Multiple Small RNAs Identified in Mycobacterium bovis BCG are Also Expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis

Share Your Story

Multiple small RNAs identified in *Mycobacterium bovis* BCG are also expressed in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*

Jeanne M. DiChiara[1], Lydia M. Contreras-Martinez[1], Jonathan Livny[2,3], Dorie Smith[1], Kathleen A. McDonough[1,4] and Marlene Belfort[1,4,*]

[1]Wadsworth Center, New York State Department of Health, PO Box 22002, Albany, NY 12201-2002, [2]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, [3]Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115 and [4]School of Public Health, State University of New York at Albany, Albany, NY 12201-0509

ABSTRACT

Tuberculosis (TB) is a major global health problem, infecting millions of people each year. The causative agent of TB, *Mycobacterium tuberculosis*, is one of the world's most ancient and successful pathogens. However, until recently, no work on small regulatory RNAs had been performed in this organism. Regulatory RNAs are found in all three domains of life, and have already been shown to regulate virulence in well-known pathogens, such as *Staphylococcus aureus* and *Vibrio cholera*. Here we report the discovery of 34 novel small RNAs (sRNAs) in the TB-complex *M. bovis* BCG, using a combination of experimental and computational approaches. Putative homologues of many of these sRNAs were also identified in *M. tuberculosis* and/or *M. smegmatis*. Those sRNAs that are also expressed in the non-pathogenic *M. smegmatis* could be functioning to regulate conserved cellular functions. In contrast, those sRNAs identified specifically in *M. tuberculosis* could be functioning in mediation of virulence, thus rendering them potential targets for novel antimycobacterials. Various features and regulatory aspects of some of these sRNAs are discussed.

INTRODUCTION

Mycobacterium tuberculosis (Mtb), the causative agent of tuberculosis (TB), is one of the world's most succesful pathogens. Treatment of TB has become increasingly difficult, due to the emergence of multiple drug resistant Mtb strains (1–5), and thus the development of new and more effective treatments for TB is imperative. One area of TB research that has only recently begun to garner attention is that of small noncoding RNAs (sRNAs) and their possible roles in virulence (6).

sRNAs are generally small and untranslated; they can originate from their own independent genes or through the processing of larger transcripts (7). They have been recognized in recent years as a major class of gene regulators in bacteria. These regulatory transcripts have been found in all three domains of life, including a diverse set of bacteria (8). sRNAs allow bacteria to respond quickly to their environment, by causing global changes in gene expression. This is particularly important for pathogenic bacteria, which regulate their virulence in response to rapidly shifting conditions and external signals in the host environment, such as temperature and pH (9). sRNA-mediated regulation has been shown to play a central role in the virulence of several bacteria. These pathogens include *Clamydia trachomatis*, *Clostridium perfringens*, *Pseudomonas aeruginosa*, *Salmonella typhimurium*, *Staphylococcus aureus*, *Streptococcus pyogenes*, *Vibrio cholerae* and *Yersinia pestis*, the causative agents of sexually transmitted genital infections, food poisoning, burn and wound infections, gastroenteritis, various respiratory and skin diseases, scarlet fever, cholera and plague, respectively (10,11).

While sRNA-mediated gene regulation has been demonstrated in both Gram-negative (G⁻) and Gram-positive (G⁺) bacteria, diffferences in the mode of sRNA action are likely to exist between G⁻ and G⁺ species. Hfq, an RNA chaperone, plays a central role in sRNA–mRNA interactions in G⁻ bacteria. However, this

does not appear to occur in all G^+ bacteria; in some cases there is no known Hfq homolog, including in the Actinomycetes-Deinococcus-Cyanobacteria clade (12,13), of which Mtb is a member. In G^+ bacteria that do have an Hfq homolog, such as *S. aureus*, the protein is not required for sRNA activity (14). Recently, evidence of a role for Hfq in sRNA antisense regulation was shown in *Listeria monocytogenes*, a low-GC G^+ bacterium (15). However, despite this breakthough, the majority of sRNAs in *L. monocytogenes* do not seem to require Hfq for stability or target interactions *in vitro*. It is therefore possible that some G^+ bacteria, including mycobacteria, with a ~65% GC content, utilize an RNA chaperone either related to or distinct from Hfq (13), or use a different mechanism for sRNA–mRNA interactions.

To begin to explore the role of sRNAs in Mtb and other G^+ bacteria, and to identify sRNAs that will provide insight into Mtb virulence, we undertook a search for sRNAs in mycobacteria. We employed a two-pronged approach that utilized both a cloning-based screen and a computational search, to identify sRNAs in the TB-complex bacterium *Mycobacterium bovis* BCG (BCG). While cloning-based approaches have proven successful in identifying novel sRNAs in various organisms, they present several limitations. First, they are ineffective in identifying non-abundant sRNAs or sRNAs expressed under specific growth conditions that differ from those used for RNA isolation (16,17). Second, these approaches are time intensive, relatively expensive and can present technical challenges. Several computational algorithms have recently been developed to identify putative sRNAs, based on the presence of a predicted Rho-independent terminator downstream of a conserved intergenic region (18-20). These algorithms have been effective in identifying sRNAs in diverse species. Moreover, computational searches can be completed quickly, are inexpensive and enable identification of putative sRNAs independent of their relative abundance or condition-dependent expression. However, unlike cloning-based approaches, these alogorithms cannot be used to identify non-canonical sRNAs, such as those not associated with a Rho-independent terminator, those located within open reading frames (ORFs) or in mis-annotated ORFs, or those not conserved in closely related species. This last limitation is particularly problematic for our analysis, as an sRNA that has recently emerged in a pathogenic species of Mycobacterium to help mediate virulence or growth in a particular host niche might not be well conserved, and thus would be missed in a bioinformatic screen.

We therefore reasoned that utilization of combined experimental and computational approaches in our analysis would yield the most comprehensive annotation for sRNAs in BCG. Indeed, using these two approaches, we identified 37 sRNAs, very few of which were identified in both screens. Thus, the present work not only identifies sRNAs in BCG, but also sheds light on the evolutionary conservation of these sRNAs: some of whose expression is restricted to TB-complex bacteria (BCG and Mtb), while others are expressed both in TB-complex and non-TB-complex mycobacteria, including the non-pathogenic *Mycobacterium smegmatis*.

MATERIALS AND METHODS

Strains and plasmids

Recombinant *M. bovis* BCG (Pasteur strain, Trudeau Institute) and Mtb H37Rv were grown in mycomedium (Middlebrook 7H9 medium supplemented with 0.5% glycerol, 10% oleic acid-albumin-dextrose-catalase (OADC) and 0.05% Tween-80) as described previously (21). Cultures were grown for the specified number of days in either 25 or $50\,cm^2$ tissue culture flasks, either shaking or standing. Low-oxygen (1.3% O_2 + 5% CO_2) and low-pH cultures were grown as described previously (21–23). *M. smegmatis* MC^2155 was grown shaking at $37^\circ C$, in trypticase soy broth supplemented with 0.05% Tween-80.

Total RNA isolation

Cells were pelleted at $4^\circ C$ and then either stored at $-80^\circ C$ or directly used for total RNA preparation. All centrifugation was performed at $4^\circ C$. The cell pellets were resuspended in 1 ml TRIzol reagent (Invitrogen), transferred to screw cap tubes containing 0.1 mm diameter zirconia beads (BioSpec Products) and incubated at $25^\circ C$ for 5 min. The cells were then lysed using a mini-beadbeater, with two 100-s pulses. The cells were kept on ice for 10 min between the two 100-s treatments. The beads and cellular debris were then spun out at $4^\circ C$ for 2 min. The supernatant was transferred to a clean, siliconized microfuge tube, 300 µl of a chloroform:isoamyl alcohol mix (v/v 24:1) was added, the samples were vortexed for 15 s, then incubated at $25^\circ C$ for 3 min. The tubes were then spun at 14 000 r.p.m. for 10 min, the aqueous phase transferred to a siliconized 1.5 ml tube, and 270 µl isopropanol and 270 µl of a sodium citrate and sodium chloride mix (0.8 and 1.2 M, respectively) was added to the tubes. The samples were mixed well, and then incubated on ice for 10 min. The RNA was sedimented by centrifugation at 14 000 r.p.m. for 15 min. The pellet was washed with 1 ml 95% ethanol and centrifuged for 5 min. The pelleted RNA was allowed to air-dry for ~5 min, and was then resuspended in 30 µl RNase-free water (Ambion); like samples were then combined. RNA concentration was measured by spectrophotometry. Samples were stored at $-20^\circ C$. Total RNA was prepared from *M. smegmatis* MC^2155 in the same manner from either log phase ($OD_{600} = 0.6$) or stationary phase cultures ($OD_{600} \geq 1$). Mtb cultures were treated with 500 µl 5 M guanidinium thiocyanate (GTC) prior to pelleting.

