GLOBAL JOURNAL OF BIO-SCIENCE AND BIOTECHNOLOGY

www.scienceandnature.org

COMPARATIVE GENOMIC ANALYSIS OF MYCOBACTERIA (ACTINOBACTERIA): REDUNDANCY IN FUNCTION AMONG SERINE THREONINE PROTEIN KINASES

Harini Laxminarayan[1*] & Sujatha Narayanan[2]
[1]*Visiting Research Fellow, University of Surrey, Guildford, United Kingdom GU2 7XH
[2]Tuberculosis Research Centre (Indian Council of Medical Research), Chennai, Tamil Nadu, India 600031
Address for correspondence: National Institute for Research in Tuberculosis/TRC
(Indian Council of Medical Research), Chennai, Tamil Nadu, India 600031
Corresponding author email: sujatha.sujatha30@gmail.com, Tel: 0091-44-28369627

ABSTRACT
Mycobacterium tuberculosis is one of the most adaptable pathogens ever know which continues to take a heavy toll despite the availability of a vaccine and good antimicrobial regimen. The pathogen's malleability is well reflected in its increasing mutability leading to emergence of drug resistance, existence of multiple species with host preferences, and the ability to sense, survive and modulate the defense mechanisms in the host. Detailed comparison with prototype STPKs (cPKAα) is useful in identifying STPKs in the Mycobacterial genus which can be grouped into a family called 'Eukaryotic like STPKs'. The homologous nature of the kinase domains among these kinases implies that all of them fold into topologically similar 3-dimensional core structures and impart phosphor transfer according to a common mechanism. There are 11 representatives each of the prokaryotic HK-RR pair and the eukaryotic like serine/threonine protein kinase system in *M.tuberculosis*. The occurrence of both eukaryotic-like and prokaryotic signaling modules may seem like a huge genetic burden and an unwanted expense of metabolic reserves especially when viewed in the backdrop of functional and mechanistic redundance among kinases. Furthermore, it has been recently reported that multiple serine threonine kinases are able to phosphorylate a single substrate and vice versa. The existence of functionally redundant kinases has always attracted the attention of researches in mycobacteriology as it presents an impediment in interpreting phenotypic effects of individual STPKs. We aimed to address this issue of redundancy through comparative genome analysis.

Keywords: Ecophysiology, Serine threonine protein kinases, PknL, Gene duplication, Copy Number variation , Phylogenetic analysis, Actinobacteria, Stress response, Gene duplication, Adaptation

INTRODUCTION

M. tuberculosis is one of the most adaptable pathogens ever know which continues to take a heavy toll despite the availability of a vaccine and good antimicrobial regimen. The pathogen's malleability is well reflected in its increasing mutability leading to the emergence of drug resistance, existence of multiple species with host preferences, and the ability to sense, survive and modulate the defense mechanisms in the host. Detailed comparison with prototype serine threonine protein kinases (STPKs), namely cPKAα, is useful in identifying members in the mycobacterial genus which can be grouped into a family called 'Eukaryotic like STPKs'. The homologous nature of the kinase domains among these kinases implies that all of them fold into topologically similar 3-dimensional core structures and impart phosphor transfer according to a common mechanism (MacDonald, 2004). There are 11 representatives each of the prokaryotic two component system (HK-RR pair) and the eukaryotic like serine/threonine protein kinase system in *M. tuberculosis* (Av-Gay, 2000). The occurrence of both eukaryotic-like and prokaryotic signaling modules may seem like a huge genetic burden and an unwanted expense of metabolic reserves especially when viewed in the backdrop of functional and mechanistic redundancy among kinases. Furthermore, it has been recently reported that multiple serine threonine kinases are able to phosphorylate a single substrate and vice versa (Grundner, 2005). The existence of functionally superfluous kinases has always attracted the attention of researchers in mycobacteriology as it presents an impediment in interpreting phenotypic effects of individual STPKs. We aimed to address this issue of redundancy through comparative genome analysis and simultaneously gain an insight into the possible function of Protein Kinase L, a STPK from *Mycobacterium tuberculosis* H37*Rv*, which has been biochemically characterized by our group (Lakshminarayan, 2008).
In this study, a comparative genome analysis was done by using bioinformatics analysis to study the distribution of STPKs among actinobacteria with the final intent of gaining a deeper insight into their role in the ecophysiology properties of mycobacterial pathogens.

