**DprE1, a new taxonomic marker in mycobacteria**

Maria Loreto Incandela¹, Elena Perrin², Marco Fondi², Ana Luisa de Jesus Lopes Ribeiro¹, Giorgia Mori¹, Alessia Moiana³, Maurizio Gramegna³, Renato Fani², Giovanna Riccardi¹ & Maria Rosalia Pasca¹

¹Department of Biology and Biotechnology “Lazzaro Spallanzani”, University of Pavia, Pavia, Italy; ²Department of Biology, University of Florence, Florence, Italy; and ³Sentinel CH, Milan, Italy

**Abstract**

Among the species of the *Mycobacterium* genus, more than 50 have been recognized as human pathogens. In spite of the different diseases caused by mycobacteria, the interspecies genetic similarity ranges from 94% to 100%, and for some species, this value is higher than in other bacteria. Consequently, it is important to understand the relationships existing among mycobacterial species. In this context, the possibility to use *Mycobacterium tuberculosis dprE1* gene as a new phylogenetic/taxonomic marker has been explored. The *dprE1* gene codes for the target of benzothiazinones, belonging to a very promising class of antitubercular drugs. Mutations in cysteine 387 of DprE1 are responsible for benzothiazinone resistance. The DprE1 tree, obtained with 73 amino acid sequences of mycobacterial species, revealed that concerning the benzothiazinone sensitivity/resistance, it is possible to discriminate two clusters. To validate it, a concatamer obtained from the amino acid sequences of nine mycobacterial housekeeping genes was performed. The concatamer revealed that there is no separation between the benzothiazinone-susceptible and benzothiazinone-resistant species; consequently, this parameter is not linked to the phylogeny. DprE1 tree might represent a good taxonomic marker for the assignment of a mycobacterial isolate to a species. Moreover, the concatamer represents a good reference phylogeny for the *Mycobacterium* genus.

**Introduction**

The *Mycobacterium* genus includes several medically important species that constitute an alarming toll in human mortality. Among the species of the *Mycobacterium* genus, more than 50 have been recognized as potential human pathogens.

The species belonging to the *Mycobacterium tuberculosis* complex (MTBC) are the most known and include the human pathogens *M. tuberculosis*, *Mycobacterium africans*, and *Mycobacterium canetti* and the animal-adapted pathogens *Mycobacterium bovis*, *Mycobacterium microti*, *Mycobacterium caprae*, and *Mycobacterium pinnipedii* as well as the recently discovered species *Mycobacterium mungi* (Bouakaze et al., 2011).

*Mycobacterium avium–intracellulare* complex (MAC) is another important group including nontuberculous mycobacteria responsible for opportunistic infections in immunocompromised individuals (Field & Cowie, 2006), while the unculturable *Mycobacterium leprae* persists in developing countries, and it is the causative agent of leprosy (Suzuki et al., 2012).

In spite of the different diseases that can be caused by mycobacteria, the interspecies genetic similarity ranges from 94% to 100%, and for some mycobacterial species, this value is higher than in other bacteria (Devulder et al., 2005). Therefore, further analysis would be important to improve our understanding of mycobacterial taxonomic identity, in the context of evolution and speciation.

Even though the 16S rRNA gene is the most used molecular marker for phylogenetic analysis in bacteria, alternative markers have been proposed for mycobacteria, such as *hsp65*, *recA*, *sodA*, and *rpoB* genes (Adékambi & Drancourt, 2004). However, the phylogenetic analysis obtained with these markers individually is limited because no gene amplification is obtained for a few
species, and some closely related species are difficult to differentiate, like those belonging to MTBC complex (Mignard & Flandrois, 2007).

Recently, multigene sequence analysis was applied to mycobacteria, revealing novel insights into the phylogenetic relationships between the various Mycobacterium species. Devulder et al. (2005) developed a multigene sequence database incorporating four genes (16S rRNA gene, hsp65, rpoB, and sodA) within the Mycobacterium genus. The concatenation of four genes provides a good tool of increasing the robustness of the final tree, but presented some inaccuracies, especially for MTBC complex (Devulder et al., 2005). Another tree based on the combination of 16S rRNA gene, rpoB, recA, hsp65, and sodA was performed by Adékambi & Drancourt (2004) with good bootstrap support; however, the same authors showed that the trees based only on one of these genes were not very robust. Recently, the tuf gene coding for EF-TU factor was proposed as phylogenetic marker, and the corresponding tree was quite robust (Mignard & Flandrois, 2007).

