Journal of Computational Intelligence in Bioinformatics
ISSN: 0973-385X Volume 4 Number 2 (2011) pp. 135-148

http://www.ripublication.com/jcib.htm

Subtractive Proteomic Approach for in Silico Identification and Characterization of Novel Drug Targets in *Leishmania major.*

[*1]Randhawa Vinay, [2]Satveer Kaur and [2]Nanda Ashina

[1]Department of Bioinformatics, ADI Biosolution,
Mohali, Punjab-160059, India.
**Corresponding Author E-mail: vinay.plp@gmail.com*
[2]Hans Raj Mahila Maha Vidyalya, Jalandhar, Punjab, India.

Abstract

The whole genome sequencing technology is providing expensive information for the identification of new therapeutic drug targets in pathogens. The availability of the complete proteome of Leishmania major has made it possible to carry out the *in silico* analysis of its proteome for identification of potential therapeutic targets. Our study revealed the essential proteins that are non-homologous to human genome and that are essential for the survival of the pathogen. The essentiality of the proteins of *L.major* was checked using the search against Database of Essential Genes (DEG). Subcellular localization prediction of these essential proteins was performed by CELLO server to find out proteins that lie on the outer membrane of the pathogen which could be potential drug targets. Metabolic pathway analysis of the identified essential surface proteins was performed using the KEGG Automated Annotation Server (KAAS) housed at KEGG pathways database. Comparative metabolic pathway analysis of *H.sapiens* and *L.major* was performed using KEGG database to find proteins involved in unique pathways of *L.major*. These essential proteins of *L.major* might be used as potential drug targets, as they belong to pathways present only in the pathogen and not in humans. Screening of the functional inhibitors against these novel targets may result in discovery of novel therapeutic compounds that can be effective against *L.major.* In present study subtractive genomic approach has been used to identify therapeutic target in *L.major*, which causes disfiguring cutaneous and muco-cutaneous lesions that can cause destruction of mucous membranes to visceral disease affecting the haemopoetic organs in humans.

Keywords: Homologs; Pathogen; Leishmania major; Paralogs; Subtractive genomics.

Abbreviations

DEG: Database of Essential Gene; KEGG: Kyoto Encyclopedia of Genes and Genomes; KAAS: KEGG Automated Annotation Server; PIR: Protein Information Resource.

Introduction

Leishmania are protozoal parasites of the family of Trypanosomatids responsible for significant worldwide human morbidity and mortality. The clinical features vary from mild cutaneous lesions to fatal mucocutaneous or visceral involvement depending on the causative species [1]. Leishmaniasis is endemic in >85 developing countries with >1.5 million estimated cases occurring each year and an additional 350 million people at risk of infection [2]. There has been a progressive expansion of leishmaniasis endemic regions as well as a concomitant increase in the total number of reported leishmaniasis cases, often in epidemic proportions (i.e., with 100,000–200,000 individuals infected) [3, 4, 5, 6]. The availability of genome-scale sequenced data in the past decade and the completion of the human genome project have revolutionized the field of drug-discovery against threatening human pathogens [7]. To date, sequence information from approximately 292 complete genomes has been deposited in various public databases and a number of bioinformatics tools have been developed to facilitate *in silico* analysis of the gene sequence information [8].

Comparative proteomics and bioinformatics provide new opportunities for finding novel drug targets among previously unexplored cellular functions based on an understanding of their related biological processes in pathogens and their hosts. In general, a target should provide adequate selectivity; yielding a drug which is specific or highly selective against the pathogen with respect to the human host.

A proteomic approach of exploiting the *L.major* and the *H.sapiens* proteomes has been used to find potential drug targets in *L.major*. The entire approach is built on two assumptions (i) the potential target protein must play an essential role in the survival of *L.major* and constitute a critical component in its metabolic pathway (ii) The target protein should not have any well conserved homolog protein in the *H.sapiens* host keeping the cytotoxicity issues into consideration. This approach for identification of potential targets is subtractive because we have used a subtraction dataset while comparing the two proteomes. The focus is on the complement of the proteome of the *L.major,* which is essential for it but is not present in human host. There are a number of approaches to locate essential genes in a given organism, some of which focus on the concept that essential genes tend to be evolutionarily conserved over species [9, 10, 11, 12]. This approach has been used to identify a set of essential proteins in *L.major*.

Most studies on essential genes were reported on experimental data acquired from gene disruption and systematic mutagenesis studies on the bacterium [13]. Gene disruption and transposon based mutagenesis are powerful tools to identify essential genes. Inactivation of essential genes results in the lethal phenotype in the bacteria [14]. These experimental approaches are laborious and technically demanding. Essential genes are those indispensable for the survival of an organism, and therefore

are considered a foundation of life. DEG (Database of Essential Genes) [15, 16] hosts records of currently available essential genes among a wide range of organisms. For eukaryotes, DEG contains essential genes those are present in yeast, humans, mice, worms, fruit flies, zebra fish and the plant *A. thaliana*. This resource facilitates drug target discovery by seeking non-human orthologous targets in the DEG. The potential of the available datasets and the application of the subtractive approach may be considered as an experimental approach for the identification of essential proteins candidates for drug discovery against leishmaniasis. Here, we present an *in silico* approach to identify drug target proteins that are essential to *L.major* based on subtractive proteomics, essentiality and comparative pathway analysis.

