2005

Comparative analysis of 87,000 expressed sequence tags from the fumonisin-producing fungus *Fusarium verticillioides*

Daren W. Brown
USDA-ARS, browndw@ncaur.usda.gov

Foo Cheung
The Institute for Genomic Research

Robert H. Proctor
USDA-ARS

Robert A. E. Butchko
USDA-ARS

Li Zheng
The Institute for Genomic Research

See next page for additional authors

Follow this and additional works at: http://digitalcommons.unl.edu/usdaarsfacpub

Authors

Daren W. Brown, Foo Cheung, Robert H. Proctor, Robert A. E. Butchko, Li Zheng, Yuandan Lee, Teresa Utterback, Shannon Smith, Tamara Feldblyum, Anthony E. Green, Ronald D. Plattner, David F. Kendra, Christopher D. Town, and Catherine A. Whitelaw

ELSEVIER

Comparative analysis of 87,000 expressed sequence tags from the fumonisin-producing fungus *Fusarium verticillioides*

Daren W. Brown [a,*], Foo Cheung [b], Robert H. Proctor [a], Robert A.E. Butchko [a], Li Zheng [b], Yuandan Lee [b], Teresa Utterback [c], Shannon Smith [c], Tamara Feldblyum [c], Anthony E. Glenn [d], Ronald D. Plattner [a], David F. Kendra [a], Christopher D. Town [b], Catherine A. Whitelaw [b]

[a] *Mycotoxin Research Unit, U.S. Department of Agriculture-ARS, Peoria, IL 61604, USA*
[b] *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA*
[c] *J. Craig Venter Science Foundation Joint Technology Center, 5 Research Place Rockville, MD 20850, USA*
[d] *Toxicology and Mycotoxin Research Unit, U.S. Department of Agriculture-ARS, Athens, GA 30605, USA*

Abstract

Fusarium verticillioides (teleomorph *Gibberella moniliformis*) is a pathogen of maize worldwide and produces fumonisins, a family of mycotoxins that have been associated with several animal diseases as well as cancer in humans. In this study, we sought to identify fungal genes that affect fumonisin production and/or the plant–fungal interaction. We generated over 87,000 expressed sequence tags from nine different cDNA libraries that correspond to 11,119 unique sequences and are estimated to represent 80% of the genomic complement of genes. A comparative analysis of the libraries showed that all 15 genes in the fumonisin gene cluster were differentially expressed. In addition, nine candidate fumonisin regulatory genes and a number of genes that may play a role in plant–fungal interaction were identified. Analysis of over 700 *FUM* gene transcripts from five different libraries provided evidence for transcripts with unspliced introns and spliced introns with alternative 3′ splice sites. The abundance of the alternative splice forms and the frequency with which they were found for genes involved in the biosynthesis of a single family of metabolites as well as their differential expression suggest they may have a biological function. Finally, analysis of an EST that aligns to genomic sequence between *FUM12* and *FUM13* provided evidence for a previously unidentified gene (*FUM20*) in the *FUM* gene cluster.
Published by Elsevier Inc.

Keywords: *Fusarium verticillioides*; Fumonisin; Expressed sequence tag; EST; Regulation; Alternatively spliced isoforms; Secondary metabolite; Gene cluster; *FUM20*

1. Introduction

Fusarium verticillioides (teleomorph *Gibberella moniliformis*) is a pathogen of maize worldwide and is capable of causing disease of ears, stalks, and seedlings (Munkvold and Desjardins, 1997). During this process, the fungus may also synthesize fumonisins, a family of poly-ketide-derived mycotoxins that are frequently found contaminating maize and are associated with a number of fatal animal diseases, including cancer (Howard et al., 2001). Recently, the association between fumonisins and neural tube birth defects in mice has led to the proposal that fumonisins are a potential risk factor for similar birth defects in humans (Marasas et al., 2004). The biological effects of fumonisin ingestion or exposure are attributed to the disruption of sphingolipid metabolism

---

via the inhibition of the enzyme ceramide synthase (Marasas et al., 2004).

Characterization of the biochemical and regulatory pathways leading to fumonisin biosynthesis may lead to novel methods to limit fumonisin contamination of maize. At present, the fumonisin biosynthetic gene (*FUM*)[1] cluster consists of 15 co-regulated genes that span 42 kb of chromosome 1 (Proctor et al., 2003). Predicted amino acid sequence of the *FUM* cluster genes suggested that 11 genes encode biosynthetic enzymes, two encode transporters, and two encode proteins putatively involved in self-protection (Proctor et al., 2003). A majority of the cluster genes have been functionally characterized and have been found to play a role in fumonisin biosynthesis (Bojja et al., 2004; Butchko et al., 2003a,b; Butchko and Proctor, unpublished observations).

Although most other fungal secondary metabolite gene clusters (e.g., sterigmatocystin/aflatoxin, trichothecene, and lovastatin gene clusters) thus far characterized include pathway-specific regulatory genes (Brown et al., 1996; Hohn et al., 1999; Kennedy et al., 1999; Yu et al., 2004a), sequence analysis of the fumonisin gene cluster did not reveal the presence of a regulatory gene. Two genes that regulate fumonisin biosynthesis and that are not linked to the cluster have been characterized (Flaherty and Woloshuk, 2004; Shim and Woloshuk, 2001). *FCC1* encodes a cyclin-like protein that is thought to be part of a signal transduction pathway that affects both fumonisin biosynthesis and development (Shim and Woloshuk, 2001). *ZFR1* encodes a zinc finger protein that appears to be a positive regulator of *FUM* genes: fumonisin production was reduced by 90% compared to wild type in *ZFR1* deletion mutants (Flaherty and Woloshuk, 2004). A regulatory gene that is solely dedicated to fumonisin biosynthesis has yet to be identified.

Differential expression of messenger RNA has been used extensively to identify genes or gene products involved in specific biological processes. EST studies with *F. graminearum* cDNA libraries identified 2110 genes while expressed sequence tag (EST) studies with *Magnaporthe grisea* cDNA libraries identified 8177 genes of which some shared identity with pathogenicity-associated genes (Ebbole et al., 2004; Trail et al., 2003). Studies with *Aspergillus flavus* and microarrays using a selected set of 753 genes differentially expressed during aflatoxin production identified a subset of genes poten-tially involved in secondary metabolism or aflatoxin production (OBrian et al., 2003). In this study, we generated over 87,000 ESTs from nine different cDNA libraries to identify candidate genes involved in fumonisin regulation as well as candidate genes important for the plant–fungal interaction.

2. Materials and methods

2.1. cDNA library construction

Total RNA from *F. verticillioides* was purified with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) following the protocol described by the manufacturer for samples with high polysaccharide content. For libraries prepared from growth in liquid media (Table 1), cultures were vacuum filtered and the resulting mycelial mats were frozen in liquid nitrogen, ground to a powder, and then added to TRIzol at ≈1 g mycelia per 10 ml TRIzol. For libraries prepared from growth on solid media (e.g., maize meal, maize seedling tissue, or developing kernels; Table 1), both the medium and the *F. verticillioides* mycelia were frozen in liquid nitrogen, ground to a powder, and then added to TRIzol at ≈1 g ground material per 10 ml TRIzol. Messenger RNA was isolated from total RNA with the Poly(A)Purist MAG RNA Purification Kit (Ambion, Austin, TX). All libraries were prepared in the pBlue-Script II SK(+) XR vector from purified mRNA with the cDNA Synthesis Kit (Stratagene, La Jolla, CA). Libraries FvF and FvG were prepared from RNA obtained from strain M-3125 grown in liquid GYAM (8 mM L-asparagine, 1.7 mM NaCl, 4.4 mM K_2HPO_4, 2 mM $MgSO_4$, 8.8 mM $CaCl_2$, 0.05% yeast extract, 0.24 M glucose, and 5.0 mM malic acid) medium for 24 and 96 h, respectively. Library FvM was prepared from pooled RNA obtained from a 48- and 72-h liquid GYAM culture from strain M-3125. Library FvO was prepared from a 90-h liquid GYAM culture from the fumonisin-nonproducing strain 57-7-7 (Desjardins et al., 1996) by the same procedures described above. Library FvN was obtained from RNA derived from a corn meal medium culture of strain M-3125. These cultures were prepared by inoculating an autoclaved mixture of 25 g corn meal and 5 ml distilled water with 5 ml water containing 5×10^7 conidia. The inoculated medium was mixed thoroughly, distributed equally into two 100-mm plastic petri dishes, and incubated at room temperature for 4 or 6 days. Library FvH was prepared from germinating conidia of strain M-3125 in an aqueous extract of maize seedlings. The aqueous extract was obtained by germinating surface-sterilized maize seeds for 4 days on 1% water agar supplemented with 100 µg/ml streptomycin sulfate. Under sterile conditions, 70 g of excised roots and shoots from the

---

[1] *Abbreviations used:* EST, expressed sequence tag; NR, nonredundant database; NCBI, National Center for Biotechnology Information; BLAST, basic local alignment search tool; AA, amino acid; ORF, open reading frame; ASF, alternative splice form; TIGR, The Institute for Genomic Research; TGICL, TIGR Gene Indices clustering tools; FUM#p, fumonisin protein; TC, tentative consensus sequence.