sRNA cloning

The sRNA cloning was performed using the miRCat^{TM} microRNA Cloning Kit (IDT), with slight modifications. Total RNA (~100 µg) was separated on a 10% denaturing polyacrylamide gel (Sequagel, National Diagnostics) and RNA in the ranges of <80 nt, between 100 and 200 nt,

between 200 and 300 nt and >300 nt were gel extracted using the DTR column method as described in the miRCat™ technical manual. The 3′ cloning linker was ligated to the gel-extracted RNA using T4 RNA ligase. This cloning linker is preactivated for ligation due to a 5′ adenylation (rApp), which allows the ligation to occur in the absence of ATP (24), thus eliminating the need to dephosphorylate the RNA prior to the first ligation, which is otherwise necessary to prevent RNA circularization. The 3′ cloning linker also contains a blocking group at its 3′ end (ddC) to minimize multimerization of the oligonucleotide. Following gel extraction, as above, of the 3′-linkered RNA, the 5′-linker was then ligated to the gel extraced RNA, again using T4 RNA ligase. The primer complementary to the 3′ linker sequence was then used to reverse transcribe the RNA into cDNA, utilizing SuperScript III Reverse Transcriptase (Invitrogen). The cDNA produced was amplified by PCR, using Fermentas PCR Master Mix, with the PCR cycling as described in the miRCat™ technical manual. The PCR products were then pooled, extracted with phenol:chloroform:isoamyl alcohol (v/v 25:24:1), and ethanol precipitated, followed by agarose gel extraction to remove any ligation and PCR artifacts. The gel-extracted cDNA library was subsequently reamplified by another round of PCR. The resulting cDNA product was then either cloned directly into the pCR®II TOPO vector (Invitrogen) or concatemerized, as described in the miRCat™ technical manual, and then cloned. Colonies resulting from chemical transformation of TOP10 cells (Invitrogen) were screened for inserts both by blue/white screening and by colony PCR, using the M13 Forward (−20) Primer (IDT) and M13 Reverse Primer (IDT) to the pCR®II vector. While the blue/white screen facilitated the detection of cloned inserts, any biologically active and/or potentially toxic cloned BCG fragments would be counterselected for, and thus missed in our screen. Plasmids containing cloned inserts based on the colony PCR screen were then sequenced. The resulting sequences were used to perform BLAST searches to the nucleotide collection on NCBI.

Northern blot analysis

Northern blots were used to verify expression of both the potential sRNA sequences determined by the linker cloning and the computationally predicted candidates. DNA oligonucleotide probes specific for each candidate sRNA (Supplementary Table S3) were end-labeled using 20 pmoles of oligonucleotide in a 20-µl kinase reaction containing 25 µM γ-P^{32} ATP and 20 units T4 polynucleotide kinase (NEB) at 37°C for 1 h. Markers were labeled in the same manner.

Total RNA (~5 µg, except for 10 µg for Mcr7 and Mcr9) was separated on a 10% denaturing polyacrylamide gel alongside either labeled 1 kb Plus DNA ladder (Invitrogen) or labeled ΦX174 DNA/*Hinf* I ladder (Promega), which was transferred to a positively charged membrane (Hybond N+, GE Life Sciences) for blotting. Hybridization was performed using Amersham Rapid-hyb buffer (GE Healthcare), following their recommended

protocol for oligonucleotide probes, with a 3-h incubation, and moderately stringent conditions, as described in Supplementary materials and methods. Membranes were exposed to a phosphor screen overnight and visualized with a phosphorimager (Typhoon 9400 Variable Mode Imager, Amersham Biosciences). Quantitation was performed using ImageQuant® software, Version 5.2 (Molecular® Dynamics). Statistical analysis was performed using InStat® software, Version 3.0b (GraphPad).

Phylogenetic seletion of computationally predicted sRNA candidates

Candidate sRNAs were predicted using the SIPHT program, as described previously (20). Briefly, SIPHT identifies candidate regulatory RNA-encoding based on the presence of intergenic sequence conservation upstream of a putative Rho-independent terminator. These loci are then annotated for several features, including their position relative to their flanking ORFs and whether they share conserved primary sequence or synteny with previously annotated regulatory RNAs.

The database nucleotide collection (nr/nt) was used with the preset parameters for blastn searches, with the computationally predicted candidate sRNA sequence. Only results with E values <1 were used in this study, unless otherwise indicated. The query search was limited to *Mycobacterium* (taxid1763) only if results from using the nr/nt collection did not include *M. smegmatis*.

Co-transcription experiments

Total RNA, prepared as above, was treated with 2 U of TURBO™ DNase (Ambion) according to the manufacturer's protocol, followed by extraction with phenol: chloroform:isoamyl alcohol (v/v 25:24:1), and ethanol precipitation. The DNase treatment was then repeated, and the final RNA pellet was resuspended and the RNA was quantitated as above. Approximately 200 ng of RNA was used for first strand cDNA synthesis, at 45°C, using the RevertAid™ First Strand Synthesis Kit (Fermentas), with primers designed to amplify the junction between the sRNA and adjacent ORF (Supplementary Table S3). The cDNA produced was amplified by PCR, using Fermentas PCR Master Mix (10 µl RT reaction in a 50 µl PCR reaction). The PCR cycling was as follows: 95°C/3 min, then 95°C/30 s, 60°C or 56°C/30 s (Mcr, Mpr annealing temperatures, respectively), 72°C/45 s for 30 cycles, followed by a 9 min 15 s final extension at 72°C. The PCR products were then visualized on 1 or 1.5% agarose gels.

RESULTS

sRNAs identified by cloning

Several independent cDNA libraries were constructed using total BCG RNA from cells grown at 37°C with shaking for 7 days (late log phase). From these libraries, 116 clones with inserts were obtained and sequenced (Figure 1). Of these 116 candidates, 56 were eliminated: 26
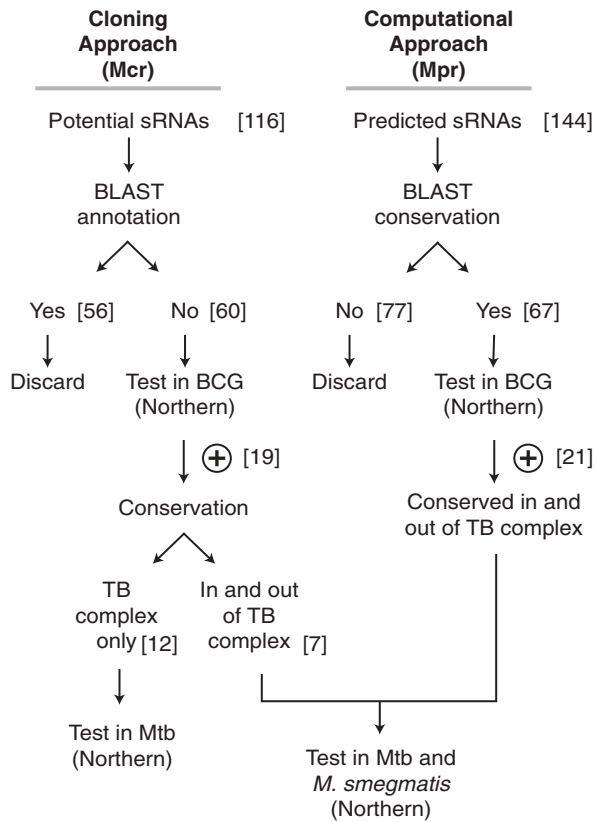
Figure 1. Strategy for sRNA discovery. This schematic shows the combination of cloning and computational approaches used to identify sRNAs in BCG, Mtb and *M. smegmatis.*

contained rRNA fragments, 22 contained tRNAs, 4 contained the 10Sa RNA, 3 contained *rnpB* fragments (RNA component of RnaseP) and 1 contained the rRNA internal transcribed spacer (ITS1) sequence (25,26). Of the remaining 60 candidates, 13 contained fragments located in intergenic regions of the BCG genome and 47 contained sequences located within annotated ORFs, as determined by megablast (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi). Some candidate sRNAs were cloned more than once (Supplementary Table S1), suggesting that these sRNAs may be more abundant or easier to clone than others.

All 60 of the sRNA candidates were tested for expression using northern blot hybridization. The Mycobacterium cloned RNAs (Mcr) verified by northern blot are designated Mcr1–Mcr19 (Figure 2). Each strand of the potential sRNA was probed independently, to ensure both that the sRNA was expressed and that expression was only observed from one strand. For candidates located within ORFs this was particularly important, to establish whether a positive Northern blot signal resulted from a *bona fide* sRNA, rather than an mRNA degradation product. We have defined the 'class' of sRNA based on several criteria, including its position relative to adjacent ORFs, evidence of processing versus an independent transcript, and its direction relative to any known or predicted overlapping transcripts. sRNAs that are >60 bp

downstream of the closest 3′ ORF and >100 bp upstream from the closest 5′ ORF are designated as intergenic, and are presumed to act *in trans* to a distal target. sRNAs that are cotranscribed with the upstream ORF and/or the downstream ORF are designated as 3′ or 5′UTRs, respectively. Finally, RNAs that are <60 bp away from the 3′ end of an opposing ORF are potentially antisense to a 3′UTR, while sRNAs <100 bp away from the 5′-end of an opposing ORF are potentially antisense to a 5′UTR. The northern analysis indicated that many of the sRNAs were not full length (Supplementary Table S1). For example, only 46 nt of the Mcr4 sRNA were cloned, but the northern blot signal corresponded to an >200 nt species (Figure 2A and B). Where possible, we therefore mapped the 5′-end utilizing 5′RLM-RACE (RNA ligase-mediated rapid amplification of cDNA ends) and/or primer extension (Supplementary Table S1). Additionally, we checked whether there was any co-transcription with neighboring ORFs, using reverse transcription and PCR amplification (RT–PCR). The RT–PCR was designed to amplify the junction between the sRNA and its adjacent ORF(s). Seven sRNAs were co-transcribed with either one or both of their adjacent ORFs, as listed in Figure 2A and shown in Figure 2C.

