Trends in Bioinformatics

ISSN 1994-7941

science alert

ANSInet
an open access publisher
http://ansinet.com

Trends in Bioinformatics 8 (2): 52-62, 2015
ISSN 1994-7941 / DOI: 10.3923/tb.2015.52.62

Identification of Putative Therapeutic Targets in *Candida tropicalis*: An *in silico* Approach

[1]Haribalaganesh Ravinarayanan, [2]Richard Coico and [1]Krishnan Sundar
[1]Department of Biotechnology, Kalasalingam University, Krishnankoil, 626126, Tamilnadu, India
[2]Department of Cell Biology, SUNY Downstate Medical Center, Brooklyn, NY, 11203, USA

Corresponding Author Krishnan Sundar, Department of Biotechnology, Kalasalingam, University Krishnankoil, 626126, Tamilnadu, India Tel: +91-4563-289042/105 Fax: +91-4563-289322

ABSTRACT

The prolonged use of antibiotics results in drug resistance in pathogenic microorganisms. This necessitates the identification of novel drug targets that are useful for the development of effective antimicrobial drugs. In the post-genomic era, computational tools and methods have contributed immensely in identifying such novel targets, thereby accelerating the drug discovery process. In the present study, an extensive *in silico* analysis of the proteome of the pathogenic yeast, *Candida tropicalis* was performed to identify potential drug targets. The complete proteome of *C. tropicalis* retrieved from Uniprot was analysed using the CD-HIT algorithm followed by BLAST for eliminating proteins homologous to human proteome. The selected proteins were then analyzed using DEG database for identifying critical genes for the survival. The identified essential proteins were subjected to pathway analysis using KEGG to predict their involvement in metabolism. This approach resulted in the identification of 20 potential drug targets present in *C. tropicalis*.

Key words: DEG, drug target, CD-HIT, essential genes, KEGG, cluster database, candidiasis

INTRODUCTION

Among the *Candida* species afflicting humans, *C. albicans* and *C. tropicalis* are by far the most common. *Candida* is present as a harmless commensal in many different parts of the body including skin and is carried by almost half of the population. However, changes in the host environment may lead to opportunistic infections caused by *Candida* particularly oral, genital and gastrointestinal infections (Bodey, 1993). Infection caused by *C. albicans* and *C. tropicalis* can be broadly divided into two categories: Superficial mucocutaneous infections (Ashman and Papadimitriou, 1990) and systematic infections involving the spread of *C. albicans* to the blood stream (candidemia) and to other major organs (Bodey and Anaissie, 1989). Superficial infections affect various mucous membranes such as in oral and vaginal thrush. Approximately 75% of all women experience a clinically significant episode of vulvovaginal candidiasis (VVC) at least once during the reproductive period (Odds, 1988). It is believed that the illness is due to minor changes in epithelial environment such as pH, altered glucose/glycogen concentration or changes in epithelial integrity. *Candida tropicalis* is considered to be one of the leading *Candida* species, next only to *C. albicans,* to cause fungemia in patients with cancer (Bodey and Anaissie, 1989; Wingard, 1995). During the 1970s and 1980s, several studies reported that *C. tropicalis* fungemia, was common in patients with leukemia and in those who had received bone marrow transplants (Abi-Said *et al*., 1997; Ashman and Papadimitriou, 1990). Introduction of fluconazole, an effective triazole, in the early 1990s, altered the frequency and distribution of *Candida* species in patients with cancer (Abi-Said *et al*., 1997).

Classical laboratory experiments to identify candidate drug targets are labour intensive, time consuming and expensive. Genomics and bioinformatics methodologies offer alternative approaches and provide new insights in finding alternate drug targets to combat antibiotic resistance in pathogens at a reduced cost. Subtractive genomics deal with the utilization of the whole proteome of host and pathogen to identify proteins exclusively present in the pathogen by deducing the homologous proteins (Abi-Said *et al*., 1997). This process has been successfully used to identify novel drug targets of bacterial pathogens such as *Pseudomonas aeruginosa* (Dutta *et al*., 2006), *Helicobacter pylori* (Perumal *et al*., 2007) and *Listeria monocytogenes* (Hossain *et al*., 2013).

The present study aimed at finding new drug targets for *C. tropicalis* using a subtractive proteomics approach. This study employed the use of bioinformatics tools and databases including those mentioned above as well as Basic Local Alignment Search Tool for Proteins (BLASTP), Database of Essential Genes (DEG) and KEGG Automatic Annotation Server (KAAS) to identify, characterize and analyze the essential genes of *C. tropicalis*. Collectively, these tools of computational biology significantly enhanced the speed and efficiency of the current investigation, allowing us to pursue antimicrobial discovery that may have significant clinical utility.

