Biochimie 127 (2016) 241–248

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Hypothesis

Functional divergence of HBHA from *Mycobacterium tuberculosis* and its evolutionary relationship with TadA from *Rhodococcus opacus*

Mariana P. Lanfranconi, Héctor M. Alvarez*

Instituto de Biociencias de la Patagonia (INBIOP), Universidad Nacional de la Patagonia San Juan Bosco y CONICET, Km 4-Ciudad Universitaria 9000, Comodoro Rivadavia, Chubut, Argentina

ARTICLE INFO

Article history:
Received 19 February 2016
Accepted 6 June 2016
Available online 7 June 2016

Keywords:
HBHA
TadA
Lipid bodies
Virulence factor
Mycobacterium
Rhodococcus

ABSTRACT

Rhodococcus opacus PD630 and *Rhodococcus jostii* RHA1 are oleaginous bacteria able to synthesize and accumulate triacylglycerols (TAG) in lipid bodies (LB). Highly relevant to the structure of LB is a protein homologous to heparin-binding hemagglutinin (HBHA) (called TadA in rhodococci), which is a virulence factor found in *Mycobacterium tuberculosis*. HBHA is an adhesin involved in binding to non-phagocytic cells and extrapulmonary dissemination. We observed a conserved synteny of three genes encoding a transcriptional regulator (TR), the HBHA protein and a membrane protein (MP) between TAG-accumulating actinobacteria belonging to *Rhodococcus*, *Mycobacterium*, *Nocardia* and *Dietzia* genera, among others. A 354 bp-intergenic spacing containing a SigF-binding site was found between *hbha* and the TR genes in *M. tuberculosis*, which was absent in genomes of other investigated actinobacteria. Analyses of available "omic" information revealed that TadA and TR were co-induced in rhodococci under TAG-accumulating conditions; whereas in *M. tuberculosis* and *Mycobacterium smegmatis*, HBHA and TR were regulated independently under stress conditions occurring during infection. We also found differences in protein lengths, domain content and distribution between HBHA and TadA proteins from mycobacteria and rhodococci, which may explain their different roles in cells. Based on the combination of results obtained in model actinobacteria, we hypothesize that HBHA and TadA proteins originated from a common ancestor, but later suffered a process of functional divergence during evolution. Thus, rhodococcal TadA probably has maintained its original role; whereas HBHA may have evolved as a virulence factor in pathogenic mycobacteria.

© 2016 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

1. Introduction

The term Terrabacteria was coined to refer to aquatic bacteria colonizing and occupying ancient terrestrial environments [1]. Terrabacteria comprise a subset of phyla, including Actinobacteria, Cyanobacteria, *Deinococcus-Thermus* and Firmicutes [2]. The process of land colonization must have been one of the most challenging aspects of bacterial evolution, which involved significant genomic changes and metabolic innovations to cope with the terrestrial environment. These microorganisms have developed diverse mechanisms in response to the main terrestrial challenges, such as ultraviolet radiation, oxidative and osmotic stresses, and desiccation, among others [3–5]. Battistuzzi et al. [1] suggested that land has been colonized multiple times in different lineages. In this context, mycolic-acid containing actinobacteria, including *Mycobacterium*, *Rhodococcus*, and *Nocardia* genera, among others, may have evolved some highly conservative strategies for adapting to terrestrial environments; such as the formation of a very robust mycolic acid-containing cellular envelope, the synthesis of carotenoid pigments, and the production of lipid esters. Interestingly, Finkelstein and collaborators [6] suggested that the ability to accumulate neutral lipids could have been a strategy of microorganisms to survive in earliest terrestrial environments. Long-chain wax esters (WE) and triacylglycerols (TAG) are less volatile and protect against desiccation [6]. Stored lipids also provide a source of endogenous metabolic water during the oxidation of fatty acids under desiccation conditions [3]. In a previous study, we demonstrated that TAG were required for the tolerance against UV-radiation and essential for the response of the extremophile *Rhodococcus* sp. A5 to desiccation conditions [4]. In general, mycolic

* Corresponding author.
E-mail address: halvarez@unpata.edu.ar (H.M. Alvarez).

http://dx.doi.org/10.1016/j.biochi.2016.06.002
0300-9084/© 2016 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

acid-producing actinobacteria are able to synthesize and accumulate variable amounts of TAG, whereas some species of *Rhodococcus*, such as *Rhodococcus opacus* and *Rhodococcus jostii*, are specialists for this trait [7]. The ability to accumulate TAG demands the presence of a complete set of genes/proteins integrated in a specialized metabolic network. TAG-accumulating machinery in actinobacteria, which includes specific enzymes involved in metabolism, transporter proteins, structural components of lipid inclusions, and transcriptional regulators; was probably designed during evolution in water-to-land transition. The processes and enzymatic machinery involved in TAG and WE synthesis have been largely studied in Gram positive actinobacteria (reviewed in Refs. [7,8]). Different genes from these microorganisms directly related to TAG synthesis and accumulation have been cloned and functionally characterized. MacEachran et al. [9] identified and characterized a lipid body-associated protein (called TadA) with structural role in *Rhodococcus opacus* PD630. This 276-residues structural protein participates in the lipid body assembly in the oleaginous strain PD630. The mutation of *tadA* led to a significant reduction of TAG accumulation, and to an alteration of lipid body size, in comparison to the wild type strain [9]. Its ortholog in *R. jostii* RHA1 has also been studied [10], and its structural role in lipid bodies confirmed [11]. TadA-like protein was one of the three most abundant proteins occurring during TAG accumulation, as revealed a proteome analysis applied to *R. jostii* RHA1 [10,12]. This protein exhibits high similarity to the heparin-binding hemagglutinin (HBHA) from *Mycobacterium tuberculosis* and binds heparin [9]. HBHA from pathogenic mycobacteria is a cell-wall protein that binds to heparin sulphate glycosaminoglycans found on the surface of human epithelial cells during infection [13]. It has been demonstrated that HBHA is an adhesin protein which promotes entry and dissemination of *M. tuberculosis* into host cells [13,14]. This protein is expressed on the surface of mycobacterial cells and is recognized by the immune system of the host [15]. Its role in bacterial-bacterial aggregation and the interaction with host cells and matrix components has been demonstrated [16]. The homologous HBHA protein of *Mycobacterium smegmatis*, a non-pathogenic mycobacterium, presents a diverging sequence in comparison to that of *M. tuberculosis* [15]. Despite being also expressed on the cell surface by *M. smegmatis*, it is not involved in epithelial adherence and has a low affinity for heparin [17]. Interestingly, Delogu et al. [18] suggested that *M. tuberculosis*-HBHA may have evolved as an adhesin in pathogenic mycobacteria from a homolog that serves a different but unknown function in a saprophytic mycobacterium. All these studies suggest a functional promiscuity or functional diversification for HBHA in *Mycobacterium* and its homolog TadA in *Rhodococcus*, with a possible role in the transition to soil-to-human host jump in pathogenic actinobacteria. The role and function of HBHA in *Mycobacterium*, as well as of its TadA-homolog in *Rhodococcus*, have been previously elucidated. This available information provides the possibility for a comprehensive comparative analysis based principally on *Mycobacterium* and *Rhodococcus*, by combining evidence from the taxonomic distribution of these proteins, their genomic organization, sequences and expression during different conditions using the available "omic" information. We also provide a unified view on the possible function diversification within an evolutionary framework for these soil-adapted actinobacteria.

