Accumulation of Stress and Inducer-Dependent Plant-Cell-Wall-Degrading Enzymes During Asexual Development in *Aspergillus nidulans*

Rolf A. Prade, Patricia Ayoubi, Shobana Krishnan, Sunita Macwana and Hugh Russell

Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma 74078-3020

Manuscript received October 10, 2000
Accepted for publication December 15, 2000

ABSTRACT

Determination and interpretation of fungal gene expression profiles based on digital reconstruction of expressed sequenced tags (ESTs) are reported. A total of 51,524 DNA sequence files processed with PipeOnline resulted in 9775 single and 5660 contig unique ESTs, 31.2% of a typical fungal transcriptome. Half of the unique ESTs shared homology with genes in public databases, 35.8% of which are functionally defined and 64.2% are unclear or unknown. In *Aspergillus nidulans* 86% of transcripts associate with intermediate metabolism functions, mainly related to carbohydrate, amino acid, protein, and peptide biosynthesis. During asexual development, *A. nidulans* unexpectedly accumulates stress response and inducer-dependent transcripts in the absence of an inducer. Stress response genes in *A. nidulans* ESTs total 1039 transcripts, contrasting with 117 in *Neurospora crassa*, a 14.3-fold difference. A total of 5.6% of *A. nidulans* ESTs implicate inducer-dependent cell wall degradation or amino acid acquisition, 3.5-fold higher than in *N. crassa*. Accumulation of stress response and inducer-dependent transcripts suggests general derepression of *cis*-regulation during terminal asexual development.

ANALYSIS of gene expression was pioneered by the Northern blot technique in 1977 (Alwine *et al.* 1977). Expression patterns are investigated by the interpretation of mRNA (Berk and Sharp 1977) or protein-derived expression maps. mRNA expression maps are currently constructed from expressed sequenced tags (ESTs; Schmitt *et al.* 1999; Ohlrogge and Benning 2000) or evaluated from microarray chips (Cho and Campbell 2000; Richmond and Somerville 2000).

The study of differences in gene expression patterns is a promising approach for genetic, biochemical, cellular, and morphogenetic systems. Because it is now technologically feasible to measure global gene expression levels, it is possible to determine precise gene/function outlines affected by broad environmental cues, determine the onset of morphogenesis, or establish gene clusters involved in cellular processes and regulatory networks (Doebley and Lukens 1998; Tavazoie *et al.* 1999).

Comprehensive gene expression survey methodologies arrange into analog and digital resolution of gene expression levels. Analog methods are based on physical measurements of DNA/DNA or RNA/DNA hybridization between probes and tags (gene-specific DNA fragments or oligonucleotide collection) while digital methods derive expression levels from absolute counts of randomly generated tags from large condition-, stage-, organ-, or tissue-specific cDNA populations (Audic and Claverie 1997).

EST-based gene expression analysis requires a few assumptions. Underlying digital expression profiling is the assumption that *in vivo* populations of a given gene transcript from a particular tissue or organ of origin are proportionally represented in *in vitro* synthesized cDNA libraries from which tags are randomly sampled and sequenced. Thus, counting ESTs and relating them to the total sequenced population of ESTs provides absolute estimates of mRNA expression levels (Kozian and Kirschbaum 1999). The requirement for EST expression maps is random sequencing of nonnormalized libraries. As a result, qualified EST data sets provide exact transcriptional profiles and quantitatively compare mRNA expression levels of different organisms.

There are more than 5,609,655 entries in dbEST (database of Expressed Sequence Tags) release September 8, 2000 sampled from more than 5318 libraries (June 21, 2000). This public source of data retains information of genetic significance supporting the understanding of specific biological processes and phenomena.

Relative frequency variations of condition-, stage-, tissue-, or organ-specific tags (of any kind) is the basis for determination of differential gene expression levels in one, a few, or all genes of a given genome. Accuracy and range of expression profiles are dependent on limitations imposed by the method of choice. For digital expression profiles, the size of the EST collection under consideration needs to be validated through statistical models (Audic and Claverie 1997).

Fungi have simple and mostly nonredundant ge-

Corresponding author: R. A. Prade, Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078-3020. E-mail: prade@okstate.edu

nomes, with *Saccharomyces cerevisiae* being the exception (Seoighe and Wolfe 1998). Most fungal genomes are haploid in the predominant phases of the life cycle; gene redundancy is observed but is not common, and pseudogenes are rare. Thus, a single-pass DNA sequence derived from RNA is a reliable source for quantitative and qualitative analysis.

The number of genes encoded by fungal genomes has been estimated by a number of methods and varies from 5800 for *S. cerevisiae* to 8100 or 9200 for *Aspergillus nidulans* and *Neurospora crassa*, respectively (Kupfer *et al.* 1997; Kelkar *et al.* 2001; Smulian *et al.* 2001).

It has become increasingly apparent that expression of morphogenetic programs, activation or inactivation of physiological processes, and structural reorganization of cellular components are exerted through rearrangements of gene expression levels of entire transcriptomes, and where observations have been made, they seem to follow modular implementation (Doebley and Lukens 1998). Modular implementation of transcripts coding structural and metabolic functions points toward the predominance of coordination through *cis*-regulatory elements.

In this study partitioned cellular functions are correlated with expression levels for genes with rationalized activities. Transcripts with unclear or unknown functional assignments are not considered. This information is examined, and metabolic conditional profiles are reconstructed from various fungal tissues on the basis of digital associations of function with digital expression data gathered from large EST collections.

Initially, different fungi are shown parsing broad functional categories into quantitatively similar clusters, regardless of tissue, organ, or physiological origin. However, when specific functional clusters were analyzed within a functional category, it is found that during *A. nidulans* asexual development, a significant number of transcripts are related to stress responses and metabolism, suggesting preventive or protective roles.

Interestingly, there is a significant and revealing number of inducer-dependent transcripts in the absence of inducer at the time the mRNA sample was collected. To explain inducer-dependent transcripts in the absence of inducer and stress response transcripts in the absence of an obvious source of stress, it is proposed that during the final stages of asexual development, vegetative *cis*-acting regulatory networks are no longer functional and result in general derepression. These findings indicate that during asexual development, *A. nidulans* overrides vegetative regulatory controls and produces a series of transcripts that may add protective and adaptive advantages to the dormant spore.

MATERIALS AND METHODS

Fungal EST data manipulation: ESTs and source descriptions were obtained from the National Center for Biotechnology Information dbEST using batch-ENTREZ (Schuler *et al.* 1996). Only EST collections of fungal origin exceeding 1000 records were used in this study. ENTREZ-downloaded, FASTA-formatted files were processed using PipeOnline, an EST optimized DNA sequence analysis, and a database construction package with automated gene function annotation (described by Ayoubi, and available from the http://aspergillus-genomics.org website). Edited, assembled, and annotated ESTs can be inspected using GeneBrowser, surveyed with simple or structured queries, and results can be retrieved as FASTA- or TAB-delimited files. Databases used for this study are available from the aspergillus-genomics.org website.

