GBE

The *Mycobacterium phlei* Genome: Expectations and Surprises

Sarbashis Das[1], B. M. Fredrik Pettersson[1], Phani Rama Krishna Behra[1], Malavika Ramesh[1], Santanu Dasgupta[1], Alok Bhattacharya[2,3], and Leif A. Kirsebom[1,*]

[1]Department of Cell and Molecular Biology, Box 596, Biomedical Centre, Uppsala, Sweden

[2]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

[3]School of Life Sciences, Jawaharlal Nehru University, New Delhi, India

*Corresponding author: E-mail: leif.kirsebom@icm.uu.se.

Accepted: February 27, 2016

Data deposition: This project has been deposited at GenBank under the accessions: CP014475, ANBO00000000, JPUH00000000, ATHW00000000, and ATDR00000000.

Abstract

Mycobacterium phlei, a nontuberculosis mycobacterial species, was first described in 1898–1899. We present the complete genome sequence for the *M. phlei* CCUG21000^T type strain and the draft genomes for four additional strains. The genome size for all five is 5.3 Mb with 69.4% Guanine-Cytosine content. This is $\approx$0.35 Mbp smaller than the previously reported *M. phlei* RIVM draft genome. The size difference is attributed partly to large bacteriophage sequence fragments in the *M. phlei* RIVM genome. Comparative analysis revealed the following: 1) A CRISPR system similar to Type 1E (*cas*3) in *M. phlei* RIVM; 2) genes involved in polyamine metabolism and transport (*potAD*, *potF*) that are absent in other mycobacteria, and 3) strain-specific variations in the number of σ-factor genes. Moreover, *M. phlei* has as many as 82 *mce* (mammalian cell entry) homologs and many of the horizontally acquired genes in *M. phlei* are present in other environmental bacteria including mycobacteria that share similar habitat. Phylogenetic analysis based on 693 *Mycobacterium* core genes present in all complete mycobacterial genomes suggested that its closest neighbor is *Mycobacterium smegmatis* JS623 and *Mycobacterium rhodesiae* NBB3, while it is more distant to *M. smegmatis* mc2 155.

Key words: Mycobacterium phlei genome sequence, mycobacterial growth, comparative genome analysis, mycobacterial phylogeny.

Introduction

The grass bacillus, *Mycobacterium phlei*, was first described in 1898–1899 as a member of the order Actinomycetales and it is found in the environment (Gordon and Smith 1953; Wayne et al. 1969; Stackebrandt et al. 1981). *Mycobacterium phlei* belongs to the rapidly growing mycobacteria and it can grow at 52 °C (Gordon and Mihm 1959a; Saito et al. 1977). It was used as an early model system to study the biology of mycobacteria. The mycobacteria-specific iron-chelating compound mycobactin was first identified in *M. phlei* (Francis et al. 1953). It is rod shaped but earlier reports showed that *M. phlei* is pleiomorphic and can exist in a coccoid form under certain environmental conditions (Wyckoff and Smithburn 1933; Gordon and Mihm 1959b; Juhasz 1962; Csillag 1970). The coccoid form represented a resting stage in aging cultures as suggested by "Time lapse" microscopy; when exposed to fresh media these coccoid forms reverted back to rod-shaped bacteria (Wyckoff and Smithburn 1933). As other *Mycobacterium* spp. it also forms biofilms (Bardouniotis et al. 2001). It is considered to be nonpathogenic but *M. phlei* can cause infections (Aguilar et al. 1989; Spiegl and Feiner 1994; Paul and Devarajan 1998; Karnam et al. 2011). Interestingly, the *M. phlei* cell wall DNA complex (MCC) has been shown to promote anticancer activity against a wide range of cancer cell lines and MCC has been included as an adjuvant in anticancer vaccines (Filion and Phillips 2001). On the basis of 16S ribosomal DNA (rDNA) gene sequences, *M. phlei* has been positioned close to *Mycobacterium smegmatis* (Pitulle et al. 1992).

The importance of this group of bacteria that includes both environmental and highly pathogenic species such as *Mycobacterium tuberculosis*, the causative agent of tuberculosis, provided incentive for a comparative genomic analysis of different *M. phlei* strains. This would expand our knowledge about the genomic content of one member of this group of bacteria and provide insight into its evolutionary path.

We provide the complete genome sequence of one *M. phlei* type strain and the draft genomes for four additional strains. Comparative genomic analysis, including the published draft genome of the *M. phlei* RIVM strain (Abdallah et al. 2012), revealed the presence of common, as well as, strain-specific genes. The genome of the latter strain is substantially larger than the five *M. phlei* genomes presented here suggesting that these strains might represent the *M. phlei* group better than the RIVM strain. Interestingly, genes involved in polyamine synthesis are present in *M. phlei* but were not identified in other *Mycobacterium* species.

Materials and Methods

Genome Sequencing, Assembly, and Annotation

The genome of *M. phlei* CCUG21000^T (MPHL21000^T; ^{T} refers to "type" strain) was sequenced at the NGI-Uppsala Genome Center (PacBio technology), while the *M. phlei* DSM43239^T, DSM43070, DSM43071, and DSM43072 genomes (referred to as MPHL43239^T, MPHL43070, MPHL43071, and MPHL43072) were done at the SNP&SEQ Technology Platform (HiSeq2000—Illumina—platform) at Uppsala University.

The PacBio-generated reads were assembled using the SMRT-analysis HGAP3 assembly pipeline (Chin et al. 2013) and polished using Quiver (Pacific Biosciences, Menlo Park, CA). Assembly of the Illumina-generated reads was done using SOAPdenovo (version 1.05) (Li et al. 2010) with a minimum contig size of 200 bases. Whole-genome alignment of assembled genomes were generated using the MAUVE program (Darling et al. 2004).

The genomes were annotated and functionally classified into different subsystems (functional roles) using Rapid Annotation using Subsystem Technology (Aziz et al. 2008, see also Das et al. 2015). Noncoding RNA genes were predicted using the INFERence RNA ALignment tool (INFERNAL 1.1), and the Rfam database (version 11.0) with a minimum energy cutoff at 34 (Nawrocki and Eddy 2013).

For further details, see supplementary information.

Plasmids and Foreign DNA

To predict the presence of plasmid fragments, the scaffolds of the five *M. phlei* genomes were aligned pairwise using the NCBI plasmid database (ftp://ftp.ncbi.nlm.nih.gov/genomes/Plasmids/, last accessed March 2015).

Prophage sequences were predicted using the PHAST server (Zhou et al. 2011).