METHODOLOGY

The completed genome sequences of *Streptomyces coelicolor* A3(2), *Bifidobacterium longum* DJO10A, *Clavibacter michiganensis* subsp. michiganensis NCPPB 382, *M. tuberculosis* H37Rv, *Mycobacterium tuberculosis* CDC1551, *Mycobacterium tuberculosis* F11, *M. bovis* AF2122/97, *Mycobacterium bovis* BCG str. Pasteur

1173P2, *M. marinum* M, *M. smegmatis* MC2 155, *M. ulcerans* Agy99, *M. abscessus, M. leprae* TN, *Mycobacterium* sp. MCS, *M. avium* subsp. paratuberculosis K-10,M. avium 104, *Mycobacterium gilvum* PYR-GCK, *Mycobacterium vanbaalenii* PYR-1, *Mycobacterium* sp. JLS, Mycobacterium sp. KMS, *C. jeikeium*K411, *C. diphtheriae* NCTC 13129, *C. glutamicum* ATCC 13032, *Corynebacterium urealyticum* DSM 7109, *Geobacillus kaustophilus*, *Lactobacillus casei*, *Burkholderia pseudomallei* K96243 chromosome 1, *Pseudomonas fluorescens* SBW25, *Salmonella enterica* subsp. enterica serovar Typhi str. CT18, *Yersinia pestis* CO92, *Myxococcus xanthus* DK 1622, *Plesiocystis pacifica* SIR-1, *Stigmatella aurantiaca* DW4/3-1 available from the NCBI Genome database were used in performing comparisons.

Cyclic AMP dependent protein kinase A (cPKAα –mouse) is the progenitor of serine/threonine protein kinases. The protein sequences of each genome were queried with the protein sequence of cPKAα using the BLASTp analysis software available from NCBI Server. The search identified individual STKs in each genome. The protein kinases were selected based on annotation and in some instances through sequence alignment. Absence of conservation in Sub domains I, IV was tolerated as they are least conserved. To identify the orthologs to PknL from *M. tuberculosis* H37*Rv*, the Swiss–Prot protein database was queried with the PknL-H37*Rv* sequence downloaded from Tuberculist server (http://genolist. pasteur.fr/ TubercuList/). The BLAST Hits were aligned using CLUSTALW software integrated into CLC Protein Workbench, Jalview and other cross-platform bioinformatics algorithms. The output alignment was used to construct a phylogenetic tree using the Neighbor Joining algorithm. Sequences showing a >40% alignment with the query sequence were reanalyzed by Pair wise BLAST and the percentage identity was determined. The protein kinases were selected based on annotation and in some instances through sequence alignment. Absence of conservation in Sub domains I, IV was tolerated as they are least conserved.

RESULTS

In order to stream line the results, only BLAST hits from the classes, Actinobacteria and Proteobacteria were used for comparison. This is because *Mycobacterium tuberculosis* is a high G+C gram positive bacteria belonging to the phylum Actinobacteria and class Actinobacteria. Members of class proteobacteria were also included as they rank next in order of G+C content. The representative members of each class with their genome size, protein coding capacity, number of serine/threonine protein kinases encoded in their genome and the existence of PknL homologs are listed in Table 1. A compilation of the closest prokaryotic orthologs of H37*Rv*-PknL was used to construct a dendrogram showing the relatedness **(**Fig. 1). It was observed that PknL-H37*Rv* clusters with STPKs from other mycobacterial species. Besides, the STPK from *Nocardia farcinia* and Rhodococcus were also identified as being related to PknL.

