Mycobacterium avium subsp. hominissuis: 
The importance of genetic and metabolic diversity

Inaugural-Dissertation
zur Erlangung des Grades eines
PhD in Biomedical Sciences

an der
Freien Universität Berlin

vorgelegt von
Andrea Sanchini

Biologe aus Rom, Italien

Berlin 2017

Journal-Nr.: 3955
Gedruckt mit Genehmigung
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Dekan: Univ.-Prof. Dr. Jürgen Zentek
Erster Gutachter: Prof. Dr. Lothar H. Wieler
Zweiter Gutachter: PD Dr. Soroush Sharbati
Dritter Gutachter: Univ.-Prof. Dr. Thomas Alter

Deskriptoren (nach CAB-Thesaurus): Mycobacterium avium, genetic diversity, genetic variation, expressivity, genetic analysis

Tag der Promotion: 10.07.2017
Index

Figures.............................................................................................................................. 5
List of abbreviations.......................................................................................................... 6
1. Introduction .................................................................................................................. 9
   1.1 Introduction to Mycobacteria ................................................................................. 10
   1.2 Nontuberculous mycobacteria (NTM) ................................................................. 12
      1.2.1 Pathogenesis of NTM ...................................................................................... 12
      1.2.2 Metabolism of NTM ......................................................................................... 13
      1.2.3 Ecology of NTM ............................................................................................... 15
      1.2.4 Epidemiology of NTM ...................................................................................... 15
      1.2.5 Classification of NTM ...................................................................................... 16
   1.3 The Mycobacterium avium complex (MAC) ............................................................ 18
      1.3.1 Mycobacterium avium subsp. hominissuis (MAH) ............................................ 19
      1.3.2 The genetic diversity of MAH ............................................................................ 20
      1.3.3 Evolution of MAH ............................................................................................ 23
   1.4 Genomic Island (GI) ............................................................................................... 23
      1.4.1 Genomic Islands in MAH ................................................................................. 26
2. Aim and justification of the study .................................................................................. 28
3. Publications .................................................................................................................... 29
   3.1 First Publications ...................................................................................................... 29
   3.2 Second Publications ................................................................................................. 29
4. Discussion ....................................................................................................................... 67
   4.1 Implications of the hvGI for MAH ........................................................................... 67
   4.2 Implications of the metabolic studies on MAH ....................................................... 72
5. Conclusions and outlook ............................................................................................... 74
6. Summary ......................................................................................................................... 75
7. Zusammenfassung .......................................................................................................... 76
8. References ....................................................................................................................... 78
9. List of publications ......................................................................................................... 93
   9.1 Scientific articles ...................................................................................................... 93
   9.2 Poster presentations at conferences ......................................................................... 93
   9.3 Scientific presentations at conferences ................................................................. 94
Figures

Figure 1. **Cell wall of mycobacteria.** The plasma membrane and the peptidoglycan (PG) have the same composition as in other bacteria. However, the layer of PG is thinner in mycobacteria than in other Gram-positive bacteria. The PG is connected to arabinogalactan (AG), a heteropolysaccharide. The mycomembrane is the specific part of the mycobacterial cell wall and is composed of glycolipids (GL), phospholipids (PL) and several mycolic acids such as trehalose monomycolates (TMM) and trehalose dimycolates (TDM, also known as cordal factor). The structure of the long-chain fatty acid TDM is shown. The outer layer is composed by glucans, proteins and other lipids. Source (Marrakchi et al. 2014).

Figure 2. **Phylogenetic analysis of the most frequent NTM species based on the sequence of the 16S rRNA gene.** Adapted from (Tortoli 2014).

Figure 3. **Typical structure of a genomic island (GI).** Typically GIs are characterized by flanking tRNA genes. The presence of an own integrase gene (int) is responsible for the GI mobility. GIs are flanked by DNA direct repeats (DR). They are characterized by the presence of Insertion sequences (IS) and other genes that might be integrated during the GI evolution. Source (Dobrindt et al. 2004).
List of abbreviations

AIDS: Acquired immunodeficiency syndrome
AG: Arabinogalactan
CF: cystic fibrosis
CRISPR: Clustered regularly interspaced short palindromic repeats
DNA: Deoxyribonucleic acid
DR: Direct repeat
GI: Genomic Island
GL: Glycolipids
IS: Insertion sequence
ITS: Internal transcribed spacer
HAART: Highly active antiretroviral therapy
HIV: Human immunodeficiency virus
HGT: Horizontal gene transfer
hvGI: Hypervariable genomic island
LSP: Large sequence polymorphisms
MAA: *Mycobacterium avium* subsp. *avium*
MAC: *Mycobacterium avium* complex
MAH: *Mycobacterium avium* subsp. *hominissuis*
MAP: *Mycobacterium avium* subsp. *paratuberculosis*

MAS: *Mycobacterium avium* subsp. *silvaticum*

MCE: mammalian cell entry

MIC: Minimal inhibitory concentrations

MIRU-VNTR: Mycobacterial interspersed repetitive units variable number of tandem repeat

MmpL: Mycobacterial membrane protein Large

mRNA: Messenger ribonucleic acid

MTB: *Mycobacterium tuberculosis*

MTBC: *Mycobacterium tuberculosis* complex

NTM: Nontuberculous mycobacteria

ORF: Open reading frame

PCR: Polymerase chain reaction

PG: Peptidoglycan

PL: Phospholipids

PM: Phenotype microarray

qPCR: quantitative polymerase chain reaction

rDNA: Ribosomal deoxyribonucleic acid

RFLP: Restriction fragment length polymorphism

RGM: Rapid growing mycobacteria

RNA: Ribonucleic acid
rRNA: Ribosomal ribonucleic acid
SGM: Slow growing mycobacteria
SNP: Single nucleotide polymorphisms
TB: Tuberculosis
TDM: Trehalose dimycolates
TMM: Trehalose monomycolates
tRNA: Transfer ribonucleic acid
WGS: Whole genome sequencing
WHO: World Health Organization
1. Introduction

Substantial progresses have been made in the prevention and control of infectious diseases in the last decades. However, the World Health Organization (WHO) estimates that in 2010 about 15 millions of people still died because of infectious diseases (Dye 2014). In low-income countries there has been a decrease in the mortality and morbidity due to infectious diseases, but diseases such as diarrhoea, pneumonia, Human Immunodeficiency Virus (HIV), Tuberculosis (TB) and Malaria are still the leading causes of deaths (Lozano et al. 2012; Murray et al. 2014). Several recent environmental and behavioural changes affect the distribution of infectious diseases. Globalization increases the opportunities for pathogens to spread from one continent to another (2006; Gushulak & MacPherson 2004). Since the Industrial Revolution more people are living in urbanized areas in more dense conditions, allowing pathogens to better spread among the populations (Neiderud 2015). The climate changes led to the rise of temperatures and to changing rainfall distribution patterns, affecting vector-borne and water-borne infectious diseases (Hasnain et al. 2012; Shuman 2010). Another factor is the increase of the antibiotic-resistance in bacteria, which complicates the therapeutic options for the treatment of bacterial diseases (Fair & Tor 2014; Ventola 2015a; Ventola 2015b). For all these reasons, efforts and resources in fighting infectious diseases must be reinforced.

Knowledge on bacterial genetic characteristics and pathogenicity mechanisms is crucial to understand how bacteria evolved, how they spread in the population and what their pathogenic potential is. In the last years there has been an increase of information regarding genetic characteristics of bacteria due to the technology of Whole genome sequencing (WGS) (Land et al. 2015). Subsequently, comparative genomic allows to understand genetic variability and to identify the most adapted bacterial strains in epidemiological settings, such as the successful “Beijing clone” of Mycobacterium tuberculosis (MTB) (Ramazanzadeh & Sayhemiri 2014). We live now in a genomic era: the availability of deoxyribonucleic acid (DNA) sequence information of numerous bacterial species allows to discover the fundamental basis of pathogenicity and to discover new approaches to fight infectious diseases (Guttmacher & Collins 2003; Yang et al. 2008).
1.1 Introduction to Mycobacteria

The genus *Mycobacterium* belongs to the *Actinobacteria* phylum and to the *Mycobacteriaceae* family. These bacteria are classified as Gram-positive, they have a rod-shaped morphology, are non-motile and aerobic (Ventura et al. 2007). A special feature of mycobacteria is that they have a thick and hydrophobic cell wall compared to other bacteria, which is rich in long-chained fatty acids called mycolic acids (Fig. 1) (Kaneda et al. 1986; Marrakchi et al. 2014; Nataraj et al. 2015; Zuber et al. 2008). This thick cell wall makes mycobacteria acid-fast, meaning that during staining methods (e.g. Ziehl-Neelsen staining) they resist to acid de-colorization (Van Deun et al. 2008). Moreover, mycobacteria are usually difficult to treat because the thick cell wall makes them resistant to antibiotics (Jarlier & Nikaido 1994; Lambert 2002). Consequently the anti-mycobacterial therapy requires multiple antibiotics for a long time-period. This leads to high cost and high probability to develop antibiotic side effect and patients' in-adherence to treatment regimens (Kasperbauer & De Groote 2015; Philley & Griffith 2015). The hydrophobic cell wall contributes also to biofilm formation, adherence to surfaces, impermeability and slow growth of mycobacteria (Brennan & Nikaido 1995; Hoffmann et al. 2008). Another distinct feature of mycobacteria is their high GC content of their DNA (Levy-Frebault & Portaels 1992).
Figure 4. Cell wall of mycobacteria. The plasma membrane and the peptidoglycan (PG) have the same composition as in other bacteria. However, the layer of PG is thinner in mycobacteria than in other Gram-positive bacteria. The PG is connected to arabinogalactan (AG), a heteropolysaccharide. The mycomembrane is the specific part of the mycobacterial cell wall and is composed of glycolipids (GL), phospholipids (PL) and several mycolic acids such as trehalose monomycolates (TMM) and trehalose dimycolates (TDM, also known as cordal factor). The structure of the long-chain fatty acid TDM is shown. The outer layer is composed by glucans, proteins and other lipids. Source (Marrakchi et al. 2014).