MATERIALS AND METHODS

The complete proteome of *C. tropicalis* was retrieved from Uni-Prot (http://www.uniprot.org/). The proteins were subjected to CD-HIT analysis (http://weizhong-lab.ucsd.edu/cdhit_ suite/cgi-bin/index.cgi?cmd=cd-hit) (Huang *et al*., 2010). The program takes protein sequence in FASTA format as input and delivers a set of non-redundant, descriptive sequences as output. The procedure was carried out with a sequence identity cut-off of 0.6, thus eradicate redundant sequences of more than 60% identity (Sarangi *et al*., 2009). The consequential proteins were grouped. BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis was carried out for the non-redundant proteome against the proteome of human genome tax id: 9606. Proteins with an E-value (expectation value) of 10^{-4} were eliminated, assuming that they have a certain level of homology with the host genome (Zhang *et al*., 2004). The homologous proteins were eliminated and non-homologous proteins were separated for further analysis.

Further the non-homologous proteins were analysed using DEG (http://tubic.tju.edu.cn/deg/), which includes all the essential genes currently available (Zhang *et al*., 2004). A random E-value was kept at 10^{-4} and the minimum bit-score cut-off value of 100, BLOSUM62 matrix and gapped alignment mode were selected to identify the essential genes (Rathi *et al*., 2009). These selected proteins of *C. tropicalis* could be considered as drug targets, because they are not present in the host but only in *Candida*. Further analysis was carried out on these essential proteins to determine their sub-cellular localization and functions. The KAAS server of Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/tools/kaas/) was used to determine the functions.

The KAAS server provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways (Moriya *et al*., 2007). KEGG pathway studies were also conducted to analyze the occurrence of alternate pathways, following which the proteins were selected as potential targets. This enabled us to predict the function of the proteins and genome annotation that resulted in the identification of potential targets. Screening of the potential drug targets was carried out by similarity search using protein sequence of all the potential targets against the Drug Bank (Knox *et al*., 2011), TTD (Chen *et al*., 2002), PDTD (Gao *et al*., 2008) and HIT (Ye *et al*., 2011) to reach the novel drug targets. Further the outer membrane proteins were

predicted using Trans-Membrane prediction using Hidden Markov Models (TMHMM), that identify surface membrane proteins which could be used as potential drug targets and vaccine candidates (Krogh *et al*., 2001). TMHMM is a program for predicting transmembrane helices based on a hidden Markov model, it reads a FASTA format protein sequence and predicts the locations of trans-membrane, intracellular and extracellular locality.

RESULTS AND DISCUSSION

The pharmaceutical industry is under constant pressure for discovering new antimicrobial drugs due to the ever increasing development of drug resistance among the microorganisms that cause disease. Identification of microbe-specific proteins through drug discovery approaches and designing of new drugs aimed at known targets are the two popular means to combat drug resistance. The objective of this study was to identify proteins which are potentially useful as drug targets using genomic data and a subtractive genomic approach.

The proteome of *C. tropicalis* has a total of 6429 proteins. Of these, 310 are cluster proteins. Cluster proteins, proteins below 100aa sequence and proteins with redundant sequence, as analyzed by the CD-HIT program at 60% identity, were not considered for further evaluation to avoid redundancy. This resulted in 5623 candidate proteins for further study. BLASTP analysis of these proteins against the human genome revealed 5373 proteins that were non-homologous with the host genome (Fig. 1). A specific homology between the host and pathogen protein chosen as drug targets might lead to redundant cross-reactions and cytotoxicity. Hence, only non-homologous proteins were selected for further identification of essential genes.

Essential genes are those that are crucial to support cellular life. Because most antimicrobials target vital cellular processes in pathogenic microorganisms, essential gene products of microbes are promising new targets for anti-microbial drugs (Zhang *et al*., 2004). Essential genes unique to an organism can be measured as species-specific drug targets (Judson and Mekalanos, 2000). The Database of Essential Genes (DEG), hosts records of currently available essential genes among a wide range of organisms. By using the DEG, 73 essential proteins in *Candida* species were identified (Fig. 1). All the 73 proteins were analyzed using the KAAS server. Detailed pathway analysis revealed the participation of 64 proteins out of 73 (Table 1). Among the 64 proteins, 37 proteins were found to be crucial for the survival of the organism and participate in 30 different pathways (Table 2).

