



Insilico analysis of hypothetical proteins unveils putative metabolic pathways and essential genes in *Leishmania donovani*

Nithin Ravooru¹, Sandesh Ganji^{1†}, Nitish Sathyanarayanan^{2*} and Holenarsipur G. Nagendra¹

¹ Department of Biotechnology, Sir Mokshagundam Visvesvaraya Institute of Technology, Bangalore, India

² The National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

Edited by:

Prashanth Suravajhala, Bioclues Organization, Denmark

Reviewed by:

Franca Fraternali, King's College London, UK

Prashanth Suravajhala, Bioclues Organization, Denmark

*Correspondence:

Nitish Sathyanarayanan, National Center for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560065, India

e-mail: nitishs@ncbs.res.in

† Present address:

Sandesh Ganji, Cermer Healthcare Solutions, Manyata TechPark, Bangalore, India

Leishmaniasis is a parasitic disease caused by the protozoan *Leishmania*, which is active in two broad forms namely, *Visceral Leishmaniasis* (VL or Kala Azar) and *Cutaneous Leishmaniasis* (CL). The disease is most prevalent in the tropical regions and poses a threat to over 70 countries across the globe. About 200 million people are estimated to be at risk of developing VL in the Indian subcontinent, and this refers to around 67% of the global VL disease burden. The Indian state of Bihar alone accounts for 50% of the total VL cases. While no vaccination exists, several pentavalent antimonials and drugs like Paromomycin, Amphotericin, Miltefosine etc. are used in the treatment of Leishmaniasis. However, due to their low efficacies and the resistance developed by the bug to these medications, there is an urgent need to look into newer species specific targets. The proteome information available suggests that among the 7960 proteins in *Leishmania donovani*, a staggering 65% remains classified as a hypothetical uncharacterized set. In this background, we have attempted to assign probable functions to these hypothetical sequences present in this parasite, to explore their plausible roles as druggable receptors. Thus, putative functions have been defined to 105 hypothetical proteins, which exhibited a GO term correlation and PFAM domain coverage of more than 50% over the query sequence length. Of these, 27 sequences were found to be associated with a reference pathway in KEGG as well. Further, using homology approaches, four pathways viz., Ubiquinone biosynthesis, Fatty acid elongation in Mitochondria, Fatty Acid Elongation in ER and Seleno-cysteine Metabolism have been reconstructed. In addition, 7 new putative essential genes have been mined with the help of Eukaryotic Database of Essential Genes (DEG). All these information related to pathways and essential genes indeed show promise for exploiting the select molecules as potential therapeutic targets.

Keywords: *Leishmania donovani*, hypothetical proteins, ubiquinone biosynthesis, seleno-cysteine metabolism, fatty acid elongation, drug targets

INTRODUCTION

Leishmaniasis is a parasitic disease caused by the protozoan belonging to the genus *Leishmania* and is transmitted by the vector phlebotomine or sand fly (Alvar et al., 2013). The disease is active in two broad forms namely Visceral Leishmaniasis (VL or Kala Azar) and Cutaneous Leishmaniasis (CL) (Croft et al., 2006). Visceral Leishmaniasis is the more severe form of the disease, characterized by anemia, splenohepatomegaly, depressed immune response and several secondary infections finally leading to death (Alvar et al., 2013). VL is primarily caused by the protozoan *Leishmania donovani* in the South Asia and African regions and by *Leishmania infantum* in the central and South American regions (Croft and Olliaro, 2011).

The disease is most prevalent in the tropical regions and poses a threat to over 70 countries across the globe. Approximately, there are 0.7–1.2 million cases of VL and CL respectively, recorded each year and about 20,000–40,000 Leishmaniasis deaths occur

per year. The 10 countries with the highest estimated cases namely Afghanistan, Algeria, Colombia, Brazil, Iran, Syria, Ethiopia, North Sudan, Costa Rica and Peru, together account for 70–75% of global estimated CL incidence (Alvar et al., 2012). In the Indian subcontinent, about 200 million people are estimated to be at risk of developing VL and this area harbors an estimated 67% of the global VL disease burden. The north Indian state of Bihar alone has captured almost 50% of the total cases in the Asian region (Bhunja et al., 2013).

Several pentavalent antimonials and drugs like Amphotericin and Paromomycin are currently available as intramuscular injections, while Miltefosine is used as an oral drug, for the treatment of Leishmaniasis. Vector control measures and the first line of drugs have proved incapable of suppressing the disease, especially in India where two thirds of the patients did not respond to these pentavalent antimonials (Lira et al., 1999; Croft et al., 2006; Sundar et al., 2009). The medications are not satisfactory mainly

due to their toxicity effects, drug resistance due to their long half-life and the costs associated with the treatment (Desjeux, 2004; Monzote, 2009; Singh et al., 2012). Thus, there is an immense & immediate requirement to look at species specific drug targets to tackle this pathogen (Guerin et al., 2002). The proteome information available suggests that amongst the 7960 protein sequences, a staggering 65% of it remains to be annotated with clarity.

Hence, as a step toward characterization of these hypothetical sequences as plausible drug targets, computational approaches have been employed toward analysing these molecules. Literature suggests that several *insilico* approaches have been adopted in order to assign functional information for such hypothetical sequences in various organisms. More than half of the uncharacterized proteins in *M. tuberculosis* are functionally correlated via computational approaches (Doerks et al., 2012). *Insilico* analysis of hypothetical proteins present in human fetal brain has been predicted to contain many sequences which function in DNA-protein binding and ligase activity (Sharma et al., 2013). Also, *insilico* characterization of hypothetical proteins in *Plasmodium falciparum* suggests that several sequences can be considered as biomarkers in Malaria (Oladele et al., 2011). Recent studies on hypothetical proteins in Trypanosomatids have predicted protein-protein interactions on a genome scale, which could be used to explore new potential drug targets (Rezende et al., 2012). In another recent study in *Trypanosoma cruzi*, attempts have been made to computationally annotate the hypothetical membrane proteins in order to identify putative drug targets (Silber and Pereira, 2012). However, there has been no comprehensive study on the hypothetical protein dataset in *Leishmania donovani* hence; a detailed investigation has been attempted.

MATERIALS AND METHODS

DATABASES EMPLOYED

Hypothetical sequences were retrieved from UNIPROT-KB (Release 2014_02) (The UniProt Consortium, 2013). KEGG (Release 69.0) database was used for assigning pathway information (Kanehisa et al., 2014). Eukaryote specific Database of Essential Genes (DEG) (Version 10.0) (Luo et al., 2013) was used to search for putative essential genes within *Leishmania donovani*.

TOOLS FOR FUNCTIONAL ANNOTATION

HMMscan, both web version and standalone (HMMER 3.1b1—Finn et al., 2011) and Batch CDD search (Marchler-Bauer et al., 2011) were used to assign PFAM domain information to the query sequences. Blast2GO (Conesa et al., 2005; Götz et al., 2008) was used to assign functional Gene Ontology (GO) terms (Ashburner et al., 2000) to the protein sequences. KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007) was used to predict the pathway associations of the protein sequences. String DB (version 9.1) (Franceschini et al., 2013) was used to identify and analyse the COG (cluster of Orthologous groups) (Tatusov et al., 1997) networks between protein families.