There are three groups of mycobacteria: the *Mycobacterium tuberculosis* complex (MTBC), *Mycobacterium leprae* and the group of nontuberculous mycobacteria (NTM). Species belonging to the MTBC, including the MTB discovered by Robert Koch in 1882, cause TB in humans (Grange 1982); it is estimated that each year there are 9 million of new active TB cases and 1.5 millions of deaths (Lonnroth et al. 2015). In addition, it has been estimated that there are every year about 500.000 cases of infections caused by multi-drug resistant (MDR) TB cases, complicating the treatment (Pontali et al. 2013). *M. leprae* causes leprosy, a human disease characterized by the loss of the sensitization to pain and consequently the loss of anatomical parts of the body because of repeated injuries. In 2011 *M. leprae* caused about 220,000 leprosy cases in the world (2012; Lastoria & Abreu 2014). The third group of mycobacteria is represented by the NTM.
1.2 Nontuberculous mycobacteria (NTM)

The group of NTM is the largest and the more heterogeneous among the mycobacteria: more than 150 different species have been described (Tortoli 2014).

1.2.1 Pathogenesis of NTM

NTM are opportunistic pathogens able to cause infections in immune-compromised humans (Tortoli 2014). They are responsible for pulmonary infections, lymphadenitis, soft-tissue infections, bone-joint infections and disseminated infections. In rare cases, NTM cause infections also in immunocompetent humans (Breda et al. 2009; Fowler & Mahlen 2014; Piersimoni & Scarparo 2009).

Humans can acquire NTM infections in different ways. The first and presumably most common transmission route is via aerosol; humans inhale droplets containing NTM, the small droplets enter the alveoli and humans might develop pulmonary NTM disease (Falkinham 2009; Falkinham et al. 2008; Parker et al. 1983). NTM and especially *Mycobacterium abscessus* have been increasingly isolated from patients with cystic fibrosis (CF), where they damage the lung and represent a risk for lung transplantation (Esther et al. 2010; Orens et al. 2006).

Another route of infection is via direct contact with contaminated water, soil or dust that is common sources of NTM (see chapter 1.2.3). This way of transmission is particularly common in children, since often they bring hands and objects to their mouth while exploring the environment (Tortoli 2009). This transmission route might be responsible for the lymphadenitis, which indeed are common among young children (Hazra et al. 1999). Following to NTM infections, the lymph nodes become enlarged and inflammation processes occur (Perdikogianni & Galanakis 2014). The direct contact with NTM may also cause cutis and soft tissue infections of hands, knees or feet. The most common clinical manifestation in these infections is the formation of granulomas (Tortoli 2009). Regarding bone and joint infections, a NTM transmission typically occurs via traumatic wounds or after surgical intervention (Brown et al. 1999; Tortoli 2009). Disseminated infections can occur in patients
with strong immuno-deficiency such as HIV-infected patients (Inderlied et al. 1993; Tortoli 2009). Such patients might acquire the NTM infection via the gastro-intestinal tract or the respiratory tract. Previous studies showed that NTM affected approximately 50% of the Acquired Immune Deficiency Syndrome (AIDS) patients with a CD4 count less than 100 / µl (Inderlied et al. 1993). However, with the development of the highly active antiretroviral therapy (HAART) for AIDS patients, it is possible to keep the level of CD4 above 100 / µl and therefore the proportion of disseminated NTM diseases in these cases decreased (Karakousis et al. 2004).

1.2.2 Metabolism of NTM

Environmental NTM can survive in a wide range of environments, such as water, soil or dust. On the other hand clinical NTM are human intracellular pathogens, therefore they must adapt to life within human macrophages. Once internalized by human macrophages, NTM can inhibit the phagosome maturation allowing for bacterial survival. In all these conditions (environmental or clinical lifestyle) NTM survive and persist in various stress conditions such as lack or scarceness of nutrients (for example in water), anoxia (low oxygen level) or acid environment (low pH), which are characteristics of the environment in macrophages. Consequently, NTM must regulate gene expression and metabolism in order to persist and survive. They switch their metabolic activities depending on the stress conditions (Archuleta et al. 2005). In bacteria there are several strategies to survive in stress conditions: accumulation of storage compounds such as glycogen, formation of spores and metabolic dormancy. The latter is the strategies that NTM use to survive in nutrient-depletion conditions. Studies on Mycobacterium avium complex (MAC) showed that bacteria in starving conditions first undergo regulatory and physiological changes to adapt to the new conditions. Second, the MAC enters in a persistence phase, which is characterized by the shutdown of the general metabolism, which permits the survival in absence of nutrients (Archuleta et al. 2005). Characteristics of this metabolic shutdown are: usage of internal reserve of lipids, alterations of mycolic acids synthesis and therefore rearrangement of the cell wall, reduction in catalase and urease activities, reduction of the messenger ribonucleic acid
(mRNA) activity (only essential genes are transcribed), and antibiotic tolerance. The rearrangement of the cell wall is responsible for a reduced permeability of the cell wall, leading to a more protected state of the bacteria (Archuleta et al. 2005; Liu et al. 1996; Yuan et al. 1998). The metabolic shutdown during the persistence phase is an adaptation of NTM to stress conditions such as starvation, low pH or low oxygen. Other bacteria also show persistence mechanisms, such as species of the genus *Vibrio* or *Streptococcus* (Armada et al. 2003; Curras et al. 2002).

At genetic level, several genes are regulated to permit persistence of NTM in stress conditions: for example genes encoding for respiratory enzymes or for enzymes involved in the fatty acids metabolism (Honer zu Bentrup & Russell 2001). The genes encoding enzymes involved in fatty acids metabolism are of particular relevance. It has been shown that members of NTM possess higher numbers and multiple copies of genes encoding enzymes involved in metabolism of lipids and fatty acids, when compared to other mycobacteria such as MTB or *M. leprae*. This redundancy suggests that NTM are capable to generate a greater diversity of lipids and fatty acids for the cell wall compared to other mycobacteria (Li et al. 2005a; Marri et al. 2006; Trivedi et al. 2004). The cell wall plays a key role in the host-pathogen interaction, since it constitutes the interface and the first point of contact. Therefore, variations in the composition of proteins but especially of lipids in the cell wall might reflect differences in adaptation to survival in the environment or in humans (Marri et al. 2006).

More information is available on the utilization of metabolic substrates by MTB than by NTM. Apart from lipids and fatty acids that are used as major carbon sources, mycobacteria metabolize amino acids as source of nitrogen, such as asparagine, glutamate and aspartate (Cook et al. 2009). They can also assimilate ammonium, if no other nitrogen source is present. In order to decide which nutrient to take, NTM must have a sensor system able to detect nutrients, as shown for MTB (Cook et al. 2009). The major source of phosphate for mycobacteria is inorganic phosphate: in MTB the *pstS1* and *pstS2* genes regulating phosphate uptake were associated with virulence: indeed, in a model of mouse infection, mutants with deleted *pstS1* and *pstS2* genes showed reduced virulence when compared with the wild type.
strains (Peirs et al. 2005). The sulphate assimilation depends also on a regulatory system able to detect lack of sulphate resulting in the activation of sulphate uptake (Cook et al. 2009).

