Deciphering the virulence factors of the opportunistic pathogen
Mycobacterium colombiense

M. N. Gonzalez-Perez1,3, M. I. Murcia1, C. Parra-Lopez1, J. Blom2 and A. Tauch1,3
1) Microbiology Department, School of Medicine, National University of Colombia, Bogotá, Colombia, 2) Bioinformatics and Systems Biology, Justus Liebig University Giessen, Giessen, Germany and 3) Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

Abstract

Mycobacterium avium complex (MAC) contains clinically important nontuberculous mycobacteria worldwide and is the second largest medical complex in the Mycobacterium genus after the Mycobacterium tuberculosis complex. MAC comprises several species that are closely phylogenetically related but diverse regarding their host preference, course of disease, virulence and immune response. In this study we provided immunologic and virulence-related insights into the M. colombiense genome as a model of an opportunistic pathogen in the MAC. By using bioinformatic tools we found that M. colombiense has deletions in the genes involved in p-HBA/PDIM/PGL, PLC, SL-1 and HspX production, and loss of the ESX-1 locus. This information not only sheds light on our understanding the virulence mechanisms used by opportunistic MAC pathogens but also has great potential for the designing of species-specific diagnostic tools.

© 2016 The Author(s). Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Keywords: Genomic islands, Mycobacterium colombiense, nontuberculous mycobacteria, tuberculosis, virulence factors

Original Submission: 21 June 2016; Revised Submission: 2 August 2016; Accepted: 6 September 2016

Article published online: 10 September 2016

Introduction

Mycobacterium avium complex (MAC) contains clinically important nontuberculous mycobacteria and is the second largest medical complex in the Mycobacterium genus after the Mycobacterium tuberculosis (Mtb) complex. MAC affects patients with chronic obstructive pulmonary disease and cystic fibrosis, as well as immunosuppressed individuals with HIV/AIDS [1]. MAC is currently composed of 12 species which are very close phylogenetically related, but they are diverse regarding to their host preference, course of disease, virulence and immune response, such as M. avium, M. intracellulare and M. colombiense. The reasons of this variability among very closely related species are not yet clear.

M. colombiense (Mcol) infections were initially described in HIV patients in Bogotá, Colombia [2]. In this initial study, Mcol was confirmed to be the causative agent of pulmonary disease and bacteraemia in this group of people who died of HIV/Mcol coinfection. After this study Mcol has been isolated in other scenarios such as lymphadenopathy, subcutaneous infections and disseminated disease in HIV-negative patients in countries like Spain, France, China, Canada and Russia. In this study we attempted to find the Mcol virulence factors from the genome sequence to decipher its possible skills as a MAC opportunistic pathogen.

Methods

The Mcol CECT 3035 genome was sequenced using PacBio technology (Institute for Genomes Sciences, University of Maryland, College Park, MD, USA), then automatically annotated using the GenDB 2.4 platform [3]. To study the Mcol pathogenome, we used different bioinformatics tools to search for prophages sequences (Phast tool [4]), genomic islands (IslandViewer [5] and GIPSy/PIPS [6] tools), protein families...
PathogenFinder [7]) and mycobacterial virulence factors (VFDataBase (VFDB) [8]). The islands detected with IslandViewer were classified using the National Center for Biotechnology Information functional annotation and the Conserved Domain Database, and for the GIPSy/PIPS analysis we used the *M. indicus pranii* genome as the closest related nonpathogenic species. To search for the Mcol singletons, we created a mycobacterial database using the genomes of *M. avium* subsp. *hominissuis* 104 (RefSeq NC_008595), *M. avium* subsp. *paratuberculosis* K-10 (RefSeq NC_002944), *M. avium* subsp. *paratuberculosis* MAP4 (RefSeq NC_021200), *M. intracellulare* ATCC 13950 (RefSeq NC_016946), *M. intracellulare* MOTT-36Y (RefSeq NC_017904), *M. intracellulare* MOTT-02 (RefSeq NC_016947), *M. intracellulare* MOTT-64 (RefSeq NC_016948), *M. indicus pranii* MTCC 9506 (RefSeq NC_018612), *M. yongonense* 051390 (RefSeq NC_021715), *Mtb* H37Rv (RefSeq NC_000962) and Mcol CECT 3035 with the EDGAR tool [9]. Then we detected the secretory singletons using PRED-TAT [10] and PRED-LIPO [11] tools. For the detection of transmembrane singletons we used TMHMM server 2.0 [12], and for the detection of the cell wall singletons we used CW-PRED [13].