TOOLS FOR SEQUENCE SEARCH AND PHYLOGENY

Sequence search algorithm, BLAST was used for identification of homologs where ever necessary. Jackhmmmer (HMMER 3.1b1)

standalone version was used to search homologs against the Eukaryote DEG database. MEGA 5.0 (Tamura et al., 2011) was used for building phylogenetic tree where Maximum likelihood approach was used. Fig-tree (version 1.4) was used to visualize the phylogenetic tree.

SEQUENCE ANALYSIS

The protocol used in this study is depicted in **Figure 1**, as a flowchart. 5299 hypothetical protein sequences belonging to *Leishmania donovani* were retrieved from Uniprot. These sequences were analyzed for domain information using HMMscan with an *e*-value of 10^{-3} against the PFAM database (version 27.0) (Finn et al., 2013). We obtained 1898 sequences, which had at least one PFAM domain associated with the query. Further, these 1898 sequences were analyzed for possible GO term associations using Blast2GO tool. The sequences were queried against Swissprot database at an *e*-value of 10^{-10} , which resulted in 727 sequences being associated with one or more GO terms. Of the 727 sequences, 105 sequences had a PFAM domain covering more than 50% of the query's length. Hence, these 105 sequences which were outcome of the two filters viz., GO term association and PFAM domain coverage formed the final dataset used for a detailed analysis. The sequences that did not clear the threshold

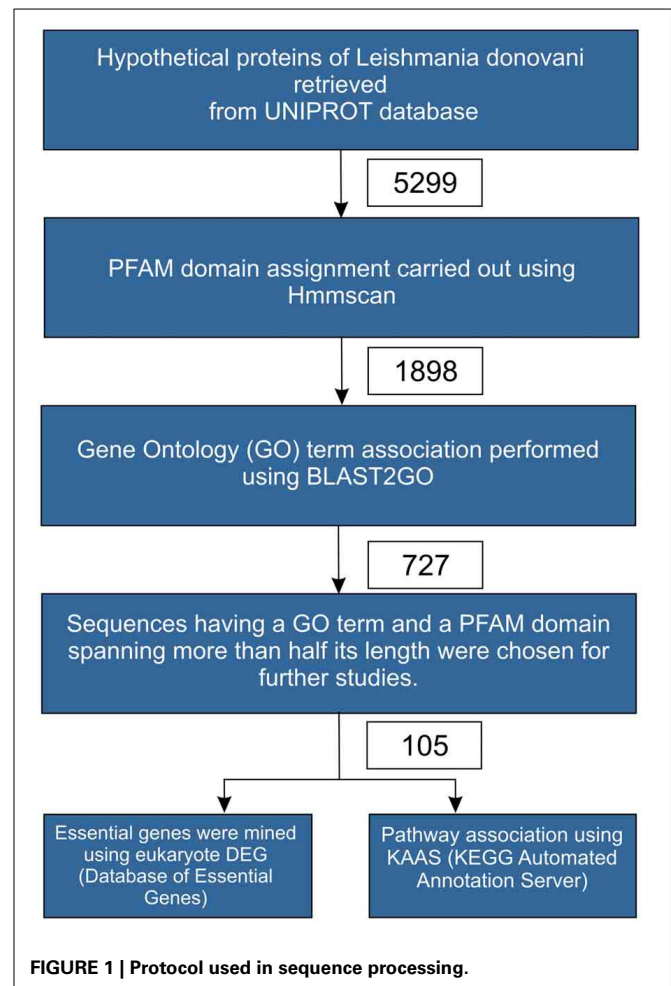
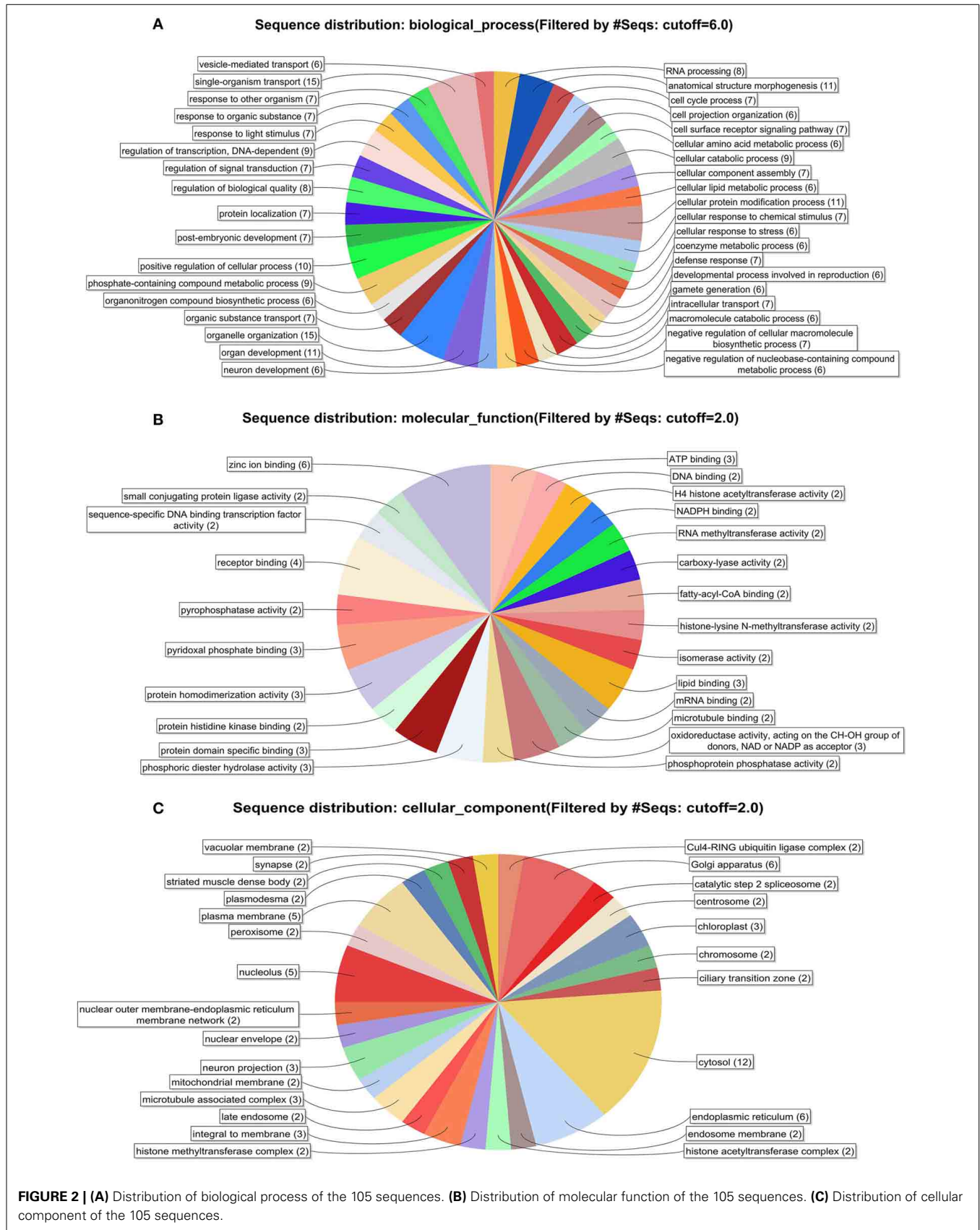


FIGURE 1 | Protocol used in sequence processing.



parameters at various filter, may have a higher likelihood of being false positives, and this may warrant a separate in depth analysis of the same.