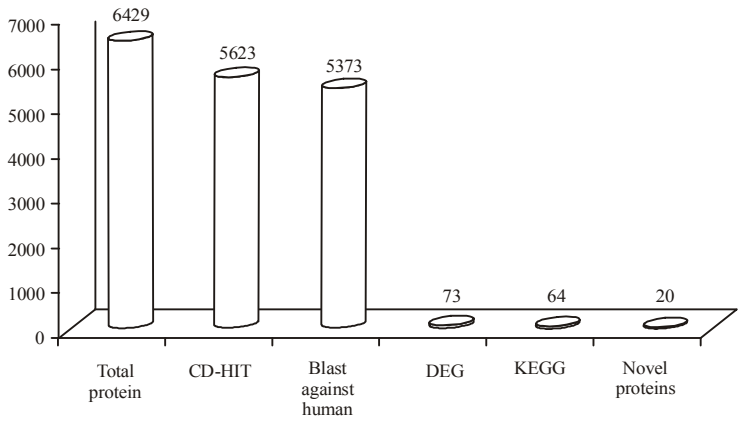

Fig. 1: Identification of drug targets in the proteome of *Candida tropicalis*

Table 1: List of proteins involved in pathways

UniProt accession No.	KEGG orthology Id
tr|C5M9J9|C5M9J9_CANTT	
tr|C5MHA9|C5MHA9_CANTT	K02543
tr|C5MCD9|C5MCD9_CANTT	
tr|C5M7X3|C5M7X3_CANTT	K10706
tr|C5MGN7|C5MGN7_CANTT	K00668
tr|C5MAM4|C5MAM4_CANTT	K00888
tr|C5M7X9|C5M7X9_CANTT	K00667
tr|C5MC92|C5MC92_CANTT	K14550
tr|C5M6C1|C5M6C1_CANTT	K14792
tr|C5MGN3|C5MGN3_CANTT	K01768
tr|C5MFY0|C5MFY0_CANTT	K11292
tr|C5M9D7|C5M9D7_CANTT	K10742
tr|C5M9N5|C5M9N5_CANTT	K14772
tr|C5M261|C5M261_CANTT	K14401
tr|C5M3U1|C5M3U1_CANTT	K06636
tr|C5MFN9|C5MFN9_CANTT	K14794
tr|C5MCS3|C5MCS3_CANTT	K14569
tr|C5M6W6|C5M6W6_CANTT	K15192
tr|C5M6B3|C5M6B3_CANTT	K14544
tr|C5M754|C5M754_CANTT	K06677
tr|C5MCQ5|C5MCQ5_CANTT	
tr|C5M8F6|C5M8F6_CANTT	K03235
tr|C5MGA7|C5MGA7_CANTT	K00698
tr|C5M6G0|C5M6G0_CANTT	K05288
tr|C5M9Y5|C5M9Y5_CANTT	K14832
tr|C5MCM2|C5MCM2_CANTT	
tr|C5MAT5|C5MAT5_CANTT	K14556
tr|C5MGU6|C5MGU6_CANTT	K00888
tr|C5M8J3|C5M8J3_CANTT	K15436
tr|C5MJJ5|C5MJJ5_CANTT	K00888
tr|C5M6B1|C5M6B1_CANTT	K14554
tr|C5MEJ5|C5MEJ5_CANTT	
tr|C5MAY4|C5MAY4_CANTT	K14007
tr|C5M800|C5M800_CANTT	K14808
tr|C5M566|C5M566_CANTT	
tr|C5MI29|C5MI29_CANTT	K01535
tr|C5MF02|C5MF02_CANTT	K14293
tr|C5MCI4|C5MCI4_CANTT	
tr|C5MA72|C5MA72_CANTT	K14787
tr|C5MB39|C5MB39_CANTT	K02999
tr|C5M998|C5M998_CANTT	K14781
tr|C5MFA5|C5MFA5_CANTT	K14824
tr|C5MG47|C5MG47_CANTT	K01886
tr|C5M2X1|C5M2X1_CANTT	K03130
tr|C5MIV7|C5MIV7_CANTT	K14555
tr|C5M353|C5M353_CANTT	K14776
tr|C5MIH7|C5MIH7_CANTT	
tr|C5MAJ2|C5MAJ2_CANTT	K14799
tr|C5M3F6|C5M3F6_CANTT	K14856
tr|C5M8U5|C5M8U5_CANTT	K15449
tr|C5MHD0|C5MHD0_CANTT	K03240
tr|C5M6B4|C5M6B4_CANTT	K01852
tr|C5M3D6|C5M3D6_CANTT	K14833
tr|C5MGI6|C5MGI6_CANTT	K14788
tr|C5M947|C5M947_CANTT	K14539
tr|C5MB89|C5MB89_CANTT	K08286
tr|C5MFE8|C5MFE8_CANTT	K13126
tr|C5M2U2|C5M2U2_CANTT	K14835
tr|C5M914|C5M914_CANTT	K14810
tr|C5M7P5|C5M7P5_CANTT	K14843
tr|C5MJA4|C5MJA4_CANTT	K01687

Table 1: Continue

UniProt accession No.	KEGG orthology Id
tr|C5MH04|C5MH04_CANTT	K14655
tr|C5MC80|C5MC80_CANTT	K00654
tr|C5MDZ3|C5MDZ3_CANTT	K09500
tr|C5MBE5|C5MBE5_CANTT	K14768
tr|C5M9D6|C5M9D6_CANTT	K14565
tr|C5M9P2|C5M9P2_CANTT	K14855
tr|C5MEB9|C5MEB9_CANTT	K06874
tr|C5MIH3|C5MIH3_CANTT	K14564
tr|C5M378|C5M378_CANTT	K00963
tr|C5M834|C5M834_CANTT	K01875
tr|C5MCC5|C5MCC5_CANTT	
tr|C5MBS9|C5MBS9_CANTT	K01867

