Analysis of Expressed Sequence Tags from the Fungal Banana Pathogen *Mycosphaerella fijiensis*

Yangrae Cho[1], Shaobin Hou[2], and Shaobin Zhong[*,1,3]

[1]*Department of Plant and Environmental Protection Sciences, University of Hawaii at Manoa, 3190 Maile Way, Honolulu, HI 96822, USA*

[2]*Center for Genomics, Proteomics and Bioinformatics Research Initiative, University of Hawaii at Manoa, Honolulu, HI 96822, USA*

[3]*Department of Plant Pathology, North Dakota State University, 306 Walster Hall, Fargo, ND 58105, USA*

Abstract: *Mycosphaerella fijiensis* is a fungal pathogen causing black leaf streak disease in banana plants. In order to develop our understanding of the molecular mechanisms of host infection by *M. fijiensis*, we initiated a first step towards large-scale gene discovery through the generation and analysis of expressed sequence tags (ESTs). Three cDNA libraries representing different culture conditions were constructed. A subset of clones was sequenced at the 5' end and a total of 3,771 ESTs were generated. These ESTs were clustered into a set of 1,945 unique sequences. BLASTX analysis indicated that 410 of the unisequences showed homology to genes with known function, including those related to pathogenicity and virulence. Meanwhile, 66% and 89% of the ESTs matched to the predicted transcripts and genome sequences of *M. fijiensis*, respectively. The diversity of the EST collection developed in this study will facilitate genome annotation as well as identification and characterization of genes involved in host infection and fungal development through a comparative and functional genomics approach.

INTRODUCTION

Mycosphaerella fijiensis causes black leaf streak, commonly known as black Sigatoka and the most destructive fungal disease of banana worldwide. It can affect both old leaves and metabolically active young leaves, cause yield losses up to 50% or more, and incur serious export problem due to premature fruit ripening [1, 2]. The disease was first reported in the Sigatoka Valley of Fiji in 1963 and subsequently detected throughout the Pacific and Southeast Asia. It was later identified at Honduras in 1972, Zambia in 1973, and throughout Latin American, African, and Australian banana growing regions by 2000 [3-5].

Disease management is one of the most expensive costs faced by banana growers in the high-rainfall tropics and subtropics, where bananas are of great economic and nutritional importance. Fungicides have been extensively utilized in commercial plantations to manage black Sigatoka. After many years of fungicide-dependent disease control on large-scale farms, however, *M. fijiensis* strains tolerant or resistant to diverse fungicides have appeared [6-8]. These resistant strains have created the need for more frequent fungicide applications. In some areas up to 60 spray applications may be needed to control the disease during a 9-month growing season. The systematic application of fungicides raises concern about their impact on the environment and human health [9-12]. Cultural practices, such as the removal of affected leaves, adequate spacing of plants, and efficient drainage within plantations, have played an important role in removing or reducing optimum conditions for disease development. However, these measures are relatively expensive and insufficient to control the disease without fungicides.

Commercial banana cultivars, the major hosts of *M. fijiensis*, are hybrids of two wild species, *Musa acuminata* (A genome) and *M. balbisiana* (B genome). These cultivars have evolved largely *via* asexual propagation, resulting in limited genetic variation and an increased susceptibility to pests and diseases [13]. Development of disease-resistant plants from commercial cultivars is difficult through traditional breeding methods because they are mostly polyploid hybrids and sterile [14, 15]. Nevertheless, great efforts have been made in the past decades to develop new banana cultivars resistant to black Sigatoka by introducing resistance genes from wild diploid bananas [16, 17]. Some of the first high-yielding, disease-resistant hybrids are available for wide-scale testing and distribution to farmers [18]. However, due to the genetic variability of the pathogen, some resistant banana accessions have become susceptible to *M. fijiensis* isolates in certain regions [19]. Although there have been extensive studies on the genetic diversity of *M. fijiensis* using molecular markers, little is known about the molecular mechanism of the host-pathogen interaction in this pathosystem.

In the past several years, EST analyses have been conducted for many filamentous fungi, including *Fusarium graminearum* [20], *Heterobasidion annosum* [21], *Magnaporthe grisea* [22], *Mycosphaerella graminicola* [23], *Neurospora crassa* [24], and *Ustilago maydis* [25]. In conjunction with fungal genome databases, specific EST databases have appeared and enabled researchers to comprehensively analyze ESTs in order to identify putative pathogenic-

*Address correspondence to this author at the Department of Plant Pathology, North Dakota State University, 306 Walster Hall, Fargo, ND 58105, USA; Tel: 701-231-7427; Fax: 701-231-7851;
E-mail: shaobin.zhong@ndsu.edu

1874-4370/08

ity and virulence genes by comparative genomics approaches [20, 23]. However, analysis of large scale ESTs has not been reported for *M. fijiensis*. The objective of this project was to discover genes through EST analysis of three cDNA libraries from *M. fijiensis* growing under three different culture conditions in order to accomplish our long-term goal to identify and characterize the genes associated with pathogenicity, fungicide resistance, and the life cycle of *M. fijiensis*. In this report we present an analysis of the ESTs from three cDNA libraries and these ESTs will be useful for the whole genome annotation as well as for the understanding of the genetic and molecular interactions between *M. fijiensis* and its banana host.

MATERIAL AND METHODS

Fungal Isolates and Growth Conditions

M. fijiensis isolates were obtained from perithecia using the method described by Stover [26]. Briefly, banana leaves infected by *M. fijiensis* were collected in the field and incubated in plastic bags with moist paper towels for 24 hours. Infected leaf tissue was cut into 10 cm^2 pieces and stapled to filter papers. The filter papers were wetted, stuck beneath the lids of Petri dishes containing 7.5% water agar, and incubated 6-12 h until ascopsores were discharged. Individual ascospores were transferred to Potato Dextrose Agar (PDA) using a fine scalpel. Typical pink-colored colonies formed after incubation for 2-3 weeks at an ambient temperature (22-25°C) and continuous, cool-white fluorescent light (1200-1250 lux). Identification of the *M. fijiensis* isolates was confirmed by PCR using *M. fijiensis*-specific primers [27].

Two *M. fijiensis* isolates (MF5-Hilo and MF11-Hilo) were used to construct cDNA libraries. Mycelia for construction of the first cDNA library (PA) were obtained from MF11-Hilo grown on PDA for 4 weeks under continuous, cool-white, fluorescent light (1200-1250 lux). For construction of the second cDNA library (FL), a colony of MF5-Hilo on PDA was transferred into a sterile mortar containing 5 ml of Fries liquid medium [5.0 g ammonium tartrate, 1.0 g ammonium nitrate, 0.5 g magnesium sulfate, 1.3 g potassium phosphate (dibasic), 2.6 g potassium phosphate (monobasic), 30.0 g sucrose, 1.0 g yeast extract, dissolved in 1000 ml water] and ground with a pestle. The ground mycelia were used to inoculate 100 ml of Fries liquid medium in a 250 ml flask, which was then incubated at 26°C on a shaker at 125 rpm. After 4 days, the mycelia were harvested by filteration through two-layers of MiraCloth (CalBiochem, La Jolla, CA), washed once with distilled water, then snap-frozen in liquid nitrogen before storing at –80°C. Mycelia for construction of the third cDNA library (HX) were prepared as described for cDNA library FL, but 30 ml of banana leaf extract was added to the 4-day-old mycelium, then cultured for an additional 24 hours. The banana leaf extract was prepared by homogenizing 30 g of young banana leaves in 50 ml sterile water, filtering it through four layers of cheesecloth, then filter-sterilizing it using a 0.22 μm filter.

RNA Isolation and cDNA Library Construction

Total RNA was isolated from fungal mycelia with the RNeasy Plant Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. All three cDNA libraries were constructed using the SMART cDNA Library Construction Kit (DB-Bioscience-ClonTech, Palo Alto, CA) and the Gateway (Invitrogen, Carlsbad, CA) cloning system with modifications from manufacturers' protocols (Fig. **1**). For first-strand cDNA synthesis, 0.5 to 1.0 μg of total RNA was mixed with 2 μl of each anchored attB2-oligo-dT primer (10 mM) (GGGGACCACTTTGTACAAGAAAGCTGGGTTTT TTTTTTTTTTTTTTTVN) and attB1-Switch primer (10 mM) (GGGGACAAGTTTGTACAAAAAAGCAGGCTrGr GrG), where rG represents ribonucleotide. The mixture was incubated at 68°C for 2 minutes, then cooled on ice for 2 minutes before adding 4 μl of 5× buffer, 2 μl of DTT (100 mM), 1μl of dNTPs (10 mM), and 1 μl of Superscript III reverse transcriptase (5 units/μl). Water was added to a final volume of 20 μl. The reverse transcription reaction was incubated for 1 hour in a thermocycler at 50°C. Two microliters of the first-strand cDNA was used as a template in a 100 μl PCR reaction containing 1x Advantage 2 PCR buffer, 2.0 U of Advantage 2 Polymerase Mix (Clontech, CA), 0.25 μM each of primers attB1 (GGGGACAAGTTTGTACAAAAAAGCAGGCT) and attB2 (GGGGACCACTTTGTACAAGAAAGCTGGGT), and 2 μl of dNTP (10 mM). PCR was performed at 94°C for 2 minutes for the initial denaturation, followed by ~18 to 22 cycles at 94°C for 20 seconds, 61°C for 30 seconds, and 72°C for 5 min. PCR products were visualized on 1% agarose gels. The amplified cDNA products were processed on a Chroma spin column (DB-Bioscience-ClonTech, Palo Alto, CA) to remove fragments smaller than 500 base pairs. The purified cDNA was then incorporated into the plasmid pDONR221 by recombination using BP clonase as described in the manufacturer's protocol (Invitrogen, Carlsbad, CA). The plasmids were transformed into DH10B *E. coli* strain.