RESULTS AND DISCUSSIONS

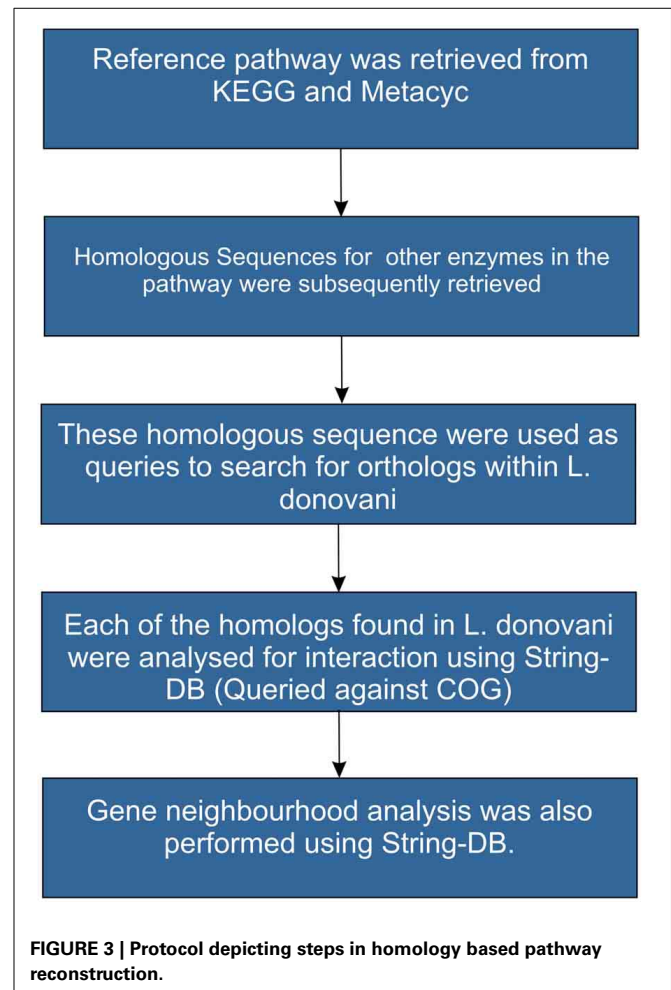
In the present work, we have integrated the available sequences and functional information from various resources to assign putative function to hypothetical proteins. Of the total proteome of 7960 sequences, 5299 proteins in this pathogen are termed hypothetical/uncharacterized, which amounts to a monumental 65% of the proteome that needs to be characterized with structural and functional information. Thus, under the current study 105 sequences have been putatively annotated using sequence/domain information from PFAM, functional information from GO, pathway information from KEGG and essential gene information from DEG.

All the information related to GO terms, Interproscan domain associations, homologs picked during the BLAST step in the BLAST2GO analysis of the 105 sequences, are presented in the **Supplementary Table 1** (as an Excel file). Similarly, **Supplementary Figure 1** indicates the sequence similarity distribution among the BLAST hits obtained in the first step of Blast2Go analysis. No BLAST hits were present with less than 30% sequence similarity with respect to the query, indicating a good homology with the query. **Supplementary Figure 2** shows the *E*-value distribution of the hits from the BLAST step. It is evident from the plot that only the hits with a significant *E*-value were considered for further analysis. **Figures 2A–C** show combined graphs depicting the biological processes, molecular functions and cellular components of the 105 sequences respectively. Presence of important class of proteins such as Zinc binding, Receptor binding, DNA binding etc., has been highlighted through GO annotation. This information can be extended to investigate the functional roles of these less characterized molecules in greater detail. Further, KEGG Automatic Annotation Server (KAAS) was used to predict Pathway associations for the 105 sequences that have an associated GO term and a domain spanning more than half of its length. KAAS was performed using the Best bidirectional Hit (BBH) method which resulted in 27 sequences associated with 33 KEGG pathways. The complete list of pathways is given in **Supplementary Table 2**, which provides crucial information related to basic metabolic actions within the protozoa. The 105 sequences when queried against the Eukaryote Database of Essential Genes (DEG), using Jackhammer with an *e*-value of 10^{-20} , resulted in 26 sequences that had 93 hits from eukaryote DEG.

Here, attempts have been made to reconstruct 4 pathways, viz., Ubiquinone biosynthesis, Fatty acid elongation in Mitochondria, Fatty acid elongation in ER and Seleno-cysteine Metabolism. Sequence homology information is derived from String DB, COG and sequence search tool such as BLAST. Reference pathway information was obtained either from KEGG or MetaCyc databases.

HOMOLOGY BASED PATHWAY RECONSTRUCTION

The protocol involved in homology based pathway reconstruction is depicted in **Figure 3**. Complete sequence information of the pathways was retrieved from MetaCyc (Caspi et al., 2011). Protein



sequence catalyzing each step in the reference pathway was used as a query to search for homologs in *Leishmania donovani* (taxid: 5561) using BlastP at an *e*-value of 10^{-3} . Hits which had a coverage of greater than 70% and an identity of $>30\%$ were considered as true positives. Further, homologs found within *Leishmania donovani* were used as query to understand the conservation of gene neighborhood using String DB. In a recent study, Doerks et al. (2012) has demonstrated the use of gene neighborhood approach to mine a putative cell envelope biogenesis operon in *Mtb*. Such gene neighborhood, co-occurrence patterns and conservation of proteins in a pathway across evolutionary space signifies the importance of the pathway analysis.

CASE STUDY 1: UBIQUINONE BIOSYNTHESIS PATHWAY

In studies involving *Plasmodium* parasites, arrested oocyte maturation is seen in the lack of NADH-Ubiquinone oxidoreductase, which is a part of the electron transport chain (Boysen and Matuschewski, 2011). This shows the importance of ubiquinone and its role within the plasmodium parasite. Subtractive genomic studies in *Mycobacterium tuberculosis* have revealed the possibility of Ubiquinone biosynthesis pathway as targets in multidrug variants of the bacteria (Anishetty et al., 2005).