Functional annotation: PipeOnline utilizes the ontological functional organization and the gene names (function) from the Metabolic Pathways database (MPW; Selkov *et al.* 1998; Overbeek *et al.* 2000). The MPW dictionary is grouped into six basic categories, each of which is subdivided into two to nine subcategories containing 2727 nonredundant standard enzyme, protein, and gene names at the lowest level. Functional assignment of records in a PipeOnline database is accomplished through a predefined gene name/gene index lookup table. The functional annotating algorithm identifies a function independent from BLASTX homolog description matching. A detailed description of the functional sorting process is being published elsewhere and is available from the http://aspergillus-genomics.org website.

Quantitative analysis and functional distribution: Each file (also identified as a clone in this study) downloaded from the dbEST database was treated as being a single-pass cDNA sequence derived from a randomly chosen clone of a nonnormalized source cDNA library. Table 1 contains additional biological information about the libraries. The term "contig" means multiple input files from a single organism and/or library assembled into a single, homologous overlapping consensus output sequence using default PHRAP arguments; *i.e.*, minmatch = 14, minscore = 30, and maxgap = 30 (compared with other assembly programs in Chen and Skiena 2000). Singlets (clones) depict single-pass, nonhomologous, and nonoverlapping sequencing reads.

Clone equivalency from differentially sized clone collections: Quantitative comparisons between *N. crassa* and *A. nidulans* whose EST collections are of different sizes were done by using a clone equivalency conversion factor on the basis of the discussion by Audic and Claverie (1997). One clone from the *A. nidulans* EST collection (12,485) equals 1.5 clones from the *N. crassa* (20,172) library. Conversely, one clone from the *N. crassa* library equals 0.7 clones in *A. nidulans*. These equivalency factors fall within a confidence interval >0.993 and <0.994.

Assembly of digital expression profiles: The FASTA- and TAB-delimited export features of PipeOnline were used extensively to download data from functionally categorized queries and upload into a local spreadsheet program for human expert verification, validation, and final tag counting.

RESULTS

Evaluation of EST-derived DNA sequence information from fungal origin: Table 1 provides a summarized overview of all publicly available fungal EST collections (retrievable from GenBank, dbEST subset). Employing PipeOnline (available from the http://aspergillus-genomics.org website), out of 51,524 submitted ESTs belonging to eight fungal species, 15,435 unique (nonhomologous) sequence files were recovered, of which 5660 (36%) were contigs and 9775 (64%) were singlets.

TABLE 1

EST-derived DNA sequence information from fungal origin in the public domain

		Database DNA sequence content						
		Public records (GenBank)						
		dbEST files	Unique ESTs			Ratios		
Fungus	Relevance	dGB	Contigs C	Singlets S	Total U	U/dGB	C/dGB	EST source library description
N. crassa	Model system	20,172	1,292	850	2,142	0.11	0.06	Nonnormalized, morning (tissue harvested after 22-hr growth in light) and evening (tissue harvested after 22-hr growth in dark)
A. nidulans	Model system	12,485	1,730	2,865	4,595	0.37	0.14	Nonnormalized, 24-hr asexual developmental and vegetative cDNA lambda zap library
S. pombe	Model system	8,157	953	2,051	3,004	0.37	0.12	Nonnormalized, late-log phase cDNA
Pneumocystis carinii	Human pathogen	3,896	666	984	1,650	0.42	0.17	Nonnormalized, *P. carinii* organisms (3×10e9) from a single rat
S. cerevisiae	Model system	2,811	504	1,094	1,598	0.57	0.18	Nonnormalized, *S. cerevisiae* strain X2180-1A from TIGR
M. anisopliae	Insect pathogen	1,693	158	690	848	0.50	0.09	Nonnormalized, grown on insect cuticle for 24 hr
M. graminicola	Plant pathogen	1,158	208	592	800	0.69	0.18	Nonnormalized, constructed from cultures utilizing ammonium as a source of nitrogen
M. grisea	Plant pathogen	1,152	149	649	798	0.69	0.13	Nonnormalized, grown at 23° in the dark with constant gyratory shaking (100 rpm) in Vogel's medium containing 0.5% isolated rice cell walls as the sole carbon source
Total:	Fungi	51,524	5,660	9,775	15,435	0.30	0.11	ESTs available from public databases
Average:	Fungi	6,441	708	1,222	1,929	0.46	0.13	U and C ratios are library indicators, reporting occurrence of DNA sequence diversity and redundancy, respectively

GenBank is a public repository for DNA sequence information maintained and curated by the National Center for Biotechnology Information, National Institutes of Health (http://www.ncbi.nim.nhi.gov); dGB, GenBank public files available from the dbEST, release 090800; EST, expressed sequence tag; C, contigs; S, singlets; U, unique sequence files.

TABLE 2

Automated functional annotation of fungal ESTs

Surveyed EST libraries	ESTs			POL_db record with hsp >100[b]					
				Total		Function		No function	
	All	Unique	% of genes[a]	No.	%	No.	%	No.	%
N. crassa	20,172	2,142	23.3	842	39.3	333	39.5	509	60.5
A. nidulans	12,485	4,595	56.7	1,863	40.5	831	44.6	1,032	55.4
S. pombe	8,118	3,003	42.9	1,319	43.9	346	26.2	973	73.8
P. carinii	3,896	1,649	41.2	976	59.2	294	30.1	682	69.9
S. cerevisiae	2,799	1,587	27.4	1,343	84.6	373	27.8	970	72.2
M. anisopliae	1,693	847	—	514	60.7	213	41.4	301	58.6
M. graminicola	1,157	799	—	462	57.8	205	44.4	257	55.6
M. grisea	1,058	727	8.6	372	51.2	159	42.7	213	57.3
Total:	51,378	15,349	31.2	7,691	50.1	2,754	35.8	4,937	64.2
Average:	6,422	1,919		961		344		617	

[a] Gene number based on the assumption that all unique ESTs represent different genes. Genome complexity estimates are from Kupfer *et al.* (1997), Kelkar *et al.* (2001), and Smulian *et al.* (2001).

[b] POL_db; PipeOnline automated DNA sequence processing and functional sorting package are described by Ayoubi and available from the http://aspergillus-genomics.org website.