Methodology

Protein sequence retrieval of host and pathogen

Complete proteomes of *H.sapiens* (ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/protein/) and *L.major* (ftp://ftp.sanger.ac.uk/pub/databases/L.major_sequences/DATASETS/LmjFwholegenome_20070731V5.2.pep) were retrieved from NCBI (www.ncbi.nlm.nih.gov) and The Wellcome Trust Sanger Institute (www.sanger.ac.uk) respectively. From the entire proteomes, the shorter protein sequences of length less than 100 amino acids were filtered out assuming that proteins of length below 100 do not represent the products of essential genes.

Identification of paralogs

The *L.major* proteins were eliminated at 60% identity using CD-HIT suite [17] to identify the paralogs or duplicate proteins within the proteome of *L.major* to reduce the redundancy of the dataset in search of more significant results. The paralogs in the dataset were excluded and the remaining sets of protein were used for further analysis.

Similarity search

Standalone BlastP [18] was performed with the non-paralogous (proteome after removing paralogs) proteins against Homo sapiens protein sequences using threshold expectation value (E-value) 1e-10 as parameter. The human homologous protein sequences were excluded (with identity above 60%) and the list of non-human homologous proteins was compiled. The DEG Database identifies essential proteins on the basis of sequence homology. The selected non human-homologues proteins were then subjected to BlastP of Database of Essential Genes (DEG) (http://tubic.tju.edu.cu/deg1) at the expectation value parameter 1e-4 and a minimum bit-score cut-off of 100 were used to screen out proteins that appeared to represent essential proteins.

In silico sub cellular localization prediction of essential proteins

Protein sub cellular localization prediction involves the computational prediction of location of a protein in the cell. Knowledge of the sub cellular localization of a protein

can significantly improve target identification during the drug-discovery process [19, 20]. For example, secreted proteins and plasma membrane proteins are easily accessible by drug molecules because of their localization in the extra cellular space or on the cell surface. Important therapeutics have been created that target proteins present on the cell surface in a specific cell type or disease state [21]. Sub cellular localization of essential proteins of *L.major* was explored by two avenues: (a) annotation transfer from homologous sequences (b) method based on amino acid composition.

Sub cellular localization prediction by annotation transfer from homologous sequences

Sequence similarity is perhaps the most frequently used method to annotate function for unknown proteins and accounts for the majority of annotations about function in public databases [22, 23, 24]. DBSubLoc [25], protein sub cellular localization database was used to predict the sub cellular localization of *L.major* essential proteins. DBSubLoc contains proteins from primary protein database SWISS-PROT (http://www.expasy.ch/sprot/) and PIR (http://pir.georgetown.edu/). A threshold of 60% sequence identity was used for transferring functional annotations. A major limitation of sequence homology based methods is that they are only applicable when another sequence similar protein with experimentally known function is available. Hence, only a small fraction of sequences were annotated using this approach.

Sub cellular localization prediction by sequence coding schemes

High accuracy methods like those based on homology are plagued by the problem of low coverage and can provide annotations for less than one-third of known sequences. Currently the best solution available is to combine low-coverage methods with high-coverage methods, like those based on amino acid composition. CELLO v.2.5: Sub cellular Localization Predictive System [26, 27], a multi-class SVM classification server was used to predict the sub cellular location of essential proteins of *L.major* to identify the surface membrane proteins which could be probable drug targets. Four types of sequence coding schemes are used by CELLO server: (a) amino acid composition, (b) di-peptide composition, (c) partitioned amino acid composition and (d) sequence composition based on the physico-chemical properties of amino acids.

Classifying functions of uncharacterized essential proteins

The family classification of the putative uncharacterized essential proteins of *L.major* was performed by PANTHER 6 web server [28], a comprehensive software system for relating protein sequence evolution to the evolution of specific protein functions and biological roles.

Metabolic Pathway analysis

Metabolic pathway analysis of the essential proteins of *L. major* sorted by CELLO server was done by KAAS server [29] for the identifying the role of potential targets in the parasite's metabolism and thus depicting the metabolic essentiality of the proteins. KAAS (KEGG Automatic Annotation Server) provides functional

annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The unique proteins involved in metabolic pathways in *L.major* were analyzed by comparative metabolic pathway analysis of host and pathogen was performed by using Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Discussion

In silico subtractive analysis is a powerful approach for identifying proteins that might be used as potential drug targets. By this method, one searches for proteins/genes present in one organism and absent in another. In the present study, non-human homolog essential proteins of *L.major* were identified by applying subtractive genomic approach. Figure 1 shows the flow diagram of step by step approach used in the current study.

