Bioinformatics
Research Article – 66743

Sezer Okay

In silico characterization of the citrate synthase family in *Mycobacterium tuberculosis*

Mycobacterium tuberculosis'te sitrat sentaz ailesinin in silico karakterizasyonu

DOI 10.1515/tjb-2016-0019
Received November 13, 2015; accepted January 4, 2016

Abstract: Objective: *Mycobacterium tuberculosis* (MTB) is an obligate aerobe bacterial pathogen. Here, the citrate synthase (CS) family, an important component of aerobic respiration, was investigated in MTB.

Methods: MTB genome was analyzed in silico to reveal the members of CS family. The nucleotide and amino acid sequences were retrieved from the NCBI database, and searched for the similarity using the NCBI BLAST tool. Sequence alignment and phylogenetic analysis were performed using MEGA6. The physicochemical parameters, cellular localization, HMM profiles, motif structure, 3D modeling, and the interactions of the proteins were analyzed using GPMAW, PSORTb, Pfam and SMART, MEME, Phyre2, and STRING databases, respectively.

Results: The members of CS family in MTB were identified as CitA, GltA2, and PrpC. The CitA and PrpC were found to be closer in phylogeny than GltA2, and the trees of three proteins were shown to be similar to that constructed based on 16S rRNA in mycobacteria. The CitA contains two CS domains while a single CS domain is found in GltA2 and PrpC. Besides, LHGGA and MGFGHRVY motifs are conserved in MTB and various bacteria. The molecular weight and pI values of CitA, GltA2, and PrpC were calculated as 40.1, 47.9, and 42.9 kDa, and 5.41, 5.35, and 9.31, respectively. Cellular localization of the proteins was predicted as cytoplasm. The highest expression ratio was found to be for *gltA2* followed by *prpC* and *citA*, respectively, in the retrieved RNA-seq datasets obtained from the aerobic log phase of MTB H37Rv.

Conclusion: This comprehensive bioinformatics analysis of CS family in MTB has a contribution to the knowledge of the genetics and physiology of this pathogen.

Keywords: Aerobic respiration, citrate synthase, *Mycobacterium tuberculosis*, pathogenicity

Özet: Amaç: *Mycobacterium tuberculosis* (MTB) zorunlu aerop olan bakteriyel bir patojendir. Burada, oksijenli solunumun önemli bir bileşeni olan sitrat sentaz (SS) ailesi MTB'de incelenmiştir.

Metod: SS ailesinin üyelerini belirlemek için MTB genomu in silico analiz edildi. Nükleotit ve aminoasit dizileri NCBI veri tabanından alındı ve benzerlikleri NCBI BLAST aracıyla tarandı. Dizi hizalama ve filogenetik analiz MEGA6 kullanılarak gerçekleştirildi. Fizikokimyasal parametreler, hücresel lokalizasyon, HMM profilleri, motif yapısı, üç boyutlu modelleme ve diğer proteinlerle etkileşimler sırasıyla GPMAW, PSORTb, Pfam ve SMART, MEME, Phyre2, ve STRING veri tabanları kullanılarak belirlendi.

Bulgular: MTB'deki SS ailesinin üyeleri CitA, GltA2 ve PrpC olarak tespit edildi. CitA ve PrpC, filogenide birbirine GltA2'den daha yakın olup, mikobakterilere ait üç proteinin filogenetik ağacı, 16S rRNA ile hazırlanana benzer olarak bulundu. CitA iki SS domaini içerirken GltA2 ve PrpC tek domain içermektedir. LHGGA ve MGFGHRVY motifleri MTB ve birçok bakteride korunmuştur. CitA, GltA2 ve PrpC'nin moleküler ağırlık ve pI değerleri sırasıyla 40.1, 47.9 ve 42.9 kDa, ve 5.41, 5.35 ve 9.31 olarak hesaplandı. Proteinlerin hücresel lokalizasyonu sitoplazma olarak tahmin edildi. Aerobik log fazındaki MTB H37Rv'den elde edilen RNA-seq veri setleri toplandı ve en yüksek ifade olan gen *gltA2*, bunu takiben sırasıyla *prpC* ve *citA* olarak tespit edildi.

Corresponding author: Sezer Okay: Çankırı Karatekin University Faculty of Science, Department of Biology, Çankırı, Turkey, e-mail: sezerokay@gmail.com

Sonuç: MTB'ye ait SS ailesinin bu geniş çaplı biyoinformatik analizi, patojenin genetik ve fizyoloji bilgisine bir katkı sağlamaktadır.

Anahtar Kelimeler: Aerobik solunum, sitrat sentaz, *Mycobacterium tuberculosis*, patojenite

Introduction

Tuberculosis (TB) is an infectious fatal disease caused by the intracellular bacterial pathogen *Mycobacterium tuberculosis* (MTB) [1]. According to the World Health Organization (WHO) Global Tuberculosis Report 2014 [2], the incidence of global TB cases in 2013 was 9.0 million and the mortality was 1.5 million including 360.000 people who were HIV-positive. Deciphering the genetics and physiology of the disease agent, MTB, will contribute to the prevention of the high rate of TB mortality.

The tubercle bacillus is obligate aerobe in acute infection but can also survive indefinitely under hypoxic conditions causing latent infection. The latency is associated with limited oxygen in the environment, in which MTB does not replicate or grow very slowly [3,4]. Besides, respiration is an important component of the MTB infection providing flexibility to the pathogen for the adaptation to the environmental stresses [5–7].