The unigene (U/dGb) and multiple sequence assembly (C/dGb) ratios are indicators asserting DNA sequence diversity and redundancy, respectively. For example, from a total of 51,524 fungal EST files (clones), 30% (U ratio) are unique, and 11% (C ratio) have been sequenced at least twice. U and C ratio comparisons among EST collections account for representation differences between cDNA libraries and the randomness of each sequencing effort, respectively. U and C ratios are noninformative when the total number of EST files is low. The *N. crassa* (U/dGB 0.11), *Mycosphaerella graminicola* (U/dGB 0.69), and *Magnaporthe grisea* (U/dGB 0.69) EST collections deviate from the average 0.46 U/dGB ratio.

Table 2 displays an assessment reproducing probable transcriptome coverage with fungal ESTs and aggregation of biochemical information assigned to these DNA sequences on the basis of the PipeOnline automated functional assignment algorithm. On average, 31.2% of a typical fungal transcriptome is represented in the listed databases, whereas the *A. nidulans* EST collection retains the largest number of the predicted transcriptome (56.7%) and *M. grisea* keeps the least, at 8.6%. On average, half (50.1%) of the unique ESTs share homology [high scoring pair (hsp) >100] with other entries in GenBank; 35.8% of those are functionally annotated by PipeOnline, and 64.2% of unique ESTs remain unclear or have no homolog (hsp < 100).

Through the analysis of the functional outline of EST collections, this study detects qualitative and quantitative differences in functional content between collections and suggests stage-specific properties unique to *A. nidulans* asexual development and vegetative growth in *N. crassa*. All potential functions that could not be found in the *A. nidulans* EST library employing the PipeOnline functional sorting algorithm were retrieved and are displayed in Table 3. It was found that the vast majority of missing functions are related to active transport of amino acids, nucleotides, sugars, and other basic components missing from the minimal medium from which the source tissue was harvested. Functions other than transport genes that were found missing include DNA adenine methylation, tRNA anabolism genes, seven group tRNA methyltransferases, nitrate respiration, signal peptide trimming, and catabolism genes.

At first, it was surprising to discover that transport functions were missing in *A. nidulans* ESTs; however, the tissues that served as sequencing templates were grown in mineral, glucose-only-containing medium (Aramayo and Timberlake 1990), affirming this unexpected result. Moreover, this strongly indicates that false annotations are not common, thus authenticating digital profiling based on transcript counting and automated annotation.

Metabolic profile of *A. nidulans*: Table 4 shows the overall metabolic activity distributed into functional categories among the three largest EST collections, *A. nidulans*, *N. crassa*, and *S. pombe*. All three collections sorted roughly into a similar functional pattern with intermediate metabolism accounting for ~84% (82% the lowest and 86% the highest) of all surveyed transcripts. Also included were information pathways, transmembrane transport, signal transduction, and electron transport accounting for 8, 2, 3, and 3% of the transcripts, respectively. The *A. nidulans*, *N. crassa*, and *Schizosaccharomyces pombe* ESTs sorted all functional categories into similar patterns.

Figure 1 shows that in *A. nidulans* 86% of all the

TABLE 3

Biochemical functions missing in the *A. nidulans* ESTs

Category	Function
Transport-related functions	
Active transport	33
ATP-driven alanine transport, ATP-driven proline transport, ATP-driven oligopeptide transport, ATP-driven arginine transport, ATP-driven phenylalanine transport, ATP-driven arabinose transport, ATP-driven cystine transport, ATP-driven serine transport, ATP-driven lactose transport, ATP-driven glutamine transport, ATP-driven threonine transport, ATP-driven ribose transport, ATP-driven glycine transport, ATP-driven tryptophan transport, ATP-driven betaine transport, ATP-driven histidine transport, ATP-driven tyrosine transport, ATP-driven lipoate transport, ATP-driven isoleucine transport, ATP-driven valine transport, ATP-driven pantothenate transport, ATP-driven leucine transport, ATP-driven heme transport, ATP-driven pyridoxine transport, ATP-driven lysine transport, ATP-driven putrescine transport, ATP-driven riboflavin transport, ATP-driven methionine transport, ATP-driven spermidine transport, ATP-driven thiamin transport, ATP-driven ornithine transport, ATP-driven dipeptide transport, ATP-driven ribosylnicotinamide transport	
Transport of ions	12
Proton transport, ATP-driven co(′2+) pump, ATP-driven molybdate transport, hg(′2+) pump, k(′+) pump, phosphate transport, ATP-driven cu(′2+) pump, na(′+) pump, transport of halogen ions, k(′+)/cu(′2+)-dependent ATPase, Mit. ATP-driven proton pump, sulfate transport	
Secondary active transport	75
H(′+)/phosphate symporter, H(′+)/guanosine symporter, imp/phosphate antiporter, Na(′+)/aspartate symport, H(′+)/deoxyguanosine symporter, tmp/phosphate antiporter, Na(′+)/L-glutamate symport, H(′+)/cytidine symporter, dtmp/phosphate antiporter, H(′+)/L-leucine symport, H(′+)/deoxycytidine symporter, ump/phosphate antiporter, H(′+)/L-lysine symport, H(′+)/uracil symporter, dump/phosphate antiporter, H(′+)/L-ornithine symport, H(′+)/thymidine symporter, cmp/phosphate antiporter, Na(′+)/glutathione symport, H(′+)/deoxythymidine symporter, dcmp/phosphate antiporter, Na(′+)/glucose symport, H(′+)/uridine symporter, citrate/lactate antiporter, H(′+)/2-keto-3-deoxygluconate symport, H(′+)/deoxyuridine symporter, decanoate/h(′+) antiporter, H(′+)/lactose symport, H(′+)/xanthine symporter, H(′+)/methylamine antiporter, H(′+)/mannose symport, H(′+)/hypoxanthine symporter, K(′+)/h(′+) glutathionereg antiporter, Na(′+)/melibiose symport, H(′+)/allantoine symporter, erythromycin/h(′+) antiporter, H(′+)/sucrose symport, H(′+)/cyaNate symporter, ethidium bromide/h(′+)antiporter, H(′+)/xylose symport, glycerol-facilitated diffusion, H(′+)/actinomycin d antiport, Na(′+)/citrate symport, nitrate/nitrite-facilitated diffusion, H(′+)acriflavin antiport, Na(′+)/fumarate symport, propanedio-facilitated diffusion, H(′+)vancomycin antiport, H(′+)/lactate symporter, k(′+)/H(′+) symport, H(′+)/crystal violet antiport, Na(′+)/malate symport, glycerol 3-pho/pho antiporter, Ca(′2+)/h(′+) antiporter, Na(′+)/succinate symport, arginine/ornithine antiport, Co(′2+)/h(′+) antiporter (via "cnrabc"), Na(′+)/hco3(′−) symport, cadaverin/lysine antiporter, Co(′2+)/h(′+) antiporter (via "czcabc"), H(′+)/phosphate symporter, putrescin/ornithine antiporter, Cd(′2+)/h(′+) antiporter, Na(′+)/phosphate symport (reNal), amp/phosphate antiporter, Ni(′2+)/h(′+) antiporter, H(′+)/adenosine symporter, damp/phosphate antiporter, Zn(′2+)/h(′+) antiporter, H(′+)/deoxyadenosine symporter, dgmp/phosphate antiporter, ammonium uniporter	
Group translocation transport systems	3
Phosphotransferase system (pts), binding-protein-dependent phosphonate	
Other functions	57
Transport of purine bases and nucleosides, coenzyme f420 reduction, pyridoxal 5′-phosphate-dep. inactivation, DNA adenine methylation, electron transport site 0, glycoprotein biosynth and palmitoylation, tRNA isopentenyltransferase, glycerophosphate shuttle, nitrogen reg. proteins covalent modification, tRNA methyltransferases (7 groups), oxidized glutathione reduction, calmodulin lysine *n*-methyltransferase, tRNA-cca turnover, trypanothione reduction, cytochrome-"*c*"-lysine *n*-methyltransferase, peptidyl-tRNA degradation, protein-disulfide reduction, signal peptide trimming, selenocysteinyl-tRNA biosynthesis, 6-phosphofructokinase reduction, signal peptide catabolism, queuine tRNA-ribosyltransferase, isocitrate dehydrogenase phosphorylation, lipoprotein maturation, cysteinyl-tRNA anabolism, nitrogen reg. proteins covalent modification, formylmethionine catabolism, glutaminyl-tRNA anabolism, holo-dihydrolipoamide acetyltransferase, flavodoxin reductase reaction, leucyl-tRNA anabolism, holo-dihydrolipoamide succinyltransferase, 5,10-methyltetrahydrofolate oxidation, methionyl-tRNA anabolism phenylalanyl-tRNA anabolism, lipoate reduction, formylmethionyl-tRNA anabolism threonyl-tRNA anabolism, NADPH cytoch P-450 oxidoreductase, tyrosyl-tRNA anabolism NH(4)(′+)-dep. dinitrogenase reductase, NADH-camphor 5-hydroxylase electron transp, valyl-tRNA anabolism, protein disulfide-isomerase reaction, trimetylamine-"*n*"-oxide reduction, asparaginyl-tRNA anabolism, ribosomal-protein s5 alanine acetylation, nitrate respiration, lysyl-tRNA anabolism, rib. protein s6 C terminus glutamylation, rubredoxin reductase reaction, as(5+) reduction, ribosomal protein s11 methylation, trimetylamine-"*n*"-oxide reductase reaction, hg(2+) reduction, acyl-carrier protein anabolism	

