FEBS Letters 588 (2014) 3906–3911

journal homepage: www.FEBSLetters.org

# Identification and comparative analysis of a genomic island in *Mycobacterium avium* subsp. *hominissuis*

Annesha Lahiri [a], Andrea Sanchini [a], Torsten Semmler [b], Hubert Schäfer [a], Astrid Lewin [a,*]

[a] Robert Koch Institute, Division 16 Mycotic and Parasitic Agents and Mycobacteria, Nordufer 20, 13353 Berlin, Germany
[b] Freie Universität Berlin, Centre for Infection Medicine, Institute of Microbiology and Epizootics, Robert-von-Ostertag-Str. 7-13, 14163 Berlin, Germany

ARTICLE INFO

*Article history:*
Received 28 July 2014
Revised 27 August 2014
Accepted 29 August 2014
Available online 12 September 2014

Edited by Takashi Gojobori

*Keywords:*
Virulence
Genomic island
Horizontal gene transfer
Genome diversity
*Mycobacterium avium*

ABSTRACT

***Mycobacterium avium* subsp. *hominissuis* (MAH) is an environmental bacterium causing opportunistic infections. The objective of this study was to identify flexible genome regions in MAH isolated from different sources. By comparing five complete and draft MAH genomes we identified a genomic island conferring additional flexibility to the MAH genomes. The island was absent in one of the five strains and had sizes between 16.37 and 84.85 kb in the four other strains. The genes present in the islands differed among strains and included phage- and plasmid-derived genes, integrase genes, hypothetical genes, and virulence-associated genes like *mmpL* or *mce* genes.**


## 1. Introduction

Non-tuberculous Mycobacteria (NTM) are natural inhabitants of soil, dust and water. Recent years have shown a substantial number of NTM as potential pathogens for humans [1]. NTM are opportunistic pathogens that cause lymphadenitis, lung infections, skin and soft tissue infections mainly in immune-compromised hosts [2]. NTM disease is found to increase in some parts of the world [3–6]. One of the clinically most important NTM is *Mycobacterium avium* [7], which together with *Mycobacterium intracellulare*, *Mycobacterium marseillense*, *Mycobacterium timonense*, *Mycobacterium bouchedurhonense*, *Mycobacterium colombiense*, *Mycobacterium vulneris*, *Mycobacterium chimaera*, and *Mycobacterium arosiense* belongs to the *M. avium* complex (MAC) [8–12]. *M. avium* comprises of the four subspecies *M. avium* subsp. *avium* (MAA), *M. avium* subsp. *silvaticum* (MAS), *M. avium* subsp. *hominissuis* (MAH), and *M. avium* subsp. *paratuberculosis* (MAP) [13]. While MAA and MAS cause tuberculosis-like diseases in birds [14], MAP is better known for causing Johne's disease in ruminants and potentially Crohn's disease in humans [15]. MAH is an important intracellular human pathogen affecting immune-compromised populations like older patients and children [16].

Among the *M. avium* subspecies, MAH exhibits the highest degree of sequence variability [17], which is reflected for example by the presence of "Long Sequence Polymorphisms" (LSPs), multiple copies of insertion sequences (e.g. IS1245 and IS900) and a heterogeneous pattern of *hsp65* types. Sequence variability generally originates either from Single Nucleotide Polymorphisms (SNPs) or from the presence of mobile genetic elements such as plasmids, phages, insertion elements and genomic islands (GI)s [16,17]. GIs are genetic entities of horizontally transferred genes that can vary in size from 10 kb to 200 kb [18]. They contribute to rapid evolution and confer survival advantages to the bacteria. GIs can be identified in bacteria based on differential GC content, the presence of foreign genes and flanking regions containing short direct repeats and transfer ribonucleic acid (t-RNA) genes [19,20]. The contribution of GIs to the evolution of *Mycobacterium tuberculosis* has been demonstrated by Becq et al. [21], who showed that around 5% of the *M. tuberculosis* genome has been acquired by horizontal gene transfer. They found virulence genes to be slightly overrepresented in the GIs (8.2% of the GI genes) compared to their proportion in the whole genome (6.5% of the whole genome genes).

Though a thorough comparative study of the two complete MAH genomes, the MAH 104 and MAH TH135 has already been performed [22], the individual sequence polymorphisms within

*Abbreviations:* GI, genomic island; LSP, Long Sequence Polymorphism; MAA, *Mycobacterium avium* subsp. *avium*; MAC, *Mycobacterium avium* complex; MAH, *Mycobacterium avium* subsp. *hominissuis*; MAP, *Mycobacterium avium* subsp. *paratuberculosis*; MAS, *Mycobacterium avium* subsp. *silvaticum*; SNP, Single Nucleotide Polymorphism

* Corresponding author. Fax: +49 (30) 18754 2110.
*E-mail address:* LewinA@rki.de (A. Lewin).

http://dx.doi.org/10.1016/j.febslet.2014.08.037

the different MAH strains need a detailed investigation. The objective of this study was to identify regions with flexible gene pools in different MAH strains isolated from different sources such as patients, animals, or environment. Exploring such flexible gene pools in MAH will help us understand the evolution of MAH from an environmental bacterium to an opportunistic pathogen.

