Paradoxical conservation of a set of three cellulose-targeting genes in *Mycobacterium tuberculosis* complex organisms

Felix Mba Medie,^{1,2} Iskandar Ben Salah,¹ Michel Drancourt¹ and Bernard Henrissat²

Correspondence

Bernard Henrissat

bernard.henrissat@afmb.univ-mrs.fr

¹Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UMR CNRS 6236, IRD 198, IFR 48, Faculté de Médecine, Université de la Méditerranée, Marseille, France

²Architecture et Fonction des Macromolécules Biologiques, UMR 6098, IFR 48, CNRS, Université de la Méditerranée et Université de Provence, Marseille, France

The genome of the tuberculosis agent *Mycobacterium tuberculosis* encodes a putative cellulose-binding protein (*CBD2*), one candidate cellulase (*Cel12*), and one fully active cellulase (*Cel6*). This observation is puzzling, because cellulose is a major component of plant cell walls, whereas *M. tuberculosis* is a human pathogen without known contact with plants. In order to investigate the biological role of such cellulose-targeting genes in *M. tuberculosis* we report here the search for and transcription analysis of this set of genes in the genus *Mycobacterium*. An *in silico* search for cellulose-targeting orthologues found that only 2.5 % of the sequenced bacterial genomes encode the *Cel6*, *Cel12* and *CBD2* gene set simultaneously, including those of the *M. tuberculosis* complex (MTC) members. PCR amplification and sequencing further demonstrated the presence of these three genes in five non-sequenced MTC bacteria. Among mycobacteria, the combination of *Cel6*, *Cel12* and *CBD2* was unique to MTC members, with the exception of *Mycobacterium bovis* BCG Pasteur, which lacked *CBD2*. RT-PCR in *M. tuberculosis* H37Rv indicated that the three cellulose-targeting genes were transcribed into mRNA. The present work shows that MTC organisms are the sole mycobacteria among very few organisms to encode the three cellulose-targeting genes *CBD2*, *Cel6* and *Cel12*. Our data point toward a unique, yet unknown, relationship with non-plant cellulose-producing hosts such as amoebae.

Received 30 December 2009

Revised 4 February 2010

Accepted 7 February 2010

INTRODUCTION

Cellulose, a linear polymer of glucose residues connected by β -1,4 linkages, makes up about half of the cell walls of plants and is both the most abundant polysaccharide and the major source of photosynthetically fixed carbon on Earth (Doi and Kosugi, 2004; Gilbert *et al.*, 2008; Lynd *et al.*, 2002). The natural resistance of cellulose to microbial digestion has probably guided plants to use it to build thick

cell walls that can last up to several thousand years. Indeed, only a few saprophytic micro-organisms of fungal or bacterial origin have evolved cellulases, i.e. enzymes able to bind to cellulose and hydrolyse it to glucose (Lynd *et al.*, 2002). A noteworthy feature of cellulases is that their structure is frequently modular, so that the catalytic domain carries one or more cellulose-binding modules.

The recent discovery that *Mycobacterium tuberculosis*, the pathogen responsible for human tuberculosis, encodes a functional cellulase, *Cel6* (*Rv0062*), was therefore totally unexpected (Varrot *et al.*, 2005). The analysis of the *M. tuberculosis* genome shows that gene *Rv1090* potentially encodes a second cellulase (*Cel12*) and that gene *Rv1987* encodes a candidate cellulose-binding protein (*CBD2*). These three cellulose-targeting genes are completed by the presence of a putative β -glucosidase gene, *Rv0186*.

Cellulases are found in the Carbohydrate-Active EnZymes database (www.cazy.org) (Cantarel *et al.*, 2009) in nine sequence-based families of glycoside hydrolases (GHs), namely GH5, GH6, GH7, GH8, GH9, GH12, GH44, GH45 and GH48. Of these, the two mycobacterial cellulases

Abbreviations: MAC, *Mycobacterium avium* complex; MTC, *Mycobacterium tuberculosis* complex; NTM, non-tuberculosis *Mycobacterium*.

The GenBank/EMBL/DDBJ accession numbers for the *Cel6*, *Cel12* and *CBD2* nucleotide sequences of non-sequenced MTC members determined in this study are shown in Table 3.

A supplementary table, showing a list of all sequenced bacteria in the CAZy database with one or more cellulose-targeting gene(s) in their genome, and two supplementary figures, showing a phylogenetic tree of *Cel6* gene sequences in 17 bacteria using the neighbour-joining method with K2P distance correction model, and a phylogenetic tree of *CBD2* gene sequences in 16 bacteria using the neighbour-joining method with K2P distance correction model, are available with the online version of this paper.

(*Cel6* and *Cel12*) belong to families GH6 and GH12, respectively.

