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

CrossMark

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

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.
© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

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

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://pipe-line.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 Island-Viewer 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	254 394–294 226	39.83	MAV_0253–MAV_0298	252 926–294 892	1
Region 2	461 330–493 978	32.64	MAV_0471–MAV_0508	483 932–488 134	1
Region 3	746 939–794 035	47.95	MAV_0779–MAV_0841	No island predicted	0
Region 4	1 424 505–1 463 494	38.98	MAV_1458–MAV_1506	1 456 328–1 463 365	1
Region 5	1 788 529–1 987 820	199.29	MAV_1793–MAV_2005	1 801 703–1 805 910 1 809 699–1 816 301 1 823 660–1 831 906 1 878 666–1 884 543 1 919 399–1 927 530 1 956 188–1 961 917 1 977 459–1 990 545	7
Region 6	2 548 507–2 724 198	175.65	MAV_2515–MAV_2689	2 559 817–2 583 641 2 639 752–2 645 183 2 683 125–2 703 539 2 683 128–2 703 539 2 708 730–2 713 220	5
Region 7	3 916 471–3 939 322	22.85	MAV_3789–MAV_3809	3 919 180–3 939 322	1

3.2. Characterization of a genomic island in strain MAH 104

Region 3 indeed contained direct repeats of 45 bp at the extremities of the potential GI. The nucleotide sequences of the repeats are given in Table 2. The potential GI in region 3 was composed of 66.8% GC compared to 69.0% in the whole genome from strain 104. It comprised of 63 genes from MAV_0779 to MAV_0841 (47.95 kb) as shown in Fig. 1. Among these genes are phage-element specific genes, which are typical genes within GIs, like the site-specific recombinase, phage integrase family protein, phage terminase protein, the phage tail protein genes and glycoprotein genes. 39 genes in this region were hypothetical. Four other miscellaneous genes namely a PPE gene, the 30 kDa protein gene, the 17 kDa surface antigen family protein gene and the excisionase family gene were also present. Similarity searches of draft genomes revealed that the region 3 shared close homology with a region from *Mycobacterium xenopi* RIVM700367 (Table 3 and Supplementary Table 1). As shown in Table 3 the region 3 from MAH 104 thus fulfills all the criteria applied to define a GI.

3.3. Characterization of the new genomic island in other MAH strains

The genes flanking the region 3, namely the methyltransferase (MAV_0778) and carveol dehydrogenase (MAV_0846) were given importance as these genes are conserved in MAH strains. The flanking genes and their homology facilitated the identification of the new GI in different MAH strains. Table 3 and Fig. 1 give an overview of the GIs and their composition in the analyzed strains and also about the similarities between the GIs of different isolates.

We first searched this region in the MAH TH135. While both the methyltransferase gene (MAH_0683) and the carveol dehydrogenase gene (MAH_0686) as well as the four tRNA genes flanking the GI from strain 104 could be identified, other features defining a GI such as direct repeats, foreign genes, phage genes, integrase genes, were absent (Table 3 and Fig. 1). Only two hypothetical genes namely the MAH_0684 and MAH_0685 were identified in the region. Overall the region in strain TH135 did not show the properties qualifying it as GI and we concluded that this specific GI was missing in TH135.

The GI was examined thereafter in the draft genome of MAH 27-1. The study exposed a completely dissimilar 16.37 kb GI

(Accession Number KM206567) in this case which was different from the one observed in MAH 104. It was flanked by the four tRNA genes on one side and direct repeats of 49 bp on both sides and exhibited a drop in GC content from 69.0% in the whole genome to 61.9% in the GI. Also genes typically occurring in GIs such as integrase and phage- and plasmid-derived genes were present. While the four tRNA genes were found next to the gene encoding the carveol dehydrogenase in strain 104, they were located next to the gene encoding the methyltransferase in strain 27-1. The GI when exposed to BLAST analysis with complete and draft genomes showed high levels of homology with *M. intracellulare* 05-1390, with *Mycobacterium yongonense* MOTT36Y, with *Mycobacterium indicus pranii* MTCC 9506 and with *Mycobacterium marinum* M (Table 3 and Supplementary Table 1). Details about the structure of the GI and the genes present are listed in Table 3 and Fig. 1.

