



Identification of Novel Drug Targets in Pathogenic *Aspergillus Fumigatus*: An *in Silico* Approach

Reena Gupta and Chandra Shekhar Rai(✉)

University School of Information, Communication and Technology,
Guru Gobind Singh Indraprastha University, New Delhi, India
{reena, csrai}@ipu.ac.in

Abstract. Invasive aspergillosis due to fungus *Aspergillus fumigatus* is considered a major human disease infecting immunosuppressed patients. Screening of novel drug targets for this opportunistic pathogen is a need of the hour due to the constraints of antifungal remedies, reactions, drug resistance & toxicities, expense and drug-drug interactions. In order to overcome these limitations, novel antifungal drug targets are needed. Thus, to find putative drug targets, we explored combination of subtractive and comparative genomics approach in the current work. Whole proteome of *A. fumigatus* is screened for homology analysis via target identification tool 'TiDv2'. TiDv2 classifies proposed drug targets as new and virulent in less time and at low cost. Genes from homology analysis are compared with humans and gut flora to achieve non homology with them to attain broad spectrum drug target. Thereafter, druggability and virulence factor analysis is performed. The resultant dataset is prioritized for metabolic pathway analysis. In addition, functional annotation and subcellular localization is accomplished. The results reveal that 5 genes namely His6, FasA, PabaA, Ftma and erg6 might work as promising wide-spectrum drug targets for *A. fumigatus*. Hence, these genes may improve current curative failures in the medication of invasive aspergillosis. However, these possible drug targets should be verified to confirm before target based lead discovery.

Keywords: Drug targets · *Aspergillus fumigatus* · Subtractive genomic analysis · Target identification · TiD · Comparative genomics

1 Introduction

Invasive aspergillosis is a fatal infection caused by pervasive fungus *A. fumigatus* in immunosuppressed hosts [1]. Qualitative or quantitative deficiencies are manipulated by *Aspergillus fumigatus* in the host resistant defence to induce intrusive infection. It is extremely difficult to treat this infection in immunocompromised hosts. The fatality from intrusive aspergillosis surpasses 50% even post chemotherapy. In 1996, in U.S., the estimate of curing diagnosed patients of intrusive aspergillosis was \$633 M, where the amount of each case was roughly \$64,500 [1]. In acute immunosuppressed patients,

presently accessible antifungals have moderate efficiency [2]. According to a 2009 article in the journal *Emerging Infectious Diseases*, exponential increasing issue in curing *Aspergillus* infection is drug resistance [1, 3, 4]. Moreover, therapeutic medication is usually unsuccessful and is obstructed by harsh side effects of available antifungals. Hence, new drug targets are required which would be successful in curing aspergillosis and have less side effects on host [5].

Traditional approach has been used to screen drug targets and vaccines against aspergillosis for last two decades [5]. The screening of novel potent drug targets is limited due to lack of data, methodology and information. All these major problems have necessitated to screen possible drugs against this hazardous pathogen. Search of new curative drug targets resulted in screening of novel drug candidates. Due to the online availability of newer genes, genomes and protein databases, comparative genomic analysis is considered one of the most reliable strategy for potential drug designing [6, 7]. In the present work, we have attempted to develop a procedure to resolve the restraints of studies done so far. Our computational procedure incorporates whole genome analysis, subtractive/comparative analysis, metabolic pathway analysis and sub cellular localization analysis.

In the present work, we have used subtractive/comparative genome analysis where deduction of identical protein set is accomplished followed by comparison of pathogen and human host genome. Essential gene set is the smallest set of genes necessary for the growth and sustenance of the pathogen. In the proposed work we have adopted in silico approach through subtractive/comparative genomics on the whole proteome of *A. fumigatus* to identify unique gene set in the metabolic pathways using KEGG and gene set related with the membrane using CELLO [8]. This computational approach is successfully adopted to identify drug targets in many other fungi and bacteria such as *Helicobacter pylori* [9], *Staphylococcus aureus* [10], I [11], *Bacillus anthracis* [12], *Neisseria meningitides* [13] etc. Hence, we have adopted in silico approach to detect the potential novel drug targets which are virulent and have unique metabolic pathways. However, these putative drug targets should be verified before drug designing.