Of particular interest are the following sRNAs (with genes referred to using their Mtb H37Rv annotations): First, Mcr9 lies upstream of *ilvB1*, which has recently been shown to play a role in Mtb virulence in mice (27). Mcr9 is co-transcribed with *ilvB1* (Figure 2A and C). This RNA most likely is derived from an mRNA leader containing a T-box. The 14-nt conserved sequence of the T-box is part of a tRNA-directed antitermination mechanism, where $tRNA_{leu}$ acts as a direct effector (a riboswitch). This was first described in *Bacillus subtilis*, but is also present in other G^+ bacteria (28–30). Second, Mcr11, which is located between an adenylyl cyclase (*Rv1264*) and the cyclic AMP-induced gene *Rv1265* (23), whose function is unknown. Finally, there are three potentially interesting sRNAs located within ORFs: Mcr16, located in *fabD*, which plays a role in fatty acid synthesis in Mtb (31); Mcr18, located in *nuoC* (*Rv3147*), which is involved in respiration (32,33); and sRNA Mcr19, located in the gene for the transcriptional regulatory protein Rv0485. This protein is required for virulence in Mtb (34). Two of the intergenic sRNAs identified, Mcr6 and Mcr14, were previously identified in Mtb, as sRNAs C8 (4.5S RNA) and F6, respectively (6).

sRNAs identified in silico

We reasoned that predicted loci that are conserved among species that are more distantly related represented strong candidates for functional transcripts. We therefore initially used the SIPHT program (20) to identify 144 candidate sRNAs in BCG (Refseq: NC_008769). We then applied an additional 'filter' using BLAST, and compiled a list of 67 BCG candidate sRNAs that were partially conserved in mycobacterial species outside of the TB-complex (Figure 1; Supplementary Table S2). All potential candidates were tested using multiple probes (Supplementary Table 3), with RNA prepared
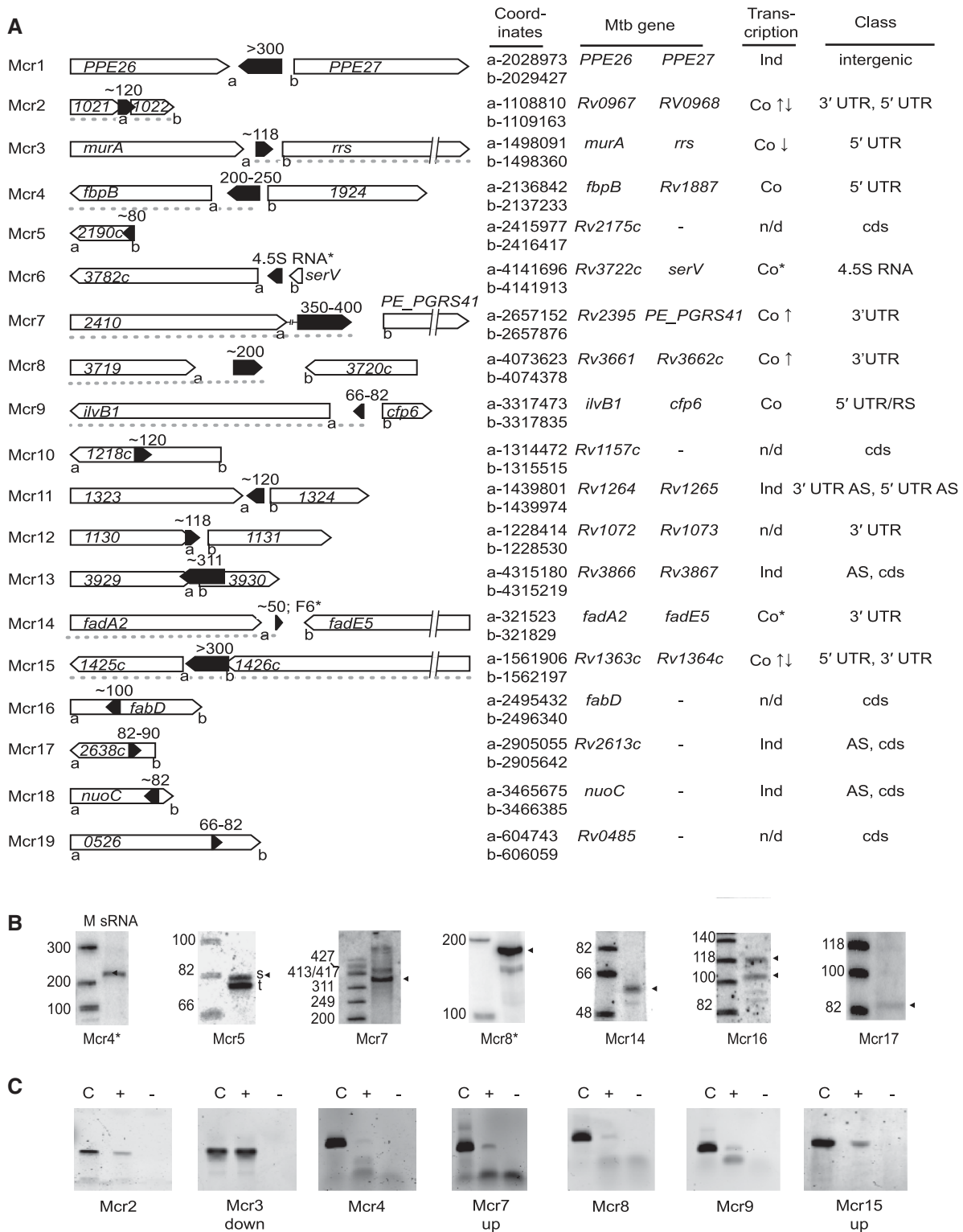
A

	Coordinates	Mtb gene		Transcription	Class
Mcr1 *PPE26* / >300 / *PPE27*	a-2028973 b-2029427	*PPE26*	*PPE27*	Ind	intergenic
Mcr2 *1021* ~120 *1022*	a-1108810 b-1109163	*Rv0967*	*RV0968*	Co ↑↓	3′ UTR, 5′ UTR
Mcr3 *murA* ~118 *rrs*	a-1498091 b-1498360	*murA*	*rrs*	Co ↓	5′ UTR
Mcr4 *fbpB* 200-250 *1924*	a-2136842 b-2137233	*fbpB*	*Rv1887*	Co	5′ UTR
Mcr5 *2190c* ~80	a-2415977 b-2416417	*Rv2175c*	-	n/d	cds
Mcr6 *3782c* 4.5S RNA* *serV*	a-4141696 b-4141913	*Rv3722c*	*serV*	Co*	4.5S RNA
Mcr7 *2410* 350-400 *PE_PGRS41*	a-2657152 b-2657876	*Rv2395*	*PE_PGRS41*	Co ↑	3′UTR
Mcr8 *3719* ~200 *3720c*	a-4073623 b-4074378	*Rv3661*	*Rv3662c*	Co ↑	3′UTR
Mcr9 *ilvB1* 66-82 *cfp6*	a-3317473 b-3317835	*ilvB1*	*cfp6*	Co	5′ UTR/RS
Mcr10 *1218c* ~120	a-1314472 b-1315515	*Rv1157c*	-	n/d	cds
Mcr11 *1323* ~120 *1324*	a-1439801 b-1439974	*Rv1264*	*Rv1265*	Ind	3′ UTR AS, 5′ UTR AS
Mcr12 *1130* ~118 *1131*	a-1228414 b-1228530	*Rv1072*	*Rv1073*	n/d	3′ UTR
Mcr13 *3929* ~311 *3930*	a-4315180 b-4315219	*Rv3866*	*Rv3867*	Ind	AS, cds
Mcr14 *fadA2* ~50; F6* *fadE5*	a-321523 b-321829	*fadA2*	*fadE5*	Co*	3′ UTR
Mcr15 *1425c* >300 *1426c*	a-1561906 b-1562197	*Rv1363c*	*Rv1364c*	Co ↑↓	5′ UTR, 3′ UTR
Mcr16 *fabD* ~100	a-2495432 b-2496340	*fabD*	-	n/d	cds
Mcr17 *2638c* 82-90	a-2905055 b-2905642	*Rv2613c*	-	Ind	AS, cds
Mcr18 *nuoC* ~82	a-3465675 b-3466385	*nuoC*	-	Ind	AS, cds
Mcr19 *0526* 66-82	a-604743 b-606059	*Rv0485*	-	n/d	cds

B

M sRNA

Mcr4*: 300, 200, 100

Mcr5: 100, 82, 66 (s, t)