1.2.3 Ecology of NTM
NTM are widely distributed in the environment, they can form biofilm and adhere in different types of surfaces (Faria et al. 2015; Sousa et al. 2015). So far NTM have been isolated from dust, soil, plants and natural water sources such as streams or rivers. In addition, NTM have been isolated also from drinking water distribution systems such as household plumbing or aerosols and hospital water systems (Dailloux et al. 2003; Falkinham 2009; Falkinham et al. 2001; Lahiri et al. 2014a; Le Dantec et al. 2002; Whittington et al. 2004). The colonization of drinking-water distribution systems brings NTM very easily in contact with humans (Tortoli 2014). Why can NTM survive in many different environmental conditions? The following properties make this possible: the hydrophobicity and impermeability of their cell wall, their tolerance to an acid environment and to high temperature and their oligotrophy, allowing them to survive in an environment with low level of nutrients, where most of other bacteria cannot survive (Norton et al. 2004).

1.2.4 Epidemiology of NTM
In the recent years the incidences of infections caused by NTM are increasing in many areas of the world such as Germany (Ringshausen et al. 2016), US (Adjemian et al. 2012), Japan (Morimoto et al. 2014), Ontario (Canada) (Marras et al. 2013), UK (Moore et al. 2010), the Netherlands (van Ingen et al. 2009), Italy (Rindi & Garzelli 2016) and Queensland (Australia) (Thomson et al. 2010). For example in Germany the prevalence of pulmonary NTM infections increased from 2.3 to 3.3 cases/100,000 population during the period 2009-2014. This increase might be due partly to the increased awareness regarding NTM, and partly to the improvement of NTM isolation and identification techniques. In addition, it needs to be considered that in countries with high TB incidences, NTM infections are often neglected (Gopinath & Singh 2010). However, also in countries where the NTM infections are a
reportable disease, an increase in the incidences of NTM-infections has been observed, as for example in Oregon State in the US and in the Queensland state in Australia (Adjemian et al. 2012; Thomson et al. 2010). The epidemiology of NTM became even more interesting because of two recent NTM outbreaks occurring in humans. The first report is from Bryant and colleagues: they analyzed by WGS all *M. abscessus* isolates from patients with CF attending the same clinic between the years 2007 and the 2011 in the UK. They identified two prolonged outbreaks caused by an almost identical *M. abscessus* isolate in 11 patients (Bryant et al. 2013). This study suggested a possible human-to-human transmission of NTM species between CF patients, while in previous studies the issue of person-to-person transmission was never proved (Harris et al. 2012; Huang et al. 2010; Olivier et al. 2003; Sermet-Gaudelus et al. 2003).

The second report of an outbreak caused by NTM is from Haller and colleagues. They reported five cases of *Mycobacterium chimaera* disseminated infections in five patients previously undergoing open chest surgery in Germany between April 2015 and February 2016. The epidemiological, environmental and molecular investigations suggested that *M. chimaera* infections have been caused by the exposure of patients to a contaminated heater-cooler unit (Haller et al. 2016).

The assumed increase of NTM infections, the capacity to cause outbreaks, the variety of clinical pictures, the potential and unknown nature of human-to-human transmission and their spread in the environment make NTM an emerging public health problem that need to be further investigated.

### 1.2.5 Classification of NTM

The NTM can be divided in two further groups based on their rapidity to grow: rapid growing mycobacteria (RGM) form visible colonies in agar plates within 3-7 days, while slow growing mycobacteria (SGM) require more than 7 days to form visible colonies in agar plates (Kim et al. 2013). Genetically those two groups are clearly distinguished, based on the phylogenetic analysis of their 16S ribosomal RNA (rRNA) gene sequences (Fig. 2).
Among the RGM the species most frequently responsible for human infection are *Mycobacterium chelonae*, *Mycobacterium fortuitum* and *M. abscessus*. While among the SGM the species most frequently responsible for human infection are *Mycobacterium xenopi*, *Mycobacterium kansasii*, *Mycobacterium gordonae* and the species belonging to the MAC. The MAC is the clinically most relevant NTM for humans (Hoefsloot et al. 2013; Tortoli 2014).
1.3 The *Mycobacterium avium* complex (MAC)

Until the beginning of the 1990s only two major species were included in the MAC; *M. avium* and *Mycobacterium intracellulare* (Inderlied et al. 1993; Wayne et al. 1991). With the availability of more advanced molecular and phenotypic tests it became clear that within the MAC there was more diversity; additional species and sub-species were identified, with different pathogenicity and host specificity. Today the MAC includes 11 species: *M. chimaera*, *Mycobacterium indicus pranii*, *M. intracellulare*, *Mycobacterium arosiense*, *M. avium*, *Mycobacterium vulneris*, *Mycobacterium bouchedurhonense*, *Mycobacterium colombiense*, *Mycobacterium marseillense*, *Mycobacterium yongonense* and *Mycobacterium timonense*. Within the MAC, *M. avium* is the most frequently isolated from human or animal infections (Rindi & Garzelli 2014; Tortoli 2014). *M. avium* is further divided in four subspecies that show differences in lifestyle, ranging from environmental bacteria that are opportunistic pathogen for humans to animal pathogens (Rue-Albrecht et al. 2014). *Mycobacterium avium* subsp. *paratuberculosis* (MAP), is responsible for paratuberculosis (also known as Johne’s Diseases) mostly in ruminants (Windsor 2015). However, MAP infections have been reported also in animals in closer contact to humans such as pigs, dogs or cats and also in other primates such as macaques or gibbons (Singh et al. 2011; Stevenson et al. 2009). The infections are chronic and characterized by the presence of a granuloma in the intestine of the animal and the outcome of those infections may be fatal (Naser et al. 2014). MAP is one of the slowest growing mycobacteria: it takes up to one month or more to see colonies in an agar plate. This slowness makes the diagnosis and etiological confirmation of MAP disease challenging (Turenne et al. 2007; Whittington et al. 1999). MAP-infected animals shed the bacteria in the faeces, and other animals acquire the infection via MAP-contaminated water or milk (Koets et al. 2015; Rindi & Garzelli 2014). Regarding infections in humans, it has been suggested that MAP is the etiological agent of the Crohn’s disease, a chronic disease of the bowel. However, this issue is controversially discussed (Davis 2015; Sechi & Dow 2015). The second subspecies is *Mycobacterium avium* subsp. *avium* (MAA). MAA is responsible for TB in birds (Turenne et al. 2007). Occasionally MAA has been isolated from other animals such as pigs or cattle (Rindi & Garzelli 2014). The transmission of MAA from infected birds to other animals occurs via the faecal-oral route (Schrenzel et al.
Reports of MAA infections in humans are rare (Thegerstrom et al. 2005). The third subspecies is *Mycobacterium avium* subsp. *silvaticum* (MAS). MAS is known to cause TB-like lesions in wood pigeons (Thorel et al. 1990). Rarely MAS has been isolated in other animals such as horses (Chiers et al. 2012). The fourth subspecies is the *Mycobacterium avium* subsp. *hominissuis* (MAH).

1.3.1 *Mycobacterium avium* subsp. *hominissuis* (MAH)

MAH is the clinically most important MAC and one of the clinically most important species among all the NTM (Hoefsloot et al. 2013; Kolb et al. 2014; Tran & Han 2014). In humans MAH isolates are responsible for lymphadenitis in children, pulmonary infections and disseminated infections in adults with immune-suppression. With the increase of the number of immune-compromised humans due to ageing of the population and better treatment of patients with underlying risk factors, the public health relevance of MAH infections increased (Falkinham 2013; Hoefsloot et al. 2013). Although rarely, MAH have been isolated from domestic animals such cats, dogs, deer, pigs, or horses (Rindi & Garzelli 2014). Along with its pathogenic lifestyle, MAH has also been isolated from different environmental sources: water, soil or dust (Falkinham 2013; Kolb et al. 2014; Lahiri et al. 2014a; Rindi & Garzelli 2014). The spread of MAH in the environment suggests that humans can acquire the infection by environmental exposure, especially through contaminated water or soil (Alvarez et al. 2008; De Groote et al. 2006; Falkinham 2010).

MAH is a relatively new proposed subspecies of *M. avium*: in 2002 Mijs and co-authors proposed the MAH nomenclature in order to differentiate the *M. avium* infecting birds (later re-defined as MAA and MAS) from the *M. avium* infecting humans (MAH) (Mijs et al. 2002). Along with this host difference, MAH is differentiated from the other subspecies also by its capacity to grow at temperatures from 24 to 45°C. With the use of molecular genotyping methods, such as restriction fragment length polymorphism (RFLP) of multiple insertion sequences (IS) and genome sequencing, it became clear that MAH has distinct genetic features when compared to the other *M. avium* subspecies. In particular, MAH can be
identified and differentiated from the other *M. avium* subspecies by the presence of multiple copies of the IS1245, by the sequence of the 16S-23S rDNA Internal transcribed spacer (ITS) and by the absence of IS901 which is on the contrary found in the subspecies MAA and MAS. The IS900 is a specific marker of the subspecies MAP. In addition, the amplification of the IS1311 is a specific marker of all the *M. avium* subspecies, and can distinguish them from the *M. intracellulare* (Rindi & Garzelli 2014).