2 Materials and Methods

Pathogenic proteins of *A. fumigatus* were screened to a wide-spectrum anti-aspergillus drug target detection procedure that incorporates comparative and subtractive genomics technique. The complete description of protocol of the target detection procedure is shown in Fig. 1.

2.1 Program and Data Acquisition of Pathogen Protein

Complete proteome of *A. fumigatus* was fetched from NCBI protein database (<http://www.ncbi.nlm.nih.gov/>) and subjected to TiDv2, a target identification software. Fasta file format was chosen among various file formats for drug target identification. TiDv2 is a standalone software developed by us. It is an extension of TiD developed by our team [14]. It includes functions for paralog analysis, essentiality analysis and non-homolog analysis that excludes paralogous proteins, includes essential genes from DEG, CEG or

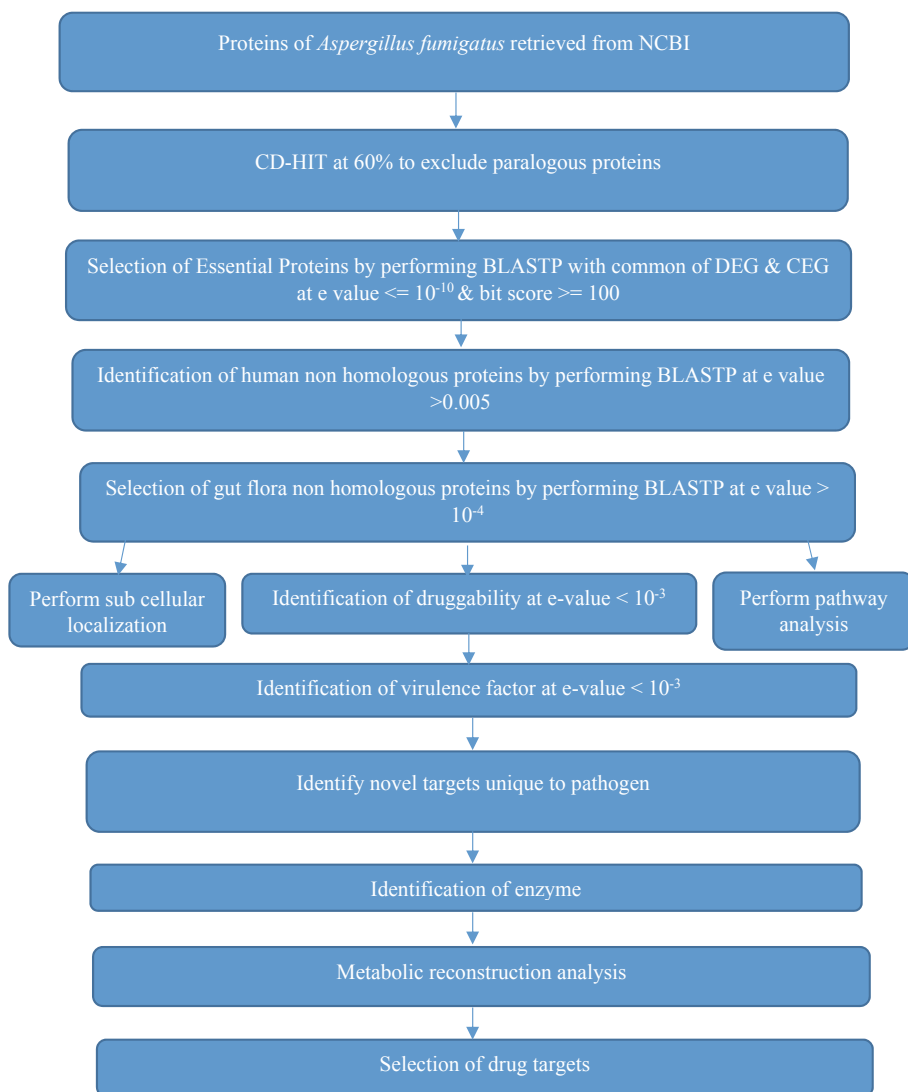


Fig. 1. Complete description of protocol of target detection procedure

common of both DEG and CEG and excludes host and gut flora homologous proteins from the essential gene dataset [12, 15].

