Open access *Mycobacterium tuberculosis* clone repository: a community resource by OSDD members

Mycobacterium tuberculosis (*Mtb*), the causative organism of tuberculosis (TB), is often described as one of the most successful pathogens. With the release of the complete genome sequence of *Mtb*, a complete and comprehensive understanding of the pathogen is eagerly anticipated, so that novel strategies for tackling the disease could be evolved. The Open Source Drug Discovery (OSDD) initiative of the Council of Scientific and Industrial Research (CSIR) has conceptualized and implemented integrated platforms where innovation in drug discovery can be taken forward through a process-oriented activity for discovering novel therapeutic strategies against TB[1].

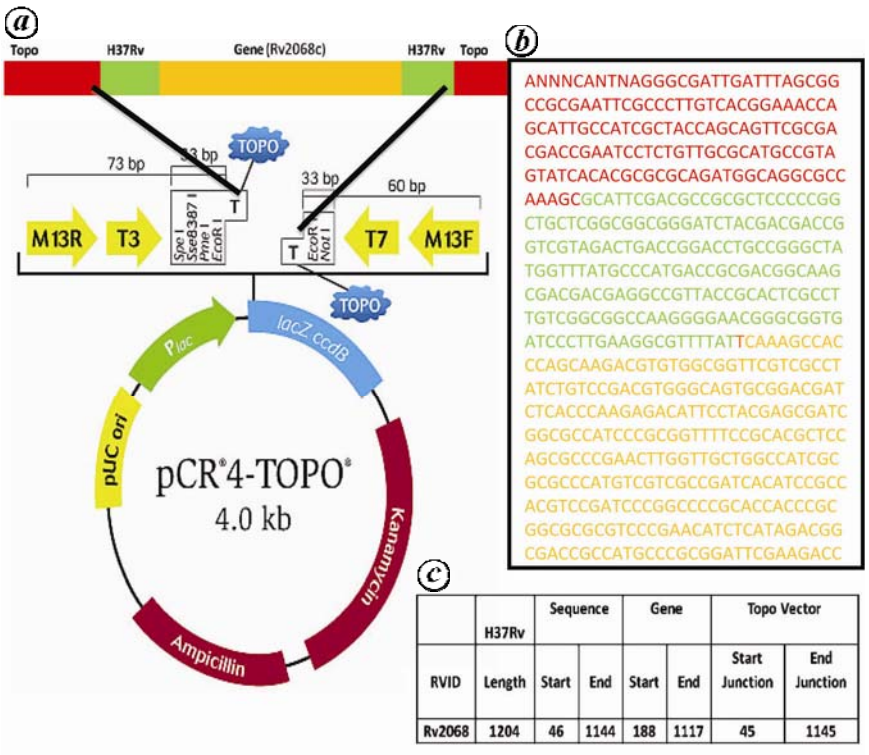
We realize that it is imperative to establish a centralized repository of *Mtb* gene clones, which would serve as a resource for researchers working on the OSDD platform, and for those working on various other aspects of TB biology. Using homogenous starting material for any type of downstream experiments will allow comparison of data generated across various laboratories and also ensure reproducibility, which is often a matter of concern in drug discovery research.

Under the 'Connect to Decode' or C2D collaborative programme of the OSDD, a comprehensive re-annotation of the *Mtb* genome was undertaken[2]. Systems-level approach was employed for identification of potential drug targets in *Mtb*. This was accomplished through an on-line collaborative approach which successfully generated the most comprehensive protein–protein functional interaction map and reactome of *Mtb*[3]. Experimental validation of these central proteins and metabolic chokepoints is crucial for taking forward these proteins further in the drug discovery pipeline. To this end, the first step was to clone these genes and make them available for experimental validation of their significance, and for assay development for screening against potent inhibitors. Also, the OSDD community members have listed reported drug targets in *Mtb*, which were also selected for cloning. Altogether, we initiated cloning of 200 selected genes of *Mtb* strain $H_{37}R_v$ with active OSDD community participation.

The activity was broadly divided into (i) primer design and validation, (ii) cloning of target *Mtb* genes and (iii) clone mapping. The involvement of over 100 students from universities and colleges across India to achieve this task in a unique crowd-sourcing approach, in addition to dry and wet laboratory work, makes this significant. The project participants were screened through an open call followed by on-line assignments.

Dedicated standard operating procedures (SOPs) and protocols were prepared along with worksheet formats for reporting results. This was done to ensure uniformity of procedures and adherence to quality parameters. These documents were uploaded on the project site on the OSDD server and were accessed by the participating individuals. To ensure experimental consistency, the same raw material (*Mtb* strain $H_{37}R_v$), reagents and plastic ware in the form of custom cloning kits were distributed to the cloning centres. For the purpose of training selected participants from geographically different locations, training documents including tutorials, videos and assignments were made available on the project website.

