed by 35 cycles of 94°C for 30 s, 45°C for 45 s, 72°C for 1 min 30 s with a final extension at 72°C for 10 min. The quality of the PCR products was examined by running 5 µl of the products on 1% agarose gel where 98% of the PCR products revealed a single band of 500 bp or longer and were chosen for microarray analysis. The unpurified PCR products [27] were resuspended in Mirco Spotting Plus Solution (Telechem). They were printed at a final concentration of 250 ng/µl on amino-silane coated slides (Corning Cat # 40005) in quadruplicate at room temperature and 50% relative humidity using a PixSys 5500 microarrayer (Genomic Solutions) fitted with Majer Precision Pins. Various heterologous external controls, control sets from Ambion, as well as different negative (empty vectors) and internal positive controls

including pooled cDNAs which were PCR amplified from the non-subtracted cDNA library were used in the arrays. After printing, each slide was rehydrated by holding the slide with the array side down over a beaker of steaming water for 1 s and snap-dried on a warm hot plate. The slides were baked at 85°C for 3 hr followed by UV cross-linking using a Stratalinker. Prior to hybridization with labeled cDNAs, the features immobilized on the slides were denatured by immersing the slides in boiling water for 2 min followed by snap-cooling on a bed of ice. The slides were dried by brief centrifugation followed by pre-hybridization at 42°C in 0.5% SDS, 0.01% BSA, 6 X SSC and 25% formamide for 1 hr and washed with nanopure water at room temperature.

### Preparation of labeled cDNAs and the hybridization procedure

Total RNA was isolated from *A. nidulans* grown in different complex carbohydrates as described before, and 5 μl of RNA isolated from each condition was run in a denaturing formaldehyde/agarose gel. RNA samples with an $A_{260}/A_{280}$ ratio between 1.8 and 2.0 and sharp ribosomal RNA bands were selected for use in preparing labeled cDNAs for hybridization. Twenty-five micrograms of total RNA was used for labeling using Genisphere's Array 350 hybridization kit according to the manufacturer's instructions. Each hybridization for a particular condition tested was repeated three times for statistical validation. Labeled cDNAs for the replicate hybridizations were prepared from three independent fungal cultures. They were incubated at 80°C for 10 min followed by 42°C for 10 min before applying to the prewarmed and prehybridized array. The formamide based hybridization buffer (provided with the kit) used for the experiments enabled us to perform the hybridization at 42°C. A 22 × 40 mm cover slip (Grace Bio-Lab, Bend, OR) was carefully placed on the slide taking care not to create any bubbles, and the slides were incubated overnight in a CMT-Hybridization chamber (Corning Inc., Corning, NY). The humid atmosphere inside the chamber was maintained by applying 15 μl of 3 X SSC in the reservoir wells. The hybridization and the stringency washes were also done according to the manufacturer's instructions. The arrays were scanned using Scan Array Express from Perkin-Elmer.

### Image extraction and data analysis

Scanned images were analyzed using the software package, Gene Pix Pro 4.0 (Axon Inc.). Spots with signal intensities lower than background, scratched spots and spots covered with dust were flagged and excluded from further analysis. Local background was subtracted from the signal intensity of each spot on the array. For between-slide normalization, a normalization factor was calculated from the mean of the background subtracted median pixel intensities of the *A. nidulans* pooled non-subtracted

cDNAs which were used as internal controls. This normalization factor was then used in Gen Pix Pro 4.0 to normalize the feature signal intensities in each slide. Normalized signals from replicate spots within each treatment were averaged, and the values were used to determine the detection of cDNAs under particular conditions. Detection of cDNAs was done by comparing the normalized signal intensities of each spot to the mean of the background subtracted median pixel intensities of the negative controls. Visualization of the intensities of cDNAs across the wide spectrum of inducing substances was done by using GENESIS software [28]. All the array results are deposited at NCBI Gene Expression Omnibus under the platform accession number GPL566 and the series number GSE783 [29].

### Northern analysis

Northern blots were prepared following standard methods [24] using 10 μg of *A. nidulans* total RNA per lane. The cDNAs used as probes were selected from our collection of negative clones, digested with EcoRI and Hind III, run on a 1% agarose gel, gel-extracted (Qiagen Gel Extraction Kit) and labeled with DIG random prime labeling system from Roche Molecular Biochemicals according to the manufacturer's instructions. The membranes were hybridized with the labeled probe in DIG Easy Hyb at a final concentration of 20 ng/ml according to the manufacturer's instructions. Before re-use, each blot was stripped twice in 100 ml of boiling water containing 0.1% SDS and shaken on a rocking platform for 10 min, and washed twice in 1 X wash buffer (100 mM maleic acid pH 7.5; 150 mM NaCl; 0.3% (v/v) Tween 20) for 10 min.