Citrate synthase (CS) (EC 2.3.3.1) enzyme catalyses the Claisen condensation of oxaloacetate and acetyl-coenzyme A to produce citrate and coenzyme A, the first step in the Krebs tricarboxylic acid (TCA) cycle [8–10]. The CS enzymes are characterized into two types according to their regulatory and structural properties. Type I CSs are dimeric molecules without allosteric regulation, and found in Gram-positive bacteria, archaea and eukaryotes. On the other hand, type II CSs, found in Gram-negative bacteria, have homohexameric structure strongly inhibited by NADH [10,11]. There is only a single study on the structural analysis of a member of CS family from MTB. Ferraris et al. [7] reported the crystal structure of dimeric GltA2 protein having a typical common α-structure.

The increased number of genome sequences belonging to MTB strains and other mycobacteria enables the comparative bioinformatic analysis of gene families in a genome and among the genomes from different sources. Here, the citrate synthase family covering CitA, GltA2, and PrpC proteins from MTB was in silico characterized exhibiting their phylogenetic relationships, physicochemical properties, cellular localization, HMM profiles, motif structure, 3D modeling, and interaction network. Moreover, expression profiles of *citA*, *gltA2*, and *prpC* in MTB under aerobic conditions were analyzed using retrieved RNA-seq datasets.

Material and Methods

Sequence information

The nucleotide and amino acid sequences were retrieved from the NCBI Nucleotide (http://www.ncbi.nlm.nih.gov/nuccore) and Protein (http://www.ncbi.nlm.nih.gov/protein) databases. The sequences of *citA* (Rv0889c), *gltA2* (Rv0896), and *prpC* (Rv1131) genes as well as the corresponding amino acid sequences used for the analysis belongs to *Mycobacterium tuberculosis* H37Rv complete genome (GenBank accession number: AL123456.3). The accession numbers of CitA, GltA2, and PrpC proteins used in this study are AFN48770.1, AFN48778.1, and NP_215647.1, respectively.

Sequence alignment and phylogenetic analysis

The similarity search for the sequences were performed using the NCBI Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi). The multiple alignment of the sequences was conducted with ClustalW, and the phylogenetic relationships were analyzed using the Neighbor-Joining method implemented in MEGA6 software [12] on the basis of uncorrected p-distance. The phylogeny was tested via bootstrap method with 1000 replications [13]. Malate dehydrogenase protein from *M. tuberculosis* H37Rv (accession number NP_215756.1) was used as the outgroup.

Bioinformatic analysis of the sequences

The physicochemical parameters of the proteins belonging to MTB citrate synthase family were determined using the online General Protein Mass Analysis for Windows (GPMAW) protein bioinformatics tool (http://www.alphalyse.com/gpmaw_lite.html) rapidly analyzing the protein information, amino acid composition, and protease digest. The cellular localization of the proteins were predicted via the PSORTb v3.0.2 (http://www.psort.org/psortb; [14]), the most precise bacterial protein subcel-

Table 1: The CS family members in *M. tuberculosis* H37Rv genome.

CS family member	Gene	Gene synonym	Locus tag	Localization	Gene length (bp)	Protein length (aa)
Citrate synthase II	*citA*	*gltA*	Rv0889c	988740..989861 (complement)	1122	373
Citrate synthase I	*gltA2*	–	Rv0896	999472..1000767	1296	431
Methylcitrate synthase	*prpC*	*gltA1*	Rv1131	1256132..1257313	1182	393