Therefore, phylogenetic analyses based on the combined dataset of a panel of gene sequences could help to delineate new mycobacterial species, as well as enabling the monitoring of drug resistance-conferring mutations (Adékambi & Drancourt, 2004). In this context, the M. tuberculosis dprE1 might represent an interesting candidate as it is an essential gene (Sassetti & Rubin, 2003), and it encodes a ‘hot’ target of five new antitubercular agents, including the benzothiazinones (BTZs; Makarov et al., 2009; Christophe et al., 2009; Magnet et al., 2010; Stanley et al., 2012; Wang et al., 2013). DprE1 enzyme works in concert with DprE2, and it is involved in the biosynthesis of arabinogalactan, an essential component of the mycobacterial cell wall core (Wolucka, 2008). It has been demonstrated that point mutations responsible for the substitution of Cys387 residue of M. tuberculosis DprE1 are responsible for BTZ resistance (Makarov et al., 2009). This cysteine residue is highly conserved in orthologous DprE1 proteins from various BTZ-susceptible Actinobacteria; on the other site, in Mycobacterium avium and Mycobacterium aurum, the Cys387 residue is replaced by serine or alanine, respectively (Makarov et al., 2009); this achievement renders bacteria belonging to these species naturally resistant to BTZ. Accordingly, DprE1 might represent a valuable phylogenetic and/or taxonomic marker, and its possible role in this context was investigated in this work.

Materials and methods

Bacterial strains and growth conditions

The type strains of nine mycobacterial species, Mycobacterium africanum (ATCC 25420), Mycobacterium xenopi (ATCC 192509), Mycobacterium intracellulare (ATCC 13209), Mycobacterium avium subsp. paratuberculosis (ATCC 19698), Mycobacterium scrofulaceum (ATCC 19073), Mycobacterium chelonae (ATCC 14472), Mycobacterium celatum (ATCC 51130), Mycobacterium gastri (ATCC 15754), and Mycobacterium simiae (ATCC 25273) were purchased from the American Type Culture Collection (ATCC). These strains were grown either in Middlebrook 7H9 broth (Difco) with 0.05% Tween 80 or on Middlebrook 7H11 agar (Difco) with 0.5% glycerol, both supplemented with 10% (vol/vol) OADC or on Lowenstein–Jensen medium, following the instructions of ATCC Web site (http://www.lgcstandards-atcc.org/). Mycobacterial cultures were usually grown at 37 °C without shaking for 3–4 weeks, with the exception of M. chelonae, which was grown in the same conditions for about 4 days.

PCR primers and cloning

Gene-specific and species-specific PCR primers were designed (Table 1) and dprE1 orthologous genes were amplified by PCR, using the cell lysates of each mycobacterial strain as template. PCR experiments were carried out using STAT-NAT DNA-Mix kit (Stabilized Amplification Technology Nucleic Acid Testing, Sentinel CH SpA, Italy; composition: 3.0 mM MgCl2, 0.8 mM dNTPs, 2u Hot Start Taq Polymerase). Amplicons were purified using the Wizard SV Gel and PCR clean-up system (Promega) and then cloned in pGEM-T Easy vector (Promega). The nucleotide sequence of the insert of recombinant plasmids (dprE1/pGEM-T) was determined by Sanger sequencing (www.bmr-genomics.it/).

The nine dprE1 gene sequences not available in databases were deposited at NCBI Web site (http://www.ncbi.nlm.nih.gov/) and were assigned the accession numbers reported in Table 1.

Sequence analysis

Amino acid sequences of putative DprE1 proteins were retrieved using the M. tuberculosis DprE1 amino acid sequence (GI:15610926) as a query to probe the nonredundant protein sequences (nr) database at NCBI site using BLASTP (Altschul et al., 1997). Only those sequences retrieved at an E-value below the 0.05 threshold were taken into account. Amino acid sequences of Hsp65 and RpoB proteins were retrieved from the NCBI database and trimmed as found in bibliography (Telenti et al., 1993; Adékambi et al., 2003). The MUSCLE program was used to perform the amino acid multialignment (Edgar, 2004). Alignments were manually checked, and misaligned regions were removed.
The ConSurf server was used for the evaluation of the evolutionary conservation of amino acid positions in the DprE1 proteins, based on the phylogenetic relations between orthologous sequences (Glaser et al., 2003). The multialignment of the 73 DprE1 amino acid sequences was used to perform a phylogenetic tree by the neighbor-joining algorithm as implemented in the Rate4Site program (Pupko et al., 2002). Position-specific conservation scores were then computed by the empirical Bayesian approach (Mayrose et al., 2004). Finally, the conservation scores were projected onto the Mycobacterium smegmatis DprE1 protein structure (PDB: 4f4q; Neres et al., 2012).