Table 1
Libraries, culture conditions, and sequencing statistics for ESTs generated in this study

Library ID	Description	TIGR Cat #	ESTs generated[a]	Unique sequences[b]	Average read length (bp)
FvF	Growth for 24 h in GYAM media	T10964	7,647	2951	636
FvG	Growth for 96 h in GYAM media	T11287	9,120	2983	655
FvH	Growth for 10 h in water extracts of maize seedlings	T11852	9,603	3114	654
FvI	Growth on excised maize seedling roots and shoots	T12440	10,953	4445	679
		T13474			
FvJ	Growth on developing maize kernels	T12441	156	139	632
FvKL	BOA subtracted library; Ligation 1 and 2	T12984	191	58	484
		T12985	156	47	476
FvM	Growth for 48 and 72 h in GYAM media	T13047	12,861	2669	846
FvN	Growth on corn meal medium for 4 and 6 days	T14017	17,648	4353	812
FvO	Fumonisin minus strain grown for 90 h in GYAM media	T14018	18,093	3534	803

[a] Total number of ESTs generated by sequence analysis of clones from each cDNA library.
[b] Number of unique sequences (TCs and singletons) within individual library.

resulting seedlings was blended with 250 ml water, filtered once through Miracloth (Calbiochem), and then again through Whatman #1 filter paper. The extracts were inoculated to a spore concentration of 5×10^6 conidia/ml and then incubated for 10 h at 30 °C with shaking at 200 RPM. Library FvI was prepared from growth of strain M-3125 on excised maize seedling roots and shoots. The roots and shoots were prepared as described above and then inoculated by dipping briefly in a suspension of 5×10^6 conidia/ml. The inoculated roots and shoots were then incubated on moistened Whatman #1 filter paper at room temperature under sterile conditions for 4 days. Libraries FvK and FvL, referred to hereafter after as library FvKL, were constructed employing suppression subtractive hybridization (SSH) based on the PCR Select cDNA Subtraction Kit (Clonetech, USA). These libraries were created to specifically identify *F. verticillioides* genes up-regulated in response to maize antimicrobial defense compounds (Glenn et al., 2001, 2002) and will be described in detail separately.

2.2. Sequencing methodology and EST clustering

Plasmid DNA was prepared using the 5-Prime DNA Purification Workstation (Eppendorf—5 Prime, Boulder, CO) and sequenced using Big Dye Terminator Cycle Sequencing Ready Reaction reagents and ABI 3700 or 3730xl Genetic Analyzers (Applied Biosystems, Foster City, CA, USA). Following electrophoresis and fluorescence detection, quality values were assigned to each base call by TraceTuner software (Paracel, Pasadena, CA, USA), and Lucy (Chou and Holmes, 2001) was used to automatically process the raw sequence data, including trimming of low quality bases, vector and *Escherichia coli* sequences, and those sequences

<100 basepairs (bp). ESTs were grouped into clusters using sequence similarity and clone links and the TGICL clustering utilities (Pertea et al., 2003). Each cluster was assembled at high stringency using the Paracel Transcript Assembler (Paracel) to produce tentative consensus sequences (TCs). Those ESTs that did not assemble into a TC were termed "singletons."

2.3. Sequence analysis

EST libraries were compared with each other using the TIGR Library Expression Search tool (LES) located within the *F. verticillioides* Gene Index web pages (http://www.tigr.org/tdb/tgi/). This tool compares ESTs from two or more libraries and generates a list of all sequences with a significant value (R) that is a measure of the likelihood that differences in EST abundance in different libraries are due to differences in the level of gene expression rather than to random sampling (Stekel et al., 2000). Sequences with R values >9, a value expected to produce a false positive rate of 1 in 1000, were examined further. Sequence similarity searches of the nonredundant (NR) database maintained by the National Center for Biotechnology Information (NCBI) were performed using BLAST (Altschul et al., 1997; Zhang et al., 1998). Lower E values reflect a higher probability that the assigned function indicated in the tables may be correct. Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. DR595215-DR681653.

2.4. RT-PCR of FUM14

Reverse transcription was carried out with SuperScript Reverse Transcriptase (Invitrogen) and primer 728 (5′-TCGAATCGTCCGCTTCACCAGT-3′) according

to the specifications of the manufacturer. PCR was carried out with *Taq* DNA polymerase (Qiagen), and primers 728 and 727 (5′-GTCAAAGCCATCCAACTCTTATC-3′).

3. Results and discussion

3.1. Sequence and library quality

Nine *F. verticillioides* cDNA libraries were generated from eight different growth conditions for this study (Table 1). A total of 115,657 sequencing reactions of cDNA clones were performed and, after quality assessment, 87,138 ESTs, with an average edited read length of 744 bp, were generated. All libraries sequenced with an average success rate of >80% except for libraries FvJ (61%) and FvKL (17%). Library FvKL was prepared by subtractive hybridization and probably sequenced poorly due to the small size of many of the cDNA inserts produced by the subtraction process as sequence reads <100 nts were automatically excluded by the quality control software. It is not clear why FvJ sequenced poorly. Due to the low success rate of these libraries, they were not sequenced extensively. Preliminary assessment of cDNAs from each library indicated that FvI and FvJ contained a significant number of maize cDNA clones. Based on this observation, all FvI and FvJ ESTs were screened against the TIGR maize Gene Index and those with >95% nucleotide identity over 95% of the sequence length were removed. The *F. verticillioides* ESTs from this project and sequences extracted from GenBank were clustered into 7198 TC sequences and 3928 singleton sequences and ETs (Expressed or non-human Transcripts from TIGR's Expressed Gene Anatomy Database (EGAD)) to produce a Gene Index of 11,126 unique sequences (www.tigr.org/tdb/tgi/) (Pertea et al., 2003; Quackenbush et al., 2001, 2000) (Table 1).

Initial analyses of the EST data revealed that many ESTs did not occur in all cDNA libraries indicating that the libraries included differentially expressed sequences. The observation that 6048 TCs and singletons were derived from only one of the nine libraries suggests that they may correspond to genes that were expressed under one culture condition but not others (Table 2). In contrast, 1130 TCs consisted of ESTs derived from four or more cDNA libraries and 16 TCs consisted of ESTs from seven libraries (Table 3). The three TCs resulting from the largest number of cDNAs were TC26125 (722), TC29728 (611), and TC26141 (379), and ESTs were present in all libraries except FvJ and FvKL. TC26125 was most similar to DNA damage-responsive protein 48 from *Saccharomyces cerevisiae*, TC29728 was most similar to elongation factor 1-α, and TC26141 was most similar to glyceraldehyde 3-phosphate dehydrogenase. The sequences generated in this study were assigned a putative function based on the best hit to a non-redundant

Table 2
Number of library-specific unique sequences

Library ID	Singletons[a]	TCs	Total sequences
FvF	516	137	653
FvG	375	198	573
FvH	416	215	631
FvI	1127	465	1592
FvJ	66	7	73
FvKL	22	2	24
FvM	282	179	461
FvN	829	367	1196
FvO	527	318	845
Total	4160	1888	6048

[a] TCs arising from reads from opposite ends of the same clone were considered singletons for this analysis.

Table 3
Number of TCs that occur in libraries

Libraries represented[a]	No. of TCs[b]
1	1888
2	1944
3	1885
4	492
5	353
6	269
7	16

[a] Number of different libraries in which ESTs for specific TCs were observed. For example, 1888 TCs were derived from ESTs from one library where 1944 TCs were derived from ESTs from two libraries.
[b] TCs arising from reads from opposite ends of the same clone have been excluded from this analysis.

GO_PEP database using DPS (a DNA alignment program) with a cutoff threshold score of 300 (Huang et al., 1997). Almost 87% of the unique sequences had a protein hit and 19% could be assigned to a functional classification based on the Gene Ontology assignments (Ashburner et al., 2000).