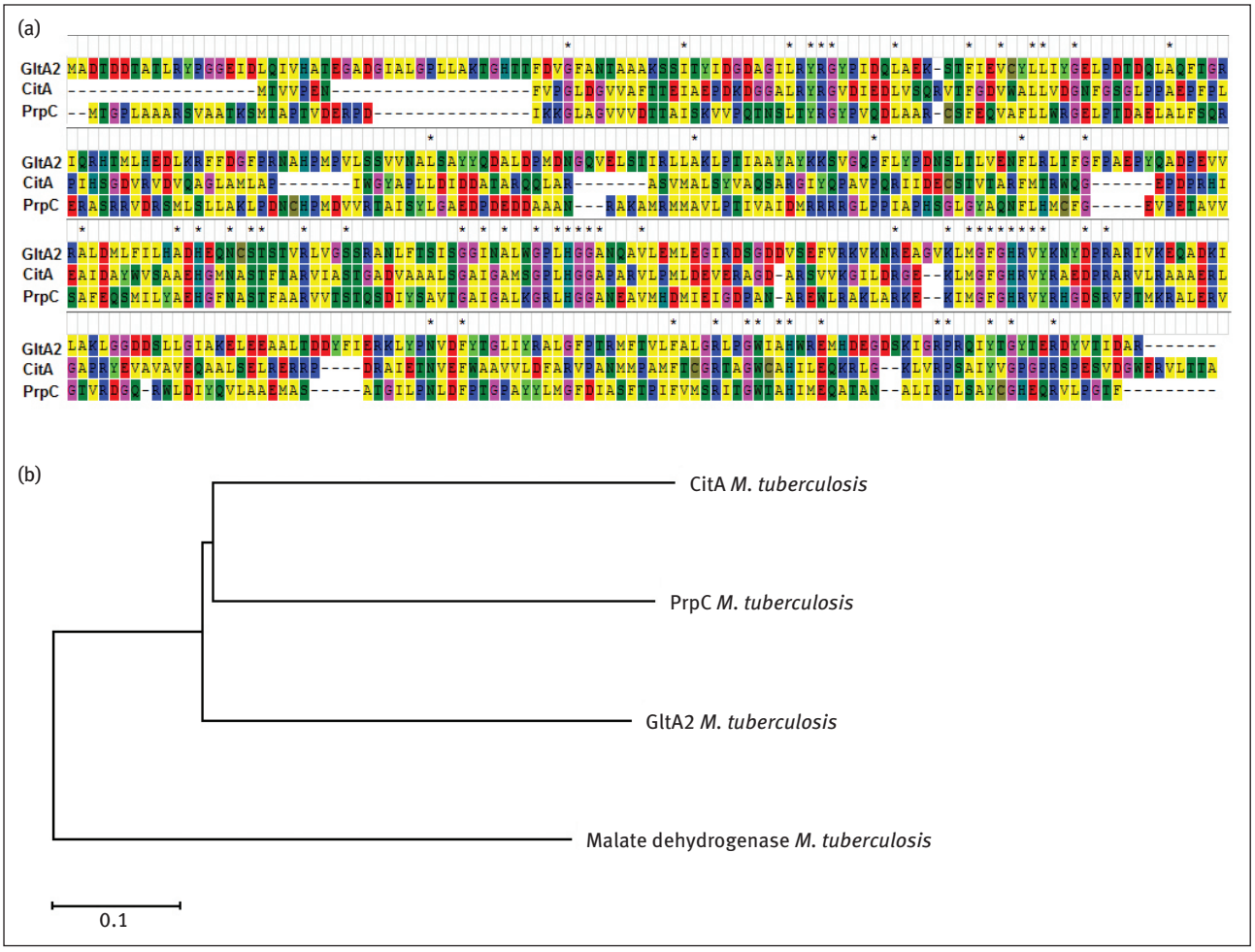

Figure 1: The citrate synthase family members in *M. tuberculosis*. Multiple alignment (a) and phylogenetic relationships (b) of GltA2, CitA, and PrpC proteins. The malate dehydrogenase sequence was used as outgroup. The stars show the conserved residues, the colors indicate same or similar amino acids, and the dash lines represent the gaps.

lular localization (SCL) predictor. The Hidden Markov Model (HMM) profiles of the proteins were analyzed using the Pfam v27.0 (http://pfam.xfam.org; [15]), and the SMART (http://smart.embl-heidelberg.de; [16]) databases. The motif analysis was conducted using the motif-based sequence analysis tools MEME Suite v4.10.0 (http://meme-suite.org). The 3D modeling of the proteins was analyzed via the Phyre2 web portal (http://www.sbg.bio.ic.ac.uk/phyre2; [17]) utilizing PSI-BLAST (Position-Specific Iterated BLAST) for similarity search, and HHpred 1.51 [18] for template detection. The interactions of the proteins were investigated by network analysis using STRING v10 database (http://string-db.org) employing the partner databases TISSUES (http://tissues.jensenlab.org) and DISEASES (http://diseases.jensenlab.org) for enrichment, and NCBI Gene Expression Omnibus (NCBI GEO; http://www.ncbi.nlm.nih.gov/geo) for co-expression scores. Additionally, the gene expression values are correlated by Pearson correlation and calibrated against KEGG pathway maps [19].