**Results**

**Prophages and genomic island detection**

Using the tools Phast, IslandViewer and GIPSy/PIPS, we did not detect phage sequences in the Mcol genome. In addition, we did not find structures compatible with CRISPR/Cas systems by applying the CRISPR Finder tool [14]. However, we detected a total of 45 genomic islands (nine of these are classified as pathogenicity island), which contain a total of 320 predicted genes. Interestingly, 92.8% of these genes code for hypothetical proteins with unknown functions (Fig. 1). In addition, the Mcol pathogenicity islands contain genes that...
code for the DrrA-C transporter, which is involved in the transportation of phthiocerol dimycocerosates (PDIMs) and also contains the rmt2 and mmpL4 genes, which are involved in the synthesis and transmembrane transportation of glycopeptidolipids (GPLs).

Singleton detection with EDGAR

Using the EDGAR tool for comparative genomics analysis of a selected set of mycobacterial genomes we detected a total of 452 singletons, which represent 8.4% of the Mcol genome. According to a classification of these singletons into functional Clusters of Orthologous Groups (COGs) categories 87 genes are involved in cellular processing and signaling, 75 are involved in information storage and processes, 163 are involved in metabolism of Mcol and 122 are grouped as poorly characterized (Fig. 2). The majority of singletons (85%) contain typical Sec signal peptides for recognition of the preprotein by the general secretory pathway, and 22% contain varying numbers of transmembrane helical segments for a putative integration into the cytoplasmic membrane. We also detected one singleton with an LPXTG motif that is an indication of cell wall localization of the respective protein. Among these Mcol singletons we found a gene coding for a cysteine desulfurase protein. It should be noted that a recent study shows that the cysteine desulfurase–encoding gene sufS2 is inducible in response to oxidative stress and is important for the survival of Agrobacterium tumefaciens [15]. We also found a gene coding for the urease accessory protein UreD2. The presence of urease in pathogenic bacteria strongly correlates with pathogenesis in some human diseases and is an important virulence factor. Additionally, Cryptococcus gattii knockout mutants for genes encoding the urease accessory proteins Ure4 and Ure6 showed reduced multiplication within macrophages [16].

VFDB analysis

The VFDB analysis allowed us to detect a total of 204 Mcol genes that are homologous with previously described mycobacterial virulence factors. As shown in Fig. 3, the largest groups of genes comprise deduced virulence factors assigned to the class’s cell surface components and secretion system.

Analysis of protein families

Analysis of protein families with the PathogenFinder 1.1 tool allowed us to predict 32 virulence factors in the Mcol genome sequence. Of these genes 65.6% (n = 21) encoded hypothetical proteins and 5.5% (n = 2) encoded potentially secreted proteins with characteristic Sec signal peptides. Among these ten were also found in the VFDB, and 23 are new potential virulence factors.

