

Identification and Assignment of Function to the Genes of *Aspergillus fumigatus* Expressed at 37 °C

JATA SHANKAR, SHAILLY NIGAM, SHWETA SAXENA, TARUNA MADAN and P. USHA SARMA¹

Molecular Biochemistry and Diagnostics Division, Institute of Genomics and Integrative Biology, Council for Scientific and Industrial Research, Mall Road, Delhi 110007, India

ABSTRACT. *Aspergillus fumigatus*, a fungal pathogen, causes a spectrum of allergic and invasive disorders. In order to rapidly identify genes of this fungus relevant for pathogenesis and as potential antifungal drug targets, 125 expressed sequence tags (ESTs) were generated from 200 phage clones of a non-normalized cDNA library. Out of a novel 68 ESTs, 45 were assigned putative functions based on the sequence similarity. The identities of some of these genes suggest that they may be involved in pathogenesis or autoimmune reactions. Additional genes were identified that are possible targets for the development of antifungal drugs or that may be of use in diagnosing fungal infections.

Key Words. Autoimmune reaction, blast, cDNA library, clone, drug target, expressed sequence tag, sequence similarity.

ASPERGILLUS fumigatus (*Afu*), an opportunistic fungus, is known to cause a spectrum of allergic diseases in the immunocompetent host. The invasive form of aspergillosis is frequently encountered in the immunocompromised host and transplant cases and in patients with leukemia and HIV. Use of currently available antifungal drugs has several limitations owing to their cytotoxicity, hepatotoxicity, and nephrotoxicity. Attempts are being made to understand *Afu* biology and its mechanisms of pathogenesis, in order to facilitate the identification of novel drug targets and development of safer drugs. While *Afu* has an approximately 30-MB genome with 8,000–9,000 genes, limited information is available for *Afu* genes in the public domain (Brookman and Denning 2000). Till now, only 22 recombinant allergens of *Afu* have been reported (Kurup and Banerjee 2000). Some immunodominant allergens/antigens of *Afu* have also been identified and characterized by conventional purification methods (Sarma and Madan 2001). While the *Afu* genome sequence project has been completed recently, the annotation of the genome sequence is still under progress. Thus, there is a need to undertake a simple and rapid approach like random sequencing of cDNA clones, which can effectively identify *Afu* genes.

Comparative analysis of *Afu* genes with non-virulent *Aspergillus* species and other fungi may lead to identification of virulence factors and novel antifungal drug targets. Identification of *Afu* genes with significant sequence similarity to their human counterparts may facilitate the understanding of their possible role in the molecular mimicry triggering autoimmune reactions. Recently, an effort was made to sequence clones from a non-normalized cDNA library of *Afu* (Kessler et al. 2002). However, the sequence data of the 1500 ESTs, reported in the study, are not yet available in the public domain.

In the current study, ESTs were identified by sequencing randomly selected phage clones from a non-normalized cDNA library. Comparative analysis of these ESTs with human and fungal databases has resulted in the identification of a number of genes that may be involved in pathogenesis or that might serve as potential drug targets.

MATERIALS AND METHODS

Isolation and random sequencing of cDNA clones. A cDNA library made from the mycelial extract of *Aspergillus fumigatus* (*Afu*), grown at 37 °C was obtained from Stratagene (La Jolla, CA). The library was expanded in *Escherichia coli* XL-1 Blue MRF' cells according to the manufacturer's instruc-

tions. The Uni-ZAP XR lambda cDNA clones in the library were converted to pBluscript SK (+) phagemid by in vivo excision as described by the manufacturer. The *E. coli* SOLR cells were infected with the phagemid and were plated onto LB-ampicillin (50 µg/ml) agar plates and incubated at 37 °C overnight. Colonies appearing on the plate containing the phagemid having the cDNA insert were used for polymerase chain reaction (PCR) using T3 (5' AATTAACCCCTCACTAAAGGG 3') and T7 (5' CGGGATATCACTCAGCATAAT 3') primers. PCR cycling conditions were 94 °C for 4 min and 28 cycles of 94 °C for 1 min, 58 °C for 1.5 min, and 72 °C for 2 min, followed by a terminal extension cycle at 72 °C for 7 min. PCR amplification product was purified with GFX[®] PCR and Gel band purification kit (Amersham Pharmacia Biotech Inc., Piscataway, NJ) and sequenced two times using T3 primer by ABI-377 DNA Sequencer (Applied Biosystems, Rotkreuz, Switzerland).

