Available online at http://www.journalcra.com

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

International Journal of Current Research
Vol.11, pp.123-128, December, 2010

ISSN: 0975-833X

RESEARCH ARTICLE

COMPARATIVE ANALYSIS OF PLANT AND ANIMAL FUNGAL PATHOGENS TO IDENTIFY PUTATIVE DRUG TARGETS- A *SUBTRACTIVE GENOMICS* APPROACH

Asheesh Shanker[1], Ruchi Jain[1], Amit. R. Rupani*[2], Anushree Mukerji[1] and Tanushree Tiwari[1]

[1]Banasthali University, Rajasthan, India.
[2]The Oxford College of Engineering, Bangalore, India.

ARTICLE INFO

Article History:
Received 18th September, 2010
Received in revised form
17th October, 2010
Accepted 15th November, 2010
Published online 15th December, 2010

Key words:

Subtractive genomic approach,
Homolog identification,
Cellular localization,
BLASTp,
tBLASTn,
CELLO v.2.5,
C6 transcription factor.

Corresponding author:
Amit.R.Rupani,
The Oxford College of Engineering, Bangalore, India,
Email: amit4biotek@gmail.com

ABSTRACT

The recent proliferation and the availability of the complete sequence information of the pathogenic fungi has made it possible to carry out the in silico *subtractive genome* analysis for identification of novel putative drug targets. In this present study subtractive genomic approach is employed to identify novel targets in Ascomycete (*Aspergillus fumigatus, Aspergillus nidulans, Gibberella zeae, Ashbya gossypii*) and Basidiomycete species (*Cryptococcus neoformans*). These fungal species act as plant and animal pathogens. Our study reveals that 52 common core proteins are found in the 5 fungal pathogens of our interest from whole proteome comparisons. Further homolog identification was performed against plants and animals using BLASTp and tBLASTn. 17 out of 52 proteins gave considerable results. Unlike all the proteins, C6 transcription factor (XP_753029.1) was found to have no homolog (hits) in plants and animals. In the later step cellular localization of those 17 proteins was found using CELLO v.2.5. The results show that among those 17 proteins, 3 were nuclear proteins, 7 were cytoplasmic proteins, 5 were mitochondrial proteins & 2 were membrane proteins. Membrane proteins were considered for further analysis as they represent the largest group (70%) of effective drug targets. It was found that C6 transcription factor (XP_753029.1) was identified as a membrane protein. Homolog identification and cellular localization prediction results render C6 transcription factor (XP_753029.1) as promising putative drug target as it was involved in unique biochemical pathways in all the 5 fungal pathogens considered for the study and not present in animal or plant host. Molecular Modeling & further screening of the functional inhibitors against these novel targets may result in discovery of novel broad spectrum anti fungal therapy.

INTRODUCTION

Genome sequencing and bioinformatics have provided the full hypothetical proteome of many pathogenic organisms. Advances in microarray and mass spectrometry have also yielded large output datasets of possible target proteins/genes.

However, the challenge remains to identify new targets for drug discovery from this wealth of information. Further analysis includes bioinformatics and/or molecular biology tools to validate the findings. This is time consuming and expensive, and could fail to yield novel drugs if protein purification and crystallography is impossible (Toomey *et al.,* 2009). Subtractive genomic strategy is developed by assuming that the novel targets identified in the pathogen should be essential for the pathogen, that is, it should be involved in the replication, survival and a important component of various metabolic pathways and mechanisms occurring in the pathogen while at the same time should be absent on the host that is human and should have no homolog in human, so that when a drug or a lead compound is designed considering the potential target it should only be against the mechanism and functionality of the pathogen and not the host (Bhawna Rathi *et al.,* 2009). Fungi cause more plant disease than any other groups of microbes; among fungi, the ascomycetes constitute the largest group of pathogens. Not only plants, fungi have proven to be harmful pathogens for animals too. They cause a number of plant and animal diseases ranging from aspergillosis in humans to stigmatomycosis in plants (Whittaker, 1969). *Aspergillus fumigatus* is an opportunistic pathogen, a saprotroph that is widespread in nature, typically found in soil and decaying organic matter where it plays an essential role in carbon and nitrogen recycling .It causes more infections (Aspergillosis disease) worldwide than any other mold. It is particularly harmful for bone marrow transplant patients, AIDS patients, and other immune compromised individuals (Jean-Paul Latge, 1999).

