Serine threonine protein kinases of mycobacterial genus: phylogeny to function

Azeet Narayan,1,2 Preeti Sachdeva,1 Kirti Sharma,1 Adesh K. Saini,1 Anil K. Tyagi,2 and Yogendra Singh1

1Institute of Genomics and Integrative Biology, Delhi; and 2Department of Biochemistry, University of Delhi South Campus, New Delhi, India

Submitted 9 October 2006; accepted in final form 29 November 2006

The genus Mycobacterium includes acid fast bacilli with more than 90 species, including slow-growing pathogens of Mycobacterium tuberculosis complex [M. tuberculosis (Mt), M. bovis (Mb), M. africanum, M. microti, and M. canetti], M. avium-intracellulare complex (M. intracellulare and M. avium), M. marinum (Mma), M. leprae (Mle), and nonpathogenic, fast-growing species such as M. smegmatis (Msm). Human tuberculosis is caused mainly by Mt; however, the etiological agent of bovine tuberculosis, Mbo can also cause human disease. Two genetically distinct species within M. avium-intracellulare complex are M. avium, which tends to infect human immunodeficiency virus-infected patients, and M. intracellulare, more common among immunocompetent individuals (27). Unlike M. intracellulare, which relates to a single species, M. avium is separated into three subspecies; M. avium ssp. avium, a major opportunistic pathogen leading to a disseminated disease among terminal acquired immune deficiency syndrome patients; M. avium ssp. paratuberculosis (Map), causing Johne’s disease among ruminants and implicated in Crohn’s disease among humans; and M. avium ssp. silvaticum, a pathogen affecting birds that may cause chronic enteritis among calves but has not yet been shown to be associated with human disease (27). Mle, a close relative of tubercle bacillus, causes leprosy, a chronic human neurological disease. Mma, the mycobacterium most closely related to the members of Mt complex, causes systemic chronic tuberculosis in ectotherms and peripheral granulomatous disease in humans (9). Msm, a fast-growing saprophytic mycobacterium, is used as a heterologous model system to study gene expression and virulence mechanisms of Mt.

Microbial genomics has provided a foundation for understanding the basic biological processes, protein-protein interactions, and host-pathogen interactions and for designing novel antimicrobial compounds and vaccines. A recent study using comparative analysis of various microbial species demonstrated that the microbial genome is a dynamic entity shaped by multiple forces (19). Therefore, comparative functional genomics of the sequenced mycobacterial species can serve as an important tool in predicting the physiological importance of certain distinct set of genes. Six mycobacterial species, including pathogenic and nonpathogenic, have been sequenced: Mt H37Rv (7), Mt CDC1551 (18), Mt F11, and Mt C (www.broad.mit.edu/seq/msc/); Mle (8); Msm mc²155 (www.tigr.org); Mbo AF212297 (20); Map K-10 (35); and Mma (www.sanger.ac.uk).

Serine/threonine protein kinases (STPKs) have been shown to be important virulence factors in various pathogenic bacteria (11). Reversible protein phosphorylation by these STPKs plays a key role in regulating plethora of cellular processes including stress response, regulation of cell cycle, and development. This regulatory phenomenon is unambiguously preserved during the course of evolution in all forms of life (31, 34). The genome sequence of Mt H37Rv revealed the presence of 11 STPKs (7). Most of the mycobacterial STPKs have been biochemically characterized (3, 6, 22, 23, 33, 37–40, 46, 52, 54, 61), and some of them have been investigated for their physiological roles as well (10, 12, 14, 15, 30, 41, 55, 61), and some of these kinases have been implicated in the pathogenesis and survival of the tubercle bacillus within host (45, 63).