### 1.3.2 The genetic diversity of MAH

The genomic variability and plasticity is an expression of the bacterial adaptation to different environmental conditions: gene duplications, gene loss, horizontal gene transfer (HGT) and gene regulation are all manifestations of adaptation (Vicente & Mingorance 2008). With the advance in molecular genotyping methods, the field of comparative genomics greatly expanded. This expansion allowed researchers to better understand the epidemiology and distribution of a specific bacterial strain, biotype, serotype etc. In addition, the recent advances in molecular methods permitted to better discriminate between species and strains determining an improvement in diagnostic accuracy and identification procedures (Franca et al. 2013).

Regarding *M. avium*, specific genetic differences have been identified between isolates apparently similar but occupying difference niches or different hosts. In particular, the MAH has been found to be more genetically diverse than the other *M. avium* subspecies (Ignatov et al. 2012; Rindi & Garzelli 2014; Semret et al. 2004; Turenne et al. 2008; Turenne et al. 2007; Wu et al. 2006). The sources of genetic variability of MAH are multiple: single nucleotide polymorphisms (SNP), large sequence polymorphisms (LSP), minisatellites sequences and presence of phages, plasmids and other mobile genetic elements. SNPs are a very common source of genetic diversity in *M. avium*. Turenne and co-authors in 2008 performed multi-locus sequence analysis of 10 housekeeping genes corresponding to 8064 bp with 56 strains of *M. avium* (MAH n= 24, MAP n= 21, MAA n=8 and MAS n=3). They found a total of 205 SNPs, and the great majority of these SNPs has been identified in the MAH subspecies.
In addition, recombination events among the different alleles of the housekeeping genes also occurred, in contrast to the other *M. avium* subspecies analyzed that showed no evidence of allele recombination.

Another interesting gene that has been subjected to SNP studies with the purpose of differentiating the *M. avium* subspecies is the *hsp65* housekeeping gene. It encodes for the heat shock protein of 65 kDa (Slany & Pavlik 2012). In 2006 Turenne and co-authors showed that a variable region of the *hsp65* gene located at the 3’ extremity has enough discriminatory power to differentiate between all the *M. avium* subspecies and also *M. intracellulare* included in the MAC. They obtained distinct *hsp65* SNP profiles corresponding to each subspecies, and once again the MAH subspecies was the one with the highest number of different *hsp65* profiles identified (Turenne et al. 2006).

Another source of genetic diversity that has been investigated in *M. avium* isolates are LSP. Differently to the SNPs that represent polymorphisms in a single nucleic acid position, the LSPs are larger regions of sequence polymorphisms that are due to events of insertion or deletion of genetic material or horizontal gene transfer (Alland et al. 2007; Brosch et al. 2001). The identification of LSPs became possible when WGS data were available. The first *M. avium* that was fully sequenced was the *M. avium* 104 (later re-defined as belonging to the MAH subspecies) in the 1980s. This strain was isolated from the blood of an AIDS patient with disseminated infection in the United States (Horan et al. 2006). Its genome size is 5.48 MB and it encodes for 5313 Open reading frames (ORF)s. Semret and co-authors in 2004 published the first comparative genomic studies with several *M. avium* subspecies, by using whole genome DNA microarray (then confirmed by Polymerase chain reaction (PCR) and sequencing) with the strains MAH 104, MAP K10, MAP Ln20 and MAS 49884. The comparison of DNA microarrays revealed that 14 LSPs were present in the MAH 104, and not in the other subspecies. The range of these 14 LSPs was from 21 to 197 kb, including altogether 527 genes for a total of 727 kb. The genes involved in metabolism represented almost 40% of all genes contained in the LSPs. These 14 LSPs altogether represent 13,5% of the WGS of MAH 104 (Semret et al. 2004). Therefore, within the species *M. avium*, MAH showed the highest degree of genetic diversity. Another more recent study used a similar
approach but has a more sensitive definition of LSP compared to the first study; the authors considered at least three consecutive ORFs to be present in a LSP. A total of 24 LSPs were identified in MAH isolates that were absent in the other \textit{M. avium} subspecies. Within the 24 LSPs also the previously identified 14 LSPs were included. The size range of these 24 LSPs was from 3 to 196 kb for a total of 846 kb, accounting for 17\% of the WGS of the MAH 104 (Wu et al. 2006).

Another genetic element characteristic of mycobacteria and specifically of the \textit{M. avium} genomes is the presence of repetitive sequences. They are called minisatellites, and are tandem repeats of a length of 10-100 nucleotides spread in several parts of the mycobacterial genome. The number and the variability of these repetitive sequences are used for molecular typing purposes by the Mycobacterial interspersed repetitive units variable number of tandem repeat (MIRU-VNTR) approach (Mears et al. 2015). The MIRU-VNTR typing method has high discriminatory power and allows for epidemiological and comparative genomics studies within MAH isolates from different sources (Inagaki et al. 2009; Radomski et al. 2010). Several studies using the MIRU-VNTR typing methods on MAH isolates revealed that human and pig MAH isolates shared similar MIRU-VNTR profiles, highlighting a possible source of transmission for human infection or common environmental source of infection (Iwamoto et al. 2012; Rindi et al. 2013; Tirkkonen et al. 2010).

In 2013 another MAH strain has been fully sequenced, and thus serves as second reference strain. The MAH TH135 strain has been isolated from the sputum of an AIDS-negative patient with pulmonary infection in Japan. Comparative genomic analysis with the other reference strain MAH 104 revealed that the genome of the TH135 strain is smaller (4.96 Mb) than the genome from MAH 104 (5.48 Mb) and encodes for 4636 ORFs. A total of 4012 genes are shared by the two reference strains, and 624 and 1108 genes are unique for the TH135 and for the MAH 104 strain, respectively. This highlighted a certain degree of diversity between the two strains, which have been isolated from different diseases and countries as well. This diversity suggests that pulmonary (MAH TH135) and disseminated (MAH 104) strains might have acquired different genes during evolution, resulting in different pathogenesis characteristics. The GC content of the two strains is almost the same.
(69.3 % for the MAH TH135 and 69.0 % for the MAH 104 strain) and they have the same numbers of transfer RNA (tRNA) genes (n = 46). Interestingly, the MAH 104 strain possesses four time more IS elements than the MAH TH135 (129 vs. 30), suggesting that the genome of MAH 104 underwent more genetic re-arrangements and recombination events (Uchiya et al. 2013). The availability of more WGS of MAH isolates from different sources and countries might clarify the genetic relations among isolates and the possible routes of infections for humans.

1.3.3 Evolution of MAH
By analyzing the results of the above mentioned comparative genomic studies, it was possible to hypothesize the evolutionary history of MAH within the species M. avium. Especially the studies focusing on SNPs and LSP permitted to understand the evolution of MAH. The higher SNPs variation and recombination events in MAH permitted to this subspecies access to a wider range of environmental conditions (Rindi & Garzelli 2014). Two major lineages evolved from the MAH, with different and more restricted host-specificity compared to the MAH: one lineage corresponding to the MAP subspecies and the other lineage including both MAA and MAS subspecies. A phylogenetic analysis of the SNPs of ten housekeeping genes by Turenne and co-authors revealed that the rates of nonsynonymous (dN) to synonymous (dS) substitutions was higher for MAP (0.67) and for MAS and MAA (0.50) when compared to the dN / dS ratio of MAH (0.08). A higher dN/dS ratio indicates positive selection of nucleotide substitutions from a common ancestor with a lower dN/dS ratio. This suggests that the MAP, MAA and MAS evolved more recently from a common MAH ancestor (Jeffares et al. 2015; Mugal et al. 2014; Turenne et al. 2008).

1.4 Genomic Island (GI)
The increasing number of available bacterial genome sequences expanded our understanding on genome and bacterial evolution. Research on genetics revealed that genetic events such as gene acquisition via HGT, gene loss and gene mutations contribute to pathogenicity, host-
specificity and adaptation of bacteria to different conditions (Doolittle 2002; Treangen & Rocha 2011). The acquisition of new genetic elements via HGT, in forms of plasmids, phages or Genomic Islands (GI)s, is a mechanisms found especially in extracellular bacteria: they use the new information acquired to adapt to a new lifestyle. The part of the genome acquired by HGT is called “accessory genome”, because it is distinct from the “core genome”, which represents the part of the genome essential for bacterial survival. On the contrary the loss of genes is a typical feature of intracellular pathogens, because in this lifestyle they use the cellular resources of the host to compensate the loss of genetic material. Mutations are responsible for the development of new gene variants, which may confer evolutionary advantages (Dobrindt et al. 2004).

Pathogenicity islands are mobile genetic elements that carry virulence genes and thus confer pathogenic potential to the recipient. For example, a pathogenicity island might carry genes encoding for bacterial toxins, for adherence factors facilitating host-cell invasion, for capsule biosynthesis which avoid phagocytosis or for siderophores allowing for a better up-take of ions necessary for the bacteria during infection (Boyd & Brussow 2002; Shankar et al. 2002).