Among these new potential Mcol virulence factors we found several genes homologous to the M. avium subsp. paratuberculosis str. k10 genome: MAP_0847, MAP_1945c, MAP_0092, MAP_3697c, MAP_1909, MAP_3223c, MAP_0908c, MAP_2811c and MAP_2636. According to Basic Local Alignment Search Tool (BLAST) searches, these genes (except MAP_1909) are likely not to be present in the Mtb H37Rv genome. Interestingly, the MAP_1909 gene encodes the LppM lipoprotein, which in Mtb is an important virulence factor involved in the manipulation of the phagosomal maturation in macrophages [17]. We also found several genes homologous to the M. avium subsp. hominisuis 104 genome: MAV_4325 (codes for ABC transporter, ATP-binding protein), MAV_1611 (codes for transcriptional regulator, TetR family protein), MAV_4258 (codes for the RsbW protein) and MAV_2193 (codes for acyl carrier protein), MAV_2835 (codes for enoyl-CoA hydratase/isomerase family protein), as well as the MAV_3699, MAV_0045, MAV_1178, MAV_1177, MAV_4234, MAV_0139, and MAV_3009 genes (code for hypothetical proteins). According to BLAST search, all these genes (except MAV_4258) are likely to not be present in the Mtb H37Rv genome. In addition, the RsbW protein in Mtb is involved in the response to heat, cold, oxidative, starvation and anaerobic stresses [18]. Finally, we found a homologue to the MMAR_4989 gene from the M. marinum M genome. This gene encodes the YrbE4A conserved membrane protein, which has an unknown role in virulence but is predicted to be involved in lipid catabolism in Mtb. Experimental studies are necessary to decipher the role of these new potential virulence factors in opportunistic pathogens such as Mcol.
Discussion

One of the hostile environments to which mycobacteria have to adapt during infection in the host is the macrophage. Therefore, among all of the probable Mcol virulence factors these have special importance because they are involved or are possibly implicated in macrophage modulation mechanisms (Fig. 4).

*M. colombiense* virulence factors that are also present in *Mtb*

Mcol has genes that encode the trehalose-recycling ABC transporter LpqYSugABC. Trehalose is an important compound for several mycobacterial cell wall glycolipids, including sulfolipid-1 (SL-1), trehalose monomycolate and trehalose dimycolate (TDM) [19]. On the other hand, Mcol has the gene that encodes the WhiB3 protein, which is an important virulence factor that is involved in several processes, including up-regulation of the synthesis of SL-1 and TDM, probably through the sensing of the reactive oxygen species (ROS) and nitric oxide [20].

Mcol has genes that encode the antigen 85 complex. This complex constitutes a group of extracellular mycolyl transferases (Ag85A, Ag85B and Ag85C), which play important roles in the biosynthesis of major components of the mycobacterial cell envelope, such as TDM and mycolylarabinogalactan [21].

Mcol has the genes that code for the Sec export systems, secA1 and secA2. The housekeeping SecA1 protein of *Mtb* is involved in the secretion of the virulence factors LprG and LpqH. These lipoproteins are TLR2 ligands involved in the manipulation of macrophage responses. In addition, the SecA2 export system secretes important virulence factors such as superoxide dismutase (SodA) involved in the survival inside macrophages [22] and catalase–peroxidase protein (KatG) involved in isoniazid resistance [23].

Mcol has the gene encode protein kinase G (PknG), which prevents phagosome–lysosome fusion by blocking lysosomal delivery [24]. This process blocks the intracellular degradation of mycobacteria in lysosomes and mediates intracellular survival of mycobacteria within macrophages [24].

Mcol has genes that code for proteins that in *Mtb* are associated with protection against ROS; these are Ndk, NuoG, SOD and KatG. In *Mtb* Ndk damages the phagocyte nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex in macrophages [25]. NADPH oxidase is responsible for the production of ROS, which are radical molecules that are crucial for the control of mycobacterial infections. In addition Ndk also prevents phagosome–lysosome fusion through a mechanism that involves the inactivation of macrophages Rab proteins [26]. It has been suggested that the Ndk protein could be exported through ESX-1, which is lacking in MAC members, including Mcol. Therefore, experimental research is necessary to investigate Ndk secretion in Mcol. In *Mtb* NuoG is an important virulence factor that can neutralize ROS produced by NADPH oxidase and with this inhibits macrophage apoptosis [27]. In addition, SodA and KatG proteins are secreted by the SecA2 export system [28]; they protect against ROS by forming H2O2, which is further detoxified by KatG [20].