In this study, a comprehensive approach has been taken to identify specific functional ortholog(s) of each Mt kinase in mycobacterial species. This is particularly important in the context of various recent reports giving evidence of in vitro phosphorylation of a substrate by multiple kinases, implying that the signaling in mycobacterium could be more complex.
than expected (23, 30, 37, 61). This study, in particular, would add another dimension to the in vitro phosphotransfer studies in identification of physiological substrates and thereby function, while resolving the existing complexities. Another practical utility of this approach would be to help determine whether STPKs are discrete genomic units or instead a continuum across various mycobacterial species that varies according to a number of properties, including but not limited to virulence and host preference. So far, there is no comparative genomic study elucidating the signaling network in mycobacterial genus. Also, the number of kinases annotated in the available databases does not reflect the correct number of kinases with characteristic structural motifs. The principles applied herein toward mining the mycobacterial kinome gamut may be extended to other prokaryotes in assigning functions to orphan kinases.

**EXPERIMENTAL PROCEDURES**

**Database Search**

We searched the completely sequenced genomes of the *Mycobacterium* species (MtB H37Rv, MtB C, MtB F11, MtB CDC1551, Mle, Mbo, Map, Msm, and Mma) for the presence of STPKs using each of 11 full-length STPK sequences of MtB as well as a consensus kinase domain sequence (calculated from kinase domains of 11 MtB STPKs using a script available at http://www.bork.embl-heidelberg.de/Alignment/consensus.html) as query. In addition, individual genomes were also searched using the words “Pkn” and “serine-threonine kinase” to identify the annotated kinase sequences. The Pkn2 type sequences (34) thus retrieved were further critically evaluated for the presence of 11 Hank’s type kinase subdomains. MtB H37Rv, MtB C, MtB F11, MtB CDC1551, Mle, Map, and Mbo genome sequences, available at National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi) were searched for the presence of STPKs by employing PSI-BLAST (1) and BLASTP into the NR database using the default settings. We limited the Entrez to a particular mycobacterial species, while performing PSI-BLAST and BLASTP at NCBI. Msm and Mma genome sequences available at TIGR (http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi) and Sanger Centre (http://www.sanger.ac.uk/Projects/Microbes/), respectively, were searched with BLASTP. Mma kinases were searched against the MM.pep database (http://www.sanger.ac.uk/DataSearch/blast.shtml), while Msm kinases were searched against the MM.pep database (http://cmrblast.tigr.org/cmr-blast/) using the default settings. The critical residues in kinase subdomains were carefully spotted in each retrieved sequence, such as Rossmann-fold motif (GXGXXGXV), lysine in subdomain II, and aspartic acid residues in subdomain VIb and VII known to be involved in the ATP binding/catalysis (5, 24). Domain boundaries and structural organization of the retrieved STPK sequences were analyzed using SMART (51), InterProScan (http://www.ebi.ac.uk/InterProScan/), and Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Transmembrane segments were predicted using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and HMMTOP (http://www.enzim.hu/hmmtop/). Multiple sequence alignments were constructed using ClustalX (1.83) (59) or T-COFFEE (v. 4.59) (44). Full-length protein kinase sequences, which give a better representation of the catalytic domain as well as regulatory regions compared with domains alone, were used for the phylogenetic tree analysis. The phylogenetic analysis was carried out using ClustalX and Phylib (ver. 3.6) programs PROTDIST, NEIGHBOR, SEQBOOT, and CONSENSE (17). The Jones-Taylor-Thomson amino acid substitution matrix was used in PROTDIST. The input order of sequences for phylogenetic analysis was randomized, wherever this option was given. Phylogenetic tree was visualized with TREEVIEW. A region of at least 10-Kb on either side of each kinase gene in all mycobacterial genomes was gleaned to identify the neighboring genes, particularly any cluster of genes proposed to be involved in an particular process/pathway. The genes/clusters thus obtained were then compared across all mycobacterial genomes for the conservation of genomic location with respect to a given kinase.