2. Material and methods

2.1. TadA distribution

Homologs of TadA were searched at the BLAST server of the National Center for Biotechnology Information (NCBI) using BLASTP. A data file containing the most related sequences, representatives of every genus retrieved, was used to conduct sequence alignments with CLUSTALX with default parameters [19]. The similarity of amino acid sequences was analyzed using the sequence identity matrix incorporated in BioEdit software. NJ, with Kimura approximation for calculation of distance matrices, and parsimony (P) treeing methods for phylogenetic analysis were done using the software package PHYLIP [20]. Bootstrap resampling of the NJ and P trees (1000 and 100 replicates, respectively) was performed to provide statistical confidence to the topologies inferred.

2.2. Genomic context

tadA/hbha and their adjacent genes were screened in order to analyze their co-occurrence and predict functional associations. For this, we considered available complete sequenced genomes of relevant strains detected in 2.1. Promotor sequences and transcription factor-binding sites in intergenic regions were searched by BPROM program [21].

2.3. Sequence comparison

Sequences of TadA/HBHA from *R. opacus* PD630, *R. jostii* RHA, *Rhodococcus equi* (the only known animal pathogen of the genus), and their homologs in *M. tuberculosis* H37Rv and *M. smegmatis* mc^2 155, which were previously used to construct the phylogenetic tree in 2.1, were aligned with CLUSTALX and visually inspected to detect motifs and regions that could differentiate or relate TadA and HBHA. Two programs, COILS [22] and MARCOIL [23] with default parameters, were used to predict coiled-coil regions in TadA and HBHA.

2.4. Analysis of "omic information" related to TadA and HBHA expression patterns

Transcriptomic and proteomic studies analyzed in this work correspond to those reports of the strains studied in 2.3: *R. opacus* PD630 [10], *R. jostii* RHA1 [9,11], *M. tuberculosis* H37Rv [24–27], and *M. smegmatis* mc2 155 [27]. In this study we retrieved the expression profiles of TadA/HBHA and its flanking genes, under stress conditions found during cell infection (in the case of mycobacteria), or those that promoted TAG accumulation (in the case of rhodococci).

3. Results

3.1. Distribution of tadA/hbha

tadA/hbha genes are widely distributed among Gram positive Actinobacteria as a single-copy gene. In contrast, these genes were not detected in neutral lipid-synthesizing Proteobacteria. Phylogenetic analysis showed high sequence heterogeneity among *tadA/hbha* genes from the analyzed actinobacteria (Fig. 1). *tadA/hbha* sequences belonging to these microorganisms were dispersed in the phylogenetic tree with two major clusters. Sequences from rhodococci formed a major cluster, containing several sub-clusters with the different rhodococcal species (Fig. 1). *tadA/hbha* from *Nocardia farcinica* was located within this major cluster. Sequences from oleaginous rhodococcal species, such as *R. opacus* PD630 and *R. jostii* RHA1, together with *Rhodococcus wratislaviensis* and *Rhodococcus imtechensis*, were almost identical and formed a tight subcluster (Fig 1). *tadA/hbha* gene from *R. equi* and *Rhodococcus fascians*, the only animal and plant pathogens respectively, appeared separated from those of these oleaginous rhodococci (Fig. 1). The