DNA Sequencing and Data Analysis

The cDNA libraries were spread on Luria-Bertani (LB) agar plates containing 50 μg/ml kanamycin. Individual colonies from each of the three libraries were transferred into 96-well microtiter plates with LB containing kanamycin and 18% glycerol. The plates from each library were sent to the Center for Genomics, Proteomics and Bioinformatics Research Initiative, University of Hawaii at Manoa, for sequencing on an ABI Model 377 DNA sequencer. Sequencing reactions were performed with standard Big Dye (Applied Biosystems, Foster City, CA) protocols and sequences were determined only at the 5′ end of each clone.

Raw sequence chromatogram trace files were processed by the program Phred (http://www.phrap.org) using its trimming options in order to remove low quality ends [28]. All automated sequences were manually examined in order to ensure high quality; incorrectly called bases were corrected and unassigned bases were correctly identified. Cross_match in the Phred-Phrap was used to automatically remove vector sequences with the aid of a custom database containing vectors and adapters used for cloning in the libraries. The program was also used to assemble the resulting EST sequences into overlapping contigs and singlets. Redundancy was calculated with the formula [(total number of sequences – total number of unisequences) / (total number of sequences)]. All ESTs have been submitted to the NCBI GenBank dbEST (accession numbers FB531204-FB535042).

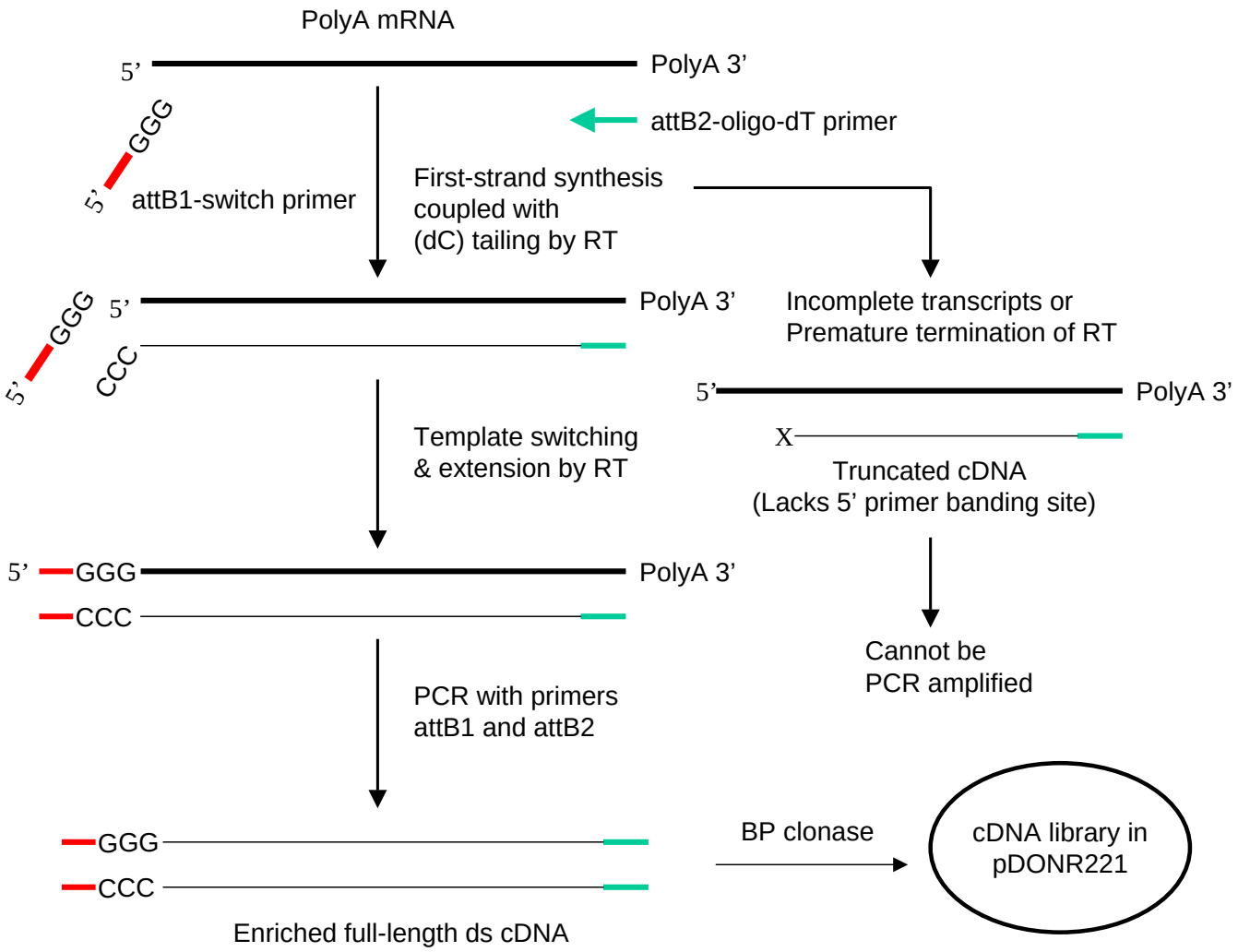

Fig. (1). Schematic diagram depicting how cDNA libraries were constructed, starting from mRNA in the total RNA in order to enrich full-length cDNA clones.

BLAST searches were conducted with stand-alone BLAST programs available from ftp://ftp.ncbi.nih.gov/blast/executables/. ESTs were searched against the non-redundant NCBI database for homologs using the Blastx algorithms [29]. Sequences showing significant similarity (E-value < 1e-05) to characterized proteins were categorized based on the MIPS scheme outlined in the MIPS database [30]. Comments from the header section of the NR records were the primary sources of information used to assign putative functions to the *M. fijiensis* unisequences. The unisequences were also queried against the predicted protein sequences of 18 sequenced fungal genomes using BLASTX. The predicted transcript sequences were downloaded on June 1, 2008 from two websites: the Fungal Genome Initiative at the Broad Institute (http://www.broad.mit.edu/annotation/fungi/fgi/) and Eukaryotic Genomics at the DOE Joint Genome Institute (JGI) (http://genome.jgi-psf.org/euk_cur1.html). For each genome analyzed, we calculated the number and percentage of *M. fijiensis* unisequences that had a homolog with E-values of <1e-05, <1e-40, and <1e-100. Further more, we compared the ESTs to the draft assembly of whole genome sequence reads of *M. fijiensis*, which included 382 scaffolds and 10,313 predicted transcripts. First, the 3,771 ESTs were queried against the database of the predicted transcripts using BLASTN. Queried sequences that had a minimum 90% identity over at least 50 base pairs of alignment to a predicted transcript sequence were considered significant (E-value < 1e-5) and exported to an Excel database for further analysis. Sequences that did not have hits were used with the BLASTN algorithm to query the database of 382 scaffolds representing the draft sequence of the whole genome. The results of these BLAST searches were used to classify the origin of each EST and evaluate its usefulness for gene prediction. All scripts used for the above analyses were kindly provided by Dr. Zheng Jin Tu and run in the Unix system of the Supercomputer Institute of the University of Minnesota, St. Paul, MN.

RESULTS

Construction and Characteristics of cDNA Libraries

We constructed separate cDNA libraries using the total RNA isolated from fungal tissues grown under three culture conditions, i.e., on the PDA media (PA library), Fries liquid media (FL library), and Fries liquid media supplemented by host plant extract (HX library). Our primary goal was to understand the similarities and differences in gene expression patterns under these three different growth conditions. The secondary goal was to identify the infection related genes

among abundantly expressed ones by comparing EST collections to public databases.

Whole fungal tissues grown on PDA were used as RNA sources in order to identify genes expressed during growth on a hard agar surface. Meanwhile, mycelia grown in the Fries medium as well as in media supplemented by host plant extract were used as RNA sources in order to discover genes expressed under similar environmental conditions during host plant infection. Titers of 8.6×10^6, 5.2×10^6, and 3.5×10^6 colonies/ml were obtained from libraries PA, FL, and HX, respectively. The average titer for the three libraries was 5.8×10^6 colonies/ml. Examination of sixty randomly picked clones from each library by PCR with the primers attB1 and attB2 showed that over 98% of the clones had inserts ranging in size from 0.3kb to 3.5 kb. Average insert size was 1.2kb for PA, 1.3kb for FL and 1.1kb for HX.