Expressed sequence tag data analysis. To validate the sequences of the ESTs obtained, the sequences were subjected to BlastN against the *Afu* whole genome shotgun assembly at Sanger Institute (<http://www.sanger.ac.uk/projects/A.fumigatus>). ESTs were assigned a putative function based on the match with highest sequence similarity, using the BlastX program against National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) non-redundant (nr) protein database of fungi (Class 1). Sequences, that did not show a match in NCBI database were subjected to BlastX in Comprehensive Yeast Genome Database (CYGD) and *A. nidulans* genome database available at The Broad Institute. Sequences (<http://www.broad.mti.edu/cgi-bin/annotation/aspergillus>), that did not show a match in NCBI or CYGD or genome sequence of *A. nidulans* were categorized as Class 2. The annotated sequences were assigned a functional category as per the yeast functional classification catalogue developed by the Munich Information Centre for Protein Sequences (MIPS; <http://www.mips.gsf.de/fungi>). Sequence similarity of *Afu* ESTs with human or other fungal counterparts was identified by using BlastX with the protein databases of *Homo sapiens*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Magnaporthe grisea* and *Candida albicans* available at NCBI. All the analyses were carried out at an expect value less than 1×10^{-3} . Cellular localization of the translated open reading frame (ORF) of each EST was predicted using online software for prediction of protein localization sites in cells (PSORT), available at <http://psort.ims.u-tokyo.ac.jp>.

RESULTS AND DISCUSSION

One hundred and twenty-five (125) sequences were obtained from 200 randomly selected clones and 68 novel ESTs were submitted to the EST database (dbEST) at GenBank (Table 1). Difference between the number of ESTs generated and number

¹ Current Address: Department of Plant Pathology, Indian Research Institute, Pusa, New Delhi-110012, India. E-mail: u.sarma@hotmail.com
Corresponding Author: P. Sarma—Telephone number: 91-11-27666156; FAX number: 91-11-27667471; Email: u.sarma@hotmail.com

Table 1. Bioinformatic analysis of expressed sequence tags of *Aspergillus fumigatus*.