**RESULTS AND DISCUSSION**

The wealth of sequence information compiled from microbial genome projects has opened new avenues for studying the importance of a protein in survival and virulence of tuberculosis bacillus and develop strategies to target these molecules/pathways for therapeutic interventions (53, 62). In the present study, we aim to assign a parallel nomenclature to all the mycobacterial STPK-like sequences as ortholog pairs, thereby aiding the prediction of their functions in mycobacterium. We used the symmetrical/bidirectional best hits (SymBet) method (32) to define an orthologous relationship between STPKs across mycobacterial species. To resolve issues, where more than one SymBet was obtained, auxiliary criteria such as percentage identity cut-off values, conservation of order of genes (synteny), and domain composition and organization within the protein were used to define orthologs and/or paralogs. Synteny is particularly important in defining correct orthologs, when there are a large number of highly related paralogs present at different genomic locations. As synteny is known to be considerably disrupted, when the protein sequence identity is <50% (26), a cutoff of >50% identity in full-length sequence was used to define candidate orthologs. Simultaneously, a stringent cut off value of ~70% was used for kinase domain sequence identity, because intracellular kinase domain particularly is under much tighter selective constraint than the juxtamembrane and extracellular domains (29). Aforementioned strategy helped us annotate certain kinases, which are currently either misannotated with different symbols or, in some cases, not even annotated as a kinase. Analysis of neighboring genes in the vicinity of STPKs revealed conservation of certain gene clusters involved in physiological processes/pathways. Dandekar et al. (13) proposed that the proteins encoded by conserved gene pairs appear to interact physically. Based on the conservation of gene clusters with respect to a particular kinase across mycobacterial species, we propose role of some of the STPKs in regulation of mycobacterial physiology.

**Distribution of STPK-like Sequences in Mycobacterial Genomes**

Earlier reports suggested the presence of 11 STPKs in MtB (H37Rv) and 4 STPKs in Mle (7, 8). Using a combination of search tools (PSI-BLAST and BLASTP), we identified 13 STPKs in Msm, 11 STPKs in Mbo, 10 STPKs in Map, 24 STPKs in Mma, and 11 STPKs in all MtB strains (CDC1551, F11, and C) (Fig. 1). A careful analysis of kinase domains and their subdomain characteristic residues in all the sequences retrieved in our search helped us annotate certain hypothetical proteins as kinases and also discard some of the sequences annotated as kinases in the genome database in our study.

**Classification and Analysis of Mycobacterial STPKs**

The full-length kinase sequence of all the retrieved STPKs of MtB (H37Rv), Mbo, Msm, Map, Mle, and Mma was used to
Fig. 1. Original genome annotations (protein identifiers) have been listed, and the names of organisms have been abbreviated as follows: Mtb, *M. tuberculosis*; Mbo, *M. bovis*; Mle, *M. leprae*; Msm, *M. smegmatis*; Map, *M. avium* subsp. *paratuberculosis*; Mma, *M. marinum*. The absence of orthologs of various kinases is shown by (–). The table is split into clades separated by dashed lines as described in Fig. 2. Representative schematic domain compositions are shown where vertical blue colored bars represent transmembrane helices and other abbreviations are as follows: S_TKc, serine threonine protein kinase catalytic domain; PASTA, penicillin-binding protein and serine/threonine kinase-associated domain; DsbG, protein-disulfide isomerase domain; NHL, NHL repeat; PBPb, periplasmic binding proteins; cyclophilin, cyclophilin-type peptidylprolyl cis, trans isomerase domain; WD40, protein-protein interaction module; DUF, domain of unknown function; MalT, ATP-dependent transcriptional regulator domain; COG5903, F-loop containing ATPase domain; Ad_Cycl_assoc, adenylate cyclase associated; COG3118, thioredoxin domain-containing protein. *Truncated protein kinases. † Sequencing error that resulted in a codon and misannotation of pknA ORF as 2 ORFs.