To estimate how well the ESTs represent genome coding capacity, we compared the 11,126 sequences in the Gene Index to a 43-kb region of *F. verticillioides* genomic sequence not linked to the fumonisin gene cluster. Initial characterization of the genomic sequence by BLASTX had identified 13 putative genes. In contrast, 16 potential genes were identified by comparison with the Gene Index. Twelve of the 13 genes identified by BLASTX matched the genes identified by the ESTs. A total of four additional genes were identified in the genomic sequence. The predicted function of most of the genes (e.g., glutamine synthetase, *glsA*; methylcitrate synthase, *mcsA*; isocitrate synthase, *icsA*) as well as their occurrence in multiple libraries suggest that they are involved in general housekeeping functions. Only one of the 13 predicted genes was not represented in the EST collection. This analysis indicates that the EST collection represents a high percentage of the *F. verticillioides* coding capacity.

Table 4
Library expression search (LES) of fumonisin production libraries

LES comparison	Total DE[a]	Fumonisin+[b]	*FUM* genes	Possible regulatory	Unknown function
FvF vs. FvG	55	12 in FvG	8	1	1
FvF vs. FvM	84	22 in FvM	8[c]	3	4
FvM vs. FvF/FvG	183	38 in FvM	0	5	10
FvG vs. FvF/FvM	179	10 in FvG	0	0	5

[a] Total number of putatively differentially expressed (DE) genes with $R > 9$.
[b] Number of DE genes observed in the library generated from culture conditions where fumonisins were produced.
[c] Over all, a total of 10 different *FUM* genes were detected. Two *FUM* genes were detected in the FvF vs. FvG comparison that were not detected in the FvF vs. FvM comparison and two *FUM* genes were detected in the FvF vs. FvM comparison that were not detected in the FvF vs. FvG comparison.

3.2. Confirmation of the presence of FUM genes in the EST libraries

In a previous study, *FUM* gene transcripts were shown to be differentially expressed over time in liquid GYAM cultures of *F. verticillioides* (Proctor et al., 2003). *FUM* gene transcripts were not detected at 24 h, prior to detection of fumonisins, but were detected in abundance at 88 and 100 h, when fumonisin production was detected. We reasoned that comparisons of ESTs from the libraries prepared from GYAM cultures harvested at different time points could identify differentially expressed transcripts that are important for fumonisin biosynthesis. In this study, fumonisins were not detected in the 24-h culture, low levels of fumonisin B₁ (<10 μg/ml) were detected in the 48- and 72-h cultures, and higher levels (>50 μg/ml) were detected in the 96-h culture. Comparisons among 24-h (FvF), 48/72-h (FvM), and 96-h (FvG) libraries using Gene Index LES tool revealed differential expression of 10 of the 15 *FUM* cluster genes that correlated with fumonisin biosynthesis (Table 4).

Since LES identifies differentially expressed genes but not genes that are simply expressed, we used the Sequencher (version 4.1.4) assemble function to further explore the occurrence of *FUM* genes in libraries FvF, FvM, and FvG. For this analysis, the entire Gene Index was assembled with 42.3 kb of *F. verticillioides* genomic sequence (Accession No. AF155773) that contains the *FUM* cluster (Fig. 1). All 10 *FUM* gene TCs detected in the LES analysis as well as TCs and singleton ESTs corresponding to the remaining five *FUM* genes were detected in this analysis. Overall expression levels, as measured by the percent total of all 15 *FUM* gene ESTs in each library, increased 2.2-fold between 48/72 h (FvM) and 96 h (FvG) of growth in GYAM (Table 5). The greatest increase in EST abundance (12-fold) was observed for *FUM1*, which encodes the polyketide synthase thought to catalyze the first step in fumonisin biosynthesis. The failure to detect the differential expression of five *FUM* genes by LES most likely reflects the limited abundance of ESTs corresponding to these genes. The finding that *FUM* transcripts were present at or after 48 h of growth is consistent with previous Northern analyses (Proctor et al., 2003) that revealed the differential expression of *FUM* genes over time in GYAM.

3.3. Identification of potential fumonisin regulatory genes

Two-library LES comparisons between the GYAM libraries also identified a number of putatively differentially expressed genes that may play a role in fumonisin
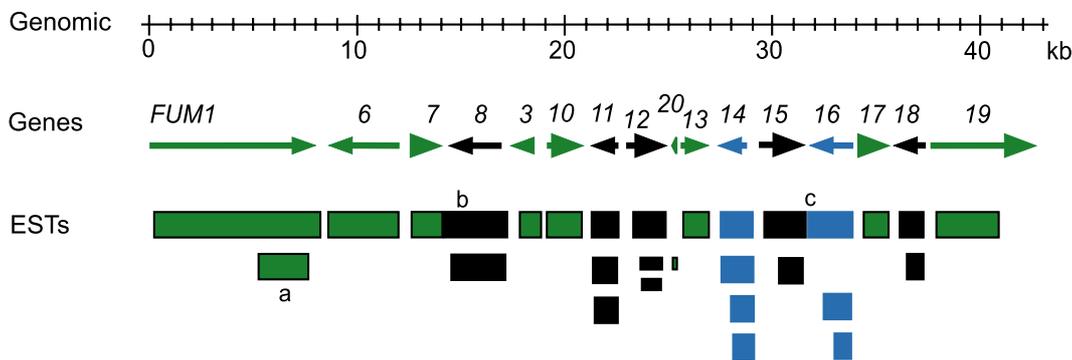
Fig. 1. Comparison of ESTs with the *F. verticillioides* genomic sequence containing the fumonisin 15-gene cluster. Relative scale is shown in kilobases. Organization of the predicted ORFs. Gene names are indicated above each large arrow. Arrowheads refer to the direction of gene transcription. All boxes represent TCs except for the two boxes under *FUM12* and the box under *FUM20* that represents singletons. Black boxes/arrows indicate TCs with alternative spliced form transcripts with retained introns and blue boxes/arrows indicate TCs with alternative spliced form transcripts with introns with alternative 3′ borders. (a) TC26329 identified a new intron in *FUM1*. (b) TC29823 contains ESTs for both *FUM7* and *FUM8*. (c) TC26133 contains ESTs for both *FUM15* and *FUM16*.

Table 5
Description of *FUM* genes and distribution of ESTs in GYAM culture-derived libraries[a]

Gene	Putative function	Predicted No. AAs	Total No. of ESTs	No. ESTs in FvF	No. ESTs FvM[b]	No. ESTs FvG[b]	Change in ESTs FvM to FvG[h]	Change in % TO[c] FvM to FvG
FUM1	Polyketide synthase	2586[d]	41	0	1	11	12.0 ↑	
FUM6[e]	Cytochrome P450 monooxygenase	1115	44	0	19	13	No change	
FUM7[e,f]	Dehydrogenase	424	17	0	11	4	2.0 ↓	
FUM8[e,f]	Oxoamine synthase	836[d]	65	0	16	36	3.3 ↑	
FUM9	Dioxygenase	300	41	0	9	17	2.7 ↑	
FUM10[f]	Fatty acyl-CoA synthetase	552[d]	91	0	20	61	4.5 ↑	
FUM11	Tricarboxylate transporter	306[d]	19	0	5	7	2.0 ↑	2.5 ↑
					2 TO	7 TO		
FUM12[e]	Cytochrome P450 monooxygenase	502[d]	85	0	17	63	5.4 ↑	
FUM20	Unknown	na	1	0	0	1		
FUM13[e]	Reductase	369	40	0	20	10	1.4 ↓	
FUM14[e]	AA condensation domain	553[d]	150	0	55	65	1.7 ↑	8.9↑
					1 TO	11 TO		
FUM15[e,g]	Cytochrome P450 monooxygenase	596[d]	21	0	8	10	1.8 ↑	
FUM16[e,g]	Fatty acyl-CoA synthetase	676[d]	90	0	46	37	1.1 ↑	2.7 ↑
					10 TO	24 TO		
FUM17	Longevity assurance factor	388	6	0	0	3		
FUM18	Longevity assurance factor	384[d]	9	0	5	3	1.3 ↓	3.3 ↑
					1 TO	2 TO		
FUM19	ABC transporter	1489	1	0	0	1		
Total					232	342	2.2 ↑	4.4 ↑

[a] FvO *FUM* transcripts were not considered since there were only four.
[b] The percent of all ESTs that correspond to *FUM* genes in FvM was 1.7 (13,315 ESTs) and in FvG was 3.7 (9172 ESTs).
[c] TO, truncated ORF transcript.
[d] Updated predicted protein sequence (Accession No. AF155773).
[e] Genes differentially expressed as identified by LES when compared with FvF.
[f] Transcripts to *FUM7* and *FUM8* were assembled into TC29823.
[g] Transcripts to *FUM15* and *FUM16* were assembled into TC26133. The potential functionality of four *FUM16* transcripts could not be assessed based on available sequence.
[h] Change calculated as fold difference in number of transcripts observed between FvM and FvG.

biosynthesis (Table 4). In addition to the *FUM* genes, the FvF/FvG comparison identified four FvG-specific TCs one of which shared similarity to a putative regulatory gene: TC30021 shared similarity with the portion of the *Homo sapiens EST1A* gene (Accession No. AAO17581) that encodes a domain thought to be involved in mRNA decay. The FvF/FvM comparison identified 14 TCs, in addition to the *FUM* genes, that occurred only in FvM. Ten of these share similarity to genes with known functions ($E < 10^{-5}$), and three of these (TC30131, TC26574, and TC26727) share similarity to genes encoding known regulatory proteins and therefore have the potential to be involved in the regulation of fumonisin biosynthesis (Table 6).