Mcr7: 427, 413/417, 311, 249, 200

Mcr8*: 200, 100

Mcr14: 82, 66, 48

Mcr16: 140, 118, 100, 82

Mcr17: 118, 100, 82

C

C + −

Mcr2   Mcr3 down   Mcr4   Mcr7 up   Mcr8   Mcr9   Mcr15 up

Figure 2. sRNAs identified in BCG by cloning. (**A**) Properties of Mcr1-Mcr19. The approximate sRNA size observed by northern blot (predominant band), orientation relative to its flanking genes, corresponding Mtb H37Rv genes, co-transcription information and the type of sRNA are given. All of the gene names and numbers in the arrows are BCG designations. The black arrows represent the sRNA. ↑ = upstream ORF; ↓ = downstream ORF; Ind = independent transcript; Co = co-transcribed; n/d = not determined; intergenic = intergenic/*trans*-acting sRNA; 5′UTR = 5′ untranslated region; 3′UTR = 3′ untranslated region; 5′UTR AS = antisense to potential 5′ UTR; 3′UTR AS = antisense to potential 3′UTR; cds = overlapping coding sequence; RS = riboswitch. *See ref. (6). (**B**) Representative northern blots of cloned sRNAs. In each case the marker lane is on the left, labeled in nucleotides, while the sRNA lane is on the right. For Mcr5, s is the sRNA, while t is a tRNA control. All northerns shown were performed with 7-day shaking BCG RNA run on a 10% denaturing PAGE, except for Mcr4 and Mcr8, which were run on a 6% denaturing PAGE. Probes are given in Supplementary Table S3. (**C**) Representative co-transcription. The seven co-transcribed RNAs are shown, with the chromosomal DNA positive PCR control (C), and RT-PCR reactions spanning the junction of the sRNA and adjacent ORF with (+) or without (−) reverse transcriptase. The RT-PCR results for sRNAs that did not display co-transcription are not shown.

from 6- (mid-log phase) and 8-day (early stationary phase) shaking and standing cultures, as well as 7-day late log phase shaking cultures, in an effort to maximize our chances of detecting expression. All candidates were also tested in both orientations.

Northern blot analysis of the potential sRNAs resulted in the confirmation of 21 sRNAs expressed in BCG, Mpr1–Mpr21 (Mycobacterium predicted RNA; Figure 3). Of the 21 sRNAs identified and confirmed through the use of computational algorithms, only three were also identified by cloning (Mcr3/Mpr7, Mcr9/Mpr14 and Mcr14/Mpr13; Figure 3A). Therefore, the complementary computational approach led to 17 novel sRNA candidates that were not identified using our cloning alone, possibly because a relatively small number of clones were employed [Mpr19 was previously identified in Mtb as sRNA B11 (6)]. Interestingly, the three sRNAs identified by both methods produced strong signals by northern blots, relative to the other sRNAs identified *in silico* (Figure 3 and Supplementary Figure 1), suggesting a potential bias of the cloning-based screen for abundant transcripts.

A comparison of the northern blot results and the computational predictions of the sRNAs revealed discrepancies in their sizes and orientations. Only five sRNAs detected were close to their predicted sizes; most of the candidates were significantly larger or smaller than their estimated sizes. An underestimate of sRNA size could result from only a portion of the sRNA sequence being conserved. In contrast, overestimates of sRNA size could result from sequence conservation that extends beyond the actual sRNA gene. Moreover, for the two sRNAs Mcr14/Mpr13 and Mpr17, two distinct bands were detected for each by northern analysis (Figure 3B), suggesting the presence of multiple, possibly processed, sRNAs. In these cases, only one sRNA was predicted computationally. Furthermore, we found that seven of the 21 sRNAs ran in the opposite orientation to the one predicted. These results likely arise from the GC-richness of the genome, leading to false prediction of a putative Rho-independent terminator associated with a real conserved intergenic locus on the opposite strand.

We again checked for co-transcription with adjacent ORFs. As shown in Figure 3A and C, several of the computationally identified sRNAs were also co-transcribed. Mpr8 is co-transcribed with its downstream ORF, *infC* (Figure 3A; data not shown). Rfam (http://rfam.sanger.ac.uk) shows that Mpr8 is likely a member of the L20 leader family. The L20 leader was shown to control expression of the *infC* operon in *Bacillus subtilis* through transcriptional attenuation (35).

Evolutionary conservation of sRNAs

Out of the 37 total sRNAs identified in BCG, 15 were also expressed in the fast-growing, non-pathogenic bacterium *M. smegmatis* (Figure 4A and B). *M. smegmatis* is related to Mtb and BCG, but is not a member of the TB-complex bacteria (Figure 4D) (36). The majority of sRNAs that are also expressed in *M. smegmatis* were only identified through computational means (12 out of 15), and are

also predicted to be in a wide range of mycobacterial species (Figure 4A and D). From this evolutionary conservation analysis we hypothesize that these sRNAs likely regulate highly conserved cellular functions, but that they may not be involved in regulating pathogenic activities or specific functions exclusive to other mycobacteria (e.g. hydrocarbon degradation).

Many sRNAs in BCG are also expressed in Mtb, as one might expect in closely related bacteria (Figure 4A, B and D). Based on homology with BCG, 17 novel candidates were confirmed in Mtb by northern blot. Of these, eight were solely from the computational predictions, thus underscoring the success of our phylogenetic approach in aiding the identification of novel sRNAs among evolutionarily related organisms.

Mcr11 is differentially expressed

We selected Mcr11 to explore differential expression of sRNAs. This sRNA lies between two genes, *Rv1264* and *Rv1265* (corresponding to BCG1323 and BCG1324 in Figure 2A), that are of particular interest to us, due to their relationship to cAMP metabolism (23). The cyclic AMP-induced gene *Rv1265* is upregulated at low pH in BCG, but not Mtb; *Rv1265* is also upregulated in both BCG and Mtb during macrophage infection (37,38). We therefore tested whether Mcr11 is regulated by conditions Mtb encounters during macrophage infection.

Indeed, Mcr11 showed differential expression in both BCG and Mtb under conditions associated with the host environments of macrophages and granulomas during infection, such as low pH and hypoxia (39,40). Expression of Mcr11 in BCG is ~3-fold higher ($P = 0.02$) in 8-day cultures than 7-day ones, when the cells are transitioning from late log phase into stationary phase (Figure 5A). Eight-day cultures grown under CO_2-supplemented low oxygen conditions would still be in an extended log phase (22), and showed an approximately 3-fold decrease ($P = 0.002$) in Mcr11 expression relative to the 8-day stationary phase cultures, but not the 7-day log phase cultures (Figure 5A). These results indicate that the expression of Mcr11 in BCG is responsive to the growth phase and possibly also to a hypoxic environment. However, expression of Mcr11 in Mtb is only dependent on the growth phase (Figure 5B). As in BCG, expression in Mtb was approximately 2-fold higher in 8-day than 7-day cultures ($P = 0.05$).

DISCUSSION

Importance of sRNAs in mycobacteria

We have identified and described 37 sRNAs in BCG, 34 of which are novel. We have further shown that out of all the sRNAs listed in Figure 4A, eight (21%) were expressed in BCG and Mtb only (TB-complex bacteria), with no expression observed in *M. smegmatis*; three (8%) were expressed in BCG and *M. smegmatis*, with no expression observed in Mtb; 12 (32%) were expressed in all three bacteria, leaving 14 (38%) that were expressed in BCG alone under the conditions tested. All 37 RNAs are predicted to be in Mtb, but only 20 (54%) were expressed
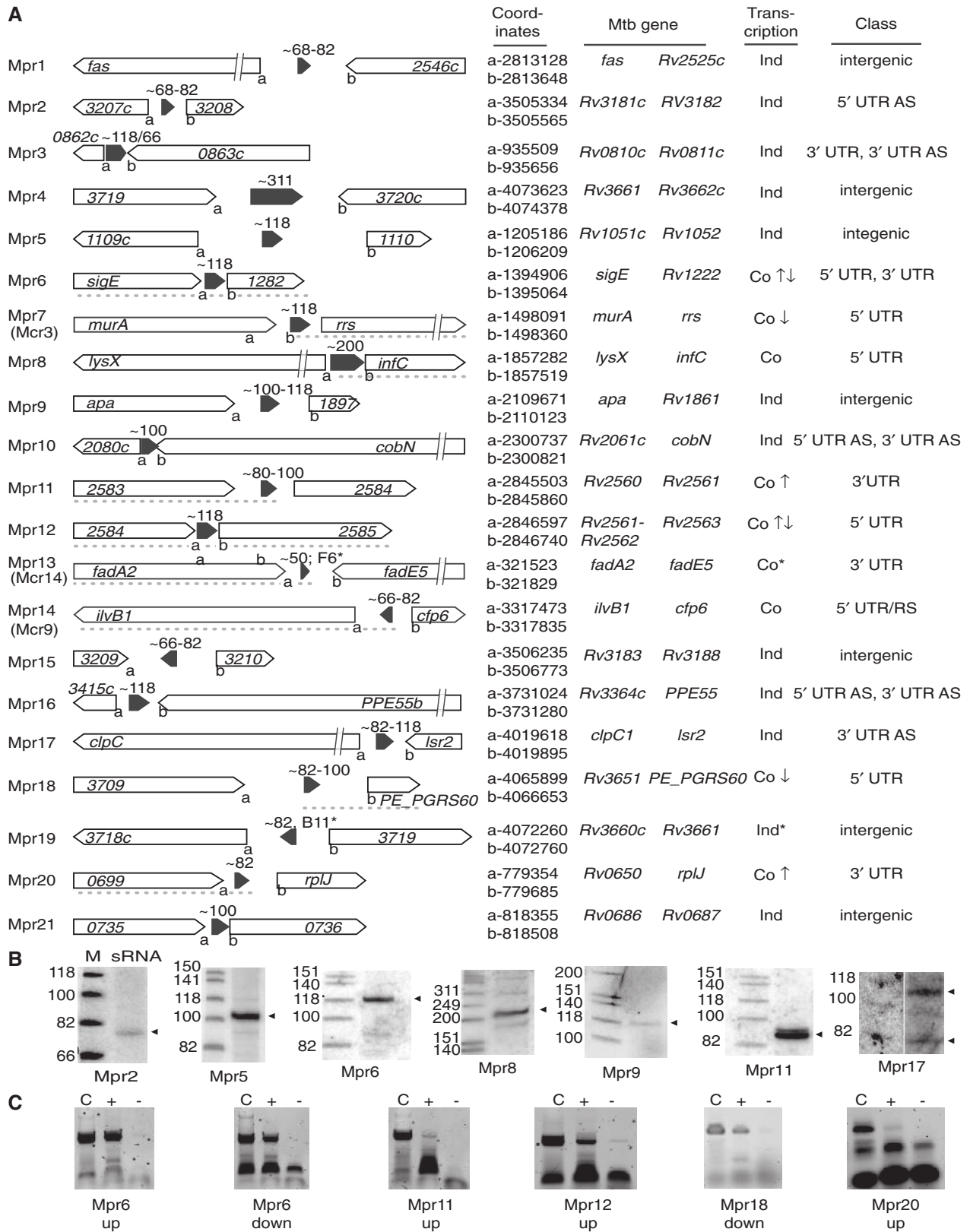
A