Identification of Orthologous Genes

To predict orthologous genes present in the six genomes, we used PanOCT (version 1.09) (Fouts et al. 2012) which uses sequence homology and gene synteny to classify a gene as orthologous. The parameters used are sequence identity $\geq$45%, query coverage $\geq$70%, and *e*-value cutoff 1×10^{-5}.

Horizontal Gene Transfer

To predict horizontally transferred genes, we used the HGTector software, which follows a hybrid between "BLAST-based" and phylogenetic approaches, with the following stringency criteria: *e*-value set at $<1 \times 10^{-100}$ for the BLAST hits, self = *Mycobacterium* (taxonomic_id 1763), and close = *Corynebacteriales* (taxonomic_id 85007) groups (Zhu et al. 2014). The distal group includes all other organisms that are phylogenetically distant to *M. phlei*. Note that BLAST hits with organism names related to phage and plasmid are not included in the analysis. Common and unique putative HGT genes among the six genomes were identified using BLASTP with percentage identity of $\geq$45% and query coverage of >70%.

Phylogenetic Analysis

Phylogenetic analysis was performed using 1) 16S rDNA and 2) 693 *Mycobacterium* core genes from 36 complete genomes as of June 2015 (supplementary table S1, Supplementary Material online; Das et al. 2015). Briefly, 16S rDNA sequences for *M. phlei* and other *Mycobacterium* spp. were aligned using MAFFT (version 5; Katoh et al. 2005). Phylogenetic trees were computed using the neighbor-joining method from the multiple sequence alignment. For core gene phylogeny, protein sequences of core genes (orthologous genes among the strains compared) from each genome were concatenated and multiple alignments were performed. Phylogenetic trees were derived using neighbor-joining method from the multiple alignments. All the phylogenetic trees were validated using 1,000 cycles of bootstrapping.

Results

Genome Assembly and Annotation

The sizes of the MPHL21000^T, MPHL43239^T, MPHL43070, MPHL43071, and MPHL43072 genomes were $\approx$5.3 Mbp (fig. 1), which is $\approx$0.35 Mbp smaller than the draft genome of the *M. phlei* RIVM strain (MPHLRIVM; Acc: AJFJ00000000; Abdallah et al. 2012). The Guanine-Cytosine content was calculated to be around 69.4% for all six *M. phlei* strains. The number of predicted protein-coding genes varies from 5,061 to 5,526. For MPHL21000^T, 39% of the genes were assigned to different functional classes and 27% with hypothetical functions (table 1). All strains carry 46 tRNA genes with the exception of MPHLRIVM, which encodes 50 tRNA genes (table 1; supplementary table S2 and fig. S1, Supplementary