In plant cell walls, cellulose is found embedded in various complex polysaccharides (hemicelluloses and pectins) and lignin (Lynd *et al.*, 2002; Varner & Lin, 1989), and the cellulolytic genes are usually accompanied by genes that encode a range of enzymes active against xylans and pectins.

Orthologues of the *M. tuberculosis* genes *Cel6*, *Cel12* and *CBD2* have been found in all currently available mycobacterial genomes, including three of the seven species of the *M. tuberculosis* complex (MTC), i.e. *M. tuberculosis* H37Rv (AL123456), *M. tuberculosis* H37Ra (ABQ72837), *M. tuberculosis* CDC1551 (AE000516), *M. tuberculosis* F11 (ABR05454), *M. tuberculosis* KZN 1435 (CP001658), *Mycobacterium bovis* (BX248337) and *M. bovis* BCG Tokyo (AP010918). It is currently unknown whether other MTC members or non-tuberculous groups such as the *Mycobacterium avium* complex (MAC) or *Mycobacterium abscessus* complex also have the same putative cellulose-targeting genes. Likewise, there is no information regarding the transcription of these genes, hampering further investigation of the potential role of cellulose-targeting genes in *Mycobacterium* species. In an effort towards understanding the role of cellulose-targeting genes in *Mycobacterium*, we have searched for the presence of the *Cel6*, *Cel12* and *CBD2* gene set in representatives of the *Mycobacterium* complexes and have examined their transcription.

METHODS

Bioinformatics analyses. The presence or absence of the *Cel6*, *Cel12* and *CBD2* genes in 907 bacterial genomes was extracted from the CAZy database (www.cazy.org), which is continuously updated in our laboratory based on day-to-day examination of the daily releases from GenBank (Cantarel *et al.*, 2009). The survey was then completed by BLAST (Altschul *et al.*, 1990) searches of additional genomes of the MTC, MAC and other non-tuberculosis *Mycobacterium* (NTM) organisms that are available as reference sequences at the NCBI (<http://www.ncbi.nlm.nih.gov/>). These genomes included three MTC members, i.e. *M. tuberculosis* H37Rv (AL123456) (Cole *et al.*, 1998), *M. bovis* (BX248333) (Garnier *et al.*, 2003) and *M. bovis* BCG (Pasteur and Tokyo strains) (NC_008769, AP010918) (Brosch *et al.*, 2007; Seki *et al.*, 2009), two MAC members, i.e. *M. avium* 104 (CP00479) and *M. avium* subsp. *paratuberculosis* K-10 (AE016958) (Li *et al.*, 2005), *Mycobacterium leprae* Br4923 (FM211192), *M. leprae* TN (AL450380), *M. abscessus* (CU458896), *Mycobacterium smegmatis* (CP000480), *Mycobacterium ulcerans* (CP000325) (Stinear *et al.*, 2007, 2008), *Mycobacterium marinum* (CP000854) (Stinear *et al.*, 2008), *Mycobacterium vanbaalenii* (CP000511), *Mycobacterium* sp. MCS (CP000384), *Mycobacterium* sp. JLS (CP000580), *Mycobacterium* sp. KMS (CP000518) and *Mycobacterium gilvum* (CP000656). The genomic organization of the *Cel6*, *Cel12* and *CBD2* genes was derived from inspection of the above-mentioned *Mycobacterium* genomes. Signal peptide prediction was done using Phobius (<http://phobius.cbr.su.se/>).

PCR amplification and sequencing. In order to avoid artefacts, we controlled the quality of the genomic DNA and of any PCR primer