## Authors' contributions

AR, SM and LTH optimized and performed the NSH procedure, PA developed the informatics. AR and SM performed the PCR amplification, gel analysis and quality control for printing the arrays; AR and PA printed the arrays, AR performed the microarray hybridizations. AR and PA carried out the analysis of the microarray results. AR carried out the northern blots. AR, PA, RP and AJM organized the design of the project and prepared the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

# References

1.  Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science* 1991, **252:**1651-1656.
2.  Diatchenko L, Chris Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD: **Suppression subtractive hybridization: A method for generating differentially regulated or tissue specific cDNA probes and libraries.** *Proc Natl Acad Sci* 1996, **93:**6025-6030.
3.  Diatchenko L, Lukyanov S, Lau YF, Siebert PD: **Suppression subtractive hybridization a versatile method for Identifying differently expressed genes.** *Methods Enzymol* 1999, **303:**349-380.
4.  Lisitysn N, Lisitysn L, Wigler M: **Cloning the differences between two complex genomes.** *Science* 1993, **259:**946-951.
5.  Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A: **Construction and characterization of a normalized cDNA library.** *Proc Natl Acad Sci* 1994, **91:**9228-32.
6.  Carninci P, Shibata Y, Hayatsu N, Sugahara Y, Shibata K, Masayoshi I, Konno H, Yasushi O, Masami M, Yoshihide H: **Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.** *Genome Res* 2000, **10:**1617-1630.
7.  Hedrick SM, Cohen DI, Nielsen EA, Davis MM: **Isolation of cDNA clones encoding T cell-specific membrane-associated proteins.** *Nature* 1984, **308:**149-153.
8.  Welford SM, Gregg J, Chen E, Garrison D, Sorensen PH, Denny CT, Nelson SF: **Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization.** *Nucleic Acids Res* 1998, **26:**3069-3065.
9.  Hubank M, Schatz DG: **Identifying differences in mRNA expression by representational difference analysis of cDNA.** *Nucleic Acids Res* 1994, **22:**5640-5648.
10. Liang P, Pardee A: **Differential display of the eukaryotic messenger RNA by means of the polymerase chain reaction.** *Science* 1992, **257:**967-970.
11. Welsh J, Chada K, Dalal SS, Ralph D, Cheng L, McClelland M: **Arbitrarily primed PCR fingerprinting of RNA.** *Nucleic Acids Res* 1992, **20:**4965-4970.
12. Bauer D, Warthoe P, Rohde L, Strauss M: **Detection and differential display of expressed genes by DDRT-PCR.** *PCR Methods Appl* 1994, **4:**S97-108.
13. Malhotra K, Foltz L, Mahoney WC, Schueler PA: **Interaction and effect of annealing temperature on primers used in differential display RT-PCR.** *Nucleic Acids Res* 1998, **26:**854-856.
14. Wan JS, Sharp SJ, Poirier GM, Wagaman PC, Chambers J, Pyati J, Hom YL, Galindo JE, Huvar A, Peterson PA, Jackson MR, Erlander MG: **Cloning differentially expressed mRNAs.** *Nature Biotechnology* 1996, **14:**1685-1691.
15. Prade RA, Ayoubi P, Krishnan S, Macwana S, Russell H: **Accumulation of stress and inducer-dependent plant-cell-wall-degrading enzymes during asexual development in *Aspergillus nidulans*.** *Genetics* 2001, **157:**957-967.
16. Ayoubi P, Jin X, Leite S, Liu X, Martajaja J, Abduraham A, Wan Q, Yan W, Misawa E, Prade R: **Pipeonline 2.0: automated EST processing and functional data sorting.** *Nucleic Acids Res* 2002, **30:**4761-4769.
17. **The PipeOnline web site** [http://bioinfo.okstate.edu/pipeonline]
18. **The NCBI dbEST web site** [http://www.ncbi.nlm.nih.gov/dbEST]
19. Kupfer DM, Reece CA, Clifton SW, Roe BA, Prade RA: **Multicellular ascomycetous fungal genomes contain more than 8000 genes.** *Fungal Genet Biol* 1997, **21:**364-372.
20. **The Center for Genome Research *Aspergillus nidulans* Database** [http://www-genome.wi.mit.edu/annotation/fungi/aspergillus/]
21. Nelson P, Hawkins V, Schummer M, Bumgarner R, Ng WL, Ideker T, Ferguson C, Hood Leroy: **Negative selection: a method for obtaining low-abundance cDNA clone arrays.** *Genetic Analysis: Biomolecular Engineering* 1999, **15:**209-215.
22. Pontecorvo G: **Genetic analysis of "somatic" cells in filamentous fungi.** *Wistar Inst Symp Monogr* 1969, **9:**1-8.
23. Chomczynski P: **A reagent for the single step simultaneous isolation of RNA, DNA and proteins from tissue samples.** *Biotechniques* 1993, **15:**532-534.
24. Sambrook J, Fritsch EF, Maniatis T: In *Molecular Cloning: A laboratory manual* Cold Spring Harbor; 1989.
25. Buitkamp J, Kollers S, Durstewitz G, Fries R, Welzel K, Schafer K, Kellermann A, Lehrach H: **Construction and characterization of a gridded cattle BAC library.** *Animal Genetics* 2000, **31:**347-351.
26. Gong Z, Koiwa W, Cushman MA, Ray A, Bufford D, Kore-eda Shin, Matsumoto TK, Zhu Jianhua, Cushman JC, Bressan RA, Hasegawa PM: **Genes that are uniquely stress regulated in salt overly sensitive mutants.** *Plant Physiol* 2001, **126:**363-375.
27. Diehl F, Beckmann B, Kellner N, Hauser NC, Diehl S, Hoheisel JD: **Manufacturing DNA microarrays from unpurified PCR products.** *Nucleic Acids Res* 2002, **30:**16-e79.
28. Sturn A, Quackenbush J, Trajanoski Z: **Genesis: Cluster Analysis of microarray data.** *Bioinformatics* 2002, **18:**207-208.
29. **The NCBI GEO web site** [http://www.ncbi.nlm.nih.gov/geo]