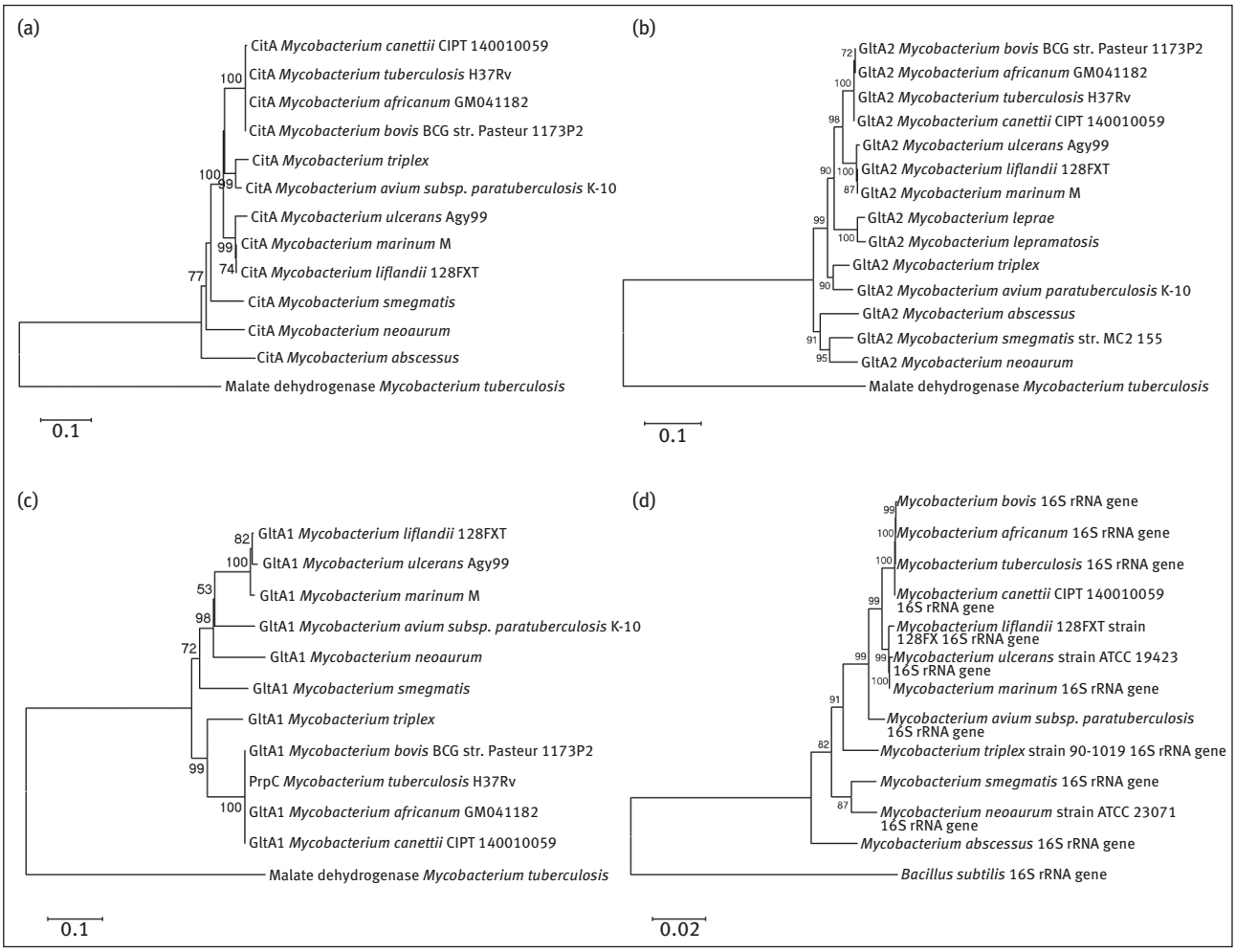

Figure 2: The phylogenetic relationships of CitA (a), GltA2 (b), and PrpC (c) proteins among the various mycobacterial species. The phylogenetic tree based on *16S rRNA* gene (d) was constructed for comparison. Numbers next to the branches show the percentages of replicate trees in which the associated taxa clustered together in the bootstrap test.

In silico analysis of the RNA-seq datasets

The RNA-seq datasets belonging to the wildtype *M. tuberculosis* H37Rv were retrieved from the NCBI Sequence Read Archive (SRA) database (http://www.ncbi.nlm.nih.gov/sra) with the accession numbers of SRX727249, SRX727250, and SRX727251. The datasets were belonging to the three biological replicates of log phase cultures grown under aerobic conditions. The transcripts representing *citA*, *gltA2*, and *prpC* were detected using the sequences encoding the LHGGA or MGFGHRVY motifs. Selected transcripts were verified by BLAST analysis. The number of transcripts for the CS family members was determined separately in each RNA-seq datasets, and their percentages with respect to the total number of transcripts in each dataset were calculated.

Results and Discussion

Phylogenetic relationships of the MTB citrate synthase family

Aerobic respiration is critical for the pathogenesis of *M. tuberculosis* (MTB) [7]. Investigation of the MTB H37Rv genome for the citrate synthase (CS) family revealed three members, namely *citA*, *gltA2*, and *prpC* (Table 1). The CitA, GltA2, and PrpC proteins are composed of 373, 431, and 393 amino acids, respectively. The CS family members in MTB take role in the pathogenesis of this microorganism. Muñoz-Elías et al. [20] showed that the PrpC is involved in the replication of MTB in macrophages. Moreover, Baek et al. [21] reported that the overexpression of *citA* in MTB caused an increased in the antibiotic sensitivity. The number studies

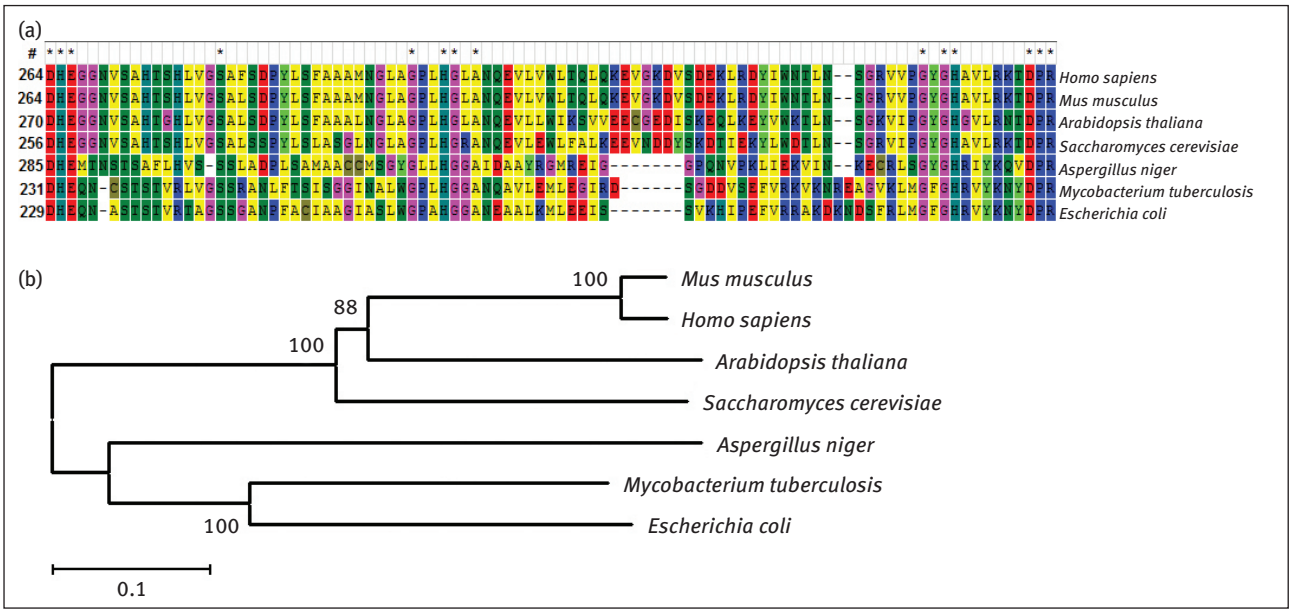

Figure 3: The citrate synthase proteins in phylogenetically unrelated organisms. Multiple alignment of the partial sequences (a) and the phylogenetic tree (b) based on the citrate synthase sequences. GltA2 sequence from *M. tuberculosis* was used. The stars and number sign show the conserved residues and the start position of the partial sequence, respectively. The colors indicate same or similar amino acids, and the dash lines represent the gaps. Numbers next to the branches show the percentages of replicate trees in which the associated taxa clustered together in the bootstrap test

Table 2: Physicochemical parameters of the CS family members in MTB.