Genome Biology and Evolution
SMBE

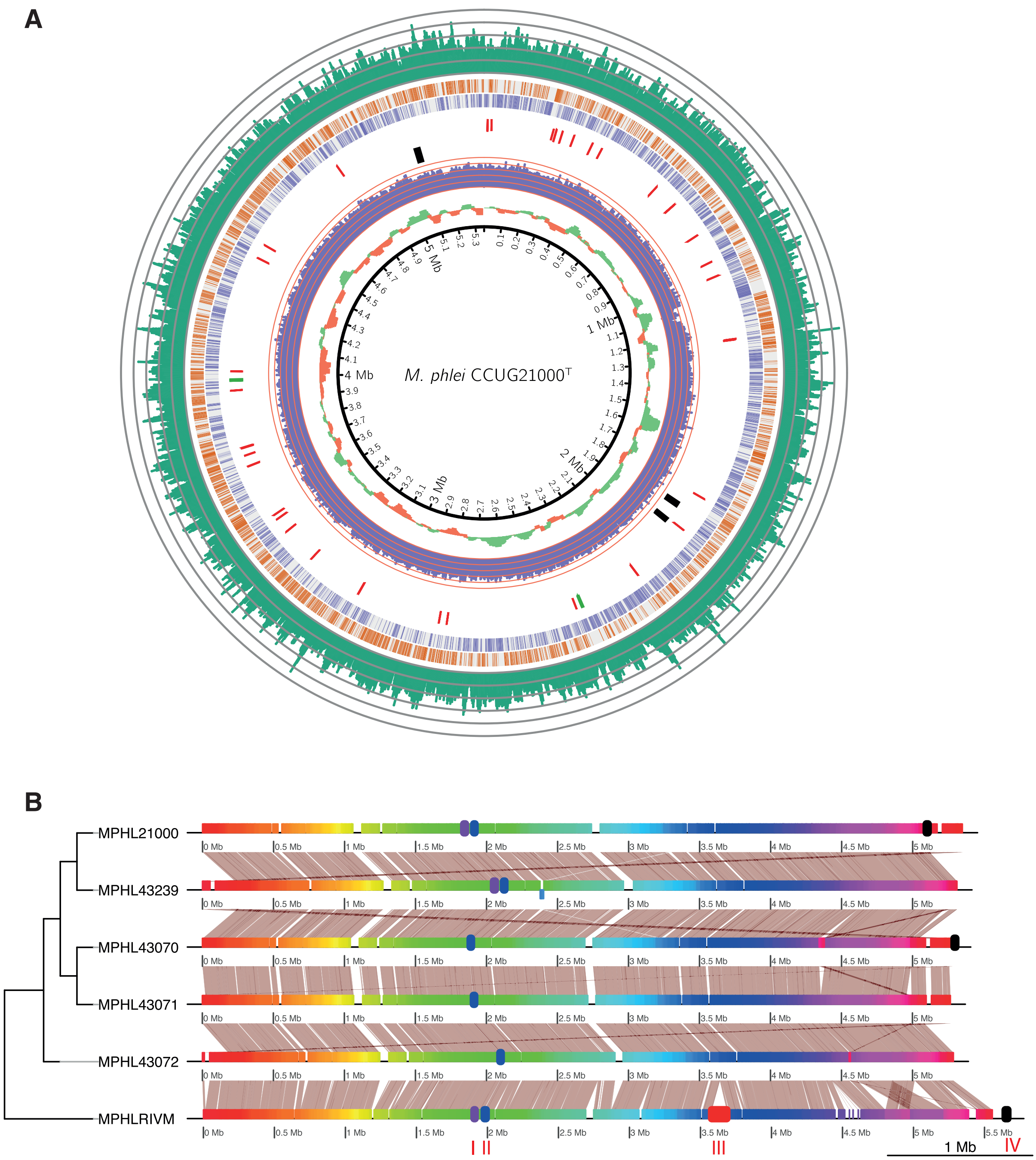

Fig. 1.—Complete genome sequence and alignment of the different *Mycobacterium phlei* genomes. (*A*) Circos plot showing the complete genome sequence of *M. phlei* CCUG21000^T. From outer to inner: Green histogram track represents the average sequencing read depth for the complete genome. Gray circles overlapping the histogram track are scale for read depth and the distance between the two circles is 50×. The brown and violet blocks in the two subsequent circles represent genes in forward and reverse strands, respectively. Next, red and green blocks show genome-wide distribution of tRNA and rRNA, respectively, while the three black blocks represent prophage sequences. The violet circles show histograms of Guanine-Cytosine (GC) content distribution of the genome sequence. The GC content (%) was calculated using a sliding window of 1,000 bp. In the histogram track, each of the orange circles represents a scale of 20. Next track shows GC skew of the genome generated using a sliding window of 1,000 bp. Positive and negative skew are represented by green and brown color, respectively. The innermost track shows scale along the genome length. (*B*) Whole-genome alignment of the six *M. phlei* strains where each of the colored horizontal blocks represents one genome and the vertical bars represent homologous regions. Diagonal lines represent genomic rearrangements, whereas white gaps represent insertions/deletions. The larger blocks of color purple, blue, red, and black indicate prophage sequence regions and are marked with I–IV, respectively. Same color blocks (except black blocks) represent the same prophage sequences where black blocks indicate nonconserved prophage sequences. Left side of the genome alignment shows a phylogenetic tree generated based on core genes.

Table 1
Summary of Assembly, Annotation, and Horizontally Transferred Genes of the *Mycobacterium phlei* Genomes

Properties	MPHL21000^T	MPHL43070	MPHL43071	MPHL43072	MPHL43239	RIVM
Strain source	CCUG	DSM	DSM	DSM	DSM	–
Assembly						
Read pairs	–	8258784	10128605	9409525	6367209	NA
Read length	**10,747**	100	100	100	100	100
Scaffold N50 (Kb)	–	119.7	53.14	220.77	299.7	155851
Number of scaffolds	**1**	117	215	88	45	102
Average GC content	**69.44**	69.4	69.4	69.4	69.4	69.24
Genome length (Mb)	**5,349,645**	5,304,064	5,313,441	5,312,278	5,322,335	5,681,954
Average read depth	**100**	~200	~200	~200	~200	–
Annotation						
Coding sequence	**5118**	5061	5068	5074	5085	5526
tRNA	**46**	46	46	46	46	50
rRNA operon	**2**	2	2	2	2	2
Noncoding RNA	**49**	38	37	37	36	39
No. of sigma factor	**19**	19	19	19	19	21
Predicted horizontally transferred genes						
No. of HGT genes	**125**	127	125	126	127	133

Note.—CCUG = Culture Collection at University of Göteborg, Sweden; DSM = Deutsche Sammlung von Mikroorganismen und Zellkulturen, Germany; GC = Guanine-Cytosine; NA = not applicable, bold column indicates complete genome.

Material online). As for other rapidly growing mycobacteria, our combined data suggested that all *M. phlei* strains have two rRNA operons (supplementary fig. S2, Supplementary Material online), which is in agreement with previous data (Bercovier et al. 1986). Moreover, the five *M. phlei* genomes have 19 sigma transcription factor genes while MPHLRIVM has 21 (table 1). This is fewer compared with the closely related *M. smegmatis* mc2 155 (supplementary fig. S3, Supplementary Material online. For further details, see supplementary material.

Mycobacterium phlei Genome Alignment and Prophage Analysis

Alignment of the six *M. phlei* genomes suggested no major genomic rearrangements but we could identify several insertion–deletion events. For example in MPHLRIVM (fig. 1*B*), 1) a segment of ≈ 60 kb (present in the other strains) is replaced with ≈20 kb near the 1.2 Mb position and 2) a 105.5-kb long segment is inserted in the vicinity of the 3.6-Mbp position (region III; fig. 1*B*). The 105.5-kb segment and other insertions (marked I–IV) were predicted to be of bacteriophage origin. In MPHL21000^T, three fragments (5–15 kb in length) were detected, while MPHLRIVM carries four (marked I–IV; fig. 1*B*). The region I insertion was predicted to be present in MPHL21000^T and MPHL43239^T, while 7.7 (of 12.1 kb present in MPHLRIVM) kb of region II is present in all six strains. The region III insertion (present only in MPHLRIVM) code for 174 proteins, 3 tRNA genes (Asn-GTT, Gln-CTG, and Trp-CCA), and interestingly, a putative tRNA^{His} guanylyltransferase gene, *tgh* (a likely homolog of the *Bacillus* phage Bcp1 gene; fig. 1*B* and supplementary table S3, Supplementary Material online; Jackman et al. 2012; Schuch et al. 2014). The tRNA$^{Asn(GTT)}$, tRNA$^{Gln(CTG)}$, and tRNA$^{Trp(CCA)}$ genes are also present in an *M. phlei* phage isolated in 1958 (Marton et al. 2016) supporting that these tRNA genes are of phage origin. The bacteriophage sequences predicted in region IV (17 kp) are not conserved between the strains. Together this accounts partly for the larger size of the MPHLRIVM genome.

Core and Unique Genes and Functional Classification

Core genes, which represent genes having 1:1 orthologs in all six strains, cover almost 89.3% (4,572) of the total predicted genes in MPHL21000^T (fig. 2*A*). MPHLRIVM displayed the highest number of strain-specific genes (n = 690; fig. 2*B*) and these clustered in just a few genomic regions (fig. 2*A*). Two of these clusters overlapped with regions III and IV discussed above. We predicted that 222 genes are present in all strains except MPHLRIVM and these genes are spread all over the genome. Moreover, 66 and 117 of the predicted genes were unique to MPHL21000^T and MPHL43071, respectively (fig. 2*B*).

Functional classification of 1,692 genes revealed that the distribution of these into different subsystems is very similar to other environmental mycobacteria (Das et al. 2015) with >33% comprising the subsystems "Amino Acids and Derivatives" and "Carbohydrates" (supplementary fig. S4, Supplementary Material online). Classification of the auxiliary ("noncore") genes gave a similar pattern with a few exceptions such as the "Virulence, Disease, and Defense" subsystem (supplementary fig. S4, Supplementary Material online).