TABLE 1: Comparative genome analysis of Mycobacterial and allied species for putative serine/threonine protein kinases

Name of species	Key feature	Genome size	Total protein	No. of STPks	PknL (annotated gene)	PknL homolog (% identity)
Streptomyces coelicolor A3(2),	Actinobacteria, complex developmental life cycle	8,667,507 nt	7769	31	_	53%
Nocardia farcinica IFM 10152	Actinobacteria, complex developmental life cycle	6,021,225	5683	20	_	56%
Bifidobacterium longum DJO10A	Actinobacteria, human gut flora	2,375,792 nt	1990	9	_	43%
Clavibacter michiganensis subsp. michiganensis NCPPB 382	Actinobacteria, Phytopathogen	3,297,891 nt	2984	5	_	51%
M. tuberculosis $H_{37}Rv$	Actinobacteria	4,411,532 nt	3989	11	+	100%
Mycobacterium tuberculosis CDC1551	Actinobacteria	4,403,837 nt	4189	11	_	99 %
Mycobacterium tuberculosis F11	Actinobacteria	4,424,435 nt	3941	11	_	99 %
M. bovis AF2122/97	Actinobacteria	4,345,492 nt	3920	11	+	99 %
Mycobacterium bovis BCG str. Pasteur 1173P2	Actinobacteria	4,374,522 nt	3952	11	+	99 %
M. marinum M	Actinobacteria	6,636,827 nt	5423	23	+	78%
M. smegmatis MC^2 155	Actinobacteria	6,988,209 nt	6716	13	_	63%
M. ulcerans Agy99	Actinobacteria	5,631,606 nt	8753	12	+	78%
M. abscessus	Actinobacteria	5,067,172 nt	4920	6	_	
M. leprae TN	Actinobacteria	3,268,203 nt	1605	4	+	74%
Mycobacterium sp. MCS	Actinobacteria	5,705,448 nt	5391	14	_	
M. avium subsp. paratuberculosis K-10	Actinobacteria	4,829,781 nt	4350	11	+	73%
M. avium 104	Actinobacteria	5,475,491 nt	5120	10	-	73%
Mycobacterium gilvum PYR-GCK	Actinobacteria	5,619,607 nt	5241	12	_	65%
Mycobacterium vanbaalenii PYR-1	Actinobacteria	6,491,865 nt	5979	11	_	65%
Mycobacterium sp. JLS	Actinobacteria	6,048,425 nt	5739	13	_	68%
Mycobacterium sp. KMS	Actinobacteria	5,737,227 nt	5460	15	_	68%

C. jeikeium K411	Actinobacteria	2,462,499 nt	2104	5	+	53%
C. diphtheriae NCTC 13129	Actinobacteria	2,488,635 nt	2272	4	_	47%
C. glutamicum ATCC 13032	Actinobacteria	3,309,401 nt	2993	4	_	57%
Corynebacterium urealyticum DSM 7109	Actinobacteria	2,369,219nt	2024	4	+	57%
Geobacillus kaustophilus	Firmicutes	47,890 nt	42	2	_	36%
Lactobacillus casei	Firmicutes, human gut flora	3,079,196 nt	3044	1	_	32%
Burkholderia pseudomallei K96243 chromosome 1	Betaproteobacteria	4,074,542 nt	3399	2	_	35%
Pseudomonas fluorescens SBW25	Gammaproteobacteria	425,094 nt	474			36% ppkA
Salmonella enterica subsp. enterica serovar Typhi str. CT18,	Gammaproteobacteria	4,809,037 nt	4395	2	_	28%
Yersinia pestis CO92	Gammaproteobacteria	4,653,728 nt	3885	2	_	_
Myxococcus xanthus DK 1622	Deltaproteobacteria, complex developmental life cycle	9,139,763 nt	7331	~95	_	41%
Plesiocystis pacifica SIR-1	Deltaproteobacteria, complex developmental life cycle	10,587,646 nt	8450	101	+	39%
Stigmatella aurantiaca DW4/3-1	Deltaproteobacteria, complex developmental life cycle	10,265,408 nt	8543	100	+	39%