Target prioritization analysis tab characterized the resultant essential gene dataset on druggability and virulence. Virulent proteins are vital for the progression, pathogenicity and sustenance of the pathogen [1].

Target prioritization was done on gutflora non homologous proteins with the help of online target prioritization tools such as Kyoto Encyclopedia of Genes and Genomes (KEGG), INTERPROSCAN and Cello for pathway analysis, functional annotation and

subcellular localization respectively. BioPython v2.7.10 scripts were written and integrated with executables and datasets in Microsoft Visual Studio 2015 platform for designing TiDv2. In TiDv2, we incorporated fungal radio button in DEG tab to identify essential genes of fungus.

2.2 Identification of Paralogous Proteins

Target detection and characterization parameters and procedures were followed as mentioned in our earlier work [14]. CD-HIT suite was used to eliminate redundant paralogous proteins from downloaded fungal proteome. The sequence identity cut off value is 60%.

2.3 Selection of Essential Proteins

The selected paralogous proteins were screened to perform BLASTP with DEG at e-value 10^{-10} and bit score ≥ 100 . The resultant genes are vital for the sustenance and progression of the pathogen.

2.4 Identification of Human and Gut Flora Non-homologs

We identified human non-homolog proteins from essential fungal genes based on threshold e-value at 0.005 and bit score ≥ 100 . Gutflora non homologous proteins were obtained at e-value 10^{-4} and bit score ≥ 100 . The screened dataset was postulated as possible drug targets and characterized for the possibility of drug and virulence. The resultant gene set was further outlined with UniProt identifiers (<http://www.uniprot.org/>).

2.5 Metabolic Pathway Analysis

To check the presence of mapped gene set in the metabolic pathways, this dataset of potential targets was characterized at the KAAS server (KEGG Automatic Annotation Server). Further, these targets were screened through BLAST in KEGG database to achieve functional annotation of these putative targets. KAAS server assigns K numbers to sequences that are similar and has bi-directional best hit through programmed procedure that allows establishment of KEGG pathways. The outcome gene set has KO (KEGG Orthology) assignments of sequences that determine metabolic proteins [16].

2.6 Subcellular Localization

We have analysed the metabolic set of proteins in CELLO v2.5 for subcellular localization and biological significance. Subcellular localization detects the potential proteins as cytoplasmic, periplasmic, outer membrane, extra cellular and inner membrane protein [16].

2.7 Identification of Druggability and Virulence Factors

The genes obtained after non homologous analysis with gut flora were further subjected to druggability analysis and virulence analysis. Virulence factor provides new insight in the development of potential anti-fungal drugs [17].

2.8 Selection of Anti Fungal Drug Targets

Finally, proteins obtained after subtractive analysis are putative anti-fungal drug targets. These targets would be experimentally validated before lead discovery. These targets would be potential for the screening of novel anti-fungal drug targets.

3 Results and Discussion

Due to active progression of bacterial proteins, resistance to the present antibiotics is expanding which leads to global health hazard. This entails the exercise to design antibacterial candidates targeted at novel drug targets. The correlation of genomes and gene products have been promoted by the desegregated databases, computerized sequencing of genes, algorithms and tools, which further recommend genome-based drug target designing. Moreover, a computerized, quick and effective methodology to identify drug targets for a specific pathogen from whole proteome of bacteria or fungi develop an effective way to deal with bacterial or fungal proteins expanding with drug resistance in present scenario [14].

The present work is based on advanced comparative and subtractive genomic approach. The unique and vital proteins are critical for *A. fumigatus* progression, pathogenicity and survival [6]. A compelling approach to tackle the demanding aspergillosis infection is to find target proteins and their essential metabolic pathways. The results of the systematic procedures for mining putative drug targets are presented in Fig. 2.

3.1 Identification of Paralogous Proteins

Whole genome sequence (*.faa) of *A. fumigatus* which is responsible for causing infectious disease was successfully retrieved from NCBI and screened through TiDv2 tool. In the prevailing examination, 56757 proteins were retrieved from NCBI. These proteins were screened in CD-HIT suite at 60% identity value as threshold which produce 10815 non paralogous genes.