![FIG. 3. Mycobacterium colombiense virulence factors detected by VFDataBase (VFDB). Shown are main VFDB category distributions for virulence factors found in this bacterium.](image-url)
Two-component systems (TCSs) are major regulatory systems for bacterial adaptation to environmental changes. The PrrA/B, MprA/B, PhoP/R and DosR/S TCSs are the most studied and are involved in important virulence mechanisms in Mtb. All of them are present in the Mcol genome. In Mtb the PrrA/B system has been implicated in macrophage phagocytosis and is transiently required for the early stages of macrophage infection [29]. The MprA/B is a stress-responsive TCS that directly regulates expression of several sigma factors, and it also regulates the espA operon in Mtb and modulates ESX-1 function [30]. In Mtb PhoP/R controls 30 genes directly, whereas regulatory cascades are responsible for signal amplification and downstream effects through proteins like EspR, which controls Esx1 function [31]. PhoP also controls the noncoding RNA (ncRNA) Mcr7, which controls the secretion of the twin arginine translocation (Tat) substrates, such as antigen Ag85 complex proteins (Ag85A and Ag85c) [31]. In Mtb the DosR/S TCS regulates the expression of around 48 genes under hypoxic and nitric oxide stress and is required for full virulence and pathogenicity in different animal models [32].

Mcol has genes that code for proteins that are involved in the transportation and regulation of iron, such as irtA, irtB, viuB, ESX-3 and IdeR. To obtain iron Mtb synthesizes a cell-associated siderophore named mycobactin and a secreted one, carboxymycobactin, which sequesters iron and delivers it to the bacterium via specialized Fe$^{3+}$ siderophore transporters [33]. In Mtb Fe-carboxymycobactin can be secreted through IrtA transporters and is translocated by the IrtB-ViuB ABC transporter [34]. The secretion and translocation of mycobactin is not entirely understood, but in M. smegmatis the ESX-3 secretion system plays an important role for the mycobactin-mediated iron uptake [35].

Mtb virulence factors that are probably absent in M. colombiense

Sulfolipid-1 (SL-1) is a cell wall glycolipid and important virulence factor that restricts Mtb growth in macrophages [36]. SL-
I has been also proposed to alter superoxide (O2−) production [37] and blocks phagosome–lysosome fusion [38]. Several enzymatic steps are necessary for the SL-1 biosynthesis; MmpL8 and Sap proteins then transport SL1 across the membrane [39]. Mcol possesses the genes involved in the SL-1 biosynthesis and transport, except the pks2. It has been confirmed that Mtb pks2 mutant does not produce sulfolipids [40]. Therefore, it is highly probable that Mcol does not produce SL-1. Additionally, the production of this glycolipid is suggested to be restricted to pathogenic mycobacteria, especially to members of the Mtb complex.

Mtb produces phospholipase C (PLC), which is an important virulence factor that induces the inhibition of cyclooxygenase-2 expression (COX-2), and this blocks prostaglandin E2 (PGE2) synthesis by macrophages. (PGE2 is an essential factor involved in the activation of the membrane repair mechanism, and therefore its inhibition is associated with necrosis of macrophages [41].) Analysis of the Mcol genome suggests that the genes involved in the phospholipase C production (plcA, plcB, plcC and plcD) are missing. Therefore, it is a feasible that this species does not utilize this virulence mechanism during macrophage infection.

PDIMs and phenolglycolipids (PGLs) are two groups of complex cell wall–associated lipids that are important mycobacterial virulence factors. Members of Mtb complex also produce p-hydroxybenzoic acid derivatives (p-HBADs), which are precursors of PGL biosynthesis. According to VFDB, several genes participating in PDIM/PGL/p-HBAD synthesis and transportation are lacking in Mcol.