Although the three-library LES comparisons among FvF, FvM, and FvG did not identify any additional *FUM* genes, a number of TCs were found that were specific to FvM or FvG that may play a role in fumonisin biosynthesis. For example, the comparison of FvM with FvF and FvG libraries combined identified 183 putatively differentially expressed TCs, 38 of which were present only in FvM (Table 4). Twenty-eight of the 38 are similar ($E < 10^{-5}$) to genes with known functions. Five of these (TC26530, TC26635, TC30015, TC30558, and TC30268) are similar to genes involved in different aspects of regulation (Table 6). The comparison of FvG with FvF and FvM combined identified 179 putatively differentially expressed TCs, 10 of which were present only in FvG (Table 4). Five of the 10 share significant similarity ($E < 10^{-5}$) with genes with known function but none to genes with a known regulatory function. It is possible that one or more of the five other FvG-specific genes, which are not similar to any genes with known functions, could be involved in fumonisin regulation. Understanding the role, if any, these genes play in fumonisin biosynthesis will require further analysis.

3.4. Alternatively spliced forms (ASFs)

The comparison of the ESTs to the *FUM* cluster sequence also indicated that some *FUM* genes matched multiple TCs that differed only by likely intron sequence(s). For example, TCs TC30353, TC30354, and

Table 6
TCs that correspond to putative fumonisin biosynthetic regulatory genes based on their differential occurrence in libraries FvF, FvG, and FvM

TC	BLAST homology[a]	Accession No.	E value	F or G[b]	M
30131	Replication factor C gene	CAA07618	3×10^{-97}	2	20
26574	Src Homology 3 gene	CAE47916	2×10^{-19}	0	18
26727	Zinc-binding transcription factor	AAO64397	1×10^{-07}	0	21
26530	Elongation factor-2 kinase	NP_037434	7×10^{-22}	0	19
26635	Activating transcription factor 4 (ATF-4)	NP_990211	6×10^{-07}	0	17
30015	Ubiquitination/deubiquitination factor	AAO26698	7×10^{-72}	0	17
30558	Frequency clock protein (frq)	T42013	4×10^{-86}	0	13
30268	Molecular chaperone; assists protein folding	NP_036205	0.0	0	11

[a] TC sequences were BLASTX searched against the NCBI-NR database and the hit with the best E (expected) value is listed. Lower E values reflect a higher probability that the function of the protein encoded by the TC is similar.

[b] Letters refer to specific library (i.e., F = Library FvF). The number indicates the number of ESTs that comprise each TC from the indicated library.

TC30355 all align to *FUM11*. The initial description of *FUM11* indicated that it had four introns (Proctor et al., 2003). Sequences corresponding to these introns were absent from TC30354 indicating that the intron sequences were spliced out of the corresponding transcripts. In contrast, sequences for the first *FUM11* intron were present in TC30353 and TC30355, indicating that the intron was not spliced. Likewise, the sequence corresponding to the third *FUM11* intron is present in TC30355 (Fig. 2).

To further explore *FUM* gene ASFs in *F. verticillioides*, we examined all the *FUM* ESTs (>700 ESTs) from all libraries (Table 5). ESTs from most *FUM* genes were present in three libraries: 232 in FvM, 342 in FvG, and 137 in FvN. There are 40 predicted introns in 12 of the 15 *FUM* genes. The *FUM* gene ESTs span 35 of these introns for an average of 22.4 ESTs per intron. The coverage ranges from a minimum of once (i.e., the third and fourth *FUM19* introns) to a maximum of 94 (i.e., the fourth *FUM14* intron). Examination of the 15 *FUM* gene EST sets revealed ASFs for seven *FUM* genes, where at least one intron had been retained or had an alternative 3′ border(s) (Table 7 and Fig. 3). The vast majority of ASFs are unlikely to encode full-length proteins because they contain stop codons and/or frame shifts in the coding sequence.

To confirm the presence of ASFs in mRNA, we carried out RT-PCR with primers 727 and 728, which were expected to amplify a fragment that spanned the third intron of *FUM14*. Electrophoretic and sequence analysis of the RT-PCR product revealed the presence of two fragments. The first was 186 bp in length and consisted of the expected fragment that lacked the sequence corresponding to the third intron sequence. The second fragment was 245 bp in length and was identical to the first except that it included the 59-bp sequence for the third intron. The detection of *FUM14* transcripts with a non-excised third intron in mRNA is consistent with the EST data.

3.5. ASF analysis

Two observations provide circumstantial evidence that the ASFs are not random occurrences of mis-splicing and may serve a function. First is the sheer number and distribution of ASFs. We examined over 700 *FUM* gene ESTs and found 78 ASFs that corresponded to seven of the 15 *FUM* genes. For the most part, a normal or correct splice form is any transcript where all possible introns have been excised, whereas ASFs retain one or more introns that introduce a stop codon or a frame-shift. The percent ASFs for *FUM11*, *FUM14*, *FUM16*, and *FUM18* were 47, 11, 51, and 44% based on 19, 134, 90, and 9 ESTs, respectively, and the percent ASFs for *FUM8* and *FUM12* were 3 and 2% based on 65 and 85 ESTs, respectively. The *FUM15* ASFs are unique in that the retention of the third intron in eight of nine ESTs does not interfere with the ORF.

Second, the ASFs were differentially expressed over time. More ASFs were found at 96 h (FvG) than at 48/72 h (FvM) for *FUM11*, *FUM14*, *FUM16*, and *FUM18* (Table 5). In the case of *FUM11*, two of five transcripts at 48/72 h were ASFs (e.g., retained an intron) while all seven transcripts at 96 h retained an intron. Thus,
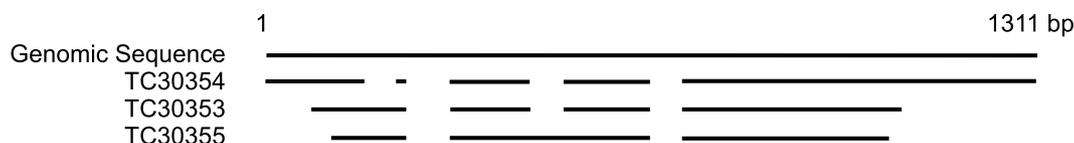
Fig. 2. Diagram of multiple sequence alignment spanning 1310 bp of the *FUM11* genomic sequence and three tentative consensus sequences generated during the assembly of the TIGR Gene Index. TC30353 is derived from six ESTs, TC30354 is derived from nine ESTs, and TC30355 is derived from four ESTs. Failure to excise the first intron introduces two stop codons in the predicted protein coding ORF. Each gap represents sequence that is missing from the respective TC.

Table 7
FUM gene introns, TCs/ESTs, and ASFs

Gene	# introns	TCs and singletons	Total ESTs	% TOs[a]
FUM1	6	26328, 26329[b]	41	
FUM6	3	26332	44	
FUM7	0	29823[c]	17	
FUM8	5	29823, 29825	65	3
FUM3	0	29917	41	
FUM10[d]	2	29844	91	
FUM11	4	30353, 30354, 30355	19	47
FUM12	3	26246, FVNDE14TV, FVGAW46TH	85	2
FUM13	0	29927	40	
FUM14[e]	4	29764, 29765, 29767, 29768	150	11
FUM15	1	26133, 26137	21	Note[f]
FUM16	3	26133, 26134	90	51
FUM17	1	26780	6	
FUM18	4	26779, 26781	9	44
FUM19	4	33164	1	

[a] TO, truncated ORF.
[b] All five ESTs that make up TC26329 identified a new *FUM1* intron.
[c] TC29823 includes EST for both *FUM7* and *FUM8*.
[d] FVNB289TH and FVNB547TH share partial identity to *FUM10* but are artifacts. Only about half of each EST sequence matches well with *FUM10* genomic sequence.
[e] TC29766 shares partial identity to *FUM14* but is an artifact. The first 864 bp of this TC is not similar to any sequence in the fumonisin cluster while the last 723 bp shares 98% identity with *FUM14*.
[f] The predicted protein from the *FUM15* ASF with the intron is 26 AAs longer than the predicted protein from the *FUM15* ASF that lacks the intron.