The 16.37 kb region identified in MAH 27-1 was also found in the GI from strain MAH 2721 (Accession Number KM105871). The investigation of the draft genome of MAH 2721 revealed a GI of 84.85 kb. Adjacent to the region in MAH 2721 homologous to the GI from strain 27-1, a second direct repeat and an integrase gene were identified followed by a 68.48 kb region containing 74 CDS. Adjacent to this 68.48 kb region we found a third copy of the direct repeat followed by the four tRNA genes. The genes present in this GI are listed in Fig. 1. The drop in GC content in this GI amounts to 63.5% compared to 68.6% in the whole genome (Table 3).

Most of the genes present in the GI from strain 27-1 could be found in a 16.40 kb GI from strain 10-4249 with four additional genes present in 10-4249 compared to 27-1 (Fig. 1). This GI was identified on contig 582 and contig 583 of the 10-4249 draft genome. The drop of GC percentage was from 69.2% GC in the whole genome to 62.5% in the GI. Flanking tRNA genes and presence of integrase genes and direct repeats of 51 bp additionally confirmed this region as GI (Table 3). Homology searches revealed similar DNA sequences to be present in *M. indicus pranii* MTCC 9506, *M. marinum* M, *M. intracellulare* MOTT-02, and *M. yongonense* MOTT36Y. See Fig. 1 for the list of genes.

4. Discussion

Members of *M. avium* are found to be phenotypically and genotypically diverse. Our intention was to further characterize the flexible gene pool from MAH. Several GIs have been identified in *M. avium* until now but only a few of them have been designated specific functions [23,24]. One of the most recent findings was the identification of a GI in *M. avium* specific for macrophage and amoeba infections [23]. Another of its kind was the identification of a 38 kb pathogenicity island in MAP that encodes cell surface proteins expressed in hosts [24]. Numerous LSPs have also been identified within the *M. avium* but the functions associated with these polymorphisms have not been labeled [25].

Seven regions of diversity were identified in MAH 104 in this study by whole genome comparative analysis of MAH 104, MAP-K10 and MAA ATCC 25291 and *M. intracellulare* ATCC 13950. This result was analogous to a study performed by Uchiya et al. in Japan [22] who performed a similar study with MAH 104 and the MAH TH135. Eleven regions specific to the MAH 104 were identified in their comparative analysis with MAH TH135 and seven of the eleven regions were identical to the regions of specificity identified by this study. The regions SR11, SR12, SR14, SR15, SR16, SR19 and SR20 identified in the study by Uchiya and colleagues correspond to the regions 1, 2, 3, 4, 5, 6 and 7 shown here.

Within region 3 we identified a 47.95 kb sequence that exhibited all features characterizing GIs, such as the location next to tRNA genes, the presence of direct repeats on the flanks of the

Table 2
Direct repeats flanking the genomic islands from strains MAH 104, 27-1, 2721 and 10-4249.

MAH strain	Sizes of direct repeats [bp]	Sequences of direct repeats
104	45	GTGCCCCCACTAGGACTCGAACCTAGGACCTGCGGATTAAAAGTC
	45	GTGCCCCCACCAGGGCTCGAACCTGGGACCTGCGGATTAAAAGTC
27–1	49	GGTGGCCAGGGGCGGGATCGAACCGCCGACCTTCCGCTTTTCAGGCGGA
	49	GGTGGCCAGGGGCGGGATCGAACCGCCGACCTTCCGCTTTTCAGGTGGA
2721	49	GGTGGCCAGGGGCGGGATCGAACCGCCGACCTTCCGCTTTTCAGACGGA
	49	GGTGGCCAGGGGCGGGATCGAACCGCCGACCTTCCGCTTTTCAGGCGGA
	44	GGTGGCCAGGGGCGGGATCGAACCGGCGACCTTCCGCTTTTCAG
10–4249	51	GGTGGCCAGGGGCGGGATCGAACCGCCGACCTTCCGCTTTTCAGGTGGGCG
	51	GGTGGCCAGGGGCGGGATCGAACCGCCGACCTTCCGCTTTTCAGGCGGACG

Fig. 1. Comparison of genes in genomic islands from different MAH strains.