TABLE 4

Metabolic activity distribution of three fungi among major functional categories

	A. nidulans		*N. crassa*		*S. pombe*		Total	
Metabolic category	Clones	%	Clones	%	Clones	%	Clones	%
Intermediate metabolism	7,769	87	6,490	82	5,354	84	19,613	84
Information pathways	542	6	648	8	622	10	1,812	8
Transmembrane transport	244	3	262	3	73	1	579	2
Signal transduction	283	3	208	3	156	2	647	3
Electron transport	240	3	355	4	184	3	779	3
Total:	9,078	100	7,963	100	6,389	100	23,430	100
ESTs (%):	73		39		79		57	

transcripts correlate with functions that fall within the category of intermediate metabolism. Within that category, half (47%) of the transcripts encode functions related to synthesis and degradation of carbohydrate and 16% encode functions related to amino acid, peptide, and protein metabolism.

Carbohydrate synthesis (58%) is the most active portion of the intermediate metabolism category. Main carbohydrate pathways account for 21% of the activity and production of mono-, di-, and polysaccharides: 17, 4, and 9%, respectively. Anabolic processing of sugars such as the production of sugar alcohol, alcohols, and organic acids account for 37% of carbohydrate-related activity. Production of aminosugars and other carbohydrates fill the remaining 10%.

Amino acid anabolism is another predominant activity (~10% of intermediate metabolism) while a moderate amount (4.3%) of recycling via proteolysis can be detected. The contributions of other categories—nitrogen, sulfur, and phosphorus metabolism, fatty acids membrane and related metabolite production, vitamins, heme, coenzymes, and other prosthetic groups—is modest, amounting to 35% of the overall metabolic activity.

Misappropriated gene expression: During the functional distribution analysis of the *A. nidulans* ESTs transcripts were noticed whose predicted function indicates

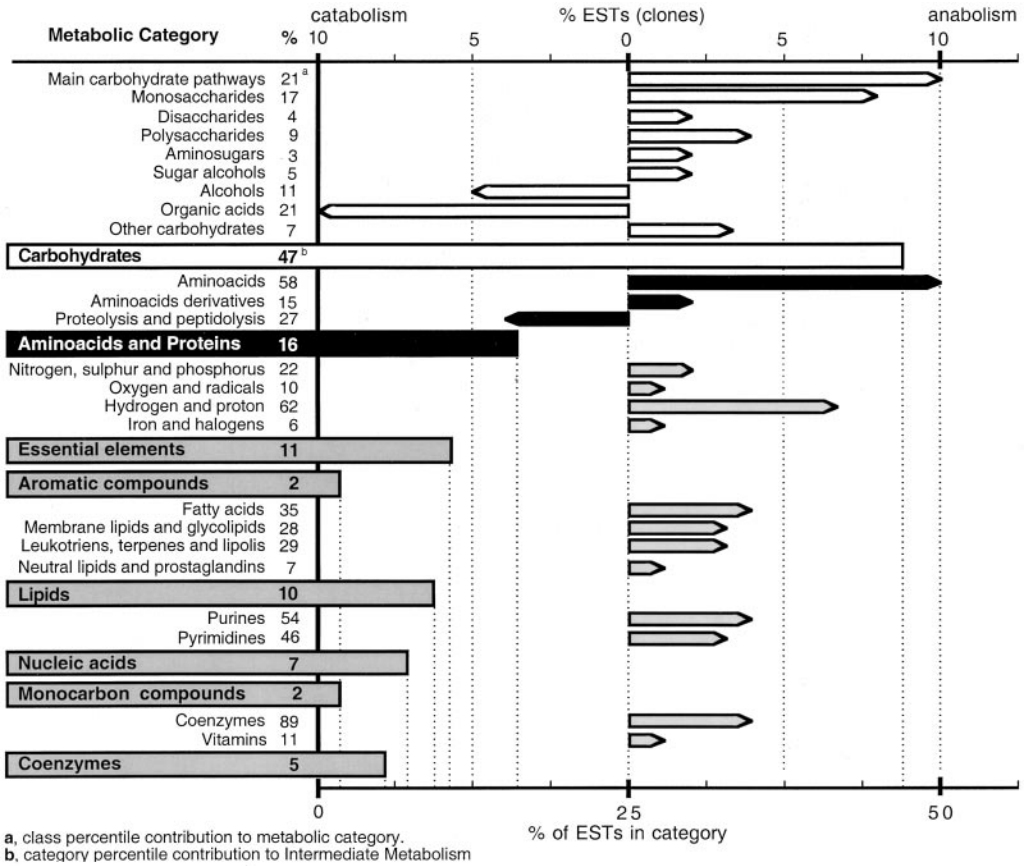

FIGURE 1.—Intermediate metabolism functional distribution of *A. nidulans* asexually developing cultures. Metabolic activity is expressed as a function of the number of ESTs (all clones) sequenced with functions that classify into the depicted major metabolic groups. Bars with arrowheads indicate the biochemical trend and anabolism of catabolism within groups that make a category. Blunt-ended bars indicate the total contribution of a given category.