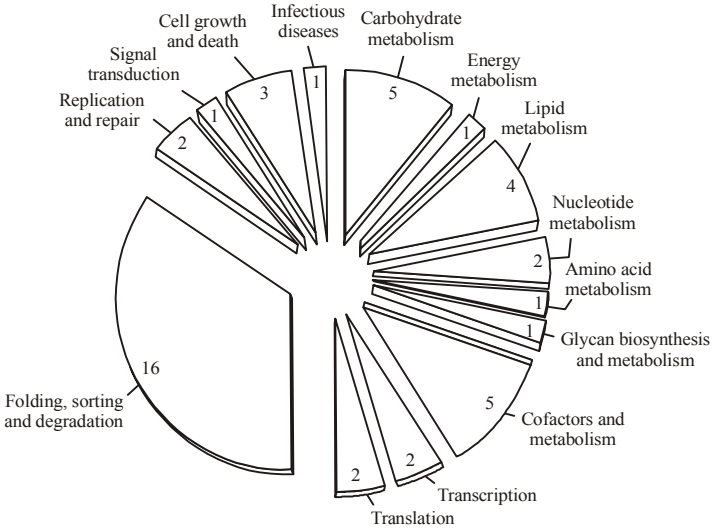

Fig. 2: Distribution of novel drug targets involved in various metabolism and cellular process

Finally, screening of drug targets was carried out using Drug Bank, Therapeutic Target Database (TTD) and Herbal Ingredients' Targets (HIT) for the 73 essential proteins identified using DEG. When analyzed by Drug Bank, among the 73 proteins, 12 were identified as approved targets and 15 are being analyzed experimentally as targets. Among the remaining 46 proteins, 6 were identified as successful targets, 6 were identified as clinical trial targets and 14 proteins were identified as research targets when analyzed using TTD. The remaining 20 are hitherto unreported novel proteins which were identified in this study as new potential drug targets.

The pathway analyses indicate that among the 37 essential proteins of *C. tropicalis*, five proteins are involved in carbohydrate metabolism, one protein is involved in energy metabolism, four proteins are involved in lipid metabolism, two proteins are involved in nucleotide metabolism, one protein is involved in amino acid metabolism, one protein is involved in glycan biosynthesis and metabolism and five proteins are involved in metabolism of cofactors and vitamins. Moreover, two proteins were found to be involved in transcription, two proteins were involved in translation, sixteen proteins were involved in folding, sorting and degradation, two proteins are involved in replication and repair; In environmental processing one protein is involved in signal transduction, In cellular process three proteins are involved in cell growth and death and one in infectious diseases. Some of the proteins are involved in multiple pathway groups (Fig. 2).

Table 2: List of drug targets involved in various metabolic pathways and other cellular activities

Metabolism	KEGG orthology and protein ID	Pathways	Genes and proteins	EC No.
Carbohydrate metabolism	K00963 (C5M378)	Pentose and glucuronate interconversions [PATH:ko00040] Galactose metabolism [PATH:ko00052] Starch and sucrose metabolism [PATH:ko00500]	UGP2, galU, galF; UTP--glucose-1 -phosphate uridylyltransferase	2.7.7.9
	K00698 (C5MGA7)	Amino sugar and nucleotide sugar metabolism	CHS1, chitin synthase	2.4.1.16
	K00963 (C5M378)	[PATH:ko00520]	UGP2, galU, galF, UTP--glucose-1-phosphate uridylyltransferase	2.7.7.9
	K00888 (C5MAM4, C5MGU6, C5MJJ5)	Inositol phosphate metabolism [PATH:ko00562]	PI4K, phosphatidylinositol 4-kinase	2.7.1.67
Energy metabolism	K01535 (C5MI29)	Oxidative phosphorylation [PATH:ko00190]	H+-transporting ATPase	3.6.3.6
Lipid metabolism	K00668 (C5MGN7)	Fatty acid biosynthesis [PATH:ko00061]	FAS1, fatty acid synthase subunit beta, fungi type	2.3.1.86
	K00667 (C5M7X9)		FAS2, fatty acid synthase subunit beta, fungi type	2.3.1.86
	K01852 (C5M6B4)	Steroid biosynthesis [PATH:ko00100]	LSS, ERG7, lanosterol synthase	5.4.99.7
	K00654 (C5MC80)	Sphingolipid metabolism [PATH:ko00600]	serine palmitoyltransferase	2.3.1.50
Nucleotide metabolism	K02999 (C5MB39)	Purine metabolism [PATH:ko00230]	RPA1, POLR1A, DNA-directed RNA polymerase I subunit	2.7.7.6
	K01768 (C5MGN3)		RPA1 adenylate cyclise	4.6.1.1
	K02999 (C5MB39)	Pyrimidine metabolism [PATH:ko00240]	RPA1, POLR1A, DNA-directed RNA polymerase I subunit RPA1	2.7.7.6
Amino acid metabolism	K01687 (C5MJA4)	Valine, leucine and isoleucine biosynthesis [PATH:ko00290]	ilvD, dihydroxy-acid dehydratase	4.2.1.9
Glycan biosynthesis and metabolism	K05288 (C5M6G0)	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis [PATH:ko00563]	PIGO, phosphatidylinositol glycan, class O	
Metabolism of cofactors and vitamins	K14655 (C5MH04)	Riboflavin metabolism [PATH:ko00740]	RIB2, PUS8, tRNA pseudouridine synthase 8/2,5-diamino-6-(5-phospho-D-ribitylamino)-pyrimidin-4(3H)-one deaminase	5.4.99.-
	K01687 (C5MJA4)	Pantothenate and CoA biosynthesis [PATH:ko00770]	ilvD, dihydroxy-acid dehydratase	4.2.1.9
Transcription translation	K02999 (C5MB39)	RNA polymerase [PATH:ko03020]	RPA1, POLR1A; DNA-directed RNA polymerase I subunit RPA1	2.7.7.6
	K03130 (C5M2X1)	Basal transcription factors [PATH:ko03022]	TAF5; transcription initiation factor TFIID subunit 5	
	K01886 (C5MG47)	Aminoacyl-tRNA biosynthesis [PATH:ko00970]	QARS, glnS, glutaminyl-tRNA synthetase	6.1.1.18
	K01875 (C5M834)		SARS, serS, seryl-tRNA synthetase	6.1.1.11
	K01867 (C5MBS9)		WARS, trpS, tryptophanyl-tRNA synthetase	6.1.1.2
	K14293 (C5MF02)	RNA transport [PATH:ko03013]	KPNB1, importin subunit beta-1	
	K03240 (C5MHD0)		EIF2B5, translation initiation factor eIF-2B subunit epsilon	
	K13126 (C5MFE8)		PABPC, polyadenylate-binding protein	