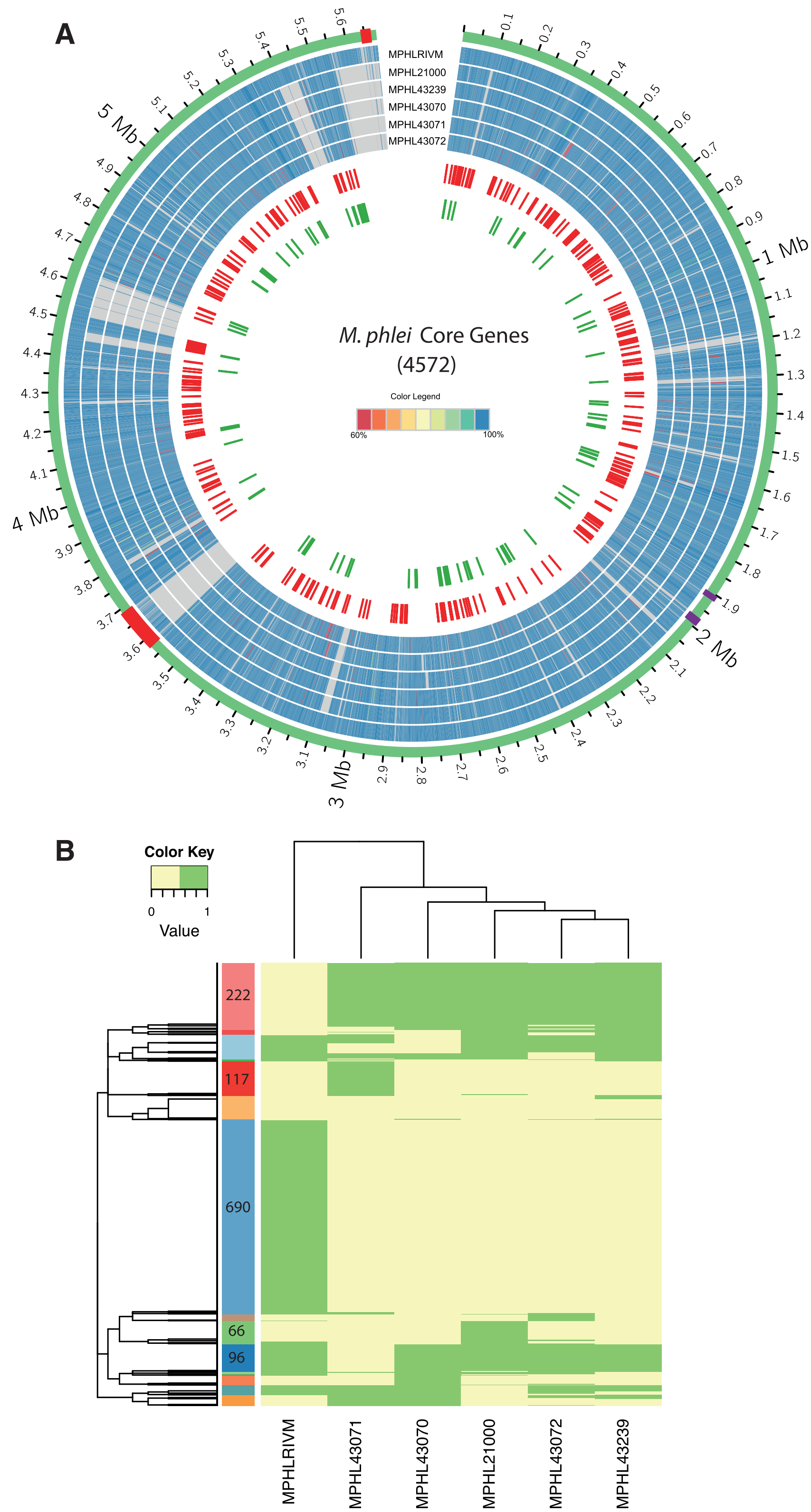

FIG. 2.—Core and auxiliary (noncore) genes identified in *Mycobacterium phlei* genomes. (*A*) The predicted orthologous genes of MPHLRIVM in the other five strains of *M. phlei*. The outer green track shows the MPHLRIVM genome as reference with scale. The blocks overlapping the reference genome indicate predicted prophage regions. Next six tracks comprise blocks representing orthologous genes predicted to be present in the different strains as indicated.

(continued)

Horizontally Transferred Genes

The total number of putative horizontally transferred genes ranged from 125 (MPHL21000^T) to 133 (MPHLRIVM) and 104 HGT genes are common in all the six strains (table 1). Among these, >50% belong to the functional categories Amino Acids and Derivatives and Carbohydrates (fig. 3*A*). Detailed analysis of the category Amino Acid and Derivatives suggested that genes involved in "Polyamine Metabolism," "Arginine and Ornithine Degradation," and "Glycine and Serine Utilization" are the most common HGT genes (fig. 3*B*).

We next compared the distribution of the *M. phlei* HGT genes orthologous in other mycobacteria. More than 60% (*n* = 63) of the HGT genes in *M. phlei* were predicted to be present in its closest neighbors with *M. smegmatis* sharing the highest number of HGT genes (fig. 3*C*). Conceivably this is because their habitats are similar ecological niches. Subsequently, we identify possible donors of the *M. phlei* HGT genes and members of the order *Streptomycetales* and *Pseudonocardiales* were predicted to be the most likely donors (fig. 3*D*).

Phylogenetic Analysis

The 16S rDNA-based phylogenetic tree suggested that the *M. smegmatis* mc2 155 and JS623 strains are the closest neighbors and that *M. phlei*, *M. smegmatis*, and *Mycobacterium* spp. (JLS, KMS, and MCS) share a common ancestor (fig. 4*A*). In contrast, the tree generated using *Mycobacterium* core genes revealed that the closest neighbors of *M. phlei* are *M. smegmatis* JS623 and *M. rhodesiae* NBB3, while *M. smegmatis* mc2 155 and *Mycobacterium* spp. (JLS, KMS, and MCS) were positioned on a different branch than *M. phlei* (fig. 4*B*).

Mycobacterium phlei Genes

Polyamine Metabolism

Polyamines such as putrescine, spermidine, and cadaverine are essential for bacterial growth and influence biofilm formation (Patel et al. 2006) and we predicted several genes in *M. phlei* involved in polyamine metabolism (fig. 5*A*; supplementary table S4, Supplementary Material online). Among these, the ornithine decarboxylase, arginine decarboxylase, and agmatine ureohydrolase genes are involved in the biosynthesis of putrescine. Several genes were also predicted to be part of transport systems of extracellular polyamines, the two ATP-binding cassette (ABC) transporters encoded by *potABCD* and *potFGHI* (*potI* only predicted in MPHLRIVM), which are specific for uptake of spermidine and putrescine, respectively. Moreover, the arginine/ornithine antiporter gene, *arcD*, was also predicted to be present in all *M. phlei* strains.