<table>
<thead>
<tr>
<th>Clade</th>
<th>Mtb H37Rv</th>
<th>Mtb CDC 1551</th>
<th>Mtb F11</th>
<th>Mtb C</th>
<th>Mbo</th>
<th>Mle</th>
<th>Msm</th>
<th>Map</th>
<th>Mma</th>
<th>Domain Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>PsATk1015c</td>
<td>MT0045</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
</tr>
<tr>
<td>II</td>
<td>PsATk0051c</td>
<td>MT0595</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
</tr>
<tr>
<td>III</td>
<td>PsATk0051c</td>
<td>MT0595</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
</tr>
<tr>
<td>IV</td>
<td>PsATk0051c</td>
<td>MT0595</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
</tr>
<tr>
<td>V</td>
<td>PsATk0051c</td>
<td>MT0595</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
<td>gH823209</td>
</tr>
</tbody>
</table>

Physiol Genomics • VOL 29 • www.physiolgenomics.org

Downloaded from http://physiolgenomics.physiology.org/ by 10.220.32.246 on June 19, 2017
generate the phylogenetic tree. The phylogenetic tree suggested the clustering of all mycobacterial STPKs into five groups or clades (Fig. 2) as described below:

**Clade I (PknA, PknB, PknL, ABL group).** Clade I represents the group of transmembrane sensor kinases from different mycobacterial species that cluster along with PknA, PknB, and PknL of Mtb. Signal transduction mediated by PknA and PknB of Mtb has recently been shown to regulate cell shape in mycobacteria, a process that may be conserved among gram-positive bacteria (2, 14, 30).

Transposon mutagenesis experiments by Sassetti et al. (50) propose that pknA and pknB genes are individually required for growth of Mtb in culture, suggesting them to be essential genes. Interestingly, search for orthologs of pknA in Map indicated the presence of two successive open reading frames (ORFs), MAP0018 and MAP0017, showing homology with the amino and carboxy termini of pknA, respectively (35). A closer examination revealed that these two ORFs in Map are separated by a stop codon generated from the deletion of a nucleotide that resulted in the formation of a new ORF MAP0017. Since the Mtb pknA gene is essential, resequencing of the region containing deletion was done, and indeed a sequencing error was found, confirming the existence of an intact pknA gene in Map (Supplemental Fig. S1). The kinase domain of this gene has no flanking sequences at either NH2 or COOH-terminal ends but has an insertion sequence of 22 amino acids between subdomains VII and VIII. The serine/threonine protein kinases (STPKs) were divided into 5 clades according to a number of criteria as described in text. The STPKs in *M. tuberculosis* strains CDC1551, F11, and C are identical or highly similar to those in *M. tuberculosis* strain H37Rv and have not been enlisted.

**Clade II**

**Fig. 2.** Phylogenetic analysis of serine/threonine kinases from 6 completely sequenced mycobacterial genomes. Analysis was performed by the neighbor-joining method of PHYLIP (ver. 3.63), and the tree was viewed by TREEVIEW. The number at the nodes indicates bootstrap support values after 500 bootstrap cycles. Unrooted tree was drawn using most distinct *M. smegmatis* protein kinase MSMeg5417 as out-group. (The kinase domain of this gene has no flanking sequences at either NH2 or COOH-terminal ends but has an insertion sequence of 22 amino acids between subdomains VII and VIII.) The serine/threonine protein kinases (STPKs) were divided into 5 clades according to a number of criteria as described in text. The STPKs in *M. tuberculosis* strains CDC1551, F11, and C are identical or highly similar to those in *M. tuberculosis* strain H37Rv and have not been enlisted.

**Clade III**

**Clade IV**

**Clade V**

**Clade I (PknA, PknB, PknL, ABL group).** Clade I represents the group of transmembrane sensor kinases from different mycobacterial species that cluster along with PknA, PknB, and PknL of Mtb. Signal transduction mediated by PknA and PknB of Mtb has recently been shown to regulate cell shape in mycobacteria, a process that may be conserved among gram-positive bacteria (2, 14, 30).