Aspergillus nidulans (aka *Emericella nidulans*) is a filamentous fungus that is the only member of the genus *Aspergillus* that forms sexual spores through meiosis (Alexander *et al.,* 2008). It is also a homothallic fungus, meaning it can form asexual spores by producing conidiospores that bear chains of asexual spores. It produces the toxin, sterigmatocysin, which leads to food contamination. Repeated inhalation of *A. nidulans* spores also causes allergic aspergillosis, a hypersensitive lung reaction in people who have asthma. Mycetoma (Madura foot) may result when this fungus infects subcutaneous tissue, bone and skin (Hammond *et al.,* 2008). *Cryptococcus neoformans* is an encapsulated fungal organism that can cause disease in apparently immune competent, as well as immune compromised hosts. Since the 1980s, the number of Cryptococcus infections has increased sharply – mainly among people with impaired immunity, including those who have HIV/AIDS or who receive cancer chemotherapy, steroid treatments, or therapy to prevent rejection of transplanted organs."In developing countries, cryptococcosis has emerged as one of the most common opportunistic infections, and a leading cause of meningitis and bloodstream infection" (Kwon-Chung *et al.,* 1996; Mitchell and Perfect, 1995).

Gibberella zeae is a plant pathogen causing the wheat headblight disease which is a devastating disease on wheat and barley. Infection causes the kernels to shrivel and contaminates the remaining grain with mycotoxins, which inhibits protein biosynthesis. *Gibberella zeae* (telomorph *Fusarium graminearum*) is the causal agent of several destructive crop diseases. Identifying interactions among *Gibberella zeae* proteins and understanding their functions can provide insights into pathogenic mechanisms underlying *Gibberella zeae* host interactions. *F. graminearum* protein-protein interaction (FPPI) database provides comprehensive information of protein-protein interactions. Despite great efforts to find resistance genes against *Gibberella zeae*, no completely resistant variety is currently available (Zhao *et al.,* 2009)

Eremothecium gossypii (aka *Ashbya gossypii*) is a filamentous fungus or mold, originally isolated from cotton as a pathogen causing stigmatomycosis. This disease affects the development of hair cells in cotton bolls and can be transmitted to citrus fruits, which thereupon dry out and collapse. *A. gossypii* is a natural overproducer of riboflavin (vitamin B2) which protects its spores against ultraviolet light and therefore made it an interesting organism for industries. Researchers record that, more than 90% of *A. gossypii* genes show both homology and a particular pattern of synteny with *Saccharomyces cerevisiae* (*Schmitz HP 2005*) (Peter Philippsen *et al.,* 2005).

MATERIALS AND METHODS

Retrieval of sequences

In this study, we performed the comparative analysis of whole proteome of selected fungal pathogens. We took the complete proteome of five fungal pathogens from the FTP site of National Center for Biotechnology Information (NCBI), out of which four are ascomycetes *(Aspergillus fumigatus, Aspergillus nidulans, Gibberella zeae, Ashbya gossypii)* and one basidiomycetes *(Cryptococcus neoformans*) (Sarangi *et al.,* 2009). The choice of organisms was such that they cover both plant and animal pathogens.

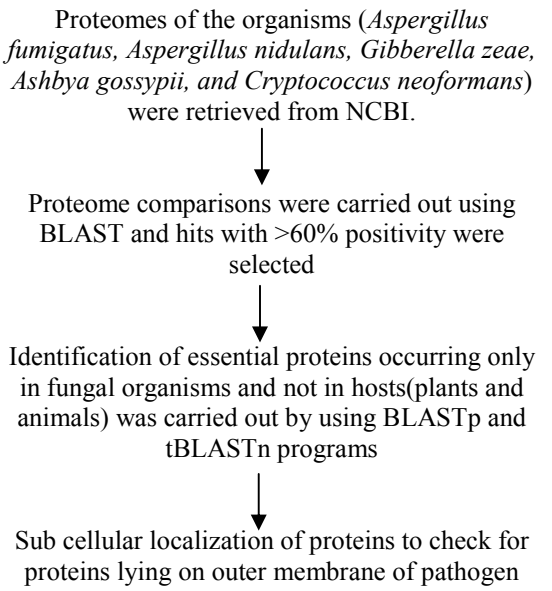


Fig. 1. Overall work flow

Whole proteome comparisons

For proteome comparison the dataset was prepared by separating annotated proteins from the hypothetical and putative ones using a Perl program. Proteome comparisons of annotated proteins of respective organisms were carried out using BLAST (*www.ncbi.nlm.nih.gov/BLAST/*). The parameter for E-value was slightly modified from its default value to 10e-15. Hits with more than 60% positivity and length greater than 50% of the query sequence were selected. Comparisons were performed among all the organisms to select the common proteins that might be responsible for the virulence in these pathogens (Sarangi *et al.,* 2009; Bhawna Rathi *et al.,* 2009).