Table 2: Continue

Metabolism	KEGG orthology and protein ID	Pathways	Genes and proteins	EC No.
	K14401 (C5M261)	mRNA surveillance pathway [PATH:ko03015]	CPSF1, CFT1, cleavage and polyadenylation specificity factor subunit 1	
	K13126 (C5MFE8)		PABPC; polyadenylate-binding protein	
	K14544 (C5M6B3)	Ribosome biogenesis in eukaryotes PATH:ko03008]	UTP22, NOL6, U3 small nucleolar RNA-associated protein 22	
	K14550 (C5MC92)		UTP10, HEATR1, U3 small nucleolar RNA-associated protein 10	
	K14554 (C5M6B1)		UTP21, WDR36, U3 small nucleolar RNA-associated protein 21	
	K14555 (C5MIV7)		UTP13, TBL3, U3 small nucleolar RNA-associated protein 13	
	K14556 (C5MAT5)		DIP2, UTP12, WDR3, U3 small nucleolar RNA-associated protein 12	
	K14564 (C5MIH3)		NOP56, nucleolar protein 56	
	K14565 (C5M9D6)		NOP58, nucleolar protein 58	
	K14569 (C5MCS3)		BMS1, ribosome biogenesis protein BMS1	
	K14539 (C5M947)		LSG1, large subunit GTPase 1	3.6.1.-
Folding, sorting and degradation	K14007 (C5MAY4)	Protein processing in endoplasmic reticulum [PATH:ko04141]	SEC24, protein transport protein SEC24	
	K13126 (C5MFE8)	RNA degradation [PATH:ko03018]	PABPC, polyadenylate-binding protein	3.6.4.12
Replication and repair	K10742 (C5M9D7)	DNA replication [PATH:ko03030]	DNA2, DNA replication ATP-dependent helicase Dna2	
Signal transduction	K00888 (C5MAM4, C5MGU6, C5MJJ5)	Phosphatidylinositol signaling system [PATH:ko04070]	PI4K, phosphatidylinositol 4-kinase	2.7.1.67
Cell growth and death	K06636 (C5M3U1)	Cell cycle [PATH:ko04110]	SMC1, structural maintenance of chromosome 1	
	K02543 (C5MHA9)	Cell cycle - yeast [PATH:ko04111]	MEC1, cell cycle checkpoint protein MEC1	
	K06636 (C5M3U1)		SMC1, structural maintenance of chromosome 1	
	K06677 (C5M754)		YCS4, CNAP1, CAPD2; condensin complex subunit 1	
	K01768 (C5MGN3)	Meiosis - yeast [PATH:ko04113]	adenylate cyclase	4.6.1.1
	K02543 (C5MHA9)		MEC1, cell cycle checkpoint protein MEC1	
	K06636 (C5M3U1)		SMC1, structural maintenance of chromosome 1	
	K06636 (C5M3U1)	Oocyte meiosis [PATH:ko04114]	SMC1, structural maintenance of chromosome 1	
Infectious diseases	K03130 (C5M2X1)	Herpes simplex infection [PATH:ko05168]	TAF5, transcription initiation factor TFIID subunit 5	