No	Accession No (Frequency per 125)	Putative identification (Function & predicted cellular localization)	No of matches in <i>Afu</i> genome	Sequence similarity to human genome	Sequence similarity to other fungal genomes
Class 1					
1	BG310457 (1)	Ribosomal protein S13 (Protein metabolism, cytoplasmic)	1	+	Nc, Sc, Mg, An
2	BG695820 (1)	Ribosomal protein S9 (Protein metabolism, cytoplasmic)	1	+	Nc, Sc, Mg, An
3	BG695822 (1)	Ribosomal protein L 24 (Protein metabolism, cytoplasmic)	1	+	Nc, Sc, Mg, An
4	BG695823 (1)	Ribosomal protein L3 (Protein metabolism, cytoplasmic)	1	+	Sc, Sp, Nc, An
5	BG695824 (2)	Translation elongation fac- tor 1 alpha (Protein me- tabolism, cytoplasmic)	2	+	Nc, Mg, Sc, Sp, An
6	AF363787 (1)	Protein disulfide isomerase (Protein metabolism, endoplasmic reticulum)	1	+	Nc, Sc, Mg, Sp, An
7	BF325537 (1)	Ribosomal protein L7A (Protein metabolism, cytoplasmic)	1	+	Nc, Mg, Sc, Sp, Ca, An
8	BI962093 (2)	Ubiquitin (Protein metabo- lism, nuclear)	2	+	Nc, Mg, Sc, Sp, Ca, An
9	BI962096 (3)	Ribosomal protein S8 (Protein metabolism, cytoplasmic)	3	+	Mg, Sc, Sp, An
10	BM005702 (1)	Ribosomal protein S9 (Protein metabolism, cytoplasmic)	1	+	Sc, Sp, An
11	BM378043 (1)	Ribosomal protein L18 (Protein metabolism, cytoplasmic)	1	+	Nc, Sc, Mg, Sp, An
12	BM500122 (1)	Ribosomal protein L9-B (L8) (Protein metabo- lism, cytoplasmic)	1	+	Nc, Mg, Sp, Sc, An
13	BM500124 (1)	26S proteasome subunit p28 (Protein metabo- lism, cytoplasmic)	1	+	Nc, Sc, Sp, An
14	CB351760 (1)	Alanyl-tRNA synthetase, (Protein metabolism, cytoplasmic)	1	+	Nc, Sp, Sc, An
15	CB351763 (1)	Ribosomal protein S28 (Protein metabolism, cytoplasmic)	1	+	Mg, Sc, Sp, An
16	BM500121 ^a (1)	Malate synthase (Energy, peroxisomal)	1	—	Nc, Sc, Mg, Ca, An
17	BQ079382 (2)	Transaldolase (Energy, cy- toplasmic)	1	+	Nc, Sc, Mg, Sp, An
18	CB331859 (2)	Nucleoside diphosphate kinase (Nucleotide me- tabolism, cytoplasmic)	2	+	Nc, Mg, Sc, Sp, Ca, An
19	CB331856 (1)	Phosphoribosyl pyrophos- phate kinase (Nucleo- tide metabolism, cyto- plasmic)	1	+	Nc, Sc, Mg, Sp, An
20	CB331860 (2)	Phosphatidylserine decar- boxylase 1 (Amino acid metabolism, cytoplasmic)	2	+	Nc, Mg, Sc, Sp, An
21	CB351756 (1)	Asparagine synthetase (Amino acid metabo- lism, cytoplasmic)	1	+	Nc, Sc, Mg, Sp, An
22	CB351759 (1)	5-methyl tetra hydropter- oyltryglutamate homo- cysteine methyl trans- ferase (Amino acid me- tabolism, cytoplasmic)	1	+	Nc, Sc, Mg, Sp, An

Table 1. Continued.