Protein name	Average mass (Dalton)	Monoisotopic mass (Dalton)	Molar absorbance (cm^{-1})	Molar extinction coefficient at 280 nm ($cm^{-1}M^{-1}$)	Isoelectric point	Hydrophobicity index
CitA	40.143,89	40.118,50	1,25	50.190,00	5,41	-0,01
GltA2	47.976,49	47.946,38	0,92	44.070,00	5,35	-0,20
PrpC	42.965,54	42.937,82	0,86	37.080,00	9,31	-0,06

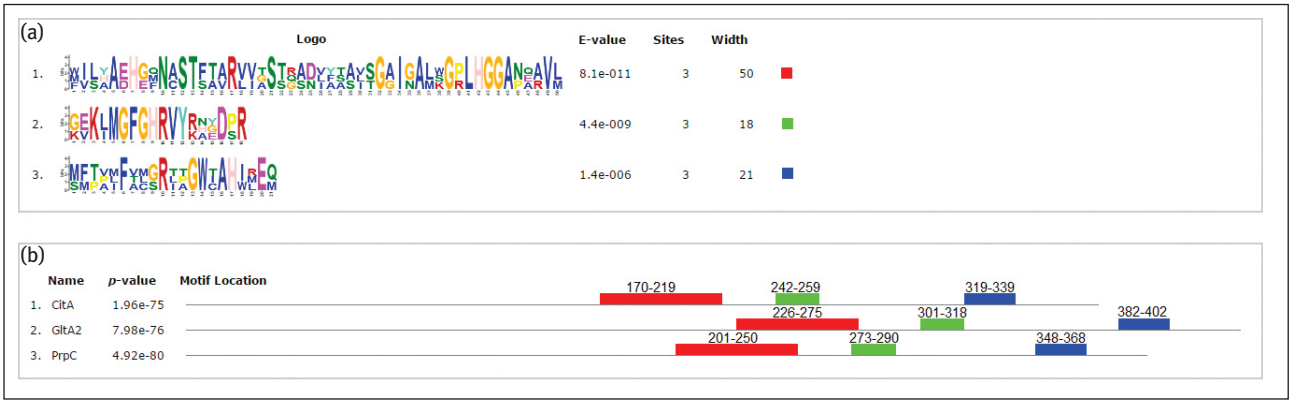

Figure 4: Motif analysis of the citrate synthase family members in *M. tuberculosis*. The logo representation of the motifs (a) and the localization of these motifs (b) in CitA, GltA2, and PrpC proteins. The numbers on the localizations indicate the residue positions.

on the family of CS proteins in bacteria is scarce. Jin and Sonenshein [22] investigated the CS family in *Bacillus subtilis* genome, and reported two CS members, namely CitA and CitZ, having 42% identity and 41 kDa molecular mass.

The ClustalW program was used in this study for the sequence alignment due to its advantages over the other

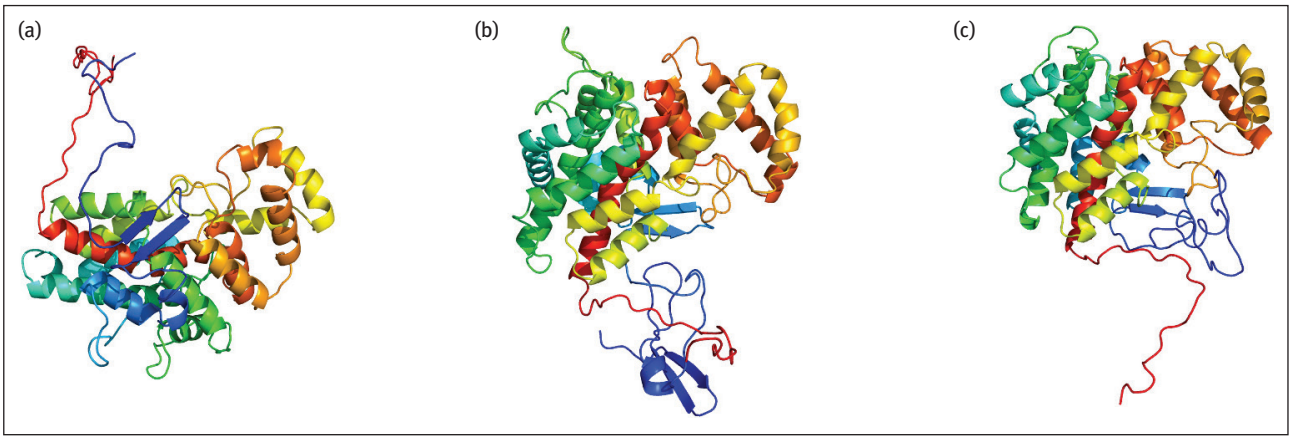

Figure 5: 3D modeling of the CitA (a), GltA2 (b), and PrpC (c) proteins from *M. tuberculosis* showing the α-helices and β-sheets in the tertiary structures.