		Coord-inates	Mtb gene		Trans-cription	Class
Mpr1		a-2813128 b-2813648	*fas*	*Rv2525c*	Ind	intergenic
Mpr2		a-3505334 b-3505565	*Rv3181c*	*RV3182*	Ind	5′ UTR AS
Mpr3		a-935509 b-935656	*Rv0810c*	*Rv0811c*	Ind	3′ UTR, 3′ UTR AS
Mpr4		a-4073623 b-4074378	*Rv3661*	*Rv3662c*	Ind	intergenic
Mpr5		a-1205186 b-1206209	*Rv1051c*	*Rv1052*	Ind	integenic
Mpr6		a-1394906 b-1395064	*sigE*	*Rv1222*	Co ↑↓	5′ UTR, 3′ UTR
Mpr7 (Mcr3)		a-1498091 b-1498360	*murA*	*rrs*	Co ↓	5′ UTR
Mpr8		a-1857282 b-1857519	*lysX*	*infC*	Co	5′ UTR
Mpr9		a-2109671 b-2110123	*apa*	*Rv1861*	Ind	intergenic
Mpr10		a-2300737 b-2300821	*Rv2061c*	*cobN*	Ind	5′ UTR AS, 3′ UTR AS
Mpr11		a-2845503 b-2845860	*Rv2560*	*Rv2561*	Co ↑	3′UTR
Mpr12		a-2846597 b-2846740	*Rv2561-Rv2562*	*Rv2563*	Co ↑↓	5′ UTR
Mpr13 (Mcr14)		a-321523 b-321829	*fadA2*	*fadE5*	Co*	3′ UTR
Mpr14 (Mcr9)		a-3317473 b-3317835	*ilvB1*	*cfp6*	Co	5′ UTR/RS
Mpr15		a-3506235 b-3506773	*Rv3183*	*Rv3188*	Ind	intergenic
Mpr16		a-3731024 b-3731280	*Rv3364c*	*PPE55*	Ind	5′ UTR AS, 3′ UTR AS
Mpr17		a-4019618 b-4019895	*clpC1*	*lsr2*	Ind	3′ UTR AS
Mpr18		a-4065899 b-4066653	*Rv3651*	*PE_PGRS60*	Co ↓	5′ UTR
Mpr19		a-4072260 b-4072760	*Rv3660c*	*Rv3661*	Ind*	intergenic
Mpr20		a-779354 b-779685	*Rv0650*	*rplJ*	Co ↑	3′ UTR
Mpr21		a-818355 b-818508	*Rv0686*	*Rv0687*	Ind	intergenic

B

C

Figure 3. sRNAs identified in BCG through computational predictions. (**A**) Properties of Mpr1-Mpr21 are depicted as for the Mcr RNAs in Figure 2A. Mpr7, Mpr13, and Mpr14 coincide with the cloned sRNAs Mcr3, Mcr14 and Mcr9, respectively. *See ref. (6). (**B**) Representative computationally predicted sRNAs verified in BCG by northern blot, depicted as in Figure 2B. Mpr5-11 were identified with 7-day shaking BCG RNA, while Mpr13 and Mpr17 were identified with 8-day shaking BCG RNA and 6-day standing BCG RNA, respectively. All northerns shown were run on 10% denaturing PAGE. Probes are given in Supplementary Table S3. (**C**) Representative co-transcription. Six of the nine co-transcribed RNAs are shown, with the chromosomal DNA positive PCR control (C), and RT-PCR reactions spanning the junction of the sRNA and adjacent ORF with (+) or without (–) reverse transcriptase.
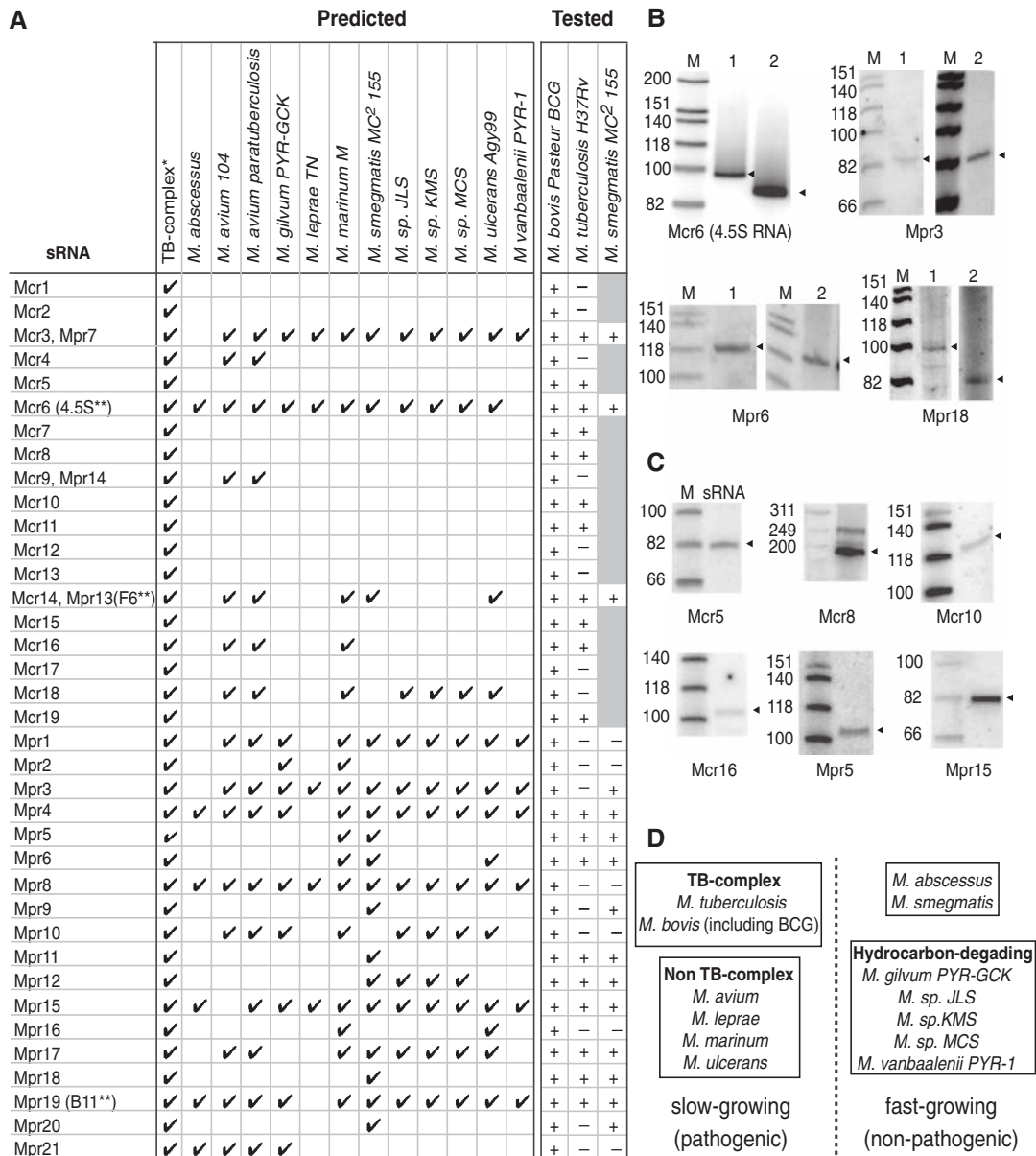
A — Predicted / Tested