Transposon mutagenesis experiments by Sassetti et al. (50) propose that pknA and pknB genes are individually required for growth of Mtb in culture, suggesting them to be essential genes. Interestingly, search for orthologs of pknA in Map indicated the presence of two successive open reading frames (ORFs), MAP0018 and MAP0017, showing homology with the amino and carboxy termini of pknA, respectively (35). A closer examination revealed that these two ORFs in Map are separated by a stop codon generated from the deletion of a nucleotide that resulted in the formation of a new ORF MAP0017. Since the Mtb pknA gene is essential, resequencing of the region containing deletion was done, and indeed a sequencing error was found, confirming the existence of an intact pknA gene in Map (Supplemental Fig. S1). (The online version of this article contains supplemental material.) A pknA and pknB locus, which includes pstP, rodA, pbpA, and two FHA-domain-containing genes, was found to be conserved near the origin of replication throughout the mycobacterium genus (Figs. 1, 3).

Bioinformatic analysis revealed that pknL of Mtb is conserved in all sequenced mycobacterial species (Fig. 1). Mycobacterial PknL orthologs lack a substantial extracellular domain (Fig. 1); however, PknL orthologs in other actinobacteria like *Corynebacterium glutamicum* ATCC13032 (CG2388, a PknL ortholog) show the presence of a penicillin-binding protein and serine/threonine kinase-associated (PASTA) domain at the carboxy terminus, an extracellular sensor domain characteristic of Mtb PknB. The significant bootstrap value (434) of the tree branch (Fig. 2) and the presence of common domains suggest a common origin and lineage-specific expansion of PknL and PknB. Interestingly, we identified a 30-kb locus encompassing pknL and several genes encoding a dcw cluster (division cell wall gene cluster involved in cell envelope biosynthesis and cell division) that was found to be conserved throughout the genus (Fig. 4). The conservation of regions proposed to be involved in cell division and cell wall synthesis in the neighborhood of both pknB and pknL substantiates the notion that the two kinases may regulate common processes. A recent study indicates that an STPK mutant lacking an extracellular sensor domain comprising three PASTA domains displayed distinct pleiotropic effects on cell growth/division/septation and virulence factor expression in...
Streptococcus pyogenes (28). In view of this report, it seems plausible that PknL in mycobacteria may have adapted a different mechanism to activate itself in vivo in the absence of a sensor domain. A recent report by Kang et al. (30) suggested that PknB and PknA can cross-phosphorylate each other in vitro. In the light of the evidences of cross-phosphorylation in mycobacterial signaling network and the necessity of phosphorylation in activation of kinase activity (4, 16), a cross-talk of PknL with PknA and PknB cannot be ruled out. Importantly, the activation loop of PknL has a consensus sequence of a preferred phosphorylation site (I178LGTAAYLSP187) for PknA and PknB (30).

The conservation of dcw cluster in the vicinity of pknL across genus implicates a possible regulation of this cluster by PknL in mycobacterium. The conservation of an ~30-kb locus encompassing a pknL and dcw gene cluster (Fig. 4), phylogenetic clustering of PknL with PknA/PknB (Fig. 2), and possible cross talk between these kinases suggest that clade I is involved in the signal relay regulating cell division. Moreover, the presence of a putative DNA binding transcriptional regulator (Rv2175) upstream of pknL across the mycobacterium genus as a conserved gene pair suggests a probable mechanism of regulation of the dcw gene cluster by Rv2175 in a phosphorylation-dependent manner. The possibility of Rv2175 serving as a substrate of PknL is further augmented by various reports of in vitro as well as in vivo phosphorylation of proteins coded by genes conserved in the vicinity of a kinase in mycobacterium (12, 30, 39, 40).

A distinctive member of clade I, MSMEG6166, shows genomic organization quite similar to that of PknA and PknB. Genes located in close proximity to MSMEG6166 encode putative transcriptional regulators, MSMEG6158 and MSMEG6159 (whiB1), and the predicted cell division and elongation protein, MSMEG6160 (pdbA or ponA). The orthologs of all these genes except for STPK (MSMEG6166) were found to be present in all mycobacterial species. The genetic organization of Msm-specific STPK, MSMEG6166, suggests its plausible role in regulation of cell division.