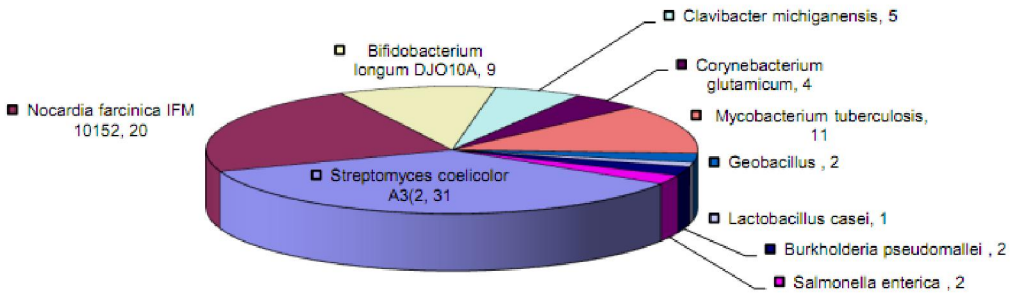

FIGURE 1: Distribution of STPKs among Actinobacteria and Proteobacteria

DISCUSSION

Correlation between the ecophysiology and STPK numbers

Comparative genome analysis has been employed in the study on STKs in archaea (Kennelly PJ, 2003), mycobacteria (Av-Gay, Y. and Everett, M., 2002), streptomyces (Petrickova, K .and Petricek, M., 2003) and cyanobacteria (Xiaowen Zhang *et al.*, 2007). Earlier studies with cyanobacteria have shown that the number of STK genes in the genome is the result of the genome size, ecophysiology, and physiological properties of the organism.

This analysis is different from the previous work (Av-Gay, Y. and Everett, M., 2002) in that its coverage and completeness is extensive with the inclusion of representative genera under the class actinobacteria and the related proteobacteria. The study does not merely identify putative STKs; it also correlates their numbers in the respective genomes with the ecophysiological properties of the organism. We have also proposed our perspective on the redundancy of the signaling molecules in mycobacterial genomes but it is maintained that the proposal needs to be reaffirmed by detailed computational analysis.

Structure of the comparative genomic analysis

This survey includes the genera under the subclass Actinobacteridae and order Actinomycetales, which is in turn subdivided into 10 suborders. Three important suborders have been considered namely, Sreptomycinae, Corynebacterinaeae and Actinomycinaeae, as they contain the ecologically and medically important genera.

Suborder Corynebacterineae

Phylogenetic analysis based on 16SrRNA gene sequence indicate that the genus Corynebacterium, Mycobacterium, Nocardia and Rhodococcus are related and are hence clubbed under one suborder in Vol. 4 of Bergeys Manual of Systematic Bacteriology.

Genus Corynebacterium

They are gram positive slightly curved rods widely distributed in soil and water, and reside on the mucous membranes of humans and other animals. *C. diptheriae*

causes diphtheria by colonizing the oropharynx and elaborating an exotoxin that causes necrosis and inflammation of the mucosa leading to pseudo membrane formation. *C. jeikeium* causes septicemia and prosthetic valve endocarditis in immunecompromised patients and *C. urealyticum* is a strong urease producer, which degrades urea into ammonia, leading to alkalinization of urine and deposition of AMP crystalline crusts in the bladder wall. The absence of a complex developmental cycle or the need for extensive modulation of the host signaling systems has resulted in slashing the numbers of ESTPKs in the genome to 4 (**Fig. 2**). These kinases are primarily involved in regulation cell growth and division (Fiuza, 2008).