Comparative analysis using complete mycobacterial genomes revealed that several genes related to polyamine metabolism and transport could not be predicted in pathogenic mycobacteria, including *M. tuberculosis*, using *M. phlei* genes as reference. Moreover, ≈50% of the genes predicted to be present in *M. phlei* were not detected in other environmental species (fig. 5*A*) suggesting that these genes might be unique to *M. phlei* (see below).

Glycerol Utilization

Mycobacterium phlei and *M. smegmatis* can use glycerol as a carbon source (McKenzie et al. 2012). Comparative analysis of genes involved in glycerol uptake and utilization pathways in these two species revealed that several genes are missing in the *M. phlei* genomes (supplementary fig. S5*A*, Supplementary Material online): *ugpB*, encoding a subunit of the ABC transporter GlpF (involved in glycerol transport); *dhaF* and *dhaKLM*, involved in conversion of glycerol to dihydroxyacetone (DHA); and phosphorylation of DHA. Given that *M. phlei* grows on media with glycerol as the sole carbon source (Tepper 1968; not shown) suggest that the uptake of glycerol is mediated by an alternative pathway(s) or diffuses through the membrane. In this context, we also noted that addition of glycerol to the media resulted in a *M. phlei* strain-dependent variation in the growth rate (supplementary fig. S5*B*, Supplementary Material online).

Mammalian Cell Entry Genes

The mammalian cell entry (*mce*) genes encode proteins involved in cell invasion (Arruda et al. 1993). The predicted numbers of "complete" *mce* clusters and genes in *M. phlei* vary with MPHLRIVM having the highest numbers: 10 clusters (I–X) comprising 82 genes (fig. 5*B*). The *mce*I and *mce*II clusters are conserved in both environmental and nonenvironmental mycobacteria with the exception of *Mycobacterium abscessus* and *Mycobacterium massiliense* in which only a few *mce*I and *mce*II genes are present (fig. 5*B*). The *mce* III–X clusters are partially conserved within many environmental mycobacteria while several *mce* genes are only present in mycobacteria belonging to the MTB complex (four in *M. tuberculosis*): *Mycobacterium avium* and *Mycobacterium avium* subsp. *paratuberculosis* (fig. 5*B* marked in red; see also Casali and Riley 2007). It therefore appears that *M. phlei* harbors a diverse set

Fig. 2.—Continued
Orthologous genes are colored based on the percentage of identity as indicated in the color legend. The red and green blocks represent *M. phlei* genes not predicted to be present in other mycobacteria and putative horizontally acquired genes in all the six strains, respectively. (*B*) Clustering of auxiliary (noncore) genes using hierarchical clustering. Green and yellow color represent gene present or absent, respectively. The vertical colored bands marked heat map on the left show different clusters and also indicate the number of genes in some major clusters.