Distribution of ESTs to Clusters, Contigs, and Singletons

A subset (1,344, 1,344, and 1,152) of cDNA clones from the three independent cDNA libraries (PA, FL, and HX) were sequenced at the 5' ends and 1,310, 1,322, and 1,139 individual ESTs were generated, respectively. The length of the EST sequences ranged from 214 to 998 bp with an average of 748 bp. The resulting 3,771 ESTs were assembled into 562 contigs and 1,383 singletons generating a total of 1,945 unique sequences. The rate of redundancy for all three cDNA libraries combined was 45%, while the redundancy for each library ranged from 31 to 44% (Table **1**). More than one contig may correspond to the same transcript, and some degree of redundancy may not be detected. Because all three libraries were constructed directionally and the sequences were determined only in one direction, from the 5' side, the undetected redundancy theoretically originated from the incomplete synthesis of the first strand. The relatively high redundancy rate in all three libraries suggests that EST sequencing efforts passed the cost efficient stage with these libraries to find novel sequences although further sequencing of the library would yield more of them.

Table 1. Summary of EST Collections from *Mycosphaerella fijiensis* Grown Under Three Different Culture Conditions

Library	Total Sequences	Unisequences	Redundancy
PA	1310	901	31%
FL	1322	797	40%
HX	1139	643	44%
Combined	3771	1945	45%

Nonetheless, about 95% (1841/1945) of the total unisequences were either singletons or contigs formed by less than five overlapping ESTs. Only 5% (104/1945) of unisequences were contigs generated by 5 or more overlapping ESTs, suggesting that the redundancy was mainly caused by a few highly expressed genes. As we determine the nucleotide sequences of additional cDNA clones, we will find more ESTs belong to the already identified sequences, especially these highly expressed genes. Subtraction by those small numbers of highly expressed genes would be helpful to increase the gene discovery rate [21, 31]. Alternatively, the suppression subtractive hybridization (SSH) method could be utilized to create the infection-specific library containing a mixture of both plant and fungal genes that are responsible for host and microbe interaction with an unbiased and equal representation [32].

Comparisons of the unisequences between and among the three libraries discovered only 99 contigs consisting of ESTs derived from all three libraries. There were 105 contigs overlapping between libraries PA and FL; 92 between libraries FL and HX and 81 between libraries PA and HX. A large number of unisequences was specific to each library, accounting for 68, 63 and 58% of the total unisequences generated from libraries PA, FL and HX, respectively (Fig. **2**). Although EST sequencing from each library was only a little over 1000, they were likely to represent expression profiles because all libraries were constructed without subtraction or normalization procedures. The high numbers of library-specific ESTs suggest different fungal reactions to the three different environmental conditions. However, the differences among the three libraries may have been partially due to other factors, such as the difference between the two isolates (MF-11 and MF-5), the use of 4-week-old mycelia for the PA library and 4- to 5-day-old mycelia for the FL and HX libraries, and two different culture media. Alternatively, the EST collection was not deep enough from each library to detect transcripts that are expressed at a low level. Consistent to this hypothesis, numbers of genes in each MIP category is similar with limited overlaps among three libraries (See below).

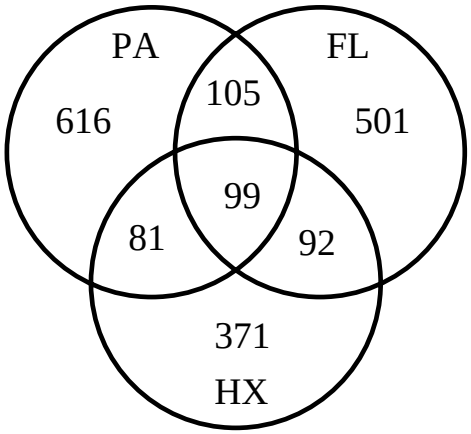

Fig. (2). Distribution of unisequences among three libraries. The numbers in the overlapping segments indicate the number of unisequences composed of ESTs derived from the designated libraries.

Highly Expressed Sequences

To examine the highly expressed sequences, we analyzed 45 contigs consisting of 10 or more ESTs (Table **2**). BLASTX analysis showed that fourteen contigs had no matching sequences in the NCBI database when a significant threshold of E-value <1e-05 was used as a cutoff. Among the 31 contigs with homologs, 10 were in the unknown functional category and seven were ribosomal proteins. The other 14 contigs showed similarities to genes with known functions. These known products include: glucose repressible gene protein, intracellular protease/amidase, opsin or opsin-like protein, superoxide dismutase, cation-transport related

Table 2. Top 40 Abundantly Expressed Genes Measured by EST Redundancy

MF_Contig	T	PA	FL	HX	Gene_Description	Organism	Accession_No	E-Value
MF_contig518	10	1	3	6	ribosomal protein	*Gibberella zeae*	XP_381414	1.20E-34
MF_contig519	10	2	3	5	glucose repressible gene protein	*Magnaporthe grisea*	mg[0624]	3.58E-22
MF_contig520	10	4	2	4	-	-	-	-
MF_contig521	10	4	4	2	ribosomal protein	*Aspergillus fumigatus*	XP_755254	6.16E-43
MF_contig522	10	9	1	0	hypothetical protein	*Aspergillus fumigatus*	XP_746739	1.88E-56
MF_contig523	10	1	4	5	ribosomal protein	*Chaetomium globosum*	EAQ85719	3.03E-112
MF_contig524	10	9	1	0	-	-	-	-
MF_contig525	11	6	2	3	cation transport-related protein	*Cryptococcus neoformans*	AAW42114	1.66E-21
MF_contig526	11	1	2	8	myo-inositol-phosphate synthase	*Aspergillus fumigatus*	XP_749237	6.45E-118
MF_contig527	11	6	4	1	-	-	-	-
MF_contig528	11	2	4	5	ribosomal protein	*Aspergillus fumigatus*	XP_755074	1.63E-61
MF_contig529	12	1	6	5	ribosomal protein	*Magnaporthe grisea*	AAW69354	2.61E-76
MF_contig530	12	9	1	2	opsin	*Leptosphaeria maculans*	AAG01180	1.44E-82
MF_contig531	12	2	7	3	unnamed protein product	*Aspergillus oryzae*	BAE59062	3.11E-47
MF_contig532	12	6	2	4	histone H4	*Aspergillus nidulans*	XP_658338	1.37E-34
MF_contig533	12	4	6	2	translation elongation factor	*Candida albicans*	XP_711870	0
MF_contig534	13	8	4	1	-	-	-	-
MF_contig535	13	0	9	4	unknown	*Magnaporthe grisea*	mg[0708]	4.98E-28
MF_contig536	13	2	6	5	ribosomal subunit	*Aspergillus fumigatus*	XP_754922	2.13E-131
MF_contig537	14	7	2	5	ribosomal protein	*Aspergillus fumigatus*	XP_751010	4.07E-51
MF_contig538	14	1	9	4	zinc metalloproteinase	*Aspergillus fumigatus*	XP_752634	1.38E-44
MF_contig539	15	12	0	3	hypothetical protein	*Aspergillus nidulans*	XP_660742	5.60E-32
MF_contig540	16	0	4	12	cation transport-related protein	*Cryptococcus neoformans*	AAW42114	1.52E-21
MF_contig541	16	0	7	9	-	-	-	-
MF_contig542	16	6	4	6	Superoxide dismutase	*Yersinia enterocolitica*	CAA65596	7.64E-70
MF_contig543	16	3	9	4	hypothetical protein	*Gibberella zeae*	XP_390148	1.42E-15
MF_contig544	18	4	8	6	hypothetical protein	*Gibberella zeae*	XP_388414	7.69E-28
MF_contig545	19	4	8	7	hypothetical protein	*Aspergillus nidulans*	XP_661903	4.50E-58
MF_contig546	20	3	10	7	glucose repressible gene protein	*Magnaporthe grisea*	mg[0624]	1.51E-23
MF_contig547	21	0	12	9	-	-	-	-
MF_contig548	21	13	5	3	-	-	-	-
MF_contig549	22	6	7	9	-	-	-	-
MF_contig550	23	13	8	2	-	-	-	-
MF_contig551	27	22	4	1	-	-	-	-
MF_contig552	28	1	15	12	intracellular protease/amidase	*Gibberella zeae*	XP_389155	4.31E-61
MF_contig553	28	1	9	18	unknown	*Cladosporium fulvum*	Cf8667490	3.98E-10
MF_contig554	31	26	2	3	hypothetical protein	*Neurospora crassa*	T49762	3.85E-24
MF_contig555	33	30	3	0	opsin-like protein	*Gibberella fujikuroi*	CAD97459	1.72E-83
MF_contig556	36	24	2	10	-	-	-	-
MF_contig557	36	8	12	16	-	-	-	-
MF_contig558	38	14	7	17	histone H3	*Chaetomium globosum*	AAT74576	2.71E-68
MF_contig559	59	3	34	22	-	-	-	-
MF_contig560	64	0	36	28	glucose repressible gene protein	*Magnaporthe grisea*	mg[0624]	1.23E-24
MF_contig561	74	10	20	44	-	-	-	-
MF_contig562	170	33	84	53	unnamed protein product	*Aspergillus oryzae*	BAE59652	6.47E-27

protein, zinc metalloproteinase, translation-elongation factor, myo-inositol-phosphate synthase, and histone H4 protein. Comparison of the three libraries indicates that the gene expression is more similar between FL and HX than between the other two pairs. For example, library PA has more expression of opsin or opsin-like proteins than FL and HX, but glucose repressible gene is highly expressed in FL and HX. This is likely due to the similar growth conditions in liquid media for FL and HX tissues. However, we do not exclude another possibility that the discrepancies reflect the differences between the two strains, Hilo-11 and Hilo-5.