TABLE 5

Misappropriated inducer-dependent transcripts in *A. nidulans* and *N. crassa*

Misappropriated inducer-dependent transcripts (MIT)	Transcript count				Induction (fold)
	A. nidulans		*N. crassa*		
	Clones	% MIT	Clones	% MIT	
Inducer-dependent transcripts					
Plant cell wall catabolism					
Pectin-degrading enzymes	20	2.9	—	—	—
Cellulases	15	2.1	—	—	—
Xylanases	5	0.7	—	—	—
Glucuronidases	4	0.6	—	—	—
Xylosidases	1	0.1	—	—	—
Others	5	0.7	—	—	—
Total plant cell wall catabolism	50	7.1			
Spore cell wall metabolism					
Chitin synthases	4	0.6	16	5.0	0.1
Chitinases	92	13.1	—	—	—
1,3-Glucanases	66	9.4	10	3.1	3.0
Total spore cell wall metabolism	158	22.6	26	8.1	2.8
Protein and amino acid catabolism					
Neutral protease	127	18.1	—	—	—
Alkaline protease	35	5.0	—	—	—
Amino acid oxidase	22	3.1	26	8.1	0.4
Total protein and amino acid catabolism	184	26.3	26	8.1	3.2
Total inducer-dependent metabolism	392	56.0	52	16.3	7.5
Glucose-derepression transcripts					
Alcohol dehydrogenase	143	20.4	185	55.1	0.4
Sucrose metabolism	70	10.0	4	1.2	8.4
Amylases	66	9.4	64	19.0	0.5
Glucosidases	18	2.6	22	6.5	0.4
Mannosidases	11	1.6	9	2.7	0.6
Total glucose-derepression metabolism:	308	44.0	284	84.5	1.1
		% EST		% EST	
Total MITs:	700	5.6	336	1.7	3.4

% MIT, percentage of misappropriated inducer-dependent transcripts (MIT); % EST, percentage of total ESTs in database.

a requirement for a specific inducer that was not present in the medium from which the EST source tissue was collected. Thus, it is not expected that these transcripts would be observed, yet they are clearly present.

Table 5 reports a survey of inducer-regulated activities detected in *A. nidulans* and compared with *N. crassa*. In *A. nidulans*, a total of 700 misappropriated inducer-dependent transcripts (MIT), 5.6% of the entire clone collection whose expression (based on the predicted function) requires a specific inducer, was found. *N. crassa* contains 336 such transcripts, representing 1.7% of that EST collection. Furthermore, these unexpected transcripts fall into two major groups: inducer-dependent and glucose-derepression-like transcripts (derepression of transcription by the absence of glucose). Inducer-dependent transcripts are frequent in *A. nidulans*, with 392 transcripts defining eight discrete functions, and rare in vegetative *N. crassa* tissues, with only 52 transcripts (26 amino acid oxidase, 16 chitin synthase, and 10 glucanase transcripts). *A. nidulans* transcripts produced in response to an inducer (56%) are implicated in plant cell wall degradation (7.1%), remodeling of the spore cell wall (22.6%), or acquisition of amino acids (26.3%). Glucose-derepression-like transcripts are detectable in *A. nidulans* and *N. crassa*, and they represent 44 and 84.5% of MIT transcripts, respectively.

Stress response genes are differentially expressed: Another observation was the frequent detection of stress-related genes in the *A. nidulans* asexually developing EST collection. Thus, *A. nidulans* stress-related transcripts were scored and compared with the *N. crassa* vegetative EST collection. A summary of the findings is shown in Figure 2.

Heat shock, DNA repair, trehalose recycling, and starvation response genes are predominant in *A. nidulans* (development), with 10.0-, 6.2-, 4.2-, and 3.0-fold higher expression levels when compared to *N. crassa* (vegeta-

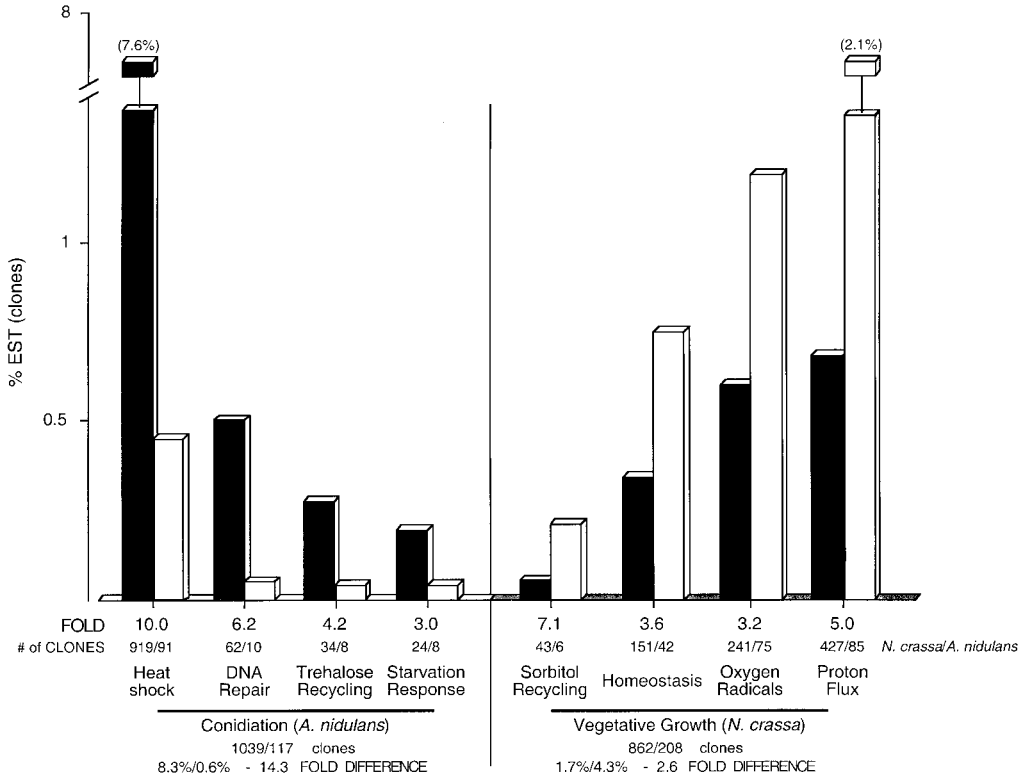

FIGURE 2.—Digital gene expression comparison of stress response genes in *A. nidulans* and *N. crassa*. Open and solid boxes indicate gene clusters associated with *N. crassa* and *A. nidulans* ESTs, respectively.

