

Computational Annotation for Hypothetical Proteins of *Mycobacterium Tuberculosis*

S.Anandakumar and P.Shanmughavel^{1*}

¹Computational Biology and Bioinformatics Laboratory,
Department of Bioinformatics, Bharathiar University,
Coimbatore – 641046, TamilNadu, India

*Corresponding author: P. Shanmughavel, E-mail: shanvel_99@yahoo.com

Received August 28, 2008; Accepted November 10, 2008; Published December 26, 2008

Citation: Anandakumar S, Shanmughavel P (2008) Computational Annotation for Hypothetical Proteins of *Mycobacterium Tuberculosis*. J Comput Sci Syst Biol 1: 050-062. doi:10.4172/jcsb.1000004

Copyright: © 2008 Anandakumar S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

There is rising death of humans worldwide by reason of tuberculosis. The current sequencing of the *Mycobacterium tuberculosis* genome holds assure for the development of new vaccines and the design of new drugs. In this view, the functions prediction of genomic sequences for hypothetical proteins will invigorate our knowledge with reference to the identification of new drugs for tuberculosis. There are various function prediction methods available based on the on the assumption. The process accurate annotation for genes in newly sequenced genomes currently has been based on sequence similarity. In this work about 250 hypothetical proteins of *Mycobacterium tuberculosis* taken functions were predicted using Bioinformatics web tools, BLAST, INTERPROSCAN, PFAM and COGs.

Keywords: Tuberculosis; Hypothetical proteins; Sequence similarity; Bioinformatics web tools

Introduction

The current research on sequencing of the *Mycobacterium tuberculosis* genome holds assure for the development of new vaccines and the design of new drugs (Prachee Chakhaiyar and Hasnain, 2004) The functions for genomic sequences of hypothetical proteins are unknown because this is a protein whose being has been predicted (Edward et al., 2000). In depth learn of function prediction on such proteins will offer opportunity for novel applications and help the researchers to Identify new drug molecules for tuberculosis. *Mycobacterium tuberculosis* organism has totally 3887 number of proteins. In these proteins 1985 hypothetical proteins were present Out of the 250 hypothetical proteins taken for this work. All hypothetical proteins were analyzed for function prediction using Bioinformatics web tools such as BLAST, INTERPROSCAN, PFAM and COGs. The results indicates 100% confidence for only 86 proteins, with 75% confidence for 92 proteins and some proteins function could

not be predicted with much confidence (unknown function).

Methodology

Complete genome sequence of pathogenic bacteria *Mycobacterium tuberculosis* sequences were downloaded from the PIR Database (<http://pir.georgetown.edu/>) and NCBI Database (www.ncbi.nlm.nih.gov/). In complete genome sequence of *Mycobacterium tuberculosis*, 1985 hypothetical proteins were present. Only 250 hypothetical proteins of genome sequence were analyzed and then downloaded from the site (<http://www.ncbi.nih.gov/genomes/lproks.cgi>). Finally genomics sequences of each protein were submitted to functions prediction web tools such as NCBI-BLAST2 (Wendy et al., 2000), INTERPROSCAN (Zdobnov and Rolf Apweiler, 2001), PFAM (Bateman et al., 2002) and COG (Roman et al, 2000). The confidence level can be measured on the basis of above tools.