Functional Annotation of ESTs

In order to determine the putative function of each EST in all three libraries, multi-dimensional sequence analysis was performed based on a series of BLAST searches [29]. BLASTX comparisons to the Genbank NR database enabled us to classify 2,259 ESTs (~60%) based on sequence similarities to genes with known and unknown functions in the databases. The remaining non-hit ESTs were further analyzed with TBLASTX against over 50,000 ESTs originated from fifteen pathogenic fungi and oomycete species, and three non-pathogenic fungi present in the COGEME phytopathogen EST database (http://cogeme.ex.ac.uk/) [33]. This approach led us to identify 445 more ESTs with similarities to functionally unclassified genes. Combining BLASTX and TBLASTX analyses indicated that approximately 72% of the ESTs from the three libraries had similar sequences in the public databases.

Based upon the MIPS classification system (http://mips.gsf.de/), we annotated each EST exhibiting the most significant BLASTX hit according to a functional category. The proportions of functional categories expressed as a percentage of the total collection for each of the three individual EST libraries are displayed in Table **3**. We were especially interested in the ESTs derived from the FL library as many fungal pathogenesis genes have previously been suggested to be expressed under nitrogen-limiting, or starvation conditions [34]. None of the specific functional categories, except slightly higher numbers of genes with unknown functions, showed noticeable differences in the FL library compared to the others. On the other hand, ESTs derived from HX library showed slightly higher numbers of genes expressed in three categories; energy, cellular transport, and biogenesis of cellular components. These are thought to be important components for nutrient uptake and nutrient metabolism, but not for infection-related functions (see next section). In general, the percentage of expressed genes in each library was similar, except for the category of response to the cellular environment. The PA library had a higher percentage of genes responding to the cellular environment.

ESTs Predicted to be Involved in Pathogenicity

Using similarity as an indicator of functions, we identified 50 genes putatively involved in pathogenicity (Table **4**). These genes encode proteins for infection structures, toxin synthesis and efflux, detoxification, nutrient and metabolite

Table 3. Functional Classification of ESTs for Each Library and the Total Collection Based on The Munich Information Center for Protein Sequences (MIPS) Scheme

Code	Function	PA	FL	HX	Total
01	Metabolism	59 (7.3%)	78 (9.6%)	56 (8.0%)	193 (8.3%)
02	Energy	16 (2.0%)	21 (2.6%)	32 (4.6%)	69 (3.0%)
10	Cell Cycle and DNA Processing	2 (0.2%)	3 (0.4%)	1 (0.1%)	6 (0.3%)
11	Transcription	12 (1.5%)	1 (0.1%)	8 (1.1%)	21 (0.9%)
12	Protein Synthesis	43 (5.3%)	68 (8.3%)	54 (7.7%)	165 (7.1%)
14	Protein Fate (folding, modification, destination)	11 (1.4%)	5 (0.6%)	12 (1.7%)	29 (1.2%)
16	Protein with Binding Function	13 (1.6%)	12 (1.5%)	4 (0.6%)	29 (1.2%)
18	Protein Activity Regulation	3 (0.4%)	7 (0.9%)	9 (1.3%)	19 (0.8%)
20	Cellular Transport, Transport Facilitation and Transport Routes	25 (3.1%)	17 (2.1%)	28 (4.0%)	70 (3.0%)
30	Cellular Communication/Signal Transduction Mechanism	15 (1.9%)	8 (1.0%)	10 (1.4%)	74 (3.2%)
32	Cell Rescue, Defense and Virulence	25 (3.1%)	26 (3.2%)	19 (2.7%)	70 (3.0%)
34	Interaction with the Cellular Environment	38 (4.7%)	5 (0.6%)	7 (1.0%)	8 (0.3%)
36	Interaction with the Environment (Systemic)	0 (0.0%)	0 (0.0%)	1 (0.1%)	1 (0.0%)
40	Cell Fate	0 (0.0%)	0 (0.0%)	3 (0.4%)	3 (0.1%)
41	Development (Systemic)	1 (0.1%)	1 (0.1%)	0 (0.0%)	2 (0.1%)
42	Biogenesis of Cellular Components	31 (3.8%)	20 (2.5%)	46 (6.5%)	97 (4.2%)
98	Classification Not Yet Clear-cut	7 (0.9%)	10 (1.2%)	2 (0.3%)	19 (0.8%)
99	Unknown protein	507 (62.7%)	534 (65.4%)	410 (58.3%)	1451 (62.4%)
	Total	809 (100%)	816 (100%)	703 (100%)	2327 (100%)

Table 4. *M. fijiensis* Genes that Share Significant Sequence Similarities to Verified Pathogenicity Genes in Other Fungi

EST	Library	Gene_Description	Accession_No	Evalue
Infection structure				
MF11-04_M13F-B03	PA(1)[a]	conidiophore development protein HymA [*Aspergillus nidulans*]	XP_660699	5.75E-23
MF56-01_M13F-F02	HX(1)	related to hard surface induced protein 3 [*Neurospora crassa*]	CAE85497	4.85E-23
MF11-14_M13F-A11	PA(1)	hydrophobic surface binding protein B [*Aspergillus oryzae*]	BAE53530	1.24E-09
MF11-03_M13F-B09	PA(1)	hydrophobin [*Cladosporium fulvum*]	CAA67187	4.22E-08
Fungal toxin synthesis and efflux				
MF05-14_M13F-B01	FL(1)	branched-chain amino acid aminotransferase [*Aspergillus fumigatus*]	XP_752208	1.08E-81
MF11-12_M13F-G01	PA(1)	trichothecene efflux pump [*Gibberella zeae*]	AAM49050	1.24E-08
MF56-07_M13F-F10	HX(1)	fatty acid elongase (FEN1) [*Neurospora crassa*]	CAD70918	4.17E-69
Detoxification, drug resistance and metabolite transport				
MF05-03_M13F-D04	FL(1)	avenacinase [*Gaeumannomyces graminis*]	AAN39015	5.06E-82
MF05-01_M13F-H11	FL(2)	MSF multidrug transporter [*Aspergillus fumigatus*]	XP_747772	9.94E-42
MF05-04_M13F-F10	FL(1)	major facilitator superfamily [*Leptosphaeria maculans*]	AAO49453	2.48E-105
MF05-03_M13F-A04	FL(1)	neutral amino acid permease [*Aspergillus fumigatus*]	XP_747312	2.45E-36
MF56-08_M13F-A03	HX(1)	putative MFS transporter [*Gibberella moniliformis*]	CAI60779	2.19E-126
MF05-10_M13F-C10	PA(1), FL(1)	related to brefeldin A resistance protein [*Neurospora crassa*]	CAF06031	1.44E-75
MF05-09_M13F-A01	FL(1)	related to neomycin resistance protein NEO1 [*Neurospora crassa*]	CAC18258	1.94E-62
MF56-01_M13F-A01	HX(1)	sucrose transporter, putative [*Aspergillus fumigatus*]	EAL91974	1.06E-54
MF11-07_M13F-C07	PA(1)	transmembrane transporter [*Aspergillus fumigatus*]	XP_748898	2.55E-56
MF05-01_M13F-C05	FL(1)	mitochondrial carrier protein [*Aspergillus fumigatus*]	XP_751597	7.65E-38
Signal cascade components				
MF11-14_M13F-F04	PA(1)	cAMP-dependent protein kinase regulatory subunit [*Magnaporthe grisea*]	O14448	1.42E-77
MF05-06_M13F-H05	FL(FL), PA(1)	calmodulin [*Magnaporthe grisea*]	AAC96324	2.48E-78
MF05-10_M13F-A06	FL(1)	pathogenicity MAP kinase 1 [*Gibberella zeae*]	AAL73403	1.49E-62
MF11-02_M13F-E09	PA(3)	GTP-binding protein Drab11 [*Neurospora crassa*]	CAD21237	1.05E-87
MF56-04_M13F-C04	FL(1), HX(1)	serine/threonine protein kinase [*Thermomyces lanuginosus*]	AAT84066	2.52E-102
MF05-02_M13F-C12	PA(1), FL(1)	signal recognition particle protein [*Aspergillus fumigatus*]	XP_751808	1.38E-56
MF11-10_M13F-B09	PA(1)	related to serine/threonine-protein kinase [*Neurospora crassa*]	CAD79666	7.35E-56
MF11-10_M13F-E10	PA(3)	protein serine/threonine kinase Ran1 [*Aspergillus fumigatus*]	XP_754566	8.99E-56
Protection agaist oxidative stress				
MF11-13_M13F-G12	PA(1)	catalase [*Gibberella zeae*]	AAO34682	1.40E-20
MF11-13_M13F-H11	PA(1), FL(1)	Superoxide dismutase [*Haemophilus somnus*]	ZP_00132526	8.44E-26
MF11-08_M13F-D04	PA(1)	Cu,Zn superoxide dismutase SOD1 [*Aspergillus fumigatus*]	XP_753715	7.09E-68
MF56-11_M13F-D12	HX(1)	cytochrome c oxidase polypeptide V [*Aspergillus nidulans*]	XP_659920	2.38E-48
MF05-02_M13F-D09	FL(4)	cytochrome c1 precursor [*Neurospora crassa*]	CAA28860	2.05E-93
MF05-03_M13F-G10	FL(1)	cytochrome P450 monooxygenase [*Aspergillus fumigatus*]	XP_746946	9.88E-35
MF11-12_M13F-C12	PA(2)	monooxigenase [*Aspergillus fumigatus*]	XP_751255	3.18E-80
MF56-12_M13F-E06	HX(2)	cytochrome-c oxidase chain VI precursor [*Neurospora crassa*]	CAD70919	2.75E-42
MF05-08_M13F-H09	PA(3), FL(3), HX(6)	sodA [*Yersinia enterocolitica*]	CAA65596	2.30E-70
MF11-05_M13F-A03	PA(1)	Superoxide dismutase (Mn) [*Escherichia coli*]	AAN83287	1.64E-61
MF56-09_M13F-E03	HX(2)	Tyrosinase [*Monophenol monooxygenase*]	Q92396	3.12E-22
MF05-07_M13F-F07	FL(1)	tyrosinase precursor [*Neurospora crassa*]	CAE81941	2.26E-24
MF05-08_M13F-A02	FL(3)	epoxidase subunit A [*Penicillium decumbens*]	BAA75924	1.23E-71