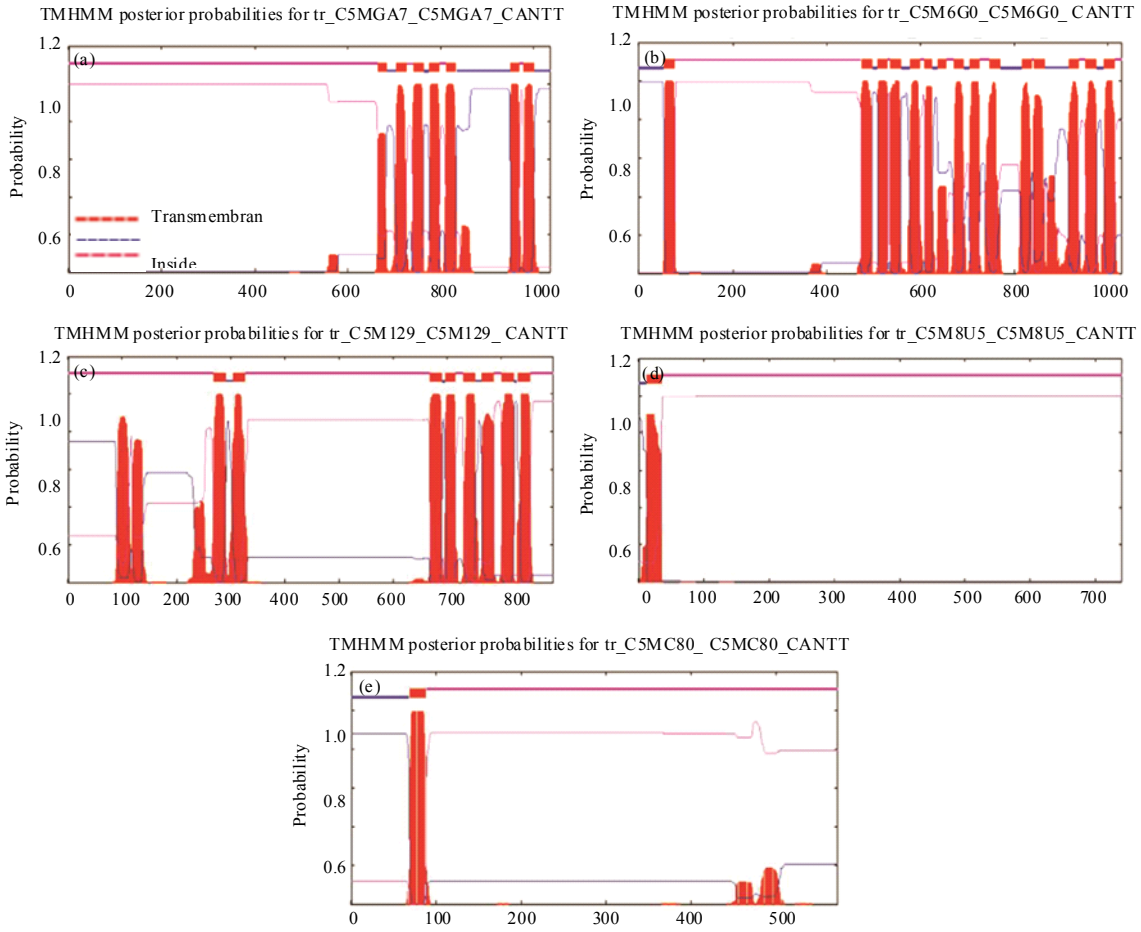

Fig. 3(a-e): Transmembrane analysis of potential drug targets using TMHMM, (a) A newly identified potential drug target that shows 7 transmembrane helix, (b) Drug target showing 15 transmembrane helix, (c) Drug target showing 8 transmembrane helix, (d) 1 transmembrane helix and (e) 1 transmembrane helix

Out of the 73 essential proteins, 34 were found to be uncharacterized. The functional classification of these 34 uncharacterized essential proteins was performed using the SVM-Prot Web server (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi) for transmembrane and functional identification. SVM-Prot Web server classified these 34 proteins as zinc-binding proteins (11), transmembrane proteins (6), metal binding proteins (3), DNA-binding proteins (3), transferases (4), hydrolases (1), motor protein (1), actin binding (1), mRNA splicing (1), RNA binding (1) mRNA binding (1) and unknown (1) (Table 3). Further, the Tied Mixture Hidden Markov Model: (TMHMM) tool was used for specific transmembrane identification. The TMHMM prediction analysis identified 5 proteins as transmembrane proteins (Fig. 3). Among these 34 uncharacterized 12 proteins are identified as novel drug targets that include Zinc binding proteins (5), transmembrane proteins (2), DNA binding proteins (2) and others (3) (Table 4). All the twelve proteins are as same as identified as novel targets in earlier analysis using drug bank, TTD and HIT. The remaining 8 proteins that were identified as novel proteins earlier were found to be involved in ribosome biogenesis (C5MCS3), TATA binding protein (C5M6W6), condensing complex

Table 3: Functional identification of the uncharacterized essential proteins of *Candida tropicalis* using SVM-Prot web server