Access no	NCBI BLAST2	INTERPROSCAN	PFAM	COG	Percentage (%)
F70590	GTPase EngC.	ENG_C_GTPASE	Protein of unknown function, DUF258	Predicted GTPases	50%
G70650	Integral membrane protein possibly involved in chromosome condensation	Camphor resistance CrcB protein	CrcB-like protein	Integral membrane protein possibly involved in chromosome condensation	50%
A70759	Ubiquinone/menaquinone biosynthesis methyltransferase	UBIQUINONE/MENAQUINONE METHYLTRANSFERASE-RELATED	Methyltransferase	Methylase involved in ubiquinone/menaquinone biosynthesis	100%
H70797	Dihydroorotase (EC 3.5.2.3) (DHOase)	PEROXIDASE_1	Protein of unknown function	Uncharacterized ACR	25%
D70506	2-methylthioadenine synthetase	TRAM	TRAM domain	2-methylthioadenine synthetase	50%
E70627	Hydantoinase/oxoprolinase.	Hydantoinase B/oxoprolinase	Hydantoinase B/oxoprolinase	N-methylhydantoinase A	75%
H70685	Nicotinate-nucleotide adenyltransferase-like (EC 2.7.7.18)	CTP_transf_2	Cytidylyltransferase	Nicotinic acid mononucleotide adenyltransferase	50%
F70660	Holliday junction resolvase YqgF	Ribonuclease H-like	Uncharacterised protein family (UPF0081)	Predicted endonuclease involved in recombination (possible Holliday junction resolvase in Mycoplasmas and B. subtilis)	25%
B70738	Peptidase M22, glycoprotease	Peptidase_M22	Glycoprotease family	Inactive homologs of metal-dependent proteases, putative molecular chaperones	75%
G70591	IMP dehydrogenase/GMP reductase:TrkA-N:TrkA-C:Sodium/hydrogen exchanger	TrkA_C	TrkA-C domain	ligand-binding protein related to C-terminal domains of K ⁺ channels	75%
G70927	Nucleic acid binding protein, containing KH domain	Prokaryotic type KH domain (KH-domain type)	KH domain	Predicted RNA-binding protein (KH domain)	100%
B70903	ATP-binding protein.	Predicted P-loop kinase	P-loop ATPase protein family	Predicted P-loop-containing kinase	50%
F70977	LPPG:Fo 2-phospho-L-lactate transferase (EC 2.7.1.-).	F420_cofD: LPPG:Fo 2-phospho-L-lactate tran	Uncharacterised protein family UPF0052	Uncharacterized ACR	50%
E70729	General substrate transporter:Major facilitator superfamily MFS_1	MFS_1	Major Facilitator Superfamily	Permeases of the major facilitator superfamily	100%
A70792	GatB/YqeY family protein.	GatB_YqeY	GatB/YqeY domain	Uncharacterized ACR	75%
E70980	Iron-sulfur cluster biosynthesis protein SufE	SufE	Fe-S metabolism associated domain	SufE protein probably involved in Fe-S center assembly	100%
C70910	Rv0623-like transcription factor	PSK_trans_fac	Rv0623-like transcription factor	NO related COG (3 BeTs)	50%
C70740	Siderophore-interacting protein.	FAD_binding_9	Siderophore-interacting FAD-binding domain	Siderophore-interacting protein	75%
D70740	ABC transporter, transmembrane region:ABC transporter	ABC_TRANSPORTER_2	ABC transporter	ABC-type multidrug/protein/lipid transport system, ATPase component	100%
F70705	Drug resistance transporter EmrB/QacA subfamily	NAD_BINDING	Zinc-binding dehydrogenase	Permeases of the major facilitator superfamily	25%
G70796	Zinc-containing alcohol dehydrogenase, long-chain (EC 1.1.1.-).	NAD_BINDING	Zinc-binding dehydrogenase	Threonine dehydrogenase and related Zn-dependent dehydrogenases	75%
D70517	Zinc-containing dehydrogenase	NAD_BINDING	Alcohol dehydrogenase GroES-like domain	Threonine dehydrogenase and related Zn-dependent dehydrogenases	50%
H70617	Alcohol dehydrogenase, zinc-containing	ADH_zinc_N	Zinc-binding dehydrogenase	NADPH:quinone reductase and related Zn-dependent oxidoreductases	100%

A70667	Short-chain dehydrogenase/reductase SDR.	SDRFAMILY	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
A70597	Short-chain dehydrogenase/reductase SDR	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
B70640	Short-chain dehydrogenase/reductase SDR	adh_short	ADH_SHORT	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
B70569	7-alpha-hydroxysteroid dehydrogenase.	adh_short	NAD dependent epimerase/dehydratase family	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
B70649	Short-chain dehydrogenase/reductase SDR	SHORT-CHAIN DEHYDROGENASES/REDUCTASE	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
A70637	Short-chain dehydrogenase/reductase SDR precursor	SHORT-CHAIN DEHYDROGENASES/REDUCTASE	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
A70853	Alcohol dehydrogenase.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
C70863	IMP dehydrogenase/GMP reductase:NAD-dependent epimerase/dehydratase:Short-chain dehydrogenase/reductase SDR	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
C70814	Short-chain dehydrogenase/reductase SDR:Glucose/ribitol dehydrogenase	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
D70635	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100).	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
D70948	Short-chain dehydrogenase/reductase SDR.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
E70677	Short-chain dehydrogenase/reductase SDR.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
E70604	Oxidoreductase.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
D70707	7-ALPHA-HYDROXYSTEROID DEHYDROGENASE (EC 1.1.1.159).	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
G70743	Serine 3-dehydrogenase (EC 1.1.1.276).	adh_short	short chain dehydrogenase	Short-chain dehydrogenases of various substrate specificities	75%
D70953	Alcohol dehydrogenase.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
F70547	Fatty acyl-CoA reductase.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
G70617	17beta-estradiol dehydrogenase.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%

G70715	3-oxoacyl-(Acyl carrier protein) reductase (EC 1.1.1.100).	adh_short	short chain dehydrogenase	Short-chain dehydrogenases of various substrate specificities	75%
C70675	3-ketoacyl-acyl carrier protein reductase.	adh_short	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
F70677	2-deoxy-D-gluconate 3-dehydrogenase.	SHORT-CHAIN DEHYDROGENASES/REDUCTASE	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	100%
H70829	Dehydrogenase/ reductase (EC 1.1.1.-) 1 (EC 1.1.1.-).	NAD(P)-binding Rossmann-fold domains	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain	75%