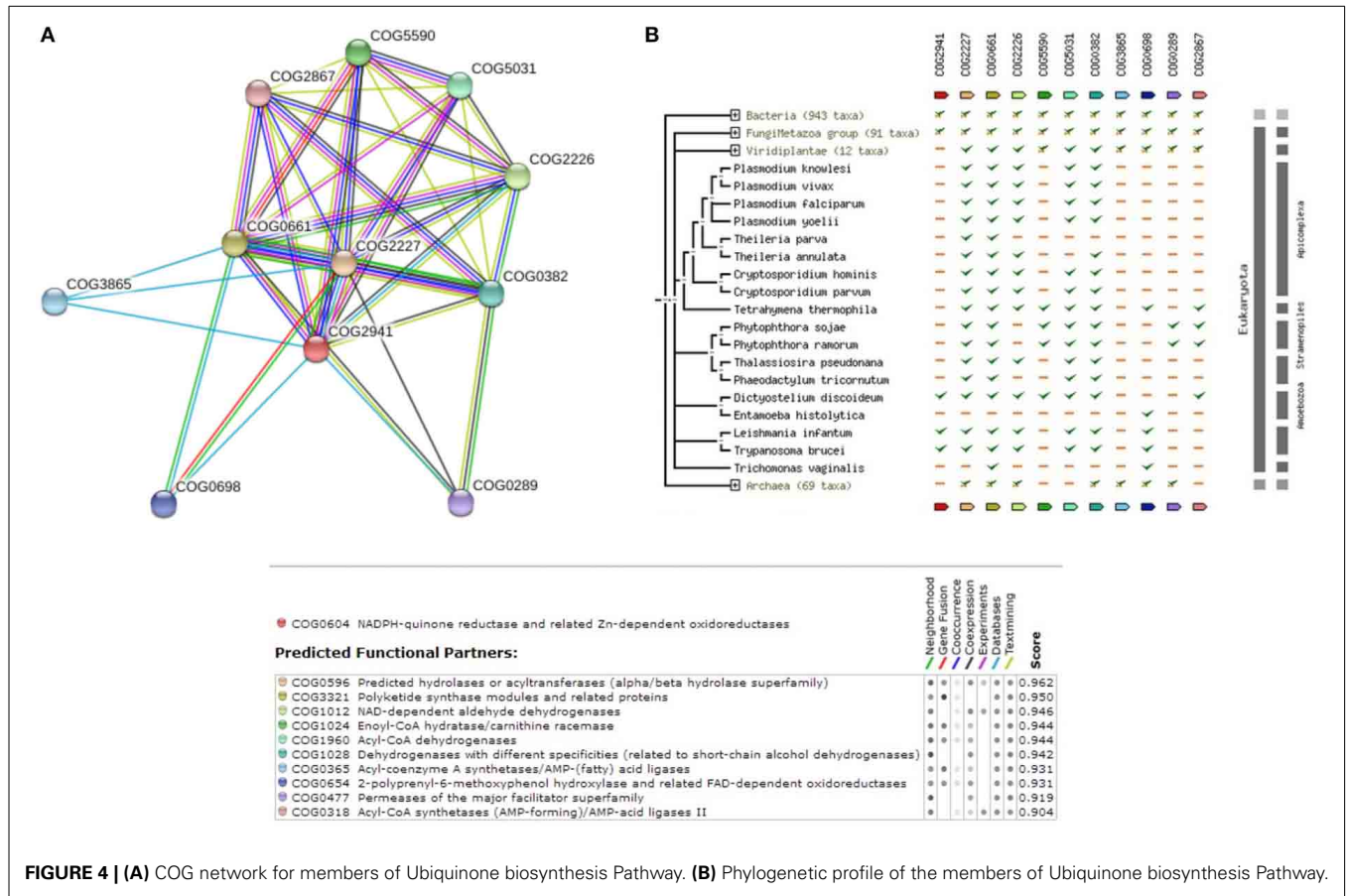


FIGURE 4 | (A) COG network for members of Ubiquinone biosynthesis Pathway. **(B)** Phylogenetic profile of the members of Ubiquinone biosynthesis Pathway.

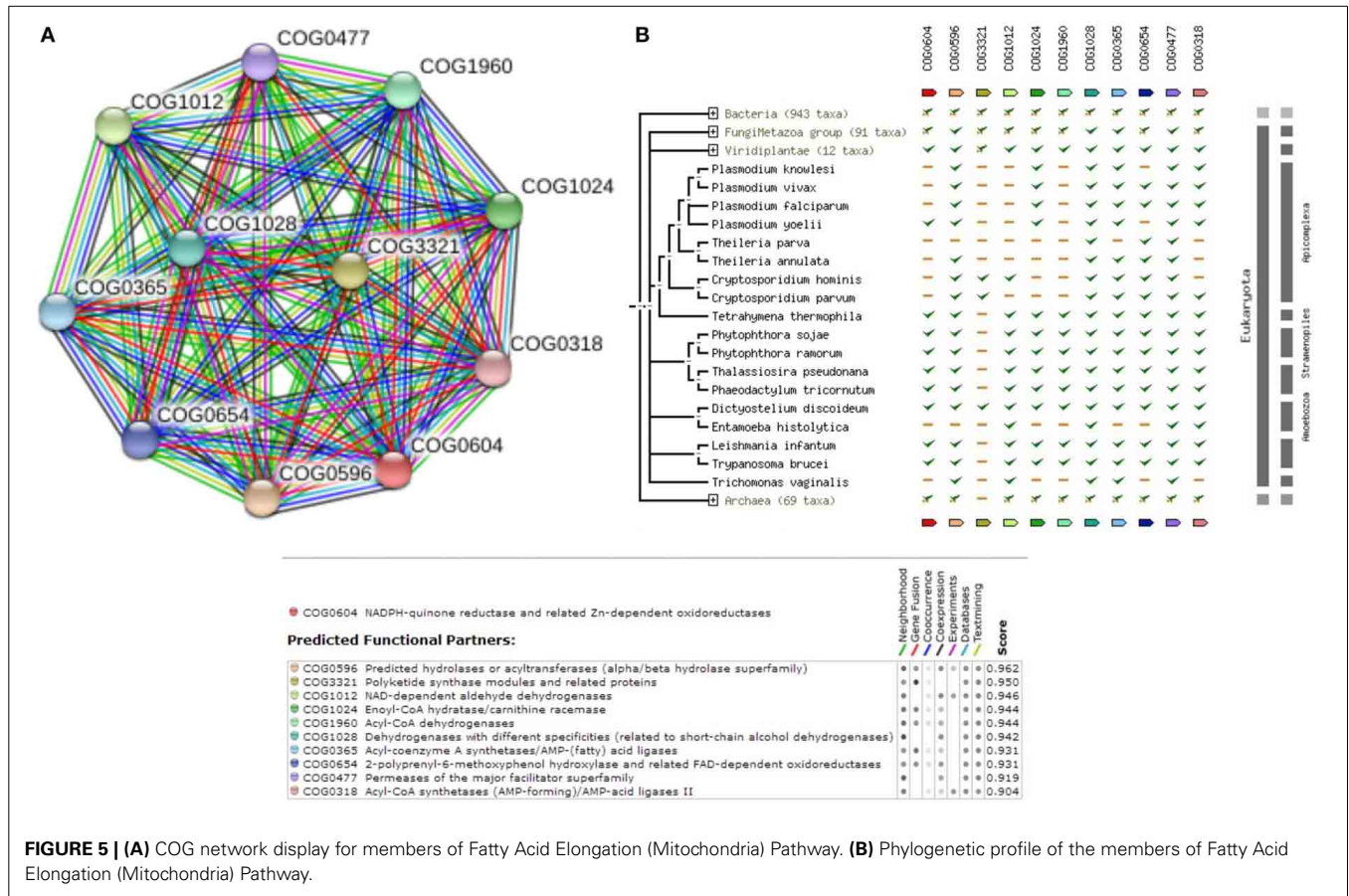
Hence, considering the possibility of ubiquinone biosynthesis pathway as a potential drug target, we explored to reconstruct the pathway in *Leishmania donovani* as well. The complete list of enzymes that are involved in ubiquinone biosynthesis within *Leishmania donovani* is also illustrated in **Supplementary Table 3**. **Supplementary Figure 3** shows the reference pathway and steps involved in ubiquinone biosynthesis. **Figure 4A** depicts the String-DB interaction of E9BL43 as queried against COG, while **Figure 4B** illustrates the Phylogenetic profile of query and other members of COG. As can be appreciated from **Figure 4B**, interaction of 5 out of 7 enzymes [E9BL43 (COG2941), E9BJL4 (COG0382), E9BSV2 (COG2227), E9BAB0 (COG0661), E9BUR6 (COG2226)], catalyzing various reactions within the pathway remain conserved in the genus *Leishmania*. Additionally, the enzymes in the Ubiquinone biosynthesis pathway are well conserved in other species of *Leishmania* (sharing >90% similarity and identity) as shown in **Supplementary Table 3**. Furthermore, it is seen that of the 7 enzymes established in the pathway, 4 are termed uncharacterized (E9BLP8, E9B8Y8, E9BAB0, E9BL43) in UNIPROT. The query sequence E9BL43, which is among the dataset of 105 sequences, is closely related to coq7, which in turn is shown to be closely interacting with structural components like coq4 and coq9 and the functional components like coq5 and coq6 in *Saccharomyces cerevisiae* (Hsieh et al., 2007). This highlights the role of coq7 (E9BL43) in the COQ (Coenzyme Q) biosynthesis pathway and its interactions with