pair used by including negative controls in all our PCR mixes. Also we took care to include all pathogenic *Mycobacterium* species, including the MTC organisms, the MAC organisms and the NTM organisms. Genomic DNA isolated from the MTC members (*M. tuberculosis* H37Rv CIP103471, *M. bovis* CIP105050, *Mycobacterium africanum* CIP105147^T (type 1), *M. bovis* BCG vaccine strain type 105060, *Mycobacterium microti* CIP104256, *Mycobacterium canettii* CIP140060001^T, *Mycobacterium pinnipedii* ATCC BAA-688 and *Mycobacterium caprae* CIP105776^T), from the MAC members (*M. avium* subsp. *hominissuis* IWGMT49, *Mycobacterium intracellulare* CIP104243, *Mycobacterium chimaera* CIP107892^T and *Mycobacterium colombiense* CIP108962), and from *M. abscessus* CIP104536, a *Mycobacterium marinum* clinical isolate and *Mycobacterium fortuitum* ATCC 49404 was used as a template for the PCR amplification and sequencing of *M. tuberculosis* *Cel6*, *Cel12* and *CBD2* gene orthologues. PCR primer pairs were designed after alignment of *Mycobacterium* genome sequences from GenBank (Table 1) by using Primer3 Input 0.4.0 (<http://frodo.wi.mit.edu/primer3/>). PCRs were carried out in a 2720 Thermal Cycler (Applied Biosystems) in a 50 µl final volume containing 25 µl H₂O, 5 µl 10× buffer (Qiagen), 25 µM MgCl₂, 100 µM of each dNTP, 5 µM each primer (Eurogentec), 2.5 U *Taq* DNA polymerase (Invitrogen) and 5 µl mycobacterial DNA, using the following program: 5 min at 95 °C, followed by 35 cycles consisting of 95 °C for 30 s, 60 °C for 30 s and 72 °C for 90 s, and a 10 min elongation step at 72 °C. For each experiment, a negative control consisting of PCR mix without target DNA was included. PCR products were resolved by electrophoresis on a 1.5% (w/v) agarose gel and stained with ethidium bromide. PCR products were then purified by adding 50 µl distilled H₂O to a purification plate (Millipore) that was agitated for 10 min. Forward and reverse sequencing mixtures contained 3 µl buffer (BigDye v1, Applied Biosystems), 10 µl distilled H₂O and 1 µl of 3.2 pmol primer µl⁻¹ in a final volume of 16 µl. The sequencing reaction comprised an initial denaturation step of 1 min at 95 °C, followed by 25 cycles of denaturation at 96 °C for 10 s, annealing at 50 °C for 5 s and elongation at 60 °C for 3 min. Sequencing products were purified using a Sephadex plate (Amersham Biosciences) that was centrifuged at 720 g for 3 min and deposited on a MicroAmp Optical 96-well reaction plate (Applied Biosystems). Sequencing electrophoresis was performed on a 3100 Genetic Analyzer (Applied Biosystems).

RT-PCR. Total RNA was extracted from 40 ml of exponential-phase cultures of *Mycobacterium* organisms. Mycobacteria were harvested by centrifugation at 1000 g for 5 min and the supernatant was discarded. The pellet was resuspended in 1 ml RLT buffer (Qiagen) containing 0.1% 2-mercaptoethanol and transferred to a 2 ml screw-capped tube containing silica beads (IEPSA Medical Diagnostics) and the mixture was homogenized in a FastPrep FP120 instrument for 45 s at speed 6.5 (Qbiogene). After centrifugation at 8000 g for 1 min at room temperature, 700 µl ethanol was added and the solution was divided and applied into two RNeasy spin columns (Qiagen). Centrifugation was carried out at 8000 g for 15 s and the sample was then washed once with 700 µl RW1 buffer and twice with 500 µl RPE buffer (Qiagen). The RNA was eluted with 50 µl RNase-free water per column, and the samples were further treated with DNase I to eliminate any genomic DNA contamination. The RNA analysis was done by formaldehyde 1.5% agarose gel electrophoresis. For RT-PCR, 10 µl RNA was mixed with 25 µl 2× Master Mix (Invitrogen), 1 µl (10 µM) of forward and reverse gene-specific primers (Table 2) and 2.5 U Platinum Pfx DNA polymerase (Invitrogen), and made up to 50 µl with distilled water. The reaction was carried out in a 2720 Thermal Cycler under the following conditions: 45 °C for 30 min, 95 °C for 2 min, followed by 25 cycles consisting of 95 °C for 30 s, 58 °C for 45 s and 72 °C for 1 min, and a 5 min elongation step at 72 °C. RT-PCR products were resolved by 1.5% agarose gel electrophoresis and stained with ethidium bromide.

Table 1. Number of *Cel6*, *Cel12* and *CBD2* genes in different bacterial species derived from the examination of their complete genome