alcohol dehydrogenases)					
H70890	Clavaldehyde dehydrogenase.	NAD(P)-binding Rossmann-fold domains	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	50%
H70805	3-oxoacyl-(Acyl-carrier-protein) reductase (EC 1.1.1.100).	SHORT-CHAIN DEHYDROGENASES/REDUCTASE	short chain dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	75%
H705231	NADPH-protochlorophyllide oxidoreductase.	GDHRDH	[No hits in Pfam]	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	25%
F70733	Aldo/keto reductase.	ALDKETRDTASE	Aldo/keto reductase family	Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)	75%
E70707	2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60) (Tartronate semialdehyde reductase) (TSAR).	3hydroxyisobu_dh	NAD binding domain of 6-phosphogluconate dehydrogenase	3-hydroxyisobutyrate dehydrogenase and related proteins	75%
C70645	D-3-phosphoglycerate dehydrogenase.	2-Hacid_dh_C	D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain	Phosphoglycerate dehydrogenase and related dehydrogenases	75%
F70796	Nucleoside-diphosphate-sugar epimerase	Epimerase	Male sterility protein	Nucleoside-diphosphate-sugar epimerases	75%
D70641	Glucose-methanol-choline oxidoreductase.	GMC_oxred_N	GMC oxidoreductase	Choline dehydrogenase and related flavoproteins	75%
E70961	Aldehyde dehydrogenase.	NAD-dependent aldehyde dehydrogenase	Aldehyde dehydrogenase family	NAD-dependent aldehyde dehydrogenases	100%
C70813	Dehydrogenase, E1 component.	E1_dh	Dehydrogenase E1 component	Thiamine pyrophosphate-dependent dehydrogenases, E1 component alpha subunit	100%
D70939	Succinate dehydrogenase (EC 1.3.99.1).	Succ_DH_flav_C	dehydrogenase flavoprotein C-terminal domain	Succinate dehydrogenase/fumarate reductase, flavoprotein subunits	100%
E70629	FAD dependent oxidoreductase.	DAO	FAD dependent oxidoreductase	Glycine/D-amino acid oxidases (deaminating)	75%
D70532	FAD-dependent pyridine nucleotide-disulphide oxidoreductase:Pyridine nucleotide-disulphide oxidoreductase dimerisation domain	Pyr_redox_dim	Pyridine nucleotide-disulphide oxidoreductase	Dihydrolipoamide dehydrogenase/glutathione oxidoreductase and related enzymes	50%
B70828	Dihydrolipoamide dehydrogenase.	PNDRDTASEI	Pyridine nucleotide-disulphide oxidoreductase	Dihydrolipoamide dehydrogenase/glutathione oxidoreductase and related enzymes	50%
C70524	Nitroreductase.	Nitroreductase	Nitroreductase family	Nitroreductase	100%
G70971	Nitroreductase	Nitroreductase	Nitroreductase family	Nitroreductase	100%
F70813	Multicopper oxidase.	Cu-oxidase	Cu-oxidase	multicopper oxidases	100%

G70948	Alpha/beta hydrolase fold.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
C70722	Alpha/beta hydroxylase.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
G70842	Alpha/beta hydrolase fold.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
D70552	Hydrolase, alpha/beta fold family precursor.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
D70733	Haloalkane dehalogenase (EC 3.8.1.5).	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	75%
E70607	Alpha/beta hydrolase fold.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
E70912	Alpha/beta hydrolase fold.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
B70722	Haloalkane dehalogenase (EC 3.8.1.5).	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	75%
F70532	Alpha/beta hydrolase fold.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
F70877	Alpha/beta hydrolase fold.	Abhydrolase_1	alpha/beta hydrolase fold	hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	100%
F70605	2, 3-dihydroxybiphenyl 1, 2-dioxygenase.	Glyoxalase	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily	Lactoylglutathione lyase and related lyases	50%
E70667	Ferredoxin reductase.	FAD/NAD(P)-binding domain	Pyridine nucleotide-disulphide oxidoreductase	Uncharacterized NAD(FAD)-dependent dehydrogenases	50%
C70957	Limonene 1,2-monoxygenase (EC 1.14.-.-).	Bac_luciferase	Luciferase-like monoxygenase	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	50%
D70636	Luciferase.	Bac_luciferase	Luciferase-like monoxygenase	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	75%
E70628	Luciferase.	Bac_luciferase	Luciferase-like monoxygenase	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	75%
B70710	Luciferase.	Bac_luciferase	Luciferase-like monoxygenase	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent	75%

				oxidoreductases	
G70615	Luciferase.	Bac_luciferase	Luciferase-like monooxygenase	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	75%
G70741	Luciferase.	Bacterial luciferase-like	[No hits in Pfam]	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	50%
G70665	Luciferase.	Bac_luciferase	Luciferase-like monooxygenase	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	75%
H70925	Luciferase.	Bac_luciferase	Luciferase-like monooxygenase	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	75%
F70593	Alkane-1-monooxygenase (EC 1.14.15.3).	FA_desaturase	Fatty acid desaturase	NO related COG	50%
G70735	DegT/DnrJ/EryC1/StrS aminotransferase.	DegT_DnrJ_EryC1	DegT/DnrJ/EryC1/StrS aminotransferase family	pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis	75%
H70977	N-6 DNA methylase.	N12N6MTFRASE	[No hits in Pfam]	Adenine-specific DNA methylase	50%
D70704	Amidinotransferase.	Amidinotransf	Amidinotransferase	N-Dimethylarginine dimethylaminohydrolase	75%
F70752	Acyltransferase.	Acyl_transf_3	Acyltransferase family	acyltransferases	100%
B70962	Acyltransferase.	Acyl_transf_3	Acyltransferase family	acyltransferases	100%
B70610	Glycosyl transferase, group 1.	Glycos_transf_1	Glycosyl transferases group	glycosyltransferases	100%
H70548	Glycosyl transferase.	Glycos_transf_1	Glycosyl transferases group 1	glycosyltransferases	100%
B70706	Histidine triad (HIT) protein.	HISTRIAD	HIT domain	Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases	100%
F70753	Histidine triad protein.	HISTRIAD	HIT domain	Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases	100%
D70571	Histidine triad (HIT) protein	Histidine triad hydrolase	HIT domain	Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family	100%