sRNA	TB-complex*	M. abscessus	M. avium 104	M. avium paratuberculosis	M. gilvum PYR-GCK	M. leprae TN	M. marinum M	M. smegmatis MC² 155	M. sp. JLS	M. sp. KMS	M. sp. MCS	M. ulcerans Agy99	M vanbaalenii PYR-1	M. bovis Pasteur BCG	M. tuberculosis H37Rv	M. smegmatis MC² 155
Mcr1	✔													+	−	
Mcr2	✔													+	−	
Mcr3, Mpr7	✔		✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	+	+	+
Mcr4	✔		✔	✔										+	−	
Mcr5	✔													+	+	
Mcr6 (4.5S**)	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔		+	+	+
Mcr7	✔													+	+	
Mcr8	✔													+	+	
Mcr9, Mpr14	✔		✔	✔										+	−	
Mcr10	✔													+	+	
Mcr11	✔													+	+	
Mcr12	✔													+	−	
Mcr13	✔													+	−	
Mcr14, Mpr13(F6**)	✔		✔	✔			✔	✔				✔		+	+	+
Mcr15	✔													+	+	
Mcr16	✔		✔	✔			✔							+	+	
Mcr17	✔													+	−	
Mcr18	✔		✔	✔			✔			✔	✔	✔	✔	+	−	
Mcr19	✔													+	+	
Mpr1	✔		✔	✔	✔		✔	✔	✔	✔	✔	✔	✔	+	−	−
Mpr2	✔					✔	✔							+	−	−
Mpr3	✔		✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	+	−	+
Mpr4	✔	✔	✔	✔	✔		✔	✔	✔	✔	✔	✔	✔	+	+	+
Mpr5	✔							✔	✔					+	+	+
Mpr6	✔							✔	✔			✔		+	+	+
Mpr8	✔	✔	✔	✔	✔		✔	✔	✔	✔	✔	✔	✔	+	−	−
Mpr9	✔							✔						+	−	+
Mpr10	✔		✔	✔	✔		✔		✔	✔	✔			+	−	−
Mpr11	✔							✔						+	+	+
Mpr12	✔							✔	✔	✔	✔			+	+	+
Mpr15	✔	✔			✔	✔	✔	✔	✔	✔	✔	✔	✔	+	+	+
Mpr16	✔							✔				✔		+	−	−
Mpr17	✔		✔	✔			✔	✔	✔	✔	✔			+	+	+
Mpr18	✔							✔						+	+	+
Mpr19 (B11**)	✔	✔	✔	✔	✔		✔	✔	✔	✔	✔	✔	✔	+	+	+
Mpr20	✔							✔						+	−	+
Mpr21	✔	✔	✔	✔	✔									+	−	−

B — Mcr6 (4.5S RNA), Mpr3, Mpr6, Mpr18

C — Mcr5, Mcr8, Mcr10, Mcr16, Mpr5, Mpr15

D

TB-complex	M. abscessus
M. tuberculosis	M. smegmatis
M. bovis (including BCG)	

Non TB-complex	Hydrocarbon-degading
M. avium	M. gilvum PYR-GCK
M. leprae	M. sp. JLS
M. marinum	M. sp.KMS
M. ulcerans	M. sp. MCS
	M. vanbaalenii PYR-1

slow-growing (pathogenic) | fast-growing (non-pathogenic)

Figure 4. The Mcr and Mpr sRNAs are conserved within mycobacteria. (**A**) Summary of the BCG sRNAs and their presence in other mycobacterial genomes. For Mcr1-19, homology was determined by megablast using the cloned sRNA sequence, while for Mpr1-21 homology was determined by blastn using the candidate sRNA sequence (only results with an *E* value <1 are shown). sRNAs that did not show homology to *M. smegmatis* (Mpr5, 6, 9, 12, 13, 17, 18 and 20) were then rechecked by blastn restricted to *Mycobacterium* (taxid1763); a few of these had *E* values >1 (Mpr9, *E* = 5.0; Mpr11, *E* = 4.5; and Mpr17, *E* = 2.4). sRNAs were tested for expression by Northern blot in the three genomes listed; shaded boxes were not tested. For clarity, duplicate sRNAs are listed only once. + = positive signal by northern; − = negative signal by northern. *In this study TB-complex includes *M. tuberculosis* H37Rv, *M. bovis* subsp. bovis AF2122/97 and *M. bovis* Pasteur BCG; **previously identified in Mtb (6). (**B**) Representative sRNAs that are expressed in *M. smegmatis*. The lanes are as follows: M. Marker, in nucleotides, 1. log-phase BCG RNA, 2. log-phase *M. smegmatis* RNA. (**C**) Representative sRNAs that are expressed in Mtb. M. Marker, in nucleotides, 1. 7-day shaking Mtb RNA. (**D**) Schematic of the general division of mycobacteria. Species correspond to those in panel A, with homology to the Mcr and Mpr sRNAs.

under the conditions tested. This could result from different regulation of the same RNA between Mtb and BCG, or from idiosyncrasies of the RNA preparations (i.e. lower sensitivity of detection in Mtb due to reduced RNA yields). It is also worth noting that, although BCG and Mtb are both TB-complex mycobacteria, Mtb is a virulent pathogen, while BCG is an attenuated vaccine strain derived from pathogenic *M. bovis*. Therefore Mtb and

BCG may use different mechanisms to regulate expression and/or stability of homologous sRNAs. Virulence-associated sRNA regulatory differences could be significant for identifying new targets for novel TB therapeutics. On a broader level, these sRNAs can be used to help elucidate how sRNAs work in a G⁺ system.

The types of transcripts identified here include the gamut of regulatory RNAs that have been described in
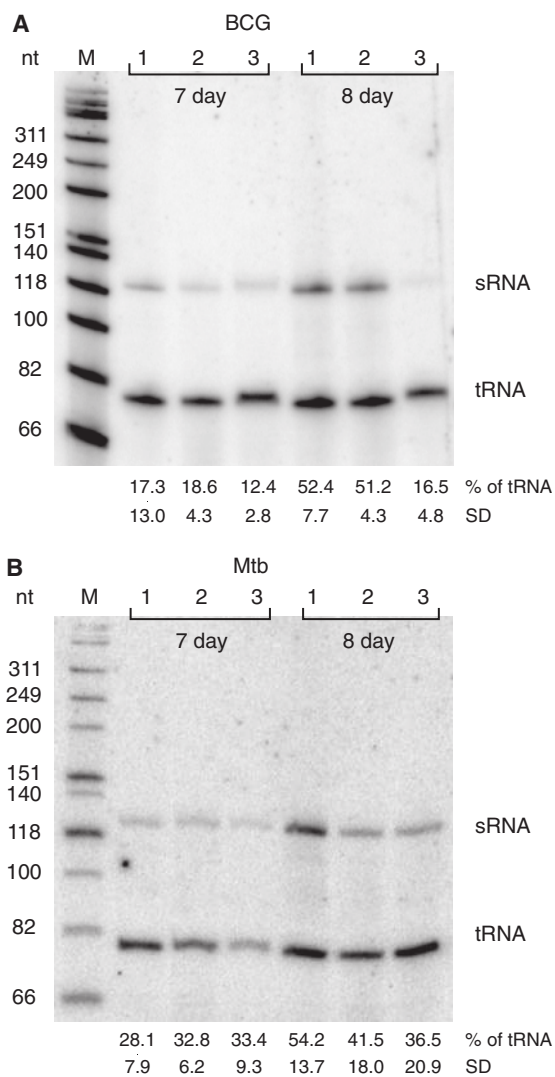
Figure 5. Differential expression of Mcr11. Representative blots showing that Mcr11 expression is responsive to environmental growth conditions, in both BCG (**A**) and Mtb (**B**). Lanes are as follows: M. Marker, in nucleotides, 1: shaking, 2: shaking, acid-treated, 3: shaking, low O_2. Differences in sRNA expression were determined by the percentage of tRNA signal for each sample; the average of three experiments is given, with the standard deviations listed below.

other bacteria, corresponding to intergenic, antisense and sense sRNAs, as well as potential regulatory UTRs and possibly multiple riboswitches. The very definition of a sRNA is 'a matter of perspective', since sRNAs can originate from independent sRNA genes, 5′UTR attenuation or processing, 3′UTR processing, and possibly even overlapping protein coding regions (7). Thus, distinct sRNAs are not necessarily due to independent RNA synthesis (7,41,42). Intergenic sRNAs usually act *in trans* to regulate a distal mRNA target, while antisense sRNAs presumably regulate their cognate mRNAs. Sense sRNAs may regulate *in cis* or *in trans*, depending on their sequence. It will be interesting to determine whether, and how, the sRNAs, found in this study that are sense to protein coding regions, such as Mcr5 and Mcr12 (Figure 2A) act as regulators.