The purpose of this project was two-pronged: (i) to train students in basic molecular biology techniques and (ii) to generate *Mtb* gene clones. The idea was to design simple experiments which could be easily executed by students and yet have the robustness to generate good-quality clones. Full-length genes were amplified by GoTaq® Flexi DNA Polymerase (Cat. No. M8295, Promega Corporation) using primers present on the flanking sequences. The amplified fragments were cloned using the commercially available TOPO® TA Cloning® Kit for Sequencing (Cat. No. K4580-01, Invitrogen Corporation) following recommended protocols from the manufacturers. Clone validation was done using



	H37Rv	Sequence		Gene		Topo Vector	
RVID	Length	Start	End	Start	End	Start Junction	End Junction
Rv2068	1204	46	1144	188	1117	45	1145

Figure 1. Example of a sequence-validated clone which was mapped for position in the vector. ***a***, Line diagram showing the origin and order of the sequences mapped on the clone. ***b***, Sequence data obtained from the clone. The sequence shown in red font is the vector sequence, green font is that of the $H_{37}R_v$ flanking region and the sequence in yellow font is the ORF of the desired *Mtb* gene. ***c***, Position of the gene and flanking regions in the *Mtb* clone.

Table 1. Validated *Mycobacterium tuberculosis* clones

RvIDs	Gene orientation in clone	Function[5,*]
Rv0046c	Reverse	Myo-Inositol-1-Phosphate Synthase Ino1
Rv0079	Forward	Possible Cytoplasmic Translation Factor
Rv0117	Reverse	Oxidative Stress Response Regulatory Protein
Rv0120c	Reverse	Probable Elongation Factor G
Rv0129c	Forward	Trehalose Dimycolyl Transferase (Antigen 85 Complex C)
Rv0142	Forward	Putative Transcriptional Regulator
Rv0170	Forward	Mce-Family Protein Mce1B
Rv0175	Reverse	Mce Associated Membrane Protein
Rv0429c	Forward	Probable Polypeptide Deformylase (Formylmethionine Deformylase)
Rv0445c	Forward	Probable Alternative RNA Polymerase Sigma Factor SigK
Rv0451c	Forward	Probable Conserved Membrane Protein MmpS4
Rv0465c	Forward	Transcriptional Repressor
Rv0466	Forward	Probable Acyl-ACP Thioesterase
Rv0470c	Forward	Mycolic Acid Synthase PcaA
Rv0475	Forward	Iron-Regulated Heparin Binding Hemagglutinin HbhA (Adhesin)
Rv0482	Forward	Probable UDP-N-Acetylenolpyruvoylglucosamine Reductase MurB (UDP-N-Acetylmuramate Dehydrogenase)
Rv0491	Reverse	Two-Component Sensory Transduction Protein RegX3 (Transcriptional Regulatory Protein) (Probably LuxR-Family)
Rv0533c	Forward	3-Oxoacyl-[Acyl-Carrier-Protein] Synthase III FabH
Rv0534c	Forward	1,4-Dihydroxy-2-Naphthoate Octaprenyltransferase MenA (DHNA-Octaprenyltransferase)
Rv0636	Reverse	(3r)-Hydroxyacyl-ACP Dehydratase Subunit HadB
Rv0639	Reverse	Probable Transcription Antitermination Protein NusG
Rv0645c	Reverse	Methoxy Mycolic Acid Synthase 1
Rv0706	Forward	50S Ribosomal Protein L22 RplV
Rv0723	Reverse	Probable 50S Ribosomal Protein L15
Rv0757	Forward	Possible Two Component System Response Transcriptional Positive Regulator PhoP
Rv0981	Reverse	Mycobacterial Persistence Regulator MrpA
Rv0982	Forward	Two Component Sensor Kinase MprB
Rv1018c	Reverse	Glucosamine-1-Phosphate *N*-Acetyltransferase
Rv1201c	Forward	Tetrahydrodipicolinate-*N*-Succinyltransferase DapD
Rv1221	Forward	Alternative RNA Polymerase Sigma Factor SigE
Rv1222	Reverse	Anti-Sigma Factor RseA
Rv1256c	Forward	Cytochrome P450 130 Cyp130
Rv1258c	Reverse	Probable Conserved Integral Membrane Transport Protein
Rv1267c	Forward	Probable Transcriptional Regulatory Protein EmbR
Rv1293	Forward	Probable Diaminopimelate Decarboxylase LysA (DAP Decarboxylase)
Rv1305	Reverse	Probable ATP Synthase C Chain AtpE (Lipid-Binding Protein) (Dicyclohexylcarbodiimide-Binding Protein)
Rv1338	Reverse	Glutamate Racemase
Rv1390	Forward	Probable DNA-Directed RNA Polymerase (Omega Chain) RpoZ
Rv1409	Reverse	5-Amino-6-(5-Phosphoribosylamino) Uracil Reductase
Rv1419	Forward	sMTL-13 (13 kda Lectin) [1326]
Rv1471	Forward	Thioredoxin TrxB1 (Thioredoxin Reductase)
Rv1553	Forward	Fumarate Reductase FrdB (Fumarate Dehydrogenase) (Fumaric Hydrogenase)
Rv1596	Reverse	Nicotinate–Nucleotide Pyrophosphorylase
Rv1631	Reverse	Dephospho-CoA Kinase
Rv1698	Forward	Outer Membrane Protein
Rv1755c	Reverse	Phospholipase-C Gene D
Rv1876	Forward	Probable Bacterioferritin BfrA
Rv1886c	Forward	Mycolyltransferase 85B
Rv1980c	Forward	Immunogenic Protein Mpt64 (Antigen Mpt64/Mpb64)
Rv2058c	Reverse	50S Ribosomal Protein L28-2
Rv2068c	Forward	Class A Beta-Lactamase BlaC
Rv2069	Forward	Probable RNA Polymerase Sigma Factor, ECF Subfamily, SigC
Rv2121c	Forward	Probable ATP Phosphoribosyltransferase HisG
Rv2130c	Forward	Cysteine: 1D-Myo-Inosityl 2-Amino-2-Deoxy-D-Glucopyranoside Ligase MshC
Rv2145c	Forward	Conserved Hypothetical Protein Wag31
Rv2150c	Reverse	Cell Division Protein FtsZ
Rv2151c	Reverse	Possible Cell Division Protein FtsQ
Rv2217	Forward	Octanoyl-[Acyl Carrier Protein]-Protein Acyltransferase
Rv2245	Reverse	3-Oxoacyl-[Acyl-Carrier Protein] Synthase 1 KasA (Beta-Ketoacyl-ACP Synthase) (KASI)
Rv2276	Forward	Cytochrome P450 121 Cyp121