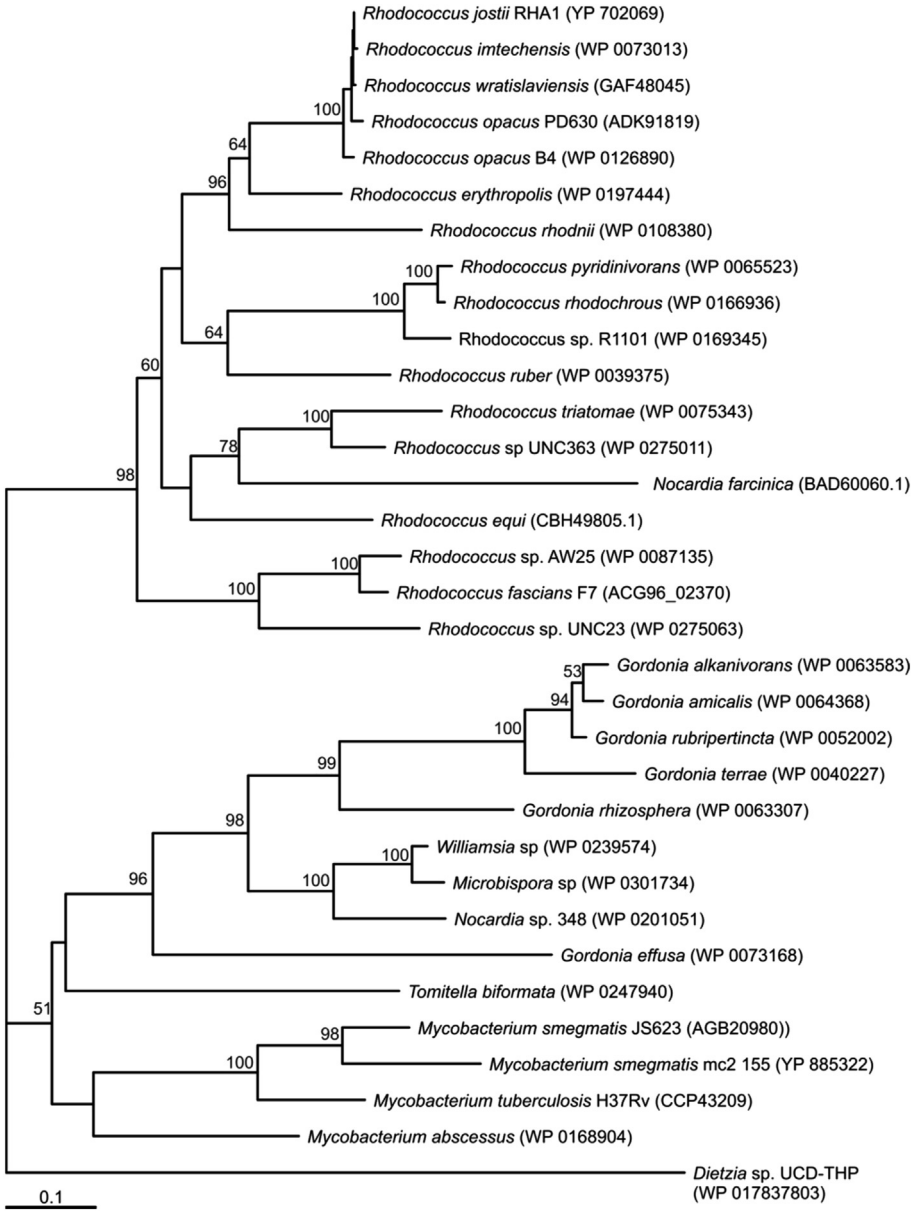

Fig. 1. Phylogenetic tree of TadA/HBHA proteins from diverse species of actinobacteria.

second major group includes *tadA*/*hbha* sequences from *Gordonia*, *Williamsia*, *Microbispora*, *Tomitella* and *Mycobacterium* species (Fig. 1). Within this heterogeneous group, *tadA*/*hbha* sequences from mycobacteria clustered separately from the rest of the actinobacteria. Interestingly, *tadA*/*hbha* gene is not present in the genome of *Streptomyces coelicolor*, which is an actinobacterial bacterium able to produce TAG [28].

Taken together, phylogenetic analysis of *tadA*/*hbha* sequences showed a high divergence between the different actinobacterial genera, with a high genetic heterogeneity inclusive between species belonging to the same genus.

3.2. Synteny and genomic organization

The occurrence and genome organization of *tadA*/*hbha* are highly conserved in the different actinobacteria, as is shown in Fig. 2. This gene was flanked by a gene encoding a transcriptional regulator (TR), and another coding for a membrane protein (MP) (Fig. 2). Sequence identities and sizes of the three genes/proteins from different actinobacterial strains were compared in Table 1. In general, the TR and MP protein lengths were well conserved. However, some differences in the sequence length were more evident between TadA/HBHA proteins (199–285 amino acids) (Table 1). Interestingly, TadA/HBHA from *M. tuberculosis* was significantly shorter than that of *M. smegmatis*.

Taking the sequences of *R. opacus* PD630 as reference, the lowest identities were obtained with TadA/HBHA and the MP from mycobacteria (46–48%) (Table 1).

Although there is synteny in these loci, the transcription orientation and the spacing between genes varied in the analyzed actinobacteria, as is shown in Fig. 2. In general, *tadA*/*hbha* gene is very close located to the gene coding for the TR in rhodococcal strains; thus, they are probably co-transcribed. Short intergenic regions appear between genes encoding TadA/HBHA and the TR in *Dietzia* sp., *Tomitella biformata* and *M. smegmatis*. Interestingly, there is an intergenic region of 354 bp between these two genes in

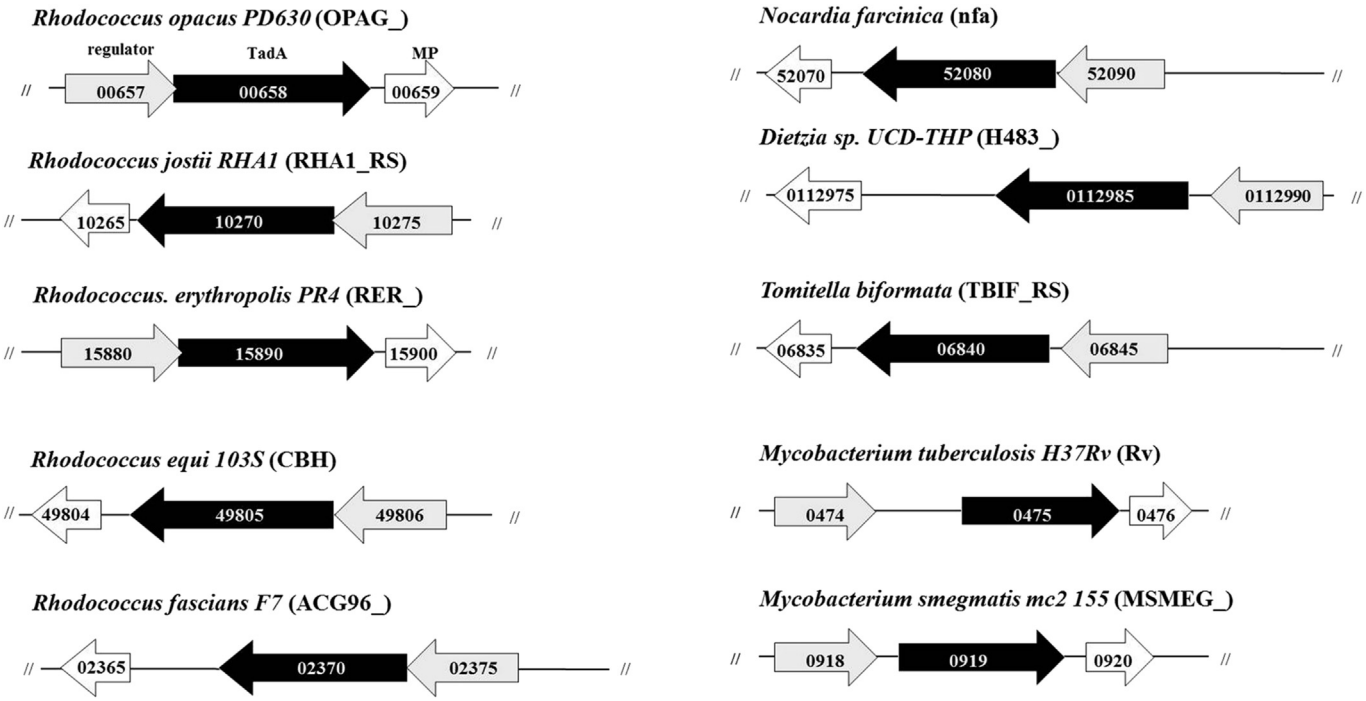

Fig. 2. Genomic organization of *tadA/hbha* genes. The arrows indicate length and transcriptional orientation. Genes encoding the putative transcriptional regulator are highlighted in grey, *tadA/hbha* genes in black, and those encoding a membrane protein in white. Genes are shown by the annotated number of the protein they encode; the prefix of the bacterial species appears in parenthesis after the name.