other members as COQ is an essential cofactor in mitochondrial respiration.

CASE STUDY 2: FATTY ACID ELONGATION (IN MITOCHONDRIA) PATHWAY

Over the years, chemotherapy has been the principle method employed in order to control the disease along with the pentavalent antimonials like Paramomycin, which are rather expensive and toxic. Miltefosine [hexadecylphosphocholine (HePC)] was the first drug to be approved for oral administration against the antimony-resistant cases and cutaneous leishmaniasis. Studies have exhibited classical modifications in the lipid composition in membranes of *Leishmania donovani* promastigotes, that are resistant to Miltefosine (Rakotomanga et al., 2005). The study also suggests that the variations in the lipid composition of the membranes (fatty acid composition and length of alkyl chains) in Miltefosine resistant *L. donovani* promastigotes, with that of its wild-type counterparts, could be helpful in identifying biochemical targets which undergo alterations due to drug resistance processes.

Hence, understanding the sequence level information of such crucial metabolic pathways related to fatty acid elongation (Mitochondria and endoplasmic reticulum), would aid in better appreciation of Miltefosine induced drug resistance. **Supplementary Table 4** shows the complete list of enzymes that are involved in Fatty acid elongation pathway (Mitochondria)



within *Leishmania donovani*. The query E9B7Z4 is found to be closely associated to enzyme of the class Trans-2-enoyl-CoA reductase (1.3.1.38). **Supplementary Table 4** also shows that the enzymes in the pathway are well conserved in other species of *Leishmania* (with >90% coverage and similarity). **Supplementary Figure 4** shows the reference pathway obtained from KEGG and steps involved in Fatty Acid elongation. **Figure 5A** shows the String-DB interaction of E9B7Z4 as queried against COG; while **Figure 5B** shows the phylogenetic profile of query and other members of COG.

CASE STUDY 3: FATTY ACID ELONGATION (IN ENDOPLASMIC RETICULUM) PATHWAY

Fatty acid elongation can occur at Endoplasmic Reticulum (ER) apart from Mitochondria. Since studies have not determined the cellular location of Miltefosine induced drug resistance, it is also important to understand the sequence level information of Fatty acid elongation pathway in ER. The query E9BQF5, which is part of the 105 sequence dataset, is closely related to Ketoacyl-coA-reductase (1.1.1.330). **Supplementary Table 5** shows that the enzymes in the pathway are well conserved with other species of *Leishmania* (with >90% coverage and similarity). **Supplementary Figure 5** depicts the reference pathway obtained from KEGG and the steps involved in the pathway. **Figure 6A** exhibits the String-DB interaction of E9BQF5 as queried against COG while, **Figure 6B** refers to the phylogenetic profile of the query and other members of COG.

CASE STUDY 4: SELENO-CYSTEINE METABOLISM PATHWAY

Selenoproteins play a wide range of roles in metabolism and oxidative stress defense in many organisms. Seleno-cysteine is synthesized by reaction of seleno-phosphate with a serine charged tRNA. The unique reactivity of selenocysteine and the specialized machinery required for selenoprotein synthesis, make selenoproteins an attractive target for antimicrobial development (Jackson-Rosario and Self, 2010). Recently, selenoproteins have been identified in a number of parasitic organisms including trypanosomes and platyhelminths (Lobanov et al., 2006; Bonilla et al., 2008). In addition, selenoproteins in *Plasmodium falciparum* have been suggested as possible targets for therapeutic development (Jackson-Rosario and Self, 2010). Studies have shown that Trypanosoma and Leishmania are sensitive to auranofin, a potent selenoprotein inhibitor; however, the probable drug mechanism is not related to selenoproteins in kinetoplastids. Latest studies have also shown that Selenium supplementation decreases the parasitemia of various Trypanosome infections and reduces important parameters associated with diseases such as anemia and parasite-induced organ damage (Da Silva et al., 2014). It is, thus, interesting to understand the sequence level information of the proteins involved in Seleno-cysteine biosynthesis in *Leishmania donovani*.

The query E9B9Y6, which is part of the 105 sequence dataset, is closely related to O-phosphoserine-tRNA:selenocysteinyl-tRNA synthase (2.9.1.2) which is involved in the synthesis of