(*Contd*)

Table 1. (*Contd*)

RvIDs	Gene orientation in clone	Function[5,*]
Rv2392	Forward	3′-Phosphoadenosine 5′-Phosphosulfate Reductase
Rv2445	Forward	Nucleoside Diphosphate Kinase NdkA (Ndk) (Ndp Kinase) (Nucleoside-2-P Kinase)
Rv2455c	Forward	Oxidoreductase Alpha Subunit
Rv2461c	Forward	Endopeptidase CLP
Rv2518c	Reverse	Probable L,D-transpeptidase LdtB
Rv2612c	Forward	Probable PI Synthase PgsA1 (Phosphatidylinositol Synthase) (CDP-Diacylglycerol–Inositol3-Phosphatidyltransferase)
Rv2623	Reverse	Stress Related ATP Binding Growth Regulating Protein, Universal Stress Protein Homolog
Rv2710	Forward	RNA Polymerase Sigma Factor SigB
Rv2719	Reverse	Possible Conserved Membrane Protein
Rv2720	Reverse	Repressor LexA
Rv2753c	Reverse	Dihydrodipicolinate Synthetase
Rv2763c	Forward	Dihydrofolate Reductase DfrA (DHFR) (Tetrahydrofolate Dehydrogenase)
Rv2794c	Reverse	Phosphopantetheinyl Transferase PptT
Rv2841c	Forward	Transcription Factor
Rv2861c	Forward	Methionine Aminopeptidase
Rv2869c	Forward	Zinc Dependent Metalloprotease
Rv2870c	Forward	1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase
Rv2882c	Forward	Ribosome Recycling Factor Frr (Ribosome Releasing Factor) (RRF)
Rv2904c	Reverse	50S Ribosomal Protein L19
Rv2939	Reverse	Acyltransferase PapA5
Rv2981c	Forward	Probable D-Alanine-D-Alanine Ligase DdlA
Rv3089	Reverse	Probable Chain-Fatty-Acid-CoA Ligase FadD13
Rv3100	Reverse	Probable SSRA-Binding Protein SmpB
Rv3132c	Reverse	Two Component Sensor Histidine Kinase DevS
Rv3133c	Forward	Two Component Transcriptional Regulatory Protein DevR
Rv3223c	Forward	RNA Polymerase Sigma Factor RpoE
Rv3246c	Forward	Two Component Sensory Transduction Transcriptional Regulatory Protein MtrA
Rv3247c	Reverse	Probable Thymidylate Kinase Tmk (dTMP Kinase) (Thymidylic Acid Kinase) (TMPK)
Rv3286c	Forward	Alternate RNA Polymerase Sigma Factor SigF
Rv3287c	Forward	Anti-Sigma Factor RsbW (Sigma Negative Effector)
Rv3307	Forward	Probable Purine Nucleoside Phosphorylase DeoD (Inosine Phosphorylase) (PNP)
Rv3310	Forward	Secreted Acid Phosphatase
Rv3315c	Reverse	Metal-Dependent Homotetrameric Cytidine Deaminase
Rv3372	Forward	Trehalose-6-Phosphate Phosphatase OtsB2
Rv3423c	Reverse	Alanine Racemase Alr
Rv3443c	Forward	50S Ribosomal Protein L13
