

Research article

Open Access

# A phylogenomic analysis of the *Actinomycetales* *mce* operons

Nicola Casali and Lee W Riley\*

Address: University of California at Berkeley, School of Public Health, 140 Warren Hall, Berkeley, CA 94720, US

Email: Nicola Casali - ncasali@berkeley.edu; Lee W Riley\* - lrwiley@berkeley.edu

\* Corresponding author

Published: 26 February 2007

Received: 9 October 2006

BMC Genomics 2007, 8:60 doi:10.1186/1471-2164-8-60

Accepted: 26 February 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/60>

© 2007 Casali and Riley; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** The genome of *Mycobacterium tuberculosis* harbors four copies of a cluster of genes termed *mce* operons. Despite extensive research that has demonstrated the importance of these operons on infection outcome, their physiological function remains obscure. Expanding databases of complete microbial genome sequences facilitate a comparative genomic approach that can provide valuable insight into the role of uncharacterized proteins.

**Results:** The *M. tuberculosis* *mce* loci each include two *yrbE* and six *mce* genes, which have homology to ABC transporter permeases and substrate-binding proteins, respectively. Operons with an identical structure were identified in all *Mycobacterium* species examined, as well as in five other *Actinomycetales* genera. Some of the *Actinomycetales* *mce* operons include an *mkl* gene, which encodes an ATPase resembling those of ABC uptake transporters. The phylogenetic profile of Mkl orthologs exactly matched that of the Mce and YrbE proteins. Through topology and motif analyses of YrbE homologs, we identified a region within the penultimate cytoplasmic loop that may serve as the site of interaction with the putative cognate Mkl ATPase. Homologs of the exported proteins encoded adjacent to the *M. tuberculosis* *mce* operons were detected in a conserved chromosomal location downstream of the majority of *Actinomycetales* operons. Operons containing linked *mkl*, *yrbE* and *mce* genes, resembling the classic organization of an ABC importer, were found to be common in Gram-negative bacteria and appear to be associated with changes in properties of the cell surface.

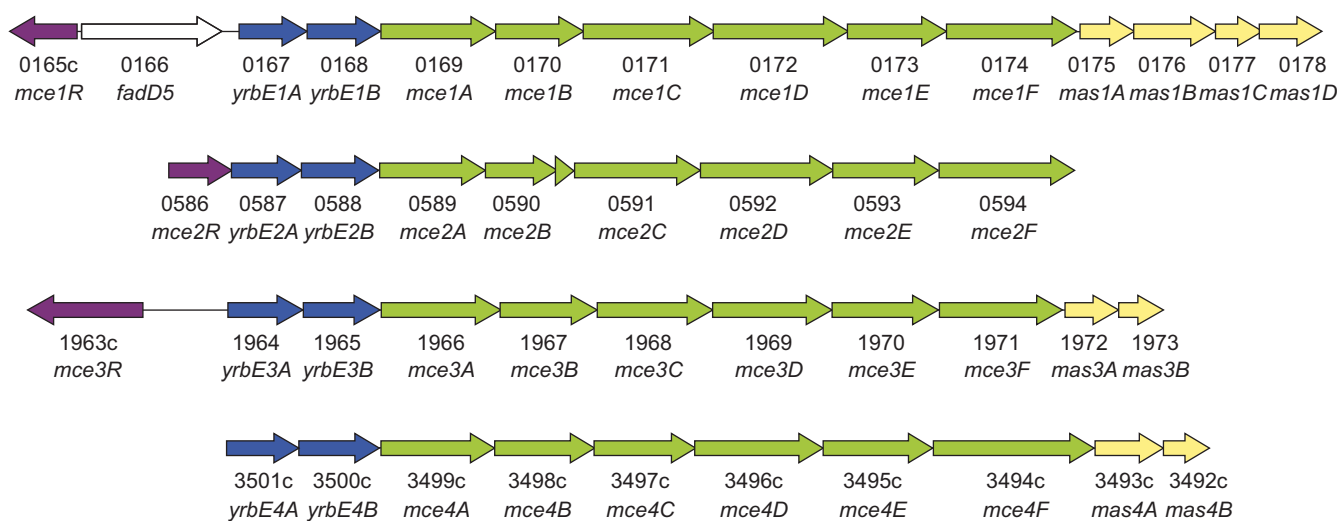
**Conclusion:** Evidence presented suggests that the *mce* operons of *Actinomycetales* species and related operons in Gram-negative bacteria encode a subfamily of ABC uptake transporters with a possible role in remodeling the cell envelope.

## Background

A putative *Mycobacterium tuberculosis* virulence gene, named *mce1A*, was originally identified because its expression in *Escherichia coli* enabled this noninvasive bacterium to enter mammalian epithelial cells [1]. Sequencing of the *M. tuberculosis* genome revealed that *mce1A* (Rv0169) was part of an operon that encoded eight putative membrane-associated proteins: YrbEA-B, MceA-F [2,3]. This operon is

present four times in the *M. tuberculosis* genome (*mce1-4*). Homologs of the genes adjacent to the *mce1* locus, Rv0175-Rv0178, are located downstream of the *mce3* and *mce4* gene clusters (Figure 1) [3].

Continued interest in the function of the *M. tuberculosis* *mce* operons stems from reports of the profound effect of disruption of *mce* operons on growth and virulence of the



**Figure 1**  
**Schematic representation of the *M. tuberculosis* H37Rv *mce* loci.** Proximal transcription regulators are colored in purple, *yrbE* genes in blue, *mce* genes in green, and genes encoding 'conserved *mce*-associated proteins' in yellow [44].

mutant strains in mice. Shimono *et al.* [4] showed that an *mce1* mutant was hypervirulent when inoculated intravenously into BALB/c mice. In the first few weeks of infection, the mutant strain multiplied more rapidly than wild-type in the mice's lungs, spleen and liver. Surprisingly, Goffre *et al.* [5] found that a *yrbE1B* mutant grew faster than wild-type in the lungs and spleens of BALB/c mice inoculated via the peritoneum, but more slowly in mice infected through the tracheal route. Sasseti and Rubin [6] reported that in competitive mixed infections *mce1* mutants exhibited a growth defect in the spleens of intravenously-infected C57BL/6J mice after one week of infection. Although the exact cause of these apparently disparate phenotypes remains to be established, the observations suggest that the fate of *mce1* mutants *in vivo* is determined by the prevailing immunological environment experienced during the first few weeks of infection.

Both *mce2* and *mce3* mutants replicated slower than wild-type in BALB/c mice infected via either the trachea or peritoneum [5]; however, neither mutant demonstrated a significant growth defect in competitive mixed infections [6]. In co-infected C57BL/6J mice, an *mce4* mutant was attenuated relative to wild-type after two to four weeks infection, whilst an *mce1-mce4* double mutant exhibited further attenuation, indicating that the *mce* operons perform non-redundant roles during infection [7].

The similarity of the YrbE and Mce proteins with  $\Delta$ ATP-binding cassette (ABC) transporter permeases and substrate-binding proteins, respectively, has been noted pre-

viously [8,9]. ABC transporters couple the energy released by ATP hydrolysis to the translocation of a substrate across a membrane. Members of the ABC transporter family are ubiquitous in living organisms and comprise one of largest superfamilies known [10].

A functional ABC transporter system minimally contains two cytoplasmic nucleotide-binding ATPase domains and two transmembrane channel-forming permease domains. These components can be homo- or heterodimers and may be encoded on separate or fused polypeptides. Both eukaryotes and prokaryotes contain ABC exporters, whereas importers have been identified only in prokaryotes. Importers additionally require substrate-binding proteins (SBPs) that provide specificity and high-affinity. Typically, SBPs are periplasmic in Gram-negative bacilli and lipoproteins in Gram-positive bacilli [11]. SBPs share a two-lobed quaternary structure with a central cleft that undergoes a large conformational change upon ligand-binding, promoting close interaction with the cognate permease. This results in hydrolysis of ATP, which energizes translocation of the substrate [12]. In Gram-negative bacteria, SBP-dependent importers also usually require porins or specific receptors to facilitate transport across the outer membrane [11].

The genes encoding the ATPase, permease and SBP components of an ABC transporter are often contiguous in the genome and comprise an operon. Phylogenetic clustering of the individual transporter components is almost always concordant, indicating that the operons have arisen from

a common ancestral transporter with minimal shuffling of constituents. In addition, sequence similarity shows good correlation with substrate specificity [13-15].

The ATPase is the most conserved component of the system and transporter function is frequently predicted solely on the basis of ATPase orthology [10,15]. These proteins contain a homologous region, of 200 amino acids, with several characteristic motifs: Walker A and B motifs in the nucleotide-binding fold [16], as well as a signature motif found only in ABC transporter-associated, or 'traffic', ATPases [17].

The permease components and SBPs have limited primary sequence similarity, and thus their identification is not facile. They are typically identified in genome sequences by their proximity to ATPases and, for permeases, possession of predicted transmembrane regions [18-20]. The inference of function through sequence comparison has traditionally relied upon similarity to close homologs of known function. The advent of the genomic age has provided invaluable new methods for the elucidation of roles of proteins with unknown function. Non-homology-based methods of genome comparison use patterns of domain fusion [21], conserved chromosomal location [22], and phylogenetic profiles [23], to predict functional interactions between proteins. In addition, the availability of hundreds of complete genome sequences permits the reliable identification of orthologs, operationally-defined as reciprocal best hits [24], enabling more precise functional prediction than sequence similarity alone. These methods are non-redundant and their application can facilitate deduction of specific function [25]. Here we endeavor to further understand the function of the *M. tuberculosis* *mce* operons, and assess the likelihood that they encode ABC transporters, through sequence and genome comparisons, database mining and the application bioinformatic methods.

## Results

### Distribution of *mce* operons in Actinomycetales

Perusal of databases of conserved domains, such as InterPro [26], Pfam [27] and TIGRFAM [28], constitutes a simple method for the identification of homologous proteins. The *M. tuberculosis* H37Rv genome encodes 24 Mce proteins, each of which contains a conserved domain of 304 amino acids defined by the TIGRFAM family: TIGR00996 (IPR005693). Members of this family are confined to the Order Actinomycetales. The corresponding Pfam family, PF02470 (IPR003399), describes a 98 amino acid sub-region of the Mce domain that is more widely distributed (see below). The *mce* genes in *M. tuberculosis* are clustered in groups of six; each cluster is preceded by two copies of a gene termed *yrbE* (Figure 1). Databases of conserved domains group the YrbE proteins

into a family called DUF140 (domain of unknown function). Pfam defines the family by a region approximately 150 amino acids long (PF02405; IPR003453). The corresponding TIGRFAM family (TIGR00056) describes a sub-family of DUF140, but excludes the mycobacterial homologs based on a stated extreme divergence at the amino end. For the sake of clarity, we refer to a cluster of genes encoding two YrbE and six Mce proteins as an '*mce* operon'.

To assess the distribution of *mce* operons in completed and draft assemblies of genomes of members of the Order Actinomycetales, we surveyed the annotation of predicted proteins for members of Pfam families PF02470 and PF02405 (Table 1). The proteomes of all 10 *Mycobacterium* species examined contained Mce proteins. The number varied from 6 in *Mycobacterium leprae* up to 66 in *Mycobacterium vanbaalenii*. Other genomes containing *mce* genes belonged to species of *Nocardia*, *Janibacter*, *Nocardiododes*, *Amycolatopsis* and *Streptomyces*. Mce homologs were absent from 18 Actinomycetales genomes, notably including those of the four sequenced *Corynebacterium* species. DUF140 proteins were found encoded within all Actinomycetales genomes that contain *mce* genes and were absent from all genomes that do not contain *mce* genes. Other completely sequenced genomes of species belonging to the Class Actinobacteria, namely *Rubrobacter xylanophilus*, *Symbiobacterium thermophilum* and *Bifidobacterium longum*, did not contain either Mce or DUF140 homologs.

Examination of the genomic location of the Mce and DUF140 homologs revealed that the *mce* genes were almost always found clustered in groups of six, located downstream from a pair of DUF140 genes (Figure 2).

### Identification of *mce*-like operons in Gram-negative bacteria

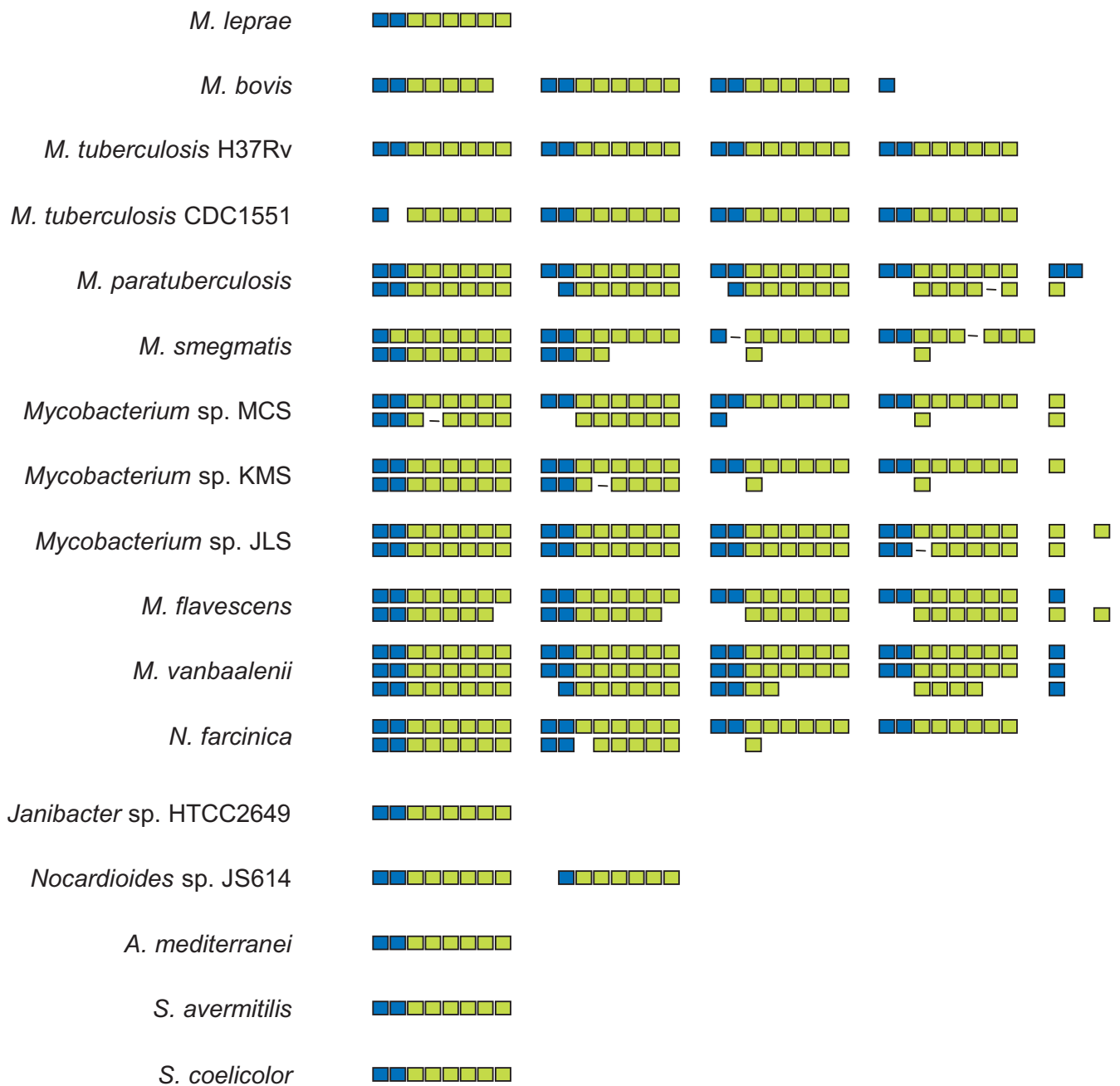
A 98 amino acid sub-region of Mce family proteins, termed the 'Mce-like' domain (PF02470), is widely distributed in Gram-negative bacteria and has also been found encoded in plant genomes. No Mce-like domains have been identified in any Archaeal or low GC-content Gram-positive bacterial genomes.