Mycobacterium avium subsp. *hominissuis* strains

Complete genomes

TH135
- Methyltransferase MAH_0683
- Hypothetical protein MAH_0684
- Hypothetical protein MAH_0685
- tRNA-Lys MAH_t013
- tRNA-Glu MAH_t014
- tRNA-Asp MAH_t015
- tRNA-Phe MAH_t016
- Carveol dehydrogenase MAH_0686

10-4249
- Methyltransferase O971_18005
- Not annotated hypothetical protein
- Not annotated hypothetical protein
- t-RNA-Lys O971_18000
- t-RNA-Glu O971_17995
- t-RNA-Asp O971_17990
- t-RNA-Phe O971_17985
- DNA Duplication (51 bp)
- integrase O971_17980
- Cell division protein FtsK O971_17975
- Plasmid replication O971_17970
- Hypothetical protein O971_17965
- Not annotated Transcriptional regulator
- Hypothetical protein O971_17960
- Regulator O971_17955
- Not annotated hypothetical protein
- TetR family transcriptional regulator O971_17950
- Hypothetical protein O971_17945
- alpha/beta hydrolase O971_17940
- Membrane protein; disrupted O971_17935
- Not annotated hypothetical protein
- Transcriptional regulator, TetR O971_17930
- Cytochrome P450 O971_17925
- Transcriptional regulator, TetR O971_17920
- Hypothetical protein O971_17915
- Not annotated hypothetical protein
- DNA Duplication (51 bp)
- Oxidoreductase O971_17910

Draft genomes

27-1
- O-Methyltransferase
- tRNA-Lys
- tRNA-Glu
- tRNA-Asp
- tRNA-Phe
- DNA Duplication (49 bp)
- Integrase
- Transfer protein traSA
- FIG00831412: hypothetical protein
- Not annotated hypothetical protein
- Transcriptional regulator KorSA, GntR family
- Nudix hydrolase family protein
- Hypothetical protein
- FIG00825226: hypothetical protein
- Hydrolase
- hypothetical protein
- hypothetical protein
- Transcriptional regulator, TetR family
- putative cytochrome P450 hydroxylase
- Transcriptional regulator, TetR family
- Hypothetical protein
- hypothetical protein
- DNA Duplication (49 bp)
- Integrase
- hypothetical protein
- hypothetical protein
- hypothetical protein
- hypothetical protein
- hypothetical protein
- hypothetical protein
- hypothetical protein
- hypothetical protein
- hypothetical protein
- putative cytochrome P450 hydroxylase
- putative cytochrome P450 hydroxylase
- NirP, putative
- Cytochrome c oxidase polypeptide III (EC 1.9.3.1)
- FIG00827853: hypothetical protein
- FIG00827853: hypothetical protein
- Transcriptional regulator, TetR family
- Conserved hypothetical integral membrane protein YrbE1A
- Conserved hypothetical integral membrane protein YrbE1B
- MCE-family protein Mce1A
- MCE-family protein Mce1B
- MCE-family protein Mce1C
- hypothetical protein
- MCE-family lipoprotein LprK (MCE-family lipoprotein Mce1e)
- MCE-family protein Mce1F
- hypothetical protein
- hypothetical protein
- hypothetical protein
- Transcriptional regulator, TetR family
- Transcriptional regulator, TetR family
- hypothetical protein
- Possible membrane protein
- Possible membrane protein
- Nitric oxide reductase activation protein NorQ
- Nitric oxide reductase activation protein NorD
- Aldehyde dehydrogenase
- Phenylpropionate dioxygenase
- FIG00821688: hypothetical protein
- hypothetical protein
- Transcriptional regulator, MarR family
- Phenylacetic acid degradation NADH oxidoreductase paaE
- 3-oxoacyl-[acyl-carrier protein] reductase
- Phenylpropionate dioxygenase
- FIG00822032: hypothetical protein
- Taurine catabolism dioxygenase TauD/TfdA
- hypothetical protein
- hypothetical protein
- putative cytochrome P450 hydroxylase
- putative cytochrome P450 hydroxylase
- FIG00820772: hypothetical protein
- Acyl-CoA dehydrogenase, short-chain specific
- FIG00822014: hypothetical protein
- FIG00822014: hypothetical protein
- Short chain dehydrogenase
- Esterase LipL
- hypothetical protein
- hypothetical protein
- hypothetical protein
- hypothetical protein
- Mobile element protein
- hypothetical protein
- putative cytochrome P450 hydroxylase
- Short-chain dehydrogenase/reductase
- Cyclohexanone monooxygenase
- putative cytochrome P450 hydroxylase
- Esterase/lipase
- 3-oxoacyl-[acyl-carrier protein] reductase
- Transcriptional regulator, TetR family
- hypothetical protein
- hypothetical protein
- Cyclohexanone monooxygenase
- oxidoreductase, short-chain dehydrogenase/reductase family
- Purine/pyrimidine phosphoribosyl transferase
- hypothetical protein
- DNA Duplication (49 bp)
- FIG00824548: hypothetical protein