Phylogeny aids in computational predictions of sRNAs

Several factors limit the use of bioinformatics and computational algorithms. For instance, Dynalign predictions, which rely on the stable formation of secondary structures from intergenic sequences (43), are challenged by genomes that are high in GC content; this is particularly problematic with mycobacterial genomes, since their GC content is >65%. Attempts to confirm sRNAs that were predicted by the Dynalign method repeatedly yielded northern blots without signals, or with signals corresponding to mycobacterial tRNAs (J. DiChiara and D. Mathews, unpublished results). Another typical challenge with computational predictions is the need to rely on transcription factor binding sites and terminators that are often species specific, and may be unknown in organisms like *M. bovis*. Although promoter and terminator regions are better studied in Mtb and *M. smegmatis* (44,45) the recent finding of unpredicted promoters for sRNA expression in Mtb indicates the immaturity of alogorithms within the mycobacterial species (6). Hence, initial results from predictive analyses often comprise long lists of potential sRNA candidates that are impractical to test, due to the inherent high numbers of false positives. The challenge then becomes how to reduce the number of predictions so as to increase the efficiency of confirming positive candidates, without reducing the sensitivity of the method. Here, a phylogeny-based strategy was used as an additional filter to select a smaller subset of BCG sRNA predictions for testing; that step aided in identifying novel sRNAs in mycobacteria. A recent study took a similar phylogenetic approach to identify noncoding RNAs on a much larger scale, by looking for conservation in 422 bacterial genomes (46). Additionally, analysis of the evolutionary conservation of sRNAs across species could be highly beneficial in future functional studies, particularly when looking at the biological relevance of an sRNA between different mycobacterial species.

Despite the shortcomings of current algorithms in mycobacteria, the methods used in this study had virtually identical success rates in identifying sRNAs when compared to the cloning-based approach (31.3 versus 31.7%, respectively; the number positive by northern blot per the number of potential sRNAs tested in BCG, Figure 1). A larger scale cloning approach, coupled with deep sequencing, would likely identify a greater proportion of sRNAs expressed under a certain condition. However, by combining small-scale cloning and computational methods we were able to identify conserved sRNAs across many different mycobacteria. Combining the data gathered on all the types of sRNAs identified in this study, there is the possibility of improving the current algorithm to have an even greater success rate. We are currently analyzing in detail the advantages of this method as a filter to large datasets of computational predictions.

Locations and implications of sRNAs identified

Highly conserved sRNAs expressed in all three species were Mcr3/Mpr7, Mcr14/Mpr13 and Mpr4, 5, 6, 11, 12, 15, 17, 18 and 19 (Figure 4A). Given that these sRNAs

occur in the relatively divergent *M. smegmatis* (Figure 4D), it will be informative to see if these sRNAs are also expressed in the other mycobacteria predicted in Figure 4A, such as the hydrocarbon-degrading mycobacteria. These species are divergent from one another and from other fast and slow-growing mycobacteria (47–49). The above-listed sRNAs may regulate conserved cellular functions, and it would be interesting to establish if they regulate the same targets, in the same manner, within the different mycobacteria. For example, the 'region of difference 1' (RD1) of Mtb encodes a specialized secretion system that is required for virulence, and absent from many non-pathogenic mycobacteria (50,51). Like the well-studied bacterial pathogen *Salmonella enterica* serovar Typhimurium, which is comprised of a mosaic genome, it is possible that sRNAs outside RD1 act to regulate virulence-associated genes within RD1 (52).

Mpr8, which likely encodes an L20 leader based on Rfam analysis, is the first such instance of an L20 leader identified in a high-GC G^+ bacterium. As mentioned above, L20 acts as a riboswitch. This 200-bp leader has only been identified in low-GC G^+ bacteria to date, including, but not limited to, Clostridia, Lactobacillales and Bacillales (http://rfam.sanger.ac.uk/family?acc = RF00558#tabview = tab4).

sRNAs were identified in, or upstream of, several ORFs that are known to affect virulence. These sRNAs include Mcr9, located upstream of *ilvB1* and Mcr19, located within the regulatory protein Rv0485. Mcr9 is co-transcribed with *ilvB1*, as mentioned in 'Results' section. IlvB1 is required for the synthesis of branched chain amino acids and a *ΔilvB1* strain cannot grow *in vitro* without all three branched chain amino acids added to the media. Additionally, this Mtb deletion strain is attenuated for virulence in mice, but persists in the spleen and lungs, making it a potential vaccine candidate (27). Mcr19 is located sense to the gene for the regulatory protein Rv0485, which was recently found to modulate the *pe13/ppe18* genes. These *pe13/ppe18* genes are unique to mycobacteria and may play a role in Mtb virulence (34).

Further studies are needed to assess Mcr11's relevance for Mtb pathogenesis. Although the target of this sRNA is not yet known, Mcr11 transcription is responsive to growth phase, and its regulation may differ between Mtb and BCG under hypoxic conditions. A previous study of *Rv1265* expression showed pH had an effect in BCG, but not in Mtb (23). Additionally, there is preliminary evidence that Mcr11 is involved in cAMP metabolism (J.M. DiChiara and K. A. McDonough, unpublished observation). cAMP plays a role in the interaction of TB-complex mycobacteria during macrophage infection (53,54) and multiple proteins are involved in cAMP-mediated gene regulation, such as CRP_{Mt} and Cmr (23,55). Thus, this work opens many possibilities for the study of sRNAs as potential virulence regulators in Mtb.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Aziz,M.A., Wright,A., Laszlo,A., De Muynck,A., Portaels,F., Van Deun,A., Wells,C., Nunn,P., Blanc,L. and Raviglione,M. (2006) Epidemiology of antituberculosis drug resistance (the Global Project on Antituberculosis Drug Resistance Surveillance): an updated analysis. *Lancet*, **368**, 2142–2154.
2. WHO Global Task Force on XDR-TB. Meeting (2006). Report of the meeting of the WHO Global Task Force on XDR-TB: Geneva, Switzerland, 9–10 October 2006. Geneva: World Health Organization, 2007. http://whqlibdoc.who.int/hq/2007/WHO_HTM_TB_2007.375_eng.pdf, doi: WHO/HTM/TB/2007.375.
3. Williams,N. (2006) Footprint fears for new TB threat. *Curr. Biol.*, **16**, R821–822.
4. Borgdorff,M. and Small,P.M. (2009) Scratching the surface of ignorance on MDR tuberculosis. *Lancet*, **373**, 1822–1824.
5. Wright,A., Zignil,M., Van Deun,A., Falzon,D., Gerdes,S.R., Feldman,K., Hoffner,S., Drobniewski,F., Barrera,L., van Soolingen,D. *et al.* for the Global Project on Anti-Tuberculosis Drug Resistance Surveillance. (2009) Epidemiology of antituberculosis drug resistance 2002-07: an updated analysis of the Global Project on Anti-Tuberculosis Drug Resistance Surveillance. *The Lancet*, **373**, 1861–1873.
6. Arnvig,K.B. and Young,D.B. (2009) Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol. Microbiol.*, **73**, 397–408.
7. Vogel,J., Bartels,V., Tang,T.H., Churakov,G., Slater-Jager,J.G., Huttenhofer,A. and Wagner,E.G.H. (2003) RNomics in *Escherichia coli* detects new sRNA species and indiates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6643.
8. Gottesman,S. (2005) Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.*, **21**, 399–404.
9. Lease,R.A., Smith,D., McDonough,K. and Belfort,M. (2004) The small noncoding DsrA RNA is an acid resistance regulator in *Eschericia coli*. *J. Bacteriol.*, **186**, 6179–6185.
10. Toledo-Arana,A., Repoila,F. and Cossart,P. (2007) Small noncoding RNAs controlling pathogenesis. *Curr. Opin. Microbiol.*, **10**, 182–188.
11. Johansson,J. (2005) RNA Molecules: More than Mere Information Intermediaries. *ASM News*, **71**, 515–520.
12. Sun,X., Zhulin,I. and Wartell,R.M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res.*, **30**, 3662–3671.
13. Chao,Y. and Vogel,J. (2010) The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.*, **13**, 24–33.
14. Bohn,C., Rigoulay,C. and Bouloc,P. (2007) No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. *BMC Microbiol.*, **7**, 10.

15. Nielsen,J.S., Lei,L.K., Ebersbach,T., Olsen,A.S., Klitgaard,J.K., Valentin-Hansen,P. and Kallipolitis,B.H. (2009) Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. *Nucleic Acids Res*, **28**, 907–919.

16. Gottesman,S. (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.*, **58**, 303–328.

17. Perez-Bello,A., Munteanu,C.R., Ubeira,F.M., De Magalhães,A.L., Uriarte,E. and González-Díaz,H. (2009) Alignment-free prediction of mycobacterial DNA promoters based on pseudo-folding lattice network or star-graph topological indices. *J. Theor. Biol*, **256**, 458–466.

18. Livny,J., Fogel,M.A., Davis,B.M. and Waldor,M.K. (2005) sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic Acids Res.*, **33**, 4096–4105.