Homolog identification

To validate that there are no corresponding homologues of these proteins in the human and plant hosts, online BLASTp and tBLASTn programs (*www.ncbi.nlm.nih.gov/BLAST/*) were used. This was necessary so that designing drugs against such sequences may not hamper the functionality of proteins in the host organisms (Sarangi *et al.,* 2009; Bhawna Rathi *et al.,* 2009).

Sub-cellular Localization prediction

Protein sub cellular localization prediction involves the computational prediction of where a protein resides in a cell. Prediction of protein sub cellular localization is an important component as it predicts the protein function and genome annotation, and it can aid the identification of targets (Sarangi *et al.,* 2009). The function and subcellular localization of each non homologous protein is identified by using online subcellular localization prediction tools, CELLO (http://cello.life.nctu.edu.tw/). This tool utilizes various protein properties such as amino acids properties, dipepetide composition, physiochemical properties, and evolutionary information using PSI BLAST. Membrane localized proteins were identified and listed as putative candidate vaccine targets (Rajendra Haribhau Mandage and Amol Shriram, 2010).

RESULTS AND DISCUSSION

The complete proteome of each of the five different fungal organisms considered was retrieved using FTP site of National Center for Biotechnology Information (NCBI). Total proteins present in each organism were identified. The results from NCBI are clearly illustrated in Table 1. As most of the sequences retrieved from FTP were unannotated and our study involves only annotated sequences, therefore, a separation step was performed in which the annotated sequences were partitioned from the rest.

Table 1. Number of annotated and hypothetical proteins parsed and Genome statistics of the organisms and their host

Organism	Total Proteins	Annotated Proteins	Hypothetical Proteins	Host
Aspergillus fumigatus	9630	2047	7583	Plant and animal
Aspergillus nidulans	2867	138	2729	Plant and animal
Cryptococcus neoformans	5615	408	5207	Plant and animal
Ashbya gossypii	4722	0	4722	Plant
Gibberella zeae	7970	82	7888	Plant

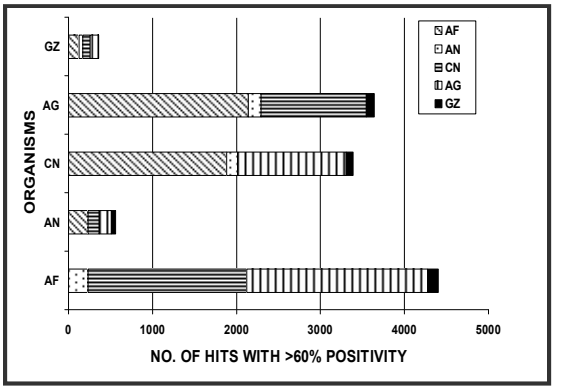
It is clear that *A. fumigatus* consisted of the highest number of proteins (i.e. 9630) out of which 22% were annotated and rest was hypothetical (Table 2). On the other hand, *A. nidulans* contained the lowest number of proteins (2867) and only 5% of them were annotated. Moreover, *G. zeae* has around 7970 proteins with lowest percentage (i.e. 1.02%) of annotated proteins. All the protein sequences in *A. gossypii* proteome were found to be unannotated.

Table 2. Inter proteome comparison results with hits more than 60% positivity

Organisms	AF	AN	CN	AG	GZ*
AF	-	235	1889	2152	125
AN	233	-	132	147	47
CN	1881	136	-	1287	86
AG	2138	153	1254	-	96
GZ	124	48	88	98	-

*[*AF: A. fumigatus, AN: A. nidulans, CN: C. neoformans, AG: A. gossypii, GZ: G. zeae]*

Whole Proteome Comparisons were carried out using BLAST. The Table 2 summarizes the number of hits that were above the set threshold of 60% positivity. This value was chosen so that significantly similar hits may be incorporated without the intervention of randomly generated alignments. This in turn ensures a high selectivity without any compromise with the sensitivity. The highest number of hits (>60% positivity) are found between *C. neoformans* and *A. fumigatus* and the lowest is between *A. nidulans* and *G. zeae.* The BLASTp output thus generated is particularly biased towards the proteomes of *A. gossypii* and *C. neoformans* (Fig. 2). This is because the data source of *A. gossypii* proteome was unannotated, therefore during the BLASTp analysis the whole proteome (and not the annotated sequences) of this organism was incorporated and thus shows the maximum number of results. The higher number of



[AF: A. fumigatus, AN: A. nidulans, CN: C. neoformans, AG: A. gossypii, GZ: G. zeae]

Fig. 2. Hits whose positivity was more than 60%

hits in case of *C. neoformans* is that it is treated as a model organism and thus has the maximum number of annotated proteins (after *A. fumigatus*). The Table 3 summarizes the proteins that were found to be bi directional best hit (BBH) during comparison. For this purpose the BLASTp outputs were compared with each other in the form A vs B and B vs A where 'A' was once treated as query and second time acted as database. This process was repeated for all organisms mentioned in table. It can be seen from the table that the number of common proteins is highest between AF vs CN and CN vs AF.