Clade II (PknH, PknE, PknD, HED group). Clade II represents the group of integral membrane receptor as well as cytoplasmic kinases that cluster along with PknH, PknE, and PknD (Fig. 2). PknH regulates intracellular bacterial growth at later stages of the infection process, particularly during the chronic phase (45). Recently, it has also been shown that PknH regulates embCAB operon, thereby affecting ethambutol resistance and LAM/LM ratio, an important determinant of virulence (55). Phylogenetic tree analysis showed that MM1982, MM4171, MM4169, MM4156, MAP2504, and Mb1297 cluster with PknH of Mtb (Fig. 2). Except for MM1982, the primary amino acid sequences of these full-length proteins...
show ≥50% similarity with PknH (Supplemental Table S1). Upstream of MM4156, an embR ortholog, MM4155, is present, suggesting the conservation of pknH synteny in Mma. Thus, the true ortholog of PknH in Mma is MM4156, although MM4171 shows higher (79%) identity compared with MM4156 (62%) with intact PknH (Supplemental Table S1). This is in agreement with an earlier report, where the investigators found that in 29–38% of investigated co-ortholog relationships, the less-similar gene pair retained the ancestral gene in the neighborhood (43). The remaining two kinases of Mma (MM4171, MM4169) can be classified as inparalogs (32), as these proteins are more similar to each other than any other kinase sequence from the other mycobacterial genomes (Supplemental Tables S1 and S2). A gene encoding FHA domain containing ABC transporter, MM4170, lies between two of the pknH coorthologs, MM4169 and MM4171, encoded in the same orientation, indicating the possible regulation of the transporter by these kinases. An earlier report has shown in vitro phosphorylation of an Mtb FHA domain containing ABC transporter Rv1747 by multiple protein kinases (23).

PknE has been shown to phosphorylate various FHA domain containing proteins like GarA and ABC transporter in vitro (23, 61). Mtb PknE orthologs have been found only in Mma (MM2581) and Mbo (Mb1772) (Fig. 1). PknD has been shown to phosphorylate a transporter of RND family, MmpL7, in addition to various FHA domain-containing proteins (23, 49, 61). The locus organization of pknD in Mtb has a downstream pstS gene shown to be involved in phosphate uptake (47). Importantly, PknD has NHL-repeat β-propeller domain at the carboxy terminus (Fig. 1), which likely functions as anchoring sensor domain (21). The pknD ortholog in Mbo (Mb0955) has been shown to be truncated due to the insertion of an A at position 830, resulting in the loss of a transmembrane and sensor domain (48). The pknD ortholog with the downstream pstS gene is conserved in Map (MAP3387) and Mma (MM4577). One of the clade II members, MM3398, has a domain organization similar to that of PknD (Fig. 1) but cannot be classified as a PknD ortholog because it does not fulfill many important criteria described earlier for assigning ortholog status. It is not a SymBet of PknD, shows <70% identity in kinase domain (Supplemental Table S2), and does not have conserved synteny. This gene might have evolved as an independent fusion event of kinase domain with β-propeller domain.

A cluster of kinases branches out in parallel to the HED group. These group members show very high percent identity with the HED group members in their kinase domain, which is evident by the bootstrap values of the nodes in phenogram.
PPIase activity and subsequent cell division in mycobacteria. A conserved gene pair of kinases hints at the existence of a similar kinase as one continuous sequence as well as an independent genome (64). Presence of a cyclophilin domain fused with a fusion of the two as one continuous sequence in another clade of two genes can be derived from the existence of the UMP biosynthesis pathway (MM3407, MM3408, and MM3409).

Interestingly, MAP1049 was found to have a distinct peptidylprolyl cis-trans isomerase (PPIase)/cyclophilin domain at its carboxy terminus (Fig. 1). Also, just 31 bp upstream of MAP1049 is a PPIase A gene (MAP1050) in the same orientation (68% similarity between two cyclophilin domains). PPIases play a role in folding, transport, and assembly of proteins and are also implicated in cell cycle control through transcriptional regulation. Their activity is directed via targeting to prior phosphorylated proteins in yeast and mammals (56). Also a PPIase, Pin 1 activity has been shown to be regulated by phosphorylation at serine residue in mammalian cells (36). Importantly, in mycobacterium, another PPIase gene, namely PPIase A (Rv0009), is conserved throughout mycobacterial species downstream from pknA and pknB operon, a locus that has already been shown to be involved in cell division regulation in mycobacteria (Fig. 3). Recent work in computational genomics has shown that a functional association between two genes can be derived from the existence of a fusion of the two as one continuous sequence in another genome (64). Presence of a cyclophilin domain fused with a kinase as one continuous sequence as well as an independent conserved gene pair of kinases hint at the existence of a similar phosphorylation mediated signaling mechanism regulating PPIase activity and subsequently cell division in mycobacteria.