although Northern analysis would suggest that the relative expression of *FUM11* at the two time points was similar, none of the *FUM11* transcripts at 96 h are capable of generating a full-length protein. Overall, the number of ESTs representing different isoforms (or ASFs) increased 4.4-fold from 48/72 to 96 h. In general, ASFs for other *F.*

verticillioides genes do not exhibit similar expression patterns. The TIGR Gene Index identified 247 genes with putative ASFs among 7198 tentative consensus sequences (http://www.tigr.org/tdb/tgi/fungi.shtml). Of 119 clusters that we examined closely, 29 contain ESTs derived from the GYAM medium libraries (FvF, FvM or FvG). Only three of these exhibited a pattern of ASF expression similar to *FUM* gene ASFs (data not shown). Thus, the pattern of expression for *FUM* gene ASFs is unusual compared to other ASFs present in the FvF, FvM, and FvG libraries. We are currently in the process of examining the relative occurrence of all *FUM* gene ASFs under different growth conditions via microarrays. It should also be noted that introns within the same gene were alternatively spliced to different degrees. For example, the first *FUM11* intron was processed in all 19 EST sequences that span the intron while the second intron was retained in five of 19 ESTs and the fourth intron was retained in nine of 15 ESTs that span the intron. The significance of this observation is not known.

Alternatively or differentially spliced transcripts can regulate gene expression positively and/or negatively in numerous mammalian systems (Black, 2003). Given this, it is possible that ASFs can regulate gene expression in fungi as well. For example, the transcripts or any resulting truncated proteins could interfere with transcription, translation or protein function. Alternatively, perhaps the non-processed transcripts serve as a reservoir for future translational potential. Intron excision may be delayed by an unknown mechanism and thus serve as a posttranscriptional regulatory control point whereby transcripts capable of generating full-length proteins are limited. In a similar way, the *S. cerevisiae HAC1* mRNA is constitutively expressed but is not functional until it is spliced (Kawahara et al., 1998). The functional Hac1p is

```
FUM14   3rd intron                                                                                          #      #
                                                                                                           ESTs   AAs
Genomic  …ACTTTCTATGATAGTATGTATCAACCTACTCAGTTGGTCTTGTGAATGCCTATGACCGACAGCCAGAATAAGACTCAACTGTAAAAGACTTT…
Ver. 1   …ACTTTCTATGATAGTATGTATCAACCTACTCAGTTGGTCTTGTGAATGCCTATGACCGACAGCCAGAATAAGACTCAACTGTAAAAGACTTT…         11
           T  F  Y  D  S  M  Y  Q  P  T  Q  L  V  L  *1                                                             69
Ver. 2   …ACTTTCTATGATA::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::ACTCAACTGTAAAAGACTTT…   71
           T  F  Y  D                                                    N  S  T  V  K  D  F                        553
Ver. 3   …ACTTTCTATGATA:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::CCAGAATAAGACTGTAAAAGACTTT…   2
           T  F  Y  D                                          T  R  I  R  L  N  C  K  R  L2                        95
Ver. 4   …ACTTTCTATGATA:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::ACTTT…              1
           T  F  Y  D                                                                       N   F3                  548

FUM16 2nd intron

Genomic  …CTTGCAGGCAGTAAGTAGAGGCTTCAACCGCCTGTTCCAAAACAAATGCTAAAGTAGAGCTGACCAGATCCTCAACAGTCACTGGACTATTAGG…
Ver. 1   …CTTGCAGGCA::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::TCACTGGACTATTAGG…          31
           L  A  G                                                         I  T  G  L  L                            676
Ver. 2   …CTTGCAGGCA:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::ATCCTCAACAGTCACTGGACTATTAGG…      11
           L  A  G                                              N  P  Q  Q  S  L  D  Y  *                           276
Ver. 3   …CTTGCAGGCA::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::ACCAGATCCTCAACAGTCACTGGACTATTAGG…  2
           L  A  G                                              N  Q  I  L  N  S  H  W  T  I  R4                    282
```

Fig. 3. Differentially spliced *FUM14* and *FUM16* introns. The predicted amino acid (AA) sequences are indicated below the nucleotide sequences and the total number of predicted AAs is indicated to the right of the AA sequence. The number of ESTs corresponding to each version (ver.) found in the EST libraries is indicated to the right of the sequence. (1) The asterisk refers to a stop codon. (2) The 27th codon following the CTT codon is a stop codon. (3) Amino acids STVKD are predicted to be missing from proteins encoded by *FUM14* Ver. 4 transcripts. (4) The fourth codon following the AGG codon is a stop codon. It should be noted that 3' border sequence for the *FUM16* Ver. 3 intron is TG rather than the typical AG and was observed in two EST sequences.

required for the activation of the unfolded protein response in the endoplasmic reticulum. In this example, the 5′ and 3′ splice sites do not match the GT–AG consensus sequences (Kawahara et al., 1998). However, all of the unspliced introns described in our study contain the consensus sequence found in fungi and higher eukaryotes (Brown et al., 2001; Proctor et al., 2003). In contrast to *FUM11* and *FUM14*, it is difficult to speculate why isoforms with incomplete or truncated ORFs for *FUM16* and *FUM18* accumulate late in culture as neither is required for fumonisin synthesis (Proctor et al., 2003; Proctor and Butchko, unpublished observations).

ASFs have been previously described for a few genes in other fungi but on a very limited basis (Boel et al., 1984; Curach et al., 2004; Larrondo et al., 2004; McColl et al., 2003; Sachs and Yanofsky, 1991; Yadav et al., 2003). In all cases so far, their physiological impact is unknown. In the yeast *Xanthophyllomyces dendrorhous*, the percentage of ASFs with incomplete ORFs of a gene involved in carotenoid biosynthesis were also found to increase with culture age (Lodato et al., 2003). In contrast to these examples, many genes with ASFs have been found in large scale EST programs described at the TIGR Fungal Gene Indices (http://www.tigr.org/tdb/tgi/fungi.shtml) including 51 in *Cryptococcus*, 96 in *A. flavus*, 92 in *M. grisea*, and 63 in *Neurospora crassa*. To our knowledge, the function and occurrence of these ASFs have not yet been described in the literature. It is interesting to note that 19% of the 21 aflatoxin cluster genes found in the *A. flavus* EST collection had ASFs (Yuet al.,2004b)(http://www.tigr.org/tigr-scripts/tgi/splnotes .pl?species=a_flavus). Perhaps the role that ASFs play is most significant for genes involved in secondary metabolism.

Although a majority of *FUM* gene ASFs result in truncated ORFs, some *FUM14* and *FUM15* ASFs are predicted to encode proteins that differ minimally from those encoded by the majority of transcripts. The synthesis of alternative proteins is a hallmark and a significant consequence of differential intron splicing in higher eukaryotes (Black, 2003). The alternative *FUM14* transcript resulting from a different 3′ border for the third *FUM14* intron (Fig. 3) could form a protein with five fewer AAs than proteins derived from the major splice form. The five AAs are located at the C-terminal part of the phosphopantetheine-binding domain ≈25 AAs from the serine attachment site. Further studies will be required to determine whether this change affects protein function. The alternative *FUM15* transcripts resulted from retention of the second intron in eight of nine *FUM15* ESTs and would encode a protein with 26 AAs more than the EST with the intron excised. The 26 AAs are located within the variable length K-helix and three AAs from an absolutely conserved sequence ExxR thought to play a critical protein stabilizing role (Pylypenko and Schlichting, 2004). The importance of the variable length is unknown as both predicted proteins match very well to two different conserved domains

(NCBI database pfam00067 and COG2124) associated with different sets of cytochrome P450s (Marchler-Bauer et al., 2005). Determining whether the activity of the alternative Fum15p form differs from that of the more abundant form will require additional studies.

3.6. Revised FUM protein predictions

The 700 plus *FUM* gene transcripts afforded us the opportunity to reassess the predicted coding region for each *FUM* gene. Complete coding sequences for each putative *FUM* gene were predicted based on a combination of similarity with previously characterized proteins, existing ORFs, and presence or absence of intron border sequences (Proctor et al., 2003). A comparison of the *FUM* gene ESTs to the fumonisin cluster genomic sequence led to the identification of new introns or intron borders in nine of the original 15 genes and enabled us to adjust the predicted protein sequences appropriately (Table 5; Accession No. AF155773). In some cases, predicted protein changes were minimal. For example, the new predicted Fum12p had nine AAs deleted near the amino terminus, while the new Fum13p had five AAs added. In other cases, the predicted protein changes are more extensive. For example, the new Fum8p had 51 AAs changed within the middle of the protein, whereas Fum10p had 59 AAs added to the carboxy terminus and Fum15p had 20 AAs deleted near the carboxy terminus. These changes could affect functional studies using expressed proteins and the new information will guide subsequent research.