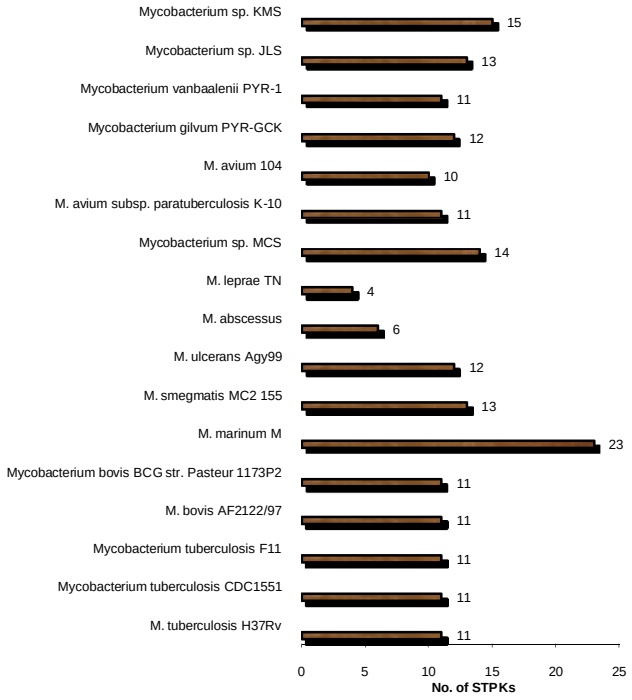

FIGURE 2: Distribution of STPKs among Mycobacteria

Genus Nocardia

Nocardia and Rhodococcus are widely distributed in the soil and acquatic habitats. Here they contribute to biodegradation of hydrocarbons and waxes (Martinkova, 2009). They form aerial mycelium and microconidia. The need to participate in xenobiotic metabolism and a complex life cycle has led to the presence STPKs in large numbers in Nocardial genome (21). *A Nocardia asteroid is* an opportunistic pathogen which causes progressive bronchopneumoniae in immunocompromised hosts. The organism is capable of invading macrophages and preventing phagolysosomal fusion. Thus the need to modulate host signaling/trafficking further justifies the extended collection of STPKs seen in this genus.

Genus Mycobacteria

Mycobacterium is a huge genus with 71 species. It has been classified base on growth rate, virulence, pigmentation, nutrititional requirements etc., although mycobacteria are best known as animal pathogens, majority of the members belonging to this genus are free living saprophytes that colonise the soil and acquatic habitats. Based on ecology and phylogeny Mycobacteria have been classified into environmentally deived mycobacteria (EDM) and obligate pathogenic mycobacteria (OPM) (Kazda J, 2000).

M. ulcerans and *M. marinum* have identical signature sequences, with nucleotide sequence identity ranging from 98.1 to 99.7% in the 16S rRNA locus. However, the two species markedly differ in their etiology and epidemiology. *M. ulcerans* is an emerging human pathogen that causes a chronic, necrotic skin lesion in humans and has an extracellular location during infection. On the other hand, the fish pathogen *M. marinum* isolated from marine habitats is an intracellular pathogen (Stinear, 2000). The robustness needed to withstand some of the extremes of aquatic environments such as sunlight exposure, varying temperatures and nutrient limitation is probably reflected in the extended collection of STPK genes (23) **(Fig. 2)**. *M. ulcerans* is similar to *M. tuberculosis* in its slow growth, UV sensitivity, and optimal 208 growths under microaerophilic conditions. The two represent species that have ably adapted to a specific environmental niche in their human host and possess the same number of STPKs (11-12) **(Fig. 2)**. The comparison between *M. marinum* and *M. ulcerans* presents a particular point in case to substantiate the view that the specific environmental niche of the microbe, in addition to its genetic constitution, forms the basis for the phenotypic differences observed.

The three Mycobacterium strains, JLS, KMS and MCS were isolated from a land treatment unit prepared for the

bioremediation of soil contaminated with wood preservatives containing poly aromatic hydrocarbons (PAHs). These unique PHA-degrading mycobacteria possess on an average, a greater number of STPK genes (13-15) **(Fig. 2)** and distinct isoenzyme patterns for catalase and superoxide dismutase
(SOD) (Miller, 2004). Subtle changes in the regulatory genes, such as those involved in signal transduction, may help the species cope with its environment. Thus, it is conjectured that STPK genes form the substrate for introducing micro-evolutionary changes among mycobacteria. *M. leprae* has retained 4 of the 11 STPKs seen in *M. tuberculosis*, namely PknA, PknB, PknG and PknL **(Fig. 2)**. The abridged repertoire of signal transduction systems in *M. leprae* may be due to the restricted tissue specificity (neural) and hence the limited adaptive needs of this pathogen. The existence of PknL in *M. leprae*, a bacterium that has undergone massive gene decay tempts one to speculate that this kinase could play a pivotal role in its growth, survival and/or pathogenesis (Cole, 2001).