**Clade III** (PknF, PknI, PknJ, FIJ group). Clade III represents the group of kinases that cluster along with PknF, PknI, and PknJ (Fig. 2). PknF has been shown to play a role in the regulation of glucose transport, cell growth, and septum formation in Mtb (15). Four STPKs of Msm (MSMEG0879, MSMEG5491, MSMEG1183, and MSMEG3682), one STPK of Map (MAP1332), seven STPKs of Mma (MM2606, MM4923, MM2941, MM2408, MM4174, MM1794, and MM1423) and three STPKs of Mtb (PknI, PknJ, and PknF) show very high similarity in their kinase domains (Supplemental Table S2). On the basis of bidirectional best hits (SymBet), locus organization (presence of ABC transporter, Rv1747 ortholog), and percent identity, it can be concluded that MM2606, Mb1775, and MAP1332 are orthologs of PknF (Fig. 1). Only syntenic evidence can determine the correct ortholog when there are a number of highly related paralogs at different genomic locations. Since none of the clade III members from Msm show conservation of synteny of PknF, it is difficult to make an ortholog assertion. MAP1332, the only representative kinase of Map in clade III, is a SymBet of PknF; however, its carboxy terminus (374–561 amino acid residues) shows 57% similarity with that of PknI. Thus, MAP1332 may represent an example of a fusion event between two kinases in Map. Moreover, upstream from MAP1332 there is an ABC transporter (Rv1747) ortholog, MAP1331, suggesting that PknI and PknF may share common substrates as well.

**pknI** is one of the genes that have been shown to be downregulated during infection of Mtb in THP-1 macrophages (58). Although the pknI ortholog has been reported in Msm based on Southern blot analysis (22), the closest gene to pknI at the DNA sequence level, MSMEG1183, shows marked differences at the protein sequence level (Supplemental Tables S1 and S2). In fact, MSMEG1183 is not a SymBet of PknI and shows only 54% identity in kinase domain at the amino acid sequence level (Supplemental Table S2). The greater similarity of MSMEG1183 to other kinases of clade III at the protein sequence level, lack of asparagine in VIb subdomain, a characteristic feature of PknI (VIb, DXKPXN-DXNPXN) (see Supplemental Table S3), and dissimilar flanking genes argue against the presence of a PknI ortholog in Msm. The analysis of PknI ortholog in Mma, MM1794, shows deleterious substitutions at critical residues in subdomain I (G1→R, G3→D), II (K→E), VIb (R→G, N→D), and VII (D→G) (Supplemental Table S3) that have been shown to be essential for catalysis (25, 60). Since these residues are involved in catalysis, the prospects of catalytic competence of MM1794 must be viewed with caution.

**Fig. 5.** Comparison of **pknG** locus across mycobacterial genomes. Arrows indicate the orientation of ORFs. Vertical bars across arrows represent pseudogenes. Original genome annotations have been listed, and the names of organisms have been abbreviated as described in Fig. 3 legend.
with a higher degree of caution. An ortholog of PknJ is absent in all sequenced mycobacterial species except for Mbo (Fig. 1).

The presence of PknJ in the sequenced members of Mtbc complex alone suggests its unique repertoire of signaling specific to the Mtbc complex.