Species	No. of genes			Accession no.
	<i>Cel6</i>	<i>Cel12</i>	<i>CBD2</i>	
<i>Catenulispora acidiphila</i> DSM 44928	4	2	30	CP001700
<i>Actinosynnema mirum</i> DSM 43827	3	2	30	CP001630
<i>Streptomyces avermitilis</i> MA-4680	4	3	11	BA000030
<i>Streptomyces coelicolor</i> A3(2)	3	2	14	AL645882
<i>Streptomyces scabiei</i> 87.22	4	3	16	FN554889
<i>Teredinibacter turnerae</i> T7901	3	1	24	CP001614
<i>Acidothermus cellulolyticus</i> 11B ATCC 43068	2	2	10	CP000481
<i>Xanthomonas campestris</i> pv. <i>campestris</i> B100	2	2	1	AM920689
<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. 8004	2	2	1	CP000050
<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	2	2	1	AE008922
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	1	2	1	AP008229
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99A	1	2	1	CP000967
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	1	2	1	AE013598
<i>Sorangium cellulosum</i> 'So ce 56'	1	1	6	AM746676
<i>Salinispora tropica</i> CNB-440	1	1	7	CP000667
<i>Salinispora arenicola</i> CNS-205	1	1	6	CP000850
<i>Mycobacterium tuberculosis</i> H37Ra ATCC 25177	1	1	1	CP000611
<i>Mycobacterium tuberculosis</i> KZN 1435	1	1	1	CP001658
<i>Mycobacterium tuberculosis</i> CDC1551	1	1	1	AE000516
<i>Mycobacterium tuberculosis</i> H37Rv	1	1	1	AL123456
<i>Mycobacterium tuberculosis</i> F11	1	1	1	CP000717
<i>Mycobacterium bovis</i> AF2122/97	1	1	1	BX248333
<i>Mycobacterium bovis</i> BCG str. Tokyo 172	1	1	1	AM746676
<i>Mycobacterium bovis</i> BCG str. Pasteur 1173P2	1	1	0	AM408590
<i>Mycobacterium gilvum</i> PYR-GCK	1	0	2	CP000656
<i>Mycobacterium vanbaalenii</i> PYR-1	1	0	2	CP000511
<i>Mycobacterium abscessus</i>	1	0	1	CU458896
<i>Mycobacterium marinum</i> M	1	0	5	CP000854
<i>Mycobacterium ulcerans</i> Agy99	0	0	5	CP000325
<i>Mycobacterium avium</i> 104	1	0	1	CP000479
<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> K-10	1	0	1	AE016958
<i>Mycobacterium</i> sp. MCS	1	0	0	CP000384
<i>Mycobacterium</i> sp. JLS	1	0	0	CP000580
<i>Mycobacterium</i> sp. KMS	1	0	0	CP000518
<i>Mycobacterium smegmatis</i> str. MC2 155	1	0	0	CP000480
<i>Mycobacterium leprae</i> Br4923	0	0	0	FM211192
<i>Mycobacterium leprae</i> TN	0	0	0	AL450380

RESULTS

Bioinformatics analyses

Out of the 907 bacterial genomes analysed in the CAZY database (November 2009), only 2.5 % harbour the *Cel6*, *Cel12* and *CBD2* genes simultaneously (Table 1). These 23 genomes include a group of 16 saprophytic organisms, and seven MTC members (Fig. 1). It is worth mentioning that most plant cell wall degraders do not appear in this list, as they rarely harbour a *Cel12* orthologue. Among the MTC members, only *M. bovis* BCG Pasteur strain 1173P2 lacks the *CBD2* gene (Mahairas *et al.*, 1996). The *Cel12* gene was found to be present in all the MTC members and absent in

all other *Mycobacterium* species under study. The structure of this gene is unusual, being composed of two overlapping fragments (*CelA2a* and *CelA2b*). Although the sequence similarity with bona fide *Cel12* proteins such as that of *Streptomyces lividans* (Sulzenbacher *et al.*, 1997) is significant, the MTC *Cel12* proteins all appear to have a truncation which has eliminated the signal peptide and the 40 first amino acids. The significance and consequences of this truncation are not clear, since they do not affect the catalytic region of the protein (data not shown). Non-MTC mycobacteria, such as *M. avium* 104, *M. avium* subsp. *paratuberculosis* K-10, *M. gilvum*, *M. marinum* and *M. vanbaalenii*, encode only two of the three genes (Table 1). A few other mycobacteria harbour only one of the three

Table 2. PCR primers used in this study

Technique used	Organism	Gene	Primers	Sequence (5'→3')	Amplicon size (bp)
PCR	<i>M. tuberculosis</i> H37Rv	<i>Cel6</i>	Cel6F/R	GTCGAATTGCGGCATTATTT/ATCGATGGCGTACTGGCTAA	980
		<i>CBD2</i>	CBD2F/R	TACTGAGGTGGGCCAGATTT/AGTTGACTTTCGCCGAGAAC	397
		<i>Cel12</i>	Cel12F/R (2b)	ATGGCCGGACTGAACATTTA/ATATTGCCCGTTGAGCAGAC	377
		<i>Cel12</i>	Cel12F/R (2a)	CTAGTTCGGGACGACTACGG/GCAGATCGAGGGGTAACCTCA	204
		<i>Cel12</i>	Cel12F/R (2ab)	GTGCTGCCCCAACCAAC/ATATTGCCCGTTGAGCAGAC	482
RT-PCR	<i>M. tuberculosis</i> H37Rv	<i>Cel6</i>	Cel6F/R	AGGAACGTTTCGACTTGGTG/CCATAGCCGATTTCTCATC	200
		<i>Cel12</i>	Cel12F/R	CTGGATTCCACACCCAAGAC/TCGACGAAACTCATCACGTC	200
		<i>CBD2</i>	CBD2F/R	TAGTTGCCATCCTCGGACTC/CAACGGTGCTATTCCATGTG	200

genes (for instance the strictly environmental *M. smegmatis* species, *Mycobacterium* sp. MCS, *Mycobacterium* sp. JLS, *Mycobacterium* sp. KMS and the Buruli ulcer agent *M. ulcerans*). Finally, the intracellular human pathogen *M. leprae* lacks the three cellulose-targeting genes altogether (Table 1).