Table 1
Comparative analysis of HBHA/TadA and adjacent coding genes in different bacterial strains. Abbreviations: TR, transcriptional regulator; MP, membrane protein; aa, amino acids; id, identity.

Strain	TR		TadA/HBHA		MP	
	Size (aa)	id[a] (%)	Size (aa)	id[a] (%)	Size (aa)	id[a] (%)
Rhodococcus opacus PD630	168	–	276	–	97	–
Rhodococcus jostii RHA1	168	99	276	99	97	100
Rhodococcus erythropolis PR4	168	90	275	79	98	66
Rhodococcus fascians F7	156	77	265	61	99	57
Rhodococcus equi 103S	155	77	285	69	97	68
Mycobacterium tuberculosis H37Rv	140	67	199	48	87	49
Mycobacterium smegmatis mc^2 155	142	74	232	46	94	46
Nocardia farcinica IFM 10152	137	85	221	55	95	53
Tomitella biformata AHU 1821	149	76	257	50	93	51

[a] In relation to *Rhodococcus opacus* PD630.

the genome of *M. tuberculosis*, as is shown in Fig. 2. After analyzing this intergenic region, the BPROM program predicted the occurrence of putative promoters, which could be recognized by at least three different DNA-binding proteins (*ihf*: TGTAAGAA at position 310; *rpoD17*: AATAGTTA at position 332 and *hns*: AAAGGAAT at position 343), and confirmed the occurrence of the binding site for SigF (TGTTT-15-GGGTA), which has been characterized in a previous study [29]. In the case of *M. smegmatis*, the program did not detect any putative binding site between both genes.

3.3. Sequence comparison

An alignment of TadA/HBHA sequences from *R. opacus* PD630, *R. jostii* RHA1, *R. equi* 103S, *M. tuberculosis* HRv37 and *M. smegmatis* mc^2 155 was performed in order to analyze the presence/absence of certain domains and their level of conservation. Fig. 3 shows that the differences of sequence lengths between TadA/HBHA proteins are not simply associated with a truncated terminal end. Instead, there is a lack of certain domains (in the case of *M. tuberculosis*), or a reduced version of them (in the case of *M. smegmatis*), in comparison to TadA/HBHA from rhodococci (Fig. 3). The multiple-sequence alignments revealed the presence of a hydrophobic region rich in leucine (L) in the N-terminal of *M. tuberculosis* (7 residues). The L-rich stretch is reduced to five residues in *M. smegmatis*. In contrast, only one L is found in *Rhodococcus* and it is not strictly conserved at the same position in the three species analyzed.

Esposito and collaborators [30] demonstrated the existence of specific domains within coiled-coil regions in the N-terminal of mycobacterial HBHA that are involved in HBHA dimerization. These coiled-coil regions dictate bacterial agglutination [31]. The existence of coiled-coil domains in TadA were predicted using COILS and MARCOIL programs, as is shown in Fig. S1. As reference, both programs also confirmed the occurrence of coiled-coil domains in the N-terminal end of mycobacterial HBHA (Fig. S1). Whether these domains detected in rhodococcal proteins can mediate lipid droplet aggregation, remains to be investigated in the future.

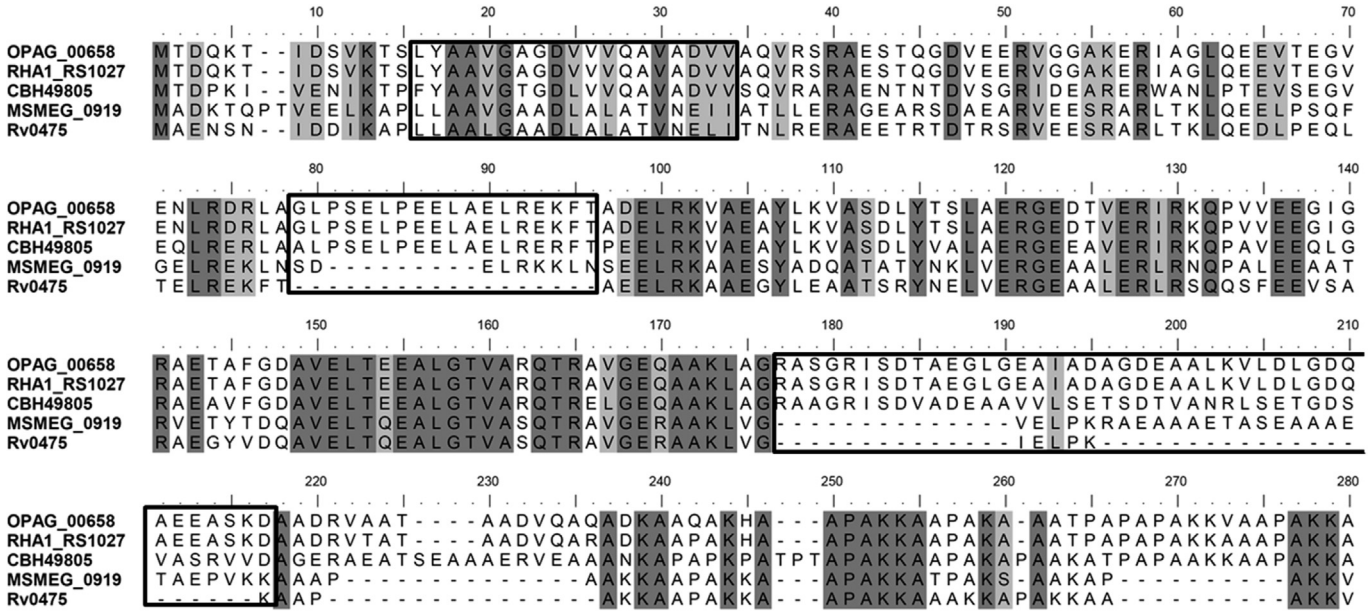

Fig. 3. Sequence alignment of TadA/HBHA proteins. Identical amino acid residues are highlighted in dark grey, similar amino acid residues appear in light grey. Regions showing the highest differences between Rv0475 and OPAG_00658 are marked as boxes. Strains/Proteins: *R. opacus* PD630 (OPAG_00658), *R. jostii* RHA1 (RHA1_RS10270), *R. equi* 103S (CBH49805), *M. smegmatis* mc2 155 (MSMEG_0919), *M. tuberculosis* H37Rv (Rv0475).