GIIs are also part of the accessory genome of a microorganism. They are mobile genetic elements (ranging from 3 to more than 200 kb) integrated in the chromosome that have been acquired via HGT (Sanchini et al. 2016). A mechanism has been proposed to explain the GI evolution. A GI might evolve from other mobile genetic elements, such as plasmids, that integrated in the chromosome by recombination. Subsequently these genetic elements lost genes essential for their mobility and for their capacity for autonomous replication and therefore became GIs integrated in the chromosome. These GIs continue to evolve by the acquisition of additional DNA elements such as transposons or IS. Each functional GI possesses an own integrase gene (int) which is responsible for the GI excision from the chromosome, thus allowing the transfer of the GI to other bacteria via HGT and re-integration in another chromosome (Ahmed et al. 2008; Dobrindt et al. 2004). Structurally, GIs are characterized by a different GC content when compared to the GC content of the whole genome, indicating that they derive from another microorganism (Blanc-Potard & Lafay 2003). In addition, GIs are associated with flanking tRNA genes, and usually are flanked by
DNA direct repeats (DR), indicating insertional events (Bellanger et al. 2014). GIs often contain other mobility genes responsible for the mobility of the GI (Fig. 3) (Che et al. 2014; Langille et al. 2010).

Figure 6. Typical structure of a genomic island (GI). Typically GIs are characterized by flanking tRNA genes. The presence of an own integrase gene (int) is responsible for the GI mobility. GIs are flanked by DNA direct repeats (DR). They are characterized by the presence of Insertion sequences (IS) and other genes that might be integrated during the GI evolution. Source (Dobrindt et al. 2004).

GIs might influence different aspects of the bacterial lifestyle, depending on the specific conditions; they might influence antibiotic-resistance, adaptation to the environment, the fitness of the bacteria, or have a role in a specific metabolic pathway. Therefore, GIs play a pivotal role in bacterial evolution (Ahmed et al. 2008; Che et al. 2014; Juhas et al. 2009). Here we report few examples of GIs that are crucial for bacterial evolution. In *Escherichia coli* the enterocyte effacement GI permitted the evolution from a commensal clone to a specific enterohaemorrhagic *E. coli* pathotype (Franzin & Sircili 2015). Moreover, GIs can shuffle resistance genes between different species or between clinical and environmental settings, leading to the formation of multi-antibiotic-resistance hospital microorganisms (Fricke et al. 2008; Juhas et al. 2009). For example, the “SCCmec” GI in *Staphylococcus aureus* is responsible for the resistance against methicillin, a crucial antibiotic against *S. aureus* (Ito et al. 2014). Another example is the SGI1 that carries multiple antibiotic-resistance genes in *Salmonella enterica* and *Proteus mirabilis* (Lei et al. 2014). In *Pseudomonas aeruginosa*, the GI1 and GI2 are responsible for the multi-antibiotic resistance phenotype (Roy Chowdhury et al. 2016).
1.4.1 Genomic Islands in MAH

In MAH, researchers described a 3.5 kb GI which is responsible for the mycobacterial invasion of macrophages and amoeba in the reference strain MAH 104. Indeed, a mutant lacking this GI lost the ability to invade human macrophages and *Acanthamoeba castellanii* cells (Danelishvili et al. 2007).

In 2014, Lahiri and co-authors compared different WGS of MAC in order to identify the accessory genome of these species. Precisely the WGS of the MAH 104, the MAP K10, the MAA ATCC 25291 and the *M. intracellulare* ATCC 13950 strains have been analyzed. The comparison of these WGS revealed the presence of seven genome regions (sizes 22.85-199.29 kb) that were specific for the MAH 104, and not found in the other species or subspecies (Lahiri et al. 2014b). Three out of the seven regions specific of the MAH 104 were flanked by tRNAs which is one indication of plausible presence of a GI (Che et al. 2014; Langille et al. 2010). Two of these 3 regions did not contain any DNA direct repeats whereas the region 3 contained two DNA direct repeats at the extremities. In addition, region 3 had a drop in the GC content when compared to the GC content of the rest of the MAH 104 genome (66.8 % vs. 69.0 %). The region 3 contained 63 genes, among which genes responsible for mobility such as transposons, integrases and phagic genes indicating HGT. Therefore, the region 3 identified in MAH 104 was classified as a new GI. This GI in the MAH 104 strain is integrated between the methyltransferase (MAV_0778) and the carveol dehydrogenase genes (MAV_0846). The GI shares 53% homology with the genome of *M. xenopi* RIVM700367.

We later searched for this GI in four other MAH genomes that were available at that time in the Genbank (strain MAH TH135, MAH 10-4249 isolated from a deer in the United States and two MAH isolated from Germany, the MAH 27-1 from dust and the MAH 2721 isolated from a child with lymphadenitis). We found that this GI was missing from the MAH TH135 strain, but it was present in the other three strains, with different size (16.37 to 84.85 kb), different gene composition and integrated always in the same insertion site in the chromosome (Lahiri et al. 2014b). Therefore we named such GI as hypervariable GI (hvGI). Such heterogeneity was classified as further element contributing to the genetic diversity of MAH. In addition, the hvGI represents a genetic marker of MAH isolates, since it was not
found in other mycobacteria. That motivated us to characterize the hvGI in other MAH isolates from different sources and to determine the degree of variability of this hvGI.
2. Aim and justification of the study

NTM are an emerging public health concern. Their wide distribution in the environment favours their contact with humans. In addition, their persistence and resistance to antibiotics makes the treatment of NTM infections challenging. Among the NTM, *M. avium* and specifically the subspecies MAH is the clinically most relevant and the most frequently isolated NTM from clinical samples. Its high genetic diversity seems to confer the ability to MAH to adapt to a wide range of environmental conditions but also to live as pathogen in human cells. Moreover, the contribution of metabolic properties to adaptation of MAH to a clinical or environmental lifestyle is not known.

In this study we investigated the genetic and the metabolic diversity of MAH isolates.

We wanted to analyze whether the new recently identified hvGI can differentiate clinical and environmental MAH isolates. Specifically we wanted to know whether the hvGI is a genetic marker for clinical or for environmental isolates. In addition, we started the investigation of the function of one of the genes contained in the hvGI in order to get first insights about functional roles of genes present in the hvGI. The pathogenicity of bacteria may originate from genetic features and plasticity, and also from metabolic adaptation to the host and to different conditions. Therefore, we addressed another question: are clinical and environmental MAH isolates different in terms of metabolic features? Moreover, having the WGS data of the MAH isolates we wanted to compare to which degree genetic and metabolic relations correlate with each other.
3. Publications

3.1 First Publications


Received: 15 April 2016

Revised: 06 June 2016

Accepted: 17 July 2016

You have to purchase this part online.

3.2 Second Publications


Received: 05 September 2016

Revised: 13 October 2016

Accepted: 27 November 2016
Metabolic phenotype of clinical and environmental *Mycobacterium avium* subsp. *hominissuis* isolates

Andrea Sanchini 1, Flavia Dematheis 2, Torsten Semmler 3, Astrid Lewin Correspond. 1

1 Division 16, Mycotic and Parasitic Agents and Mycobacteria, Robert Koch Institute, Berlin, Germany
2 Institute of Microbiology and Epizootics, Free University Berlin, Berlin, Germany
3 NG 1 Microbial Genomics, Robert Koch Institute, Berlin, Germany

Corresponding Author: Astrid Lewin
Email address: LewinA@rki.de

**Background.** *Mycobacterium avium* subsp. *hominissuis* (MAH) is an emerging opportunistic human pathogen. It can cause pulmonary infections, lymphadenitis and disseminated infections in immunocompromised patients. In addition, MAH is widespread in the environment, since it has been isolated from water, soil or dust. In the recent years, knowledge on MAH at molecular level substantially increased. On the contrary, knowledge of the MAH metabolic phenotypes remains limited.

**Methods.** In this study for the first time we analyzed the metabolic substrate utilization of ten MAH isolates, five from clinical and five from environmental source. We used the BIOLOG Phenotype Microarray™ technology. This technology permits the rapid and global analysis of metabolic phenotypes.

**Results.** The ten MAH isolates tested showed different metabolic patterns pointing to high intra-species diversity. Our MAH isolates preferred to use fatty acids such as Tween, caproic, butyric and propionic acid as a carbon source, and L-cysteine as a nitrogen source. Environmental MAH isolates resulted to be more metabolically active than clinical isolates, since the former metabolized more strongly butyric acid ($p = 0.0209$) and propionic acid ($p = 0.00307$).