(Table 4) contd…..

EST	Library	Gene_Description	Accession_No	Evalue
hydrolytic enzyme				
MF11-06_M13F-D03	PA(1)	1,3-beta glucanase [*Blumeria graminis*]	AAL26905	5.66E-65
MF11-07_M13F-D08	PA(1), HX(1)	aspartic proteinase precursor [*Botryotinia fuckeliana*]	AAR13364	5.24E-105
MF56-04_M13F-C06	HX(3)	zinc metalloproteinase, putative [*Aspergillus fumigatus*]	EAL90596	5.65E-43
Avirulence/virulence factors				
MF05-14_M13F-B11	PA(2), FL(1), HX(2)	Avr4 [*Cladosporium fulvum*]	CAA55403	1.80E-17
MF05-07_M13F-C09	FL(2), PA(2)	ecp2 [*Cladosporium fulvum*]	CAA78401	7.22E-49
MF05-11_M13F-B02	PA(2), FL(2), HX(2)	Protein SnodProt1 precursor [*Phaeosphaeria nodorum*]	O74238	2.81E-37
MF56-05_M13F-G11	PA(1), HX(2)	pSI-7 protein [*Cladosporium fulvum*]	CAA74888	1.06E-92
MF56-11_M13F-D06	PA(4), FL(2), HX(1)	virulence related protein Cap20 [*Aspergillus fumigatus*]	EAL88346	2.74E-29
Miscellaneous				
MF05-04_M13F-G02	FL(1)	isocitrate lyase [*Leptosphaeria maculans*]	AAM89498	3.25E-70
MF05-07_M13F-B12	PA(2), FL(1), HX(2)	cytosolic cyclophilin 1 [*Botryotinia fuckeliana*]	AAQ16573	1.37E-60
MF11-11_M13F-G08	PA(1), FL(1)	septin [*Aspergillus fumigatus*]	XP_748972	1.95E-113
MF05-01_M13F-F08	FL(2)	histidine kinase [*Cochliobolus heterostrophus*]	BAD07400	3.89E-31

[a]The number in the parenthesis indicates the count of ESTs with the homolog identified in the library.

transport, response to oxidative stress, hydrolytic enzymes, and signal cascade components. Three of four genes related to infection structures were identified from the PA library, while a putative homolog of hard surface induced protein was identified from the HX library. Genes involved in detoxification were discovered mainly from the FL library. Genes involved in oxidative stress responses were similarly expressed under all three conditions. Several ESTs showed sequence similarities to the genes for avirulence effectors, such as the Avr4, ecp2 and pSI-7 proteins of *Cladosporium fulvum* and the Protein SnodProt1 precursor of *Phaeosphaeria nodorum*.

Comparison of the EST Set to Sequenced Genomes of 18 Fungi

We downloaded the assembled sequences and the predicted transcripts (or proteins) of 18 fungal genomes that have been sequenced through the Fungal Genome Initiative at the Broad Institute and Eukaryotic Genomics at the DOE Joint Genome Institute (JGI). The 1945 *M. fijiensis* unisequences were qerried against these genome sequence databases using the BLASTX program. The results are summarized in Table **5**. In general, the relative overall level of homology between the unisequence set and a genome reflected the phylogenic relationship of *M. fijiensis* to the species analyzed. The highest numbers of unisequences matched to the transcripts derived from *M. fijiensis*, followed by a close relative species *M. graminicola* and other ascomycetous fungi (Table **5**). They include both plant-pathogenic fungi (*F. graminearum*, *M. grisea*, *M. fijiensis*, *M. graminicola, Nectria haematococca,* and *Stagonospora nodorum*) and saphrophytic fungi (*Aspergillus nidulans* and *N. crassa*). Less unisequences matched to the predicted transcripts derived from the fungal species belonging to the phyla Basidiomycota and Zygomycota. The *M. fijiensis* unisequence set had the least homology to the genome of *Batrachochytrium dendrobatidis, a* chytridiomycete fungus. In this case only 36% of the unisequences had a homolog with an E-value <1e-05.

To further evaluate the usefulness of the ESTs for annotation of the *M. fijiensis* genome, we performed a global BLAST analysis using the 3,771 ESTs against the assembled genome sequence (version 1.0) released on August 20, 2007. The draft includes a total of 10,327 gene models predicted and functionally annotated with the JGI annotation pipeline. Of the 3,771 ESTs from the three *M. fijiensis* cDNA libraries, 2,110 (55.6%) had an almost 100% sequence identity (perfect E-value (0)) to the predicted transcripts, suggesting an accurate prediction of genes as well as gene structure. Further, 395 ESTs (10.5%) exhibited significant (0 < E-value 1e < 05) similarity to the predicted transcripts, suggesting an accurate gene prediction, but an inaccurate prediction of gene structure. However, some of the EST sequences might have intrinsic problems, such as unclean sequences, chimeric sequences, or unknown origins. It is noteworthy that 1,327 ESTs (34%) had no similarity (no hits) to the predicted transcripts.

These 1,327 ESTs were further compared with the 382 scaffolds representing the draft whole genome sequence of *M. fijiensis* in order to identity their origin. BLASTN analysis identified an additional 337 (9%) of 1,327 ESTs that showed ~100% sequence identify to the genomic sequences in the 382 scaffolds. Another 519 (13%) ESTs had a significant (0 < E-value < 1e-5) similarity to the genome sequence, inferring the presence of an intron, or introns, within the genes or DNA polymorphism between the different isolates used. In conclusion, the additional 856 (22%) ESTs showed a sequence identity to the genomic sequences, but not to the predicted transcript sequences. This suggests that the gene prediction algorithms used for the *M. fijiensis* draft version did not recognize these regions as genes. Finally, 473 (12%) of the ESTs did not show significant similarity to any ge-

Table 5. Comparison of *M. fijiensis* Unisequeces to the Fully Sequenced Genomes of Fungi

Phylum	Fungal Species	No. of Unisequences with Hits[a]			% of Unisequences		
		<1e-5	<1e-40	<1e-100	<1e-5	<1e-40	<1e-100
Ascomycota	*Aspergillus nidulans*	1187	614	75	61.0	31.6	3.9
	Aspergillus niger	851	337	17	43.8	17.3	0.9
	Chaetomium globosum	1130	519	64	58.1	26.7	3.3
	Fusarium graminearum	1207	588	78	62.1	30.2	4.0
	Magnaporthe grisea	1184	566	72	60.9	29.1	3.7
	Mycosphaerella fijiensis	1417	1180	569	72.9	60.7	29.3
	Mycosphaerella graminicola	1358	903	213	69.8	46.4	11.0
	Neurospora crassa	1170	559	79	60.2	28.7	4.1
	Nectria haematococca	1204	620	91	61.9	31.9	4.7
	Stagonospora nodorum	1275	674	88	65.6	34.7	4.5
Basidiomycota	*Coprinus cinereus*	858	301	20	44.1	15.5	1.0
	Cryptococcus neoformans	842	300	24	43.3	15.4	1.2
	Laccaria bicolor	872	308	20	44.8	15.8	1.0
	Puccinia graminis f. sp. tritici	752	251	16	38.7	12.9	0.8
	Ustilago maydis	836	297	24	43.0	15.3	1.2
Zygomycota	*Phycomyces blakesleeanus*	853	295	22	43.9	15.2	1.1
	Rhizopus oryzae	812	283	23	41.7	14.6	1.2
Chytridiomycota	*Batrachochytrium dendrobatidis*	690	244	19	35.5	12.5	1.0

[a]Indicates the counts of unisequences with homology to the predicted protein sequences based on a top BLASTX hit at an E-value < 1e-5, 1e-40, or 1e-100, respectively.

nome sequences. These EST sequences were clean, with an average length of 852 bp, and thus possibly derived from regions of the genome yet to be sequenced. These EST sequences could add value to the gene discovery process, complementing the whole genome sequencing effort.