Genes with related functions are frequently encoded within operons and thus found clustered in the genomes of prokaryotes [22]. We investigated the gene neighborhoods of selected *mce*-like genes with the aim of obtaining clues regarding the biological role of proteins of this family (Figure 3). The Mce-like proteins in Gram-negative bacteria were frequently found clustered in the genome with a DUF140 family protein and an ATPase homolog (IPR003439) in an arrangement typical of an ABC transporter system [11]. The three components were found encoded in any order and in some instances either the

**Table 1: Distribution of Mce and YrbE proteins within the Order Actinomycetales<sup>a</sup>**

Suborder	Family	Species	Mce <sup>b</sup>	DUF140 <sup>c</sup>	Source
Actinomycinaeae	Actinomycetaceae	<i>Actinomyces naeslundii</i> MG1	0	0	UniProt
Corynebacterineae	Corynebacteriaceae	<i>Corynebacterium diphtheriae</i> NCTC 13129	0	0	UniProt
		<i>Corynebacterium efficiens</i> YS-314	0	0	UniProt
		<i>Corynebacterium glutamicum</i> ATCC 13032	0	0	UniProt
		<i>Corynebacterium jeikeium</i> K411	0	0	UniProt
		<i>Corynebacterium urealyticum</i> ATCC 49619	0	0	UniProt
	Mycobacteriaceae	<i>Mycobacterium leprae</i> TN	6	2	UniProt
		<i>Mycobacterium bovis</i> AF2122/97	18	7	UniProt
		<i>Mycobacterium tuberculosis</i> CDC1551	24	7	TIGR
		<i>Mycobacterium tuberculosis</i> H37Rv	24	8	TIGR
		<i>Mycobacterium paratuberculosis</i> K-10	48	14	UniProt
		<i>Mycobacterium smegmatis</i> MC2 155	34	11	TIGR
		<i>Mycobacterium</i> sp. MCS	38	11	JGI
		<i>Mycobacterium</i> sp. KMS	38	12	JGI
		<i>Mycobacterium</i> sp. JLS	50	16	JGI
		<i>Mycobacterium flavescens</i> PYR-GCK	48	13	UniProt
<i>Mycobacterium vanbaalenii</i> PYR-1	66	24	UniProt		
Nocardiaceae	<i>Nocardia farcinica</i> IFM 10152	36	12	UniProt	
Frankineae	Acidothermaceae	<i>Acidothermus cellulolyticus</i> 11B	0	0	UniProt
	Frankiaceae	<i>Frankia</i> sp. Ccl3	0	0	UniProt
		<i>Frankia</i> sp. EAN1pec	0	0	UniProt
	Kineosporiaceae	<i>Kineococcus radiotolerans</i> SRS30216	0	0	UniProt
Micrococcineae	Brevibacteriaceae	<i>Brevibacterium linens</i> BL2	0	0	JGI
	Cellulomonadaceae	<i>Tropheryma whipplei</i> str. Twist	0	0	UniProt
		<i>Tropheryma whipplei</i> TW08/27	0	0	UniProt
	Intrasporangiaceae	<i>Janibacter</i> sp. HTCC2649	6	2	NCBI
	Microbacteriaceae	<i>Leifsonia xyli</i> subsp. xyli str. CTCB07	0	0	UniProt
	Micrococcaceae	<i>Arthrobacter aurescens</i> TC1	0	0	UniProt
		<i>Arthrobacter</i> sp. FB24	0	0	UniProt
Propionibacterineae	Nocardioidaceae	<i>Nocardioides</i> sp. JS614	12	3	UniProt
	Propionibacteriaceae	<i>Propionibacterium acnes</i> KPA171202	0	0	UniProt
Pseudonocardineae	Pseudonocardiaceae	<i>Amycolatopsis mediterranei</i> <sup>d</sup>	6	2	Pfam
Streptomycineae	Streptomycetaceae	<i>Streptomyces avermitilis</i> MA-4680	6	2	UniProt
		<i>Streptomyces coelicolor</i> A3(2)	6	2	UniProt
Streptosporangineae	Nocardiopsaceae	<i>Thermobifida fusca</i> YX	0	0	UniProt

<sup>a</sup> Taxonomy from Bergey's Manual of Systematic Bacteriology [107]<sup>b</sup> Number of proteins classified as PF02470<sup>c</sup> Number of proteins classified as PF02405<sup>d</sup> Incomplete genome, EMBL Accession [AF040570](#)



**Figure 2**  
**Schematic representation of the organization of *mce* loci in *Actinomycetales* genomes.** Genes encoding proteins belonging to Pfam family PF02470 (Mce) are depicted as green boxes, and to family PF02405 (DUF140) as blue boxes. Dashes indicate gaps in gene numbering.

DUF140 or ATPase homolog was duplicated. In a number of  $\gamma$ -Proteobacteria the ATPase-DUF140-Mce cluster was encoded in a conserved genomic region that included a Tol protein (IPR008869), a STAS domain protein (IPR002645) and MurA (IPR005750), the product of which catalyses the first step of murein biosynthesis. Like Mce domains, Tol proteins have homology to SBPs [29]; the presence of SBPs indicates that these operons encode substrate uptake transporters. Aravind and Koonin suggested that the nucleotide-binding activity of STAS domains, found in sulfate transporters, could regulate uptake in response to intracellular ATP or GTP concentrations [30]. Several DUF140 proteins that are N-terminally fused to STAS domains have been identified [31], implying a functional linkage between these two proteins in the *mce* operons [21]. The Mce transporter clusters were also frequently found associated with homologs of a surface-exposed lipoprotein VacJ (IPR007428), and the morpho-protein BOLA (IPR002634).

The Mce homologs in these putative transporter operons each contain a single 98 amino acid Mce-like domain. Many proteobacterial genomes additionally contain Mce homologs, sometimes annotated as PqiB, that contain 2–7 copies of the Mce-like domain and are usually associated with a PqiA family protein (IPR007498) of unknown function. The *E. coli pqiAB* operon is induced by treatment with the model superoxide generator, paraquat [32].

#### **Mce-associated ATPases**

Since ABC transporters absolutely require an ATPase to provide the energy required for substrate translocation, the genes neighboring the *Actinomycetales mce* operons were inspected for ATPase homologs (IPR003439). Although none of the mycobacterial *mce* operons neighbors an ATPase, a candidate gene was identified immediately upstream of a single *mce* operon in the genome of every non-mycobacterial *Actinomycetales* species that possesses *mce* genes (Table 2). BLASTP analyses demonstrated that the corresponding protein sequences were reciprocal best hits with the *mce*-linked ATPases in Gram-negative bacteria, indicating orthology [24]. A phylogenetic analysis of ABC transporter ATPases reported by Dassa and Bouige groups these *Actinomycetales* and Gram-negative bacterial ATPases into a family termed Mkl [8].

The sequences of the *N. farcinica* and *Streptomyces mce*-linked ATPases (nfa51100, SAV5902 and SCO2422) were used as BLASTP queries in order to identify additional Mkl-like ATPases. The best hits from each of the completed *Actinomycetales* genomes (Table 1) were retrieved for further evaluation. Phylogenetic analysis of the protein sequences revealed that each *Mycobacterium* species contained a single ATPase that clustered with the Mkl family, providing strong evidence of orthology (Figure 4,

Table 2). In addition, a paralog was identified in the *N. farcinica* genome (nfa20200); this ORF is annotated in The Institute of Genome Research (TIGR) database as MetN, a D-methionine ABC transporter ATPase, but it does not cluster with other putative MetN orthologs (Figure 4).

Comparison of the most closely related ORFs in other *Actinomycetales* revealed that only those genomes that contained *mce* operons possessed an orthologous ATPase (Figure 4). Congruency of the phylogenetic profiles of the Mkl ATPases with YrbE and Mce proteins provides further evidence of functional association [23].

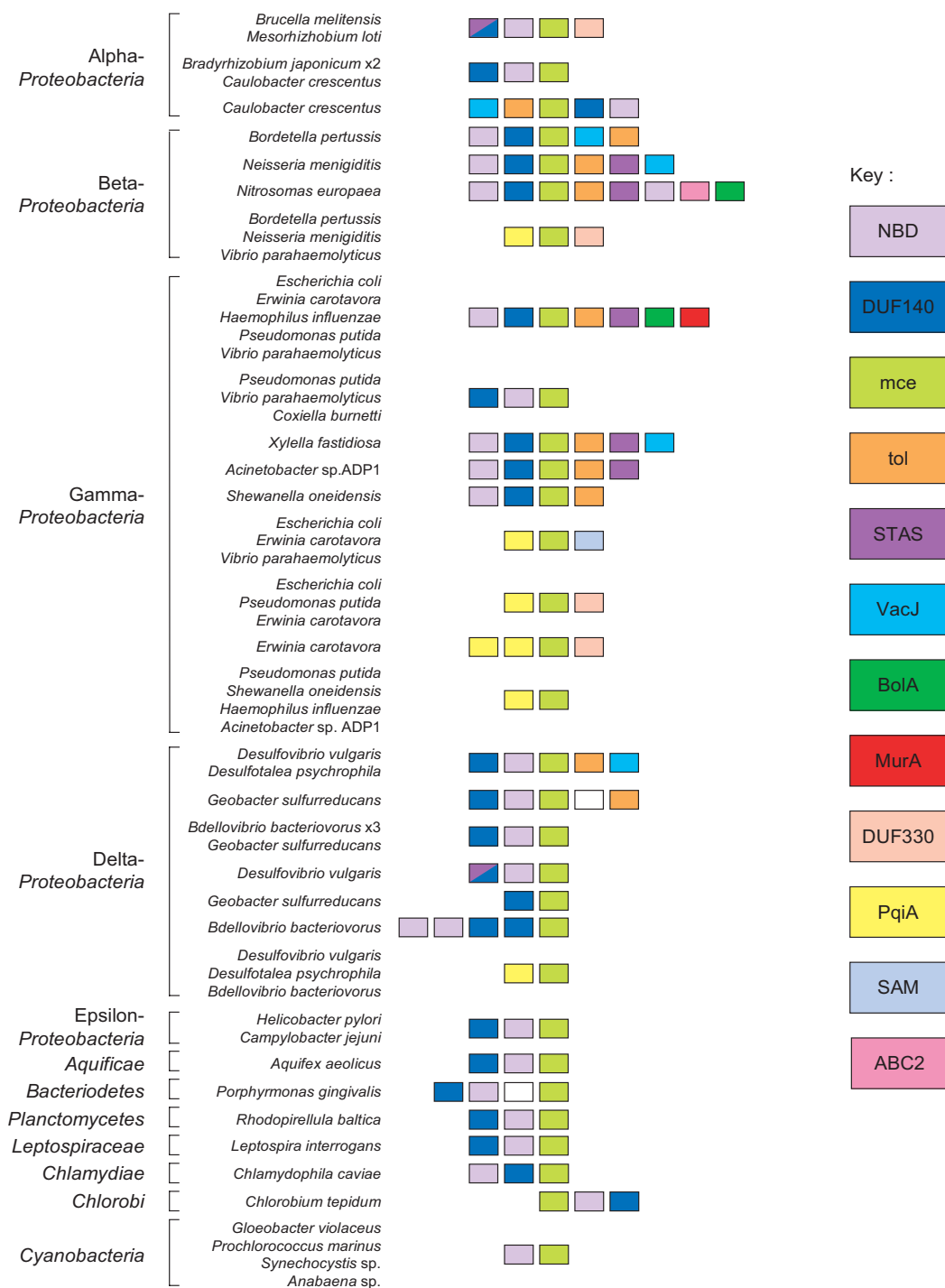
Each of the *mce*-linked ATPases and mycobacterial orthologs contain the conserved Walker A and B motifs required for ATP binding, as well as the ABC transporter family signature (LSGGQ) with no more than one mismatch [16,33]. In a published analysis of *M. tuberculosis* ABC transporters, the putative Mce ATPase, Rv0655, segregated with importers but did not fall into any of the previously described families with known substrates [20]. Similarly, in a more expansive study, the Mkl family ATPases fell into the SBP-dependent importer clade, but clustered separately from those with established specificity [8].

The mycobacterial Mkl ATPases and nfa20200 and are not genomically located near any other ABC transporter components and appear to be transcriptionally-isolated. The *M. leprae* ortholog is located adjacent to RNA polymerase *rpo* genes leading to speculation that this ATPase was involved in ribonucleotide uptake [34]. Consequently, Mkl ATPases are sometimes annotated as ribonucleotide uptake systems.