The ESX secretion system is a specialized secretion system for the transport of extracellular proteins across the mycobacterial cell wall [42]. In mycobacteria there are five ESX systems (ESX-1, ESX-2, ESX-3, ESX-4, ESX-5), and all of them are present in the Mtb genome. The ESX-1 secretion system exports important Mtb virulence factors such as ESAT-6 and CFP-10, and it also induces apoptosis in macrophages and generates a semiporous phagosome [43,44]. Mcol possesses all of the ESX systems except ESX-1, and Mcol also lacks the ESAT-6 and CFP-10 genes.

Finally, Mcol lacks the hspX gene. The heat shock protein HspX is under the control of the DosR/S TCS during hypoxic conditions, and is suggested to be involved in the long-term survival of Mtb inside macrophages [45].

Probable M. colombiense virulence factors that are not present in Mtb

The GPLs are a class of glycolipids produced by several nontuberculous mycobacteria [46]. GPLs are exported to the cell wall by Gap protein. However, the MmpS and MmpL proteins assemble into a complex that contribute to GPLs biosynthesis and export [47]. GPLs are involved in the inhibition of the phagosome–lysosome fusion [48]. In M. avium GPLs from promote macrophage activation in a TLR2- and MyD88-dependent manner [49]. Mcol possesses a GPL locus; it is highly possible that this species is able to produce GPL. However, experimental studies are necessary to establish the Mcol GPL serovar structure.

In conclusion, Mcol possesses many of the Mtb virulence mechanisms that have been involved in apoptosis, necrosis and inhibition of phagosome–lysosome fusion. However, Mcol has deletions in the genes involved in the p-HBA/ PDIM/PGL, PLC, SL-1 and HspX production, and loss of the ESX-1. The inability to produce these compounds may be related to its lower virulence in mice infected with Mcol compared to mice infected with Mtb [50]. Thus, the opportunistic nature of Mcol could be associated with its inability to produce significant virulence mechanisms used by Mtb during its adaptation to the host. In addition, the singletons and the virulence-associated genes conserved in the Mcol genome could be used to design molecular diagnostic tools.

Acknowledgements

MNG-P was supported by the National University of Colombia (Convocatoria del Programa Nacional de Proyectos para el Fortalecimiento de la Investigación, la Creación y la Innovación en Posgrados; Proyecto 19104). We thank the Institute for Genomes Sciences, University of Maryland, for the Mcol PacBio sequencing. We thank J. Yang (Institute of Pathogen Biology, CAMS and PUMC) for the detection of Mcol virulence factors by VFDB. We also thank S. Soares for the analysis of genomic islands by the GIPSy/PIPS tool.

Conflict of Interest

None declared.

References

[1] Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. Am J Respir Crit Care Med 2007;175:367–416.
[2] Murcia MI, Tortoli E, Menendez MC, Palenque E, Garcia MJ. Mycobacterium colombiense sp. nov., a novel member of the Mycobacterium avium complex and description of MAC-X as a new ITS genetic variant. Int J Syst Evol Microbiol 2006;56(pt 9):2049–54.
[3] Meyer F, Goemann A, McHardy AC, Bartels D, Beek T, Clausen J, et al. GeneDB—an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 2003;31:2187–95.

[4] Sievers F, Wilim A, Dineen D, Gibson Tj, Kaplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011;7:539.

[5] Langille MG, Brinkman FS. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics 2009;25:664–5.

[6] Soares SC, Abreu VA, Ramos RT, Ceberde L, Silva A, Baumbach J, et al. PIPS: pathogenicity island prediction software. PLoS One 2012;7:e30848.

[7] Cosentino S, Voldby Lar森 M, Moller Aarestrup F, Lund O. PathogenFinder—distinguishing friend from foe using bacterial whole genome sequence data. PLoS One 2013;8:e77302.

[8] Chen L, Xiong Z, Sun L, Yang J, Qin J. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res 2012;40(database issue):D641–5.