3.7. Identification of a new FUM gene

Our approach to sequence the cDNA libraries to significant depth led to the identification of a new *FUM* gene, designated *FUM20*, located between *FUM12* and *FUM13* in the *FUM* gene cluster (Fig. 1). The first evidence for the existence of this gene came from a comparison of TC26246 with genomic DNA sequence (Accession No. AF155773). The presence of three previously predicted *FUM12* introns was confirmed (Proctor et al., 2003) as well as a fourth intron with non-typical border sequences. Analysis of the 87 ESTs that make up TC26246 indicated that the sequence bordering the fourth intron was derived from a single EST, FVGBR83TH, consisting of 680 nts (Fig. 4). The presence of a poly(A) tract in FVGBR83TH indicates a direction of transcription opposite to that of *FUM12*, and when viewed in this orientation, the intron borders are typical and support the identification of *FUM20*. The 3′ border for FVGBR83TH includes the predicted *FUM12* stop codon and overlaps a number of the *FUM12* ESTs by up to 200 nts. The 5′ border of FVGBR83TH is 92 nts from the start codon of *FUM13* and 43 nts from the end of TC29927 (Fig. 4).
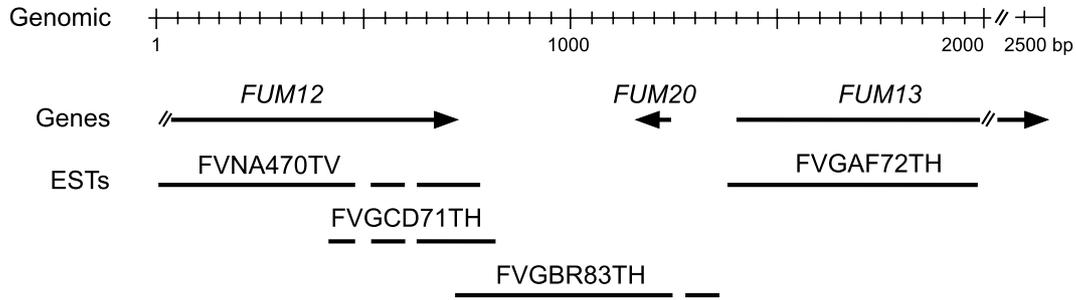
Fig. 4. Discovery of *FUM20*. Diagram shows gapped alignments of representative ESTs from *FUM12* (FVNA470TV, 907 nts, and FVGCD71TH, 670 nts) and *FUM13* (FVGAF72TH, 686 nts, and FVGBR83TH, 680 nts) with *F. verticillioides* genomic sequence within the fumonisin gene cluster. Relative scale is shown in basepairs. Gene names are indicated above each arrow and arrowheads refer to the direction of gene transcription. The gaps in the ESTs sequences indicate introns. The *FUM20* ORF could not be determined.

What role, if any, *FUM20* plays in fumonisin biosynthesis is not clear. Both BLASTX of FVGBR83TH and BLASTP of the predicted 40-AA *FUM20* protein did not suggest a possible function as neither share significant similarity with any previously described genes or proteins in the NCBI-NR database. In addition, TBLASTN of Fum20p against the *F. graminearum* genome sequence did not reveal any sequence similarity. It is possible that the *FUM20* transcript is an ASF. The excision of additional intron(s) may lead to a predicted protein that shares similarity to proteins with known function. Alternatively, the *FUM20* transcript may be non-coding RNA that regulates *FUM* gene expression. Over-expression and gene deletion studies are underway to help determine the function of *FUM20*.

3.8. Maize-tissue induced ESTs

To identify genes involved in pathogenesis of *F. verticillioides* on maize, we created four cDNA libraries (FvH, FvI, FvN, and FvJ) from fungal RNA extracted after growth on maize-based media (Table 1). Due to the small number of ESTs in library FvJ, it has limited value. A LES comparison of libraries FvH, FvI, and FvN with the four GYAM-derived libraries (FvF, FvG, FvM, and FvO; Table 1) identified 743 differentially expressed TCs ($R > 9$) some of which share similarity with previously recognized fungal virulence genes (Table 8). For example, TC27603 shares significant similarity with a copper-transporting ATPase of *Colletotrichum lindemuthianum* that is critical for infection (Parisot et al., 2002). In addition, TC31240 shares significant similarity with a chitin-binding protein from the yeast *Pichia inositovara* that is part of a killer activity toward *S. cerevisiae* (Klassen and Meinhardt, 2003).

There were three differentially expressed TCs that match previously identified *F. verticillioides* hydrophobins and another (TC26886) that shares similarity to a protein (snodprot1) produced by *Phaeosphaeria nodorum* during infection of wheat leaves and is a member of a family of proteins with hydrophobin-like properties (Pazzagli et al., 1999). The reference protein for the snodprot1 protein family, cerato-platanin, produced by *Ceratocystis fimbriata*, is phytotoxic (Boddi et al., 2004). Hydrophobins are known to be involved in recognition processes with other organisms and hence likely play an important role in pathogenesis (Ebbole, 1997). Recently, the three hydrophobin genes described above plus two additional hydrophobin genes were disrupted and none were required to cause maize seedling disease

Table 8
Fusarium verticillioides plant-associated genes with similarity to virulence genes from other fungi

TC	Putative function	Accession No.	E value	FGMO[a]	IH or N
27603	Copper-transporting	AAN62846	0.0	0	6 I, 1 H
31240	Chitin-binding protein	CAD91890	3×10^{-65}	1 F	6 H
29850	Hydrophobin 1[b]	AY155496	0.0	1 G	42 N, 27 I
26567	Hydrophobin 2	AY155497	0.0	0	27 N
30113	Hydrophobin 3	AY155498	0.0	1 O	6 I, 1 H
26886	Virulence protein	AAC26870	1×10^{-38}	0	14 N, 2 H
30234	P450	AAQ16576	8×10^{-69}	0	12 N
30242	P450	AAQ16576	7×10^{-66}	0	14 N
31401	Terpenoid cyclase	AAQ16575	5×10^{-22}	0	6 H
27406	Stress–response factor	NP_014371	4×10^{-41}	0	7 H, 2 N

[a] Letters refer to EST libraries, respectively (i.e., F = library FvF). The number indicates the number of ESTs that comprise each TC from the indicated library (i.e., 6 I = TC27603 includes six ESTs from library FvI).

[b] The three *F. verticillioides* hydrophobin genes have previously been described. See text for details.

(Fuchs et al., 2004). This finding does not preclude the possibility that hydrophobins are required for other maize diseases (e.g., ear rot) caused by *F. verticillioides*.

The LES analysis also detected four differentially expressed TCs (TC30234, TC30242, TC31401, and TC27406) that share similarity to genes (CNDs) regulated through the calcineurin pathway (Table 8). Calcineurin is an essential phosphatase and has been shown to be required for virulence in a number of fungal systems (Viaud et al., 2003). *CND5* and *CND15* are located adjacent to each other in *Botrytis cinerea* and are speculated to be involved in the synthesis of a secondary metabolite involved in virulence (Viaud et al., 2003). TC27406 shares significant similarity with a *S. cerevisiae* transcription factor (Crz1p) involved in stress response and whose nuclear localization is positively regulated by calcineurin-mediated dephosphorylation (Philippsen et al., 1997). The exact role of this or any of the potential virulence-associated genes described above in *F. verticillioides* will be determined by further studies.

3.9. Comparison of ESTs to three fungal genome sequences

Fusarium verticillioides is closely related to *F. graminearum* which can also cause ear rot and other maize diseases. BLAST analysis of the *F. verticillioides* ESTs in the Gene Index revealed that 86% of the ESTs shared significant similarity ($E < 10^{-5}$) with portions of *F. graminearum* genome sequence and 81% had significant similarity ($E < 10^{-5}$) to the ~13,700 *F. graminearum* genes predicted by the Munich Information Center for Protein Sequences (http://mips.gsf.de/genre/proj/fusarium). On the basis of the likelihood that the number of genes in *F. graminearum* is similar to *F. verticillioides*, the latter comparison suggests that our ESTs represent over 80% of all possible *F. verticillioides* genes. Experiments are in progress to determine whether some of the genes unique to *F. verticillioides* are involved in aspects of the *F. verticillioides*–maize interaction that differ from the *F. graminearum*–maize interaction.