#### **The Mce proteins**

Comparison of the amino acid sequences of the Mce proteins encoded in the genomes of *Mycobacterium bovis* and the *M. tuberculosis* strains H37Rv, CDC1551 and 210, revealed that each of the *M. tuberculosis* genomes contained 24 Mce ORFs, whilst, as noted previously, the *mce3* operon is deleted in *M. bovis* [35]. A number of genes were found to contain frameshift mutations: *mce1F* in strain 210; *mce2B* in strains H37Rv and CDC1551; *mce2C* in strain CDC1551; and *mce2D* and *mce2E* in *M. bovis*. The truncated ORFs thus conspicuously clustered within the *mce2* operon.

A non-redundant set of Mce proteins from the genomes of *M. tuberculosis*, *M. bovis*, *M. leprae*, *Mycobacterium avium* subsp. *paratuberculosis* (*M. paratuberculosis*), *Mycobacterium smegmatis*, *N. farcinica*, *S. coelicolor* and *S. avermilitis* were selected for further analysis. Examination of the genomic regions of partial operons revealed the presence of several



**Figure 3**  
**Conserved proteins encoded in the neighborhood of mce genes in Gram-negative bacteria.** Coloring reflects conserved domains identified in the key. Protein families shown are: NBD, an ABC transporter ATPase (IPR003439); DUF140 (IPR003453); Mce (IPR003399); Tol, a Ttg2 toluene tolerance protein (IPR008869); STAS, a domain found in sulfate transporters and anti-sigma factor antagonists (IPR002645); VacJ, a lipoprotein of unknown function (IPR007428); BolA, a possible regulator induced by stress (IPR002634); MurA, UDP-N-acetylglucosamine-1-carboxyvinyltransferase (IPR005750); DUF330 (IPR005586); PqiA, an integral membrane protein inducible by superoxide generators (IPR007498); SAM, an S-adenosyl methionine binding methyltransferase (IPR000051); and ABC2, an ABC-2 type permease (IPR013525).

**Table 2: Actinomycetales mce-linked ATPases and mycobacterial orthologs**

Organism	ATPase
<i>Amycolatopsis mediterranei</i>	TrEMBL: Q7BUF5
<i>Janibacter</i> sp. HTCC2649	JNB_08429
<i>Nocardia farcinica</i>	nfa51100
<i>Nocardiooides</i> sp. JS614	NocaDRAFT_4321
<i>Streptomyces avermitilis</i>	SAV5902
<i>Streptomyces coelicolor</i>	SCO2422
<i>Mycobacterium bovis</i>	Mb0674
<i>Mycobacterium flavescens</i>	MflvDRAFT_3283
<i>Mycobacterium leprae</i>	MLI892
<i>Mycobacterium paratuberculosis</i>	MAP4129
<i>Mycobacterium smegmatis</i>	MSMEG1359
<i>Mycobacterium</i> sp. JLS	MjlsDRAFT_1757
<i>Mycobacterium</i> sp. KMS	MkmsDRAFT_1059
<i>Mycobacterium</i> sp. MCS	MmcsDRAFT_0968
<i>Mycobacterium tuberculosis</i> CDC1551	MT0684
<i>Mycobacterium tuberculosis</i> H37Rv	Rv0655
<i>Mycobacterium vanbaalenii</i>	MvanDRAFT_5200

additional putative Mce homologs that were included in this analysis (Table 3).

Multiple alignment and phylogenetic analysis of the Mce homologs revealed six distinct branches, which corresponded exactly to the encoding genes in the respective operons (that is *mceA-F*; Figure 5). Within each of the six major branches, the clustering of sequences was essentially the same. This pattern indicates that each *mce* gene cluster duplicated from an ancestral operon that contained six *mce* genes and that no shuffling between or within operons has occurred.

We have classified the operons as *mce1-8* according to the clustering observed (Table 3). The *mce1* and *mce2* operons are the most closely related and duplication may have occurred after divergence of the fast- and slow-growing mycobacteria, since *M. smegmatis* contains a single copy. Although the orthology of the *M. smegmatis* operon cannot be deduced from the phylogenetic tree, we infer from synteny that it is orthologous to the *M. tuberculosis mce1* operon. Thus, *mce1* is the sole operon that is found in all, and in only, the *Mycobacterium* species examined. The *Streptomyces* operons fall into a cluster, termed *mce6*, that does not contain any mycobacterial orthologs, but is found in *N. farcinica*. The Mkl-like ATPase is located upstream of *yrbEA6* in all three of these operons. In several cases operon orthology could not be deduced from the branching pattern observed, presumably due to recent duplication events. Thus, it appears that *M. paratuberculosis* and *M. smegmatis* possess two copies of the *mce5* operon; *M. paratuberculosis* and *N. farcinica* have two copies of the *mce7* operon; and *N. farcinica* has two copies of the *mce8* operon. The *M. paratuberculosis* Mce5E protein

(MAP2193) seems to have diverged significantly from its paralog (MAP0764); examination of the encoding sequences revealed that this is a consequence of a 40bp deletion, which results in a frameshift of the N-terminal 120 amino acids.

One and two extra copies of Mce1A were found in *M. paratuberculosis* (MAP3289) and *M. smegmatis* (MSMEG5783, MSMEG6500), respectively; whilst *N. farcinica* contained a second copy of Mce4A (nfa25900). Each of the encoding genes appeared to be transcriptionally isolated, with the exception of MSMEG5783, which is located within a four-gene operon that includes pyridoxamine 5-phosphate oxidase and a putative lipoprotein.

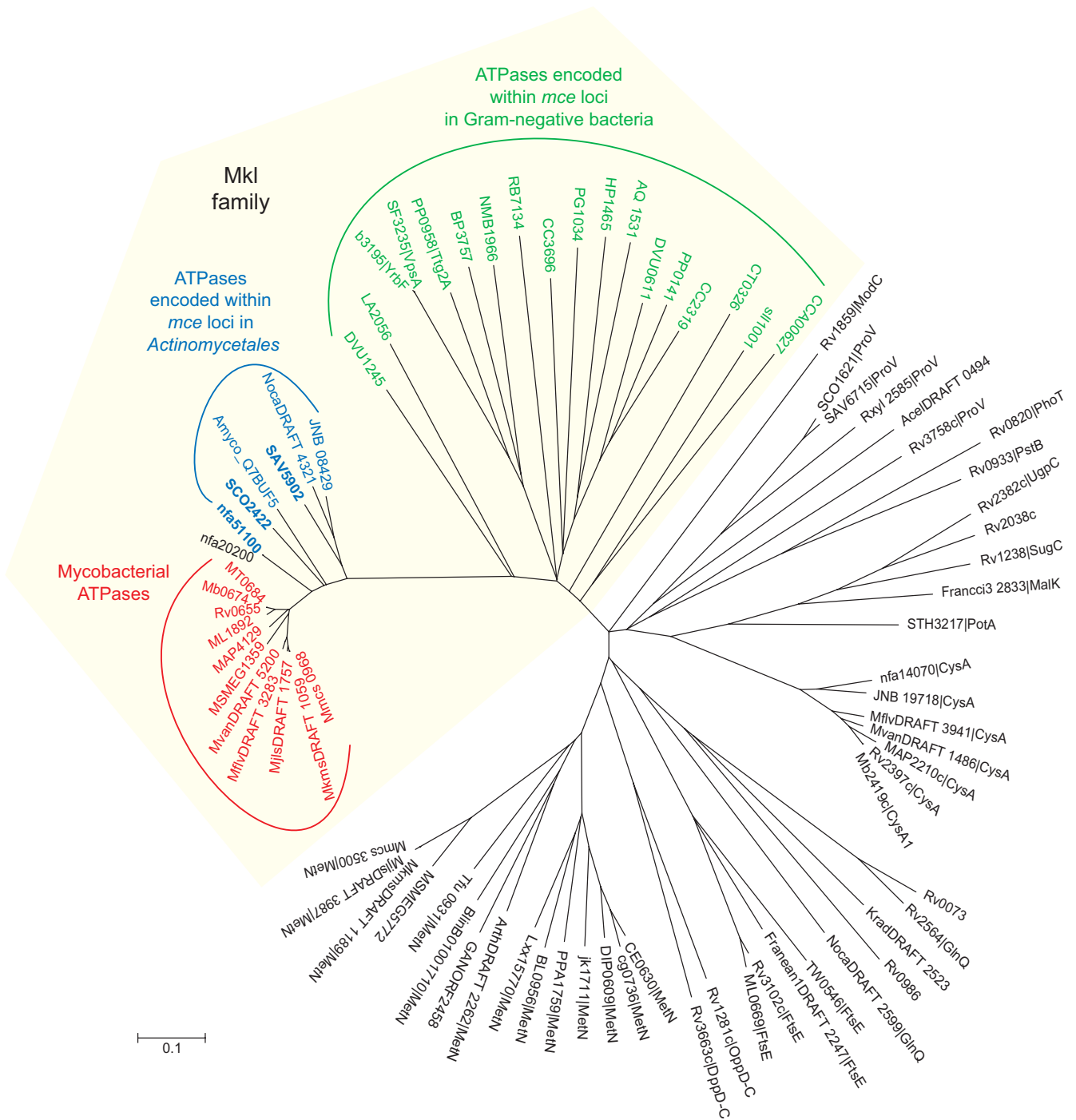
Secondary structure predictions, through the JPred server, revealed the consensus structure of the conserved Pfam region folded into five  $\beta$ -strands; the central region of *Actinomycetales* Mce proteins, included in the conserved TIGRFAM region, contains eight  $\alpha$ -helices. The C-terminal region varies in length from 10–250 amino acids, has predicted low complexity and is rich in proline residues (Figure 6). Length is not conserved within the six homologous families, with the exception of the MceB proteins in which the C-terminal region is 30–50 amino acids in all cases. On average the MceA and MceF proteins are the longest. An RGD motif was identified in the C-terminal tail of 16 (of 27) MceE sequences. This motif is known to bind integrins, as well as C2 domains [36,37].

Each of the Mce proteins contained a hydrophobic stretch at the N-terminus, likely to be a transmembrane helix. Using a neural network trained on Gram-positive bacteria the program SignalP predicted a signal peptide cleavage site for 98 of 161 of these proteins [38]. There was no correlation between prediction of secretion and Mce-type (A-F) or bacterial species. Although the Mce anchor regions frequently contained a pair of arginine residues, characteristic of Twin-arginine transporter (Tat) motifs, few (12 of 161) are recognized as Tat substrates [39]. A lipoprotein attachment site (PS00013) was present in 22 of 27 MceE proteins. The highly conserved operon structure containing six *mce* genes suggests that they associate to form a heteromeric complex [22,40], which is therefore likely to remain tethered to the cell membrane even if some proteins are cleaved. Indeed, Mce1A-1F have been shown to localize to the cell envelope of *M. tuberculosis* [4].

### The YrbE proteins

Unlike the Mce proteins, the amino acid sequences of YrbE orthologs in the *M. tuberculosis* strains H37Rv, CDC1551 and 210, as well as *M. bovis*, were found to be >99.5% identical in all cases. The sequences of the YrbE proteins associated with the *mce* gene clusters of *M. tuberculosis*, *M. leprae*, *M. paratuberculosis*, *M. smegmatis*, *N. far-*





**Figure 4**  
**Phylogenetic tree showing relationship between *mce*-linked ATPases and mycobacterial orthologs.** ATPases encoded within *mce* operons in *Actinomycetales* species are colored blue; those in Gram-negative bacterial *mce* operons are colored green. The sequences most similar to *nfa51100*, *SAV5902* and *SCO2422* (indicated in bold), in the *Actinomycetales* genomes listed in Table I, were identified by BLASTP searches and included in the tree. All of the best hits from mycobacterial species cluster within the Mkl family and are colored red. For comparison, sequences of all *M. tuberculosis* H37Rv ATPases of ABC uptake transporters were included [20]. All of the top hits from *Actinomycetales* that do not possess *mce* operons are rooted among these non-*mce*-linked ATPases, as are all of the second hits from mycobacterial species. ORFs are designated by (UniProt gene name | protein name).