tive). Sorbitol recycling (7.1×), homeostasis (3.6×), oxygen radical removal (3.2×), and proton flux (5.0×) were dominant activities detected in *N. crassa*. Stress genes account for 9.9% of all transcripts in *A. nidulans*, and they appear to be a major group present in asexually developing *A. nidulans* tissue. In *N. crassa*, however, similar stress genes account for only 4.85% of the total transcripts, a 2.0-fold reduction. Moreover, stress transcripts prevalent in *A. nidulans* (heat shock, starvation, trehalose, and DNA repair) total 1039 transcripts, which contrasts with 117 transcripts in *N. crassa*, a 14.3-fold difference. Thus, in *A. nidulans* representation of stress-related transcripts is dramatically increased in relation to *N. crassa*.

DISCUSSION

In this study, digital gene expression profiles were evaluated for asexually developing *A. nidulans* tissues and compared to *N. crassa* vegetative tissue using equivalency criteria to account for population differences. The analysis of other fungal cDNA libraries was used to gather information about functional distribution and consistency of the digital information recovered from PipeOnline databases (Table 1).

EST collections obtained from different fungi by extraction of mRNA from tissues exposed to numerous physiological conditions did not result in extreme variability (Tables 1 and 2). Library redundancy and representation of clones was comparable if the size of the EST collection was considered (Table 1). In addition, functional annotation by PipeOnline produced EST subsets with functional annotation that corresponds on average to 35.8% of all the fungal collections, 44.6% being the highest and 27.5% the lowest (Table 2). Thus, dbEST fungal ESTs are useful for quantitative analysis, producing results with biological significance (AUDIC and CLAVERIE 1997; EWING *et al.* 1999).

ESTs from different organisms cluster similarly into the main cellular metabolic and structural groups (Table 4). These results corroborate the findings by EWING and CLAVERIE (2000) that gene clusters, representing metabolic pathways, may be compared with clusters of other organisms to render significant multiconditional gene expression information. Thus, *A. nidulans* clusters, when compared with *N. crassa* EST clusters, should reveal differences significant to vegetative growth or asexual development.

Here, the focus is on the identification of gene clusters important in *A. nidulans* asexual development by employing *N. crassa* ESTs to compare the vegetative state. We found two transcript clusters of interest, which we analyze and discuss in detail: (1) stress response genes and (2) misappropriated inducible genes whose expression requires an inducer absent at the time the tissue was harvested.

A. nidulans accumulates significant levels of these misappropriated transcripts during conidiation (Table 5). Under vegetative growth conditions, these transcripts are expressed only if an inducer is present. Accumula-

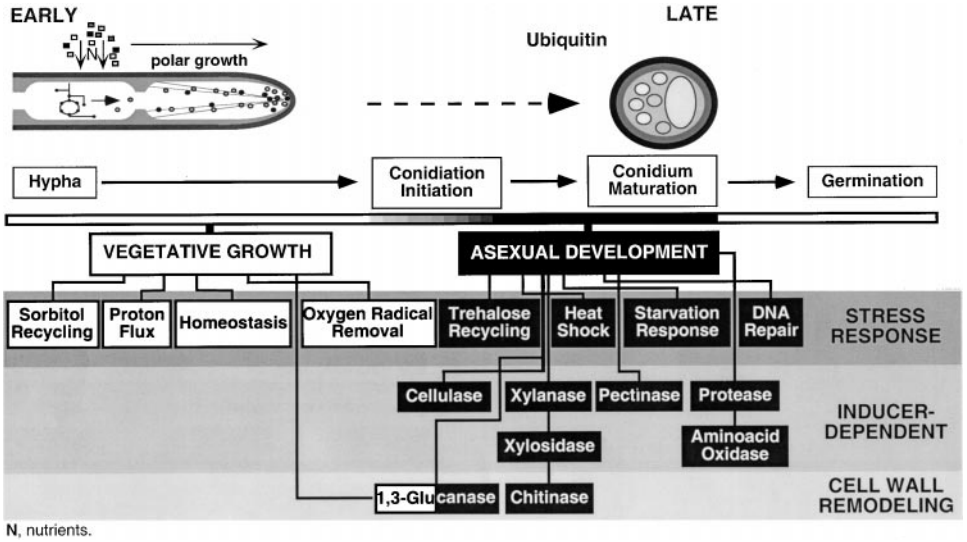

FIGURE 3.—Model illustrating the accumulation of stress response and misappropriated transcripts during asexual reproduction. Open and solid boxes indicate gene clusters associated with vegetative and asexual reproductive tissue ESTs, respectively. Lines connecting boxes indicate associations defined through EST function clustering.

tion of these transcripts may occur late during spore maturation and translate during germination. These pretranscribed mRNAs could confer a significant advantage if the spore germinates on a substratum on which free glucose is not readily available. An interesting aspect of these misappropriated transcripts is that all of them encode functions for plant cell wall or protein degradation, substrates likely to be abundant in natural habitats where Aspergillii are commonly found. The question of whether these misappropriated transcripts encode functional proteins remains unclear.

Finally, the involvement of low mRNA levels encoding catabolic functions has been suggested to explain regulation of cellulases and other plant-cell-wall-degrading enzymes (TORIGOI *et al.* 1996; CARLE-URIOSTE *et al.* 1997). Furthermore, implication of specific activities localized to the spore, or development in fungi, has also been suggested (KUBICEK 1987; BAGGA *et al.* 1989; MESSNER *et al.* 1991).

Accumulation of stress response mRNAs during conidiation produced another cluster for which a large number of transcripts has been determined (Figure 2). Heat-shock transcripts account for 7.6% (919 transcripts) of all *A. nidulans* ESTs, 10.0-fold higher than in *N. crassa*. Other stress-related clusters expressed at higher levels during asexual reproduction include DNA repair (62 transcripts, 6.2 times higher), trehalose recycling (34 transcripts, 4.2× higher) transcripts, and starvation response genes.