DISCUSSION

The purpose of this study was to discover the genes expressed by *M. fijiensis* under three different environmental conditions, with special interest in candidate genes involved in pathogenicity. We chose to use pure fungal cultures in order to increase fungal gene discovery rates without complication from host ESTs [34, 35]. We determined 3,771 EST sequences from three cDNA libraries and assembled them into 1,945 putative genes. The estimated genome size of *M. fijiensis* is about 35 Mb with a chromosome number of 17 or 18 (Andy James, personal communication). This genome size is within the range of other ascomycete fungi. It would be expected to contain around 8,000–9,000 genes [36], close to the 10,327 genes by *ab intio* gene prediction of the initial genome assembly (http://genome.jgi-psf.org/Mycfi1/Mycfi1.home.html). Based on this estimation, the 1,945 unigenes identified represent *ca*. 20% of the genome in *M. fijiensis*. Overall, the sequences identified in this study present the first large-scale EST analysis of this fungus. This work provides an awareness of the biological processes that occur during growth on a solid medium, in Fries liquid medium, and in the presence of banana plant extracts and these processes may be related to the interaction between *M. fijiensis* and its banana host.

BLASTN comparisons of the 3,771 ESTs against the predicted transcripts identified 2,110 (55%) with perfect, 395 (10%) with partial, and 1,327 (35%) with no matching partners. Among the 1,327 no-matching ESTs, 856 (22%) had a significant sequence identity to the genomic sequences. Of these, 337 (9%) had an almost 100% identical match to the genomic sequence. This result suggests that the gene prediction was fairly accurate, though it missed about 20% of the ESTs, and that the prediction of gene structure was even less accurate than the prediction of genes. When a large number of gene sequences are not available, *ab initio* gene prediction using diverse programs are employed to predict splicing probabilities of donor and acceptor sites, the coding probability of each single exon, and the probabilities of various gene models. The prediction of gene and gene structure can be significantly improved by adding more gene sequences, including ESTs, into the program training set (reviewed in [37]). The 3,771 ESTs reported in this study will be useful for improved prediction of gene structure as well as discovery of new genes in the draft genome.

Among the 45 contigs formed by more than 10 ESTs, 14 have unknown functions, indicating that these genes may be unique to *M. fijiensis*. Seven of the 45 contigs are ribosomal synthesis genes. High-level expression of ribosomal genes has been reported in several other phytopathogenic fungi, including *Heterobasidion annosum* [21], *Ustilago maydis* [25], and *Gibberella zeae* [20]. This may reflect the active growth of the fungus under the investigated conditions. In

contrast, ribosomal proteins were expressed at a low levels in a closely related species, *M. graminicola* [23]. These contradictory results might have occurred because of different culture conditions for fungal tissues that have been used as library sources. Among the highly expressed genes, glucose repressible genes (contig 519, 546, and 560) were highly expressed in the HX library, while Opsin and opsin-like protein were highly expressed in the PA library. This is probably due to the continuous lighting conditions used during incubation as opposed to the dark mycelial growth conditions used for the other two libraries. Opsin is a light responsive protein and its expression can be induced by light [38]. Cation transport-related protein was also abundant in both the FL and HX libraries derived from liquid cultures. This may reflect that the uptake of cations from Fries medium is active in fungal mycelia.

Our interpretation of ESTs is based on the sequence similarity to other genes with known functions listed in public databases. Most putative pathogenicity genes occurred in the FL library compared to the other libraries in this study. Not surprisingly, several hydrolytic enzyme genes were identified in the FL library. There is substantial evidence supporting the importance of enzymatic digestion in host cuticle penetration and subsequent colonization. Several cell wall degrading enzyme (CWDE) coding genes and their control element genes have been identified as virulence factors and their null mutation has significantly reduced the infectivity of several diverse phytopathogens [39-43].

Typically, both the pathogenic fungus and its compatible host plant secrete large numbers of hydrolytic enzymes during fungus-host plant interactions. However, it is difficult to establish the role of individual hydrolytic enzymes in pathogenesis due to functional redundancy among numerous enzymes and multiple isozymes [42, 44, 45]. This enzymatic warfare between the plant and fungus is likely to produce signaling molecules, such as oligosaccharides, from the cell walls of the plant. Further, the fungus may activate additional defense and pathogenesis responses from the host. One of the responses to the enzyme warfare is an oxidative burst. Several genes that respond to or create oxidative stress were identified in this study. These genes include catalases, peroxidases, superoxide dismutases, monooxygenase, and cytochrome elements (Table **4**). There is no supporting evidence yet, however, that any of these genes are directly associated with pathogenicity. In addition, only three hydrolytic enzymes were identified in this study, in contrast to general expectation, especially from the HX library that was prepared with mycelia grown in the presence of host extracts. The extracts might be insufficient elicitors compared to host cell wall components, which may be required for the full induction of putative CWDE gene expression demonstrated in other pathogenic fungi [46-48]. In general, it may be desirable to sequence further ESTs from libraries representing compatible host plant infection stages in order to better understand pathogen-host interactions.

We identified three genes for fungal toxin synthesis and ten genes for detoxification, drug resistance and transport protein (Table **4**). Phytopathogenic fungi have to cope with many natural toxic compounds such as those produced by host plants during pathogenesis. Membrane proteins are known to provide protection against a wide range of natural toxic compounds by pumping them out of the cell [49, 50]. An example of a transmembrane protein is MSF transporter whose activity is driven by the proton-motive force through membranes [51]. They are involved in protection against exogenous toxic compounds in case of *Candida albicans* [52] and *S. cerevisiae* [53]. One potential mechanism of infection is killing host cells with fungal phytotoxic secondary metabolites and toxic proteins. Transporters are known to mediate the secretion of endogenously produced toxins, such as aflatoxin, cercosporin, HC-toxin [54, 55]. For example, *ToxA* encodes an HC-toxin efflux pump which contributes to self-protection against HC-toxin in *Cochliobolus carbonum*. Like HTS1 encoding the central enzyme in HC-toxin biosynthesis, *ToxA* occurs only in isolates of the fungus that make HC-toxin. It is logical to consider that fungal transporters are important virulence factors, in parallel with secondary metabolites. Regardless, the function of transporters is to pump out the secondary metabolites originating from the pathogen itself or from the host plant. The necrotrophic fungus, *Alternaria alternata,* is known for its production of species-specific toxins encoded by nonribosomal polypeptide synthetases (NRPS) and polyketide synthases (PKS) [56]. Neither NRPS nor PKS have been identified in this study, in contrast to several transporters. This is probably due to their low level of expression under our trial conditions and the small numbers of EST collections from each library.

One clone in the FL library showed similarity to a signal transduction gene encoding a 'PMK1 homolog' from many pathogenic fungi. This MAP kinase and its homologs are essential for pathogenicity in many plant-pathogenic fungi [57, 58]. MAP kinase is required for appressorium formation in fungal pathogens, such as *Colletotrichum lagenarium*, *M. grisea*, *Cochliobolus heterostrophus*, and *Pyrenophora teres* [59-62]. *M. grisea* and *C. lagenarium* produce especially large and heavily melanized appressoria that generate strong turgor pressure [63]. In both fungi, MAP kinase (Fus3/Kss1 homolog) mutants are defective in appressorium formation and are nonpathogenic as a result of their inability to penetrate the plant epidermis and colonize host plant tissues [61, 62]. MAP kinase was also found to be required for full pathogenicity in the necrotrophic parasite *Botrytis cinerea* and in closely related Dothideomycete fungi, including *Stagonospora nodorum*, *P. teres*, *C. heterostrophus*, and *Alternaria brassicicola* [60-62, 64-67]. It is likely that the PMK1 homolog in *M. fijiensis* identified in this study also mediates plant infection either in the penetration and/or colonization stages. There are three additional kinases and one kinase regulatory subunit which may also be pathogenicity factors. These signal transduction pathways may be associated with the induction of downstream hydrolytic enzyme genes during pathogenicity as suggested previously [64].