hydrolases					
D70899	RNA polymerase, omega subunit.	RNA polymerase omega subunit	RNA polymerase Rpb6	DNA-directed RNA polymerase subunit K/omega	50%
D70881	Dienelactone hydrolase.	DLH	Dienelactone hydrolase family	Dienelactone hydrolase and related enzymes	100%
E70945	Dienelactone hydrolase	DLH	Dienelactone hydrolase	Dienelactone hydrolase and related enzymes	100%
G70972	Hydrolase, haloacid dehalogenase-like family.	HADHALOGNASE	haloacid dehalogenase-like hydrolase	hydrolases of the HAD superfamily	100%
H70724	Metallophosphoesterase.	Metallophos	Calcineurin-like phosphoesterase	NO related COG (3 BeTs)	50%
F70788	Phosphoserine	Hydrolase	[No hits in Pfam]	Phosphoserine	50%

	phosphatase (EC 3.1.3.3)			phosphatase	
A70632	AAA family ATPase.	AAA	ATPase family associated with various cellular activities (AAA)	ATPases of the AAA+ class	50%
F70634	Beta-lactamase.	Beta-lactamase	Beta-lactamase	Beta-lactamase class C and other penicillin binding proteins	100%
C70743	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase.	CN_hydrolase	Carbon-nitrogen hydrolase	Predicted amidohydrolase	75%
E70804	Carbonic anhydrase.	Pro_CA	Carbonic anhydrase	Carbonic anhydrase	100%
A70747	Porphobilinogen deaminase.	PORPHOBILINOGEN DEAMINASE	Porphobilinogen deaminase, dipyrromethane cofactor binding domain	Porphobilinogen deaminase	100%
H70544	Phosphoglycerate mutase.	phosphoglycerate mutase	Phosphoglycerate mutase family	Fructose-2,6-bisphosphatase	75%
B70653	Phosphoglycerate mutase.	PGAM	Phosphoglycerate mutase family	Fructose-2,6-bisphosphatase	75%
C70577	Phosphoglycerate mutase.	PGAM	Phosphoglycerate mutase family	Fructose-2,6-bisphosphatase	75%
B70716	Chorismate mutase.	Chorismate mutase II	Chorismate mutase type II	Chorismate mutase	100%
A70971	RarD	ATP_bind_1	Conserved hypothetical ATP binding protein	Predicted GTPase	25%
E70867	Single-strand binding protein.	SSB	Single-strand binding protein family	Single-stranded DNA-binding protein	100%
B70807	PE-PGRS FAMILY PROTEIN.	HMG_COA_REDUCTASE_2	no hits	No hits	25%
E70917	PE-PGRS FAMILY PROTEIN.	PE_region_N	Pericardin like repeat	NO related COG	50%
A70514	PE-PGRS FAMILY PROTEIN.	EGGSHELL	No hits	NO related COG	25%
H70846	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
E70806	PE-PGRS FAMILY PROTEIN.	PFKB_KINASES_1	PE family	NO related COG	50%
D70807	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
F70806	PE-PGRS FAMILY PROTEIN.	PFKB_KINASES_1	PE family	NO related COG	50%
A70869	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
A70934	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
A70807	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
B70812	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
E70820	PE-PGRS FAMILY PROTEIN.	PE	PE family	NO related COG	75%
H70987	PE-PGRS FAMILY PROTEIN.	PE	PE family	NO related COG	75%
F70620	PE-PGRS FAMILY PROTEIN.	CABNDNGRPT	PE family	NO related COG	50%
D70835	PE-PGRS FAMILY PROTEIN.	PHOSPHOPANTHEINE	PE family	NO related COG	50%
F70824	PE-PGRS FAMILY PROTEIN.	PHOSPHOPANTHEINE	PE family	NO related COG	50%
D70954	PE-PGRS FAMILY PROTEIN.	PHOSPHOPANTHEINE	PE family	NO related COG	50%
H70820	PE-PGRS FAMILY PROTEIN.	NHL	NHL repeat	Uncharacterized ACR	50%
G70846	PE-PGRS FAMILY PROTEIN.	No hits	Pericardin like repeat	No hits	25%
D70916	PE-PGRS FAMILY PROTEIN.	TUBULIN	PE family	NO related COG	50%
H70839	PE-PGRS FAMILY PROTEIN.	PE	PE family	NO related COG	75%
E70983	PE_PGRS 33.	PFKB_KINASES_1	PE family	NO related COG	50%
E70768	PE-PGRS FAMILY	PE	no hits	NO related COG	50%