Also in the N-terminal, there is a stretch of 18 amino acids containing acidic residues (D and E) present in *Rhodococcus* species, which is shorter in *M. smegmatis* and absent in *M. tuberculosis* (Fig. 3). From Glu99 to Gly176 in *R. opacus* PD630, all sequences exhibit a highly conserved region. The lysine (K)-rich heparin binding site found at the C-terminal of HBHA of *M. tuberculosis*, is preceded by a domain containing acidic residues in *M. smegmatis* (shorter) and in rhodococci (longer) (Fig. 3). Furthermore, the number and distribution of K pairs differ not only between TadA/HBHA from *M. tuberculosis* and its homolog in *M. smegmatis*, but also when compared to rhodococcal proteins (Fig. 3).

3.4. Omic studies

Results from previously reported transcriptomic and proteomic studies performed for *Rhodococcus* and *Mycobacterium* strains under conditions favoring TAG accumulation or infection, were collected in this study. We analyzed the expression patterns of the three genes/proteins of *M. tuberculosis*, *M. smegmatis*, *R. opacus* and *R. jostii*, which were integrated in Table 2. In general, we observed two different expression patterns of the three considered genes/proteins, when combining results and comparing between mycobacteria and rhodococci. When cells of *R. opacus* PD630 and *R. jostii* RHA1 were cultivated under TAG-accumulating conditions (nitrogen limitation), TadA/HBHA, the adjacent TR, as well as some specific genes/proteins involved in TAG synthesis (diacylglycerol acyltransferases, DGAT's), exhibited the same expression pattern (up-regulation) (Table 2). During "omic" analyses of *M. tuberculosis* and *M. smegmatis* under stress or nutrient starvation conditions, which mimic conditions found during infection, *tadA/hbha* was usually up-regulated, but not the adjacent TR. Moreover, we could observe inverted expression patterns between *tadA/hbha* and DGAT enzymes (TAG biosynthetic enzymes) in *M. tuberculosis* and *M. smegmatis*.

Taken together, these results suggested that *tadA/hbha* and the gene coding for the TR found upstream are co-transcribed in rhodococci during conditions that promote TAG accumulation; in contrast to mycobacteria, where TadA/HBHA seems to be under the control of other regulatory module.

4. Discussion

Genomic sequencing and mapping have enabled comparison of genomes sequences of many different bacterial species. Based on our results, the gene coding for TadA/HBHA protein is uniquely found in actinobacteria with the ability to accumulate TAG, but not in neutral lipid-producing proteobacteria. The co-localization of this gene together with two genes encoding a TR and a MP; is highly conserved in genomes of different TAG-accumulating actinobacteria (Figs. 1 and 2). This conserved synteny suggests that these genes are homologous and probably share a functional relationship. Phylogenetic analysis of TadA/HBHA sequences and the preservation of the precise order of these three genes on chromosomes of different actinobacteria suggest a relevant function on physiology, passed down from a common ancestor. TadA/HBHA proteins have been thoroughly characterized for *Mycobacterium tuberculosis* and *Rhodococcus opacus*, but interestingly, it was demonstrated that they possess different localization and functions in cells. MacEachran et al. [9] reported that TadA protein localizes to lipid bodies in *R. opacus* PD630 and plays a role in lipid body assembly and maturation in the intracellular milieu, promoting the aggregation of small into larger lipid bodies. In contrast, the homologous protein HBHA is associated to the cell wall and exposed on the surface in *M. tuberculosis*, promoting bacterial-bacterial aggregation and the interaction with host cells during infection [31]. Thus, the available evidence demonstrates that TadA/HBHA is a component of the TAG-accumulating machinery in rhodococci, whereas the homolog in pathogenic mycobacteria is part of the invasion mechanism to host cells. It is noteworthy that the possible role of mycobacterial HBHA in lipid body formation has never been directly investigated. In this context, when proteins associated with intracellular lipid droplets in *Mycobacterium bovis* were analyzed, HBHA was not detected [32].

In this study, we compared TadA/HBHA sequences from rhodococci and mycobacteria, and analyzed the conservation of relevant domains, since the protein structure is tightly related to its

Table 2
Expression patterns of genes/proteins under conditions favoring TAG accumulation or infection, collected from available "*omics*" studies. Abbreviations: TR, transcriptional regulator; MP, membrane protein; DGAT, diacylglycerol acyltransferase; h, hours; d, days.

Strains	Genes	Protein names	Expression	Conditions	References
Mycobacterium tuberculosis H37Rv	Rv0475	TadA/HBHA	Upregulated	Multi-stress 3 h	[25]
	Rv3130c	DGAT	Downregulated	Multi-stress 3 h	[25]
	Rv0476	MP	Upregulated	Nutrient starvation 4 d	[26]
	Rv0474	TR	Downregulated	Nutrient starvation 4 d	[26]
	Rv0475	TadA/HBHA	Upregulated	Nutrient starvation 7 d	[27]
	Rv3130c	DGAT	Downregulated	Nutrient starvation 9 d	[24]
Mycobacterium smegmatis mc^2 155	MSMEG_0919	TadA/HBHA	Upregulated	Multi-stress 3 h	[25]
	MSMEG_6322	DGAT	Downregulated	Multi-stress 3 h	[25]
Rhodococcus opacus PD630	LPD06283	TadA/HBHA	Upregulated	MSM 24 h	[11]
	LPD06283	TadA/HBHA	Upregulated	MSM 3 h	[11]
	LPD06284	TR	Upregulated	MSM 3 h	[11]
	LPD02996	DGAT	Upregulated	MSM 3 h	[11]
Rhodococcus jostii RHA1	RHA1_RS10270	TadA/HBHA	Upregulated	High C:N 24 h	[10]
	RHA1_RS10270	TadA/HBHA	Upregulated	MSM0 8 h	[12]
	RHA1_RS10275	TR	Upregulated	MSM0 8 h	[12]
	RHA1_RS07790	DGAT	Upregulated	MSM0 8 h	[12]
	RHA1_RS26160	DGAT	Upregulated	MSM0 8 h	[12]
	RHA1_RS30960	DGAT	Upregulated	MSM0 8 h	[12]

function. We observed differences in the protein lengths principally between those of *R. opacus* and *M. tuberculosis* (276 and 199 aa, respectively), but also between *M. smegmatis* and *M. tuberculosis* (232 and 199 aa, respectively). TadA/HBHA of the pathogenic mycobacterium probably became shorter during evolution, but preserved the main domains linked to its functionality. Specific coiled-coil domains found in the N-terminal end of HBHA can promote interactions among proteins leading to reversible dimerization processes [30], which contribute to formation of cell aggregates [31]. In this study, we predicted the occurrence of a putative coiled-coil domain in the N-terminal end of rhodococcal TadA proteins (Fig. S1), which may contribute to formation of LD aggregates in these oleaginous actinobacteria. This putative coiled-coil domain of TadA might promote interactions among proteins attached to different LD's, resulting in the aggregation of small into larger lipid bodies during TAG accumulation. The role of TadA in rhodococci and probably also in mycobacteria may be part of a strategy to accommodate large amounts of TAG in cells. Further studies are necessary to fully understand the functions of these proteins in actinobacteria.