2721
- O-Methyltransferase
- tRNA-Lys
- tRNA-Glu
- tRNA-Asp
- tRNA-Phe
- DNA Duplication (49 bp)
- Integrase
- Transfer protein traSA
- FIG00831412: hypothetical protein
- Not annotated hypothetical protein
- Transcriptional regulator KorSA, GntR family
- Nudix hydrolase family protein
- Hypothetical protein
- FIG00825226: hypothetical protein
- Hydrolase
- hypothetical protein
- hypothetical protein
- Transcriptional regulator, TetR family
- putative cytochrome P450 hydroxylase
- Transcriptional regulator, TetR family
- Hypothetical protein
- hypothetical protein
- DNA Duplication (49 bp)

104
- Methyltransferase MAV_0778
- DNA Duplication (45 bp)
- Recombinase MAV_0779
- Hypothetical protein MAV_0780
- Hypothetical protein MAV_0781
- Hypothetical protein MAV_0782
- Hypothetical protein MAV_0783
- Hypothetical protein MAV_0784
- Hypothetical protein MAV_0785
- gp53 protein MAV_0786
- gp50 protein MAV_0787
- Hypothetical protein MAV_0788
- Hypothetical protein MAV_0789
- ppe family protein MAV_0790
- Hypothetical protein MAV_0791
- gp34 protein MAV_0792
- gp27 protein MAV_0793
- gp28 protein MAV_0794
- Hypothetical protein MAV_0795
- Hypothetical protein MAV_0796
- gp37 protein MAV_0797
- gp36 protein MAV_0798
- gp23 protein MAV_0799
- Phage tail tape measure protein MAV_0800
- Hypothetical protein MAV_0801
- gp32 protein MAV_0802
- gp31 protein MAV_0803
- Hypothetical protein MAV_0804
- Hypothetical protein MAV_0805
- Hypothetical protein MAV_0806
- Hypothetical protein MAV_0807
- Hypothetical protein MAV_0808
- Hypothetical protein MAV_0809
- Hypothetical protein MAV_0810
- Hypothetical protein MAV_0811
- Hypothetical protein MAV_0812
- Phage terminase MAV_0813
- Hypothetical protein MAV_0814
- Hypothetical protein MAV_0815
- 30kDa protein MAV_0816
- Hypothetical protein MAV_0817
- Hypothetical protein MAV_0818
- gp75 protein MAV_0819
- Hypothetical protein MAV_0820
- Hypothetical protein MAV_0821
- gp79 protein MAV_0822
- gp78 protein MAV_0823
- Hypothetical protein MAV_0824
- Hypothetical protein MAV_0825
- Hypothetical protein MAV_0826
- Hypothetical protein MAV_0827
- 17 kDa surface antigen MAV_0828
- YqaK protein MAV_0829
- gp60 protein MAV_0830
- Hypothetical protein MAV_0831
- Hypothetical protein MAV_0832
- Hypothetical protein MAV_0833
- Hypothetical protein MAV_0834
- Hypothetical protein MAV_0835
- Hypothetical protein MAV_0836
- gp54 protein MAV_0837
- Hypothetical protein MAV_0838
- gp54 protein MAV_0839
- Hypothetical protein MAV_0840
- Excisionase DNA binding protein MAV_0841
- DNA Duplication (45 bp)
- tRNA-Lys MAV_0842
- tRNA-Glu MAV_0843
- tRNA-Asp MAV_0844
- tRNA-Phe MAV_0845
- Carveol dehydrogenase MAV_0846