	PROTEIN.				
C70816	Transglycosylase-like precursor.	Transglycosylas	Transglycosylase-like domain	NO related COG	75%
E70756	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
F70571	PE-PGRS FAMILY PROTEIN.	PE	PE family	NO related COG	75%
C70720	PE-PGRS FAMILY PROTEIN.	PE	PE family	NO related COG	75%
F70573	Glycine-rich protein precursor.	HMG_COA_REDUCTASE_2	No hits	NO related COG	25%
B70893	PE-PGRS FAMILY PROTEIN.	PE_region_N	PE family	NO related COG	75%
D70956	Pseudouridine synthase, Rsu (EC 4.2.1.70).	No hits	No hits	NO related COG	25%
G70555	Sf3a2 protein	signal-peptide	No hits	NO related COG	25%
G70701	Acyl Carrier Protein, ACP.	PP-binding	Phosphopantetheine attachment site	Acyl carrier protein	50%
C70888	Phosphoesterase, PA-phosphatase related.	PA_PHOSPHATASE	PAP2 superfamily	Membrane-associated phospholipid phosphatase	100%
F70688	Sulfate transporter.	Sulfate_transp	Sulfate transporter family	Sulfate permease and related transporters (MFS superfamily)	100%
G70943	Sugar ABC transporter, permease protein	BPD_transp_1	Binding-protein-dependent transport system inner membrane component	ABC-type sugar transport systems, permease components	100%
F70943	Sugar ABC transporter, permease protein.	BPD_transp_1	Binding-protein-dependent transport system inner membrane component	Sugar permeases	100%
G70614	Glucokinase.	GLUCOKINASE-RELATED	ROK family	Transcriptional regulators	50%
H70853	Transcriptional regulator.	SUGAR_TRANSPORT_1	Bacterial transcriptional regulator	Transcriptional regulator	100%
B70686	Regulatory proteins, IclR.	HTHASNC	Bacterial transcriptional regulator	Transcriptional regulator	75%
C70858	TRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase precursor (EC 2.1.1.61).	tRNA_Me_trans	tRNA methyl transferase	Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain	100%
B70821	Signal transduction histidine kinase.	HIS_KIN	His Kinase A (phosphoacceptor) domain	Signal transduction histidine kinase	100%
H70622	MazG protein.	Nucleoside triphosphate pyrophosphohydrolase MazG	MazG nucleotide pyrophosphohydrolase domain	Predicted pyrophosphatase	75%
F70645	Regulatory protein, LuxR:Response regulator receiver.	HTH_LUXR_1	Response regulator receiver domain	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	75%
C70710	TRANSCRIPTIONAL REGULATOR, GNTR FAMILY	HTHGNTR	Bacterial regulatory proteins, gntR family	Transcriptional regulators	75%
A70555	Regulatory protein GntR, HTH.	HTH_GNTR	Bacterial regulatory proteins, gntR family	Predicted transcriptional regulators	75%
H70791	Anti-sigma factor antagonist	ant_ant_sig	STAS domain	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	75%
B70964	Anti-sigma factor antagonist.	ant_ant_sig: anti-anti-sigma factor	STAS domain	NO related COG	50%
F70611	DedA:Rhodanese-like	signal-peptide	No hits	Uncharacterized ACR	25%
H70559	MFS permease.	MFS	Major Facilitator Superfamily	Permeases of the major facilitator superfamily	100%
B70907	Sugar efflux transporter B.	MFS_1	Major Facilitator Superfamily	Arabinose efflux permease	50%
B70709	Drug resistance transporter EmrB/QacA subfamily	efflux_EmrB: drug resistance transporter	Major Facilitator Superfamily	Permeases of the major facilitator superfamily	50%

A70954	Drug resistance transporter EmrB/QacA subfamily.	efflux_EmrB	Major Facilitator Superfamily	Permeases of the major facilitator superfamily	50%
F70556	GTPase.	ELONGATNFCT	Elongation factor Tu GTP binding domain	Predicted membrane GTPase involved in stress response	50%
H70738	IstB-like ATP binding protein	IstB-like ATP-binding protein	IstB-like ATP binding protein	DNA replication protein	75%