**Discussion.** Our study provides new insight into the metabolism of MAH. Understanding how bacteria utilize substrates during infection might help the developing of strategies to fight such infections.
1 Article Title

2 Metabolic phenotype of clinical and environmental *Mycobacterium avium* subsp. *hominissuis*

3 isolates

4

5 Authors

6 Andrea Sanchini¹, Flavia Dematheis², Torsten Semmler³, Astrid Lewin¹."  

7

8 Author affiliations

9 ¹ Division 16, Mycotic and Parasitic Agents and Mycobacteria, Robert Koch Institute, Berlin, Germany

10 ² Institute of Microbiology and Epizootics, Free University Berlin, Berlin, Germany

11 ³ NG 1 Microbial Genomics, Robert Koch Institute, Berlin, Germany

13

14 Corresponding author

15 Dr Astrid Lewin

16 Division 16, Mycotic and Parasitic Agents and Mycobacteria, Robert Koch Institute, Seestraße

17 10, 13353, Berlin, Germany.

18 Telephone: +4930187542112
Abstract

Background. Mycobacterium avium subsp. hominissuis (MAH) is an emerging opportunistic human pathogen. It can cause pulmonary infections, lymphadenitis and disseminated infections in immuno-compromised patients. In addition, MAH is widespread in the environment, since it has been isolated from water, soil or dust. In the recent years, knowledge on MAH at molecular level substantially increased. On the contrary, knowledge of the MAH metabolic phenotypes remains limited.

Methods. In this study for the first time we analyzed the metabolic substrate utilization of ten MAH isolates, five from clinical and five from environmental source. We used the BIOLOG Phenotype Microarray™ technology. This technology permits the rapid and global analysis of metabolic phenotypes.

Results. The ten MAH isolates tested showed different metabolic patterns pointing to high intra-species diversity. Our MAH isolates preferred to use fatty acids such as Tween, caproic, butyric and propionic acid as a carbon source, and L-cysteine as a nitrogen source. Environmental MAH isolates resulted to be more metabolically active than clinical isolates, since the former metabolized more strongly butyric acid ($p = 0.0209$) and propionic acid ($p = 0.00307$).

Discussion. Our study provides new insight into the metabolism of MAH. Understanding how bacteria utilize substrates during infection might help the developing of strategies to fight such infections.
42 Introduction

43 *Mycobacterium avium* subsp. *hominissuis* (MAH) is clinically one of the most relevant non-
tuberculous mycobacteria (Tortoli 2014). MAH is an opportunistic human pathogen causing
pulmonary infections, lymphadenitis in small children and disseminated infections (Despierrres et
al. 2012; Rindi & Garzelli 2014). It is of increasing public health relevance, with reports of
MAH infections increasing worldwide (Hoefsloot et al. 2013). Moreover, MAH is widespread in
the environment (Falkingham 2013; Lahiri et al. 2014). In recent years, there have been
substantial advances in the analysis of bacteria at the molecular level. Indeed, several whole
genome sequences are now available for many mycobacterial species, including MAH
(Bannantine et al. 2014; Kim et al. 2012; Uchiya et al. 2013; Wynne et al. 2010). In contrast,
there has been little concomitant advance in knowledge at the phenotypic level. Phenotype
analysis deserves greater attention, as it is the phenotype that selection pressure acts upon to
confer evolutionary advantages to the bacterial species (Plata et al. 2015). In order to address this
knowledge gap for bacterial phenotypes, BIOLOG Inc. developed the Phenotype MicroArray™
(PM) (BIOLOG, Hayward CA), a high throughput method for the rapid and global analysis of
microbial metabolic phenotypes (Bochner 2003; Bochner 2009; Bochner et al. 2001; Bochner et
al. 2008). The PM technology consists of several commercially available 96-well plates in which
every well has a different substrate, allowing nearly 2000 different microbial metabolic
phenotypes to be tested (Bochner 2003; Bochner 2009; Bochner et al. 2001; Bochner et al.
2008). PM technology has been applied to several microorganisms, including mycobacteria
(Baloni et al. 2014; Bochner et al. 2008; Borglin et al. 2012; Chen et al. 2012; Gupta et al. 2015;
application of PM is the detection of phenotype changes due to gene knock-out. For example, Chen and co-authors showed that a *leuD* mutant of *M. avium* subsp *paratuberculosis* lost the ability to use several carbon, nitrogen, sulfur and phosphorous substrates (Chen et al. 2012). Other researchers showed that the use of 12 carbon substrates differentiated *M. tuberculosis* from *M. bovis* (Khatri et al. 2013; Lofthouse et al. 2013). In this study we tested clinical and environmental isolates of MAH using the PM technology. Our aim was to describe the metabolic substrates utilized by MAH isolates and to identify any metabolic differences between clinical and environmental MAH isolates.
Materials and Methods

Bacterial isolates and BIOLOG phenotype microarray

We analyzed five clinical and five environmental MAH isolates (Table 1).

We performed the BIOLOG Phenotype Microarray™ (BIOLOG, Hayward, CA) according to the manufacturer’s recommendations. The technology is based on the measurement of bacterial respiration, which produces NADH (Bochner et al. 2001). If bacteria are able to metabolize a specific substrate, electrons from NADH reduce a tetrazolium dye in an irreversible reaction generating a purple color in the PM plate wells. This color change is measured and recorded every 15 minutes by the reporter instrument OmniLog™ (BIOLOG, Hayward, CA), generating a kinetic response curve for each well (Bochner 2003; Bochner 2009).

The ten MAH isolates were tested with the 96-wells plates PM1 to PM4, containing 190 carbon (PM1 and PM2), 95 nitrogen (PM3), 59 phosphorous (PM4) and 35 sulfur (PM4) substrates. The PM plates 1, 2 and 3 include one negative control well, in which bacteria are tested without any substrate. The PM4 plate includes two negative control wells, one for the phosphorus and one for the sulfur substrates. All isolates were tested three times. Briefly, we cultivated each MAH isolate in 30 ml of 7H10 Middlebrook medium supplemented with 10% modified ADC-enrichment (2% of glucose, 5% of BSA, 0.85% of NaCl) until an OD₆₀₀ nm of 0.3-0.6 was achieved (mid-logarithmic phase of growth). The use of liquid cultures in place of agar reduces bacterial clumping. Bacterial cultures were harvested by centrifugation for 10 minutes at 4000g and pellets were re-suspended in 10 ml of distilled water. Bacterial cells were starved for one night in water at room temperature to minimize false positive reactions due to nutrient accumulation in MAH cells and to ensure the use of the substrates provided by the PM plates.
The following day the cells were centrifuged and re-suspended using a sterile stick in tubes containing 10 ml of GN/GP-IF-0a (BIOLOG inoculating fluid), 120 μl of 100× BIOLOG Redox Dye Mix G and 1 ml of the appropriate additive (Table 2), until 85% transmittance was reached as measured using the turbidimeter provided by BIOLOG. In order to reduce bacterial clumping, the sterile stick used for inoculation was ground against the wall of the tube. A volume of 100 μl of this final suspension was added to each of the 96 wells of the PM plates. The PM plates were then sealed to avoid drying and incubated at 37°C in the OmniLog® (BIOLOG, Hayward, CA) incubator reader for 8 days.

As recommended by BIOLOG, we tested plates PM1 to PM4 using the same assay protocol but without addition of bacteria in order to identify wells with abiotic dye reduction, which can generate false positive results.

**Analysis of BIOLOG phenotype microarray data**

The raw kinetic data were exported as CSV files using OmniLog PM file Management/kinetic Analysis module (Bochner 2003; Khatri et al. 2013). Differences in the metabolism of the different substrates by the ten MAH isolates were investigated by analyzing the maximum height of the bacterial respiration curves (parameter A) using the R-package opm (Vaas et al. 2013). To allow comparisons across plates processed in different experimental runs, the A parameters were normalized by subtracting the well mean of the negative control (Vaas et al. 2013). Furthermore, the A parameters of the triplicates were combined by calculating the mean and discretized into “positive”, “moderate” and “negative” metabolism using the method “discrete” within the R-opm package. Substrates differentiating the isolates from each other were visualized as a heatmap generated using the R-packages heatmap.plus with the Euclidean algorithm. The
heatmap displays the utilization of each substrate with a color key: yellow for strong positive metabolism, green for moderate metabolism and blue for no metabolism.

**Analysis of metabolic pathways**

The metabolic pathways of the two substrates of interest butyric and propionic acid have been further analyzed. Specifically, we extracted all sequences of the genes known to be associated with the pathways related to butyric and propionic acid from the KEGG pathway database (Kanehisa et al. 2016). We extracted the genes from all the *M. avium* subspecies (n=8) present in the KEGG pathway database, namely: *M. avium* subsp. *paratuberculosis* K-10, *M. avium* subsp. *paratuberculosis* MAP4, *M. avium* subsp. *paratuberculosis* E1, *M. avium* subsp. *paratuberculosis* E93, *M. avium* subsp. *avium* DJO-44271, *M. avium* subsp. *avium* 2285 (R), *M. avium* subsp. *avium* 2285 (S) and the *M. avium* 104. The redundant genes have been excluded. Then we screened all such genes in genomes of our ten MAH isolates by performing a Custom BLAST analysis using Geneious version 9 (Kearse et al. 2012). The parameters for the screening that we used to determine if a gene was present or not were: sequence identity ≥ 90 %, sequence coverage ≥ 90 %, e value ≤ 0.01.