The disposition of the three cellulose-targeting genes within the genomes of various MTC members was similar (Fig. 2). This similarity in organization was also true as for the size and the intergenic spacers between these genes. In *M. avium* 104 and *M. avium* subsp. *paratuberculosis* K10,

the intergenic spacer between *Cel6* and *CDB2* was 2.4 Mb, similar to that of *M. vanbaalenii*, while in *M. marinum*, it was two times larger. The orientation of these genes in the genome was similar in all MTC, MAC and some NTM members, except for *M. vanbaalenii* and *M. gilvum* (Fig. 2).

Cel6 and *CBD2* putative gene products were found to possess a signal peptide across the *Mycobacterium* genus, while the *Cel12* gene product lacked a signal peptide in all MTC members. We observed that the 31.27 kb region of the *M. marinum* genome that contains the *Cel6* gene was deleted in *M. ulcerans*; a similar observation was made

Fig. 1. Clustering of bacteria according to their content of cellulose-targeting genes. Only bacteria with one or more cellulose-targeting gene(s) are shown. Black arrows and numbers indicate the position of reference bacteria for identification of species in Supplementary Table S1.

Fig. 2. Disposition of *Cel6*, *Cel12* and *CBD2* genes within the genome of MTC, MAC and NTM members. Arrows indicate the 5'→3' orientation, and distances between adjacent genes are indicated (Mb).

when aligning the *Cel12* genomic region in *M. tuberculosis* H37Rv with the corresponding region in *M. marinum* and *M. avium* genomes, which lack this gene. Within MTC members, the *Cel6* gene interspecies nucleotide sequence similarity ranged from 97.6% between *M. microti* and *M. canettii* to 99.7% between *M. microti* and *M. caprae*. We found no deletion or insertion within the *Cel6* gene in the MTC reference strains. The *CBD2* gene sequence of *M. canettii* exhibited 99.4% similarity to the MTC reference *M. tuberculosis* H37Rv sequence, while *M. caprae*, *M. microti*, *M. pinnipedii* and *M. africanum* yielded 100% *CBD2* gene sequence similarity to the MTC reference sequence; *M. africanum*, *M. caprae*, *M. microti*, *M. pinnipedii* and *M. canettii* yielded identical sequences for the *Cel12* gene. These data show that the three cellulose-targeting genes are highly conserved among MTC members.

PCR amplification, sequencing and RT-PCR

PCR amplification and sequencing showed that *Cel6*, *Cel12* and *CBD2* genes were present in the genome of *M. canettii*, *M. africanum*, *M. pinnipedii*, *M. microti*, *M. caprae*, *M. tuberculosis* H37Rv and *M. bovis*, whereas the *M. bovis* BCG Pasteur strain lacked the *CBD2* gene. All original *Cel6*, *Cel12* and *CBD2* nucleotide sequences of non-sequenced MTC members have been deposited in GenBank (Table 3). In MAC strains, *Cel6* and *CBD2* were detected in *M. intracellulare*, *M. chimaera*, *M. colombiense* and *M. avium* subsp. *hominissuis*, while in NTM strains, the *Cel6* gene was present in *M. smegmatis* and *M. fortuitum*. Experimental PCR amplification and sequencing data obtained in MTC members correlated with *in silico* data using a BLAST search with the *M. tuberculosis* H37Rv nucleotide reference sequences of the *Cel6*, *Cel12* and *CBD2* genes (GenBank

Table 3. GenBank accession numbers of original sequences determined in this study

Species	Strain	GenBank accession number		
		<i>Cel6</i>	<i>Cel12</i>	<i>CBD2</i>
<i>M. canettii</i>	CIP140060001 ^T	FJ405033	GQ287404	FJ405035
<i>M. microti</i>	CIP104256	FJ405031	GQ287405	GQ287402
<i>M. africanum</i>	CIP105147 ^T (type 1)	FJ405034	GQ287403	GQ287399
<i>M. caprae</i>	CIP105776 ^T	GQ287408	GQ287406	GQ287400
<i>M. pinnipedii</i>	ATCC BAA-688	GQ287409	GQ287407	GQ287401

accession no. NC 000962) (Cole *et al.*, 1998) as query sequences against the non-redundant database of NCBI.

The absence of genomic DNA in our RNA samples was further confirmed by a lack of PCR amplification (data not shown). RT-PCR amplification using RNA isolated from *M. tuberculosis* H37Rv CIP103471 showed that *Cel6*, *Cel12* and *CBD2* were transcribed into mRNA. A single band of the expected (200 bp) size was detected (data not shown). The RT-PCR product was sequenced and the sequence yielded 100 % similarity to the *M. tuberculosis* H37Rv reference sequence. No PCR band was detected in DNase I-treated negative controls. This indicates that the *Cel6*, *Cel12* and *CBD2* genes present in the genome of *M. tuberculosis* H37Rv are transcribed into mRNA.