[9] Blom J, Albaum SP, Doppmeier D, Pulher A, Vorholter FJ, Zakrzewski M, et al. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinformatics 2009;10:154.

[10] Bagos PG, Nikolaou EP, Liakopoulos TD, Hamodrakas SJ. Prediction of protein signal peptide and with hidden Markov models. Bioinformatics 2010;26:2811–7.

[11] Bagos PG, Tsirigos KD, Liakopoulos TD, Hamodrakas SJ. Prediction of lipoprotein signal peptide in Gram-positive bacteria with a hidden Markov model. J Proteome Res 2008;7:5082–93.

[12] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.

[13] Finnerle DK, Tsirigos KD, Litou ZI, Liakopoulos TD, Bagos PG, Hamodrakas SJ. CW-PRED: a HMM-based method for the classification of cell wall–anchored proteins of Gram-positive bacteria. Berlin: Springer; 2012.

[14] Griss I, Vergnaud G, Pourcel C. CRISPRFinder—a web tool to identify clustered regularly interspersed short palindromic repeats. Nucleic Acids Res 2007;35(Web Server issue):W52–9.

[15] Bhubhanil S, Niamyim P, Sukchawalit R, Mongkolsuk S. Cysteine Liao D, Fan Q, Bao L. The role of superoxide dismutase in the survival of Mycobacterium tuberculosis in macrophages. Jpn J Infect Dis 2013;66:480–8.

[16] Sanki AK, Boucau J, Ronning DR, Sucheck SJ. Antigen BSC–mediated acyl-transfer between synthetic acyl donors and fragments of the arabinan. Glycoconj J 2009;26:589–96.

[17] Liao D, Fan Q, Bao L. The role of superoxide dismutase in the survival of Mycobacterium tuberculosis in macrophages. Jpn J Infect Dis 2013;66:480–8.
subverts PGE2 synthesis and induces necrosis in alveolar macrophages.

BMC Microbiol 2014;14:128.

[42] Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luijink J, Vandenbroucke-Grauls CM, et al. Type VII secretion——mycobacteria show the way. Nat Rev Microbiol 2007;5:883–91.

[43] Renshaw PS, Panagiotidou P, Whelan A, Gordon SV, Hewinson RG, Williamson RA, et al. Conclusive evidence that the major T-cell antigens of the Mycobacterium tuberculosis complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. J Biol Chem 2002;277:21598–603.

[44] Simeone R, Bobard A, Lippmann J, Biter W, Majlessi L, Brosch R, et al. Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death. PLoS Pathog 2012;8:e1002507.

[45] Cunningham AF, Spreadbury CL. Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. J Bacteriol 1998;180:801–8.

[46] Ripoll F, Deshayes C, Paez S, Laval F, Beretti JL, Biet F, et al. Genomics of glycopeptidolipid biosynthesis in Mycobacterium abscessus and M. chelonae. BMC Genomics 2007;8:114.

[47] Deshayes C, Bach H, Euphrasie D, Attarian R, Coureuil M, Sougkooff W, et al. MmpS4 promotes glycopeptidolipids biosynthesis and export in Mycobacterium smegmatis. Mol Microbiol 2010;78:989–1003.

[48] Shimada K, Takimoto H, Yano I, Kumazawa Y. Involvement of mannose receptor in glycopeptidolipid-mediated inhibition of phagosome—lysosome fusion. Microbiol Immunol 2006;50:243–51.

[49] Sweet L, Schorey JS. Glycopeptidolipids from Mycobacterium avium promote macrophage activation in a TLR2- and MyD88-dependent manner. J Leukoc Biol 2006;80:415–23.

[50] Gonzalez-Perez M, Marino-Ramirez L, Parra-Lopez CA, Murcia MI, Marquina B, Mata-Espinoza D, et al. Virulence and immune response induced by Mycobacterium avium complex strains in a model of progressive pulmonary tuberculosis and subcutaneous infection in BALB/c mice. Infect Immun 2013;81:4001–12.