UniProt accession No.	Functions
C5M9J9|C5M9J9_CANTT	Metal binding
C5MHA9|C5MHA9_CANTT	Transferases
C5MCD9|C5MCD9_CANTT	Motor protein
C5M7X3|C5M7X3_CANTT	Transferases
C5MAM4|C5MAM4_CANTT	Transferases
C5MC92|C5MC92_CANTT	Transmembrane
C5M6C1|C5M6C1_CANTT	DNA binding
C5MGN3|C5MGN3_CANTT	Zinc binding
C5MFY0|C5MFY0_CANTT	Zinc binding
C5M9D7|C5M9D7_CANTT	Zinc binding
C5M9N5|C5M9N5_CANTT	Transferases
C5M261|C5M261_CANTT	Transmembrane
C5MFN9|C5MFN9_CANTT	Zinc binding
C5M6B3|C5M6B3_CANTT	Zinc binding
C5MCQ5|C5MCQ5_CANTT	Metal binding
C5M6G0|C5M6G0_CANTT	Transmembrane
C5M9Y5|C5M9Y5_CANTT	DNA binding
C5MCM2|C5MCM2_CANTT	Transmembrane
C5M8J3|C5M8J3_CANTT	Transmembrane
C5M6B1|C5M6B1_CANTT	Actin binding
C5MAY4|C5MAY4_CANTT	Zinc binding
C5M800|C5M800_CANTT	DNA binding
C5MF02|C5MF02_CANTT	Transmembrane
C5M998|C5M998_CANTT	mRNA slicing
C5MIV7|C5MIV7_CANTT	Zinc binding
C5M353|C5M353_CANTT	RNA binding
C5M3F6|C5M3F6_CANTT	Zinc binding
C5M8U5|C5M8U5_CANTT	Hydrolases
C5MGI6|C5MGI6_CANTT	mRNA-binding
C5M947|C5M947_CANTT	Metal binding
C5M914|C5M914_CANTT	Zinc binding
C5M7P5|C5M7P5_CANTT	Zinc binding
C5M9P2|C5M9P2_CANTT	Zinc binding
C5M566| C5M566_CANTT	Unknown

Table 4: Target identification of uncharacterized essential proteins of *Candida tropicalis* using drug bank, TTD and HIT

Name of the protein	UniProt accession No.	Known targets
Uncharacterized protein	C5MC92	
Uncharacterized protein	C5M6C1	
Uncharacterized protein	C5MFY0	
Uncharacterized protein	C5MFN9	
Uncharacterized protein	C5M9Y5	
Uncharacterized protein	C5M8J3	
Uncharacterized protein	C5MAY4	
Uncharacterized protein	C5M566	
Uncharacterized protein	C5M998	
Uncharacterized protein	C5M3F6	
Uncharacterized protein	C5M8U5	
Uncharacterized protein	C5M7P5	
Uncharacterized protein	C5M9J9	Successful target
Uncharacterized protein	C5M800	Experimental
Uncharacterized protein	C5M353	Experimental
Uncharacterized protein	C5M914	Experimental
Uncharacterized protein	C5MHA9	Successful target
Uncharacterized protein	C5MCD9	Research target
Uncharacterized protein	C5M7X3	Research target
Uncharacterized protein	C5MAM4	Research and clinical trial target
Uncharacterized protein	C5MGN3	Successful, clinical and research target
Uncharacterized protein	C5M9D7	Research target
Uncharacterized protein	C5M9N5	Research target
Uncharacterized protein	C5M261	Research target

Table 4: Continue

Name of the protein	UniProt accession No.	Known targets
Uncharacterized protein	C5M6B3	Successful target
Uncharacterized protein	C5MCQ5	Research target
Uncharacterized protein	C5M6G0	Successful and research target
Uncharacterized protein	C5MCM2	Research target
Uncharacterized protein	C5M6B1	Successful and research target
Uncharacterized protein	C5MF02	Successful target
Uncharacterized protein	C5MIV7	Clinical trial and research target
Uncharacterized protein	C5MGI6	Clinical trial target
Uncharacterized protein	C5M947	Clinical trial and research target
Uncharacterized protein	C5M9P2	Clinical trial and research target

subunit (C5M754), predicted protein (C5MCI4), glutaminyl t-RNA synthetase (C5MG47), nucleolar complex protein-2 (C5M3D6), di-hydroxy acid dehydratase (C5MJA4) and nucleolar protein NOP-58 (C5M9D6).

Comparative analysis of the metabolic pathways of the pathogen *C. tropicalis* and the human host revealed 30 pathways that are unique to the pathogen. Further investigations on the predicted genes will be needed to verify the reliability of the data. The Outer Membrane (OM) proteins of *Candida* play an important role in the interaction of the organism with the host and in subsequent pathogenicity; the OM proteins play a role in adherence, uptake of nutrients from the host and in countering host defense mechanisms (Seltmann and Holst, 2002). They could be protective antigens because the components of the outer membrane are easily recognized as foreign substances by immunological defense systems of hosts and therefore potential vaccine targets as well.