Sub order Streptomycineae

The genus Streptomyces is huge with around 500 species. They have a complex developmental life cycle forming substrate and aerial mycelia and spores. The natural habitat is soil (moist earth) and here they play a role in mineralization by degrading chitin, keratin, latex and aromatic compounds. They also cause actinomycotic mycetoma in humans. Because of their capability to develop into a new developmental stage, Streptomyces have many STPK-encoding genes (Fig 2).

Redundancy in STPKs: Does it represent Copy Number Variation?

Copy number variation (CNV) is defined as a DNA segment that is 1Kb or larger and present at variable copy number in comparison with a reference genome. A copy number variation can be simple in structure, such as tandem duplications, or may involve complex gains or losses of homologous sequences at multiple sites in the genome.

CNVs influence gene expression, phenotypic variation and adaptation by disrupting genes and altering gene dosage. They alter gene expression through position effects and provide substrates for chromosomal changes in evolution.

There is emerging evidence that structural variations have a role in determining the fitness of an organism, with potential evolutionary implications. Gene Ontology studies on the human genome have identified that there is a particular enrichment of genes that are involved in general defense responses, including defense response to bacteria, responses to external biotic stimuli, xenobiotic metabolism and regulation of cell organization and biogenesis. These observed variations indicate that they may have roles in the adaptability and fitness of an organism in response to external pressures. These “plastic genes” have a tendency to evolve quickly and are important for the dynamics of gene and organismal evolution (Redon, 2006).

One significant mechanism of CNV formation is duplication. Segmental duplications are defined as sequences in the reference genome assembly sharing > 90% sequence similarity over > 1Kb with genomic location (Sharp, 2005).

Extrapolations to Mycobacterial genomics

The complete sequencing of *M. tuberculosis* and the availability of the genome sequence for other mycobacterial species has uncovered the occurrence of gene duplication in this family. The PE_PGRS multigene family coding for asparagines or glycine rich proteins, members ESAT6 family coding for T cell epitopes and the signaling kinases occur in multiple copies. While the more direct and obvious consequence of such gene duplication is the profound antigenic and genetic polymorphisms, the influence on virulence of the corresponding strains has also been documented (Poulet, 1995).

Comparing the STPKs among all bacteria, in particular actinobacteria reveals that the kinase domain encompassing at least 280 amino acids is well conserved (Hanks, 1995). This accounts for at least ~1Kb in terms of nucleotide sequence and thus fits into the description of segmental duplication. For such comparisons it is required that the ancestral state of the CNV be determined to a construct a reference point for subsequent comparisons. In this study the actinobacterial genome of Streptomyces was selected as the reference genome as it represents the node of the phylogenetic tree of actinobacteria. Though *M. tuberculosis* var. canettii, the extant relative of *M. prototuberculosis*, the progenitor mycobacteria, its genome sequence was not available in the NCBI server and hence was not used for comparison (Smith NH, 2006).

There are 11 representatives each of the prokaryotic HK-RR pair and the eukaryotic like serine/threonine protein kinase system in *M. tuberculosis*. The occurrence of both eukaryotic-like and prokaryotic signaling modules may seem like a huge genetic burden and an unwanted expense of metabolic reserves especially when viewed in the backdrop of functional and mechanistic redundancy among kinases. More over the occurrence of orphan members of HK or RR pairs and the variable numbers of kinases seen among mycbacterial species seems to point towards a certain phenomenon of ‘load shedding’ where the commonality in the function is given more weightage than any one particular kinase.

Thus the eukaryotic like STPKs in *Mycobacterium tuberculosis* may be assumed to function as a web rather than assuming distinct roles.

REFERENCES

Av-Gay, Y. & Everett, M. (2000) The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* 8 (5):238-44.

Cole, S.T., Eiglmeier, K., Parkhill, J. (2001) Massive gene decay in the leprosy 293 bacillus. *Nature* 409 (6823):1007-11.