**Table 3: Classification of Actinomycetales *yrbE* and *mce* genes <sup>a</sup>**

Prefix <sup>b</sup>	<i>yrbE1A</i>	<i>yrbE1B</i>	<i>mce1A</i>	<i>mce1B</i>	<i>mce1C</i>	<i>mce1D</i>	<i>mce1E</i>	<i>mce1F</i>
<b>Rv</b>	0167	0168	0169	0170	0171	0172	0173	0174
<b>MT</b>	0176	0177 <sup>c</sup>	0178	0179	0180	0181	0182	0183
<b>Mb</b>	0173	0174	0175	0176	0177	0178	0179	0180
<b>ML</b>	2587	2588	2589	2590	2591	2592	2593	2594
<b>MAP</b>	3602	3603	3604	3605	3606	3607	3608	3609
<b>MSMEG</b>	0126	0127	0128	0129	0130	0131	0132	0133
	<b><i>yrbE2A</i></b>	<b><i>yrbE2B</i></b>	<b><i>mce2A</i></b>	<b><i>mce2B</i></b>	<b><i>mce2C</i></b>	<b><i>mce2D</i></b>	<b><i>mce2E</i></b>	<b><i>mce2F</i></b>
<b>Rv</b>	0587	0588	0589	0590	0591	0592	0593	0594
<b>MT</b>	0616	0617	0618	0619	0621	0622	0623	0624
<b>Mb</b>	0602	0603	0604	0605	0606	0607	0609	0610
<b>MAP</b>	4082	4083	4084	4085	4086	4087	4088	4089
	<b><i>yrbE3A</i></b>	<b><i>yrbE3B</i></b>	<b><i>mce3A</i></b>	<b><i>mce3B</i></b>	<b><i>mce3C</i></b>	<b><i>mce3D</i></b>	<b><i>mce3E</i></b>	<b><i>mce3F</i></b>
<b>Rv</b>	1964	1965	1966	1967	1968	1969	1970	1971
<b>MT</b>	2016	2017	2018	2019	2020	2021	2022	2023
<b>Mb</b>	1999							
<b>MAP</b>	2117c	2117c.1 <sup>d</sup>	2116c	2115c	2114c	2113c	2112c	2111c
<b>MSMEG</b>	0335	0336 <sup>e</sup>	0337	0338	0339	0340	0341	0342
	<b><i>yrbE4A</i></b>	<b><i>yrbE4B</i></b>	<b><i>mce4A</i></b>	<b><i>mce4B</i></b>	<b><i>mce4C</i></b>	<b><i>mce4D</i></b>	<b><i>mce4E</i></b>	<b><i>mce4F</i></b>
<b>Rv</b>	3451c	3450c	3499c	3498c	3497c	3496c	3495c	3494c
<b>MT</b>	3605	3604	3603	3602	3601	3600	3599	3598
<b>Mb</b>	3531c	3530c	3529c	3528c	3527c	3526c	3525c	3524c
<b>MAP</b>	0562	0563	0564	0565	0566	0567	0568	0569
<b>MSMEG</b>	5861	5860	5859.3 <sup>e</sup>	5859.2 <sup>e</sup>	5859.1 <sup>e</sup>	5859	5858	5857.1 <sup>e</sup>
<b>nfa</b>	5350	5360	5370	5380	5390	5400	5410	5420
	<b><i>yrbE5A</i></b>	<b><i>yrbE5B</i></b>	<b><i>mce5A</i></b>	<b><i>mce5B</i></b>	<b><i>mce5C</i></b>	<b><i>mce5D</i></b>	<b><i>mce5E</i></b>	<b><i>mce5F</i></b>
<b>MAP<sup>f</sup></b>	0757	0758	0759	0760	0761	0762/3 <sup>g</sup>	0764	0765
<b>MAP</b>			2189	2190	2191	2192	2193 <sup>h</sup>	2194
<b>MSMEG</b>	2855	2856	2857	2858	2859	2860	2861	2862
<b>MSMEG<sup>f</sup></b>	4785	4784	4783	4782	me <sup>i</sup>	4777	4776	4775
	<b><i>yrbE6A</i></b>	<b><i>yrbE6B</i></b>	<b><i>mce6A</i></b>	<b><i>mce6B</i></b>	<b><i>mce6C</i></b>	<b><i>mce6D</i></b>	<b><i>mce6E</i></b>	<b><i>mce6F</i></b>
<b>nfa</b>	51090	51080	51070	51060	51050	51040	51030	51020
<b>SCO</b>	5901	5900	5899	5898	5897	5896	5895	5894
<b>SAV</b>	2421	2420	2419	2418	2417	2416	2415	2514
	<b><i>yrbE7A</i></b>	<b><i>yrbE7B</i></b>	<b><i>mce7A</i></b>	<b><i>mce7B</i></b>	<b><i>mce7C</i></b>	<b><i>mce7D</i></b>	<b><i>mce7E</i></b>	<b><i>mce7F</i></b>
<b>MAP<sup>j</sup></b>	me <sup>i</sup>	0107	0108	0109	0110	0111	0112	0113
<b>MAP</b>	1849	1850	1851	1852	1853	1854	1855	1856
<b>MSMEG<sup>j</sup></b>	1131	1132	1133	1134	1135	1136	1137	1138
<b>nfa</b>	50540	50530	50520	50510	50500	50490	50480	50470
<b>nfa</b>	56330	56320	56310	56300	56290	56280	56270	56260
	<b><i>yrbE8A</i></b>	<b><i>yrbE8B</i></b>	<b><i>mce8A</i></b>	<b><i>mce8B</i></b>	<b><i>mce8C</i></b>	<b><i>mce8D</i></b>	<b><i>mce8E</i></b>	<b><i>mce8F</i></b>
<b>nfa</b>	11130	11140	11150	11160	11170	11180	11190	11200
<b>nfa</b>	29780	29770	29760 <sup>h</sup>	29750	29740	29730	29720	29710

<sup>a</sup> Operons *mce1-4* designated as in TubercuList; *mce5-8* designated herein. Gene names in organisms other than *M. tuberculosis* do not correspond to those given in genome annotation.

<sup>b</sup> Organism specific gene number prefix: Rv, *M. tuberculosis* H37Rv; MT, *M. tuberculosis* CDC1551; Mb, *M. bovis*; ML, *M. leprae*; MAP, *M. paratuberculosis*; MSMEG, *M. smegmatis*; nfa, *N. farcinica*; SCO, *S. coelicolor*; SAV, *S. avermitilis*.

<sup>c</sup> Orthologous sequence present, but ORF annotated in reverse direction.

<sup>d</sup> Orthologous sequence present, but not annotated. ORF extends ~400 bp at 5' end.

<sup>e</sup> Orthologous sequence present, but not annotated.

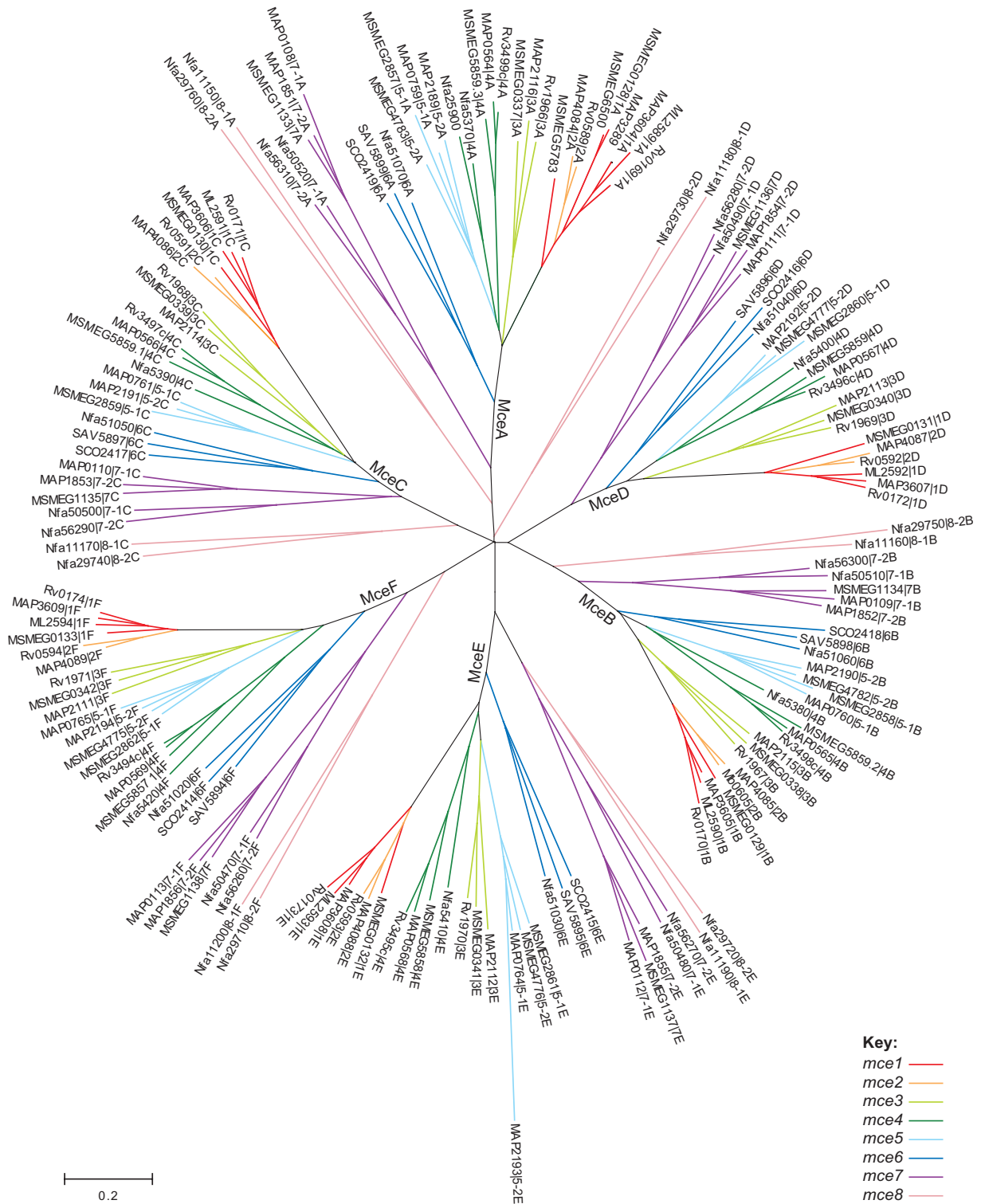
<sup>f</sup> Orthology inferred from synteny.

<sup>g</sup> Contains frameshift mutation, resulting in two ORFs.

<sup>h</sup> Not a member of IPR003399 or IPR005693.

<sup>i</sup> Insertion of mobile element.

<sup>j</sup> Orthology inferred from synteny.



**Figure 5**  
**Phylogenetic tree of Actinomycetales Mce proteins.** A non-redundant set of Mce protein sequences were aligned and an unrooted neighbor-joining tree was computed by MEGA. Coloring corresponds to the classification scheme specified in Table 3. ORFs are designated by [gene locus name | operon number (1-8) and gene position (A-F)]. Where operon orthology cannot be inferred, operons are designated: -1, -2.

**Figure 6**

**Illustration of conserved regions and predicted secondary structure of Actinomycetales Mce proteins.** Six separate alignments of the Mce proteins (A-F) listed in Table 3 were submitted to JPred and the consensus secondary structure prediction estimated manually. White boxes represent  $\alpha$ -helices and grey arrows  $\beta$ -strands. The C-terminal proline-rich region had low complexity and varied in length from 10–250 amino acids. Signal sequences were identified by SignalP and lipid attachment sites matched the ProSite motif PS00013.

*cinica*, *S. coelicolor* and *S. avermilitis* were selected for further analysis. In several cases the ORF downstream of *yrbEA* was either not annotated or annotated in the reverse direction; however, translation of the genomic sequence revealed a YrbEB homolog encoded in the expected direction (Table 3).

Phylogenetic analysis showed deep branching between the YrbEA and YrbEB sequences (Figure 7). Within each clade the clustering of sequences was almost identical demonstrating that the *yrbEA-yrbEB* genes have evolved as a pair. The clustering was comparable to that seen in the Mce protein tree, with members of the *mce1/2* and *mce3* to *mce8* operons easily distinguishable. Thus, it appears that all of the operons examined evolved from a common ancestral eight-gene cluster without shuffling of genes within or between operons.

ABC permeases typically contain six transmembrane segments with the C-terminus located on the cytoplasmic side of the membrane [11]. The consensus TMHMM-predicted structure of Actinomycetales YrbE homologs found in *mce* operons suggests the presence of five or six transmembrane helices with the C-terminus outside (Figure 8a). The presence of the N-terminal transmembrane helix was equivocal, and therefore the N-terminus may be cytoplasmic or outside. Further topological predictions using the programs HMMTOP and TopPred confirmed this model, but were unable to verify or refute the existence of the N-terminal transmembrane segment.

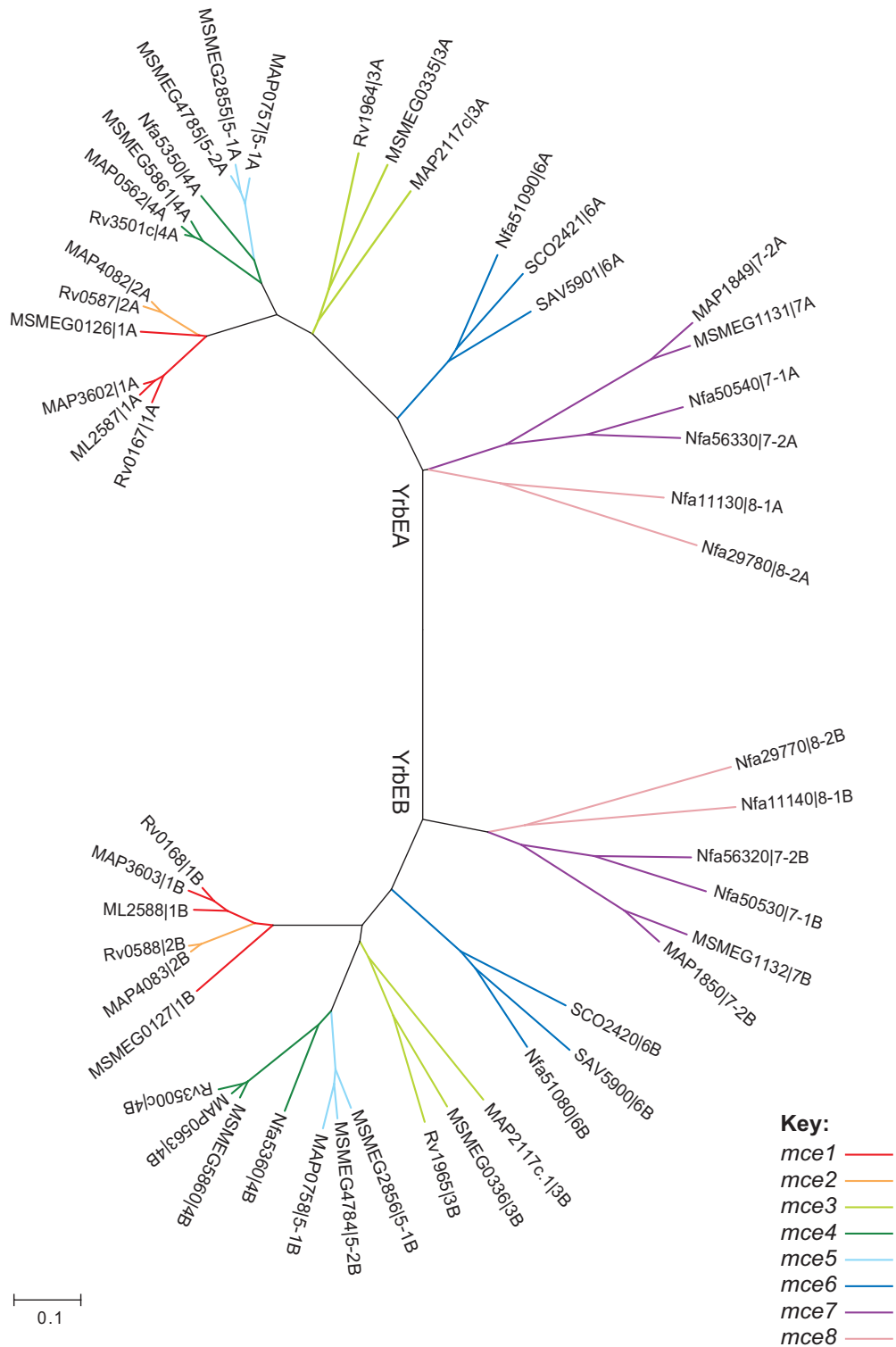
Dassa and colleagues [41,42] have described a highly-conserved sequence, the EAA motif, in the final cytoplasmic

loop of some SBP-dependent ABC permeases that is proposed to interact with the cognate ATPase [43]. Examination of the multiple alignment of YrbE proteins revealed a conserved sequence motif located in the penultimate cytoplasmic loop. The consensus deduced from 50 Actinomycetales YrbEA and YrbEB sequences is shown in Figure 8b. Alignment of Gram-negative bacterial DUF140 proteins revealed that this region was highly conserved in all family members. The consensus sequence we have deduced does not appear to be homologous to the published motifs, but does contain the common invariant glycine residue and is predicted to adopt the typical  $\alpha$ -helical structure [42]. The consensus 47 amino acid YrbE sequence, that we have termed the EExDA motif, was able to specifically retrieve Actinomycetales and Gram-negative DUF140 proteins from the National Center for Biotechnology Information (NCBI) microbial proteomes database.