Fig. 1. Comparison of genes in genomic islands from different MAH strains. Genes highlighted with the same color share more than 70% of DNA identity. White genes represent genes present only in a specific island. The annotations are available in the NCBI database (Accession Numbers: CP000479 for MAH 104, AP012555 for MAH TH135, KM206567 for MAH 27-1, KM105871 for MAH 2721, AYNQ00000000.1 for MAH 10-4249).

GI, a GC content differing from the host genome, integrase genes and other genes originating from mobile genetic elements and the presence of foreign genes originating from other bacterial species [19]. Interestingly, the program IslandViewer could not detect this GI. The most probable explanation is the relatively moderate difference in GC content in the GI from region 3 and the remaining genome of only 2.2%. Missing difference in GC content between GIs and host genomes was also reported by other authors [26] and was attributed to either adaptation to the host genome or uptake of the foreign DNA from a related organism. The most striking element of this GI from strain MAH 104 differentiating it from the other investigated GIs is the high percentage of mycobacteriophage-derived genes pointing to transfer by transduction. This is in good agreement with the data reported in the

Table 3
Characteristics of the genome regions containing the genomic islands identified among *Mycobacterium avium* subsp. *hominissuis* strains.

Characteristics	*Mycobacterium avium* subsp. *hominissuis* strains				
	Complete genomes		Draft genomes		
	104	TH135	27-1	2721	10-4249
Accession Number	CP000479	AP012555	AWXK00000000.1	AWXJ00000000.1	AYNQ00000000.1
Start and end of the genome region[a]	nt 745460–795760	nt 688025–693241	Contigs 038 and 032	Contigs 0338 and 0341	Contigs 000583 and 000582
Start and end of the genomic island[b]	nt 746457–794400	Not applicable	Contigs 038 and 032	Contigs 0338 and 0341	Contigs 000583 and 000582
Length of the genomic island (bp)	47946	Not applicable	16372	84851	16400
No. of tRNAs	4	4	4	4	4
GC content (%) of the genomic island vs.	66.8 vs.	Not applicable	61.9 vs.	63.5 vs.	62.5 vs.
GC content (%) of the strain	69.0	Not applicable	69.0	68.6	69.2
No. of integrases, recombinases	2	Not applicable	1	3	1
No. of phage or plasmid genes	16	Not applicable	1	4	1
Number of direct repeats	2	0	2	3	2
Length of direct repeats (bp)	45	Not applicable	49	49, 44	51
Other mycobacteria showing DNA identity to the genomic island[c]	M. xenopi RIVM700367	Not applicable	M. yongonense MOTT36Y M. intracellulare 05-1390 M. indicus pranii MTCC 9506 M. marinum M M. intracellulare MOTT-02	M. yongonense MOTT36Y M. intracellulare 05-1390 M. intracellulare MOTT-02 M. indicus pranii MTCC 9506 M. intracellulare MOTT-64	M. indicus pranii MTCC 9506 M. marinum M M. intracellulare MOTT-02 M. yongonense MOTT36Y M. intracellulare 05-1390

[a] The genome region is defined as the region within the methyltransferase and the carveol dehydrogenase genes. For MAH 27-1 and 2721 the genomic regions are also available in NCBI Accession Numbers KM206567 and KM105871, respectively.
[b] The genomic island is defined as the region within the direct repeats. In case of genomes fully sequenced, the nt positions are given. In case of draft genomes, the contig numbers where the island has been found are given.
[c] As compared to the Genbank database. More details regarding DNA identities in the Supplementary Table 1.