We also compared the EST sequences against the genomic sequences of the saprophytic (non-pathogenic) fungi *Aspergillus nidulans* and *N. crassa* (Fig. 5). As expected, the number of *F. verticillioides* sequences with high levels of similarity to genes from the saprophytes was less than that observed for *F. graminearum*. We then sought to identify unique *Fusarium* genes that may be critical for these fungi to cause disease. We reasoned that these genes could be identified by taking the set of *F. verticillioides* ESTs most similar to *F. graminearum* (≈2000 TCs and singletons with a BLASTX threshold $\leqslant 10^{-200}$) and removing all those that are shared with *Aspergillus* or *Neurospora* ($E \leqslant 10^{-5}$ in BLASTX). This analysis yielded 39 *F. verticillioides* ESTs and TCs. Twenty-two of these share similarity to enzyme-encod-
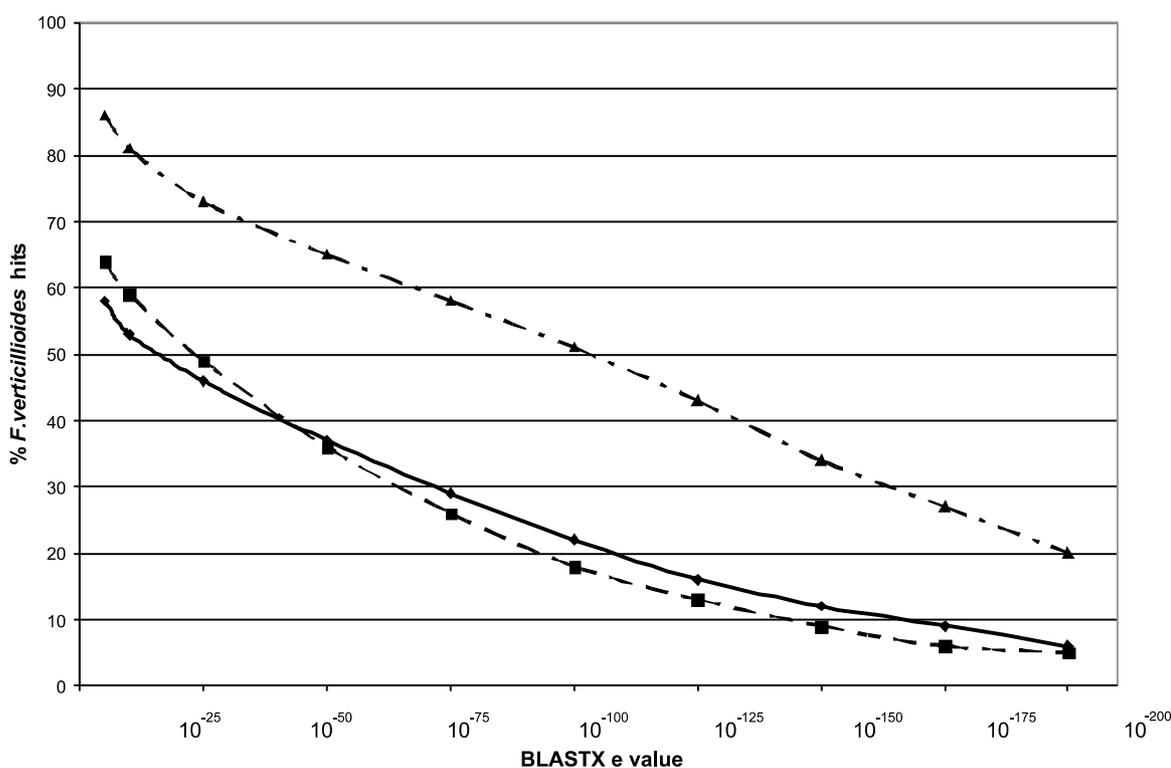

Fig. 5. Comparison of *F. verticillioides* Gene Index sequences against *F. graminearum* (▲, - – -), *Aspergillus nidulans* (■, - - -), and *Neurospora crassa* (◆, ——) protein sequences. Vertical axis represents percent *F. verticillioides* matches and the horizontal axis represents BLASTX *E* value of the matches.

ing genes involved in modifications of low molecular weight compounds or proteins (e.g., monooxygenase), and 15 do not share similarity to any previously described genes. The two remaining ESTs share similarity to putative regulatory genes; TC27771 shares similarity with a putative zinc finger protein (NP_594226) from *Schizosaccharomyces pombe* and TC32201 shares similarity with a gene encoding a WD40 repeat containing protein (ZP_00324864) from *Trichodesmium erythraeum*. We are in the process of determining whether these two genes are involved in plant pathogenesis by *F. verticillioides*.

4. Summary

Expressed sequence tags are an important tool in defining gene structure and understanding gene expression, transcript splicing, and protein structure, and are a critical complement to whole genome sequence efforts. Recently, 26,110 ESTs generated from an *A. flavus* normalized cDNA library led to the identification of 7218 putative genes which was estimated to represent 60% of the predicted genes, encoded by the *A. flavus* and genome (Yu et al., 2004b). The library was constructed from RNA from eight different media conditions seven of which support aflatoxin production (Yu et al., 2004b). The 87,059 ESTs from our nine cDNA libraries led to 11,119 putative *F. verticillioides* genes which we estimate to represent over 80% of the predicted genes. Both the *A. flavus* and the *F. verticillioides* collections should contribute significantly to a better understanding of the biology of host–pathogen interactions, and the biochemical and regulatory processes involved in mycotoxin biosynthesis.

An initial goal of our project was to find fumonisin regulatory genes by identifying genes that were differentially expressed genes within the fumonisin production medium (GYAM) libraries. Nine potential fumonisin regulatory genes have been identified and are being further characterized. The abundance and differential expression of ASFs among *FUM* genes raise the intriguing possibility that the regulation of intron splicing serves a role in fumonisin biosynthesis. Microarray experiments are in progress and may provide evidence to support our model that the *FUM* transcripts with truncated ORFs impact fumonisin production.

A second goal was to identify fungal genes that are important for the *F. verticillioides*–maize interaction. Numerous ESTs were identified that were differentially expressed by the fungus in the presence of maize material. Comparison of the ESTs to the genome sequence of two saprophytic and another plant pathogenic fungus identified genes that may be involved in the ability of *Fusarium* species to cause disease. Gene-disruption studies are in progress to determine whether any of these genes are required for toxin production or virulence of *F. verticillioides* on maize.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.

Black, D.L., 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. 72, 291–336.

Boddi, S., Comparini, C., Calamassi, R., Pazzagli, L., Cappugi, G., Scala, A., 2004. Cerato-platanin protein is located in the cell walls of ascospores, conidia and hyphae of *Ceratocystis fimbriata* f. sp. *platani*. FEMS Microbiol. Lett. 233, 341–346.

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

Bojja, R.S., Cerny, R.L., Proctor, R.H., Du, L., 2004. Determining the biosynthetic sequence in the early steps of the fumonisin pathway by use of three gene-disruption mutants of *Fusarium verticillioides*. J. Agric. Food Chem. 52, 2855–2860.

Brown, D.W., McCormick, S.P., Alexander, N.J., Proctor, R.H., Desjardins, A.E., 2001. A genetic and biochemical approach to study trichothecene diversity in *Fusarium sporotrichioides* and *Fusarium graminearum*. Fungal Gen. Biol. 32, 121–144.

Brown, D.W., Yu, J.-H., Kelkar, H.S., Fernandes, M., Nesbitt, T.C., Keller, N.P., Adams, T.H., Leonard, T.J., 1996. Twenty-five co-regulated transcripts define a sterigmatocystin gene cluster in *Aspergillus nidulans*. Proc. Natl. Acad. Sci. USA 93, 1418–1422.

Butchko, R.A., Plattner, R.D., Proctor, R.H., 2003a. *FUM9* is required for C-5 hydroxylation of fumonisins and complements the meiotically defined Fum3 locus in *Gibberella moniliformis*. Appl. Environ. Microbiol. 69, 6935–6937.

Butchko, R.A., Plattner, R.D., Proctor, R.H., 2003b. *FUM13* encodes a short chain dehydrogenase/reductase required for C-3 carbonyl reduction during fumonisin biosynthesis in *Gibberella moniliformis*. J. Agric. Food Chem. 51, 3000–3006.

Chou, H.H., Holmes, M.H., 2001. DNA sequence quality trimming and vector removal. Bioinformatics 17, 1093–1104.

Curach, N.C., Te'o, V.S., Gibbs, M.D., Bergquist, P.L., Nevalainen, K.M., 2004. Isolation, characterization and expression of the *hex1* gene from *Trichoderma reesei*. Gene 331, 133–140.