Phylogenetic analysis

Neighbor-joining (NJ) phylogenetic trees (Saitou & Nei, 1987) were obtained with Mega5, using pairwise deletion option and 1000 bootstrap replicates (Tamura et al., 2011). A concatamer was obtained adopting the following procedure: (1) the orthologs of FusA (protein chain elongation factor EF-G), IleS (isoleucyl-tRNA synthetase), LepA (back-translocating elongation factor EF4), LeuS (leucyl-tRNA synthetase), PyrG (CTP synthetase), RecA (recombinase A), RecG (ATP-dependent DNA helicase), RplB (50S ribosomal protein L2) and RpoB (RNA polymerase beta subunit; Santos & Ochman, 2004) of M. tuberculosis H37Rv strain [sequences recovered from the RDP Resource Download Area at Ribosomal Database project site (http://rdp.cme.msu.edu/)] were retrieved from 46 mycobacterial genomes; (2) each ortholog dataset was independently aligned; and (3) all the different multialignments were concatenated in a single one comprising 7833 residues. The concatenated sequences of the same genes of Corynebacterium efficiens YS314 were used as an out-group.

Results and discussion

PCR amplification and sequencing of dprE1 orthologous genes from strains belonging to nine Mycobacterium species

Nine dprE1 genes of M. africanum, M. xenopi, M. intracellulare, M. avium subsp. paratuberculosis, M. scrofulaceum, M. chelonae, M. celatum, M. gastri, and M. simiae species were amplified by PCR using ad hoc-designed gene-specific primers reported in Table 1. The dprE1 amplicons obtained were then cloned in pGEM-T Easy plasmid vector, and their nucleotide sequences were determined as described in Materials and methods. The nine dprE1 gene sequences obtained were submitted to NCBI Web site, and the corresponding accession numbers were reported in Table 1. These sequences were utilized for further analyses.
Identification of both DprE1 mycobacterial proteins and putative mycobacterial BTZ-susceptible and BTZ-resistant species

To check the phylogenetic distribution of the DprE1-like proteins in the entire Mycobacterium genus, the M. tuberculosis DprE1 amino acid sequence (GI:15610926) was used as a query to probe the nr NCBI database. In this way, a total of 64 sequences homologous to M. tuberculosis DprE1 were retrieved and aligned with the nine ones obtained in this work (DprE1 of M. africanum, M. xenopi, M. intracellulare, M. avium subsp. paratuberculosis, M. scrofulaceum, M. chelonae, M. celatum, M. gastr, and M. simiae). The corresponding multialignment of 73 DprE1 amino acid sequences is reported as Supporting Information, Data S1. The analysis revealed a high degree of sequence conservation between DprE1 proteins belonging to different mycobacterial species. This similarity is highlighted in Fig. 1, obtained using the ConSurf server (Glaser et al., 2003). In this picture, the conservation scores of each amino acid at each positions are projected onto the M. smegmatis DprE1 protein structure (PDB: 4f4q; Neres et al., 2012). In Fig. 1b, all residues having the highest degree of conservation are highlighted.

To identify the conservation degree of amino acids located into the DprE1 active site, twelve residues from M. smegmatis DprE1 structure (Tyr67, His139, Gly140, Lys141, Lys425, Gln341, Gln343, Leu370, Lys374, Phe376, Asn392, and Cys394) were analyzed in the 73 mycobacterial species (Neres et al., 2012). Eleven of these twelve residues are conserved in all analyzed sequences; the only exception is represented by Cys394 (Cys387 in M. tuberculosis) that, in some sequences, is replaced by an alanine. Because the presence of cysteine (Cys394 in M. smegmatis and Cys387 in M. tuberculosis) is clearly associated to BTZ sensitivity (Makarov et al., 2009; Neres et al., 2012), the 11 mycobacterial species (M. abscessus, M. massiliense, M. chelonae, M. rhodesiae, M. tusciae, M. neoaurum, M. parascrofulaceum, M. avium subsp. avium, M. avium subsp. paratuberculosis, M. colombiense, and M. intracellulare), having an Ala replacing a Cys in this position, are very likely resistant to BTZs. This result is in agreement with previous experimental findings, showing that strains belonging to the two species M. avium and M. aurum are naturally resistant to BTZs (Makarov et al., 2009).