ACLAME (A CLAssification of Mobile genetic Elements) database, which classifies a 40.50 kb region with 45 genes from MAV_0786 to MAV_0830 located within the 47.95 kb region described in this study as a prophage [27].

Comparative analysis with other MAH genomes revealed that this GI represented a highly dynamic genetic region and had a different gene composition in each strain. The GI was completely absent in strain MAH TH135. While the GIs from strains 27-1, 2721 and 10-4249 exhibited different degrees of homology, the GI from MAH 104 was unique. In this strain the tRNA genes were located next to the gene for the carveol dehydrogenase, whereas in all other strains they were located next to the gene for the methyltransferase. In addition, the direct repeats flanking the GI from strain 104 differed from the repeat sequences in the other strains. The MAH 104 GI contained phage-derived genes which were less abundant in the GIs form the other strains. These GIs, however, all carried a gene annotated as "plasmid replication integration and excision" and transfer of plasmid sequences might have played a role in their generation. The GI from strain 2721, which has a size of 84.85 kb seems to be composed of two segments. One part (16.37 kb) is similar to the GI found in strain 27-1, while the second segment is completely different from the other investigated GIs. Interestingly, a direct repeat and an integrase gene were found between segment one and segment two, pointing to two successive episodes of horizontal gene transfer events, a process known as tandem accretion [28]. Within the set of investigated strains neither the geographic origin nor the ecological environment correlated with the type of GI present. For example, the GI found in the strain MAH 27-1 from household dust in Germany is very similar to the GI in strain MAH 10-4249 from a deer in the US. It was also not possible to correlate the gene composition of the GIs with the virulence of the strains. Strain TH135, for example, which originated from a non-HIV patient and consequently must be classified as virulent, is lacking this GI and must have virulence-associated genes at other genome sites. Among the species showing the highest homologies to parts of the GIs were *M. xenopi*, *M. marinum*, *M. intracellulare*, *M. yongonense* and *M. indicus pranii*, which are all slow- or intermediate growing mycobacteria. *M. intracellulare*, *M. indicus pranii* and *M. yongonense* either belong to or are closely related to MAC [29,30] and *M. indicus pranii* has been proposed to be a predecessor of MAC [31]. *M. xenopi* and *M. marinum*, however, are not closely related to MAC [32]. The homologous DNA regions in *M. xenopi* and *M. marinum* are also composed by GIs supporting horizontal gene transfer to be the driving force behind the evolution of this GI.

Our focus will now be on investigating the role of the GI structure for the virulence of MAH. All analyzed GIs contain genes that have been related to pathogenicity of mycobacteria. For example, the GIs from strains MAH 27-1 and MAH 10-4249 contain a gene homologous to *mmpl10*. Mmpl proteins are involved in the transport of lipids and in cell wall biosynthesis and contribute to the virulence of *M. tuberculosis* [33]. The GI from strain MAH 2721 contains genes belonging to the *mce* (macrophage cell entry) family. This class of genes is a crucial virulence factor of *M. tuberculosis* involved in invasion and survival within macrophages [34]. Genes encoding nitric oxide reductase activation proteins are also present in this GI and might influence the resistance of the bacteria towards the damaging effects of nitric oxide synthesized by activated macrophages. The GI from strain 104 contains a PPE family gene (MAV_0790). PPE proteins are recognized as virulence factors participating in antigenic variation and host immune evasion [35]. Linkage of PPE genes to GIs has been suggested by Saini et al. [36], who described that in *M. indicus pranii* PPE clusters were in majority related to mobile genetic elements. Overall it thus can be reasonably assumed that the newly described GI has an impact on the virulence of MAH.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2014.08.037.

References

[1] Heifets, L. (2004) Mycobacterial infections caused by nontuberculous mycobacteria. Semin. Respir. Crit. Care Med. 25, 283–295.

[2] Khan, K., Wang, J. and Marras, T.K. (2007) Nontuberculous mycobacterial sensitization in the United States: national trends over three decades. Am. J. Respir. Crit. Care Med. 176, 306–313.