Finally, we have identified ESTs encoding for proteins similar to several Avr factors, including AVR4 and ECP2 from *Cladosporium fulvum*. Both proteins are highly induced during host infection and were isolated from apoplastic washing fluids of tomato [68, 69]. However, our EST analysis indicated that the AVR4 homolog was found in all three libraries (PA, FL, and HX) and the ECP 2 homolog existed in two of the libraries (FL and HX). It seems that these homologs are expressed *in vitro* although it is not known if their expression level is higher *in vivo*. It remains to be investigated whether these genes in *M. fijiensis* have similar

functions to the *Avr* genes in *C. fulvum*. Targeted gene knockout study is the most straight-forward method to test the functions of these genes. Currently, targeted gene knockout methods for either targeted gene disruption or targeted gene deletion are unavailable for *M. fijiensis*. Successful transformation of exogenous DNA fragments, especially GFP expression vectors, has been reported [70]. Further optimization of the transformation method will lead to successful targeted gene knockout methods. Similarly, pathogenicity test methods [71] may be further optimized and established for *M. fijiensis* in order to discover pathogenicity genes and advance our understanding of their modes of action.

CONCLUSION

The putative identity of the mRNA transcripts discovered in three EST libraries led to several candidate genes potentially important in determining the outcome of a compatible fungal-plant interaction. Further functional analysis of the genes identified in this study by targeted gene replacement or disruption studies, as well as gene expression profiling experiments, will provide more information on the interaction between *M. fijiensis* and its economically important host, banana. Fungal ESTs identified in this study will also facilitate the annotation of the *M. fijiensis* genome sequence. A genome sequencing project was funded by the Department of Energy (DOE) in 2005. It was initiated at the DOE-Joint Genome Institute to generate a draft whole genome sequence. The ESTs identified in this study would aid annotation, and most specifically the prediction, of genes in the *Mycosphaerella* genome. Comparative analysis of ESTs with the closely related fungus *M. graminicola* and other distantly related fungi, may facilitate the identification of candidate pathogenicity genes and improve our understanding of the evolution of fungal plant pathogens. This could lead to development of more effective systems for banana disease control.

ACKNOWLEDGEMENTS

Financial support for this study was provided by University of Hawaii at Manoa to YC and SZ and North Dakota State University to SZ. We thank Scot C. Nelson for collecting *M. fijiensis*-infected banana leaves for fungal isolation and Baojun Yang for assistance in construction of cDNA libraries, and Fred Brooks for helpful discussion. We would also like to thank Zheng Jin Tu at the University of Minnesota and Anupma Sharma at UHM for their assistance with various computational aspects of this study.

ABBREVIATIONS

EST = Expressed Sequence Tags

PA = cDNA library representing MF11-Hilo grown on PDA for 4 weeks

FL = cDNA library representing MF5-Hilo grown in Fries liquid medium for 4 days

HX = cDNA library representing MF5-Hilo grown in Fries liquid medium plus banana leaf extract

REFERENCES

[1] Ploetz RC. Black Sigatoka of Banana. The Plant Health Instructor [serial on the internet] 2001; [cited on 2008 June 1] Available from: http://www.apsnet.org/education/feature/banana/

[2] Mobambu KN, Gauhl F, Vuylsteke D, Ortiz R, Pasberg-Gauhl C, WSwennen R. Yield loss in plantain from blacks Sigatoka leaf spot and field performance of resistant hybrids. Field Crops Res 1993; 35: 35-42.

[3] Gauhl F. Epidemiology and ecology of black Sigatoka (*Mycosphaerella fijiensis* Morlet) on plantain and banana (*Musa* spp.) in Costa Rica, Central America. Ph.D. dissertation, Universität Göttingen, 1994.

[4] Ploetz RC, Mourichon X. First report of black sigatoka in Florida. Plant Dis 1999; 83: 300.

[5] Stansbury C, KMcKirdy S, Power G. Black Sigatoka disease: new technologies to strengthen eradication strategies in Australia. Aust Plant Pathol 2006; 35: 181-93.

[6] Chin KM, Wirz M, Laird D. Sensitivity of *Mycosphaerella fijiensis* from banana to trifloxystrobin. Plant Dis 2001; 85: 1264-70.

[7] Romero R, Sutton T. Characterization of benomyl resistance in *Mycosphaerella fijiensis*, cause of black Sigatoka of banana, in Costa Rica. Plant Dis 1998; 82: 931-4.

[8] Romero RA, Sutton TB. Sensitivity of *Mycosphaerella fijiensis*, causal agent of black Sigatoka of banana, to propiconazole. Phytopathology 1997; 87: 96-100.

[9] Castillo LE, Ruepert C, Solis E. Pesticide residues in the aquatic environment of banana plantation areas in the north Atlantic zone of Costa Rica. Environ Toxicol Chem 2000; 9: 1942-50.

[10] Chaves A, Shea D, Cope WG. Environmental fate of chlorothalonil in a Costa Rican banana plantation. Chemosphere 2007; 69: 1166-74.

[11] Daly GL, Lei YD, Teixeira C, Muir DCG, Castillo LE, Wania F. Accumulation of current-use pesticides in neotropical montane forests. Environ Sci Technol 2007; 41: 1118-23.

[12] Vindas R, Ortiz F, Ramírez V, Cuenca P. Genotoxicity of three pesticides used in Costa Rican banana plantations. Rev Biol Trop 2004; 52: 601-9.

[13] Janick J. Fruit breeding in the 21st century. In: Galan Sauco V, Ed. Proceedings of the First International Symposium on Banana in the Subtropics; 1997: Tenerife, Spain; 1998; pp. 39-46.

[14] Simmonds NW. Bananas. London: Longmans; 1966.

[15] Simmonds NW, Shepherd K. Taxanomy and origins of cultivated bananas. J Linn Soc Bot 1955; 55: 302-12.

[16] Persley GJ, George P, Eds. Banana Improvement: Research challenges and opportunities. Washington D.C: The World Bank; 1996.

[17] Persley GJ, George P, Eds. Banana, Breeding, and Biotechnology: Commodity advances through banana improvement project research, 1994-1998. Washington D.C: The World Bank; 1999.

[18] Bakry F, Carreel F, Caruana ML, *et al*. Banana. In: Charrier A, Jacquot M, Hamon S, Nicolas D, Eds. Tropical plant breeding. Centre de cooperation internationale en researche agronomique pour le developpement (CIRAD): France; 2001; 1-29.

[19] Fullerton RA, Olsen TL. Pathogenic variability in *Mycosphaerella fijiensis* MORELET, cause of black Sigatoka in banana and plantain. NZ J Crop Hortic Sci 1995; 23: 39-48.

[20] Trail F, Xu JR, San Miguel P, Halgren RG, Kistler HC. Analysis of expressed sequence tags from *Gibberella zeae* (anamorph *Fusarium graminearum*). Fungal Genet Biol 2003; 38: 187-97.

[21] Abu SM, Li G, Asiegbu FO. Identification of *Heterobasidion annosum* (S-type) genes expressed during initial stages of conidiospore germination and under varying culture conditions. FEMS Microbiol Lett 2004; 233: 205-13.

[22] Ebbole DJ, Jin Y, Thon M, *et al*. Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: analysis of expressed sequence tags. Mol Plant Microbe Interact 2004; 17: 1337-47.

[23] Keon J, Antoniw J, Rudd J, Skinner W, Hargreaves J, Hammond-Kosack K. Analysis of expressed sequence tags from the wheat leaf blotch pathogen *Mycosphaerella graminicola* (anamorph S*eptoria tritici*). Fungal Genet Biol 2005; 42: 376-89.

[24] Zhu H, Nowrousian M, Kupfer D, *et al*. Analysis of expressed sequence tags from two starvation, time-of-day-specific libraries of *Neurospora crassa* reveals novel clock-controlled genes. Genetics 2001; 157: 1057-65.

[25] Nugent KG, Choffe K, Saville BJ. Gene expression during *Ustilago maydis* diploid filamentous growth: EST library creation and analyses. Fungal Genet Biol 2004; 41: 349-60.

[26] Stover RH. Distribution and cultural characteristics of the pathogens causing banana leaf spot. Trop Agric (Trinidad) 1976; 53: 111-4.

[27] Johanson A, Jeger MJ. Use of PCR for detection of *Mycosphaerella fijiensis* and *M. musicola*, the causal agents of Sigatoka leaf spots on banana and plantain. Mycol Res 1993; 97: 670-4.

[28] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998; 8: 175-85.

[29] Altschul SF, Madden TL, Schaffer AA, *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25: 3389-402.

[30] Mewes HW, Frishman D, Guldener U, *et al*. MIPS: a database for genomes and protein sequences. Nucleic Acids Res 2002; 30: 31-4.

[31] Cramer RA, Lawrence CB. Identification of *Alternaria brassicicola* genes expressed in planta during pathogenesis of *Arabidopsis thaliana*. Fungal Genet Biol 2004; 41: 115-28.

[32] Diatchenko L, Lau YF, Campbell AP, *et al*. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA 1996; 93: 6025-30.

[33] Soanes DM, Skinner W, Keon J, Hargreaves J, Talbot NJ. Genomics of phytopathogenic fungi and the development of bioinformatic resources. Mol Plant Microbe Interact 2002; 15: 421-7.