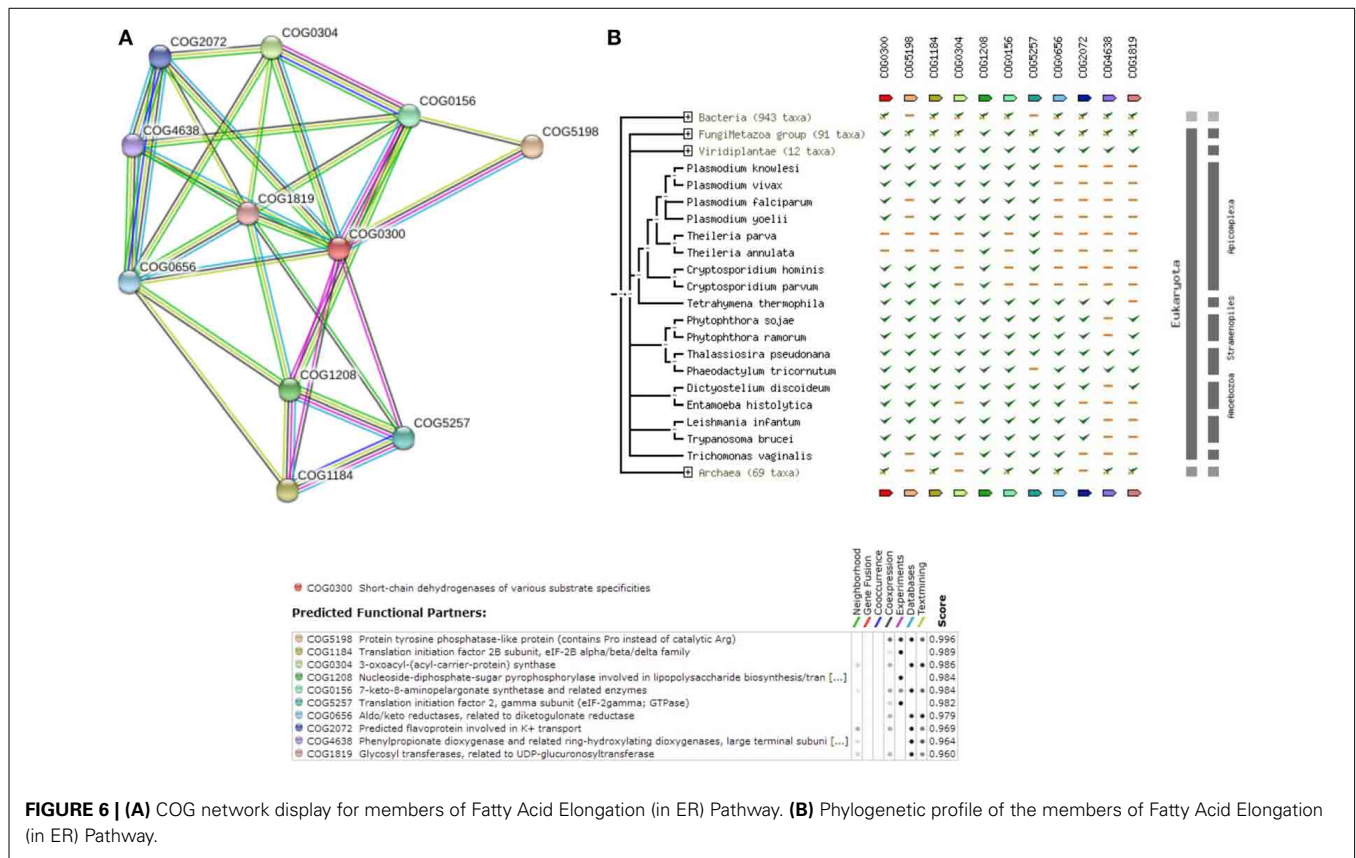


FIGURE 6 | (A) COG network display for members of Fatty Acid Elongation (in ER) Pathway. **(B)** Phylogenetic profile of the members of Fatty Acid Elongation (in ER) Pathway.

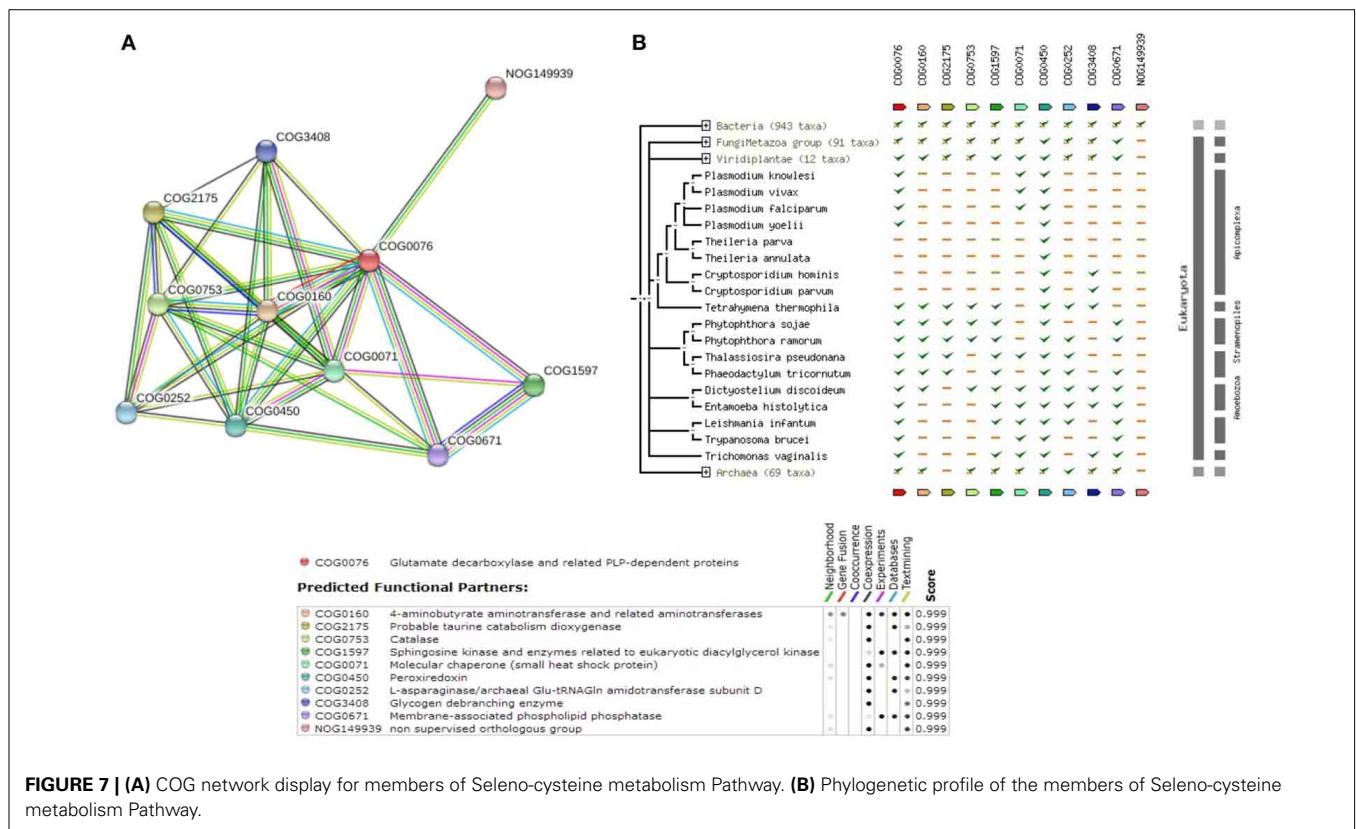
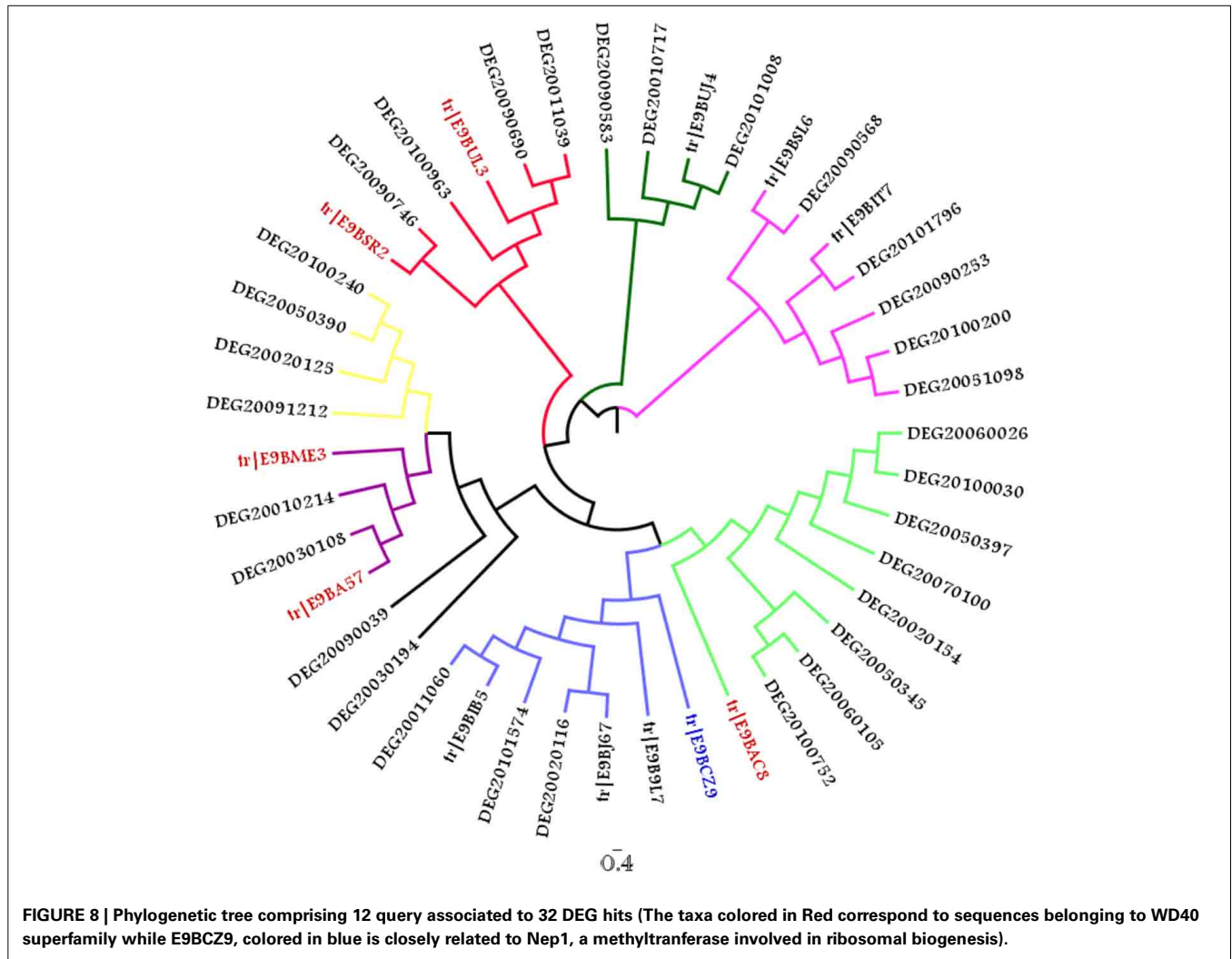