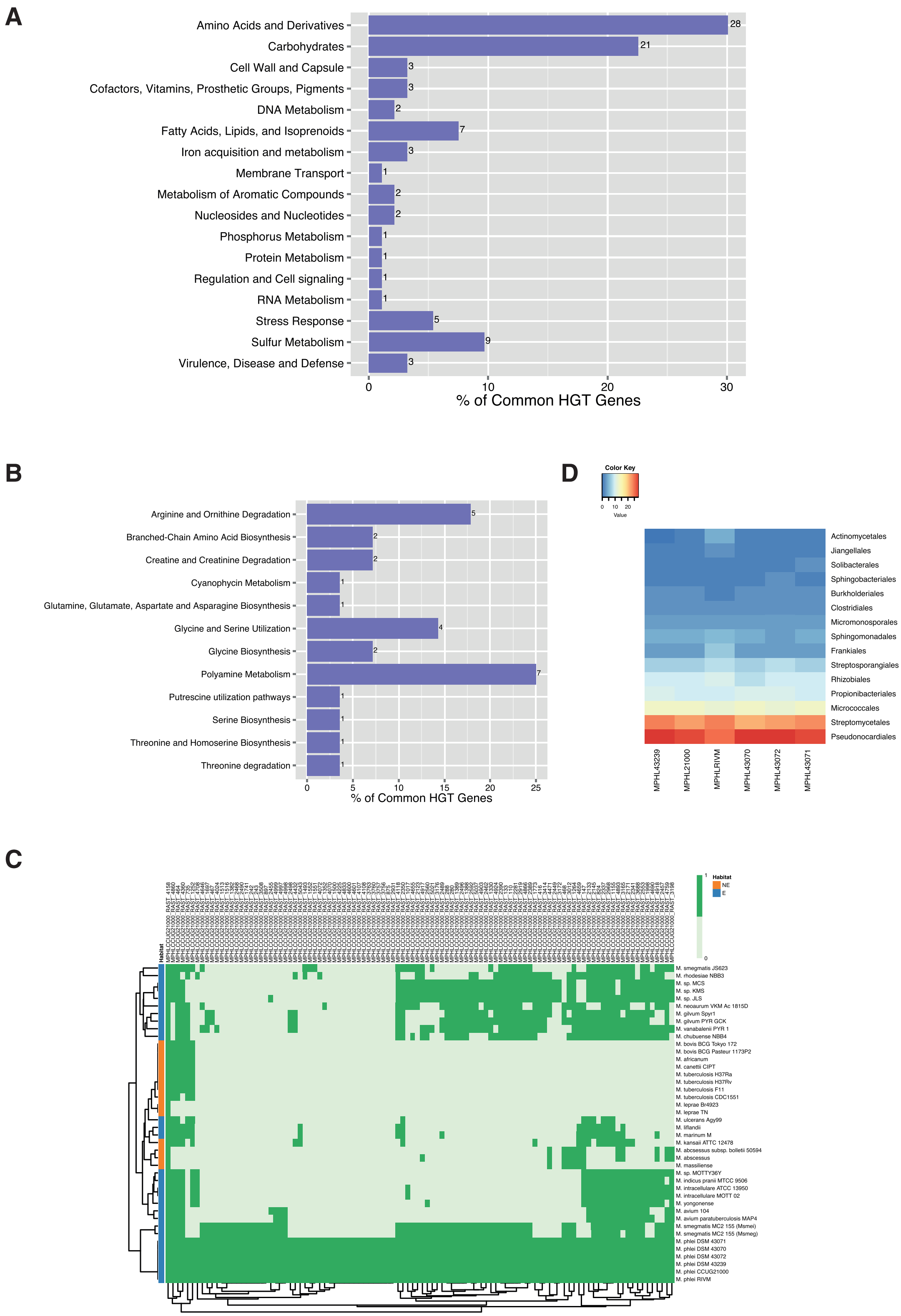

Fig. 3.—Analysis of horizontally acquired genes in *Mycobacterium phlei*. (*A*) Bar plot showing percentage of common horizontally acquired genes in different functional categories. (*B*) Percentage of horizontally acquired genes in different subsystems of the category Amino Acids and Derivatives. (*C*) Clustering of horizontally acquired genes across different environmental and nonenvironmental mycobacteria for which complete genomes are available. (*D*) Putative donors of the predicted horizontally acquired genes; color code dark to light indicates high to low number of genes acquired.

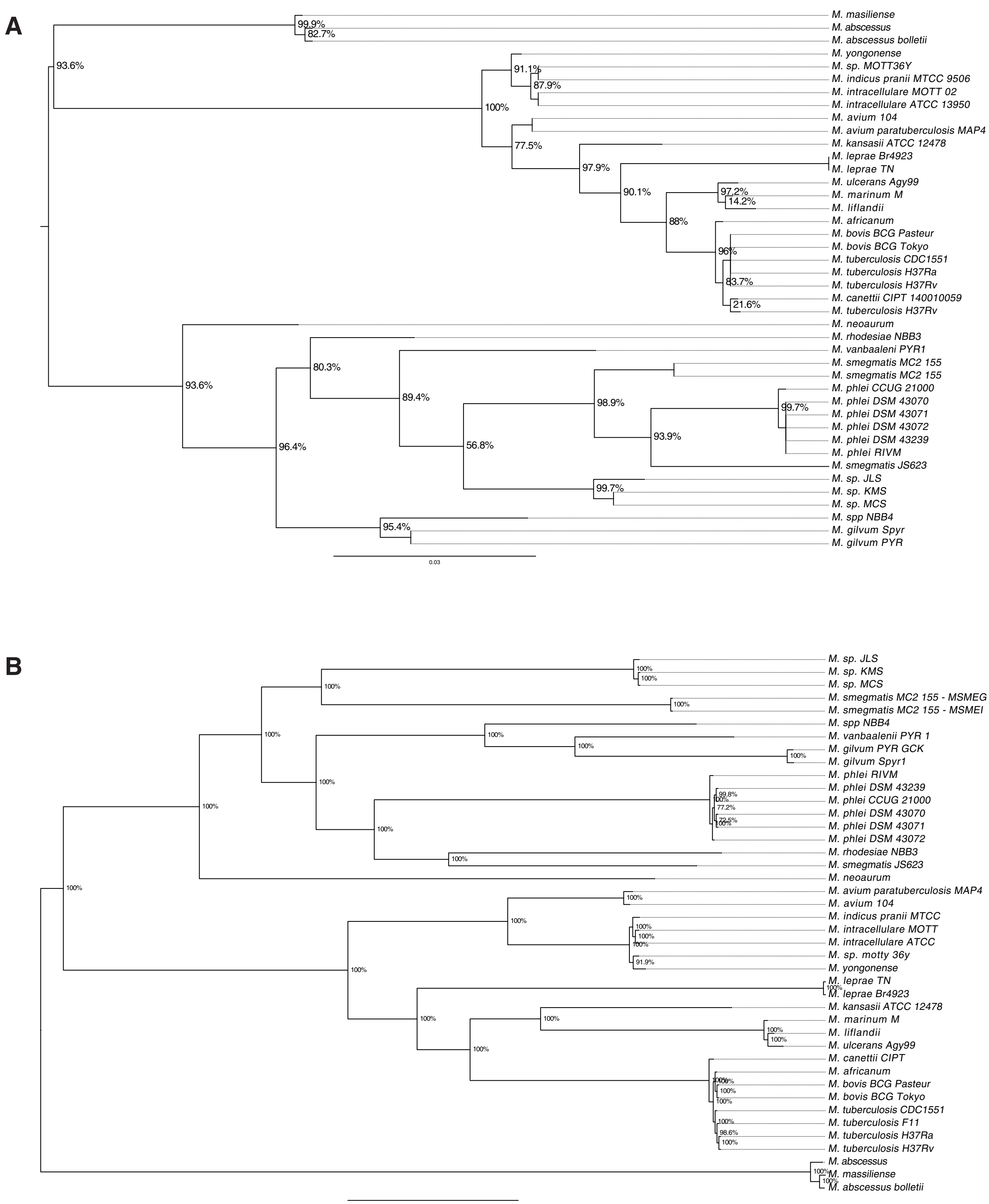

Fig. 4.—Phylogenetic analysis. Phylogenetic trees were based on (*A*) 16S rDNA and (*B*) core genes (693) predicted to be present in the *Mycobacterium* spp. for which complete genomes are available. The percentage values in the nodes represent bootstrap values generated by 1,000 cycles.