DISCUSSION

Using bioinformatics analyses we have observed the presence of a set of three cellulose-targeting genes in *Mycobacterium* species with a completely sequenced genome. The study was completed by PCR-based experiments to identify, amplify and determine the DNA and mRNA sequences of the three cellulose-targeting genes in 14 other *Mycobacterium* species. According to our database search, the set of three cellulose-targeting genes, *Cel6*, *Cel12* and *CBD2*, is present in only 2.5 % of the bacteria with a complete genome sequence (Table 1 and Supplementary Table S1). These bacteria include soil inhabitants and marine organisms in contact with decaying plants. Because cellulose almost never occurs as a pure component in plant cell walls, saprophytes have a complete spectrum of enzymes for all constituents of plant cell walls, and cellulolytic microbes usually produce a cohort of hemi-cellulase and pectinase enzymes along with their cellulases. In contrast, the MTC members stand out because the putative cellulose-targeting genes are not accompanied by other plant cell wall-digesting enzymes. The fact that in MTC members the cellulolytic machinery is so focused suggests that the plant cell wall may not be the actual target for cellulolytic activity in MTC organisms.

The lack of *Cel6*, *Cel12* and *CBD2* genes in the *M. leprae* genome is congruent with the severe genome reduction that accompanied the evolution of this species towards a strictly parasitic lifestyle (Cole *et al.*, 2001). Five *Mycobacterium* species (Table 1) encoded only one of the three cellulolytic-targeting genes, suggesting a limited activity towards cellulose, if any. The vast majority of *Mycobacterium* species (Table 1 and Supplementary Table S1) encoded *Cel6* and a putative *CBD2* gene. The fact that the MTC organisms (mammal-associated species responsible for tuberculosis) were the sole *Mycobacterium* organisms to exhibit the complete set of *Cel6*, *Cel12* and *CBD2* genes was unexpected. *M. bovis* has been shown to be transmitted from bovine to bovine and from bovine to humans (Ashford *et al.*, 2001) by the oral route, thus being potentially in close contact with cellulose-containing

foodstuffs in the digestive tract of vertebrates such as mice (Boulahrouf *et al.*, 1990) and pigs (Varel *et al.*, 1984). Likewise, we have recently shown that living *M. tuberculosis* organisms can be detected in the stools of patients with pulmonary tuberculosis (El Khechine *et al.*, 2009). *M. bovis*, however, is also efficiently transmitted directly by the aerosol route, being responsible for outbreaks of tuberculosis (Rodwell *et al.*, 2008) and bovine tuberculosis (Wilkins *et al.*, 2008). It has also been demonstrated that *M. bovis* can survive ingestion by amoebae, which suggests that protozoa could significantly enhance the survival of *M. bovis* in the soil and hence may be instrumental in the transmission of bovine tuberculosis (Taylor *et al.*, 2003). Likewise, *M. tuberculosis* is mainly transmitted directly by the aerosol route and has no known environmental reservoir (Riley *et al.*, 1995). However, the fact that these three cellulose-targeting genes are highly conserved in MTC organisms argues against the idea that these genes would be in a degenerate state.

A recent report has identified the *Legionella pneumophila* strain 130b *CelA* gene, which has been shown to degrade carboxymethyl cellulose and microcrystalline cellulose (Pearce & Cianciotto, 2009). Because *L. pneumophila* resides in amoebae which produce cellulose, it has been suggested that *CelA* could be involved in intra-amoeba growth of *Legionella*. However, it has been observed that an *L. pneumophila* *CelA*-deficient mutant and the wild-type strain both exhibit comparable growth in amoebae trophozoites and macrophages and comparable persistence in the lungs, indicating that *celA* is not required for intracellular life nor for lung infection (Pearce & Cianciotto, 2009). Based on these data, it has been suggested that the *CelA* protein could promote the growth of this organism in natural aquatic environments that contain higher levels and/or different types of polysaccharides, including cellulose produced by plants, amoebae or other bacteria.

Our observation that MTC organisms also possess cellulose-targeting genes in their genome is somewhat similar to this recent report. In *M. avium* 104, the *M. tuberculosis* 7.8 kb region that includes *Cel12* was deleted, leaving a region of 2.1 kb covered by two hypothetical proteins and one spacer (Fig. 3), neither of them matching the *M. tuberculosis* *Cel12a* and *Cel12b* genes, implying that these proteins were not *Cel12* orthologues. This observation suggests that the *Cel12* gene has been uniquely retained in the MTC strains from a *Mycobacterium* early ancestor. Phylogenetic trees based on *Mycobacterium* *Cel6* and *CBD2* sequences show that the eight MTC members cluster together (Supplementary Figs S1 and S2). This observation suggests that these genes are typical of MTC organisms and *Mycobacterium* and that no recent lateral gene transfer has occurred.