No	Accession No (Frequency per 125)	Putative identification (Function & predicted cellular localization)	No of matches in <i>Afu</i> genome	Sequence similarity to human genome	Sequence similarity to other fungal genomes
23	CB351762 (2)	H ⁺ -transporting P-type ATPase (Transport facilita- tion, mitochondrial)	2	—	Nc, Mg, Sc, Sp, Ca, An
24	BF440091 (1)	Mitochondrial f-1-ATPase beta—subunit (Energy, mitochondrial)	1	+	Nc, Sc, Mg, Sp, An
25	BF440090 (1)	Ribosomal protein L7 (Protein metabolism, cyto- plasmic)	1	+	Nc, Sc, Mg, Sp, An
26	CB351761 (2)	Alcohol dehydrogenase III (Energy, mitochondrial)	2	+	Nc, Sc, Mg, Ca, Sp, An
27	BG695825 ^a (1)	Transcriptional activator for allantoin (cytoplasmic)	1	—	Mg, Nc, An
28	BM378045 (2)	Oxysterol binding protein homolog 7 (cytoplasmic)	2	+	Nc, Mg, Sc, Sp, Ca, An
29	CB331855 (1)	Allantoate permease (Plas- ma membrane)	1	+	Nc, Sc, Mg, Sp, An
30	CB331857 (2)	R153 core histone H2B gene (Chromosomal pack- aging, nucleus)	2	+	Nc, Sc, Mg, An
31	CB351758 (2)	Flavohemoglobin (Energy, cytoplasmic)	2	+	Nc, Sc, Mg, Sp, An
32	CF413597 ^a (1)	Thiamine regulated gene (Hormone, endoplasmic reticulum)	1	—	Nc, Sc, An
33	CF413598 ^a (1)	Guanylate kinase (Nucleo- tide metabolism, mito- chondrial)	1	—	Nc, Sc, Mg, An
34	CF413599 (2)	Alcohol dehydrogenase II (Energy, cytoplasmic)	2	+	Nc, Sc, Mg, Sp, Ca, An
35	CF413600 (1)	Ribosomal protein 16A (Protein metabolism, cyto- plasmic)	1	+	Nc, Sc, Mg, Sp, An
36	CF413602 (1)	Mitochondrial phosphate (Energy, mitochondrial)	1	+	Nc, Sc, Mg, Sp, An
37	CF413603 (2)	Aldehyde dehydrogenase (Energy, cytoplasmic)	2	+	Nc, Sc, Mg, An
38	CF413604 (1)	Human IgE-dependent his- tamine-releasing factor (Cytoplasmic)	1	+	Mg, Nc, Sp, Sc, An
39	CF413605 (1)	Ser/thr protein kinase (Amino acid metabolism, cytoplasmic)	1	+	Nc, Mg, Sc, Sp, Ca, An
40	CF413606 (2)	Histone H2A (Chromosomal packaging, nucleus)	2	+	Nc, Mg, Sc, Sp, An
41	CF413607 (1)	Kinetochore protein com- plex CBF3 (Nucleus)	1	+	Nc, Mg, Sc, Sp, An
42	CF413609 (2)	26S proteasome regulatory subunit (Protein metabo- lism, cytoplasmic)	2	+	Nc, Mg, Sc, Sp, An
43	CB351757 (1)	Formyltetrahydrofolate hy- drolase (Cytoplasmic)	1	+	Nc, Sc, Mg, An
44	BF219590 (1)	Hypothetical protein (Nu- clear)	1	+	Nc, Mg, An
45	BI962094 ^a (1)	Hypothetical protein (Nu- clear)	1	—	Mg, Nc, An
46	BG695819 (1)	Hypothetical protein (Nu- clear)	1	+	Mg, Nc, An
47	BG695821 ^a (1)	Hypothetical protein (Cy- toplasmic)	1	—	Nc, An
48	BM037654 ^a (1)	Hypothetical protein (Nu- clear)	1	—	Nc, An
49	CB351755 (1)	Predicted protein (Nuclear)	1	+	Nc
50	CF413594 ^a (1)	Hypothetical protein (Mi- tochondrial)	1	—	Nc, An

Table 1. Continued.

No	Accession No (Frequency per 125)	Putative identification (Function & predicted cellular localization)	No of matches in <i>Afu</i> genome	Sequence similarity to human genome	Sequence similarity to other fungal genomes
51	CF413595 ^a (1)	Predicted protein (Cyto- plasmic)	1	—	Nc
52	CF413596 ^a (1)	Predicted protein (Endo- plasmic reticulum)	1	—	Nc, An
53	CF413601 ^a (1)	Hypothetical protein (Nu- clear)	1	—	Nc, An
54	CF413610 ^a (1)	Hypothetical protein (Nu- clear)	1	—	Nc, An
55	CF413613 ^a (1)	Hypothetical protein (Nu- clear)	1	—	Mg, Nc, An
56	BM500123 (1)	Regulatory factor XI (Cy- toplasmic)	1	+	An
57	CF413611 (1)	Hypothetical protein (Ex- tra cellular, including cell wall)	1	+	An
58	BG310456 (2)	Elastin like protein (Nu- clear)	2	—	An
59	BM378044 ^a (1)	Hypothetical protein (En- doplasmic reticulum)	1	—	An
60	CB331862 ^a (1)	Hypothetical protein (En- doplasmic reticulum)	1	—	An
61	BQ276244 ^a (1)	Hypothetical protein (Nu- clear)	1	—	An
62	CB331861 ^a (1)	Hypothetical protein (Mi- tochondrial)	1	—	An
63	CF413612 ^a (1)	Hypothetical protein (Mi- tochondrial)	1	—	An
Class 2	Accession No	Putative identification			
64	CB331858 (1)	Unknown (Nuclear)	1	—	—
65	BG310459 (1)	Unknown (Nuclear)	1	—	—
66	BI962095 (1)	Unknown (Nuclear)	1	—	—
67	BM032616 (1)	Unknown (Cytoplasmic)	1	—	—
68	CF413608 (1)	Unknown (Cytoplasmic)	1	—	—