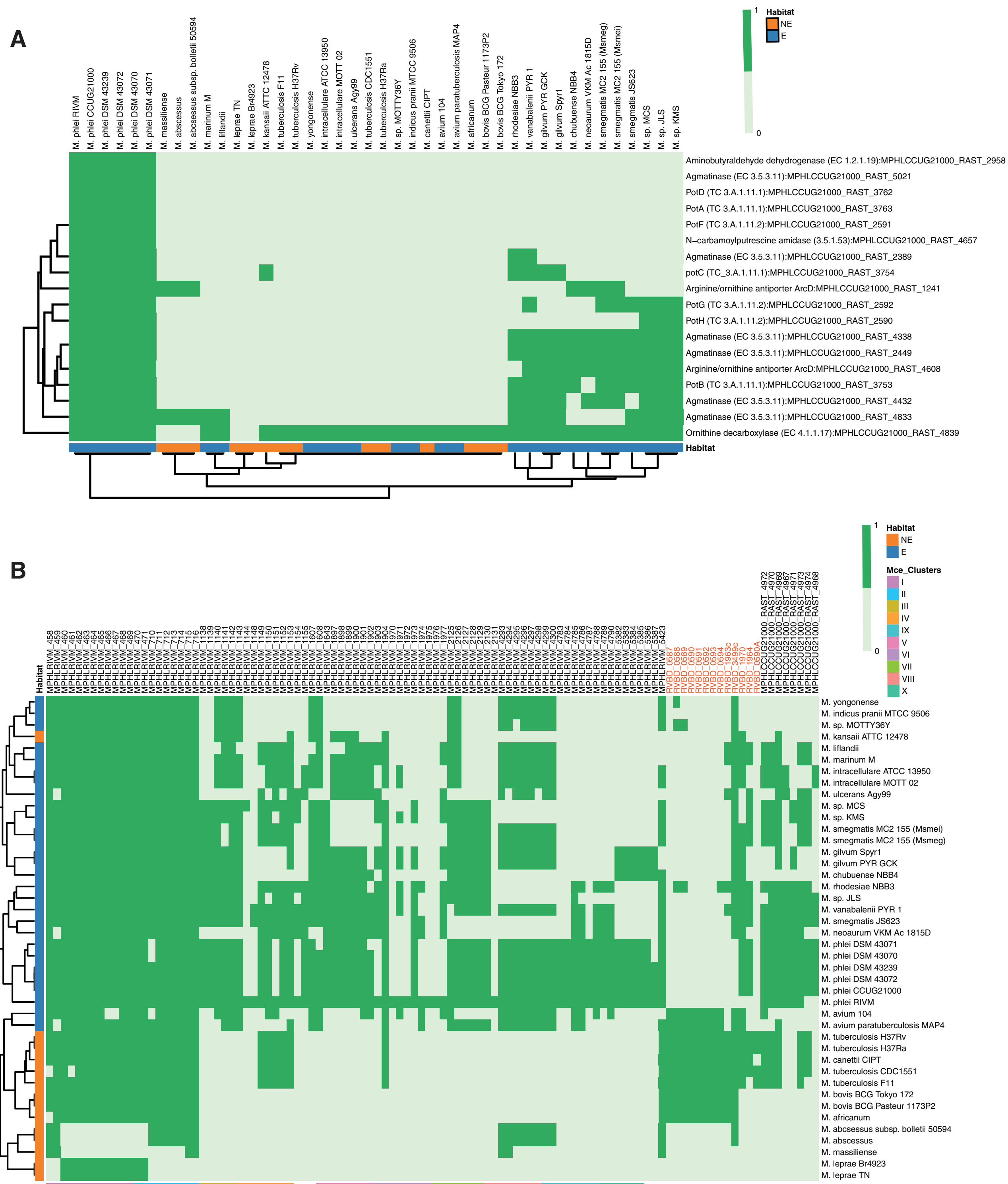

Fig. 5.—Distribution of specific genes/gene families predicted to be present in *Mycobacterium phlei* and in other *Mycobacterium* species. Heat map showing presence/absence (dark/light green) of orthologous genes in *M. phlei* and different environmental and nonenvironmental mycobacteria for which complete genomes are available. (*A*) Genes involved in uptake and metabolism of polyamines predicted to be present in *M. phlei*. (*B*) Predicted *mce* genes and operons in *M. phlei* and *Mycobacterium tuberculosis* (for details see main text).

of *mce* operons/genes; some are present in both environmental and nonenvironmental mycobacteria, while others are present only in the environmental mycobacteria.

Mycobactin Genes

Mycobacterium phlei was predicted to have the two *mbt* clusters, *mbt*-1 (*mbtABCDEFGIJ*) and *mbt*-2 (*mbtKLMN*), which encompass the genes responsible for the biosynthesis of mycobactin (supplementary fig. S6, Supplementary Material online). However, *M. phlei* contains only partial *mbtI* and *mbtJ* genes with low sequence identity (<45%). Orthologs of these two partial genes were also predicted to be present in the *M. tuberculosis* genome in addition to the longer *mbtI* and *mbtJ* genes. Moreover, we also predicted the presence of *mbtT*, which is absent in *M. tuberculosis* but present in other nonpathogenic mycobacteria such as *M. smegmatis* mc2 155 (Chavadi et al. 2011). We also noted that *irtA* and *irtB*, which encode an ABC transporter with a role in *M. tuberculosis* growth under iron-deficient conditions (Rodriguez 2006), appear to be missing in *M. phlei*.

CRISPR-Cas System

Our analysis revealed the presence of partial fragments of the adaptive immunity system Type 1E CRISPR-cas in MPHLRIVM. This system encompasses a signature gene of Type 1 (*cas*3) and several type-dependent genes, *cse*1, *cse*2, *cse*4, and *cas*5 (supplementary fig. S7, Supplementary Material online). The complete Type 1E system includes two additional genes *cas*1 and *cas*2, which are universally present in the all known CRISPR-Cas systems (Bhaya et al. 2011). However, we were unable to detect these two genes in MPHLRIVM. Neither could we detect CRISPR-Cas genes in any of the other *M. phlei* genomes.

Discussion

Although there are genomic variations comparing the six strains where MPHLRIVM differs the most, the genomes appear to be stable. The genome sizes of five *M. phlei* strains, including one complete genome (MPLH21000^T), were found to be ≈5.3 Mb, which is 350 kb smaller compared with the draft MPHLRIVM genome (Abdallah et al. 2012). The difference in size is partly due to the presence of prophage sequences in the MPHLRIVM. In conclusion, the five *M. phlei* genomes are likely to better represent the *M. phlei* group than the RIVM strain.

Phylogenetic analysis based on 16S rDNA positioned *M. phlei* close to *M. smegmatis* mc2 155 and JS623 strains while using 693 orthologous genes present in the genomes of 42 *Mycobacterium* spp. (including the six *M. phlei* strains) suggested that the closest relatives of *M. phlei* are *M. rhodesiae* NBB3 and *M. smegmatis* JS623. In this context, we raise the question whether *M. smegmatis* mc2 155 and JS623 should be considered as separate species because these two were clearly separated based on our *Mycobacterium* core genes phylogenetic tree (fig. 4*B*). Our data also revealed that 4,572 genes are common among all *M. phlei* strains and that 393 genes were only predicted to be present in *M. phlei* (not present in the other mycobacteria). Among mycobacteria these 393 genes can therefore be considered to constitute the species signature for *M. phlei*. Some of these genes relate to polyamine biosynthesis and to functions that are linked to the presence of unique *mce* genes. The majority of these genes were, however, classified as encoding hypothetical proteins and several were also classified as HGT genes that originate from other environmental bacteria belonging to, for example, *Streptomycetales* and *Pseudonocardiales* (figs. 3*A* and 4*D*). Identification of the functions of these unique *M. phlei* genes will possibly give clues to a molecular understanding of what separates *M. phlei* from other mycobacterial species. To conclude, the genomes for the different *M. phlei* strains constitute a platform to understand the biology of members of the *Mycobacterium* genus in general and *M. phlei* in particular and its use in, for example, cancer therapy.