The capacity of a gene to respond to internal/external cues can be determined by its ability to be transcribed into mRNA. We have studied the transcription of the *Cel6*,

Fig. 3. Schematic diagram of the *Cel12* genomic region in *M. tuberculosis* H37Rv and *M. avium* 104, which lack *Cel12*. The region in *M. tuberculosis* H37Rv contains six genes in tandem, including *Cel12*. The homologous region in *M. avium* 104 contains only two genes, MAV_1212 and MAR_1213, separated by a spacer of 246 bp, pointing to the complete lack of the *M. tuberculosis* *Cel12* genome region in *M. avium* 104.

Cel12 and *CBD2* genes, and all three were transcribed into mRNA. The transcription of *Cel6* was predictable, as this gene has been cloned and the encoded protein is expressed in *Escherichia coli* in an active form (Varrot *et al.*, 2005). Moreover, the *Cel6* gene has been shown to be down-regulated in an *M. tuberculosis* mutant lacking *SigF* (Geiman *et al.*, 2004), indicating that the regulation of this gene is *SigF*-dependent. The *Cel12* gene lacks a peptide signal and exhibits a potentially fragmented structure which could suggest gene decay. It has been shown that the second part of the *Cel12* gene, namely *CelA2b*, is upregulated upon treatment of *M. tuberculosis* H37Rv with mefloquine (a derivative of 4-quinoline methanol) for 24 h (Danelishvili *et al.*, 2005). This observation suggests that even though the *M. tuberculosis* *Cel12* gene presents an unusual structure, it is implicated in the mechanisms of resistance to mefloquine in *M. tuberculosis*. Many cellulases share a common basic architecture that comprises a catalytic domain linked to a cellulose binding domain (CBD), which determines the efficiency of degradation of insoluble cellulose. In our study we found that *M. tuberculosis* *CBD2* was transcribed into mRNA, and this gene also possesses a signal peptide, which certainly points to an extracellular role.

Bioinformatics searches combined with PCR-based analyses have revealed that a set of three cellulose-targeting genes (*Cel6*, *Cel12* and *CBD2*) is present in the genome of MTC members. The only exception is the *M. bovis* BCG Pasteur strain, in which the *CBD2* gene has been lost along with several other portions of DNA. We have verified that the cellulose-targeting genes are transcribed into mRNA and it has been previously shown that the *Cel6*-encoded protein can be expressed as an active protein (Varrot *et al.*, 2005). Preliminary observations from this group suggest that *Cel12* and *CBD2* can also be cloned and the proteins expressed in *E. coli* (F. Mba Medie and others, unpublished observations).

It has been shown that examination of the complete genome of a fastidious micro-organism can help to design

a suitable culture medium (Renesto *et al.*, 2003). Here we show that analysis of the genome sequence can also reveal unexpected metabolic traits or abilities. Indeed, the presence of cellulose-targeting proteins reported here raises the question of a possible, yet unknown, environmental stage in MTC bacteria, perhaps as soil saprophytes or as hosts of cellulose-producing organisms such as amoebae, as suggested for *M. bovis* (Hagedorn *et al.*, 2009; Taylor *et al.*, 2003).

ACKNOWLEDGEMENTS

F.M.M. is the recipient of an Infectiopole Sud doctoral fellowship. The enthusiastic support of Didier Raoult is gratefully acknowledged.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Ashford, D. A., Whitney, E., Raghunathan, P. & Cosivi, O. (2001). Epidemiology of selected mycobacteria that infect humans and other animals. *Rev Sci Tech* **20**, 325–337.
- Boulahrouf, A., Fonty, G. & Gouet, P. (1990). Establishment of cellulolytic bacteria in the digestive tract of conventionally reared young mice: effect of the dietary cellulose content in the adult. *FEMS Microbiol Lett* **57**, 87–90.
- Brosch, R., Gordon, S. V., Garnier, T., Eiglmeier, K., Frigui, W., Valenti, P., Dos Santos, S., Duthoy, S., Lacroix, C. & other authors (2007). Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci U S A* **104**, 5596–5601.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233–D238.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S. & other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.
- Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007–1011.