[34] Talbot N, McCafferty H, Ma M, Moore K, Hamer J. Nitrogen starvation of the rice blast fungus *Magnaporthe grisea* may act as an environmental cue for disease symptom development. Physiol Mol Plant Pathol 1997; 50: 179-95.

[35] Cramer RA, La Rota CM, Cho Y, *et al*. Bioinformatic analysis of expressed sequence tags derived from a compatible *Alternaria brassicicola-Brassica oleracea* interaction. Mol Plant Pathol 2006; 7: 113-24.

[36] Kupfer DM, Reece CA, Clifton SW, Roe BA, Prade RA. Multicellular ascomycetous fungal genomes contain more than 8000 genes. Fungal Genet Biol 1997; 21: 364-72.

[37] Cho Y, Walbot V. Computational methods for gene annotation: the *Arabidopsis* genome. Curr Opin Biotechnol 2001; 12: 126-30.

[38] Purschwitz J, Muller S, Kastner C, Fischer R. Seeing the rainbow: light sensing in fungi. Curr Opin Microbiol 2006; 9: 566-71.

[39] Ospina-Giraldo MD, Mullins E, Kang S. Loss of function of the *Fusarium oxysporum SNF1* gene reduces virulence on cabbage and Arabidopsis. Curr Genet 2003; 44: 49-57.

[40] Shieh MT, Brown RL, Whitehead MP, *et al*. Molecular genetic evidence for the involvement of a specific polygalacturonase, *P2c*, in the invasion and spread of *Aspergillus flavus* in cotton bolls. Appl Environ Microbiol 1997; 63: 3548-52.

[41] ten Have A, Mulder W, Visser J, van Kan JA. The endopolygalacturonase gene *Bcpg1* is required for full virulence of *Botrytis cinerea*. Mol Plant Microbe Interact 1998; 11: 1009-16.

[42] Tonukari NJ, Scott-Craig JS, Walton JD. The *Cochliobolus carbonum SNF1* gene is required for cell wall-degrading enzyme expression and virulence on maize. Plant Cell 2000; 12: 237-48.

[43] Voigt CA, Schafer W, Salomon S. A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. Plant J 2005; 42: 364-75.

[44] Mendgen K, Hahn M, Deising H. Morphogenesis and mechanisms of penetration by plant pathogenic fungi. Annu Rev Phytopathol 1996; 34: 367-86.

[45] Walton JD. Deconstructing the cell wall. Plant Physiol 1994; 104: 1113-8.

[46] Mendoza-Mendoza A, Pozo MJ, Grzegorski D, *et al*. Enhanced biocontrol activity of *Trichoderma* through inactivation of a mitogen-activated protein kinase. Proc Natl Acad Sci USA 2003; 100: 15965-70.

[47] Olmedo-Monfil V, Mendoza-Mendoza A, Gomez I, Cortes C, Herrera-Estrella A. Multiple environmental signals determine the transcriptional activation of the mycoparasitism related gene *prb1* in *Trichoderma atroviride*. Mol Genet Genomics 2002; 267: 703-12.

[48] Rogers LM, Kim YK, Guo W, Gonzalez-Candelas L, Li D, Kolattukudy PE. Requirement for either a host- or pectin-induced pectate lyase for infection of *Pisum sativum* by *Nectria hematococca*. Proc Natl Acad Sci USA 2000; 97: 9813-8.

[49] Andrade AC, Van Nistelrooy JG, Peery RB, Skatrud PL, De Waard MA. The role of ABC transporters from *Aspergillus nidulans* in protection against cytotoxic agents and in antibiotic production. Mol Gen Genet 2000; 263: 966-77.

[50] Del Sorbo G, Schoonbeek H, De Waard MA. Fungal transporters involved in efflux of natural toxic compounds and fungicides. Fungal Genet Biol 2000; 30: 1-15.

[51] Pao SS, Paulsen IT, Saier MH, Jr. Major facilitator superfamily. Microbiol Mol Biol Rev 1998; 62: 1-34.

[52] Calabrese D, Bille J, Sanglard D. A novel multidrug efflux transporter gene of the major facilitator superfamily from *Candida albicans* (*FLU1*) conferring resistance to fluconazole. Microbiology 2000; 146: 2743-54.

[53] Alarco AM, Balan I, Talibi D, Mainville N, Raymond M. AP1-mediated multidrug resistance in *Saccharomyces cerevisiae* requires *FLR1* encoding a transporter of the major facilitator superfamily. J Biol Chem 1997; 272: 19304-13.

[54] Pitkin JW, Panaccione DG, Walton JD. A putative cyclic peptide efflux pump encoded by the *TOXA* gene of the plant-pathogenic fungus *Cochliobolus carbonum*. Microbiology 1996; 142: 1557-65.

[55] Paulsen IT, Brown MH, Skurray RA. Proton-dependent multidrug efflux systems. Microbiol Rev 1996; 60: 575-608.

[56] Ito K, Tanaka T, Hatta R, Yamamoto M, Akimitsu K, Tsuge T. Dissection of the host range of the fungal plant pathogen *Alternaria alternata* by modification of secondary metabolism. Mol Microbiol 2004; 52: 399-411.

[57] Xu JR. Map kinases in fungal pathogens. Fungal Genet Biol 2000; 31: 137-52.

[58] Zhao X, Mehrabi R, Xu JR. Mitogen-activated protein kinase pathways and fungal pathogenesis. Eukaryot Cell 2007; 6: 1701-14.

[59] Lev S, Sharon A, Hadar R, Ma H, Horwitz BA. A mitogen-activated protein kinase of the corn leaf pathogen *Cochliobolus heterostrophus* is involved in conidiation, appressorium formation, and pathogenicity: diverse roles for mitogen-activated protein kinase homologs in foliar pathogens. Proc Natl Acad Sci USA 1999; 96: 13542-7.

[60] Ruiz-Roldan MC, Maier FJ, Schafer W. PTK1, a mitogen-activated-protein kinase gene, is required for conidiation, appressorium formation, and pathogenicity of *Pyrenophora teres* on barley. Mol Plant Microbe Interact 2001; 14: 116-25.

[61] Takano Y, Kikuchi T, Kubo Y, Hamer JE, Mise K, Furusawa I. The Colletotrichum lagenarium MAP kinase gene *CMK1* regulates diverse aspects of fungal pathogenesis. Mol Plant Microbe Interact 2000; 13: 374-83.

[62] Xu JR, Hamer JE. MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus *Magnaporthe grisea*. Genes Dev 1996; 10: 2696-706.

[63] Thines E, Weber RW, Talbot NJ. MAP kinase and protein kinase A-dependent mobilization of triacylglycerol and glycogen during appressorium turgor generation by *Magnaporthe grisea*. Plant Cell 2000; 12: 1703-18.

[64] Cho Y, Cramer RA Jr, Kim KH, *et al*. The Fus3/Kss1 MAP kinase homolog *Amk1* regulates the expression of genes encoding hydrolytic enzymes in *Alternaria brassicicola*. Fungal Genet Biol 2007; 44: 543-53.

[65] Di Pietro A, Garcia-MacEira FI, Meglecz E, Roncero MI. A MAP kinase of the vascular wilt fungus *Fusarium oxysporum* is essential for root penetration and pathogenesis. Mol Microbiol 2001; 39: 1140-52.

[66] Mey G, Held K, Scheffer J, Tenberge KB, Tudzynski P. CPMK2, an *SLT2*-homologous mitogen-activated protein (MAP) kinase, is essential for pathogenesis of *Claviceps purpurea* on rye: evidence for a second conserved pathogenesis-related MAP kinase cascade in phytopathogenic fungi. Mol Microbiol 2002; 46: 305-18.

[67] Zheng L, Campbell M, Murphy J, Lam S, Xu JR. The *BMP1* gene is essential for pathogenicity in the gray mold fungus *Botrytis cinerea*. Mol Plant Microbe Interact 2000; 13: 724-32.

[68] Joosten MHAJ, Vogelsang R, Cozijnsen TJ, Verberne MC, De Wit PJGM. The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*-mediated resistance by producing unstable AVR4 elicitors. Plant Cell 1997; 9: 367-379.

[69] Van den Ackerveken GFJM, Van Kan JAL, Joosten MHAJ, Muisers JM, Verbakel HM, De Wit PJGM. Characterization of two putative pathogenicity genes of the fungal tomato pathogen *Cladosporium fulvum*. Mol Plant Microbe Interact 1993, 6: 210-15.

[70] Balint-Kurti PJ, May GD, Churchill AC. Development of a transformation system for *Mycosphaerella* pathogens of banana: a tool for the study of host/pathogen interactions. FEMS Microbiol Lett 2001; 195: 9-15.

[71] Donzelli B, Churchill A. A quantitative assay using mycelial fragments to assess virulence of *Mycosphaerella fijiensis*. Phytopathology 2007; 97: 916-29.

Received: April 16, 2008 Revised: June 17, 2008 Accepted: June 29, 2008