It is noteworthy that among the 11 conserved amino acids, His139, Lys425, and Gln343 residues are essential for FAD binding and critical for full enzyme activity, as it was demonstrated by structural and enzymatic assays (Neres et al., 2012).

DprE1 phylogenetic analysis

The multialignment of DprE1 amino acid sequences (Data S1) was used to build the neighbor-joining tree shown in Fig. 2. The analysis of the DprE1 tree revealed that different strains of the same species shared a high degree of sequence similarity and were clustered together in the tree, except for the Mycobacterium rhodesiae and M. xenopi strains (Fig. 2). At the same time, each species is clearly separated from each other. The topology of the DprE1 tree is different from those obtained with other molecular markers for the branching order of some mycobacterial species (Tortoli, 2012). However, most nodes in the DprE1 tree were supported by high bootstrap values (Fig. 2).

All these data strongly suggest the possibility to use DprE1 as a taxonomic marker for identifying/clustering strains belonging to the same mycobacterial species. Nonetheless, in our opinion, the most important feature of the DprE1 phylogenetic analysis is related to the BTZ sensitivity/resistance of mycobacterial strains. In fact, in

Fig. 1. Conservation scores of amino acid at each position projected onto the Mycobacterium smegmatis DprE1 protein structure. (a) Conservation scores, obtained with ConSurf server (Glaser et al., 2003), of each amino acid at each positions are projected onto the M. smegmatis DprE1 protein structure (PDB: 4f4q; Neres et al., 2012). The continuous conservation scores are divided into a discrete scale of nine grades for visualization, from the most variable positions (grade 1) colored turquoise, through intermediately conserved positions (grade 5) colored white, to the most conserved positions (grade 9) colored maroon. In (b) Only residues having the highest degree of conservation are highlighted.
Fig. 2. Phylogenetic tree constructed with the amino acid sequences of 73 homologous to Mycobacterium tuberculosis DprE1. Neighbor-joining (NJ) phylogenetic tree was obtained with Mega5 (Tamura et al., 2011), pairwise deletion option, and 1000 bootstraps replicates. C, cysteine; A, alanine.
the DprE1 tree, it is possible to discriminate two different clusters, each of which including mycobacterial species having Cys or Ala in position 387. It is underlined that beside BTZs, other three molecules, DNB1, VI-9376, and 377790, have been published to form covalent bonds with the cysteine residue within the active site of DprE1, thus blocking the enzymatic activity (Christophe et al., 2009; Makarov et al., 2009; Magnet et al., 2010; Stanley et al., 2012). Until now, only one DprE1 inhibitor is able to form a noncovalent binding with Cys387 (Wang et al., 2013). The first cluster (having Cys387) comprised the following species: MTBC, *M. leprae*, *M. ulcerans*, *M. marinum*, *M. kansasi*, *M. xenopi*, *M. celatum*, *M. gastri*, *M. simiae*, *M. phlei*, *M. smegmatis*, *M. thermoresistibile*, *M. chubuense*, *M. vanbaliensi*, and *M. gilvum*. On the basis of the available data, strains belonging to these species might be susceptible to BTZs as well as to the other three DprE1 inhibitors. In fact, it is noteworthy that *M. tuberculosis* H37Rv (moreover 240 clinical isolates comprising MDR and XDR strains), *M. bovis*, *M. smegmatis*, and *M. marinum* are susceptible to BTZs (Makarov et al., 2009; Pasca et al., 2010). The second cluster (exhibiting an Ala387) embedded the following species whose

![Phylogenetic tree constructed using the concatenated sequences of nine proteins belonging to 46 different Mycobacterium strains. The concatamer was obtained using the orthologs of FusA, IleS, LepA, LeuS, PyrG, RecA, RecG, RplB, and RpoB of *M. tuberculosis* H37Rv strain.](image-url)
representatives might be resistant to BTZs: M. abscessus, M. massiliense, M. chelonae, M. abscessus subsp. bolletii, M. rhodesiae, M. tusciae, M. neoaurum, M. parascrofulaceum, and MAC (Fig. 2). Among these cited species, M. avium was tested to be resistant to BTZs (Makarov et al., 2009). The presence of either a Cys or an Ala residue might, in principle, be a powerful tool to indicate whether a mycobacterial species is susceptible/resistant to BTZs and to the other DprE1 inhibitors DNB1, VI-9376, and 377790.