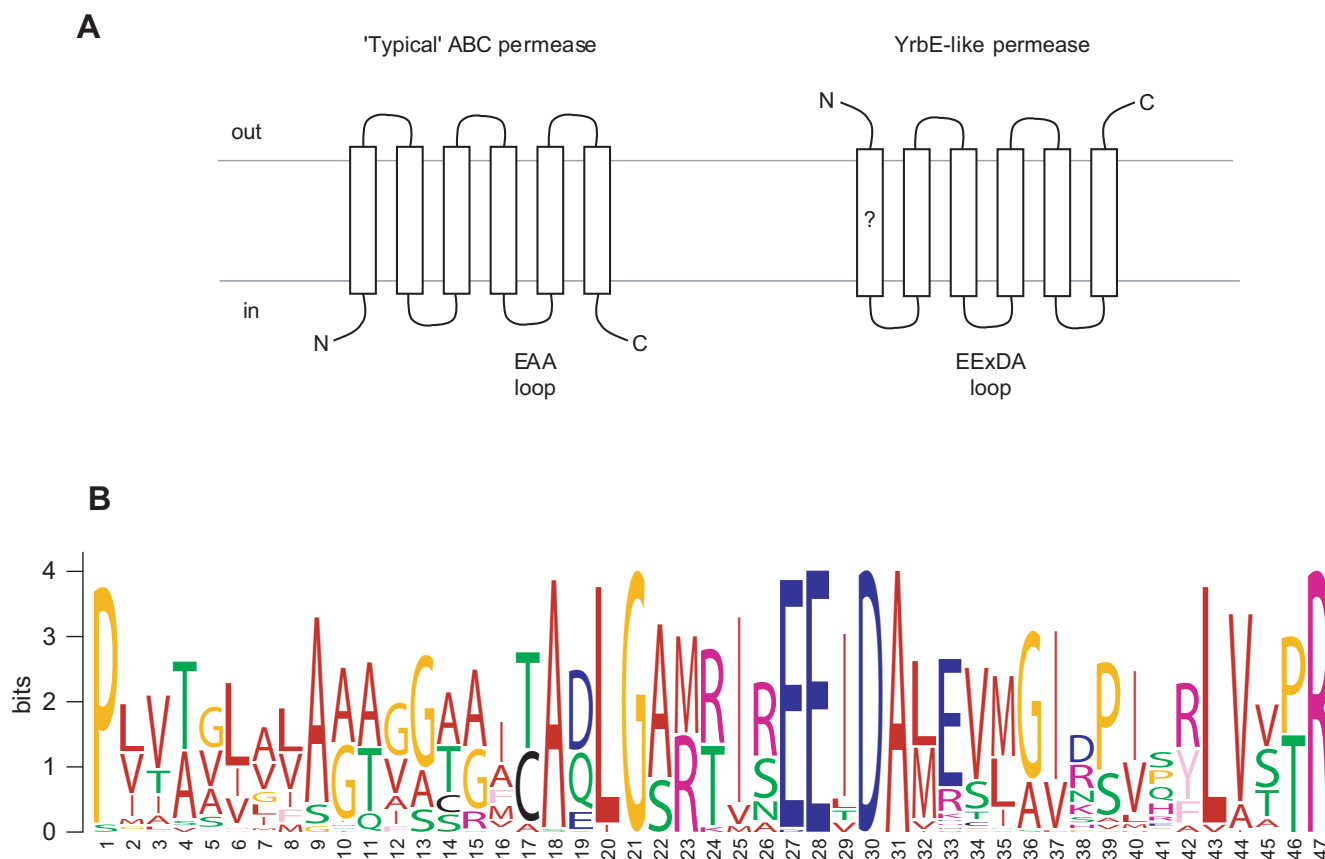
In one case (*Rhodopirellula baltica*, RB3287) a DUF140 domain is fused to an ABC ATPase domain providing evidence that the function of DUF140 proteins requires ATP hydrolysis [21].

#### The Mas proteins

The four genes downstream of the *M. tuberculosis mce1* operon, as well as two each downstream of the *mce3* and *mce4* operons, are annotated in TubercuList [44] as 'conserved mce-associated proteins' (herein termed Mas). The *mce1* operon transcript has been empirically demonstrated to include the associated *mas* genes (Rv0175-78) [45]. Examination of a multiple alignment of the protein sequences revealed that they were not conserved along



**Figure 7**  
**Phylogenetic tree of Actinomycetales YrbE proteins.** A non-redundant set of YrbE protein sequences were aligned and an unrooted neighbor-joining tree was computed by MEGA. Coloring corresponds to the classification scheme specified in Table 3. ORFs are designated by [gene locus name | operon number (1–8) and gene position (A, B)]. Where operon orthology cannot be inferred, operons are designated: -1, -2.



**Figure 8**  
**Predicted topology and conserved sequence motif of Actinomycetales YrbE proteins.** (A) The consensus topology prediction of Actinomycetales YrbE proteins analysis is shown compared to that of a typical ABC permease [42]. (B) WebLogo illustration of the conserved YrbE EExDA sequence motif identified through MEME analysis.

their entire length but shared a similar C-terminal region of approximately 160 amino acids. Pairwise sequence identity scores, generated by ClustalX, for the conserved region ranged from 12 to 25%.

To determine whether homologous domains were present in other genomes, we used each of the eight Mas C-terminal sequences as a PSI-BLAST query against the NCBI non-redundant database. A total of 137 sequences were retrieved; of these, 124 sequences were hit by all eight query sequences, and all 137 were hit by more than two queries. The proteins identified belonged to six genera: *Amycolatopsis*, *Janibacter*, *Mycobacterium*, *Nocardia*, *Nocardioles* and *Streptomyces*. Thus, the phylogenetic profile for the putative Mas homologs in *Actinomycetales* genera exactly matches that of the Mce, DUF140 and Mkl proteins. Mas homologs in the *M. smegmatis* genome, which was not covered by the NCBI database, were identified by exhaustive BLAST querying of the TIGR proteome. Nine-

teen putative Mas homologs were thus identified ( $P < 0.00001$ ).

Sequences of the putative Mas domain containing proteins from *M. tuberculosis*, *M. leprae*, *M. paratuberculosis*, *M. smegmatis*, *N. farcinica*, *S. avermitilis* and *S. coelicolor* were selected for further analysis. This resulted in a set of 66 sequences (including one hybrid sequence, MAP2107/9c, that has been disrupted by a transposase).

The Mas domain genes were typically found in pairs (58 of 66) and the majority (43 of 66) were encoded downstream of, and in the same direction, as *mce* genes (Table 4). Putative orthologs of each of the eight *M. tuberculosis* *mce* operon-associated *mas* genes were identified in the corresponding positions of those genomes carrying orthologous operons. Each of the *mce7* operons had a single Mas protein encoded downstream. The *mce6* operons of *N. farcinica* and *S. avermitilis* contained two *mas* genes,

while the corresponding *S. coelicolor* operon carried four. In *M. paratuberculosis*, a pair of *mas* homologs was located in the regions both upstream and downstream of the *mce5* operon, but transcribed from the opposite strand (MAP0750-51c, MAP0767-68c). The 23 non-*mce* operon-associated *Mas* homologs were generally located in pairs in isolated operons. An exception was Rv2390c, which TIGR predicts is part of a three-gene operon including a resuscitation promoting factor (*rpfD*, Rv2389c) and an Fe-S enzyme involved in porphyrin biosynthesis (*hemN*, Rv2388c).

The *Mas* region is not currently recognized as a conserved domain in the databases. However, within this region, InterPro recognized a lipocalin family motif (IPR002345) in Rv3492c, and a partial C2 domain signature

(IPR000008) in Rv0199 and ML2614. Notably, the corresponding Pfam families (PF00061 and PF00168) did not include these sequences as members. Nonetheless, it may be worthy of mention that the lipocalin and C2 domains share a lipid-binding function, as well as an eight-stranded anti-parallel beta sandwich structure [46,47].

The majority of pairwise identity scores for the 66 *Mas* domains were 10–20%. This low level of sequence similarity resulted in multiple sequence alignments that were extremely sensitive to input parameters. Exclusion of the 13 non-mycobacterial sequences produced a much more robust alignment. A phylogenetic tree generated from this alignment is shown in Figure 9. Examination of the tree revealed that the *Mas* proteins encoded by the first and second genes in each pair formed phylogenetically dis-

**Table 4: Mas Homologs in Selected Actinomycetales Genomes<sup>ab</sup>**

	Rv	ML	MAP	MSMEG	nfa	SAV	SCO
<b>Mas1 A</b>	0175 (213)	2595 (182)	3610 (213)	0134 (202)			
<b>B</b>	0176 (322)	2596 (325)	3611 (323)	0135 (288)			
<b>C</b>	0177 (184)	2597 (184)	3612 (184)	0136 (182)			
<b>D</b>	0178 (244)	2598 (184)	3613 (252)	0137 (296)			
<b>Mas3 A</b>	1972 (191)		2110c (203)	0343 (200)			
<b>B</b>	1973 (160)		2109/7c <sup>c</sup>	0344 (202)			
<b>Mas4 A</b>	3493c (242)		0570 (243)	5857 (233)	5430 (315)		
<b>B</b>	3492c (160)		0571 (164)	5856 (161)	5440 (162)		
<b>Mas6 A</b>					51010 (248)	5893 (177)	2413 (170)
<b>B</b>					51000 (274)	5892 (272)	2412 (219)
<b>C</b>							2411 (184)
<b>D</b>							2410 (253)
<b>Mas7-1 A</b>			0114 (198)	1139 (230)	50460 (246)		
<b>Mas7-2 A</b>			1857 (227)		56250 (321)		
<b>ClusterI A</b>	1363c (261)		0751c (295)	4759.2 (303)			
<b>B</b>	1362c (220)		0750c (187)	4759 (200)			
<b>A</b>			0768c (298)	2867 (190)			
<b>B</b>			0767c (224)	2868 (218)			
<b>ClusterII A</b>	0199 (219)	2614 (229)		0225 (206)	6070 (197)		
<b>B</b>	0200 (229)	2615 (224)		0226 (229)			
<b>A</b>	2390c (185)		0090c (212)				
<b>A</b>				0878 (167)			
<b>B</b>				0879 (496)			
<b>A</b>				5189 (231)			
<b>B</b>				5190 (192)			

<sup>a</sup> Organism specific gene number prefix: Rv, *M. tuberculosis* H37Rv; ML, *M. leprae*; MAP, *M. paratuberculosis*; MSMEG, *M. smegmatis*; nfa, *N. farcinica*; SCO, *S. coelicolor*; SAV, *S. avermitilis*.

<sup>b</sup> Each row contains putative orthologs. Length of protein in amino acids shown in parentheses.

<sup>c</sup> ORF is interrupted by a transposase, MAP2108.

tinct clusters. The Mas proteins encoded adjacent to *mce* operons were not separated from the non-*mce* associated Mas proteins. The *M. leprae*, *M. paratuberculosis* and *M. smegmatis* Mas proteins associated with the *mce1*, *mce3* and *mce4* operons are clearly orthologs of those in the corresponding genomic positions in *M. tuberculosis*. The *mce7*-associated Mas proteins also cluster together. Several pairs of non-*mce* associated Mas homologs were conserved between mycobacterial species (Figure 9; Cluster I and Cluster II).

The mycobacterial *mce*-associated Mas orthologs have greater than 50% pairwise identity. In contrast, the *Nocardia* and *Streptomyces mce6*-associated Mas proteins are highly divergent (15–20% identity). This suggests that, unlike the *mce* and *yrbE* genes, the *mas* genes have either diverged more rapidly or were independently recruited to the operons.

Comparison of JPred secondary structure predictions for orthologous clusters revealed the consensus structure of the conserved domain was  $\alpha_1\alpha_2\alpha_3\alpha_4\beta_1\beta_2\beta_3\beta_4$ . Prediction of transmembrane helices indicated that all 66 protein sequences harbored a transmembrane segment located about 140–180 amino acids from the C-terminus and corresponding to  $\alpha_1$ . Topology prediction programs, TMHMM, HMMTOP and TopPred, suggested the C-terminus was extracellular for 41, 56 and 42, of the 66 submitted sequences, respectively. In no case did all three programs predict an extracellular N-terminus for a single protein. Thus, it seems likely that all N-termini are intracellular, while the C-terminal Mas domains are located on the external side of the cytoplasmic membrane.

The length of the N-terminal region preceding the Mas domain ranged from 7 to 325 amino acids. In the majority of proteins in which the N-terminal segment was less than 30 amino acids (11 of 16),  $\alpha_1$  was predicted to be a signal peptide by SignalP (Figure 10). Consensus topology predictions indicated that the four Mas1B orthologs and three Cluster IIB proteins contained two N-terminal transmembrane helices (oriented in-out, out-in). In the Mas1B orthologs, the two N-terminal transmembrane segments correspond to an RDD domain (IPR010432). Examination of a multiple alignment revealed that although *M. smegmatis* Mas1B does not actually have the N-terminal signature RD residues, the Cluster IIB proteins do. It has been proposed that the RDD domain is involved in transport [31]; however, to date, no empirical evidence has been published to support this claim. In MSMEG0879 the 325 amino acid N-terminal region encodes a protein kinase domain (IPR000719) containing the Ser/Thr kinase active site motif (PS00108). Coiled-coils, which are known to mediate protein-protein interactions [48], were

identified in the N-terminal region of each Cluster IA sequence by the Lupas COILS algorithm.