Fiuza, M., Canova, M.J., Zanella-Cleon, I., Becchi, M., Cozzone, A.J., Mateos, L.M., Kremer, L., Gil, J.A. & Molle, V. (2008) From the characterization of the four serine/threonine protein kinases (PknA/B/G/L) of *Corynebacterium glutamicum* toward the role of PknA and PknB in cell division. *J Biol Chem* 283 (26):18099-112.

Greenstein, A.E., Grundner, C., Echols, N., Gay, L.M., Lombana, T.N., Miecskowski, C.A., Pullen, K.E., Sung, P.Y. & Alber, T. (2005) Structure/function studies of Ser/Thr and Tyr protein phosphorylation in *Mycobacterium tuberculosis*. *J Mol Microbiol Biotechnol* 9 (3-4):167-81.

Grundner, C., Gay, L.M. & Alber, T. (2005) *Mycobacterium tuberculosis* serine/threonine kinases PknB, PknD, PknE and PknF phosphorylate multiple FHA domains. *Protein Sci.* 14 (7):1918-21.

Hanks, S.K. & Hunter, T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9 (8):576-96.

Han, G. & Zhang, C.C. (2001) On the origin of Ser/Thr kinases in a prokaryote. *FEMS Microbiol Lett* 200(1):79-84.

Kazda, J. (2000) The Ecology of Mycobacteria. pp 13-15. Kluwer Academic Publishers.

Kennelly, P.J. (2003) Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry. *Biochem J* 370:373-389.

Kliebenstein, D.J. (2008) A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLoS One* 3 (3):e1838.

Koneman, E.W., Winn, W.C., Allen, S.D., Janda, W.M., Schreckenberger, P.C., Procop, G.W. & Woods, G.L. (2005) Color atlas and textbook of diagnostic microbiology. pp 783-810.Lippincott Williams & Wilkins.

Lakshminarayan, H., Narayanan, S., Bach, H., Sundaram, K.G. & Av-Gay, Y. (2008) Molecular cloning and biochemical characterization of a serine threonine protein kinase, PknL, from *Mycobacterium tuberculosis*. *Protein Expr Purif* 58 (2):309-17.

MacDonald, J.A. (2004) Signal transduction pathways and the control of cellular responses to external stimuli. Functional metabolism: regulation and adaptation, (Storey KB, ed), pp. 87-120. J.Wiley and Sons, Hoboken, NJ.

Martinkova, L., Uhnakova, B., Patek, M., Nesvera, J. & Kren, V. (2009) Biodegradation potential of the genus Rhodococcus. *Environ Int* 35 (1):162-77.

Miller, C.D., Hall, K., Liang, Y.N., Nieman, K., Sorensen, D., Issa, B., Anderson, A.J. 325 & Sims, R.C. (2004) Isolation and characterization of polycyclic aromatic hydrocarbon-degrading Mycobacterium isolates from soil. *Microb Ecol* 48 (2):230-8.

Petrickova, K. & Petricek, M. (2003) Eukaryotic-type protein kinases in *Streptomyces coelicolor*: variations on a common theme. *Microbiol* 149 (Pt 7):1609-1621.

Poulet, S. & Cole, S.T. (1995) Characterization of the highly abundant polymorphic GC-rich repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol.*, 163 (2):87-95.

Redon, R., Ishikawa, S., Fitch, K.R. (2006) Global variation in copy number in the human genome. *Nature* 444 (7118):444-54.

Sharp, A.J., Locke, D.P., McGrath, S.D. (2005) Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 77 (1):78-88.

Smith, N. H. (2006) A re-evaluation of *M. prototuberculosis*. *PLoS Pathog* 2 (9):e98.

Stinear, T.P., Jenkin, G.A., Johnson, P.D. & Davies, J.K. (2000) Comparative genetic analysis of *Mycobacterium ulcerans* and *Mycobacterium marinum* reveals evidence of recent divergence. *J Bacteriol* 182 (22):6322-30.

Zhang, X., Zhao, F., Guan, X., Yang, Y., Liang, C. & Qin, S. (2007) Genome-wide survey of putative Serine/Threonine protein kinases in cyanobacteria. *BMC Genomics* 8: 395.