Rv3456c	Forward	50S Ribosomal Protein L17
Rv3465	Reverse	dTDP-4-Dehydrorhamnose 3,5-Epimerase RmlC (dTDP-4-Keto-6-Deoxyglucose 3,5-Epimerase) (dTDP-L-Rhamnose Synthetase) (Thymidine Diphospho-4-Keto-Rhamnose 3,5-Epimerase)
Rv3526	Forward	3-Ketosteroid 9-alpha-Hydroxylase
Rv3571	Forward	Possible Hemoglobine-Related Protein HMP
Rv3582c	Forward	2-C-Methyl-D-Erythritol 4-Phosphate Cytidylyltransferase
Rv3583c	Reverse	Transcription Factor
Rv3584	Reverse	Lipoprotein
Rv3588c	Forward	Beta-Carbonic Anhydrase CanB
Rv3604c	Reverse	Probable Conserved Transmembrane Protein Rich In Alanine And Arginine And Proline
Rv3605c	Forward	Probable Conserved Secreted Protein
Rv3720	Forward	Cyclopropane-Fatty-Acyl-Phospholipid Synthase
Rv3782	Forward	Galactofuranosyl Transferase
Rv3793	Reverse	Arabinosyltransferase EmbC
Rv3802c	Forward	Thioesterase
Rv3843c	Forward	Probable Conserved Transmembrane Protein
Rv3849	Forward	DNA-Binding Transcription Factor, ESX-1 transcriptional regulatory protein EspR
Rv3875	Forward	6 kDa Early Secretory Antigenic Target EsxA (ESAT-6)
Rv3911	Forward	RNA Polymerase Sigma Factor SigM
Rv3923c	Forward	Ribonuclease P Protein Component RnpA
Rv3924c	Reverse	50S Ribosomal Protein L34 RpmH

*Updated annotations may be seen at http://crdd.osdd.net/servers/ipw/

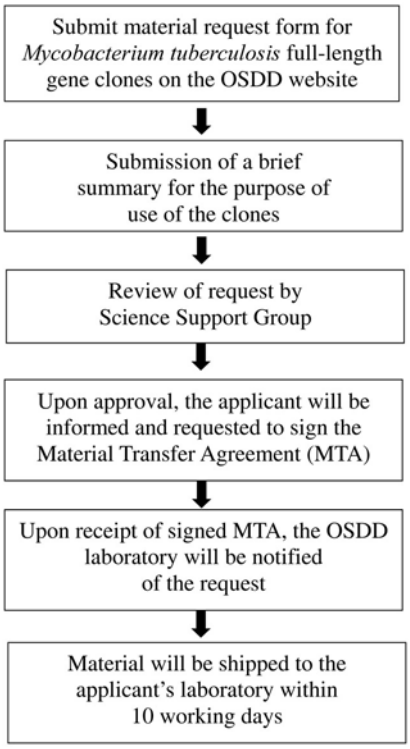


Figure 2. Process flow for requesting *Mtb* clones from the repository.

colony PCR and confirmed using standard sequencing methods. The clone sequences were then mapped onto the *Mtb* genome and vector sequence. This exercise generated a map of the respective clones (Figure 1). A total of 116 *Mtb* genes were cloned and validated by groups of students and cross-validated by scientists. The standardized protocols, gel photographs and sequence verification for each clone have been submitted by participants on the OSDD portal as open laboratory notebooks. This allows anybody to check the experimental procedures and results for each clone. The confirmed clones, transformed in *Escherichia coli* DH5α strain were preserved as glycerol stocks. These stocks are being used for DNA extraction and long-term storage for distribution among the scientific community. A complete list of these clones with their respective ordering codes is given in Table 1.