## Discussion

In this study we sought to gain insight into the function of the *M. tuberculosis mce* operons using genome comparisons and bioinformatic methods.

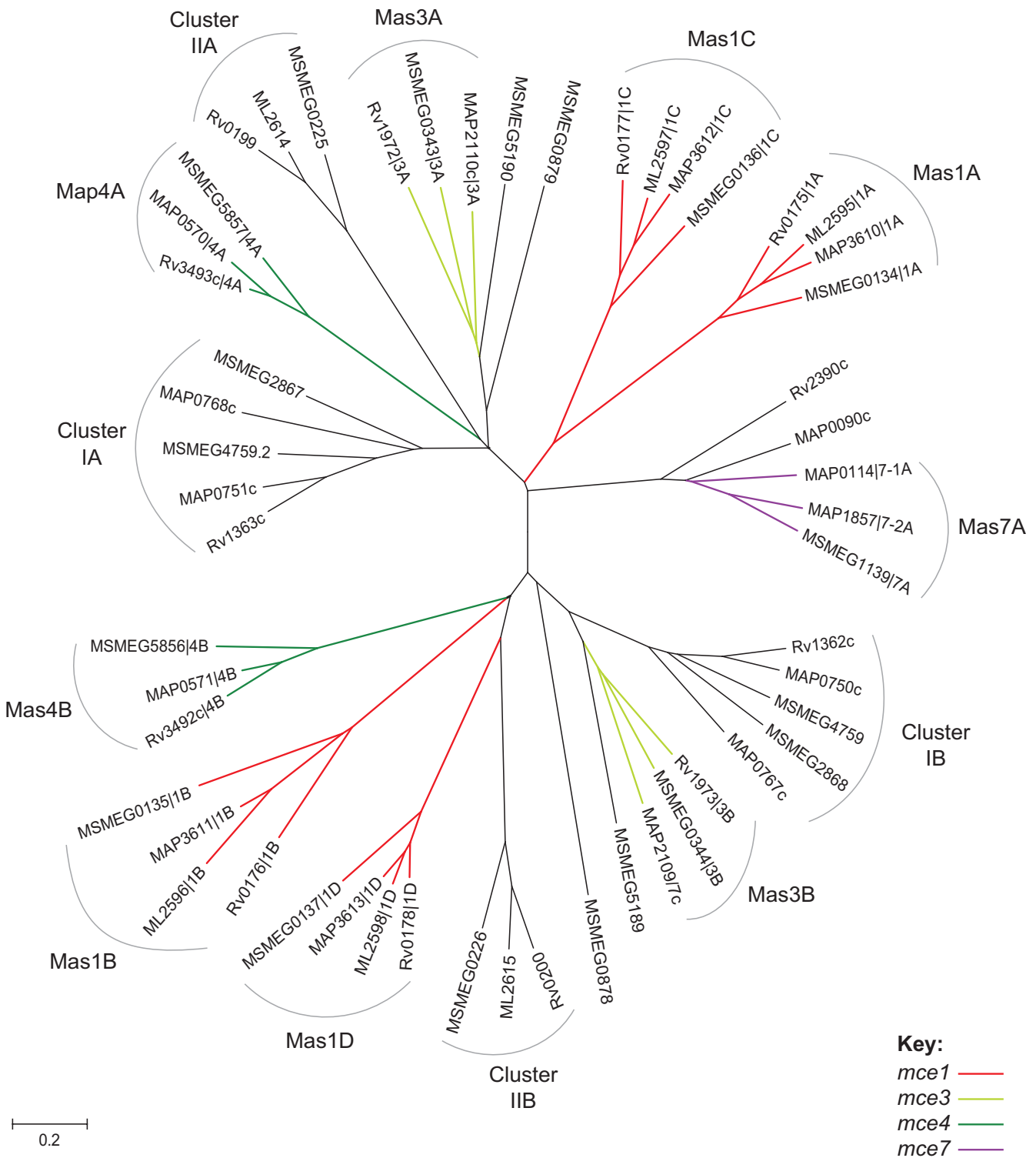
The YrbE and Mce proteins, encoded by the *M. tuberculosis mce* operons, have homology to the permease and SBP components of ABC transporters, respectively [29]. However, sequence similarity within these protein families is notoriously low, and confirmation that the *mce* operons encode ABC importers has required identification of the necessary cognate ATPase. Dassa and Bouige [8] have proposed that Rv0655, an ATPase named Mkl, might supply this function and here we provide substantial evidence that this is indeed the case.

Firstly, Mkl orthologs are encoded immediately upstream of the mycobacterial-like *mce* operons in species of *Nocardia*, *Janibacter*, *Nocardioideis*, *Amycolatopsis* and *Streptomyces*. Secondly, orthologs of Mkl are found in all, and in only, those *Actinomycetales* species that also contain Mce and DUF140 homologs. The presence of an intact *mkl* gene in the *M. leprae* genome, which has undergone extensive reductive evolution [49], is significant in this respect. Thirdly, in Gram-negative bacteria, operons containing DUF140 and *mce* homologs invariably include the orthologous *mkl* gene. Recently, Joshi *et al.* [7] observed that in competitive mouse infections an Rv0655 mutant was attenuated relative to wild-type *M. tuberculosis*, whereas an Rv0655-*mce1* double mutant showed no attenuation relative to the *mce1* mutant, providing evidence that Rv0655 and the Mce1 proteins are functionally linked. It is notable that in the *Mycobacterium* species examined, the *mkl* gene is located within the genomic region that encodes the majority of ribosomal proteins; this is generally the most conserved region in prokaryotic genomes and could facilitate high level expression of *mkl* [40].

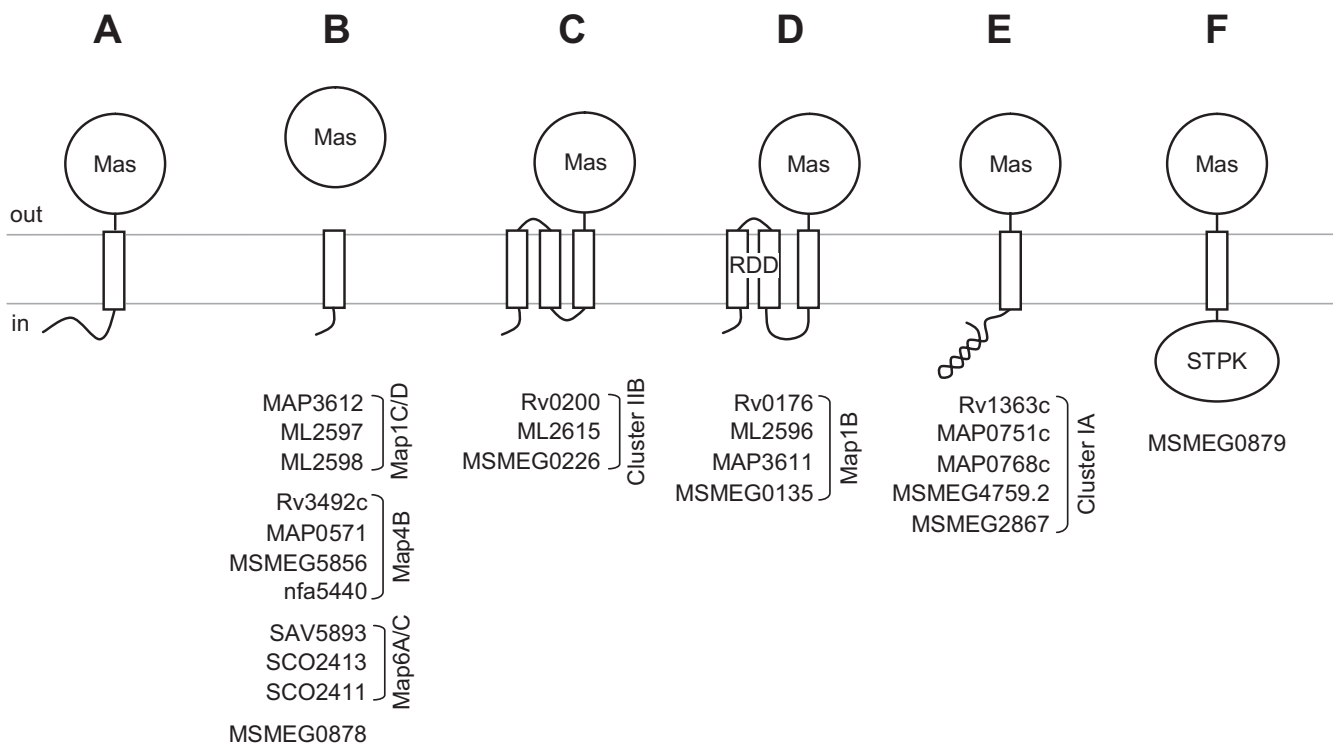
It is widely accepted that the direction of substrate transport of ABC transporters can be predicted on the basis of ATPase homology [10]. In phylogenetic analyses, Mkl ATPases fall into the importer clade [8,20]; this prediction is consistent with the proposed role of Mce proteins as SBPs, which are found exclusively in substrate import systems.

The results of topology prediction indicated that the YrbE proteins contained five to six transmembrane segments, with the C-terminal five the most conserved and the C-terminus outside. In support of this model, the periplasmic location of the C-terminus of *E. coli* YrbE has been demonstrated empirically [50]. In general, ABC permeases





**Figure 9**  
**Phylogenetic tree of mycobacterial Mas domain sequences.** The conserved Mas domains of mycobacterial proteins listed in Table 4 were aligned and an unrooted neighbor-joining tree was computed by MEGA. Coloring corresponds to the classification scheme specified in Table 3. ORFs are designated by [gene locus name | operon number (1, 3, 4, 7) and gene position (A-D)]. Where operon orthology cannot be inferred, operons are designated: -1, -2.

**Figure 10**

**Representative architectures of Mas domain-containing proteins.** Membrane topology predictions for the 66 Mas proteins listed in Table 4 indicated that the conserved domain was located on the extracellular side of the cytoplasmic membrane. The Mas domain was predicted to remain anchored in the majority of proteins (A), but cleaved in eight (B). Three transmembrane segments were identified in seven proteins and four of these were classified as RDD domains (C, D). Five proteins contained an N-terminal coiled-coil region (E), and one, a serine-threonine protein kinase domain (STPK; F).

show the highest level of sequence similarity over the C-terminal five transmembrane regions, and this is considered to be the minimal functional unit [11]. In compiled alignments of ABC permease sequences, the most conserved region localizes to the final cytoplasmic loop [42]. This motif, termed the EAA loop, likely interacts with the cognate ATPase [43]. A highly conserved motif, predicted to localize to the penultimate cytoplasmic loop, was identified in YrbE proteins from both *Actinomycetales* and Gram-negative bacteria. We propose that this motif, named the EExDA loop, serves as the site of interaction with the putative cognate Mkl ATPase, in a manner analogous to the EAA loop.

Conservation of the 'two *γrbE* plus six *mce*' operon structure suggests that these components comprise the functional unit of the canonical *Actinomycetales* Mce transporter [22,40]. We have found that mutation of either the *γrbE1A*, *mce1A* or *mce1E* genes of *M. tuberculosis* results in undetectable levels of all the Mce1 proteins, implying that these proteins are part of a hetero-octomeric complex and its formation is necessary for stability of the Mce proteins [4] (L. Morici, personal communication). It

is interesting that many *Proteobacteria* contain membrane proteins with multiple Mce domains (PqiB proteins) that could potentially interact forming a quaternary structure analogous to the putative *Actinomycetales* Mce complex. The permease components of ABC transporters, that form a channel across the cytoplasmic membrane, are frequently heterodimers; however, although present in stoichiometric excess, SBPs are generally encoded by one or two genes [11]. The presence of six SBPs is, thus far, a unique characteristic of the *Actinomycetales* Mce transporters. Using computational methods, Pajon *et al.* [51] found that the  $\beta$ -sheet region of eight of the *M. tuberculosis* Mce proteins contained patterns typical of transmembrane  $\beta$ -strands and suggested that this region could promote penetration of the outer lipid layer. Thus, it is tempting to speculate that the Mce proteins are designed to form a channel that crosses this lipid bilayer. Chitale *et al.* [52] have previously shown that Mce1A is indeed exposed on the surface of *M. tuberculosis*.

Proteins encoded downstream of three of the four *M. tuberculosis* *mce* operons exhibit significant sequence homology. Similarity is confined to the 160 amino acid C-

terminal region, we have termed the Mas domain, that is predicted to localize to the extracellular side of the cytoplasmic membrane. In each of the *Actinomycetales* genomes examined, Mas domain proteins were found linked to the majority of *mce* operons. Mas proteins show absolute phylogenetic congruency with Mkl, DUF140 and Mce proteins in the genomes of *Actinomycetales*, providing evidence that they are involved in Mce transporter function. Given that Mas domains are not found associated with all *mce* operons, their function may not always be strictly required or they may be shared between operons. The propensity of Mas homologs to be located in pairs suggests that they form heterodimers. Such an interaction would likely keep the predicted secreted Mas proteins tethered to the cell surface. The domain architectures of the Mas proteins suggest that the conserved domain plays an accessory ligand-binding role.

Several studies have shown that the  $\gamma$ -proteobacterial *mce* loci play a role in determination of structural properties of the cell envelope, which in pathogenic species affects invasive activity. In *Pseudomonas putida*, a transposon insertion within the DUF140-Mce-associated *ttg2A* ATPase (PP0958) renders the cells sensitive to toluene [53]. In addition to toluene degradation and efflux, toluene tolerance is known to be mediated by increased cell membrane rigidity resulting from changes in fatty acid and phospholipid composition [54]. In *Shigella flexneri*, mutations in the *vpsABC* locus (S\_3453-51), encoding an ABC transporter with the ATPase-DUF140-Mce configuration, result in a defect in intercellular spread through epithelial cell monolayers, altered colony morphology, increased sensitivity to detergent lysis and hypersecretion of both Sec-dependent and TypeIII-dependent virulence proteins [55]. Carvalho *et al.* have reported that in *Campylobacter* isolates, presence of *iamA*, the ATPase gene of the *mce* operon (Cj1646-48), correlated with an invasive phenotype [56], although, this association remains controversial [57-59]. In *Neisseria meningitidis* the *mce*-like operon, *glT* (NMB1966-64), belongs to the GdhR regulon, which is expressed at higher levels in invasive versus commensal isolates, and is particularly elevated in hypervirulent lineages [60].