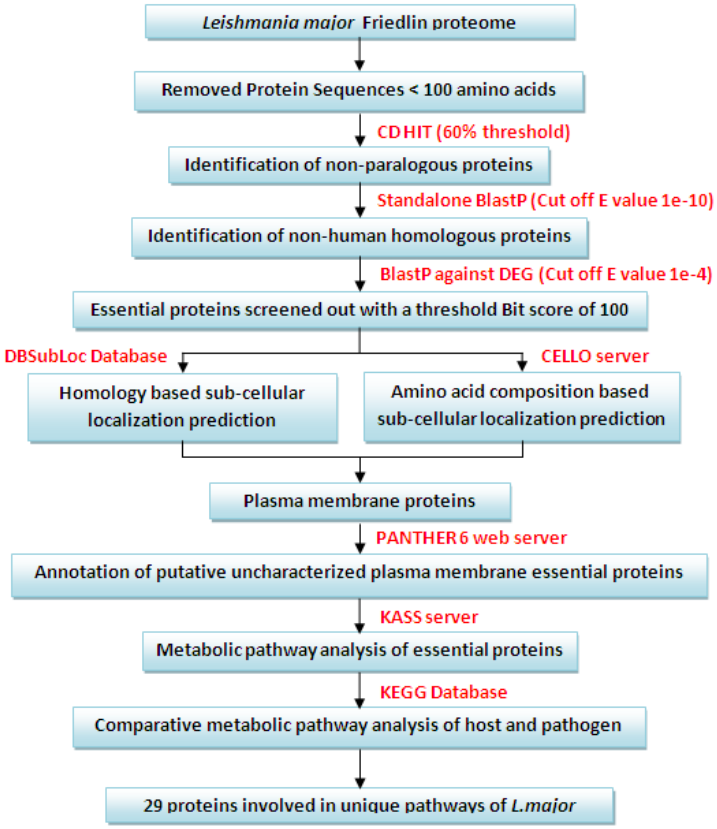

Figure 1: Flow chart representing the systematic identification and characterization of potential drug targets in *Leishmania major*.

A total of 8335 proteins sequences were calculated in *L. major* proteome. The protein sequences with length less than 100 aa were removed assuming that the

sequences with the length below 100 aa do not signify essential gene products. The proteome analysis of *L.major* revealed that out of 8335 proteins, 558 protein sequences were identified as duplicates through CD-HIT at 60% similarity. These duplicate proteins were removed to reduce the redundancy of the dataset. The remaining non paralogs were subjected to subtractive genomics which leads to 8262 proteins. These 8262 proteins of the pathogen were subjected for comparative proteome analysis against the *H.sapiens* proteome with standalone version of BlastP. A Bit score of 100 was used to filter out the non-human homologs. A set of 7590 proteins was obtained which was subjected to BlastP against DEG database showed 808 proteins that were essential for the pathogen.

The results of subtractive proteomic approach, sub cellular localization and metabolic pathway analysis are listed in Table 1. The purpose of the study was to locate those essential proteins of *L.major* that play vital roles in the normal functioning within the host and thus the resultant proteins are likely to represent potential drug targets. The sub cellular localization by CELLO server revealed 151 surface proteins, 20 out of the set were hypothetical/uncharacterized proteins. The functional classification of the uncharacterized essential proteins was done by PANTHER server. The 14 non- annotated proteins were not taken into consideration for the further process.

Metabolic pathway analysis of these essential proteins by KAAS server concluded that out of 137 (excluding non-annotated), 102 proteins might be invariably linked with essential metabolic and signal transduction pathways. Metabolic pathway analyses of the 102 essential proteins revealed that 10 proteins were involved in Carbohydrate Metabolism, 5 in Energy Metabolism, 12 in Lipid Metabolism, 2 in Nucleotide Metabolism, 7 in Amino Acid Metabolism, 3 in Glycan Biosynthesis and metabolism, 3 in Metabolism of Co-factors and Vitamins, 1 in Xenobiotics Biodegradation and Metabolism, 2 in transcription, 5 in translation, 13 in folding, sorting and degradation, 10 in replication and repair, 12 in membrane transport, 13 in signal transduction, 5 in signaling and interaction, 9 in endocrine system, 2 in nervous system, 4 in digestive system, 6 in cell growth and death, 9 in cellular transport and catabolism and 1 in cellular communication and 1 in cell mortality. The results are summarized in Table 2. Comparative metabolic pathway analysis of the host (*H.sapiens*) and the pathogen (*L.major*) was performed by using Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The results of comparative analysis revealed 29 proteins are involved in unique pathways of *L.major* summarized in Table 3. However these unique proteins involved in various metabolic pathways are essential for survival of pathogen hence these unique proteins might be used as potential targets for drug discovery.