19. Livny,J. (2007) Efficient annotation of bacterial genomes for small, noncoding RNAs using the integrative computational tool sRNAPredict2. *Methods Mol. Biol.*, **395**, 475–488.

20. Livny,J., Teonadi,H., Livny,M. and Waldor,M.K. (2008) High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE*, **3**, e3197.

21. Purkayastha,A., McCue,L.A. and McDonough,K.A. (2002) Identification of a *Mycobacterium tuberculosis* putative classical nitroreductase gene whose expression is coregulated with that of the *acr* gene within macrophages, in standing versus shaking cultures, and under low oxygen conditions. *Infect. Immun.*, **70**, 1518–1529.

22. Florczyk,M.A., McCue,L.A., Purkayastha,A., Currenti,E., Wolin,M.J. and McDonough,K.A. (2003) A family of *acr*-coregulated *Mycobacterium tuberculosis* genes shares a common dna motif and requires Rv3133c (*dos*R or *dev*R) for expression. *Infect. Immun.*, **71**, 5332–5343.

23. Gazdik,M.A., Bai,G., Wu,Y. and McDonough,K.A. (2009) Rv1675c (*cmr*) regulates intramacrophage and cAMP-induced gene expression in *Mycobacterium tuberculosis*-complex mycobacteria. *Mol. Microbiol.*, **71**, 434–448.

24. Lau,N.C., Lim,L.P., Wienstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.

25. Kempsell,K.E., Ji,Y., Estrada-G,I.C.E., Colston,M.J. and Cox,R.A. (1992) The nucleotide sequence of the promoter, 16S rRNA and spacer region of the ribosomal RNA operon of *Mycobacterium tuberculosis* and comparison with *Mycobacterium leprae* precursor rRNA. *J. Gen. Microbiol.*, **138**, 1717–1727.

26. Roth,A., Fischer,M., Hamid,M.E., Michalke,S., Ludwig,W. and Mauch,H. (1998) Differentiation of phylogentetically related slowly growing mycobacteria based on 16s-23s rrna gene internal transcribed spacer sequences. *J. Clin. Microbiol.*, **36**, 139–147.

27. Awasthy,D., Gaonkar,S., Shandil,R.K., Yadav,R., Bharath,S., Marcel,N., Subbulakshmi,V. and Sharma,U. (2009) Inactivation of *ilvB1* gene in *Mycobacterium tuberculosis* leads to branched chain amino acids auxotrophy and attenuation of virulence in mice. *Microbiology*, **155**, 2978–2987.

28. Grandoni,J.A., Zahler,S.A. and Calvo,J.M. (1992) Transcriptional regulation of the *ilv*-leu operon of *Bacillus subtilis*. *J. Bacteriol.*, **174**, 3212–3219.

29. Grandoni,J.A., Fulmer,S.B., Brissio,V., Zahler,S.A. and Calvo,J.M. (1993) Regions of the *Bacillus subtilis ilv*-leu Operon Involved in Regulation by Leucine. *J. Bacteriol.*, **175**, 7581–7593.

30. Henkin,T.M. (1996) Control of Transciption Termination in Prokaryotes. *Annu. Rev. Genet.*, **30**, 35–57.

31. Kremer,L., Nampoothiri,K.M., Lesjean,S., Dover,L.G., Graham,S., Betts,J., Brennan,P.J., Minnikin,D.E., Locht,C. and Besra,G.S. (2001) Biochemical characterization of acyl carrier protein (AcpM) and malonyl-CoA:AcpM transacylase (mtFabD), two major components of *Mycobacterium tuberculosis* fatty acid synthase II. *J. Biol. Chem.*, **276**, 27967–27974.

32. Betts,J.C., Lukey,P.T., Robb,L.C., McAdam,R.A. and Duncan,K. (2002) Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol.*, **43**, 717–731.

33. Jungblut,P.R., Schaible,U.E., Mollenkopf,H.J., Zimny-Arndt,U., Raupach,B., Mattow,J., Halada,P., Lamer,S., Hagens,K. and Kaufmann,S.H. (1999) Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol. Microbiol.*, **33**, 1103–1117.

34. Goldstone,R.M., Goonesekera,S.D., Bloom,B.R. and Sampson,S.L. (2009) The transcriptional regulator Rv0485 modulates the expresion of a *pe/ppe* gene pair and is required for *Mycobcaterium tuberculosis* virulence. *Infect. Immun.*, **77**, 4654–4667.

35. Choonee,N., Evan,S., Zig,L. and Putzer,H. (2007) Ribosomal protein L20 controls expression of the *Bacillus subtilis infC* operon via a transcription attenuation mechanism. *Nucleic Acids Res.*, **35**, 1578–1588.

36. Devulder,G., Pérouse de Montclos,M. and Flandrois,J.P. (2005) A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int. J. Syst. Evol. Microbiol.*, **55**, 293–302.

37. Hobson,R.J., McBride,A.J., Kempsell,K.E. and Dale,J.W. (2002) Use of an arrayed promoter -probe library for the identification of macrophage-regulated genes in *Mycobacterium tuberculosis*. *Microbiology*, **148**, 1571–1579.

38. Monohan,I.M., Bettes,J., Banerjee,D.K. and Butcher,P.D. (2001) Differential expression of mycobacterial proteins following phagocytosis by macrophages. *Microbiology*, **147**, 459–471.

39. Smith,I. (2003) *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin. Microbiol. Rev.*, **16**, 463–496.

40. Sturgill-Koszycki,S., Schlesinger,P.H., Chakraborty,P., Haddix,P.L., Collins,H.L., Fok,A.K., Allen,R.D., Gluck,S.L., Heuser,J. and Russell,D.G. (1994) Lack of acidification in *Mycobacterial* phagosomes produced by exclusion of the vesicular proton-ATPase. *Science*, **263**, 678–681.

41. Kawano,M., Reynolds,A.A., Miranda-Rios,J. and Storz,G. (2005) Detection of 5′- and 3′-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.*, **33**, 1040–1050.

42. Liu,J.M., Livny,J., Lawrence,M.S., Kimball,M.D., Waldor,M.K. and Camilli,A. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.*, **37**, e46, 1–10.

43. Uzilov,A.V., Keegan,J.M. and Mathews,D.H. (2006) Detection of non-coding RNAs on the basis of predicted secondary structure formation and free energy change. *BMC Bioinformatics*, **7**, 173.

44. Lee,J.H., Karakousis,P.C. and Bishai,W.R. (2008) Roles of SigB and SigF in the Mycobacterium tuberculosis sigma factor network. *J. Bacteriol.*, **190**, 699–707.

45. Agarwal,N. and Tyagi,A.K. (2006) Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity. *Nucleic Acids Res.*, **34**, 4245–4257.

46. Marchais,A., Naville,M., Bohn,C., Bouloc,P. and Gautheret,D. (2009) Single-pass classification of all noncoding sequences in a bacterial genome using phylogenetic profiles. *Genome Res.*, **19**, 1084–1092.

47. Miller,C.D., Hall,K., Liang,Y.N., Nieman,K., Sorensen,D., Issa,B., Anderson,A.J. and Sims,R.C. (2004) Isolation and Characterization of polycyclic aromatic hydrocarbon-degrading mycobacterium isolates from soil. *Microb. Ecol.*, **48**, 230–238.

48. Khan,A.A., Kim,S.J., Paine,D.D. and Cerniglia,C.E. (2002) Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Mycobacterium* sp. strain PYR-1, as Mycobacterium vanbaalenii sp. nov. *Int. J. Syst. Evol. Microbiol.*, **52**, 1997–2002.

49. Brezna,B., Khan,A.A. and Cerniglia,C.E. (2003) Molecular characterization of dioxygenases from polycyclic aromatic hydrocarbon-degrading Mycobacterium spp. *FEMS Microbiol. Lett.*, **223**, 177–183.

50. Ganguly,N., Siddiqui,I. and Sharma,P. (2008) Role of *M*. tuberculosis RD-1 region encoded secretory proteins in protective response and virulence. *Tuberculosis*, **88**, 510–517.

51. van Ingen,J., de Zwaan,R., Dekhuijzen,R., Boeree,M. and van Soolingen,D. (2009) Region of difference 1 in nontuberculous *Mycobacterium* Species adds a phylogenetic and taxonomical character. *J. Bacteriol.*, **191**, 5865–5967.

52. Vogel,J. (2009) A rough guide to the non-coding RNA world of *Salmonella*. *Mol. Microbiol.*, **71**, 1–11.

53. Agarwal,N., Lamichhane,G., Gupta,R., Nolan,S. and Bishai,W.R. (2009) Cyclic AMP intoxication of macrophages by a *Mycobacterium tuberculosis* adenylate cyclase. *Nature*, **460**, 98–102.
54. Bai,G., Schaak,D.D. and McDonough,K.A. (2009) cAMP levels within *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG increase upon infection of macrophages. *FEMS Immunol. Med. Microbiol.*, **55**, 68–73.
55. Bai,G., McCue,L.A. and McDonough,K.A. (2005) Characterization of *Mycobacterium tuberculosis* Rv3676 (CRP_{Mt}), a cyclic AMP receptor protein-like DNA binding protein. *J. Bacteriol.*, **187**, 7795–7804.