Consequently, the amino acid localized at position 387 should be critical for the resistance/sensitivity to BTZs. To check the influence of such residues (Cys or Ala) at position 387 on the topology of the DprE1 tree, an additional phylogenetic tree was constructed using the original DprE1 amino acid sequences and the ‘chimeric’ ones where the Cys387 was replaced by an Ala and vice versa (Data S2). The analysis of the DprE1 phylogenetic tree embedding also the ‘chimeric’ sequences revealed that, as might be expected, its topology was identical to that of the tree constructed with the original sequences (Data S2; Fig. 2).

Phylogenetic analysis

The analysis of DprE1 tree raised the question of the possible congruence existing between these data and the overall Mycobacterium phylogeny, an issue yet under debate. Indeed, the Mycobacterium genus is characterized by a very limited interspecies genetic variability, and this is the cause of a problematic phylogenetic reconstruction (Tortoli, 2012).

Consequently, a phylogenetic tree was constructed from the alignment of the concatenated amino acid sequences of the products of nine housekeeping genes (Santos & Ochman, 2004), as recommended on the site of Ribosomal Database Project as universally conserved genes that can be used for identification and phylogenetic analysis of bacteria (see Materials and methods; Data S3) from 46 Mycobacterium strains (Santos & Ochman, 2004; Fig. 3).

The analysis of the concatamer tree revealed that strains belonging to the same species were clustered together, whereas different species are clearly separated. It is quite interesting that nodes separating different species are supported by very high bootstrap values (99–100%; Fig. 3). Both the robust topology and high bootstraps suggest that the concatamer tree is much more reliable for other trees constructed using single genes, such as those based on complete 16S rRNA gene sequences or on the amino acid sequences coded by fragments of hsp65 and rpoB genes (Telenti et al., 1993; Adékambi et al., 2003; Data S4).

The only exception is represented by M. rhodesiae J60 and NBB3 strains that are not clustered together (Fig. 3), like in the DprE1 tree, previously described (Fig. 2). Also in this tree, the related species were clustered in MTBC or MAC complexes (Fig. 3).

Moreover, there is no a separation between the mycobacterial BTZ-susceptible and BTZ-resistant species. Consequently, this parameter is not linked to the phylogeny of mycobacteria.

Conclusions

The aim of this work was to check the possibility of using the dprE1 gene as a new molecular marker for taxonomical and/or phylogenetic studies of mycobacteria. The whole body of data obtained suggested that DprE1 tree might represent an additional good taxonomic marker for the assignment of a mycobacterial isolate to a given species. In addition, the same marker might also give insights into sensitivity/resistance of mycobacterial isolates to BTZs and to other drugs hitting DprE1 enzyme, simply checking for the presence/absence in position 387 of Cys residue, respectively.

Lastly, the phylogenetic tree, constructed using a concatamer of nine housekeeping genes, was supported by very high bootstrap values. Consequently, this tree represents a good reference phylogeny for the Mycobacterium genus.

Acknowledgement

This work was supported by European Commission (VII Framework, contract no. 260872).

Authors’ contribution

M.L.I. and E.P. contributed equally to this work.

References


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. DprE1 sequences multialignment (Multialignment of 73 DprE1 amino acid sequences used in this work).

Data S2. Phylogenetic tree of DprE1 chimeric sequences (Phylogenetic tree constructed with both the original DprE1 sequences and the chimeric DprE1 sequences).

Data S3. Concatenated sequences multialignment (Multialignment of the concatenated amino acid sequences of nine housekeeping genes of 46 Mycobacterium strains).

Data S4. Phylogenetic tree of the 16S rRNA gene, Hsp65 and RpoB sequences of 46 Mycobacterium strains (Description of data: Neighbor-Joining (NJ) phylogenetic trees were obtained with Mega5 (Tamura et al., 2011), pairwise deletion option and 1000 bootstraps replicates. Phylogenetic trees were constructed using: (a) the complete 16S rRNA gene sequences; (b) the amino acid sequences coded by fragments of Hsp65 (Telenti et al., 1993); (c) RpoB (Adékambi et al., 2003) proteins of 46 Mycobacterium strains whose genome was present in the database).