Supplementary Material

Supplementary tables S1–S4 and figures S1–S71 are available at *Genome Biology and Evolution online* (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was funded by the Swedish Research Council (M), SIDA/SAREC, the Swedish Research Council-SIDA, the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (FORMAS), and Uppsala RNA Research Center (Swedish Research Council Linneus support). A. Bhattacharya acknowledges the Department of Biotechnology, India. The SNP&SEQ Technology Platform in Uppsala performed sequencing of the genomes. The platform is part of Science for Life Laboratory at Uppsala University and supported as a national infrastructure by the Swedish Research Council. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2011072. L.A.K. is on the board of directors of Bioimics AB. L.A.K. and S.D. are holders of Swedish patent application number PCT/SE2008/051486.

Literature Cited

Abdallah AM, et al. 2012. Complete genome sequence of *Mycobacterium phlei* type strain RIVM601174. J Bacteriol. 194:3284–3285.
Aguilar JL, Sanchez EE, Carrillo C, Alarcón GS, Silicani A. 1989. Septic arthritis due to *Mycobacterium phlei* presenting as infantile Reiter's syndrome. J Rheumatol. 16:1377–1378.
Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley LW. 1993. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. Science 261:1454–1457.

Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

Bardouniotis E, Huddleston W, Ceri H, Olson ME. 2001. Characterization of biofilm growth and biocide susceptibility testing of *Mycobacterium phlei* using the MBEC assay system. FEMS Microbiol Lett. 203: 263–267.

Bercovier H, Kafri O, Sela S. 1986. Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. Biochem Biophys Res Commun. 136:1136–1141.

Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet. 45:273–297.

Casali N, Riley LW. 2007. A phylogenomic analysis of the *Actinomycetales mce* operons. BMC Genomics. 8:60.

Chavadi SS, et al. 2011. Mutational and phylogenetic analyses of the mycobacterial *mbt* gene cluster. J Bacteriol. 193:5905–5913.

Chin CS, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 10: 563–569.

Csillag A. 1970. A simple method to obtain the mycococcus form of *Mycobacterium phlei*. J Gen Microbiol. 62:251–259.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394–1403.

Das S, et al. 2015. Characterization of three *Mycobacterium* spp. with potential use in bioremediation by genome sequencing and comparative genomics. Genome Biol Evol. 7:1871–1886.

Filion MC, Phillips NC. 2001. Therapeutic potential of mycobacterial cell wall-DNA complexes. Expert Opin Investig Drugs. 10: 2157–2165.

Fouts DE, Brinkac L, Beck E, Inman J, Sutton G. 2012. PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. Nucleic Acids Res. 40:e172.

Francis J, Macturk HM, Madinaveitia J, Snow GA. 1953. Mycobactin, a growth factor for *Mycobacterium johnei*. I. Isolation from *Mycobacterium phlei*. Biochem J. 55:596–607.

Gordon RE, Mihm JM. 1959a. A comparison of four species of mycobacteria. J Gen Microbiol. 21:736–748.

Gordon RE, Mihm JM. 1959b. A comparison of *Nocardia asteroides* and *Nocardia brasiliensis*. J Gen Microbiol. 20:129–135.

Gordon RE, Smith MM. 1953. Rapidly growing, acid fast bacteria. J Bacteriol. 66:505–507.

Jackman JE, Gott JM, Gray MW. 2012. Doing it in reverse: 3′-to-5′ polymerization by the Thg1 superfamily. RNA 18:886–899.

Juhasz SE. 1962. Aberrant forms of *Mycobacterium phlei* produced by streptomycin and their multiplication on streptomycin-free media. J Gen Microbiol. 28:9–13.

Karnam S, et al. 2011. *Mycobacterium phlei*, a previously unreported cause of pacemaker infection: thinking outside the box in cardiac device infections. Cardiol J. 18:687–690.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.

Li R, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20:265–272.

Marton S, et al. 2016. Genome sequence of a cluster A13 mycobacteriophage detected in *Mycobacterium phlei* over a half century ago. Arch Virol. 161:209–212.

McKenzie JL, et al. 2012. A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria. J Bacteriol. 194:2189–2204.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 29:2933–2935.

Patel CN, et al. 2006. Polyamines are essential for the formation of plague biofilm. J Bacteriol. 188:2355–2363.

Paul E, Devarajan P. 1998. *Mycobacterium phlei* peritonitis: a rare complication of chronic peritoneal dialysis. Pediatr Nephrol. 12:67–68.

Pitulle C, Dorsch M, Kazda J, Wolters J, Stackebrandt E. 1992. Phylogeny of rapidly growing members of the genus *Mycobacterium*. Int J Syst Bacteriol. 42:337–343.

Rodriguez GM. 2006. Control of iron metabolism in *Mycobacterium tuberculosis*. Trends Microbiol. 14:320–327.

Saito H, et al. 1977. Cooperative numerical analysis of rapidly growing mycobacteria: the second report. Int J Syst Bacteriol. 27:75–85.

Schuch R, Pelzek AJ, Fazzini MM, Nelson DC, Fischetti VA. 2014. Complete genome sequence of *Bacillus cereus* sensu lato bacteriophage Bcp1. Genome Announc 2:e00334–14.

Spiegl PV, Feiner CM. 1994. *Mycobacterium phlei* infection of the foot: a case report. Foot Ankle Int. 15:680–683.

Stackebrandt E, Ludwig W, Schleifer KH, Gross HJ. 1981. Rapid cataloging of ribonuclease T1 resistant oligonucleotides from ribosomal RNAs for phylogenetic studies. J Mol Evol. 17:227–236.

Tepper BS. 1968. Differences in the utilization of glycerol and glucose by *Mycobacterium phlei*. J Bacteriol. 95:1713–1717.

Wayne LG, Runyon EH, Kubica GP. 1969. Mycobacteria: a guide to nomenclatural usage. Am Rev Respir Dis. 100:732–734.

Wyckoff RWG, Smithburn KC. 1933. Micromotion pictures of the growth of *Mycobacterium phlei*. J Infect Dis. 53:201–209.

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39:1–6.

Zhu Q, Kosoy M, Dittmar K. 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. BMC Genomics 15:717.

Associate editor: Howard Ochman