In the whole study two approaches were used to identify the suitable drug targets: (a) substractive genomic approach and (b) accessing the essentiality of target proteins for *L. major*. This approach was already successfully used in many organisms such as Pseudomonas aeruginosa, Helicobacter pylori, Burkholderia pseudomallei, Mycobacterium tuberculosis H37Rv, Salmonella typhi and Neisseria meningitides serogroup B for drug target identification, which results constructive thoughts for further drug development.

Table 1: Substractive proteomic and metabolic pathway analysis result for *L.major*.

L.major	Number
Total Number of proteins	8335
Protein sequences less than 100 aa	73
Duplicates (>60% identical) in CD-HIT	558
Non-paralogs	7703
Non-human homologous proteins (E-value 1e-10)	7590
Essential protein in DEG (E-value 1e-3)	808
Surface membrane proteins	150
Functionally annotated proteins by PANTHER	6
Proteins involved in Metabolic pathways	102
Unique pathway proteins in L.major	29

Metabolism

Carbohydrate Metabolism					
SN	KO	Protein ID	Gene Name	Pathway	EC
1	K01193	LmjF27.2340	scrB,sacA	Galactose metabolism	EC:3.2.1.26
2	K00820	LmjF06.0950	GFPT2	Amino sugar and nucleotide sugar metabolism	EC:2.6.1.16
3	K01112	LmjF20.1480	MTMR,PHPT1	Fructose and mannose metabolism	E3.1.3.-
4	K00888	LmjF29.1450	PI4K	Inositol phosphate metabolism	EC:2.7.1.67
5	K00074	LmjF33.2600	paaH	Butanoate metabolism	EC:1.1.1.157
6	K00999	LmjF26.2480	CDIPT	Inositol phosphate metabolism	EC:2.7.8.11
7	K00914	LmjF24.2010	PIK3C3, VPS34	Inositol phosphate metabolism	EC:2.7.1.137
8	K01110	LmjF34.1430	PTEN	Inositol phosphate metabolism	EC:3.1.3.67
9	K00326	LmjF27.0730	CYB5R	Amino sugar and nucleotide sugar metabolism	EC:1.6.2.2
10	K01099	LmjF15.1210	INPP5B,SYNJ2, SYNJ1	Inositol phosphate metabolism	EC:3.1.3.36
Energy Metabolism					
1	K01760	LmjF32.2640, LmjF14.0460	METC,STR3	Nitrogen metabolism, Sulfur metabolism	EC:4.4.1.8
2	K00058	LmjF03.0030	PHGDH	Methane metabolism	EC:1.1.1.95
3	K01535	LmjF18.1510	AHA	Oxidative phosphorylation	EC:3.6.3.6
4	K01760	LmjF32.2640, LmjF14.0460	metC	Nitrogen metabolism, Sulfur metabolism	EC:4.4.1.8
5	K02154	LmjF23.1510, LmjF32.0920	ATPeVI, ATP6N1A	Oxidative phosphorylation	EC:3.6.3.14
Lipid Metabolism					
1	K10256	LmjF10.1320	FAD12	Biosynthesis of unsaturated fatty acids	EC:1.14.19.-
2	K00507	LmjF14.0510, LmjF24.2250	SCD	Biosynthesis of unsaturated fatty acids, Lipid biosynthesis proteins	EC:1.14.19.1
3	K08730	LmjF14.1200	PTDSS2	Glycerophospholipid metabolism	EC:2.7.8.-
4	K00223	LmjF33.0680	ERG4	Steroid biosynthesis	EC:1.3.1.71
5	K00507	LmjF14.0510, LmjF24.2250	SCD, desC	Biosynthesis of unsaturated fatty acids, Lipid biosynthesis proteins	EC:1.14.19.1
6	K00981	LmjF26.1620	CDS1, cdsA	Glycerophospholipid metabolism	EC:2.7.7.41