G70562	IstB-like ATP binding protein	IstB-like ATP-binding protein	IstB-like ATP binding protein	DNA replication protein	75%
D70837	Trans-aconitate methyltransferase.	Methyltransferase type 11	Methyltransferase domain	SAM-dependent methyltransferases	100%
E70572	SAM-dependent methyltransferase.	Methyltransferase type 11	Methyltransferase domain	SAM-dependent methyltransferases	100%
B70527	Methyltransferase.	Generic methyltransferase	Methyltransferase domain	SAM-dependent methyltransferases	100%
F70964	Regulatory protein, ArsR precursor.	Bacterial regulatory protein, ArsR	Bacterial regulatory protein, arsR family	Predicted transcriptional regulators	100%
A70821	Response regulator receiver:Transcriptional regulatory protein, C-terminal	RESPONSE_REGULATORY	Response regulator receiver domain	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	100%
G70924	Transcriptional regulator.	Transcriptional regulatory protein, C-terminal	Transcriptional regulatory protein, C terminal	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	100%
B70810	Transcriptional regulator	Transcriptional regulatory protein, C-terminal	Transcriptional regulatory protein, C terminal	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	100%
F70801	Response regulator receiver:Transcriptional regulatory protein, C-terminal	Response regulator receiver	Response regulator receiver domain	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	100%
E70704	Transcriptional regulator (Lrp/AsnC family).	Bacterial regulatory proteins, AsnC/Lrp	AsnC family	Transcriptional regulators	75%
D70981	Regulatory proteins, AsnC/Lrp.	Bacterial regulatory proteins, AsnC/Lrp	AsnC family	Transcriptional regulators	75%
H70740	Transcriptional regulator, TetR family.	Bacterial regulatory protein, TetR	Bacterial regulatory proteins, tetR family	Transcriptional regulators	75%
B70827	Transcriptional regulator.	Bacterial regulatory protein, TetR	No hits	Transcriptional regulators	75%
G70903	Thioesterase superfamily	Thioesterase superfamily	Thioesterase superfamily	Predicted thioesterase	100%
F70608	Metabolite-proton symporter.	Citrate-proton symport	Sugar (and other) transporter	Permeases of the major facilitator superfamily	75%
C70504	Cobyrinic acid a,c-diamide synthase.	Cobyrinic acid a,c-diamide synthase	CobQ/CobB/MinD/ParA nucleotide binding domain	ATPases involved in chromosome partitioning	75%
F70595	Cobyrinic acid a,c-diamide synthase.	Cobyrinic acid a,c-diamide synthase	CobQ/CobB/MinD/ParA nucleotide binding domain	ATPases involved in chromosome partitioning	75%
F70702	Phage integrase.	Phage integrase	Phage integrase family	Integrase	100%
A70581	S-adenosyl-methyltransferase mraW (EC 2.1.1.1.-).	Bacterial methyltransferase	MraW methylase family	Predicted S-adenosylmethionine-dependent methyltransferase involved in cell envelope biogenesis	100%
G70685	Iojap-related protein	Iojap-related protein	Domain of unknown function DUF143	Uncharacterized ACR (homolog of plant Iojap proteins)	50%
H70577	Phospholipid-binding protein	TIGR00481: conserved hypothetical protein T	Phosphatidylethanolamine-binding protein	Phospholipid-binding protein	50%
H70756	Lysine exporter protein (LYSE/YGGA).	Lysine exporter protein (LYSE/YGGA)	LysE type translocator	Lysine efflux permease	100%
C70744	Lysine exporter protein (LYSE/YGGA).	Lysine exporter protein (LYSE/YGGA)	LysE type translocator	Lysine efflux permease	100%
A70897	Fructose-1,6-bisphosphatase II (EC 3.1.3.37).	Fructose-1,6-bisphosphatase, GlpX type	Bacterial fructose-1,6-bisphosphatase, glpX-encoded	Fructose-1,6-bisphosphatase/sedoheptulose 1,7-bisphosphatase and related proteins	100%
A70521	Haloacid dehalogenase-	Hydrolase	haloacid dehalogenase-like	Predicted hydrolases of the	100%

	like hydrolase.		hydrolase	HAD superfamily	
C70671	DNA methyltransferase.	N-6 Adenine-specific DNA methylase	Conserved hypothetical protein 95	N6-adenine-specific methylase	75%
E70795	UPF0233 protein Mb3743c	Collagen triple helix repeat	Uncharacterised BCR, YbaB family COG0718	Uncharacterized BCR	50%
A70768	Twin-arginine translocation protein TatA/E.	Twin-arginine translocation protein TatB	mttA/Hcf106 family	Sec-independent protein secretion pathway components	50%
G70567	Transposase.	Transposase_8	Transposase	Transposase	100%
E70845	Transcriptional regulator.	Bacterial regulatory protein, MerR	No hits	Predicted transcriptional regulators	100%
E70586	Hemolysin containing CBS domains.	CBS	CBS domain pair	Hemolysins and related proteins containing CBS domains	100%
B70664	Hemolysin containing CBS domains.	CBS	CBS domain pair	Hemolysins and related proteins containing CBS domains	100%
F70968	Peptide methionine sulfoxide reductase MsrB (EC 1.8.4.6).	Methionine sulfoxide reductase B	SelR domain	Conserved domain frequently associated with peptide methionine sulfoxide reductase	100%
D70549	Monooxygenase, FAD-binding.	FAD_binding_2	FAD dependent oxidoreductase	Dehydrogenases (flavoproteins)	75%
B70560	Universal stress protein.	Universal stress protein (Usp)	Universal stress protein family	Universal stress protein UspA and related nucleotide-binding proteins	100%
H70727		Uncharacterized conserved protein.	Uncharacterised conserved protein	Uncharacterised protein family UPF0047	Uncharacterized ACR
H70941	Cation efflux protein	Cation efflux protein	Cation efflux family	Predicted Co/Zn/Cd cation transporters	100%
C70531	Endoribonuclease L-PSP.	Endoribonuclease L-PSP	Endoribonuclease L-PSP	Putative translation initiation inhibitor	75%
A70684	CBS.	CBS	CBS domain pair	CBS domains	100%
A70573	CBS.	CBS	CBS domain pair	CBS domains	100%
C70964		Protein of unknown function UPF0060	Protein of unknown function UPF0060	Uncharacterised BCR, YnfA/UPF0060 family	Uncharacterized BCR
H70666	MOSC.	MOSC	MOSC domain	Uncharacterized BCR	75%
C70903		Uncharacterized conserved membrane-associated protein	Protein of unknown function UPF0052 and CofD	Uncharacterised protein family UPF0052	Uncharacterized ACR
D70508	Haloacid dehalogenase-like protein.	haloacid dehalogenase-like hydrolase	haloacid dehalogenase-like hydrolase	Predicted sugar phosphatases of the HAD superfamily	100%
F70959	TRNA (Guanine-N(7)-)-methyltransferase	Methyltransf_4	Putative methyltransferase	Predicted S-adenosylmethionine-dependent methyltransferase	100%
E70932	Protein yajQ.	Protein of unknown function DUF520	Protein of unknown function (DUF520)	Uncharacterized BCR	25%
A70800	Cytidine/deoxycytidylate deaminase, zinc-binding region.	CYT_DCMP_DEAMINASES	Cytidine and deoxycytidylate deaminase zinc-binding region	Cytosine/adenosine deaminases	100%
G70879	Beta-lactamase-like:RNA-metabolising metallo-beta-lactamase.	RNA-metabolising metallo-beta-lactamase	Metallo-beta-lactamase superfamily	Predicted hydrolase of the metallo-beta-lactamase superfamily	100%
G70525	Integral membrane protein.	transmembrane_regions	No hits	Predicted divalent heavy-metal cations transporter	50%
H70578	Alanine racemase, N-terminal.	Ala_racemase_N	No hits	Predicted enzyme with a TIM-barrel fold	50%
D70682	Cobalamin (Vitamin B12) biosynthesis CbiX protein.	Cobalamin (vitamin B12) biosynthesis CbiX	CbiX	Uncharacterized ACR	75%
F70626	Cobalamin (Vitamin B12) biosynthesis CbiX protein.	Cobalamin (vitamin B12) biosynthesis CbiX	CbiX	Uncharacterized ACR	75%
F70650	Camphor resistance CrcB protein.	Camphor resistance CrcB protein	CrcB-like protein	Integral membrane protein possibly involved in chromosome condensation	75%
G70812	Methyltransferase.	Methyltransf_11	Methyltransferase domain	SAM-dependent methyltransferases	100%
D70554	Phospholipid methyltransferase.	MET_TRANS	Methyltransferase domain	SAM-dependent methyltransferases	100%
H70900	Methyltransferase	Methyltransf_11	Methyltransferase domain	SAM-dependent methyltransferases	100%
B70901	Methyltransferase	Methyltransf_11	Methyltransferase domain	SAM-dependent	100%