Not all stress response clusters were overrepresented in *A. nidulans*. Proton flux (427 transcripts, 5.0×), oxygen radical removal (241 transcripts, 3.2×), metal homeostasis (151 transcripts, 3.6×), and sorbitol recycling (43 transcripts, 7.1×) were expressed at higher levels in *N. crassa*, even though the overall difference was <2.6-fold.

Association of stress responses, specifically reactive oxygen removal, with reproduction in *A. nidulans* has been frequent and in some cases corroborated with powerful experimental demonstration (SKROMNE *et al.* 1995; NAVARRO *et al.* 1996; KAWASAKI *et al.* 1997; HAFKER *et al.* 1998; NAVARRO and AGUIRRE 1998). Digital clustering is based on quantitative measurements (AUDIC and CLAVERIE 1997; EWING and CLAVERIE 2000); thus, the heat-shock cluster needs to be considered as meaningful. Interestingly, oxygen radical removal was not found as a dominant cluster in *A. nidulans* even though 7 catalase A (developmental specific) and 4 catalase B transcripts were detected (data not shown, available from the http://aspergillus-genomics.org website). In fact, 3.21 times more transcripts in the *N. crassa* oxygen radical removal cluster were counted, including 30 catalase A transcripts and no catalase B.

Heat-shock treatments in *A. nidulans* germlings has been reported to dramatically increase trehalose, mannitol, and catalase A mRNA levels (NOVENTA-JORDAO *et al.* 1999). Moreover, treatment with hydrogen peroxide increases germling viability after heat shock, and catalase-A-deficient mutants are heat sensitive. Thus, there seems to be a genetic link between these two gene clusters.

Vegetative tissues are specialized in rapid growth and environmental occupation (WESSELS 1994; BARTNICKI-GARCIA *et al.* 1995; HAROLD 1999). Polar growth happens at hyphal tips, and components are delivered mainly through a vesicular delivery system (KAMADA *et al.* 1991). Asexual development in filamentous fungi entails a three-step process in which vegetative cells initiate asexual reproduction by making a decision usually noticed by switching the mode of cell division from polar to budding (WIESER *et al.* 1994; YU *et al.* 1996; KAMINSKYJ and HAMER 1998; YE *et al.* 1999). Conidiophore assembly and conidium production in most cases involve more than one cell type specialized in channeling haploid mitotic nuclei and amplifying the outcome of the reproductive process (TIMBERLAKE 1991; PRADE

and Timberlake 1993; Karos and Fischer 1996). The final stage entails maturation of the newly synthesized conidium by the addition of protective cell wall layers to the spore (Stringer *et al.* 1991; Timberlake 1991; Prade and Timberlake 1994).

Figure 3 shows a model that summarizes our findings. It is likely that during the later stages of development stress response and misappropriated transcript cluster genes are deposited in the conidium. These transcripts may be advantageous during the germination process. Simple carbon sources may not always be present, and induction of enzymes that enable assimilation and metabolism of alternate carbon sources is essential and may require the presence of low levels of the mRNA. Thus, deposition of these misappropriated transcripts during the reproductive process may be essential to the future survival of the spore. Deposition of stress response genes can be explained similarly. Spores may germinate under conditions where temperatures, salt concentrations, or water potential are not ideal.

Detection of inducer-dependent transcripts in the absence of inducer and stress response transcripts in the absence of an obvious source of stress late in development may indicate that vegetative *cis*-acting regulatory networks are no longer functional and result in derepression. Thus, during asexual development, *A. nidulans* overrides vegetative regulatory controls and produces a series of transcripts that may add protective and adaptive advantages to the dormant spore. These observations lead to the conclusion that during terminal asexual development, vegetative *cis*-acting regulatory networks are no longer functional and result in general derepression that may result in augmented survivability upon germination.

We thank Bruce Roe and Doris Kupfer, whose scientific interest resulted in the first public, free-of-charge, and unfettered access to large-scale DNA sequence information to the fungal community. We also thank Eduardo Misawa and co-workers from the Oklahoma State University bioinformatics laboratory for expert assistance in the Pipe-Online software package development and implementation as well as administration of local computational resources. The reported research was partially funded with grants from the National Science Foundation (NSF 98-13360), the United States Department of Agriculture (USDA 97-35303-4459), and an industrial consortium: Genencor International (USA), Glaxo Wellcome (Spain), Gist-brocades (The Netherlands), Novo Nordisk (Denmark), Kikkoman (Japan), and Amano (Japan).

LITERATURE CITED

Alwine, J. C., D. J. Kemp and G. R. Stark, 1977 Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. USA **74:** 5350–5354.

Aramayo, R., and W. E. Timberlake, 1990 Sequence and molecular structure of the *Aspergillus nidulans yA* (laccase I) gene. Nucleic Acids Res. **18:** 3415.

Audic, S., and J. M. Claverie, 1997 The significance of digital gene expression profiles. Genome Res. **7:** 986–995.

Bagga, P. S., S. Sharma and D. K. Sandhu, 1989 Developmentally related changes in the production and expression of endo-beta-1,4-glucanases in *Aspergillus nidulans*. Genome **32:** 288–292.

Bartnicki-Garcia, S., D. D. Bartnicki and G. Gierz, 1995 Determinants of fungal cell wall morphology: the vesicle supply center. Can. J. Bot. **73** (Suppl. 1): S372–S378.

Berk, A. J., and P. A. Sharp, 1977 Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12:** 721–732.

Carle-Urioste, J. C., J. Escobar-Vera, S. El-Gogary, F. Henrique-Silva, E. Torigoi *et al.*, 1997 Cellulase induction in *Trichoderma reesei* by cellulose requires its own basal expression. J. Biol. Chem. **272:** 10169–10174.

Chen, T., and S. S. Skiena, 2000 A case study in genome-level fragment assembly. Bioinformatics **16:** 494–500.

Cho, R. J., and M. J. Campbell, 2000 Transcription, genomes, function. Trends Genet. **16:** 409–415.

Doebley, J., and L. Lukens, 1998 Transcriptional regulators and the evolution of plant form. Plant Cell **10:** 1075–1082.

Ewing, R. M., and J. M. Claverie, 2000 EST databases as multi-conditional gene expression datasets. Pac. Symp. Biocomput. **57:** 430–442.

Ewing, R. M., A. B. Kahla, O. Poirot, F. Lopez, S. Audic *et al.*, 1999 Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res. **9:** 950–959.

Hafker, T., D. Techel, G. Steier and L. Rensing, 1998 Differential expression of glucose-regulated (*grp78*) and heat-shock-inducible (*hsp70*) genes during asexual development of *Neurospora crassa*. Microbiology **144:** 37–43.

Harold, F. M., 1999 In pursuit of the whole hypha. Fungal Genet. Biol. **27:** 128–133.