3.2 Identification of Essential Proteins

Non-paralogous proteins were run in Essentiality Analysis Tab of TiD. These 10815 proteins were run on BLASTP with common of DEG and CEG at e-value $\leq 10^{-10}$ and bit score ≥ 100 . 1860 proteins were observed to be essential for the growth of *A. fumigatus*.

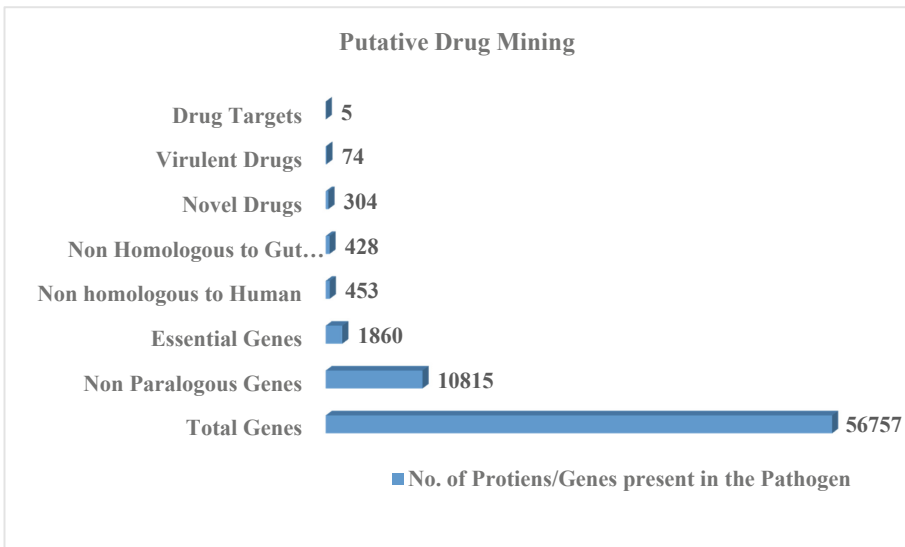


Fig. 2. Results of putative drug mining

3.3 Selection of Human and Gut Flora Non-homologs

1860 essential proteins were screened in Non-Homology Analysis Tab of TiD. Human non homologous proteins were detected by screening through BLASTP at e-value > 0.005 . By performing BLASTP, we detected ample analogies between pathogen and human host at e-value < 0.005 . Proteins identified as homologous are excluded so that target candidates are safe for human host and has less side effects. We found 453 pathogenic proteins non homologous to human host. Further, out of 453 human non homologous proteins, we identified 428 proteins as non homologous to gutflora at e-value $> 10^{-4}$.

3.4 Metabolic Pathway Analysis

We found 395 uniprot identifiers when 428 gut flora non homologous proteins were mapped with uniprot identifiers of *A. fumigatus*. To study the relationship of these genes in metabolic pathways, the mapped geneset of potential drug targets was characterized at the KAAS server (KEGG Automatic Annotation Server). Further, these targets were screened through BLAST in KEGG database to achieve metabolic pathway of these potential targets. The outcome gene set contains KO (KEGG Orthology) assignments of sequences that determine metabolic pathway. There were 13817 total KO assignments in the whole proteome.

3.5 Subcellular Localization

Subcellular localization of probable drug targets in the present examination revealed that 137 drug targets were cytoplasmic, 132 were inner membrane, 81 were periplasmic, 51 were outer membrane and 26 were extra cellular proteins (as shown in Table 1). Position

of these putative drug targets is vital in future at the time of designing drug or vaccine. A vital condition in rational drug design is the balance between the subcellular localization report of a drug target with the pharmaceutical features of lead molecules focused to it.

Table 1. Subcellular localization of putative drug targets

S.No.	Accession no.	Name of protein	KO number	Pathways	Enzyme
1.	EAL84512.1	sterol 24-c-methyltransferase	K00559	Steroid biosynthesis	Yes
2.	EAL85149.1	nonribosomal brevianamide peptide synthase FtmA	K18281	Staurosporine biosynthesis	Yes
3.	EAL85558.1	para-aminobenzoate synthase PabaA	K13950	Folate biosynthesis	Yes
4.	EAL86700.1	fatty acid synthase alpha subunit FasA	K00667	Fatty acid biosynthesis	Yes
5.	KMK59268.1	5-proFAR isomerase His6	K01814	Histidine metabolism	Yes

3.6 Identification of Druggability and Virulence Factors

We identified 304 proteins as novel drug targets, when 428 gut flora non homologous proteins were screened for druggability analysis. Out of 428 proteins, 74 proteins were found to be virulent. Virulence factor analysis is essential to identify drug targets of the pathogen [17]. The drug targets with virulence factor are vital for inception of infection and perseverance in host. Hence, experimental lead molecules which are developed on the basis of these potential drug targets are integral for formulating a novel curative procedure against pathogens.