FIGURE 7 | (A) COG network display for members of Seleno-cysteine metabolism Pathway. **(B)** Phylogenetic profile of the members of Seleno-cysteine metabolism Pathway.



Selenocysteine. **Supplementary Table 6** shows that the enzymes in the pathway are well conserved in other species of *Leishmania* (with >90% coverage and similarity). **Supplementary Figure 6** shows the reference pathway obtained from Metacyc and steps involved in the pathway. **Figure 7A** depicts the String-DB interaction of E9BQF5 as queried against COG while; **Figure 7B** illustrates the phylogenetic profile of the query and other members of COG.

DEG ANALYSIS OF THE 105 SEQUENCES

In order to identify the presence of putative essential genes within the final dataset of 105 sequences, these proteins were queried against the Eukaryote Database of Essential Genes using JACKHMMER (with an e -value of 10^{-20}) and 93 associations were found for 23 query sequences. Upon removal of false positives based on query coverage (>75%), 12 true positives were found to have associations to 32 DEG sequences. A phylogenetic tree was constructed to identify closer clustering of these true positives with their corresponding DEG hits. MUSCLE present within MEGA 5.0 was used to align the sequences with 10 rounds of iterations while Maximum likelihood approach

with JTT (Jones et al., 1992) model and 100 bootstrap replications were used to build the tree. Fig-tree was used to visualize the tree provided as **Figure 8**. The unedited tree is shown in **Supplementary Figure 7** for further information related to bootstrap values.

Among the 12 true positives, 5 sequences belonged to the super family of WD40 repeats suggesting their roles in various protein-protein interactions. A String-DB based gene neighborhood and interaction analysis was performed using the remaining 7 true positives as query. Detailed analysis of E9BCZ9, suggests plausible roles in ribosomal biogenesis in eukaryotes. All of its interacting members are conserved across eukaryotes as displayed in **Figures 9A,B**. Careful homology based sequence analysis of E9BCZ9 suggests its close association with Nep1 (methyltransferase), a protein that plays roles in the ribosome biogenesis which is conserved across Eukaryotes and Achaea. Nep1 are a class of enzymes that catalyze methylation reaction during the steps of rRNA processing necessary for the generation of 40s ribosomal subunits (Eschrich et al., 2002).

The *Leishmania donovani* counterparts (sharing >90% similarity and coverage) for the *Leishmania infantum*

Supplementary Figure 6 | MetaCyc representation of Seleno-cysteine Metabolism. E9B9Y6 (part of 105 sequence dataset) is associated to Seleno-cysteine synthase (EC 2.9.1.2) in the pathway.

Supplementary Figure 7 | Unedited Phylogenetic tree with bootstrap values for 12 query sequences associated to 32 DEG hits.

Supplementary Table 1 | Information containing GO terms, Interproscan domain association, homologs picked during the BLAST step in the BLAST2GO analysis for the 105 sequences.

Supplementary Table 2 | Sequence wise association of pathway information obtained from KASS for 27 proteins.

Supplementary Table 3 | Table showing the sequence information in the Ubiquinone biosynthesis pathway along with the sequence information for other members of the genus *Leishmania*. LD, *Leishmania donovani*; DGR, *Drosophila grimshavi*.

Supplementary Table 4 | Table showing the sequence information in the fatty acid elongation (Mitochondria) pathway along with the sequence information for other members of the genus *Leishmania*. LD, *Leishmania donovani*; DGR, *Drosophila grimshavi*.

Supplementary Table 5 | Table showing the sequence information in the Fatty Acid elongation in ER pathway along with the sequence information for other members of the genus *Leishmania*. LD, *Leishmania donovani*; DGR, *Drosophila grimshavi*.

Supplementary Table 6 | Table showing the sequence information in the Seleno-cysteine metabolism pathway along with the sequence information for other members of the genus *Leishmania*. LD, *Leishmania donovani*; DGR, *Drosophila grimshavi*.

Supplementary Table 7 | Table showing the corresponding ortholog for each of the protein within *Leishmania donovani*. The String interaction for the E9BCZ9 was performed against *Leishmania infantum* since *L. donovani* is not available in String-DB.