The need for bio-safety laboratory level-3 facilities for culturing *Mtb* serves as a limitation for researchers. The clone repository created by the CSIR–OSDD will enable researchers to perform experiments on *Mtb* as opposed to restricting themselves to *Mycobacterium smegmatis*, the preferred avirulent form of the *Mycobacterium* family more suited for research in laboratories with lower bio-safety adherence.

The need for centralized *Mtb* strain and clone repositories has already been identified for catalysing research in the domain[4]. Pathogen Functional Genomics Resource Center (PFGRC) at J. Craig Venter Institute (JCVI) houses a similar clone repository for *Mtb* genes in the Invitrogen Gateway system. The maintenance and distribution of these clones is managed by the BEI Resources. This clone resource is undoubtedly more extensive as it contains >3000 gene clones of *Mtb*. The clones provided by OSDD would however enable a faster distribution network for Indian investigators, besides providing the flexibility of choosing between different vector systems for downstream sub-cloning activities.

A Science Support Group, comprising scientific staff will periodically review the content and quality of the repository and add full-length clones of *Mtb*, if additional genes show up as potential drug targets. This group will also review requests for clones. DNA samples would be distributed to the scientific community free of cost along with the accompanying information in accordance to the Material Transfer Agreement (MTA) of OSDD. The material request process is explained in Figure 2. The material request form and MTA are available to registered users on the OSDD website. These clones will serve as a valuable resource for downstream protein characterization work on these genes to study their role in the drug discovery pipeline.

1. Bhardwaj, A. *et al.*, *Tuberculosis*, 2011, **91**(5), 479–486.
2. http://c2d.osdd.net
3. Vashisht, R. *et al.*, *PLoS One*, 2012, **7**(7), e39808.
4. Vincent, V. *et al.*, *Int. J. Tuberc. Lung Dis.*, 2012, **16**(1), 24–31.
5. Vashisht, R. and Bhardwaj, A., *Mol. Biosyst*., 2013, **9**(7), 1584–1593.

ACKNOWLEDGEMENTS. We thank the students, staff and faculty of SASTRA University, Thanjavur; CSIR-IGIB, New Delhi and the Centre for Genomic Application (TCGA), New Delhi for support during different stages of this work. The administrative support rendered by SASTRA University and TCGA, New Delhi is acknowledged. We also thank the Council for Scientific and Industrial Research, New Delhi for financial support (Grant No. HCP001).

Received 6 July 2013; revised accepted 27 August 2013

R. S. SANTHOSH[1,†]
AMIT TUTEJA[2,†]
SOUMASHREE BASU[2]
AISHWARYA SUBRAMANIAN[1]
S. SUNDHARRAMAN[1]
ANITA KUMARI[2]
M. VAIDHYANATHAN[1]
ATHUL MOHAN[1]
R. MAHALAKSHMI[1]
S. SHANKARANARAYANAN[1]
SHREYANS CHORDIA[1]
G. DIVYA[1]
AKANSHA VIRMANI[2]
KANDARP JOSHI[3]
SISHIR KUMAR YADAV[4]
LAKSHMI NARASIMHA CHARYULU[1]
R. SAVITHA[1]
C. AJITHAVALLI[1]
DEEPAK RAJENDRAN[1]
KAAMINI RAITHATHA[3]
SARFARAZ ALAM[5]
SWATI GANDHI[3]
DEEPA KAMATH[6]
SURI SOWMYA[7]
TASADUQ HUSSAIN WANI[2]
SYED SHARIQUE AHMAD[2]
OSDD CONSORTIUM[8]
M. VIJAYALAKSHMI[1,*]
ANSHU BHARDWAJ[8,*]
SWATI SUBODH[2,8,*]

[1]*School of Chemical and Biotechnology, SASTRA University, Thanjavur 613 401, India*
[2]*The Centre for Genomic Application (an IGIB-IMM collaboration), New Delhi 110 020, India*
[3]*M. S. University, Baroda 900 020, India*
[4]*MLS University, Udaipur 313 001, India*
[5]*Institute of Biomedical Sciences, Kolkata, India*
[6]*Thadomal Shahani Engineering College, Mumbai 400 050, India*
[7]*Jawaharlal Nehru Technological University, Hyderabad 500 085, India*
[8]*Council for Scientific and Industrial Research – Open Source Drug Discovery Unit, Anusandhan Bhawan, 2, Rafi Marg, Delhi 110 001, India*
**For correspondence.*
e-mail: mviji@sastra.edu; anshub@osdd.net; swati@osdd.net
[†]*Equally contributing authors.*