8	K09830	LmjF23.1300	ERG3	Steroid biosynthesis	EC:1.3.3.-
8	K00222	LmjF32.2320	TM7SF2, ERG24	Steroid biosynthesis	EC:1.3.1.70
9	K08745	LmjF24.1780	SLC27A1_4, FATP1, FATP4	Lipid biosynthesis proteins	EC:6.2.1.-
10	K01852	LmjF06.0650	LSS, ERG7	Steroid biosynthesis	EC:5.4.99.7
11	K00649	LmjF34.1090	GNPAT	Glycerophospholipid metabolism	EC:2.3.1.42
12	K00999	LmjF26.2480	CDIPT	Glycerophospholipid metabolism	EC:2.7.8.1
1.4 Nucleotide Metabolism					
1	K02335	LmjF34.1260	polA	Purine metabolism, Pyrimidine metabolism	EC:2.7.7.7
2	K01768	LmjF17.0200	ACXA, ACXB, ACXC, ACXD	Purine metabolism	EC:4.6.1.1
1.5 Amino Acid Metabolism					
1	K00147	LmjF02.0630	proA	Arginine and proline metabolism	EC:1.2.1.41
2	K00058	LmjF03.0030	PHGDH	Glycine, serine and threonine metabolism	EC:1.1.1.95
3	K01760	LmjF32.2640 LmjF14.0460	METC,STR3	Cysteine and methionine metabolism	EC:4.4.1.8
4	K00820	LmjF06.0950	GFPT2	Alanine, aspartate and glutamate metabolism	EC:2.6.1.16
5	K00815	LmjF36.2360	TAT	Cysteine and methionine metabolism, Tyrosine metabolism, Phenylalanine metabolism, Phenylalanine, tyrosine and tryptophan biosynthesis, Amino acid related enzymes	EC:2.6.1.5
Metabolism of other Amino Acids					
1	K01760	LmjF32.2640 LmjF14.0460	METC,STR3	Selenocompound metabolism	EC:4.4.1.8
2	K01256	LmjF29.2240	pepN	Glutathione metabolism	EC:3.4.11.2
Glycan Biosynthesis and Metabolism					
1	K07151	LmjF35.1150	STT3	Glycosyltransferases, N-Glycan biosynthesis, Various types of N-glycan biosynthesis	EC:2.4.1.119
2	K05546	LmjF36.4100	GANAB	N-Glycan biosynthesis	EC:3.2.1.84
3	K03434	LmjF09.0040	PIGL	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	EC:3.5.1.89
1.8 Metabolism of Cofactors and Vitamins					
1	K00815	LmjF36.2360	TAT	Ubiquinone and other terpenoid-quinone biosynthesis	EC:2.6.1.5
2	K00859	LmjF22.1530	coaE	Pantothenate and CoA biosynthesis	EC:2.7.1.24
3	K09680	LmjF28.0140	coaW	Pantothenate and CoA biosynthesis	EC:2.7.1.33
Xenobiotics Biodegradation and Metabolism					
1	K00074	LmjF33.2600	paaH	Benzoate degradation	EC:1.1.1.157

Genetic Information Processing

Transcription					
1	K12828	LmjF28.2570	SF3B1, SAP155	Transcription, Spliceosome	
2	K02202	LmjF26.0040	CDK7	Transcription; Basal transcription factors	EC:2.7.11.22
Translation					
1	K12618	LmjF06.0260	XRN1, SEP1,	Translation; Ribosome biogenesis in	EC:3.1.13.-

			KEM1	eukaryotes	
2	K14558	LmjF18.0830	PWP2, UTP1	Translation; Ribosome biogenesis in eukaryotes	
3	K14376	LmjF29.2600	PAP	Translation; mRNA surveillance pathway	EC:2.7.7.19
4	K11108	LmjF35.1700	RCL1	Translation; Ribosome biogenesis in eukaryotes	
5	K14554	LmjF36.4210	UTP21, WDR36	Translation; Ribosome biogenesis in eukaryotes	
Folding, Sorting and Degradation					
1	K07151	LmjF35.1150	STT3	Protein processing in endoplasmic reticulum	EC:2.4.1.119
2	K05546	LmjF36.4100	GANAB	Protein processing in endoplasmic reticulum	EC:3.2.1.84
3	K13280	LmjF08.0450	SEC11, sipW	Protein export	EC:3.4.-.-
4	K03869	LmjF36.3060	CUL3	Ubiquitin mediated proteolysis	
5	K03032	LmjF28.1730	PSMD1, RPN2	Proteasome	
6	K12618	LmjF06.0260	XRN1, SEP1, KEM1	RNA degradation	EC:3.1.13.-
7	K10976	LmjF16.1530	ERO1LB	Protein processing in endoplasmic reticulum	EC:1.8.4.-
8	K12581	LmjF22.1630	CNOT7_8, CAF1, POP2	RNA degradation	
9	K14442	LmjF22.1500	DHX36, RHAU	RNA degradation	EC:3.6.4.13
10	K10589	LmjF30.0910	UBE3C	Ubiquitin system	EC:6.3.2.19
11	K03355	LmjF12.0610	APC8, CDC23	Ubiquitin mediated proteolysis, Ubiquitin system	
12	K08337	LmjF07.0010	ATG7	Ubiquitin system	
13	K08860	LmjF34.2150	EIF2AK	Protein processing in endoplasmic reticulum	EC:2.7.11.1
Replication and Repair					
1	K08737	LmjF36.1950	MSH6	Mismatch repair, DNA repair and recombination proteins	
2	K02335	LmjF34.1260	polA	DNA replication, DNA replication proteins, Base excision repair, Nucleotide excision repair, Homologous recombination, DNA repair and recombination proteins	EC:2.7.7.7
3	K08736	LmjF15.1420	MSH3	Mismatch repair, DNA repair and recombination proteins	
4	K11304	LmjF36.6990	TIP60, KAT5, ESA1	Chromosome	EC:2.3.1.48
5	K10737	LmjF05.0330	MCM8	DNA replication proteins	
6	K03355	LmjF12.0610	APC8, CDC23	Chromosome	
7	K10848	LmjF08.0140	ERCC4, XPF	DNA replication proteins, Nucleotide excision repair, DNA repair and recombination proteins	EC:3.1.-.-
8	K06627	LmjF25.1470	CCNA1,CCNA2	DNA replication proteins	
9	K04728	LmjF02.0120	ATM,tefu	DNA replication proteins, DNA repair and recombination proteins	EC:2.7.11.1
10	K02202	LmjF26.0040	CDK7	Nucleotide excision repair, DNA repair and recombination proteins	EC:2.7.11.22