REFERENCES

- Alvar, J., Croft, S. L., Kaye, P., Khamesipour, A., Sundar, S., and Reed, S.G. (2013). Case study for a vaccine against leishmaniasis. *Vaccine* 31(Suppl. 2), B244–B249. doi: 10.1016/j.vaccine.2012.11.080
- Alvar, J., Vélez, I.D., Bern, C., Herrero, M., Desjeux, P., Cano, J., et al. (2012). “Leishmaniasis Worldwide and Global Estimates of Its Incidence.” Edited by Martyn Kirk. *PLoS ONE* 7:e35671. doi: 10.1371/journal.pone.0035671
- Anishetty, S., Pulimi, M., and Pennathur, G. (2005). Potential drug targets in mycobacterium tuberculosis through metabolic pathway analysis. *Comput. Biol. Chem.* 29, 368–378. doi: 10.1016/j.compbiolchem.2005.07.001
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Bhunia, G. S., Kesari, S., Chatterjee, N., Kumar, V., and Das, P. (2013). Spatial and temporal variation and hotspot detection of kala-azar disease in Vaishali district (Bihar), India. *BMC Infect. Dis.* 13:64. doi: 10.1186/1471-2334-13-64
- Bonilla, M., Denicola, A., Novoselov, S. V., Turanov, A. A., Protasio, A., Izmendi, D., et al. (2008). Platyhelminth mitochondrial and cytosolic redox homeostasis is controlled by a single thioredoxin glutathione reductase and dependent on selenium and glutathione. *J. Biol. Chem.* 283, 17898–17907. doi: 10.1074/jbc.M710609200
- Boysen, K. E., and Matuschewski, K. (2011). Arrested oocyst maturation in plasmodium parasites lacking type II NADH:ubiquinone dehydrogenase. *J. Biol. Chem.* 286, 32661–32671. doi: 10.1074/jbc.M111.269399
- Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., et al. (2011). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40, D742–D753. doi: 10.1093/nar/gkr1014
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Croft, S. L., and Olliaro, P. (2011). Leishmaniasis chemotherapy—challenges and opportunities. *Clin. Microbiol. Infect.* 17, 1478–1483. doi: 10.1111/j.1469-0691.2011.03630.x
- Croft, S. L., Sundar, S., and Fairlamb, A. H. (2006). Drug resistance in leishmaniasis. *Clin. Microbiol. Rev.* 19, 111–126. doi: 10.1128/CMR.19.1.111-126.2006
- Da Silva, M. T. A., Silva-Jardim, I., and Thiemann, O. H. (2014). Biological implications of selenium and its role in trypanosomiasis treatment. *Curr. Med. Chem.* 21, 1772–1780. doi: 10.2174/0929867320666131119121108
- Desjeux, P. (2004). Leishmaniasis: current situation and new perspectives. *Comp. Immunol. Microbiol. Infect. Dis.* 27, 305–318. doi: 10.1016/j.cimid.2004.03.004
- Doerks, T., van Noort, V., Minguéz, P., and Bork, P. (2012). Annotation of the M. Tuberculosis hypothetical orfeome: adding functional information to more than half of the uncharacterized proteins. *PLoS ONE* 7:e34302. doi: 10.1371/journal.pone.0034302
- Eschrich, D., Buchhaupt, M., Kötter, P., and Entian, K. D. (2002). Nep1p (Emg1p), a novel protein conserved in eukaryotes and archaea, is involved in ribosome biogenesis. *Curr. Genet.* 40, 326–338. doi: 10.1007/s00294-001-0269-4
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2013). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176
- Guerin, P. J., Olliaro, P., Sundar, S., Boelaert, M., Croft, S.L., Desjeux, P., et al. (2002). Visceral leishmaniasis: current status of control, diagnosis, and treatment—a and a proposed research and development agenda. *Lancet Infect. Dis.* 2, 494–501. doi: 10.1016/S1473-3099(02)00347-X
- Hsieh, E. J., Gin, P., Gulmezian, M., Tran, U.C., Saiki, R., Marbois, B.N., et al. (2007). Saccharomyces cerevisiae Coq9 polypeptide is a subunit of the mitochondrial coenzyme Q biosynthetic complex. *Arch. Biochem. Biophys.* 463, 19–26. doi: 10.1016/j.abb.2007.02.016
- Jackson-Rosario, S.E., and Self, W.T. (2010). Targeting selenium metabolism and selenoproteins: novel avenues for drug discovery. *Metallomics* 2, 112–116. doi: 10.1039/b917141j
- Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi: 10.1093/nar/gkt1076
- Lira, R., Sundar, S., Makharia, A., Kenney, R., Gam, A., Saraiva, E., et al. (1999). Evidence that the high incidence of treatment failures in Indian kala-azar is due to the emergence of antimony-resistant strains of leishmania donovani. *J. Infect. Dis.* 180, 564–567. doi: 10.1086/314896
- Lobanov, A. V., Gromer, S., Salinas, G., and Gladyshev, V. N. (2006). Selenium metabolism in trypanosoma: characterization of selenoproteomes and identification of a kinetoplastid-specific selenoprotein. *Nucleic Acids Res.* 34, 4012–4024. doi: 10.1093/nar/gkl541
- Luo, H., Lin, Y., Gao, F., Zhang, C.-T., and Zhang, R. (2013). DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* 42, D574–D580. doi: 10.1093/nar/gkt1131
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229. doi: 10.1093/nar/gkq1189
- Monzote, L. (2009). Current treatment of leishmaniasis: a review. *Open Antimicrobial Agents J.* 1, 9–19.

- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Oladele, T. O., Sadiku, J. S., and Bewaji, C. O. (2011). *In silico* characterization of some hypothetical proteins in the proteome of plasmodium falciparum. *Centrepont J.* 17, 129–139.
- Rakotomanga, M., Saint-Pierre-Chazalet, M., and Loiseau, P. M. (2005). Alteration of fatty acid and sterol metabolism in miltefosine-resistant leishmania donovani promastigotes and consequences for drug-membrane interactions. *Antimicrob. Agents Chemother.* 49, 2677–2686. doi: 10.1128/AAC.49.7.2677-2686.2005
- Rezende, A. M., Folador, E. L., Resende, D. M., and Ruiz, J. C. (2012). Computational prediction of protein-protein interactions in leishmania predicted proteomes. *PLoS ONE* 7:e51304. doi: 10.1371/journal.pone.0051304
- Sharma, P., Patil, K., Sarang, D., and Shinde, P. (2013). *In silico* structure modeling and characterization of hypothetical proteins present in human fetal brain. *Int. J. Adv. Bioinform. Comput. Biol.* 1, 22–30.
- Silber, A. M., and Pereira, C.A. (2012). Assignment of putative functions to membrane 'Hypothetical Proteins' from the trypanosoma cruzi genome. *J. Membr. Biol.* 245, 125–129. doi: 10.1007/s00232-012-9420-z
- Singh, N., Kumar, M., and Singh, R.K. (2012). Leishmaniasis: current status of available drugs and new potential drug targets. *Asian Pac. J. Trop. Med.* 5, 485–497. doi: 10.1016/S1995-7645(12)60084-4
- Sundar, S., Agrawal, N., Arora, R., Agarwal, D., Rai, M., and Chakravarty, J. (2009). Short-course paromomycin treatment of visceral leishmaniasis in India: 14-day vs. 21-day treatment. *Clin. Infect. Dis.* 49, 914–918. doi: 10.1086/605438
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. *Science* 278, 631–637. doi: 10.1126/science.278.5338.631
- The UniProt Consortium. (2013). Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42, D191–D198. doi: 10.1093/nar/gkt1140

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 May 2014; accepted: 06 August 2014; published online: 26 August 2014.
Citation: Ravooru N, Ganji S, Sathyanarayanan N and Nagendra HG (2014) Insilico analysis of hypothetical proteins unveils putative metabolic pathways and essential genes in *Leishmania donovani*. *Front. Genet.* 5:291. doi: 10.3389/fgene.2014.00291
This article was submitted to Bioinformatics and Computational Biology, a section of the journal *Frontiers in Genetics*.
Copyright © 2014 Ravooru, Ganji, Sathyanarayanan and Nagendra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.