Comparable function has been attributed to the *M. tuberculosis mce1* operon. The prototypical Mce protein, *M. tuberculosis* Mce1A, conferred invasive ability upon *E. coli* and an *M. bovis* BCG *mce1A* mutant exhibited impaired invasion of epithelial cells [1,61]. Moreover, an *M. tuberculosis mce1* operon mutant has been shown to have an overabundance of free mycolic acids in the outer lipid layer (S. Cantrell, personal communication), supporting the proposition that *mce1* and related operons play a role in remodeling the cell envelope. The presence of *mce* operons in Gram-negative bacteria and *Actinomycetales* genera

that possess a somewhat analogous outer lipid bilayer raises the possibility that the *mce* operons are involved in maintenance of outer membrane integrity. However, their presence in other *Actinomycetales* with typical Gram-positive type cell envelopes appears to preclude this hypothesis. In addition, the absence of *mce* operons in *Corynebacterium* species indicates that their function is not essential for maintenance of an outer lipid bilayer.

Based on a stated similarity of the ATPase component to GluA of *Corynebacterium glutamicum*, Meidanis *et al.* [62] proposed that the *Xylella fastidiosa mce*-like operon (XF0421-19) encoded a glutamate importer. It was subsequently shown that a mutation within the homologous *N. meningitidis glT* operon resulted in impaired glutamate-specific uptake at low sodium concentrations [63]. Glutamate is a prominent constituent of peptidoglycan; thus, disruption of its uptake in the proteobacterial *mce* operon mutants could perhaps account for the observed effect on cell envelope properties. Also relevant in this respect, is the conserved location of the peptidoglycan biosynthetic gene, *murA*, downstream of the Mce transporter genes in  $\gamma$ -Proteobacteria.

Homologs of the Mkl, Mce and DUF140 proteins have also been identified in plants [64]. The *Arabidopsis* homologs of DUF140 (TGD1, At1g19800) and Mce (TGD2, At3g20320) both localize to the inner plastid membrane, with the Mce domain located in the intramembrane space. Lipid binding studies demonstrated that TGD1 specifically bound 1,2-diacyl-*sn*-glycerol 3-phosphate (phosphatidic acid). TGD1 and TGD2 mutants exhibited identical phenotypes consistent with disruption of transport of ER-derived phosphatidic acid into chloroplasts, suggesting the TGD proteins form part of a lipid translocator [65-67].

Orthologous ABC transporters are expected to be functionally equivalent [13-15], thus the proposal of both phosphatidic acid and glutamate as possible substrates of the Mce transporters is puzzling. It is noteworthy that in sequence analyses, by us and others, the Mkl-like ATPases are not closely related to GluA [8]. If the bacterial Mce homologs have phospholipid binding function, equivalent to TGD1, this might enable interaction with host cell membranes and explain the invasive phenotype associated with the *mce* loci. It is generally accepted that host-derived lipids are the primary source of carbon utilized by *M. tuberculosis in vivo* [68]; however no mechanism of lipid import has been identified. Thus it is enticing to hypothesize that the Mce transporters might perform this role. Inclusion of the fatty-acyl CoA synthetase, *fadD5*, in the *mce1* operon and repression of the operon by a FadR-like regulator, lends some support to this conjecture [45].

The canonical eight-gene *mce* operon has undergone extensive proliferation and deletion events within certain *Actinomycetales* lineages, most notably in *Mycobacterium* and *Nocardia* species. The simplest explanation for the presence of multiple *mce* operons is that it facilitates elevated expression. However, evidence from transcriptional analyses of *M. tuberculosis* suggest that, at least in this organism, the operons are not co-regulated [69-72]; in addition, three of the four operons are associated with transcriptional regulators [45,73]. In competitive mouse infections, Sasseti and Rubin [6] found that an *mce1* mutant exhibited a growth defect during the first 1–2 weeks of infection, whilst an *mce4* mutant showed attenuation 3–4 weeks after inoculation. These observations support the proposition that the operons function at different stages of infection. Differential expression of the individual Mce transporters may reflect optimization for substrate uptake under differing conditions, such as in the low sodium intracellular environment; alternatively, they might have varying substrate specificities.

The number of *mce* operons in individual species appears to reflect the variety of environmental niches inhabited. Thus, the fast-growing, typically soil-dwelling, *Mycobacterium* species possess the greatest number, with polycyclic aromatic hydrocarbon-degrading species, isolated from bioremediation sites, containing the most [74]. In contrast, the host-specialized, slow-growing pathogenic species possess fewer operons, and the obligate intracellular pathogen, *M. leprae*, encodes a single complete *mce* operon. A high degree of sequence similarity indicates that the *mce1* operon duplicated to create *mce2* relatively recently. In *M. tuberculosis* complex strains, *mce* frameshift mutations are found conspicuously in these two operons: of the five described in this paper, four are in *mce2* and the fifth is in *mce1*. This pattern may reflect the functional divergence of the *mce1* and *mce2* operons.

With the exception of mycolic acids, the distribution of morphological and chemotaxonomic traits within the *Actinomycetales* is polyphyletic [75]. Given the incongruent taxonomic distribution of the *mce* operons and their proposed role in integrity of the cell envelope, it is pertinent to note that presence of *mce* operons does not correlate with type of peptidoglycan, menaquinones, phospholipids or fatty acids in the cell envelope [75,76]. In addition, there is no correlation with oxygen requirement, habitat or pathogenicity.

## Conclusion

The available evidence suggests that the *mce* operons encode a novel subfamily of ABC transporter uptake systems comprised of DUF140 permease components, Mce-like substrate-binding proteins, and Mkl-type ATPase domains. Disruption of *mce* operons, in both *Actinomyc-*

*etales* and Gram-negative bacteria, affects properties of the cell envelope and associated virulence phenotypes of pathogenic species. Empirical studies have implicated both glutamate and phosphatidic acid as substrates of *mce*-like transporters; thus, although the precise substrate specificity of the *M. tuberculosis* Mce transporters remains uncertain, we conclude that it is likely to be an organic acid precursor of cell envelope biogenesis.

## Methods

### Databases

Gene annotations and protein sequences were obtained from the publicly available databases: UniProt [77,78]; TIGR Comprehensive Microbial Resource (CMR) [79,80]; NCBI Microbial Genome Project [81]; Joint Genome Institute Microbial Genomics Database [82]; and TubercuList [44]. Sequences are referred to by the ordered locus name provided in these databases. Protein classification was informed by interrogation of conserved domain and motif databases: InterPro (IPR) [26,83], Pfam (PF) [27,31], TIGRFAM (TIGR) [28,79], and PROSITE (PS) [84,85]. The ABC transporter classification database, ABCISSE, was also consulted [29].

### BLAST analyses

Sequence similarity searches were performed by BLASTP against complete microbial genome sequences deposited in the TIGR-CMR and NCBI Microbial Genome Project databases [79,81,86]. To determine whether the EExDA motif identified in YrbE proteins was uniquely characteristic of the DUF140 family, we performed a BLASTP search of NCBI Microbial Genome Project with the *Actinomycetales* YrbE consensus motif (PLVTGLALAGAGGAITADLGARRIREIDALEVMGID-PISRLVVP) using the default parameters, except with no filter and expect threshold of 100. To identify homologs of the *M. tuberculosis* Mas domain, each of the eight sequences was used in a PSI-BLAST query against the NCBI non-redundant database [87]. We used an inclusion threshold of  $P < 10^{-5}$  and the scores were adjusted with composition-based statistics; these parameters resulted in convergence after 6–8 iterations.

### Multiple alignment and phylogenetic analyses

Phylogenetic analyses were conducted using the MEGA version 3.1 suite of programs [88]. Multiple alignments were constructed by CLUSTAL-W using the Gonnet weight matrix and default gap penalties [89]. Unrooted trees were computed by the neighbor-joining method. The consensus tree, after 500 bootstrap replicates, was displayed graphically with Tree Explorer. In addition, CLUSTAL-W alignments were converted to PHYLIP format and trees computed by the maximum likelihood method implemented by PROML using default parameters [90]. In all cases this resulted in a tree with topology that was essen-

tially the same as the neighbor-joining tree generated by MEGA. Percentage pairwise similarity scores were calculated by CLUSTAL-X [91].

#### Identification of conserved motifs

The MEME server was used to discover highly conserved sequence motifs within groups of homologous proteins [92,93]. Motifs were displayed graphically using WebLogo [94,95].

#### Secondary structure and topology prediction

Groups of aligned orthologs were submitted to JPred [96], a consensus secondary structure prediction server, that provides improved accuracy over single sequence prediction methods [97]. Comparison of predictions between orthologous clusters by visual inspection allowed estimation of the consensus structure for a homologous family. Coiled-coils were predicted using the Lupas COILS algorithm through the JPred server [98].

Protein sequences were analyzed by SignalP and TatP to identify Sec- and Tat-dependent signal sequences [38,39,99]. The reliability of prediction of transmembrane helices and topology of proteins increases when different methods are combined [100]. Hence, we submitted sequences to TMHMM [101,102], HMMTOP [103,104] and TopPred [105,106], and determined the consensus prediction by manual comparison.

#### Authors' contributions

NC conceived, designed and performed the study. LWR helped to interpret the data. NC drafted the manuscript; both authors read and approved the final manuscript.

#### Acknowledgements

We are grateful to Owen Solberg and Sally Cantrell for useful discussions and critical reading of the manuscript. We also thank S.C. and Lisa Morici for sharing unpublished data. This work was supported by grants from NIH (R21AI063350) and the Senior Scholar Award in Global Infectious Disease of the Ellison Medical Foundation.

#### References

1. Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley LW: **Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells.** *Science* 1993, **261**(5127):1454-1457.
2. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG: **Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.** *Nature* 1998, **393**(6685):537-544.
3. Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST: **Analysis of the proteome of Mycobacterium tuberculosis in silico.** *Tuber Lung Dis* 1999, **79**(6):329-342.
4. Shimono N, Morici L, Casali N, Cantrell S, Sidders B, Ehrst S, Riley LW: **Hypervirulent mutant of Mycobacterium tuberculosis resulting from disruption of the mceI operon.** *Proc Natl Acad Sci USA* 2003, **100**(26):15918-15923.
5. Gioffre A, Infante E, Aguilar D, De la Paz Santangelo M, Klepp L, Amadio A, Meikle V, Etchechoury I, Romano MI, Cataldi A, Hernandez RP, Bigi F: **Mutation in mce operons attenuates Mycobacterium tuberculosis virulence.** *Microbes Infect* 2005, **7**(3):325-334.
6. Sasseti CM, Rubin EJ: **Genetic requirements for mycobacterial survival during infection.** *Proc Natl Acad Sci USA* 2003, **100**(22):12989-12994.
7. Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sasseti CM: **Characterization of mycobacterial virulence genes through genetic interaction mapping.** *Proc Natl Acad Sci USA* 2006, **103**(31):11760-11765.
8. Dassa E, Bouige P: **The ABC of ABCs: A phylogenetic and functional classification of ABC systems in living organisms.** *Res Microbiol* 2001, **152**(3-4):211-229.
9. Kumar A, Chandolia A, Chaudhry U, Brahmachari V, Bose M: **Comparison of mammalian cell entry operons of mycobacteria: In silico analysis and expression profiling.** *FEMS Immunol Med Microbiol* 2005, **43**(2):185-195.
10. Saurin W, Hofnung M, Dassa E: **Getting in or out: Early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters.** *J Mol Evol* 1999, **48**(1):22-41.
11. Boos W, Eppler T: **Prokaryotic binding protein-dependent ABC transporters.** In *Microbial Transport Systems* Edited by: Winkelmann G. Weinheim, Germany, Wiley VCH; 2002:77-114.
12. Ames GF, Liu CE, Joshi AK, Nikaido K: **Liganded and unliganded receptors interact with equal affinity with the membrane complex of periplasmic permeases, a subfamily of traffic ATPases.** *J Biol Chem* 1996, **271**(24):14264-14270.
13. Tam R, Saier MH Jr.: **Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria.** *Microbiol Rev* 1993, **57**(2):320-346.
14. Saurin W, Dassa E: **Sequence relationships between integral inner membrane proteins of binding protein-dependent transport systems: Evolution by recurrent gene duplications.** *Protein Sci* 1994, **3**(2):325-344.
15. Kuan G, Dassa E, Saurin W, Hofnung M, Saier MH Jr.: **Phylogenetic analyses of the ATP-binding constituents of bacterial extracytoplasmic receptor-dependent ABC-type nutrient uptake permeases.** *Res Microbiol* 1995, **146**(4):271-278.
16. Walker JE, Saraste M, Runswick MJ, Gay NJ: **Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold.** *EMBO J* 1982, **1**(8):945-951.
17. Ames GF, Mimura CS, Holbrook SR, Shyamala V: **Traffic ATPases: A superfamily of transport proteins operating from Escherichia coli to humans.** *Adv Enzymol Relat Areas Mol Biol* 1992, **65**:1-47.
18. Linton KJ, Higgins CF: **The Escherichia coli ATP-binding cassette (ABC) proteins.** *Mol Microbiol* 1998, **28**(1):5-13.
19. Quentin Y, Fichant G, Denizot F: **Inventory, assembly and analysis of Bacillus subtilis ABC transport systems.** *J Mol Biol* 1999, **287**(3):467-484.
20. Braibant M, Gilot P, Content J: **The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis.** *FEMS Microbiol Rev* 2000, **24**(4):449-467.
21. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D: **Detecting protein function and protein-protein interactions from genome sequences.** *Science* 1999, **285**(5428):751-753.
22. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N: **The use of gene clusters to infer functional coupling.** *Proc Natl Acad Sci USA* 1999, **96**(6):2896-2901.
23. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO: **Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles.** *Proc Natl Acad Sci USA* 1999, **96**(8):4285-4288.
24. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**(5338):631-637.
25. Yanai I, DeLisi C: **The society of genes: Networks of functional links between genes from comparative genomics.** *Genome Biol* 2002, **3**(11):research0064.
26. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, Copley R, Courcelle E, Das U, Durbin R, Fleischmann W, Gough J, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lonsdale D, Lopez R, Letunic I, Mad-