CONCLUSION

To our knowledge this is the first report on the application of subtractive genome analysis of any fungal pathogen and its potential host. Using this approach, we found that among the total of 6429 proteins of *C. tropicalis,* 64 are essential proteins for survival. Further analysis identified 20 proteins, which were hitherto unreported as putative novel drug targets. This study, also identified outer membrane proteins that could be studied further as drug targets. Virtual screening of candidate drugs against these target proteins might be useful in the discovery of novel therapeutic compounds against the fungal pathogen *C. tropicalis*.

REFERENCES

Abi-Said, D., E. Anaissie, O. Uzun, I. Raad, H. Pinzcowski and S. Vartivarian, 1997. The epidemiology of hematogenous candidiasis caused by different *Candida* species. Clin. Infect. Dis., 24: 1122-1128.

Ashman, R.B. and J.M. Papadimitriou, 1990. What's new in the mechanisms of host resistance to *Candida albicans* infection? Pathol. Res. Pract., 186: 527-534.

Bodey, G.P. and E.J. Anaissie, 1989. Chronic systemic candidiasis. Eur. J. Clin. Microbiol. Infect. Dis., 8: 855-857.

Bodey, G.P., 1993. Hematogenous and Major Organ Candidiasis. In: *Candidiasis*: Pathogenesis, Diagnosis and Treatment, Bodey, G.P. (Ed.). 2nd Edn., Raven Press, New York, USA., ISBN-13: 9780881679540, pp: 279-329.

Chen, X., Z.L. Ji and Y.Z. Chen, 2002. TTD: Therapeutic target database. Nucleic. Acids. Res., 30: 412-415.

Dutta, A., S.K. Singh, P. Ghosh, R. Mukherjee, S. Mitter and D. Bandyopadhyay, 2006. *In silico* identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. *In Silico* Biol., 6: 43-47.

Gao, Z., H. Li, H. Zhang, X. Liu and L. Kang *et al*., 2008. PDTD: A web-accessible protein database for drug target identification. BMC Bioinform., Vol. 9. 10.1186/1471-2105-9-104

Hossain, M., A.T.M.J. Mosnaz, A.M. Sajib, P.K. Roy, S.K. Shakil, S.M.S. Ullah and S.H. Prodhan, 2013. Identification of putative drug targets of *Listeria monocytogenes* F2365 by subtractive genomics approach. J. BioSci. Biotechnol., 2: 63-71.

Huang, Y., B. Niu, Y. Gao, L. Fu and W. Li, 2010. CD-HIT suite: A web server for clustering and comparing biological sequences. Bioinformatics, 26: 680-682.

Judson, N. and J.J. Mekalanos, 2000. TnAraOut: A transposon-based approach to identify and characterize essential bacterial genes. Nat. Biotechnol., 18: 740-745.

Knox, C., V. Law, T. Jewison, P. Liu and S. Ly *et al*., 2011. DrugBank 3.0: A comprehensive resource for omics research on drugs. Nucleic Acids Res., 39: D1035-D1041.

Krogh, A., B. Larsson, G. von Heijne and E.L.L. Sonnhammer, 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol., 305: 567-580.

Moriya, Y., M. Itoh, S. Okuda, A.C. Yoshizawa and M. Kanehisa, 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res., 35: W182-W185.

Odds, F.C., 1988. Candidosis of the Genitalia. In: *Candida* and Candidosis, Odds, F.C. (Ed.). 2nd Edn., Elsevier Science Health Science Division, London, UK., ISBN-13: 9780702012655, pp: 124-135.

Perumal, D., C.S. Lim, K.R. Sakharkar and M.K. Sakharkar, 2007. Differential genome analyses of metabolic enzymes in *Pseudomonas aeruginosa* for drug target identification. Silico Biol., 7: 453-465.

Rathi, B., A.N. Sarangi and N. Trivedi, 2009. Genome subtraction for novel target definition in *Salmonella typhi*. Bioinformation, 4: 143-150.

Sarangi, A.N., R. Aggarwal, Q. Rahman and N. Trivedi, 2009. Subtractive genomics approach for *in silico* identification and characterization of novel drug targets in *Neisseria meningitides serogroup* B. J. Comput. Sci. Syst. Biol., 2: 255-258.

Seltmann, G. and O. Holst, 2002. The Bacterial Cell Wall. Springer Science and Business Media, Berlin, Heidelberg, ISBN-13: 9783540426080, Pages: 280.

Wingard, J.R., 1995. Importance of *Candida* species other than *Candida albicans* as pathogens in oncology patients. Clin. Infect. Dis., 20: 115-125.

Ye, H., L. Ye, H. Kang, D. Zhang and L. Tao *et al*., 2011. HIT: Linking herbal active ingredients to targets. Nucleic Acids Res., 39: D1055-D1059.

Zhang, R., H.Y. Ou and C.T. Zhang, 2004. DEG: A database of essential genes. Nucleic Acids Res., 32: D271-D272.