## 2. Materials and methods

### 2.1. Comparison of different mycobacterial genomes

A comparison of the MAH 104 genome with other members of the MAC was performed with VISTA gateway (http://pipeline.lbl.gov/cgi-bin/gateway2). The VISTA browser contains microbial genomes that are available in the form of precompiled alignments. The MAH 104 was selected as a reference genome and genomes from members of the MAC like the MAP K-10 (Accession Number AE016958) and MAA ATCC 25291 (Accession Number ACFI00000000.1) and *M. intracellulare* ATCC 13950 (Accession Number CP003322) were used for further comparison.

### 2.2. Identification of plausible genomic islands

Island Viewer (http://www.pathogenomics.sfu.ca/islandviewer/query.php) and Alien Hunter (http://bioinformatics.oxfordjournals.org/content/22/18/2196.full.pdf+html) software were used as prediction tools for computational visualization of GIs in MAH. Screening of flexible gene pools based on the presence of flanking t-RNA genes, existence of direct repeats, and GC content of the GIs was accomplished by Geneious version 7.1.4 (Biomatters, New Zealand, http://www.geneious.com/). GC profile (http://tubic.tju.edu.cn/GC-Profile/) was used to determine the GC content of the GIs. MAFFT Alignment was used for alignment of direct repeats. Analysis of GIs and similarity searches was performed by NCBI BLAST.

### 2.3. Identification of genomic islands in other M. avium subsp. hominissuis genomes

Comparative analysis was performed with MAH 104 (isolate from a HIV patient with disseminated MAH infection, U.S., Accession Number CP000479), MAH TH135 (isolate from a non-HIV patient with lung infection, Japan, Accession Number AP012555), MAH 27-1 (isolate from household dust, Germany, Accession Number AWXK00000000.1), MAH 2721 (isolate from a child with MAH lymphadenitis, Germany, Accession Number AWXJ00000000.1) and MAH 10-4249 (isolate from a deer, U.S., Accession Number AYNQ00000000.1). The flanking genes present on either sides of the region of diversity (namely O′methyltransferase gene, MAV_0778 and carveol dehydrogenase gene, MAV_0846) were used for the identification of the GIs in different annotated as well as non-annotated genomes of MAH.

## 3. Results

### 3.1. Comparison of different mycobacterial genomes and identification of plausible genomic islands in strain MAH 104

A comparative analysis with the VISTA gateway resulted in the identification of seven specific regions that were explicitly found in MAH 104 but were absent in MAP K-10, MAA ATCC 25291 and *M. intracellulare* ATCC 13950 genomes. The regions identified by the comparative analysis are presented in Table 1. Their sizes varied from 22.85 kb to 199.29 kb. Since bacterial t-RNA genes are insertion hotspots for GIs [19] these seven regions were checked for the presence of flanking t-RNA genes. Three regions (regions 1, 3 and 4) were identified with flanking t-RNA genes. Region 1 was flanked by two t-RNA genes, the t-RNA-serine and t-RNA-arginine. Region 3 was flanked by t-RNA-lysine, t-RNA-glutamine, t-RNA-aspartate and t-RNA-phenylalanine and region 4 by t-RNA-arginine. The software IslandViewer was used thereafter to explore the likelihood of GIs in the seven regions revealing that the regions 1, 2, 4, 5, 6 and 7 contained probable GIs. The results obtained from IslandViewer are shown in Table 1. No GI was predicted in region 3. It was interesting to note that region 3, identified as a region unique to MAH 104 according to VISTA Gateway, had no GI projected by IslandViewer despite 4 flanking t-RNAs. We therefore additionally used the software Alien Hunter to search for the existence of GIs in region 3 and found two stretches covering most of the region 3 containing genes putatively acquired by horizontal gene transfer. Hence we decided to analyze region 3 for presence of further features characterizing GIs. Regions 2, 5, 6 and 7 were excluded from further analysis because they did not contain any flanking tRNA genes and the regions 1 and 4 because they did not contain any DNA repeats at the extremities and exhibited no relevant difference in GC content compared to the whole genome.

**Table 1**
Seven regions specific to the MAH 104 identified by comparison with the genomes from MAP K-10, MAA ATCC 25291 and *M. intracellulare* ATCC 13950 using the VISTA gateway tool and genomic islands predicted by IslandViewer.

| Regions identified | Start position–end position of the region | Size of region (kb) | Genes | Start position–end position of predicted genomic islands | Number of predicted genomic islands |
|---|---|---|---|---|---|
| Region 1 | 254394–294226 | 39.83 | MAV_0253–MAV_0298 | 252926–294892 | 1 |
| Region 2 | 461330–493978 | 32.64 | MAV_0471–MAV_0508 | 483932–488134 | 1 |
| Region 3 | 746939–794035 | 47.95 | MAV_0779–MAV_0841 | No island predicted | 0 |
| Region 4 | 1424505–1463494 | 38.98 | MAV_1458–MAV_1506 | 1456328–1463365 | 1 |
| Region 5 | 1788529–1987820 | 199.29 | MAV_1793–MAV_2005 | 1801703–1805910<br>1809699–1816301<br>1823660–1831906<br>1878666–1884543<br>1919399–1927530<br>1956188–1961917<br>1977459–1990545 | 7 |
| Region 6 | 2548507–2724198 | 175.65 | MAV_2515–MAV_2689 | 2559817–2583641<br>2639752–2645183<br>2683125–2703539<br>2683128–2703539<br>2708730–2713220 | 5 |
| Region 7 | 3916471–3939322 | 22.85 | MAV_3789–MAV_3809 | 3919180–3939322 | 1 |