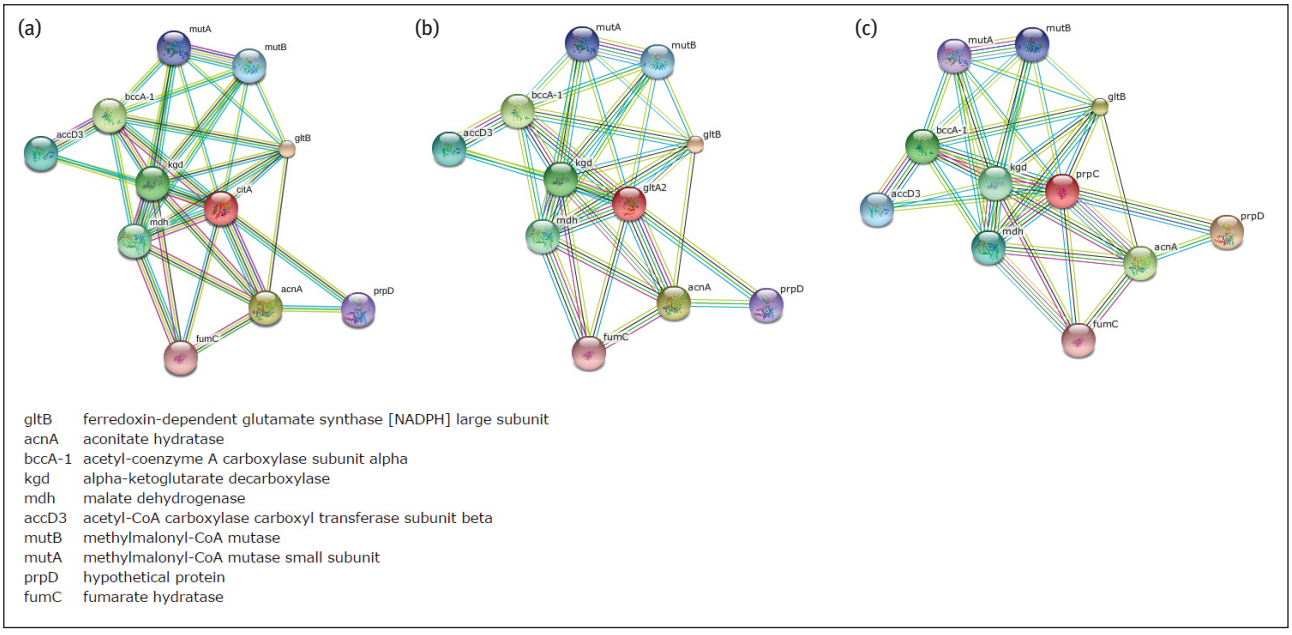

Figure 6: Interaction network of CitA (a), GltA2 (b), and PrpC (c) proteins from *M. tuberculosis* with the other proteins.

alignment tools such as MUSCLE, MAFFT or T-Coffee. The 'W' is the abbreviation of 'weights' which provides sensitive and efficient alignment of large sets of sequences even from unrelated organisms. Another advantage of ClustalW is the free and user-friendly graphical interface which is compatible with diverse operating systems and desktop environments [23,24].

The alignment of MTB CS proteins exhibited 62 conserved residues, and the CitA and PrpC were found to be more related phylogenetically than GltA2 (Fig. 1). The phylogenetic relationships of CitA (Fig. 2a), GltA2 (Fig. 2b), and PrpC (GltA1) (Fig. 2c) proteins among the mycobacteria were also investigated, and compared with the phylogenetic tree constructed based on *16S rRNA* genes (Fig. 2d). A similar pattern was observed in all of the four phylogenetic trees. The amino acid sequences of CS family members from the species in MTB complex, namely *M. tuberculosis*, *M. bovis*, *M. africanum*, and *M. canettii* were grouped together, and those from *M. marinum*, *M. ulcerans*, and *M. liflandii* were shown to be closely related, which are in the same mycobacterial complex [25]. The alignment (Fig. 3a) and phylogenetic analysis (Fig. 3b) of the CS members from diverse organisms exhibited a pattern similar to their evolutionary relationship. However, Bond et al. [26] reported that the CS sequences in *Geobacteraceae* have more similarity to eukaryotic ones as observed in *16S rRNA* sequences.