Environmental Information Processing

Membrane Transport					
1	K05672	LmjF23.0210	ABCC12	ABC Transporters	
2	K05666	LmjF23.0250	ABCC2	ABC Transporters	
3	K05658	LmjF34.0990	ABCB1	ABC Transporters	
4	K05681	LmjF36.2890	ABCG2	ABC Transporters	
5	K05666	LmjF23.0250 LmjF31.1290	ABCC2	ABC Transporters	
6	K08150	LmjF24.0680	SLC2A13, ITR	ABC Transporters	
7	K06861	LmjF29.0620	lptB	ABC Transporters	EC:3.6.3.-
8	K05677	LmjF31.0540 LmjF33.1860 LmjF31.1280	ABCD3, PMP70	ABC Transporters	
9	K05643	LmjF11.1220	ABCA3	ABC Transporters	
10	K09687	LmjF15.0760	yadG	ABC Transporters	
11	K05662	LmjF32.3080	ABCB7	ABC Transporters	
12	K05657	LmjF25.0530	ABCB10	ABC Transporters	
Signal Transduction					
1	K04428	LmjF36.3680	MAP3K4	MAPK signaling pathway	EC:2.7.11.25
2	K07203	LmjF34.4530	MTOR	ErbB signaling pathway, mTOR signaling pathway	
3	K05863	LmjF19.0200	SLC25A4	Calcium signaling pathway	
4	K04421	LmjF33.2290	MAP3K3	MAPK signaling pathway	EC:2.7.11.25
5	K00981	LmjF26.1620	CDS1, cdsA	Phosphatidylinositol signaling system	EC:2.7.7.41
6	K00981	LmjF26.1620	PI4K	Phosphatidylinositol signaling system	EC:2.7.1.67
7	K04505	LmjF15.1530	PSEN1, PS1	Wnt signaling pathway, Notch signaling pathway	EC:3.4.23.-
8	K11229	LmjF21.0130, LmjF26.1730	BCK1	MAPK signaling pathway - yeast	EC:2.7.11.1
9	K00999	LmjF26.2480	CDIPT	Phosphatidylinositol signaling system	EC:2.7.8.11
10	K00914	LmjF24.2010	PIK3C3, VPS34	Phosphatidylinositol signaling system	EC:2.7.1.137
11	K01110	LmjF34.1430	PTEN	Phosphatidylinositol signaling system	EC:3.1.3.67
12	K01099	LmjF15.1210	INPP5B, SYNJ2, SYNJ1	Phosphatidylinositol signaling system	EC:3.1.3.36
13	K07198	LmjF29.2020	PRKAA2,P RKAA1, SNF1A	mTOR signaling pathway	EC:2.7.11.11
Signaling Molecules and Interaction					
1	K05658	LmjF34.0990	ABCB1	Cellular antigens	
2	K05016	LmjF04.1000, LmjF32.3370	CLCN7	Ion channels	
3	K05681	LmjF36.2890	ABCG2	Cellular antigens	
4	K05012	LmjF01.0180	CLCN3	Ion channels	
5	K04843	LmjF34.0480	SCN11A	Ion channels	

Organismal Systems

Endocrine System					
1	K04428	LmjF36.3680	MAP3K4	GnRH signaling pathway	EC:2.7.11.25
2	K00507	LmjF24.2250, LmjF14.0510	SCD	PPAR signaling pathway	EC:1.14.19.1
3	K07203	LmjF34.4530	MTOR	Insulin signaling pathway, Adipocytokine signaling pathway	
4	K04421	LmjF33.2290	MAP3K3	GnRH signaling pathway	EC:2.7.11.25
5	K00507	LmjF24.2250, LmjF14.0510	SCD, desC	PPAR signaling pathway	EC:1.14.19.1
6	K03355	LmjF12.0610	APC8, CDC23	Progesterone-mediated oocyte maturation	
7	K08745	LmjF24.1780	SLC27A1_4, FATP1, FATP4	PPAR signaling pathway	EC:6.2.1.-
8	K06627	LmjF25.1470	CCNA1,CCNA2	Progesterone-mediated oocyte maturation	
9	K07198	LmjF29.2020	PRKAA2, PRKAA1, SNF1A	Insulin signaling pathway, Adipocytokine signaling pathway	EC:2.7.11.11
Nervous System					
1	K04421	LmjF33.2290	MAP3K3	Neurotrophin signaling pathway	EC:2.7.11.25
2	K04505	LmjF15.1530	PSEN1, PS1	Neurotrophin signaling pathway	EC:3.4.23.-
Digestive System					
1	K05666	LmjF31.1290, LmjF23.0250	ABCC2	Bile secretion	
2	K05658	LmjF34.0990	ABCB1	Bile secretion	
3	K05681	LmjF36.2890	ABCG2	Bile secretion	
4	K05667	LmjF31.1280	ABCC3	Bile secretion	