Class 1. ESTs with a match in NCBI nr protein database of fungi/CYGD/*A. nidulans* genome database with an expect value less than 1×10^{-3} .

Class 2. ESTs with no significant sequence similarity in the databases used.

Neurospora crassa (Nc); *Saccharomyces cerevisiae* (Sc); *Magnaporthe grisea* (Mg); *Schizosaccharomyces pombe* (Sp); *Aspergillus nidulans* (An); *Candida albicans* (Ca); *Aspergillus fumigatus* (*Afu*); Presence (+); Absence (—) and ^a Potential drug targets.

of ESTs submitted was due to occurrence of multiple isolates of the same gene with complete sequence identity (number = 15, redundant ESTs), short sequences (number = 27), and sequences comprising of poly A tract (number = 15). The percentage of redundant ESTs (12%) observed in the present study was less than percentage of redundant ESTs in the non-normalized library (30.8%), but more than the normalized library (8.98%) (Kessler et al. 2002). Minimum, maximum, and average length of sequences was 177, 795, and 470 bp respectively. All 125 ESTs showed complete sequence identity with the genome sequence of *Afu*.

Assignment of putative function to *Afu* ESTs. Putative functions could be assigned to 45 ESTs (Table 1). Four ESTs of the present study (i. e. Trans elongation factor 1, histone H2A, flavohemoglobin, ribosomal protein S8) were reported earlier by Kessler et al. (2002), none of these EST sequences were reported in the public domain. Five ESTs, which did not show a match in the database, could represent *Afu*-specific genes and may encode proteins necessary for infection (Table 1, Class 2). These genes may be further investigated for their role in pathogenesis and specific detection of *Afu* in clinical samples.

Comparative analysis of *Afu* ESTs with other organisms.

Sixty-one and 42 *Afu* ESTs showed sequence similarity with the proteins of *A. nidulans* (89.7%) and *S. cerevisiae* (61.7%) respectively, suggesting an evolutionary divergence between *Aspergillus* and yeast as reported earlier (Kessler et al. 2002) (Table 1). Sequence similarities of *Afu* ESTs with genes in *N. crassa*, *S. pombe*, *M. grisea*, and *C. albicans* were 76.47%, 48.52%, 61.76%, and 13.23% respectively in the present study. Those were different from those observed by Kessler et al. (2002) who reported sequence similarities of *Afu* ESTs with *S. cerevisiae*, *S. pombe* and *C. albicans* as 39%, 32%, and 38.6%, respectively). Among the sequenced 125 clones, 9.6% had sequence similarity to 18S and 25S rRNAs, which was significantly more than the percentage observed by Kessler et al. (2002) who reported 1.7% in the normalised and 1.4% in the non-normalised cDNA library. This may be due to the difference in total number of clones identified in the two studies.

Fifty three ESTs showed similarity to only one contig (single copy ESTs) and 15 ESTs showed similarity to different contigs of *Afu* genome with variable expect values (Table 1, Number of matches). It is possible that some of them may be copies of the same genes with similar function.