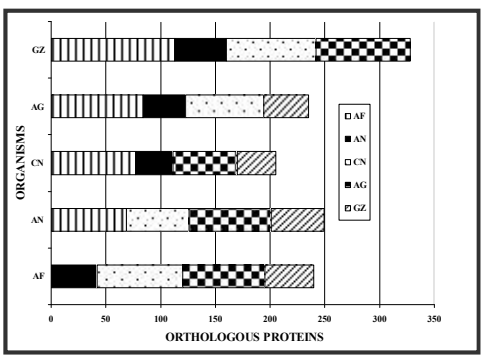
Table 3. Number of common proteins found to be bi directional best hits during comparisons

Organisms	AF	AN	CN	AG	GZ
AF		69	77	84	113
AN	42		34	39	47
CN	78	57		71	82
AG	75	75	59		86
GZ	45	48	35	41	-

Table 4. Cellular location of proteins and Significant hits of tBLASTn against the composite plant database

S. No.	Protein Name(Accession number)	Cellular Localization	Total length of Protein (aa)	Identity
1.	P-type ATPase (XP_749997.1)	Plasma Membrane	959	312
2.	G protein complex alpha subunit GpaB (XP_752663.1)	Nuclear	356	21
3.	G-protein complex alpha subunit GpaA (XP_752684.1)	Nuclear	353	23
4.	C6 transcription factor (XP_753029.1)	Plasma Membrane	503	No Hits
5.	ADP-ribosylation factor family protein (XP_752866.1)	Cytoplasmic	181	60
6.	Antigenic mitochondrial protein HSP60 (XP_755263.1)	Mitochondrial	587	117
7.	Mitochondrial Hsp70 chaperone (XP_755328.1)	Mitochondrial	685	142
8.	60S ribosomal protein L3 (XP_755517.1)	Cytoplasmic	392	92
9.	Glucose-6-phosphate 1-dehydrogenase (XP_754767.1)	Cytoplasmic	502	119
10.	Cysteine synthase (XP_748124.1)	Mitochondrial	371	31
11.	Citrate synthase (XP_747920.1)	Mitochondrial	474	61
12.	Rab small monomeric GTPase Rab7 (XP_753438.1)	Chloroplast	171	69
13.	Citrate synthase (XP_747718.1)	Mitochondrial	465	56
14.	Calcineurin Ca2+-binding regulatory subunit (XP_747624.1)	Nuclear	193	58
15.	14-3-3 family protein (XP_750568.1)	Cytoplasmic	271	79
16.	Pyruvate kinase (XP_750636.1)	Cytoplasmic	527	126
17.	DNA replication factor C subunit Rfc5 (XP_755557.1)	Mitochondrial	294	36

52 proteins were common among these five fungal organisms that might be responsible for the virulence. It was very important to check for the presence of any of these 52 proteins in host (plant and animal). BLASTp and tBLASTn online programs (*www.ncbi.nlm.nih.gov/BLAST/*) provide us with those proteins for which there were no corresponding homologues in humans and plants. These proteins if found to be essential, could further act as drug targets. Only 17 of the 52 proteins gave results that were considerable. The identity column in Table 4 implies that every protein had homologues in humans except C6 transcription factor (XP_753029.1). This protein had no homologues in plants as well and therefore, is the most significant result, as it lacks similarity with both the plants and humans, showing that it has high probability of being a drug target.



*[*AF: A. fumigatus, AN: A. nidulans, CN: C. neoformans, AG: A. gossypii, GZ: G. zeae]*

Fig. 3: Bi- directional best hits found in the cross-

These 17 proteins may be considered for identification of putative drug targets. But, in any organism, membrane localized proteins represent the largest group (70%) of effective drug targets

(Lundstrom, 2007). So the next step was to find cellular localization of these proteins and identify how many of them were membrane proteins. The sub cellular localization prediction for these 17 proteins was done by CELLO: a subcellular localization predictive system (http://cello.life.nctu.edu.tw/) that can locate the outer membrane proteins which could be probable vaccine targets. The results obtained by CELLO for 17 proteins are given by Cellular Localization column in Table 4. It is evident from the table that out of 17 proteins only 2 proteins namely P-type ATPase (XP_749997.1) & C6 transcription factor (XP_753029.1) were membrane proteins.