In addition, we analyzed the number of Single Nucleotide polymorphisms (SNP)s (both synonymous and nonsynonymous) in the sequence of the genes detected in our MAH isolates. For each gene, we also constructed a phylogenetic tree using the nucleotide sequences to determine whether any SNP was associated with clinical or environmental source of the isolates based on the Tamura–Nei model using Geneious version 9.

**Statistical analyses**
We generated two groups, one with data from all clinical isolates and the other with data from all environmental isolates. Statistical differences between clinical and environmental isolates in the metabolism of butyric acid and propionic acid were evaluated by means of 95% family-wise comparison of group means (Tukey contrast test) of the parameter A on specific wells using the function “opm_mcp” within the opm R-package. A p value less than 0.05 was considered to be statistically significant.

Whole genome sequencing of MAH isolates

Genomic DNAs were extracted from the MAH isolates as described previously (Lewin et al. 2003). Whole genome sequencing (WGS) was performed using Illumina MiSeq 300 bp paired-end sequencing, yielding a coverage that exceeded 100x. The NGS QC tool kit was used to assess the quality of the data reads, which was set as reads with a minimum of 70% of bases having a phred score greater than 20 (Patel & Jain 2012). De novo assembly of the resulting reads into multiple contigs was performed using CLC Genomics Workbench 8.0 (CLC bio, Aarhus, Denmark) and contigs annotation was done using RAST (Aziz et al. 2008).

Determination of the maximum common genome and of the accessory genome

We determined the maximum common genome (MCG), comprising those genes present in all of the ten MAH genomes, as reported previously (von Mentzer et al. 2014). All these genes were then extracted from all genomes, concatenated and aligned. The resulting alignment was used to generate a clustering tree using RAxML 8.1 (Stamatakis 2014).

For determination of the accessory genome we applied the PanGenome Pipeline – Roary. After determination of the accessory genome of the ten MAH genomes and its distribution within
them, we separated those genes that are exclusively present only in either the environmental strains or the clinical strains (Page et al. 2015).
Results

Substrate utilization of the ten MAH isolates

We tested the capability of our ten MAH isolates to metabolize 379 different substrates. In total, 334/379 (88.1 %) substrates were negative for all of the isolates (see Supplemental Table S1). A total of 23/379 (6.1 %) substrates caused abiotic reactions and were excluded from further analysis. A list of false-positive substrates is shown in the Supplemental Table S2. The kinetic curves corresponding to the control plates PM1 to PM4 tested without bacteria are presented in the Supplemental Fig. S1.

Only two carbon substrates, the fatty acid derivatives Tween 20 and Tween 40 were strongly positive for all of the ten MAH isolates. The kinetic curves for these substrates reached 250 Omnilog units, amongst the highest values recorded in our analysis (see Supplemental Fig. S2 for all kinetic curves of the ten MAH isolates). The opm analysis revealed that a total of 20/379 (5.3 %) substrates were metabolized differently among the MAH isolates (Table 3). We therefore carried out further analysis using only these substrates. The majority of these 20 substrates were carbon substrates, 15/20 (75.0 %), followed by 3 nitrogen and 2 phosphorous substrates. The heatmap in Fig. 1 shows the utilization of these 20 substrates among the ten MAH isolates. The isolates are grouped according to their substrate utilization. Isolates utilizing similar substrates appear to cluster together.

Two major clusters, each composed of five isolates, could be observed. One was rich in environmental isolates (4/5) and the other was rich in clinical isolates (4/5). The substrates predominantly contributing to this clustering were butyric acid and propionic acid and indeed,
the Tukey’s test revealed that environmental isolates metabolized more strongly butyric acid \( (p = 0.0209) \) and propionic acid \( (p = 0.00307) \) than clinical isolates with statistical significance.

**Metabolic pathways analysis**

The propionic and butyric acid are involved in three and one pathway, respectively (Table 3). A total of 151 genes have been identified in the KEGG database associated with all these pathways (Kanehisa et al. 2016). In the supplemental Table S3 we reported the distribution and SNPs analysis of those genes in the MAH genomes. Of the 151 genes, 134 (88.7 %) are present in all the ten MAH. The median gene length was 1099 bp (range 318-2253), whereas the median number of SNPs per gene is 17 (range 1-147). The phylogenetic analysis revealed that none of the SNPs could be associated with the group of the clinical or the group of environmental MAH isolates (see supplemental Table S4). In the propanoate pathway four operons have been identified: *fadAB* associated with the β-oxidation of several fatty acids (DiRusso 1990), *ech8-9* encoding for hydrogenases that play a role in energy conversion (Sant’Anna et al. 2015), *sucCD* responsible for the succinate metabolism (Cerdeno-Tarraga et al. 2003) and *mutAB* involved in the methylmalonate pathway (Schoenwolf & Alvarez 1989). In the nicotinate pathway there are two operons: the *pntAA-AB-B* responsible for the transhydrogenation between NADH and NADP (Anderlund et al. 1999) and *nadABC* involved in the biosynthesis of NAD+ (Vilcheze et al. 2010). In the degradation of aromatic compounds pathway we identified the *pcaHGB* operon involved in the β-ketoadipate pathway (Harwood & Parales 1996). In the butanoate pathway we identified the *fadAB* and *ech8-9operons*, the *sdhCDAB* encoding for the succinate dehydrogenase complex involved in the fatty acid metabolism (Nam et al. 2005) and the *ihvBN* responsible for the acetylactate synthesis, a precursor of several amino acids (Keilhauer et al. 1993).
Clustering analysis and determination of the accessory genome

The WGS of the two reference strains MAH 104 and MAH TH135 were already in the GenBank database and we submitted the remaining genomes at DDBJ/EMBL/GenBank under the BioProject Number PRJNA299461. The MCG, the maximum number of genes shared by all ten MAH isolates was 1,658, the alignment of which spanned 1.378 Mbp. The clustering analysis of the ten MAH isolates is shown in Fig. 2. By comparing the genetic clustering obtained by WGS with the phenotypic clustering obtained through BIOLOG PM we observed slight differences. For examples, the isolates MAH E-96-2 and MAH E-82-7, which share identical metabolic profiles, were genetically more distant from each other. Interestingly, at the genetic level there was no obvious clustering between the group of clinical and the group of environmental isolates.

The accessory genome is constituted by 4,067 genes. A total of 1688 genes were specific for the group of clinical isolates (supplemental Table S5). On the other hand, 698 genes were specific for the group of environmental isolates (supplemental Table S6). We found no genes that were present in all the clinical and absent in all the environmental isolates, and vice-versa. Among the most abundant specific genes of the two groups of isolates, there were no known genes associated with the pathways which involved butyric and propionic acid. However, genes annotated as hypothetical proteins represented the most abundant specific genes of the two groups.
Discussion

This study represents the first phenotypic analysis of a collection of clinical and environmental MAH isolates using the Biolog PM technology. We showed that the PM technology works well and can be performed with MAH isolates. Strong positive reactions with several substrates were observed with kinetic curves exceeding 200 Omnilog dye units. Although some substrates were metabolized only moderately by our MAH isolates (green in Fig. 1), this might be due to the fact that the use of such substrates by bacteria has a time lag.

The ten MAH isolates showed different metabolic patterns pointing to high intra-species diversity. Only two out of the ten isolates had identical heatmap profiles (MAH E-96-2 and MAH E-82-7).

Our study showed that MAH isolates prefer to metabolize fatty acids as a carbon source. Indeed, the Tween substrates were strongly metabolized by all MAH isolates tested. This is in agreement with prior studies, showing that Tween substrates were widely used by different mycobacterial species (Baloni et al. 2014; Chen et al. 2012; Hayashi et al. 2010; Khatri et al. 2013; Lofthouse et al. 2013; Wang et al. 2011). It has been reported that mycobacteria hydrolyze Tween 80 to generate the fatty acid oleic acid, which can enter the Tricarboxylic acid (TCA) cycle or can be used as a substrate for energy production (Lofthouse et al. 2013; Vandal et al. 2009). Other fatty acids used by the majority of our MAH isolates are represented by two short fatty acids, caproic acid and butyric acid (Kanehisa & Goto 2000; Kanehisa et al. 2016; Khatri et al. 2013). Caproic acid and its derivatives are involved in several mycobacterial pathways such as the degradation of aromatic compounds, oxocarboxylic acid metabolism or lysine degradation (Kanehisa & Goto 2000; Kanehisa et al. 2016). The butyric acid is the final product of butanoate metabolism.
Propionic acid is another fatty acid used by our MAH isolates and this represents the terminal product of propanoate metabolism (Kanehisa & Goto 2000; Kanehisa et al. 2016). The nitrogen source L-cysteine, used by six of our MAH isolates, is the final product of cysteine metabolism and is involved in the biosynthesis of other amino acids such as methionine and histidine (Baloni et al. 2014; Kanehisa & Goto 2000; Kanehisa et al. 2016).