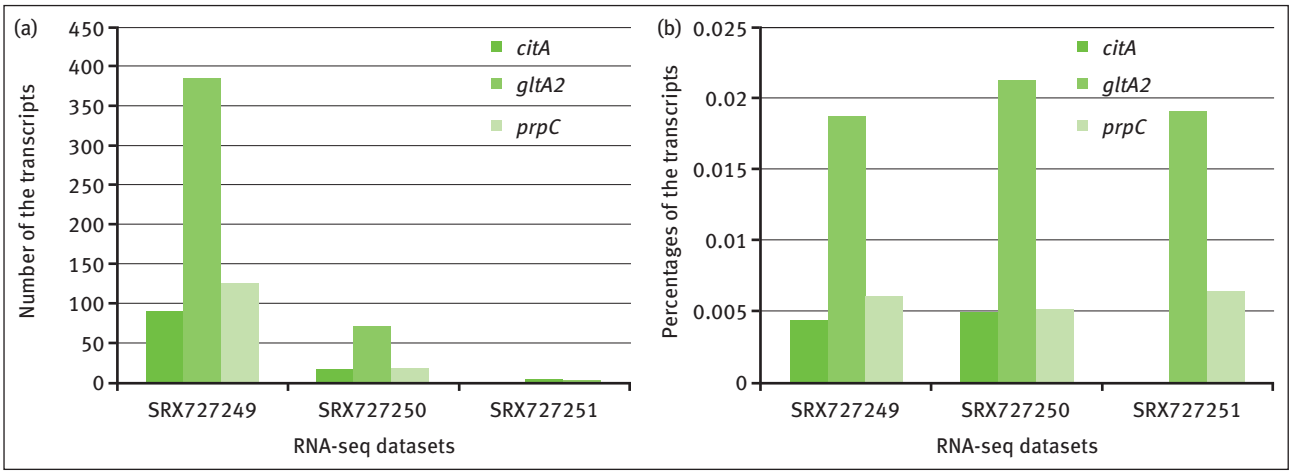

Figure 7: Expression profile of the *citA*, *gltA2*, and *prpC* in the wildtype *M. tuberculosis* H37Rv. The number of the transcripts (a) and their percentages with respect to the total number of transcripts (b) were detected in the RNA-seq datasets retrieved from the SRA database.

Structural properties of the CS family members in MTB

The protein family analysis revealed that the CitA contains two hidden Markov model (HMM) CS domains while a single CS domain was detected in GltA2 and PrpC (Table 2). Additionally, the amino acid sequences of MTB CS family members were analyzed to determine the conserved motifs, and three motifs covering LHGGA and MGFGHRVY were revealed (Fig. 4a). The locations of the motifs mainly found in *C*-terminal of the proteins (Fig. 4b). The BLAST analysis also showed that the GxIxAxxGxLHGGA and KxMGFGHRVYxxxDxR motifs are conserved in MTB CS members as well as those from the bacteria such as *Corynebacterium*, *Rhodococcus*, *Helicobacter*, *Bacillus*, and *Streptomyces*. The alignment of the CSs from the distant organisms (GltA2 from MTB) revealed that GxxHGxA and GxGHxxxxxxDPR are conserved phylogenetically (Fig. 3a).

The online, free, and user-friendly portal Protein Homology/analogY Recognition Engine v2.0 (Phyre2) was utilized for the 3D modeling of the CitA (Fig. 5a), GltA2 (Fig. 5b), and PrpC (Fig. 5c) proteins. For CitA and GltA2, the structure of *Acetobacter aceti* citrate synthase complexed with oxaloacetate and carboxymethyldethia coenzyme A (CMX) was used as template. On the other hand, crystal structure of methylcitrate synthase from MTB was utilized for PrpC homology. The tertiary structure of CS (monomer) proteins in MTB are mainly composed of α-helices with a few β-sheets. Ferraris et al. [7] reported that the GltA2 from MTB is a dimeric protein with unstructured *N*-terminal region, and consists of eight α-helices and two antiparallel small β-sheets. The major structure of CS from *Fasciola hepatica* with 469 amino acid length also has α-helix folding [10].

Analysis of the physicochemical parameters of the CS family proteins in MTB showed that the predicted molecular weight (MW) of CitA, GltA2, and PrpC monomers are approximately 40.1, 47.9, and 42.9 kDa, while isoelectric points (pI) were found to be 5.41, 5.35, and 9.31, respectively (Table 3). The MW and pI of the CS from *Geobacter sulfurreducens* were predicted as 49.8 kDa and 6.46, respectively [26]. The CS from *F. hepatica* has a MW of 52 kDa, and pI of 8.1 [10]. Min et al. [27] reported that the CitA from *Aspergillus nidulans* has 474 amino acids in length with a MW of 52.2 kDa. Moreover, Cheung et al. [28] characterized an iron-regulated CS (SbnG) from *Staphylococcus aureus*. The SbnG protein has 259 amino acids in length with a MW of 28.7 kDa, and pI of 5.46.

The cellular localization of the proteins was determined as cytoplasm with a score of 9.97 for CitA and GltA2, and 9.95 for PrpC. The interaction network of CitA (Fig. 6a), GltA2 (Fig. 6b), and PrpC (Fig. 6c) proteins with citrate synthase function revealed the connections with other proteins taking role in TCA cycle.

Gene expression profiles of the CS family members in MTB H37Rv

Transcriptome-wide expression analysis of the *citA*, *gltA2*, and *prpC* was conducted using the RNA-seq datasets obtained from the wild-type MTB H37Rv grown under aerobic conditions. The nucleotide sequences encoding LHGGA or MGFGHRVY motifs in each CS family member are varying, which are ctgcatggtggcgcg, cttcatggcggcgcc, and ctacacggcggcgcc or atggggttcgggcaccgggtctac, atgggtttcggtcatcgtgtctac, and atgggcttcgggcatcgggtgtac for CitA, GltA2, and PrpC, respectively. These sequences

and their reverse complements were used to find out the number of transcripts representing *citA*, *gltA2*, and *prpC* in three RNA-seq datasets from biological replicates. The number of transcripts (Fig. 7a) and their percentages with respect to the total number of transcripts in each dataset (Fig. 7b) were exhibited. The highest number of transcripts was detected as 458 for *gltA2*, and then 143 and 105 transcripts were found for *prpC* and *citA*, respectively. The rank order was found to be the same in all of the datasets.

This comprehensive in silico analysis provides a closer look into the CS family in MTB, which would confer a better insight into the genetics and physiology of this infectious agent. The expression profiles of CS genes in MTB under limited oxygen conditions should also be investigated to figure out their role in latency of the pathogen. Besides, the sub-cellular localization of the proteins can be visualized using a recombinant reporter protein such as GFP-tagging. Recently, isocitrate lyase, another TCA cycle enzyme, from MTB was shown to be taking role in the antibiotic tolerance of the pathogen [29]. For a future perspective, involvement of the CSs in the antibiotic tolerance of MTB may be investigated. Moreover, an attenuated strain that would be obtained via inactivation of a CS member in MTB might be used for vaccine development against tuberculosis.