Kamada, T., C. E. Bracker and S. Bartnicki-Garcia, 1991 Chitosomes and chitin synthetase in the asexual life cycle of *Mucor rouxii*: spores, mycelium and yeast cells. J. Gen. Microbiol. **137:** 1241–1252.

Kaminskyj, S. G., and J. E. Hamer, 1998 *hyp* loci control cell pattern formation in the vegetative mycelium of *Aspergillus nidulans*. Genetics **148:** 669–680.

Karos, M., and R. Fischer, 1996 *hymA* (hypha-like metulae), a new developmental mutant of *Aspergillus nidulans*. Microbiology **142:** 3211–3218.

Kawasaki, L., D. Wysong, R. Diamond and J. Aguirre, 1997 Two divergent catalase genes are differentially regulated during *Aspergillus nidulans* development and oxidative stress. J. Bacteriol. **179:** 3284–3292.

Kelkar, H. S., J. Griffith, M. E. Case, S. F. Covert, R. D. Hall *et al.*, 2001 The *Neurospora crassa* genome: cosmid libraries sorted by chromosome. Genetics **157:** 979–990.

Kozian, D. H., and B. J. Kirschbaum, 1999 Comparative gene-expression analysis. Trends Biotechnol. **17:** 73–78.

Kubicek, C. P., 1987 Involvement of a conidial endoglucanase and a plasma-membrane-bound beta-glucosidase in the induction of endoglucanase synthesis by cellulose in *Trichoderma reesei*. J. Gen. Microbiol. **133:** 1481–1487.

Kupfer, D. M., C. A. Reece, S. W. Clifton, B. A. Roe and R. A. Prade, 1997 Multicellular ascomycetous fungal genomes contain more than 8000 genes. Fungal Genet. Biol. **21:** 364–372.

Messner, R., E. M. Kubicek-Pranz, A. Gsur and C. P. Kubicek, 1991 Cellobiohydrolase II is the main conidial-bound cellulase in *Trichoderma reesei* and other Trichoderma strains. Arch. Microbiol. **155:** 601–606.

Navarro, R. E., and J. Aguirre, 1998 Posttranscriptional control mediates cell type-specific localization of catalase A during *Aspergillus nidulans* development. J. Bacteriol. **180:** 5733–5738.

Navarro, R. E., M. A. Stringer, W. Hansberg, W. E. Timberlake and J. Aguirre, 1996 *catA*, a new *Aspergillus nidulans* gene encoding a developmentally regulated catalase. Curr. Genet. **29:** 352–359.

Noventa-Jordao, M. A., R. M. Couto, M. H. Goldman, J. Aguirre, S. Iyer *et al.*, 1999 Catalase activity is necessary for heat-shock recovery in *Aspergillus nidulans* germlings. Microbiology **145:** 3229–3234.

Ohlrogge, J., and C. Benning, 2000 Unraveling plant metabolism by EST analysis. Curr. Opin. Plant Biol. **3:** 224–228.

Overbeek, R., N. Larsen, G. D. Pusch, M. D'Souza, E. Selkov, Jr. *et al.*, 2000 WIT: integrated system for high-throughput genome

sequence analysis and metabolic reconstruction. Nucleic Acids Res. **28:** 123–125.

Prade, R. A., and W. E. Timberlake, 1993 The *Aspergillus nidulans brlA* regulatory locus encodes two functionally redundant polypeptides that are individually essential for development. EMBO J. **12:** 2439–2447.

Prade, R. A., and W. E. Timberlake, 1994 The *Penicillium chrysogenum* and *Aspergillus nidulans wetA* developmental regulatory genes are functionally equivalent. Mol. Gen. Genet. **244:** 539–547.

Richmond, T., and S. Somerville, 2000 Chasing the dream: plant EST microarrays. Curr. Opin. Plant Biol. **3:** 108–116.

Schmitt, A. O., T. Specht, G. Beckmann, E. Dahl, C. P. Pilarsky *et al.*, 1999 Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. Nucleic Acids Res. **27:** 4251–4260.

Schuler, G. D., J. A. Epstein, H. Ohkawa and J. A. Kans, 1996 Entrez: molecular biology database and retrieval system. Methods Enzymol. **266:** 141–162.

Selkov, E. J., Y. Grechkin, N. Mikhailova and E. Selkov, 1998 MPW: the Metabolic Pathways Database. Nucleic Acids Res. **26:** 43–45.

Seoighe, C., and K. H. Wolfe, 1998 Extent of genomic rearrangement after genome duplication in yeast. Proc. Natl. Acad. Sci. USA **95:** 4447–4452.

Skromne, I., O. Sanchez and J. Aguirre, 1995 Starvation stress modulates the expression of the *Aspergillus nidulans brlA* regulatory gene. Microbiology **141:** 21–28.

Smulian, A. G., T. Sesterhenn, R. Tanaka and M. T. Cushion, 2001 The *ste3* pheromone receptor gene of *Pneumocystis carinii* is surrounded by a cluster of signal transduction genes. Genetics **157:** 991–1002.

Stringer, M. A., R. A. Dean, T. C. Sewall and W. E. Timberlake, 1991 Rodletless, a new *Aspergillus* developmental mutant induced by directed gene inactivation. Genes Dev. **5:** 1161–1171.

Tavazoie, S., J. D. Hughes, M. J. Campbell, R. J. Cho and G. M. Church, 1999 Systematic determination of genetic network architecture. Nat. Genet. **22:** 281–285.

Timberlake, W. E., 1991 Temporal and spatial controls of *Aspergillus* development. Curr. Opin. Genet. Dev. **1:** 351–357.

Torigoi, E., F. Henrique-Silva, J. Escobar-Vera, J. C. Carle-Urioste, O. Crivellaro *et al.*, 1996 Mutants of *Trichoderma reesei* are defective in cellulose induction, but not basal expression of cellulase-encoding genes. Gene **173:** 199–203.

Wessels, J. G. H., 1994 Developmental regulation of fungal cell wall formation. Annu. Rev. Phytopathol. **32:** 413–437.

Wieser, J., B. N. Lee, J. Fondon III and T. H. Adams, 1994 Genetic requirements for initiating asexual development in *Aspergillus nidulans*. Curr. Genet. **27:** 62–69.

Ye, X. S., S. L. Lee, T. D. Wolkow, S. L. McGuire, J. E. Hamer *et al.*, 1999 Interaction between developmental and cell cycle regulators is required for morphogenesis in *Aspergillus nidulans*. EMBO J. **18:** 6994–7001.

Yu, J. H., J. Wieser and T. H. Adams, 1996 The *Aspergillus FlbA* RGS domain protein antagonizes G protein signaling to block proliferation and allow development. EMBO J. **15:** 5184–5190.

Communicating editor: J. Arnold