F70502	NAD(+) kinase (EC 2.7.1.23).	ATP-NAD/AcoX kinase	ATP-NAD kinase	methyltransferases Predicted kinase	100%
A70774	Sua5/YciO/YrdC/YwIC.	Sua5/YciO/YrdC/YwIC	yrdC domain	Putative translation factor (SUA5)	75%
B70670	Glycosyltransferase gtfD.	Glycosyl transferase, family 2	Glycosyl transferase family 2	Glycosyltransferases involved in cell wall biogenesis	100%
H70693	Phosphoesterase, RecJ-like:Phosphoesterase, DHHA1.	Phosphoesterase, DHHA1	DHH family	Exopolyphosphatase-related proteins	75%
D70685	DegV family protein.	DegV: degV family protein	Uncharacterized protein, DegV family COG1307	Uncharacterized BCR	
D70702	UPF0301 protein yqgE.	Protein of unknown function DUF179	Uncharacterized ACR, COG1678	Putative transcriptional regulator	
B70669	Delta-1-pyrroline-5-carboxylate dehydrogenase 3.	Protein of unknown function DUF98	No hits	4-Hydroxybenzoate synthetase (chorismate lyase)	50%

B70839	Integral membrane protein.	Protein of unknown function UPF0118	Domain of unknown function DUF20	Predicted permease	50%
C70897	Predicted permease	Protein of unknown function UPF0118	Domain of unknown function DUF20	Predicted permease	50%
F70546	Glycosyl transferase, family 2.	Glycosyl transferase, family 2.	Glycosyl transferase family 2	Glycosyltransferases involved in cell wall biogenesis	100%
E70985	Rv0623-like transcription factor	Rv0623-like transcription factor	Rv0623-like transcription factor	NO related COG	75%
D70611	Rv0623-like transcription factor	Rv0623-like transcription factor	Rv0623-like transcription factor	NO related COG	75%
D70616	Peptide methionine sulfoxide reductase (EC 1.8.4.6).	Methionine sulfoxide reductase A	Peptide methionine sulfoxide reductase	Peptide methionine sulfoxide reductase	100%
F70731	Transcriptional regulator (Bacterial regulatory protein, LysR family).	Bacterial regulatory protein, LysR	Bacterial regulatory helix-turn-helix protein, lysR family	Transcriptional regulator	100%
D70561	D-alanyl-D-alanine carboxypeptidase.	Peptidase S13, D-Ala-D-Ala carboxypeptidase C	D-Ala-D-Ala carboxypeptidase 3 (S13) family	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)	100%
F70517	D-tyrosyl-tRNA(Tyr) deacylase.	D-tyrosyl-tRNA(Tyr) deacylase	D-Tyr-tRNA(Tyr) deacylase	D-Tyr-tRNA ^{Tyr} deacylase	100%
E70785	HesB/YadR/YfhF.	HesB/YadR/YfhF	Iron-sulphur cluster biosynthesis	Uncharacterized ACR	50%
D70725	Beta-lactamase-like.	Lactamase_B	Metallo-beta-lactamase superfamily	Zn-dependent hydrolases, including glyoxylases	75%
C70560	Beta-lactamase-like.	Beta-lactamase-like	Metallo-beta-lactamase superfamily	Zn-dependent hydrolases, including glyoxylases	75%
G70612	Beta-lactamase-like.	Beta-lactamase-like	Metallo-beta-lactamase superfamily	Zn-dependent hydrolases, including glyoxylases	75%
H70862	Beta-lactamase-like.	Beta-lactamase-like	Metallo-beta-lactamase superfamily	Zn-dependent hydrolases, including glyoxylases	75%
H70767	Sec-independent protein translocase protein tatC homolog.	Sec-independent periplasmic protein translocase	Sec-independent protein translocase protein (TatC)	Sec-independent protein secretion pathway component TatC	100%
F70684	DUF404.	Protein of unknown function DUF404, bacteria N-terminal	Domain of unknown function (DUF404)	Uncharacterized BCR	
G70870	Zinc metallopeptidase	Peptidase M20	Peptidase dimerisation domain	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases	75%
H70812	Proline imino-peptidase	Alpha/beta hydrolase fold-1	alpha/beta hydrolase fold	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	50%
F70674	Transferase hexapeptide	Hexapep	Bacterial	Carbonic	75%