CONCLUSION

The comparative proteome analysis of various fungal plant pathogens belongs to ascomycetes and basidiomycetes were successfully done. The main idea behind this study was to explore the knowledge about common proteins responsible for virulence and pathogenecity in the organisms considered. The comparative studies reported here are based upon the 52 common proteins that were extracted using BLAST. Using BLASTp and tBLASTn online programs only 17 proteins were identified as proteins that could be used as drug targets. Almost every protein had homologues in hosts (plant and animal) except C6 transcription factor (XP_753029.1). Membrane localized proteins represent the largest group (70%) of effective drug targets. CELLO: a subcellular localization predictive system, classified C6 transcription factor (XP_753029.1) as a membrane protein. Therefore it can be concluded that C6 transcription factor (XP_753029.1) protein is a highly promising putative drug target for the fungal pathogens under consideration. As subtractive genomic approach is applied for the identification of drug target, so the drug would be specific for the pathogen and not lethal to the host. Furthermore, there is no drug available in DrugBank having C6 transcription factor (XP_753029.1) as a target. There are no corresponding structures for this protein submitted in the PDB (*www.pdb.org/pdb/home/home.do/*). Molecular modeling of the targets will decipher the best possible active sites that can be targeted by simulations for drug design. Virtual screening against these potential targets might be useful in the discovery of potential therapeutic compounds against these fungal pathogens.

REFERENCES

Alexander, WG., Raju, NB., Xiao, H., Hammond, TD., Perdue, Metzenberg, R.L., Pukkila, PJ. And Shiu PKT, 2008. DCL-1 Colocalizes with other components of the MSUD machinery and is required for silencing. *Fungal Genet. Biol.,* 45: 719-727.

Bhawna Rathi, Aditya N. Sarangi and Nidhi Trivedi, 2009. Genome subtraction for novel target definition in Salmonella.*Typhi.Bioinformation.,* 2009; 4(4): 143–150

Hammond, TM., Bok, JW., Andrewski, MD., Domínguez, Y., Scazzocchio, C. and Keller, NP. 2008. RNA silencing gene truncation in the filamentous fungus *Aspergillus nidulans*. *Euk. Cell.,* 7: 339-349.

Jean-Paul Latge, 1999. *Aspergillus fumigatus* and *Aspergillosis*. *Clin. Microbiol. Rev.,* 12(2): 310–350.

Kwon-Chung, K.J, Chang, YC. and Penoyer, LA. 1996. The second capsule gene of *Cryptococcus neoformans*, CAP64, is essential for virulence. *Infect Immun.,* 64: 1977–1983.

Lundstrom, K. 2007. Latest development in drug discovery on G protein-coupled receptors. *Curr. Protein Pept. Sci*., 7:465-70.

Mitchell, TG and Perfect, JR, 1995. Cryptococcosis in the era of AIDS—100 years after the discovery of *Cryptococcus neoformans. Clin. Microbiol. Rev.,* 8: 515-48.

Peter Philippsen., Andreas Kaufmann and Hans- Peter Schmitz, 2005. Homologues of yeast polarity genes control the development of multinucleated hyphae in *Ashbya gossypii*. *Current Opinion in Microbiology.,* 8: (4):370-377.

Rajendra Haribhau Mandage and Amol Shriram Wadnerkar.2010.Substractive genomics approach to identify potential therapeutic targets in Leishmania Donovani. International Journal of Pharma and Biosciences., Vol.1/Issue- 3/Jul-Sep.

Sarangi, AN., Aggarwal, R., Rahman, Q. and Trivedi, N. 2009. Subtractive Genomics Approach for in Silicoidentification and Characterization of Novel Drug Targets in Neisseria Meningitidis Serogroup B. *J. Comput. Sci. Syst. Biol.,* 2: 255-258.

Toomey D, Heinrich C, Hoppe MP, Brennan KB and Nolan AJ, Chabb, 2009. Genomes 2 Drugs: Identifies target proteins and lead drugs from proteome data. plos one 4: e6195.

Whittaker, RH. 1969. New concepts of kingdoms of organisms. *Science.,* 163: 150-161.

Zhao. XM., Zhang, XW., Tang, WH. and Chen, L. 2009. FPPI: *Fusarium graminearum* protein-protein interaction database. *J. Proteome. Res.,* 8: 4714-4721.