Autoimmune reactions have been implicated in the extensive tissue damage observed in allergic bronchopulmonary aspergillosis (ABPA) patients (Purkayastha et al. 2000). Molecular mimicry between the human and pathogenic proteins is one of the proposed mechanisms of autoimmune reactions. Three ESTs encoding L3 ribosomal protein (Identities (I): 47/ 57 (82%), Expect Value (E): 6e-24), L7A ribosomal protein (I: 52/ 86 (60%), E: 6e-22), and histone H2A protein (I: 88/ 107 (82%), E: 3e-42) showed high sequence similarity with their human counter parts. These proteins from other pathogenic organisms have already been implicated in the autoimmune response (Eriksen et al. 1995; Hemmerich et al. 1998). However, these *Afu* ESTs need further evaluation to establish their role in autoimmune reactions.

Potential antifungal drug targets. Based on the comparative analysis (Table 1) we observed that malate synthase, transcriptional activator for allantoin, guanylate kinase, thiamine regulated gene, 12 hypothetical proteins, and 2 predicted proteins are present in single copy in the *Afu* genome and absent in human genome. The presence of similar sequences in other fungi, such as *A. nidulans*, *C. albicans*, and *S. cerevisiae*, suggests that these genes may be useful to explore as potential antifungal drug targets (Table 1). Since malate synthase, an enzyme of glyoxylate cycle, contributes to the virulence of intracellular bacterial (*Mycobacterium tuberculosis*) and fungal (*C. albicans*) pathogens and is absent in mammals, it may be one of the important antifungal drug targets (Lorenz and Fink 2001).

The present study has resulted in rapid and economical identification of various functionally important genes and 5 unique sequences of *Afu*. This compendium of coding sequences of *Afu* may serve as a resource for further investigations to achieve a better understanding of the molecular events of pathogenesis

and for development of effective drugs for aspergillosis in the near future.

ACKNOWLEDGMENTS

We are grateful to the entire sequencing unit and Prof. Samir K. Brahmachari, Director, Institute of Genomics and Integrative Biology for providing facilities. The financial support No. SMM003, provided by the Council of Scientific and Industrial Research, Government of India, is gratefully acknowledged.

LITERATURE CITED

- Brookman, J. L. & Denning, D. W. 2000. Molecular genetics in *Aspergillus fumigatus*. *Curr. Opin. Microbiol.*, **3**:468–474.
- Eriksen, N., Kumar, S. B., Fukuchi, K., Martin, G. M. & Benditt, E. P. 1995. Molecular mimicry: histone H3 and mycobacterial protein epitopes. *Proc. Natl. Acad. Sci.*, **92**:2150–2153.
- Hemmerich, P., Neu, E., Macht, M., Peter, H. H., Krawinkel, U. & Von, M. A. 1998. Correlation between chlamydial infection and autoimmune response: molecular mimicry between RNA polymerase major sigma subunit from *Chlamydia trachomatis* and human L7. *Eur. J. Immunol.*, **28**:3857–3866.
- Kessler, M. M., Willins, D. A., Zeng, O., Richard, G., Mastero, D., Cook, R., Doucettstamm, L., Lee, H., Caron, A., McClanahan, T. K., Wang, L., Greene, J., Hare, R. S., Cottarel, G. & Shimer Jr., G. H. 2002. The use of direct cDNA selection to rapidly and effectively identify genes in the fungus *Aspergillus fumigatus*. *Fungal Genet. Biol.*, **36**:59–70.
- Kurup, V. P. & Banerjee, B. 2000. Fungal allergens and peptide epitopes. *Peptide*, **21**:589–599.
- Lorenz, M. C. & Fink, G. R. 2001. The glyoxylate cycle is required for fungal virulence. *Nature*, **412**:83–86.
- Purkayastha, S., Madan, T., Shah, A., Krishnamurthy, H. G. & Sarma, P. U. Multifunctional antigens of *A. fumigatus* and specific antibodies. 2000. *Appl. Biochem. Biotechnol.*, **83**:271–283.
- Sarma, P. U. & Madan, T. 2001. Diagnostics for Aspergillosis: emerging trends. *Proc. Indian Natl. Sci. Acad.*, **68**:69–96.

Received 08-12-03, 01-08-04, 04-08-04, 04-17-04; accepted 04-17-04