Cellular Processes

Cell Motility					
1	K10392	LmjF34.4260	KIF13A, KIF13B, KIF13C	Cytoskeleton proteins	
Cell Growth and Death					
1	K01768	LmjF17.0200	ACXA, ACXB, ACXC ,ACXD	Meiosis - yeast	EC:4.6.1.1
2	K03355	LmjF12.0610	APC8, CDC23	Cell cycle, Cell cycle-yeast, Meiosis – yeast, Oocyte meiosis	
3	K01110	LmjF34.1430	PTEN	p53 signaling pathway	EC:3.1.3.67
4	K06627	LmjF25.1470	CCNA1,CCNA2	Cell cycle	
5	K04728	LmjF02.0120	ATM,tefu	Cell cycle, Apoptosis, p53 signaling pathway	EC:2.7.11.1
6	K02202	LmjF26.0040	CDK7	Cell cycle	EC:2.7.11.22
Transport and Catabolism					
1	K12393	LmjF22.1140	AP1M1	Lysosome	
2	K13338	LmjF34.3520	PEX1	Peroxisome	
3	K05677	LmjF33.1860, LmjF31.0540	ABCD3, PMP70	Peroxisome	

4	K12398	LmjF34.2590	AP3M	Lysosome	
5	K00649	LmjF34.1090	GNPAT	Peroxisome	EC:2.3.1.42
6	K02154	LmjF23.1510, LmjF32.0920	ATPeVI, ATP6N1A	Phagosome, Lysosome	EC:3.6.3.14
7	K00914	LmjF24.2010	PIK3C3, VPS34	Phagosome, Regulation of autophagy	EC:2.7.1.137
8	K08337	LmjF07.0010	ATG7	Regulation of autophagy	
9	K07198	LmjF29.2020	PRKAA2, PRKAA1, SNF1A	Regulation of autophagy	EC:2.7.11.11
Cell Communication					
1	K01110	LmjF34.1430	PTEN	Focal adhesion, Tight junction	EC:3.1.3.67

Table 2: Essential proteins of *L.major* involved in several metabolic pathways.

Protein ID	Protein Function
LmjF23.0210	multidrug resistance protein, putative
LmjF03.0030	D-3-phosphoglycerate dehydrogenase-like protein
LmjF23.0250	multidrug resistance protein, putative
LmjF06.0950	glucosamine-fructose-6-phosphate aminotransferase, putative
LmjF14.0510	stearic acid desaturase, putative
LmjF14.1200	phosphatidylserine synthase, putative
LmjF17.0200	RAC-A‖receptor-type adenylate cyclase a
LmjF08.0450	signal peptidase type I, putative
LmjF29.1450	phosphatidylinositol-kinase domain protein, putative
LmjF29.0620	ABC transporter, putative
LmjF06.0260	5'-3' exonuclease, putative
LmjF14.0460	cystathionine beta-lyase-like protein
LmjF04.1000	chloride channel protein, putative
LmjF05.0330	DNA replication licensing factor, putative
LmjF29.2600	poly(A) polymerase, putative
LmjF23.1300	athosterol oxidase-like protein
LmjF35.1700	RNA 3'-terminal phosphate cyclase
LmjF06.0650	anosterol synthase, putative
LmjF05.0360	ATP-dependent RNA helicase, putative
LmjF01.0180	CLC-type chloride channel, putative
LmjF07.0650	vacuolar-type Ca2+-ATPase, putative
LmjF23.1510	vacuolar proton translocating ATPase subunit A, putative
LmjF08.0140	DNA repair protein, putative
LmjF13.1530	phospholipid-transporting ATPase 1-like protein
LmjF11.1220	ABC transporter, putative
LmjF07.0010	ubiquitin activating E1 enzyme, putative
LmjF29.2240	aminopeptidase, putative
LmjF29.2020	serine/threonine protein kinase, putative
LmjF03.0500	phosphate-Repressible Phosphate Permease-like protein

Conclusion

In the past few decades, there has been a tremendous increase in amount of genomic and proteomic data generated from the sequencing projects. This data along with the developed computer aided softwares can be utilized to identify and characterize probable drug targets. The in silico approaches enable us to reduce substantially the number of such candidates to test and speed up drug discovery process with least toxicity risk. The DEG database is a better option for identification of essential genes more efficiently and frequently rather than conventional methods of gene/protein essentiality and facilitates the identification of the most relevant drug targets in the pathogens. The subtractive proteomic approach has been applied in the present study for the identification of several proteins that can be targeted for effective drug development against *L. major*. The drugs developed against the identified proteins will be specific for the pathogen and therefore less/non toxic for the *H.sapiens*. The drug development against these novel targets might be useful in the discovery of novel therapeutic compounds against *L. major*.