One of the observed differences between sequences was the occurrence of a leucine-rich hydrophobic region localized in the N-terminal region of *M. tuberculosis* HBHA, which is involved in anchoring to the cell wall, as has been reported previously [31]. Leucine-rich domains in proteins may function as a basement membrane anchor in cells [33]. Interestingly, leucine residues are not abundant in the same domain in rhodococci, which could be consistent with the different localization postulated for TadA in the intracellular lipid bodies [9]. This domain architecture in rhodococci should be adequate for the interaction of the protein with the phospholipids monolayer of lipid bodies. In contrast, the association to cell wall structures at the surface of cells may require a special sequence adaptation of the N-terminal region in *M. tuberculosis* HBHA.

On the other hand, HBHA from *M. tuberculosis* binds to host components via its C-terminal domain [34]. This domain is required to adhere to the heparan-sulphate proteoglycans present on the surface of epithelial cells, and the receptor-mediated endocytosis results in a transcellular migration of the HBHA-containing particles [18]. HBHA-mediated binding to host components involves electrostatic interactions promoted by the lysine-rich domain at the C-terminal end, which is exposed on the cell surface [31]. Within this domain in *M. tuberculosis*, two lysine-rich pentapeptides constitute the high-affinity binding site to heparan-sulphate proteoglycans present on epithelial cells [35]. KAAAK and KAPAK are present in all species analyzed, with *M. tuberculosis* H37Rv containing the highest number of repeats (two and one, respectively), followed by *R. equi* 103S with two consecutive KAPAK penta-repeat. Based on these findings, TadA/HBHA proteins from the two pathogenic representatives of mycobacteria and rhodococci exhibited the highest potential affinity to heparin.

The HBHA from the saprophytic *M. smegmatis* contains an extra acidic-rich domain prior to the C-terminal end that reduces the heparin binding capacity in comparison to the adhesin from *M. tuberculosis* [12] (Fig. 3). Based on this observation and experimental data, Delogu et al. [18] suggested that *M. tuberculosis* HBHA may have evolved as an adhesin in pathogenic mycobacteria from a homolog that serves a different but unknown function in saprophytic mycobacteria. In this context, we found a larger extra acidic-rich domain inserted prior to the C-terminal region of TadA/HBHA from rhodococci (Fig. 3). MacEachran et al. [9] reported that C-terminus of TadA/HBHA protein is essential for lipid body aggregation. Altogether, these results suggested that the differential functions of TadA/HBHA in saprophytic and pathogenic actinobacteria are supported by specific modifications of the protein architecture.

In addition to the changes in the TadA/HBHA sequences occurring in actinobacteria, some rearrangements to this locus seem to have occurred in mycobacteria during evolution. A 354 bp-intergenic spacing is found between *hbha* and the TR genes in the genome of *M. tuberculosis*; whereas these genes are closely linked in rhodococci, as is shown in Fig. 2. Interestingly, the upstream region of *hbha* contains a functional binding site recognized by the sigma factor SigF in *M. tuberculosis*. Hartkoorn et al. [29] demonstrated that SigF governs expression of *hbha* in this bacterium in response to stationary-phase stress, cold shock or nitrogen depletion. Thus, *hbha* is included in the SigF regulon in *M. tuberculosis*, together with other genes involved in lipid and intermediary metabolism and virulence [29]. Moreover, using the BPROM program we found a TATA motif and a TTGCCC box in the upstream region of *hbha* in *M. tuberculosis*, which were not found in the other investigated actinobacterial strains. Among three possible DNA-binding factors that could recognize and bind to these putative promoters in *M. tuberculosis*, the IHF protein is an integration host factor playing a central role in bacterial pathogenesis, as has been reported previously [36]. These findings suggest that *hbha*

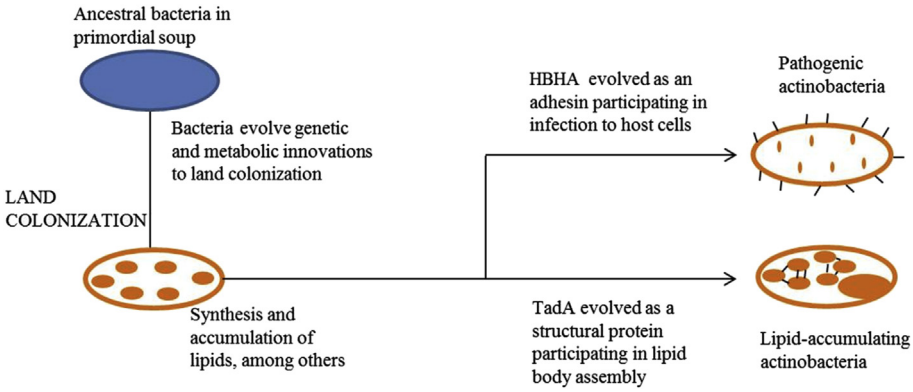

Fig. 4. Schematic representation of the possible events occurring during evolution that resulted in functional divergence between bacterial TadA and HBHA.

possesses an independent regulation in pathogenic mycobacteria; whereas in rhodococci is probable co-transcribed with the TR gene. The available "omic" information analyzed in this study supported this presumption. The *hbha* gene in *M. tuberculosis* was expressed independently of the other two adjacent genes, coding for the TR and MP, during the responses of cells against diverse stresses occurring during infection (Table 2). In addition, no correlation in the gene expression tendency between *hbha* gene and *dgat* genes (involved in TAG synthesis) was found. In contrast, transcriptomic and proteomic studies performed for *R. opacus* and *R. jostii* during TAG-accumulating conditions showed the concomitant up-regulation of TadA/HBHA, the TR and DGAT enzymes involved in TAG accumulation (Table 2). All these results demonstrated that TadA/HBHA is induced under conditions occurring during infection in pathogenic mycobacteria, and during TAG accumulation in rhodococci. We could speculate that HBHA is part of a mechanism of virulence evolution based on the cooption of core microbial traits, probably triggered by host niche-adaptive events. The recruitment or cooption of existing microbial traits may provide access to the host niche during virulence evolution. This mechanism that allows rapid adaptive change and novel trait acquisition has been reported for *R. equi* during the emergence and evolution of its pathogenicity [37].