Desjardins, A.E., Plattner, R.D., Proctor, R.H., 1996. Linkage among genes responsible for fumonisin biosynthesis in *Gibberella fujikuroi* mating population A. Appl. Environ. Microbiol. 62, 2571–2576.

Ebbole, D.J., 1997. Hydrophobins and fungal infection of plants and animals. Trends Microbiol. 5, 405–408.

Ebbole, D.J., Jin, Y., Thon, M., Pan, H., Bhattarai, E., Thomas, T., Dean, R., 2004. Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: analysis of expressed sequence tags. Mol. Plant Microbe Interact. 17, 1337–1347.

Flaherty, J.E., Woloshuk, C.P., 2004. Regulation of fumonisin biosynthesis in *Fusarium verticillioides* by a zinc binuclear cluster-type gene, *ZFR1*. Appl. Environ. Microbiol. 70, 2653–2659.

Fuchs, U., Czymmek, K.J., Sweigard, J.A., 2004. Five hydrophobin genes in *Fusarium verticillioides* include two required for microconidial chain formation. Fungal Gen. Biol. 41, 852–864.

Glenn, A.E., Gold, S.E., Bacon, C.W., 2002. *Fdb1* and *Fdb2*, *Fusarium verticillioides* loci necessary for detoxification of preformed antimicrobials from corn. Mol. Plant Microbe Interact. 15, 91–101.

Glenn, A.E., Hinton, D.M., Yates, I.E., Bacon, C.W., 2001. Detoxification of corn antimicrobial compounds as the basis for isolating *Fusarium verticillioides* and some other *Fusarium* species from corn. Appl. Environ. Microbiol. 67, 2973–2981.

Hohn, T.M., Krishna, R., Proctor, R.H., 1999. Characterization of a transcriptional activator controlling trichothecene toxin biosynthesis. Fungal Genet. Biol. 26, 224–235.

Howard, P.C., Eppley, R.M., Stack, M.E., Warbritton, A., Voss, K.A., Lorentzen, R.J., Kovach, R.M., Bucci, T.J., 2001. Fumonisin B1 carcinogenicity in a two-year feed study using F344 rats and B6C3F1 mice. Environ. Health Perspect. 109, 277–282

Huang, X., Adams, M.D., Zhou, H., Kerlavage, A.R., 1997. A tool for analyzing and annotating genomic sequences. Genomics 46, 37–45.

Kawahara, T., Yanagi, H., Yura, T., Mori, K., 1998. Unconventional splicing of HAC1/ERN4 mRNA required for the unfolded protein response. Sequence-specific and non-sequential cleavage of the splice sites. J. Biol. Chem. 273, 1802–1807.

Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C., Hutchinson, C.R., 1999. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. Science 284, 1368–1372.

Klassen, R., Meinhardt, F., 2003. Structural and functional analysis of the killer element pPin1-3 from *Pichia inositovora*. Mol. Genet. Genomics 270, 190–199.

Larrondo, L.F., Gonzalez, B., Cullen, D., Vicuna, R., 2004. Characterization of a multicopper oxidase gene cluster in *Phanerochaete chrysosporium* and evidence of altered splicing of the *mco* transcripts. Microbiology 150, 2775–2783.

Lodato, P., Alcaino, J., Barahona, S., Retamales, P., Cifuentes, V., 2003. Alternative splicing of transcripts from *crtI* and *crtYB* genes of *Xanthophyllomyces dendrorhous*. Appl. Environ. Microbiol. 69, 4676–4682.

Marasas, W.F., Riley, R.T., Hendricks, K.A., Stevens, V.L., Sadler, T.W., Gelineau-Van Waes, J., Missmer, S.A., Cabrera, J., Torres, O., Gelderblom, W.C., Allegood, J., Martinez, C., Maddox, J., Miller, J.D., Starr, L., Sullards, M.C., Roman, A.V., Voss, K.A., Wang, E., Merrill Jr., A.H., 2004. Fumonisins disrupt sphingolipid metabolism, folate transport, and neural tube development in embryo culture and in vivo: a potential risk factor for human neural tube defects among populations consuming fumonisin-contaminated maize. J. Nutr. 134, 711–716.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., Bryant, S.H., 2005. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res. 33, D192–D196.

McColl, D., Valencia, C.A., Vierula, P.J., 2003. Characterization and expression of the *Neurospora crassa nmt-1* gene. Curr. Genet. 44, 216–223.

Munkvold, G.P., Desjardins, A.E., 1997. Fumonisins in maize. Can we reduce their occurrence?. Plant Dis. 81, 556–565.

OBrian, G., Fakhoury, A.M., Payne, G.A., 2003. Identification of genes differentially expressed during aflatoxin biosynthesis in *Aspergillus flavus* and *Aspergillus parasiticus*. Fungal Genet. Biol. 39, 118–127.

Parisot, D., Dufresne, M., Veneault, C., Lauge, R., Langin, T., 2002. *clap1*, a gene encoding a copper-transporting ATPase involved in the process of infection by the phytopathogenic fungus *Colletotrichum lindemuthianum*. Mol. Genet. Genomics 268, 139–151.

Pazzagli, L., Cappugi, G., Manao, G., Camici, G., Santini, A., Scala, A., 1999. Purification, characterization, and amino acid sequence of cerato-platanin, a new phytotoxic protein from *Ceratocystis fimbriata* f. sp. *platani*. J. Biol. Chem. 274, 24959–24964.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J., 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19, 651–652.

Philippsen, P., Kleine, K., Pohlmann, R., Dusterhoft, A., Hamberg, K., Hegemann, J.H., Obermaier, B., Urrestarazu, L.A., Aert, R., Albermann, K., Altmann, R., Andre, B., Baladron, V., Ballesta, J.P., Becam, A.M., Beinhauer, J., Boskovic, J., Buitrago, M.J., Bussereau, F., Coster, F., Crouzet, M., D'Angelo, M., Dal Pero, F., De Antoni, A., Hani, J., et al., 1997. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIV and its evolutionary implications. Nature 387, 93–98.

Proctor, R.H., Brown, D.W., Plattner, R.D., Desjardins, A.E., 2003. Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in *Gibberella moniliformis*. Fungal Genet. Biol. 38, 237–249.

Pylypenko, O., Schlichting, I., 2004. Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s. Annu. Rev. Biochem. 73, 991–1018.

Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R., White, J., 2001. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res. 29, 159–164.

Quackenbush, J., Liang, F., Holt, I., Pertea, G., Upton, J., 2000. The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. 28, 141–145.

Sachs, M.S., Yanofsky, C., 1991. Developmental expression of genes involved in conidiation and amino acid biosynthesis in *Neurospora crassa*. Dev. Biol. 148, 117–128.

Shim, W.B., Woloshuk, C.P., 2001. Regulation of fumonisin B(1) biosynthesis and conidiation in *Fusarium verticillioides* by a cyclin-like (C-type) gene, *FCC1*. Appl. Environ. Microbiol. 67, 1607–1612.

Stekel, D.J., Git, Y., Falciani, F., 2000. The comparison of gene expression from multiple cDNA libraries. Genome Res. 10, 2055–2061.

Trail, F., Xu, J.R., San Miguel, P., Halgren, R.G., Kistler, H.C., 2003. Analysis of expressed sequence tags from *Gibberella zeae* (anamorph *Fusarium graminearum*). Fungal Genet. Biol. 38, 187–197.

Viaud, M., Brunet-Simon, A., Brygoo, Y., Pradier, J.M., Levis, C., 2003. Cyclophilin A and calcineurin functions investigated by gene inactivation, cyclosporin A inhibition and cDNA arrays approaches in the phytopathogenic fungus *Botrytis cinerea*. Mol. Microbiol. 50, 1451–1465.

Yadav, J.S., Soellner, M.B., Loper, J.C., Mishra, P.K., 2003. Tandem cytochrome P450 monooxygenase genes and splice variants in the white rot fungus *Phanerochaete chrysosporium*: cloning, sequence analysis, and regulation of differential expression. Fungal Genet. Biol. 38, 10–21.

Yu, J., Bhatnagar, D., Cleveland, T.E., 2004a. Completed sequence of aflatoxin pathway gene cluster in *Aspergillus parasiticus*. FEBS Lett. 564, 126–130.

Yu, J., Whitelaw, C.A., Nierman, W.C., Bhatnagar, D., Cleveland, T.E., 2004b. *Aspergillus flavus* expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops. FEMS Microbiol. Lett. 237, 333–340.

Zhang, Z., Schaffer, A.A., Miller, W., Madden, T.L., Lipman, D.J., Koonin, E.V., Altschul, S.F., 1998. Protein sequence similarity searches using patterns as seeds. Nucleic Acids Res. 26, 3986–3990.