Conflict of Interest: The author has no conflict of interest.

References

[1] Caetano LA, Almeida AJ, Gonćalves LM. Approaches to tuberculosis mucosal vaccine development using nanoparticles and microparticles: a review. J Biomed Nanotechnol 2014; 10(9):2295–316.

[2] World Health Organization. Global tuberculosis report 2014. ISBN 9789241564809. http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf?ua=1 (Last accessed: December 2015).

[3] Starck J, Källenius G, Marklund BI, Andersson DI, Akerlund T. Comparative proteome analysis of Mycobacterium tuberculosis grown under aerobic and anaerobic conditions. Microbiology 2004; 150(Pt 11):3821–9.

[4] Hartman T, Weinrick B, Vilchèze C, Berney M, Tufariello J, Cook GM, et al. Succinate dehydrogenase is the regulator of respiration in Mycobacterium tuberculosis. PLoS Pathog 2014; 10(11):1004510.

[5] Shi L, Sohaskey CD, Kana BD, Dawes S, North RJ, Mizrahi V, et al. Changes in energy metabolism of Mycobacterium tuberculosis in mouse lung and under in vitro conditions affecting aerobic respiration. Proc Natl Acad Sci U S A 2005; 102(43):15629&–34.

[6] Cook GM, Hards K, Vilchèze C, Hartman T, Berney M. Energetics of Respiration and Oxidative Phosphorylation in Mycobacteria. Microbiol Spectr 2014; 2(3).

[7] Ferraris DM, Spallek R, Oehlmann W, Singh M, Rizzi M. Structures of citrate synthase and malate dehydrogenase of Mycobacterium tuberculosis. Proteins 2015; 83(2):389–94.

[8] Daidone I, Roccatano D, Hayward S. Investigating the accessibility of the closed domain conformation of citrate synthase using essential dynamics sampling. J Mol Biol 2004; 339(3):515–25.

[9] Wilkins MJ, Callister SJ, Miletto M, Williams KH, Nicora CD, Lovley DR, et al. Development of a biomarker for Geobacter activity and strain composition; proteogenomic analysis of the citrate synthase protein during bioremediation of U(VI). Microb Biotechnol 2011; 4(1):55–63.

[10] Zinsser VL, Moore CM, Hoey EM, Trudgett A, Timson DJ. Citrate synthase from the liver fluke Fasciola hepatica. Parasitol Res 2013; 112(6):2413–7.

[11] van Ooyen J, Emer D, Bussmann M, Bott M, Eikmanns BJ, Eggeling L. Citrate synthase in Corynebacterium glutamicum is encoded by two gltA transcripts which are controlled by RamA, RamB, and GlxR. J Biotechnol 2011; 154(2-3):140–8.

[12] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013; 30(12):2725,9.

[13] Felsenstein J. Phylogenies and the comparative method. Am Nat 1985; 125: 1–15.

[14] Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 2010; 26(13):1608–15.

[15] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res 2014;42(Database issue):D222–30.

[16] Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res 2015; 43(Database issue):D257–60.

[17] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 2015; 10(6):845–58.

[18] Söding J. Protein homology detection by HMM-HMM comparison. Bioinformatics 2005; 21(7):951–60.

[19] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43(Database issue):D447–52.

[20] Muñoz-Elías EJ, Upton AM, Cherian J, McKinney JD. Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. Mol Microbiol 2006; 60(5):1109–22.

[21] Baek SH, Li AH, Sassetti CM. Metabolic regulation of mycobacterial growth and antibiotic sensitivity. PLoS Biol 2011; 9(5):e1001065.

[22] Jin S, Sonenshein AL. Identification of two distinct Bacillus subtilis citrate synthase genes. J Bacteriol 1994; 176(15):4669–79.

[23] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23(21):2947–8.

[24] Yihui L. A sequential iterative refinement optimization method to multiple sequence alignment. MSc. thesis. National

University of Singapore, 2004. http://scholarbank.nus.edu.sg/handle/10635/13895 (Last accessed: December 2015).
[25] Doig KD, Holt KE, Fyfe JA, Lavender CJ, Eddyani M, Portaels F, et al. On the origin of Mycobacterium ulcerans, the causative agent of Buruli ulcer. BMC Genomics 2012; 13:258.
[26] Bond DR, Mester T, Nesbø CL, Izquierdo-Lopez AV, Collart FL, Lovley DR. Characterization of citrate synthase from Geobacter sulfurreducens and evidence for a family of citrate synthases similar to those of eukaryotes throughout the Geobacteraceae. Appl Environ Microbiol 2005; 71(7):3858,65.
[27] Min IS, Bang JY, Seo SW, Lee CH, Maeng PJ. Differential expression of citA gene encoding the mitochondrial citrate synthase of Aspergillus nidulans in response to developmental status and carbon sources. J Microbiol 2010; 48(2):188–98.
[28] Cheung J, Murphy ME, Heinrichs DE. Discovery of an iron-regulated citrate synthase in Staphylococcus aureus. Chem Biol 2012;19(12):1568–78.
[29] Nandakumar M, Nathan C, Rhee KY. Isocitrate lyase mediates broad antibiotic tolerance in Mycobacterium tuberculosis. Nat Commun 2014; 5:4306.