Taken these results into account, we hypothesize that mycobacteria, rhodococci and closely related actinobacteria are organisms of relatively recent divergence which evolve from a common ancestor. This common ancestral organism probably developed diverse genetic and metabolic innovations to cope with terrestrial environments, since these microorganisms belong to the Terrabacteria group according to previous studies [1,2,38]. One of the traits evolved by these microorganisms was the ability to synthesize and accumulate neutral lipids (TAG and/or wax esters), which contributed considerably to the success of bacterial land colonization [6]. The capacity for accumulating neutral lipids, which is more prevalent among terrestrial than aquatic bacteria, allowed cells adapting and tolerating the main stresses found in soil environments, such as desiccation, osmotic stress, and oxidative stress, among others [3–5]. The massive biosynthesis and accumulation of TAG by actinobacteria demanded the evolution of complex and specific genetic and metabolic networks probably triggered by the harsh terrestrial conditions (Fig. 4). Among the different components of the TAG-accumulating machinery designed by actinobacteria, including diverse enzymes such as TAG synthases, mycolyltransferases, acyl-CoA synthetases, lipases, among others; there were genes coding for proteins involved in the assembly, stabilization and structure of lipid bodies, such as TadA/HBHA [7]. Thus, TadA/HBHA is induced during TAG accumulation [12], and binds to small lipid droplets during the early stages of lipid storage, promoting aggregation into larger lipid bodies, as has been reported for *R. opacus* PD630 [9]. Some actinobacteria are supposed to transit from soil to plant host or have experienced soil-to-human/animal jump [38]. These microorganisms experienced divergent evolutionary processes by which traits evolved for one purpose, were then employed for a new role. In this context, TadA/HBHA of different actinobacterial species with a common origin diverged somewhat in mycobacteria in overall structure and function (Fig. 4). The evolutionary change of TadA/HBHA occurred at two different levels: (1) the protein underwent changes in its structure (modifications and loss of domains) and function, and (2) the expression of the protein came under a different regulatory circuit. Indeed, the effect of these proteins remains the aggregation of particles in the extracellular (cell-to-cell) or intracellular milieus (lipid body-to-lipid body). These changes allow the new localization of the protein in the cell (on the cell surface rather than in the lipid bodies); and its induction in a different condition (during invasion of host cells and infection rather than during lipid storage).

This work provides a conceptual framework which connected and organized results of different studies for the elaboration of a work-hypothesis that orientates research into bacterial virulence evolution. Further studies should be focused on experimental validation of this hypothesis.

Acknowledgements

The studies are financially supported by the SCyT of the University of Patagonia San Juan Bosco (PI 007/13), Project PIP-CONICET Nro. 0764, Project PICT2012 Nro. 2031 and Project PICT2013 Nro. 1564 (ANPCyT), Argentina. Lanfranconi M.P. and Alvarez H.M. are career investigators of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biochi.2016.06.002.

References

[1] F.U. Battistuzzi, A. Feijao, S.B. Hedges, A genomic timescale of prokaryote evolution, insights into the origin of methanogenesis, phototrophy, and the colonization of land, BMC Evol. Biol. 4 (2004) 44.
[2] F.U. Battistuzzi, S.B. Hedges, A major clade of prokaryotes with ancient adaptations to life on land, Mol. Biol. Evol. 26 (2009) 335–343.
[3] H.M. Alvarez, R.A. Silva, A.C. Cesari, A.L. Zamit, S.R. Peressutti, R. Reichelt, U. Keller, U. Malkus, C. Rasch, T. Maskow, F. Mayer, A. Steinbüchel, Physiological and morphological responses of the soil bacterium *Rhodococcus opacus*