- era M, Maslen J, McDowall J, Mitchell A, Nikolskaya AN, Orchard S, Pagni M, Ponting CP, Quevillon E, Selengut J, Sigrist CJ, Silventoinen V, Studholme DJ, Vaughan R, Wu CH: **InterPro, progress and status in 2005**. *Nucleic Acids Res* 2005, **33(Database Issue):**D201-205.
27. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A: **Pfam: Clans, web tools and services**. *Nucleic Acids Res* 2006, **34(Database issue):**D247-251.
28. Haft DH, Selengut JD, White O: **The TIGRFAMs database of protein families**. *Nucl Acids Res* 2003, **31(1):**371-373.
29. **ABCISSE: Database of ABC systems** [<http://www.pasteur.fr/recherche/unites/pmtg/abc/database.iphtml>]
30. Aravind L, Koonin EV: **The STAS domain - A link between anion transporters and antisigma-factor antagonists**. *Curr Biol* 2000, **10(2):**R53-5.
31. **Pfam** [<http://www.sanger.ac.uk/Software/Pfam/>]
32. Koh YS, Roe JH: **Isolation of a novel paraquat-inducible (pqi) gene regulated by the soxRS locus in Escherichia coli**. *J Bacteriol* 1995, **177(10):**2673-2678.
33. Ames GF, Mimura CS, Shyamala V: **Bacterial periplasmic permeases belong to a family of transport proteins operating from Escherichia coli to humans: Traffic ATPases**. *FEMS Microbiol Rev* 1990, **6(4):**429-446.
34. Honore N, Bergh S, Chanteau S, Doucet-Populaire F, Eiglmeier K, Garnier T, Georges C, Launois P, Limpiboon T, Newton S, Niang K, del Portillo P, Ramesh GR, Reddi P, Ridel PR, Sittisombut N, Wu-Hunter S, Cole ST: **Nucleotide sequence of the first cosmid from the Mycobacterium leprae genome project: Structure and function of the Rif-Str regions**. *Mol Microbiol* 1993, **7(2):**207-214.
35. Zumarraga M, Bigi F, Alito A, Romano MI, Cataldi A: **A 12.7 kb fragment of the Mycobacterium tuberculosis genome is not present in Mycobacterium bovis**. *Microbiology* 1999, **145(4):**893-897.
36. D'Souza SE, Ginsberg MH, Plow EF: **Arginyl-glycyl-aspartic acid (RGD): A cell adhesion motif**. *Trends Biochem Sci* 1991, **16(7):**246-250.
37. Simoes I, Mueller EC, Otto A, Bur D, Cheung AY, Faro C, Pires E: **Molecular analysis of the interaction between cardosin A and phospholipase Da: Identification of RGD/KGE sequences as binding motifs for C2 domains**. *FEBS Journal* 2005, **272(22):**5786-5798.
38. **SignalP Server version 3.0** [<http://www.cbs.dtu.dk/services/SignalP/>]
39. **TatP Server version 1.0** [<http://www.cbs.dtu.dk/services/TatP/>]
40. Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV: **Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context**. *Genome Res* 2001, **11(3):**356-372.
41. Dassa E, Hofnung M: **Sequence of gene malG in E. coli K12: Homologies between integral membrane components from binding protein-dependent transport systems**. *EMBO J* 1985, **4(9):**2287-2293.
42. Saurin W, Koster W, Dassa E: **Bacterial binding protein-dependent permeases: Characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins**. *Mol Microbiol* 1994, **12(6):**993-1004.
43. Mourez M, Hofnung M, Dassa E: **Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits**. *EMBO J* 1997, **16(11):**3066-3077.
44. **TubercuList** [<http://genolist.pasteur.fr/TubercuList/>]
45. Casali N, White AM, Riley LW: **Regulation of the Mycobacterium tuberculosis mceI operon**. *J Bacteriol* 2006, **188(2):**441-449.
46. Bishop RE: **The bacterial lipocalins**. *Biochim Biophys Acta* 2000, **1482(1-2):**73-83.
47. Rizo J, Sudhof TC: **C2-domains: Structure and function of a universal Ca<sup>2+</sup>-binding domain**. *J Biol Chem* 1998, **273(26):**15879-15882.
48. Lupas A: **Coiled coils: New structures and new functions**. *Trends Biochem Sci* 1996, **21(10):**375-382.
49. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG: **Massive gene decay in the leprosy bacillus**. *Nature* 2001, **409(6823):**1007-1011.
50. Daley DO, Rapp M, Granseth E, Melen K, Drew D, von Heijne G: **Global topology analysis of the Escherichia coli inner membrane proteome**. *Science* 2005, **308(5726):**1321-1323.
51. Pajon R, Yero D, Lage A, Llanes A, Borroto CJ: **Computational identification of b-barrel outer-membrane proteins in Mycobacterium tuberculosis predicted proteomes as putative vaccine candidates**. *Tuberculosis* 2006, **86(3-4):**290-302.
52. Chitale S, Ehart S, Kawamura I, Fujimura T, Shimono N, Anand N, Lu S, Cohen-Gould L, Riley LW: **Recombinant Mycobacterium tuberculosis protein associated with mammalian cell entry**. *Cell Microbiol* 2001, **3(4):**247-254.
53. Kim K, Lee S, Lee K, Lim D: **Isolation and characterization of toluene-sensitive mutants from the toluene-resistant bacterium Pseudomonas putida GM73**. *J Bacteriol* 1998, **180(14):**3692-3696.
54. Ramos JL, Duque E, Rodriguez-Herva JJ, Godoy P, Haidour A, Reyes F, Fernandez-Barrero A: **Mechanisms for solvent tolerance in bacteria**. *J Biol Chem* 1997, **272(7):**3887-3890.
55. Hong M, Gleason Y, Wvckoff EE, Payne SM: **Identification of two Shigella flexneri chromosomal loci involved in intercellular spreading**. *Infect Immun* 1998, **66(10):**4700-4710.
56. Carvalho AC, Ruiz-Palacios GM, Ramos-Cervantes P, Cervantes LE, Jiang X, Pickering LK: **Molecular characterization of invasive and noninvasive Campylobacter jejuni and Campylobacter coli isolates**. *J Clin Microbiol* 2001, **39(4):**1353-1359.
57. Rozynek E, Dzierzanowska-Fangrat K, Jozwiak P, Popowski J, Korsak D, Dzierzanowska D: **Prevalence of potential virulence markers in Polish Campylobacter jejuni and Campylobacter coli isolates obtained from hospitalized children and from chicken carcasses**. *J Med Microbiol* 2005, **54(7):**615-619.
58. Muller J, Schulze F, Muller W, Hanel I: **PCR detection of virulence-associated genes in Campylobacter jejuni strains with differential ability to invade Caco-2 cells and to colonize the chick gut**. *Veterinary Microbiology* 2006, **113(1-2):**123-129.
59. Al-Mahmeed A, Senok AC, Ismael AY, Bindayna KM, Tabbara KS, Botta GA: **Clinical relevance of virulence genes in Campylobacter jejuni isolates in Bahrain**. *J Med Microbiol* 2006, **55(7):**839-843.
60. Pagliarulo C, Salvatore P, De Vitis LR, Colicchio R, Monaco C, Tredici M, Tala A, Bardaro M, Lavitola A, Bruni CB, Alifano P: **Regulation and differential expression of gdhA encoding NADP-specific glutamate dehydrogenase in Neisseria meningitidis clinical isolates**. *Mol Microbiol* 2004, **51(6):**1757-1772.
61. Flesselles B, Anand NN, Remani J, Loosmore SM, Klein MH: **Disruption of the mycobacterial cell entry gene of Mycobacterium bovis BCG results in a mutant that exhibits a reduced invasiveness for epithelial cells**. *FEMS Microbiol Lett* 1999, **177(2):**237-242.
62. Meidanis J, Braga MD, Verjovski-Almeida S: **Whole-genome analysis of transporters in the plant pathogen Xylella fastidiosa**. *Microbiol Mol Biol Rev* 2002, **66(2):**272-299.
63. Monaco C, Tala A, Spinosa MR, Progidia C, De Nitto E, Gaballo A, Bruni CB, Bucci C, Alifano P: **Identification of a meningococcal L-glutamate ABC transporter operon essential for growth in low-sodium environments**. *Infect Immun* 2006, **74(3):**1725-1740.
64. Garcia O, Bouige P, Forestier C, Dassa E: **Inventory and comparative analysis of rice and Arabidopsis ATP-binding cassette (ABC) systems**. *J Mol Biol* 2004, **343(1):**249-265.
65. Xu C, Fan J, Riekhof W, Froehlich JE, Benning C: **A permease-like protein involved in ER to thylakoid lipid transfer in Arabidopsis**. *EMBO J* 2003, **22(10):**2370-2379.
66. Xu C, Fan J, Froehlich JE, Awai K, Benning C: **Mutation of the TGD1 chloroplast envelope protein affects phosphatidate metabolism in Arabidopsis**. *Plant Cell* 2005, **17(11):**3094-3110.
67. Awai K, Xu C, Tamot B, Benning C: **A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking**. *Proc Natl Acad Sci USA* 2006, **103(28):**10817-10822.

68. McKinney JD, zu Bentrup KH, Munoz-Elias EJ, Miczak A, Chen B, Chan WT, Swenson D, Sacchettini JC, Jacobs WR, Russell DG: **Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase.** *Nature* 2000, **406(6797)**:735-738.
69. Manganelli R, Voskuil MI, Schoolnik GK, Smith I: **The Mycobacterium tuberculosis ECF sigma factor  $\sigma$ E: Role in global gene expression and survival in macrophages.** *Mol Microbiol* 2001, **41(2)**:423-437.
70. Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK: **Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding a-crystallin.** *Proc Natl Acad Sci USA* 2001, **98(13)**:7534-7539.
71. Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K: **Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling.** *Mol Microbiol* 2002, **43(3)**:717-731.
72. Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK: **Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: Insights into the phagosomal environment.** *J Exp Med* 2003, **198(5)**:693-704.
73. Santangelo MP, Goldstein J, Alito A, Gioffre A, Caimi K, Zabal O, Zumarraga M, Romano MI, Cataldi AA, Bigi F: **Negative transcriptional regulation of the mce3 operon in Mycobacterium tuberculosis.** *Microbiology* 2002, **148(Pt 10)**:2997-3006.
74. Miller CD, Hall K, Liang YN, Nieman K, Sorensen D, Issa B, Anderson AJ, Sims RC: **Isolation and characterization of polycyclic aromatic hydrocarbon-degrading Mycobacterium isolates from soil.** *Microb Ecol* 2004, **48(2)**:230-238.
75. Embley TM, Stackebrandt E: **The molecular phylogeny and systematics of the actinomycetes.** *Annu Rev Microbiol* 1994, **48**:257-289.
76. Kroppenstedt RM: **Fatty acid and menaquinone analysis of actinomycetes and related organisms.** *Soc Appl Bacteriol Tech Ser* 1985, **20**:173-199.
77. UniProt: **The Universal Protein Resource** [<http://www.ebi.uniprot.org/index.shtml>]
78. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B: **The Universal Protein Resource (UniProt): An expanding universe of protein information.** *Nucleic Acids Res* 2006, **34(Database issue)**:D187-191.
79. **The Institute for Genomic Research: Comprehensive Microbial Resource** [<http://cmr.tigr.org/>]
80. Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O: **The Comprehensive Microbial Resource.** *Nucl Acids Res* 2001, **29(1)**:123-125.
81. **National Center for Biotechnology Information: Microbial Genome Project** [<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>]
82. **Joint Genome Institute: Microbial Genomics** [<http://genome.jgi-psf.org/>]
83. **InterPro** [<http://www.ebi.ac.uk/interpro/>]
84. **PROSITE: Database of protein families and domains** [<http://www.expasy.org/prosite/>]
85. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJA: **The PROSITE database.** *Nucleic Acids Res* 2006, **34(Database issue)**:D227-230.
86. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
87. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17)**:3389-3402.
88. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5(2)**:150-163.
89. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22)**:4673-4680.
90. Felsenstein J: **PHYLIP (Phylogeny Inference Package) version 3.6.** Department of Genome Sciences, University of Washington, Seattle ; 2005.
91. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25(24)**:4876-4882.
92. **MEME: Multiple Em for Motif Elicitation** [<http://meme.sdsc.edu/meme/>]
93. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers: Menlo Park, California.** AAAI Press; 1994:28-36.
94. **WebLogo** [<http://weblogo.berkeley.edu/>]
95. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: A sequence logo generator.** *Genome Res* 2004, **14(6)**:1188-1190.
96. **JPred: A method for protein secondary structure prediction** [<http://www.compbio.dundee.ac.uk/~www-jpred/>]
97. Cuff JA, Barton GJ: **Application of multiple sequence alignment profiles to improve protein secondary structure prediction.** *Proteins* 2000, **40(3)**:502-511.
98. Lupas A, Van Dyke M, Stock J: **Predicting coiled coils from protein sequences.** *Science* 1991, **252(5010)**:1162-1164.
99. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340(4)**:783-795.
100. Nilsson J, Persson B, von Heijne G: **Consensus predictions of membrane protein topology.** *FEBS Letters* 2000, **486(3)**:267-269.
101. Sonnhammer EL, von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences.** *Proc Int Conf Intell Syst Mol Biol* 1998, **6**:175-182.
102. **TMHMM Server version 2.0** [<http://www.cbs.dtu.dk/services/TMHMM/>]
103. **HMMTOP: Prediction of transmembrane helices and topology of proteins** [<http://www.enzim.hu/hmmtop/index.html>]
104. Tusnady GE, Simon I: **The HMMTOP transmembrane topology prediction server.** *Bioinformatics* 2001, **17(9)**:849-850.
105. **TopPred: Topology prediction of membrane proteins** [<http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html>]
106. Claros MG, von Heijne G: **TopPred II: An improved software for membrane protein structure predictions.** *Comput Appl Biosci* 1994, **10(6)**:685-686.
107. Garrity GM, Bell JA, Lilburn TG: **Bergey's Manual of Systematic Bacteriology.** 2nd edition. New York, Springer; 2004.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