References

[1] Dubessay P, Ravel C, Bastien P, Crobu L, Dedet JP, Pagès M, Blaineau C. The switch region on Leishmania major chromosome 1 is not required for mitotic stability or gene expression, but appears to be essential. Nucleic Acids Res. 2002 Sep 1;30(17):3692-7.

[2] Ameen M (2007) Cutaneous leishmaniasis: therapeutic strategies and future directions. Expert Opin Pharmacother 8: 2689–2699.

[3] Roberts LJ, Handman E, Foote SJ (2000) Leishmaniasis. BMJ 321: 801–804.

[4] Leishmaniasis epidemic hits Afghanistan (2002) CMAJ 167: 536.

[5] Stephenson J (2004) Leishmaniasis epidemic. JAMA 292: 1294.

[6] Hotez P (2008) Neglected infections of poverty in the United States of America. PLoS Negl Trop Dis 2: e256. doi:10.1371/journal.pntd.0000256.

[7] Miesel, L., Greene, J. and Black, T. A. (2003). Genetic strategies for antibacterial drug discovery. Nature Rev. Genet. 4, 442-456.

[8] Kaminski, N. (2000). Bioinformatics. A user's perspective. Am. J. Respir. Cell Mol. Biol. 23, 705-711.

[9] Itaya, M. (1995). An estimation of the minimum genome size required for life. FEBS Lett. 362, 257-260.

[10] Tatusov, R. L., Koonin, E. V. and Lipman, D. J. (1997). A genomic perspective of protein families. Science 278, 631-637.

[11] Koonin, E. V., Tatusov, R. L. and Galperin, M. Y. (1998). Beyond complete genomes - from sequence to structure and function. Curr. Opin. Struc. Biol. 8, 355-363.

[12] Kobayashi, K., et al. (2003). Essential Bacillus subtilis genes. Proc. Natl. Acad. Sci. USA 100, 4678-4683.

[13] Judson, N. and Mekalanos, J. J. (2000a). TnAraOut, A transposon-based approach to identify and characterize essential bacterial genes. Nat. Biotechnol. 18, 740-745.

[14] Judson, N. and Mekalanos, J. J. (2000b). Transposon-based approaches to identify essential bacterial genes. Trends Microbiol. 8, 521-526.
[15] Ren Zhang and Yan Lin, (2009) DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Research 37, D455-D458
[16] Ren Zhang, Hong-Yu Ou and Chun-Ting Zhang, (2004) DEG, a Database of Essential Genes. Nucleic Acids Research 32, D271-D272.
[17] Ying Huang, Beifang Niu, Ying Gao, Limin Fu and Weizhong Li. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics, 2010
[18] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucl.AcidRes. 25(1997) 3389–3402 [PMID: 9254694].
[19] Ohlstein, E. H., R. Ruffolo, Jr., and J. D. Elliott. 2000. Drug discovery in the next millennium. Annu Rev Pharmacol Toxicol 40:177–91.
[20] Maliepaard, M., G. L. Scheffer, I. F. Faneyte, M. A. van Gastelen, A. C. Pijnenborg, A. H. Schinkel, M. J. van De Vijver, et al. 2001. Subcellular localization and distribution of the breast cancer resistance protein transporter in normal human tissues. Cancer Res 61:3458–64.
[21] Clark, H. F., A. L. Gurney, E. Abaya, K. Baker, D. Baldwin, J. Brush, J. Chen, et al. 2003. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: A bioinformatics assessment. Genome Res 13:2265–70.
[22] Koonin, E. V. 2000. Bridging the gap between sequence and function. Trends Genet 16:16.
[23] Devos, D., and A. Valencia. 2001. Intrinsic errors in genome annotation. Trends Genet 17:429–31.
[24] Valencia, A., and F. Pazos. 2002. Computational methods for the prediction of protein interactions. Curr Opin Struct Biol 12:368–73.
[25] Guo T, Hua S, Ji X, Sun Z. DBSubLoc: database of protein subcellular localization. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D122-4. PMID14681374
[26] Yu CS, Lin CJ, Hwang JK: Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Science 2004, 13:1402-1406.
[27] Yu CS, Chen YC, Lu CH, Hwang JK: Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics 2006, 64:643-651.
[28] Mi H, Guo N, Kejariwal A, Thomas PD. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res. 2007 Jan;35(Database issue):D247-52. Epub 2006 Nov 27.
[29] Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M.; KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182-W185 (2007).