[3] Van Ingen, J., Bendien, S.A., De Lange, W.C.M., Hoefsloot, W., Dekhuijzen, P.N.R., Boeree, M.J. and Van Soolingen, D. (2009) Clinical relevance of non-tuberculous mycobacteria isolated in the Nijmegen-Arnhem region, The Netherlands. Thorax 64, 502–506.

[4] Lai, C.C. et al. (2010) Increasing incidence of nontuberculous mycobacteria, Taiwan, 2000–2008. Emerg. Infect. Dis. 16, 294–296.

[5] Al Houqani, M., Jamieson, F., Chedore, P., Mehta, M., May, K. and Marras, T.K. (2011) Isolation prevalence of pulmonary nontuberculous mycobacteria in Ontario in 2007. Can. Respir. J. 18, 19–24.

[6] Moore, J.E., Kruijshaar, M.E., Ormerod, L.P., Drobniewski, F. and Abubakar, I. (2010) Increasing reports of non-tuberculous mycobacteria in England, Wales and Northern Ireland, 1995–2006. BMC Public Health 10.

[7] Tortoli, E. (2009) Clinical manifestations of nontuberculous mycobacteria infections. Clin. Microbiol. Infect. 15, 906–910.

[8] Bang, D., Herlin, T., Stegger, M., Andersen, A.B., Torkko, P., Tortoli, E. and Thomsen, V.O. (2008) Mycobacterium arosiense sp. nov., a slowly growing, scotochromogenic species causing osteomyelitis in an immunocompromised child. Int. J. Syst. Evol. Microbiol. 58, 2398–2402.

[9] Salah, I.B., Cayrou, C., Raoult, D. and Drancourt, M. (2009) Mycobacterium marseillense sp. nov., Mycobacterium timonense sp. nov. and Mycobacterium bouchedurhonense sp. nov., members of the Mycobacterium avium complex. Int. J. Syst. Evol. Microbiol. 59, 2803–2808.

[10] Murcia, M.I., Tortoli, E., Menendez, M.C., Palenque, E. and Garcia, M.J. (2006) 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. 56, 2049–2054.

[11] van Ingen, J., Boeree, M.J., Kosters, K., Wieland, A., Tortoli, E., Dekhuijzen, P.N. and van Soolingen, D. (2009) Proposal to elevate Mycobacterium avium complex ITS sequevar MAC-Q to Mycobacterium vulneris sp. nov. Int. J. Syst. Evol. Microbiol. 59, 2277–2282.

[12] Tortoli, E. et al. (2004) Proposal to elevate the genetic variant MAC-A, included in the Mycobacterium avium complex, to species rank as Mycobacterium chimaera sp. nov. Int. J. Syst. Evol. Microbiol. 54, 1277–1285.

[13] Mijs, W., de Haas, P., Rossau, R., Van Der Laan, T., Rigouts, L., Portaels, F. and van Soolingen, D. (2002) Molecular evidence to support a proposal to reserve the designation Mycobacterium avium subsp. avium for bird-type isolates and 'M. avium subsp. hominissuis' for the human/porcine type of M. avium. Int. J. Syst. Evol. Microbiol. 52, 1505–1518.

[14] Dvorska, L. et al. (2003) A standardised restriction fragment length polymorphism (RFLP) method for typing Mycobacterium avium isolates links IS901 with virulence for birds. J. Microbiol. Methods 55, 11–27.

[15] Feller, M. et al. (2007) Mycobacterium avium subspecies paratuberculosis and Crohn's disease: a systematic review and meta-analysis. Lancet Infect. Dis. 7, 607–613.

[16] Ignatov, D., Kondratieva, E., Azhikina, T. and Apt, A. (2012) Mycobacterium avium-triggered diseases: pathogenomics. Cell. Microbiol. 14, 808–818.

[17] Rindi, L. and Garzelli, C. (2014) Genetic diversity and phylogeny of Mycobacterium avium. Infect. Genet. Evol. 21, 375–383.

[18] Juhas, M., Van Der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W. and Crook, D.W. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. 33, 376–393.

[19] Dobrindt, U., Hochhut, B., Hentschel, U. and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. 2, 414–424.