3.7 Selection of Anti Fungal Drug Targets

Using subtractive genomic analysis, we detected 5 proteins as novel drug targets having virulence factor and unique metabolic pathway. Out of 5 novel drug targets, 1 was outer membrane protein and 4 were cytoplasmic proteins. The detailed information of proteins with uniprot identifier and gene name is shown in Table 2.

This subtractive/comparative genome analysis has been efficiently practiced in various bacterial and fungi genes such as *Mycoplasma pneumoniae* M129 [18], *Staphylococcus aureus* N315 [8], *Mycobacterium tuberculosis* F11 [19], *Neisseria meningitidis* serogroup B for screening of drug targets [13]. The proteins screened in the present study will fairly lead to a positive way in developing potential therapeutic drugs for future researchers.

Table 2. Detailed information of putative drug targets

S.No.	Description	Gene name	Uniprot identifier	Subcellular localization
1.	EAL84512.1 sterol 24-c-methyltransferase, putative [<i>Aspergillus fumigatus</i> Af293]	erg6	Q4W9V1	cytoplasmic
2.	EAL85149.1 nonribosomal brevianamide peptide synthase FtmA [<i>Aspergillus fumigatus</i> Af293]	FtmA	Q4WAW3	cytoplasmic
3.	EAL85558.1 para-aminobenzoate synthase PabaA [<i>Aspergillus fumigatus</i> Af293]	PabaA	Q4WDI0	Outer membrane
4.	EAL86700.1 fatty acid synthase alpha subunit FasA [<i>Aspergillus fumigatus</i> Af293]	FasA	Q4WEX7	cytoplasmic
5.	KMK59268.1 5-proFAR isomerase His6 [<i>Aspergillus fumigatus</i> Z5]	His6	A0J5PRW7	cytoplasmic

4 Conclusion

In this era of bioinformatics, previous traditional methods of drug discovery and designing are becoming obsolete. Mammoth of biological database restructures the process of drug discovery and designing procedures. In order to reduce the hazards of dangerous aspergillosis, we have to design novel probable drug targets. In this study, we have executed subtractive genomics and comparative genomics analysis on pathogenic proteins of *A. fumigatus* and identified 5 novel potential drug targets (His6, FasA, PabaA, FtmA and erg6) for designing of drug and development of vaccine. Moreover, these putative drug targets execute an important role in the vital metabolic pathways. However, outcomes of the present work can be authenticated by further clinical research to confirm their execution in restraining the production and disturbing the virulence factor of *A. fumigatus* pathogen. Promisingly, this study would help to design new anti-*A. fumigatus* drugs against aspergillosis. Hence, this study would be undoubtedly favourable because of recent findings which state the growing risk of resistance of *A. fumigatus* pathogen.