- Danelishvili, L., Wu, M., Young, L. S. & Bermudez, L. E. (2005). Genomic approach to identifying the putative target of and mechanisms of resistance to mefloquine in mycobacteria. *Antimicrob Agents Chemother* **49**, 3707–3714.
- Doi, R. H. & Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat Rev Microbiol* **2**, 541–551.
- El Khechine, A., Henry, M., Raoult, D. & Drancourt, M. (2009). Detection of *Mycobacterium tuberculosis* complex organisms in the stools of patients with pulmonary tuberculosis. *Microbiology* **155**, 2384–2389.
- Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C. & other authors (2003). The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A* **100**, 7877–7882.
- Geiman, D. E., Kaushal, D., Ko, C., Tyagi, S., Manabe, Y. C., Schroeder, B. G., Fleischmann, R. D., Morrison, N. E., Converse, P. J. & other authors (2004). Attenuation of late-stage disease in mice infected by the *Mycobacterium tuberculosis* mutant lacking the SigF alternate sigma factor and identification of SigF-dependent genes by microarray analysis. *Infect Immun* **72**, 1733–1745.
- Gilbert, H. J., Stalbrand, H. & Brumer, H. (2008). How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr Opin Plant Biol* **11**, 338–348.
- Hagedorn, M., Rohde, K. H., Russell, D. G. & Soldati, T. (2009). Infection by tubercular mycobacteria is spread by nonlytic ejection from their amoeba hosts. *Science* **323**, 1729–1733.
- Li, L., Bannantine, J. P., Zhang, Q., Amonsin, A., May, B. J., Alt, D., Banerji, N., Kanjilal, S. & Kapur, V. (2005). The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc Natl Acad Sci U S A* **102**, 12344–12349.
- Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**, 506–577.
- Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. (1996). Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* **178**, 1274–1282.
- Pearce, M. M. & Cianciotto, N. P. (2009). *Legionella pneumophila* secretes an endoglucanase that belongs to the family-5 of glycosyl hydrolases and is dependent upon type II secretion. *FEMS Microbiol Lett* **300**, 256–264.
- Renesto, P., Crapoulet, N., Ogata, H., La, S. B., Vestris, G., Claverie, J. M. & Raoult, D. (2003). Genome-based design of a cell-free culture medium for *Tropheryma whippelii*. *Lancet* **362**, 447–449.
- Riley, R. L., Mills, C. C., Nyka, W., Weinstock, N., Storey, P. B., Sultan, L. U., Riley, M. C. & Wells, W. F. (1995). Aerial dissemination of pulmonary tuberculosis. A two-year study of contagion in a tuberculosis ward. 1959. *Am J Epidemiol* **142**, 3–14.
- Rodwell, T. C., Moore, M., Moser, K. S., Brodine, S. K. & Strathdee, S. A. (2008). Tuberculosis from *Mycobacterium bovis* in binational communities, United States. *Emerg Infect Dis* **14**, 909–916.
- Seki, M., Honda, I., Fujita, I., Yano, I., Yamamoto, S. & Koyama, A. (2009). Whole genome sequence analysis of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG) Tokyo 172: a comparative study of BCG vaccine substrains. *Vaccine* **27**, 1710–1716.
- Stinear, T. P., Seemann, T., Pidot, S., Frigui, W., Reyssset, G., Garnier, T., Meurice, G., Simon, D., Bouchier, C. & other authors (2007). Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res* **17**, 192–200.
- Stinear, T. P., Seemann, T., Harrison, P. F., Jenkin, G. A., Davies, J. K., Johnson, P. D., Abdellah, Z., Arrowsmith, C., Chillingworth, T. & other authors (2008). Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res* **18**, 729–741.
- Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C. & Davies, G. J. (1997). The *Streptomyces lividans* family 12 endoglucanase: construction of the catalytic core, expression, and X-ray structure at 1.75 Å resolution. *Biochemistry* **36**, 16032–16039.
- Taylor, S. J., Ahonen, L. J., de Leij, F. A. & Dale, J. W. (2003). Infection of *Acanthamoeba castellanii* with *Mycobacterium bovis* and *M. bovis* BCG and survival of *M. bovis* within the amoebae. *Appl Environ Microbiol* **69**, 4316–4319.
- Varel, V. H., Fryda, S. J. & Robinson, I. M. (1984). Cellulolytic bacteria from pig large intestine. *Appl Environ Microbiol* **47**, 219–221.
- Varner, J. E. & Lin, L. S. (1989). Plant cell wall architecture. *Cell* **56**, 231–239.
- Varrot, A., Leydier, S., Pell, G., Macdonald, J. M., Stick, R. V., Henrissat, B., Gilbert, H. J. & Davies, G. J. (2005). *Mycobacterium tuberculosis* strains possess functional cellulases. *J Biol Chem* **280**, 20181–20184.
- Wilkins, M. J., Meyerson, J., Bartlett, P. C., Spieldenner, S. L., Berry, D. E., Mosher, L. B., Kaneene, J. B., Robinson-Dunn, B., Stobierski, M. G. & Boulton, M. L. (2008). Human *Mycobacterium bovis* infection and bovine tuberculosis outbreak, Michigan, 1994–2007. *Emerg Infect Dis* **14**, 657–660.

Edited by: D. W. Ussery