	repeat		transferase hexapeptide (three repeats)	anhydrases/acetyltransferases, isoleucine patch superfamily	
H70585	Di-trans-poly-cis-decaprenylcistransferase (EC 2.5.1.31).	Di-trans-poly-cis-decaprenylcistransferase	Putative undecaprenyl diphosphate synthase	Undecaprenyl pyrophosphate synthase	50%
D70895	Di-trans-poly-cis-decaprenylcistransferase (EC 2.5.1.31).	Di-trans-poly-cis-decaprenylcistransferase	Putative undecaprenyl diphosphate synthase	Undecaprenyl pyrophosphate synthase	50%
C70570	SNO glutamine amidotransferase.	SNO glutamine amidotransferase	SNO glutamine amidotransferase family	Predicted glutamine amidotransferase involved in pyridoxine biosynthesis	75%

Table 1: functional genomics of *Mycobacterium tuberculosis*.

No. of Proteins	84	92	56	12	6
Percentage of similarity	100 %	75 %	50 %	25 %	0 %

Table 2: Percentage of similarity.

(In 250 proteins, 100% confidence levels present in eighty-four proteins, 75% in Ninety-two proteins, 50% in fifty-six proteins, 25% in twelve proteins and 0% in six proteins).

1. If the given four tools indicate the same functions then the confidence level were to be 100 percent.

2. If the given three tools indicate the same functions other is different functions then the confidence level were to be 75 percent.

3. If the given two tools indicate the same functions other two given different functions then the confidence level were to be 50 percent.

4. If the given four tools indicate different functions then the confidence level were to be 25 percent.

5. If the given tool doesn't indicate any functions then the confidence level were to be 0 percent

Result and Discussion

There is rising death of humans worldwide by reason of tuberculosis (Smith et al., 2004). Central goal of Bioinformatics is recognized as the major area of research to determining protein functions from their genomic sequences and to develop personalized medicine. Functional annotations of genomic sequences for hypothetical proteins

are of major importance in providing insights into their molecular functions and will help in the identification of new drugs for tuberculosis. Table 1 shows the functional genomics of *Mycobacterium tuberculosis* by using tools such as BLAST, INTERPROSCAN, PFAM and COG. *Mycobacterium tuberculosis* organism has totally 3887 number of proteins. In this 3887 proteins 1985 were hypothetical proteins from which 250 hypothetical proteins were retrieved for this study. Those hypothetical proteins were submitted to above tools, which help to determine the confidence level. Among 250 proteins, 244 proteins only were obtained the function such as DEHYDROGENASES/REDUCTASE, HYDROLASES, LUCIFERASES & METHYL TRANSFERASES were in more in number.

References

- Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, et al. (2002) The Pfam protein families database. *Nucleic Acids Res* 30: 276-80. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
- Edward E, Gilliland GL, Herzberg O, Moulton J, Orban J, et al. (2000) Biological function made crystal clear — annotation of hypothetical proteins via structural

- genomics. *Current Opinion in Biotechnology* 11: 25-30.
» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
3. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* 96: 4285-4288. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 4. Prachee C, Hasnain SE (2004) Defining the Mandate of Tuberculosis Research in a Postgenomic Era. *Medicinal principles and practice* 13: 177-184. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 5. Roman L, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* 28: 33-36. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 6. Smith CV, Sharma V, Sacchettini JC (2004) TB drug discovery: addressing issues of persistence and resistance. *Tuberculosis* 84: 45-55. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 7. Wendy B, van den Broek A, Camon E, Hingamp P (2000) The EMBL Nucleotide Sequence Database. *Nucleic Acids Research*. 28: 19-23. » [CrossRef](#) » [Google Scholar](#)
 8. Zdobnov EM, Rolf A (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847-848. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)