**Clade IV (PknK).** Clade IV represents soluble kinases that cluster along with PknK (Fig. 2). Our analysis suggests that orthologs of PknK are present in both slow- and fast-growing mycobacteria, with the exceptions of Map and Mle (Fig. 1). The PknK orthologs, like their Mtbc counterpart, have regulatory regions of the LuxR family and AAA domain at the carboxy terminus. pknK is located adjacent to oxidoreductases and various transcriptional regulators including *virs* (Rv3082c), which has been shown to regulate *mymA* operon (57). Four STPKs from Mma (MM2576, MM3347, MM3066, and MM3065), one from Msm (MSMEG0519), and one from Mbo (Mb3107) cluster with PknK. The locus organization of *pknK* is not conserved in other species; however, in Msm, various oxidoreductase/dehydrogenase family genes (MSMEG0515, MSMEG0516, MSMEG0517, MSMEG0518, and MSMEG0520) and transcriptional regulators (MSMEG0514, MSMEG0525, and MSMEG0529) are present in the vicinity of the kinase. MM2576 is the only kinase from Mma that fulfills the aforementioned criteria for ortholog assertion (Fig. 1). Importantly, InterProscan result shows an AraC type HTH region (amino acid residues from 966–1009) and TPR motif (from 735 to 983 amino acid residues, 6.2 e−7) in the Mma PknK ortholog, MM2576 (Supplemental Fig. S3). These regions are found to be conserved in Mtbc and Msm PknK orthologs as well. Other clade members from Mma show the presence of autoregulatory regions like ATPase domains. Also, two of the members, MM0366 and MM3547, have an additional terminally located adenyl cyclase-associated region (Fig. 1).

**Clade V (PknG).** Clade V represents soluble kinases that cluster along with PknG. The *pknG* locus in Mtbc includes *glnH* (Rv0411c) as an operon member (data not shown) and a conserved hypothetical membrane protein (Rv0412c). PknG was reported to inhibit phagosome-lysosome fusion, thus allowing intracellular survival of mycobacteria (63). Reduced in vitro growth and disturbed glutamine metabolism in *pknG* mutants of Mtbc as well as *Corynebacterium glutamicum* imply that *pknG* function is not restricted to the pathogenic lifestyle (10, 42). In agreement with this notion, our analysis suggested the conservation of *pknG* and its locus throughout mycobacterial genus (Figs. 1, 5). Moreover, we also noticed the conservation of few genes of thiamine biosynthesis pathway upstream of *pknG* and *glnH* (Fig. 5). MSMEG0781 of Msm is 79, 77, and 76% similar to PknG orthologs of Mtbc (H37Rv), Map, and Mle, respectively (Supplemental Table S1). Multiple sequence alignment of PknG orthologs suggests that they have similar kinase domains and carboxy termini, while PknG of Mtbc CDC1551, Mle, Map, and Msm has an extended amino terminus (Supplemental Fig. S4). The dissimilarity observed at amino terminus of PknG (Supplemental Fig. S4) may account for the functional differences that may be an underlying reason for enhanced survival observed in the case of Mtbc PknG overexpressing Msm (63). The ubiquitous presence of *pknG* and operon member, *glnH*, across both pathogenic and non-pathogenic species strongly supports the notion that *pknG* has a role to play in the maintenance of general mycobacterial physiology.

In conclusion, we have endeavored to provide an in-depth analysis and functional ortholog designation for mycobacterial STPKs based on a combination of criteria. This functional classification into clades helped us to predict certain novel physiological functions of kinases. With knowledge recently gained about the physiological role of mycobacterial STPKs, kinases have emerged as good drug targets for mycobacteria. Hence a holistic approach is inevitable in targeting these molecules, where functionally linked/similar kinases, as classified in the study, can constitute a regulatory network.

**ACKNOWLEDGMENTS**

The *M. marinum* sequence data were generated by the Sanger Institute Pathogen Sequencing Unit and can be obtained from: http://www.sanger.ac.uk/Projects/M_marinum.

Thanks to Professor Vivek Kapur, University of Minnesota, for resequencing Map *pknG* gene.

**GRANTS**

The work was supported by Council of Scientific and Industrial Research Grant No. SMM0007.

**REFERENCES**