References

1. Hohl, T.M., Feldmesser, M.: *Aspergillus fumigatus*: principles of pathogenesis and host defense. *Eukaryot. Cell* **6**, 1953–1963 (2007). <https://doi.org/10.1128/EC.00274-07>
2. Fair, R.J., Tor, Y.: Antibiotics and bacterial resistance in the 21st century. *Perspect. Med. Chem.* **6**, 25–64 (2014). <https://doi.org/10.4137/PMC.S14459>
3. Cramer, R.A., et al.: Calcineurin target CrzA regulates conidial germination, hyphal growth, and pathogenesis of *Aspergillus fumigatus*. *Eukaryot. Cell* **7**, 1085–1097 (2008). <https://doi.org/10.1128/EC.00086-08>
4. Davison, J.: Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. *PLOS Pathog*
5. Ebel, F., Schwienbacher, M., Beyer, J., Heesemann, J., Brakhage, A.A., Brock, M.: Analysis of the regulation, expression, and localisation of the isocitrate lyase from *Aspergillus fumigatus*, a potential target for antifungal drug development. *Fungal Genet. Biol.* **43**, 476–489 (2006). <https://doi.org/10.1016/j.fgb.2006.01.015>
6. Abadio, A.K.R., Kioshima, E.S., Teixeira, M.M., Martins, N.F., Maigret, B., Felipe, M.S.S.: Comparative genomics allowed the identification of drug targets against human fungal pathogens. *BMC Genom.* **12**, 75 (2011). <https://doi.org/10.1186/1471-2164-12-75>
7. Ou-Yang, S.-S., Lu, J.-Y., Kong, X.-Q., Liang, Z.-J., Luo, C., Jiang, H.: Computational drug discovery. *Acta Pharmacol. Sin.* **33**, 1131–1140 (2012). <https://doi.org/10.1038/aps.2012.109>
8. Hossain, M., et al.: Identification of potential targets in *Staphylococcus aureus* N315 using computer aided protein data analysis. *Bioinformatics* **9**, 187–192 (2013). <https://doi.org/10.6026/97320630009187>
9. Dutta, A., Singh, S.K., Ghosh, P., Mukherjee, R., Mitter, S., Bandyopadhyay, D.: In silico identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. *Silico Biol.* **6**, 43–47 (2006)
10. Uddin, R., Saeed, K., Khan, W., Azam, S.S., Wadood, A.: Metabolic pathway analysis approach: identification of novel therapeutic target against methicillin resistant *Staphylococcus aureus*. *Gene* **556**, 213–226 (2015). <https://doi.org/10.1016/j.gene.2014.11.056>
11. Shanmugham, B., Pan, A.: Identification and characterization of potential therapeutic candidates in emerging human pathogen mycobacterium abscessus: a novel hierarchical in silico approach. *PLoS ONE* **8**, e59126 (2013). <https://doi.org/10.1371/journal.pone.0059126>
12. Rahman, A., et al.: Identification of potential drug targets by subtractive genome analysis of *Bacillus anthracis* A0248: an in silico approach. *Comput. Biol. Chem.* **52**, 66–72 (2014). <https://doi.org/10.1016/j.compbiolchem.2014.09.005>
13. Narayan Sarangi, A.: Subtractive genomics approach for in silico identification and characterization of novel drug targets in *Neisseria Meningitidis* Serogroup B. *J. Comput. Sci. Syst. Biol.* **02** (2009). <https://doi.org/10.4172/jcsb.1000038>
14. Gupta, R., Pradhan, D., Jain, A.K., Rai, C.S.: TiD: standalone software for mining putative drug targets from bacterial proteome. *Genomics* **109**, 51–57 (2017). <https://doi.org/10.1016/j.ygeno.2016.11.005>
15. Yeh, I., Hanekamp, T., Tsoka, S., Karp, P.D., Altman, R.B.: Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. *Genome Res.* **14**, 917–924 (2004). <https://doi.org/10.1101/gr.2050304>
16. Gupta, R., Verma, R., Pradhan, D., Jain, A.K., Umamaheswari, A., Rai, C.S.: An in silico approach towards identification of novel drug targets in pathogenic species of *Leptospira*. *PLoS ONE* **14**, e0221446 (2019). <https://doi.org/10.1371/journal.pone.0221446>
17. Gauwerky, K., Borelli, C., Korting, H.C.: Targeting virulence: a new paradigm for antifungals. *Drug Discov. Today* **14**, 214–222 (2009). <https://doi.org/10.1016/j.drudis.2008.11.013>

18. Kumar, G.S.V., Sarita, S., Kumar, G.M., Kk, P., Pk, S.: Definition of potential targets in mycoplasma pneumoniae through subtractive genome analysis (2010). <https://doi.org/10.4172/jaa.1000020>
19. Hosen, M.I., et al.: Application of a subtractive genomics approach for in silico identification and characterization of novel drug targets in Mycobacterium tuberculosis F11. *Interdiscip. Sci. Comput. Life Sci.* **6**, 48–56 (2014). <https://doi.org/10.1007/s12539-014-0188-y>