[20] Schmidt, H. and Hensel, M. (2004) Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. 17, 14–56.

[21] Becq, J., Gutierrez, M.C., Rosas-Magallanes, V., Rauzier, J., Gicquel, B., Neyrolles, O. and Deschavanne, P. (2007) Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. Mol. Biol. Evol. 24, 1861–1871.

[22] Uchiya, K.I. et al. (2013) Comparative genome analysis of mycobacterium avium revealed genetic diversity in strains that cause pulmonary and disseminated disease. PLoS One 8.

[23] Danelishvili, L. et al. (2007) Identification of Mycobacterium avium pathogenicity island important for macrophage and amoeba infection. Proc. Natl. Acad. Sci. U.S.A. 104, 11038–11043.

[24] Stratmann, J. et al. (2004) A 38-kilobase pathogenicity island specific for Mycobacterium avium subsp. paratuberculosis encodes cell surface proteins expressed in the host. Infect. Immun. 72, 1265–1274.

[25] Semret, M., Turenne, C.Y., De Haas, P., Collins, D.M. and Behr, M.A. (2006) Differentiating host-associated variants of Mycobacterium avium by PCR for detection of large sequence polymorphisms. J. Clin. Microbiol. 44, 881–887.

[26] Blanc-Potard, A.B. and Lafay, B. (2003) MgtC as a horizontally-acquired virulence factor of intracellular bacterial pathogens: evidence from molecular phylogeny and comparative genomics. J. Mol. Evol. 57, 479–486.

[27] Leplae, R., Hebrant, A., Wodak, S.J. and Toussaint, A. (2004) ACLAME: a classification of mobile genetic elements. Nucleic Acids Res. 32, D45–D49.

[28] Bellanger, X., Payot, S., Leblond-Bourget, N. and Guédon, G. (2014) Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol. Rev. 38, 720–760.

[29] Saini, V., Raghuvanshi, S., Talwar, G.P., Ahmed, N., Khurana, J.P., Hasnain, S.E., Tyagi, A.K. and Tyagi, A.K. (2009) Polyphasic taxonomic analysis establishes Mycobacterium indicus pranii as a distinct species. PLoS One 4.

[30] Kim, B.J., Math, R.K., Jeon, C.O., Yu, H.K., Park, Y.G., Kook, Y.H. and Kim, B.J. (2013) Mycobacterium yongonense sp. nov., a slow-growing non-chromogenic species closely related to Mycobacterium intracellulare. Int. J. Syst. Evol. Microbiol. 63, 192–199.

[31] Ahmed, N., Saini, V., Raghuvanshi, S., Khurana, J.P., Tyagi, A.K., Tyagi, A.K. and Hasnain, S.E. (2007) Molecular analysis of a leprosy immunotherapeutic bacillus provides insights into Mycobacterium evolution. PLoS One 2.

[32] Roth, A., Fischer, M., Hamid, M.E., Michalke, S., Ludwig, W. and Mauch, H. (1998) Differentiation of phylogenetically related slowly growing mycobacteria based on 16S–23S rRNA gene internal transcribed spacer sequences. J. Clin. Microbiol. 36, 139–147.

[33] Domenech, P., Reed, M.B. and Barry Iii, C.E. (2005) Contribution of the Mycobacterium tuberculosis MmpL protein family to virulence and drug resistance. Infect. Immun. 73, 3492–3501.

[34] Zhang, F. and Xie, J.P. (2011) Mammalian cell entry gene family of Mycobacterium tuberculosis. Mol. Cell. Biochem. 352, 1–10.

[35] Akhter, Y., Ehebauer, M.T., Mukhopadhyay, S. and Hasnain, S.E. (2012) The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? Biochimie 94, 110–116.

[36] Saini, V., Raghuvanshi, S., Khurana, J.P., Ahmed, N., Hasnain, S.E., Tyagi, A.K. and Tyagi, A.K. (2012) Massive gene acquisitions in Mycobacterium indicus pranii provide a perspective on mycobacterial evolution. Nucleic Acids Res. 40, 10832–10850.