strain PD630 to water stress, FEMS Microbiol. Ecol. 50 (2004) 75–86.
[4] S. Bequer Urbano, C. Di Capua, N. Cortez, M.E. Farías, H.M. Alvarez, Triacylglycerol accumulation and oxidative stress in *Rhodococcus* species: differential effects of pro-oxidants on lipid metabolism, Extremophiles 18 (2014) 375–384.
[5] J.C. LeBlanc, E.R. Goncalves, W.W. Mohn, Global response to desiccation stress in the soil actinomycete *Rhodococcus jostii* RHA1, Appl. Environ. Microbiol. 9 (2008) 2627–2636.
[6] D.B. Finkelstein, S.C. Brassell, L.M. Pratt, Microbial biosynthesis of wax esters during desiccation: adaptation for colonization of the earliest terrestrial environments? Geology 38 (2010) 247–250.
[7] H.M. Alvarez, Triacylglycerol and wax ester-accumulating machinery in prokaryotes, Biochimie (2015), http://dx.doi.org/10.1016/j.biochi.2015.08.016.
[8] A. Röttig, A. Steinbüchel, Acyltransferases in bacteria, Microbiol. Mol. Biol. Rev. 77 (2013) 277–321.
[9] D.P. MacEachran, M.E. Prophete, A.J. Sinskey, The *Rhodococcus opacus* PD630 heparin-binding hemagglutinin homolog TadA mediates lipid body formation, Appl. Environ. Microbiol. 76 (2010) 7217–7225.
[10] Y. Ding, L. Yang, S. Zhang, Y. Wang, Y. Du, J. Pu, G. Peng, Y. Chen, H. Zhang, J. Yu, H. Hang, P. Wu, F. Yang, H. Yang, A. Steinbüchel, P. Liu, Identification of the major functional proteins of prokaryotic lipid droplets, J. Lipid Res. 53 (2012) 399–411.
[11] Y. Chen, Y. Ding, L. Yang, J. Yu, G. Liu, X. Wang, S. Zhang, D. Yu, L. Song, H. Zhang, C. Zhang, L. Huo, C. Huo, Y. Wang, Y. Du, H. Zhang, P. Zhang, H. Na, S. Xu, Y. Zhu, Z. Xie, T. He, Y. Zhang, G. Wang, Z. Fan, F. Yang, H. Liu, X. Wang, X. Zhang, M.Q. Zhang, Y. Li, A. Steinbüchel, T. Fujimoto, S. Cichello, J. Yu, P. Liu, Integrated omics study delineates the dynamics of lipid droplets in *Rhodococcus opacus* PD630, Nucleic Acids Res. 42 (2014) 1052–1064.
[12] J.S. Dávila Costa, O.M. Herrero, H.M. Alvarez, L. Leichert, Label-free and redox proteomic analyses of the triacylglycerol-accumulating *Rhodococcus jostii* RHA1, Microbiology 161 (2015) 593–610.
[13] K. Pethe, S. Alonso, F. Biet, G. Delogu, M.J. Brennan, C. Locht, et al., The heparin-binding haemagglutinin of *M. tuberculosis* is required for extrapulmonary dissemination, Nature 412 (2001) 190–194.
[14] M.A. Forrellad, L.I. Klepp, A. Gioffré, J. Sabio y Garcia, H.R. Morbidoni, et al., Virulence factors of the *Mycobacterium tuberculosis* complex, Virulence 4 (2013) 3–66.
[15] L.A. Sechi, N. Ahmed, G.E. Felis, I. Dupre, S. Cannas, et al., Immunogenicity and cytoadherence of recombinant heparin binding haemagglutinin (HBHA) of *Mycobacterium avium* subsp. *paratuberculosis*: functional promiscuity or a role in virulence? Vaccine 24 (2006) 236–243.
[16] F.D. Menozzi, J.H. Rouse, M. Alavi, M. Laude-Sharp, J. Muller, R. Bischoff, M.J. Brennan, C. Locht, Identification of a heparin binding hemagglutinin present in mycobacteria, J. Exp. Med. 184 (1996) 993–1001.
[17] F. Biet, M.M. Angela de Melo, M. Grayon, E.K. Xavier da Silveira, P.J. Brennan, H. Drobecq, D. Raze, M.C. Vidal Pessolani, C. Locht, F.D. Menozzi, *Mycobacterium smegmatis* produces an HBHA homologue which is not involved in epithelial adherence, Microbes Infect. 9 (2007) 175–182.
[18] G. Delogu, G. Fadda, M.J. Brennan, Impact of structural domains of the heparin binding hemagglutinin of *Mycobacterium tuberculosis* on function, Protein Pept. Lett. 19 (2012) 1035–1039.
[19] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.
[20] J. Felsenstein, J. Phylip, Phylogeny Inference Package (Version 3.2) Cladistics, 5, 1989, pp. 164–166.
[21] V. Solovyev, A. Salamov, Automatic annotation of microbial genomes and metagenomic sequences, in: R.W. Li (Ed.), Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies Nova Science Publishers, 2011, pp. 61–78.
[22] A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences, Science 252 (1991) 1162–1164.
[23] M. Delorenzi, T. Speed, An HMM model for coiled-coil domains and a comparison with PSSM-based predictions, Bioinformatics 18 (2002) 617–625.
[24] C. Deb, C.-M. Lee, V.S. Dubey, J. Daniel, B. Abomoelak, T.D. Sirakova, et al., A novel in vitro multiple-stress dormancy model for *Mycobacterium tuberculosis* generates a lipid-loaded, drug-tolerant, dormant pathogen, PLoS One 4 (2009) e6077.
[25] A. Cossu, L.A. Sechi, E. Bandino, S. Zanetti, V. Rosu, Expression profiling of *Mycobacterium tuberculosis* H37Rv and *Mycobacterium smegmatis* in acid-nitrosative multi-stress displays defined regulatory networks, Microb. Pathog. 65 (2013) 89–96.
[26] J.C. Betts, P.T. Lukey, L.C. Robb, R.A. McAdam, K. Duncan, Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling, Mol. Microbiol. 43 (2002) 717–731.
[27] J. Albrethsen, J. Agner, S.R. Piersma, P. Højrup, T.V. Pham, K. Weldingh, C.R. Jimenez, P. Andersen, I. Rosenkrands, Proteomic profiling of *Mycobacterium tuberculosis* identifies nutrient-starvation-responsive toxin-antitoxin systems, Mol. Cell. Proteom. 12 (2013) 1180–1191.
[28] A. Arabolaza, E. Rodriguez, S. Altabe, H.M. Alvarez, H. Gramajo, Multiple pathways for triacylglycerol biosynthesis in *Streptomyces coelicolor*, Appl. Environ. Microbiol. 74 (2008) 2573–2582.
[29] R.C. Hartkoorn, C. Sala, S. Uplekar, P. Busso, J. Rougemont, et al., Genome-wide definition of the SigF regulon in *Mycobacterium tuberculosis*, J. Bacteriol. 194 (2012) 2001–2009.
[30] C. Esposito, M. Cantisani, G. D'Auria, L. Falcigno, E. Pedone, S. Galdiero, R. Berisio, Mapping key interactions in the dimerization process of HBHA from *Mycobacterium tuberculosis*, insights into bacterial agglutination, FEBS Lett. 586 (2012) 659–667.
[31] G. Delogu, M.J. Brennan, Functional domains present in the mycobacterial hemagglutinin, HBHA, J. Bacteriol. 181 (1999) 7464–7469.
[32] K.L. Low, G. Shui, K. Natter, W.K. Yeo, S.D. Kohlwein, T. Dick, S.P. Rao, M.R. Wenk, Lipid droplet-associated proteins are involved in the biosynthesis and hydrolysis of triacylglycerol in *Mycobacterium bovis* bacillus Calmette-Guerin, J. Biol. Chem. 285 (2010) 21662–21670.
[33] E. Bengtsson, M. Morgelin, T. Sasaki, R. Timpl, D. Heinegard, et al., The leucine-rich repeat protein PRELP binds perlecan and collagens and may function as a basement membrane anchor, J. Biol. Chem. 277 (2002) 15061–15068.
[34] K. Pethe, M. Aumercier, E. Fort, C. Gatot, C. Locht, F.D. Menozzi, Characterization of the heparin-binding site of the mycobacterial heparin binding hemagglutinin adhesin, J. Biol. Chem. 275 (2000) 14273–14280.
[35] P. Lebrun, D. Raze, B. Fritzinger, J.M. Wieruszeski, F. Biet, A. Dose, M. Carpentier, D. Schwarzer, F. Allain, G. Lippens, C. Locht, Differential contribution of the repeats to heparin binding of HBHA, a major adhesin of *Mycobacterium tuberculosis*, PLoS One 7 (2012) e32421.
[36] M.W. Mangan, S. Lucchini, V. Danino, T.O. Cróinín, J.C. Hinton, C.J. Dorman, The integration host factor (IHF) integrates stationary-phase and virulence gene expression in Salmonella enterica serovar Typhimurium, Mol. Microbiol. 59 (2006) 1831–1847.
[37] M. Letek, P. González, I. Macarthur, H. Rodríguez, T.C. Freeman, et al., The genome of a pathogenic rhodococcus: cooptive virulence underpinned by key gene acquisitions, PLoS Genet. 6 (2010) e1001145.
[38] H. Wu, Y. Fang, J. Yu, Z. Zhang, The quest for a unified view of bacterial land colonization, ISME J. 8 (2014) 1358–1369.