The question of whether bacteria of the same species originating from either clinical or environmental sources differ from each other is still a matter of discussion. Li and co-authors showed that comparative genome analysis clearly distinguished clinical and environmental *Vibrio parahaemolyticus* isolates from each other (Li et al. 2014). In contrast, other researchers have reported no difference between clinical and environmental *Pseudomonas aeruginosa* isolates with regard to virulence and metabolic properties (Alonso et al. 1999; Vives-Flores & Garnica 2006). Although our study did not reveal any clear distinction between clinical or environmental MAH isolates at the level of the whole genome, we observed differences between clinical and environmental isolates with regard to substrate utilization. The most intriguing difference is that the two fatty acids butyric acid and propionic acid are metabolized more by the environmental than by clinical isolates.

We observed no difference in the presence / absence of genes associated with butyric or propionic acid pathways among the group of clinical and the group of environmental MAH isolates. The SNPs analysis of the genes involved in the pathways revealed that no SNPs were associated with clinical or environmental origin of the MAH isolates. These evidences suggest that the metabolic differences observed among clinical and environmental MAH isolates might be due to difference in gene regulation. However, we screened all the genes that up to now have
been associated with the pathways of interest. We can speculate that there might be additional
genes, of unknown function, that might play a role in the above pathways.

The analysis of the accessory genome revealed that none of the genes specific for the clinical or
for the environmental isolates could be associated with the pathways of interest. However, future
studies on the high number of hypothetical proteins might clarify whether they have a role in the
pathways which involved the butyric and propionic acid.

The higher metabolic activity observed among environmental MAH isolates might be
advantageous for survival in an environment presenting a wider range of nutritional conditions
than the host cells alone. Further studies testing a larger number of isolates from different origins
might clarify this. In addition, it has been showed that in bacteria the fatty acids have a role in
adaptation to different environmental conditions (de Sarrau et al. 2012; de Sarrau et al. 2013;
Diomande et al. 2015).
279 **Conclusions**

280 Our study contributes to the understanding of the emerging pathogen MAH at the phenotypic and metabolic level. Understanding how bacteria utilize their own or host-derived substrates during infection might help the development of strategies to fight such infections. We encourage phenotypic testing of microbial isolates from different ecological niches to identify key substrates or pathways that can be used as targets for drug development or for selective growth media development.

286
Acknowledgments

We would like to thank Barry Bochner (President of BIOLOG) for his invaluable input to this study and Brian Weinrick (Albert Einstein College of Medicine, New York City) for his support in developing the BIOLOG laboratory protocol. Elvira Richter (National Reference Center for Mycobacteria, Borstel, Germany) and Roland Schulze-Röbbecke (University Hospital Düsseldorf) provided a number of MAH isolates, Carsten Schwarz (Christiane Herzog Zentrum, Charité, Berlin) provided respiratory samples from cystic fibrosis patients and Kei-ichi Uchiya provided the reference strain MAH TH135. We thank Katharina Schaufler (Free University Berlin) for her support with the BIOLOG data analysis and Inga Eichorn (Free University Berlin) for her help with the whole genome sequencing data. We thank Steve Norley (Robert Koch Institute, Berlin) for the English revision of the manuscript.
References


Manuscript to be reviewed


annotation and revised genome sequence. *J Bacteriol* 192:6319-6320. 10.1128/JB.00972-10
Table 1 Characteristics of the ten MAH isolates analyzed in this study.

<table>
<thead>
<tr>
<th>MAH Isolate name</th>
<th>Year of isolation</th>
<th>Source</th>
<th>Provider or reference</th>
<th>Accession of whole genome sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-10091-06</td>
<td>2006</td>
<td>Clinical - Child with lymphadenitis</td>
<td>NRC for Mycobacteria, Borstel, Germany</td>
<td>LNAV00000000</td>
</tr>
<tr>
<td>2721</td>
<td>2004</td>
<td>Clinical - Child with lymphadenitis</td>
<td>NRC for Mycobacteria, Borstel, Germany</td>
<td>AWXJ00000000</td>
</tr>
<tr>
<td>P-9-13</td>
<td>2013</td>
<td>Clinical - Adult pulmonary infection</td>
<td>Charité Hospital, Berlin, Germany</td>
<td>LNB600000000</td>
</tr>
<tr>
<td>104</td>
<td>1983</td>
<td>Clinical - Adult pulmonary infection</td>
<td>Reference strain, USA</td>
<td>CP000479</td>
</tr>
<tr>
<td>TH135</td>
<td>2013</td>
<td>Clinical - Adult pulmonary infection</td>
<td>Reference strain, Japan</td>
<td>AP012555</td>
</tr>
<tr>
<td>E-128</td>
<td>2010</td>
<td>Environmental - Soil</td>
<td>Friedrich Löffler Institute, Jena, Germany</td>
<td>LVCS00000000</td>
</tr>
<tr>
<td>E-96-2</td>
<td>2010</td>
<td>Environmental - Soil</td>
<td>This study</td>
<td>LMWV00000000</td>
</tr>
<tr>
<td>E-82-7</td>
<td>2010</td>
<td>Environmental - Dust</td>
<td>This study</td>
<td>LNAF00000000</td>
</tr>
<tr>
<td>27-1</td>
<td>2010</td>
<td>Environmental - Dust</td>
<td>This study</td>
<td>AWXK00000000</td>
</tr>
<tr>
<td>E-2514</td>
<td>na</td>
<td>Environmental - Water</td>
<td>University of Düsseldorf, Germany</td>
<td>LNBJ00000000</td>
</tr>
</tbody>
</table>

MAH: Mycobacterium avium subsp. hominissuis; NRC: National reference center; na: not available
Table 2 Additives used for each PM plates. As additive are usually provided nutrient that are absent to the PM minimal media, but present in a standard MAH growth conditions. We used additives to make a complete minimal medium but omitted anything that could act as a source of the substrates of interest (for example, we did not include nitrate additives in the nitrogen source plates).

<table>
<thead>
<tr>
<th>PM plate usage</th>
<th>Additive a</th>
<th>Additive b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingredients</td>
<td>24mM MgCl2</td>
<td>24mM MgCl2</td>
</tr>
<tr>
<td></td>
<td>12mM CaCl2</td>
<td>12mM CaCl2</td>
</tr>
<tr>
<td></td>
<td>0.0012% ZnSO4</td>
<td>0.0012% ZnSO4</td>
</tr>
<tr>
<td></td>
<td>0.06% ferric ammonium citrate</td>
<td>0.01% tween 80</td>
</tr>
<tr>
<td></td>
<td>1.2% NH4Cl</td>
<td>1.2% NH4Cl</td>
</tr>
<tr>
<td></td>
<td>0.01% tween 80</td>
<td>0.01% tween 80</td>
</tr>
</tbody>
</table>
### Table 3: The 20 substrates differentiating the ten MAH isolates analyzed in this study.

<table>
<thead>
<tr>
<th>PM Plate</th>
<th>Substrate and well number</th>
<th>Pathway involved</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>Acetic acid – C08</td>
<td>Pyruvate metabolism</td>
<td>(Baloni et al. 2014; Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>Acetoacetic acid – G07</td>
<td>Pyruvate metabolism</td>
<td>(Baloni et al. 2014; Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>Methyl pyruvate – G10</td>
<td>Pyruvate metabolism</td>
<td>(Baloni et al. 2014; Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>Mono–methyl Succinate – G09</td>
<td>Tricarboxylic acid cycle</td>
<td>(Baloni et al. 2014; Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>D–psicose – H05</td>
<td>Glycolysis and branches</td>
<td>(Baloni et al. 2014; Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>Pyruvic acid – H08</td>
<td>Pyruvate metabolism</td>
<td>(Baloni et al. 2014; Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>Tween 80 – E05</td>
<td>Fatty acid metabolism</td>
<td>(Baloni et al. 2014; Nai et al. 2013)</td>
</tr>
<tr>
<td>Carbon</td>
<td>L–alaninamide – G02</td>
<td>Amino acid metabolism</td>
<td>(Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>Caproic acid – E02</td>
<td>Carboxylic acid metabolism</td>
<td>(Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>L–histidine – G06</td>
<td>Amino acid metabolism</td>
<td>(Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>γ–hydroxy–butyric acid – E09</td>
<td>Succinate metabolism</td>
<td>(Breitkreuz et al. 2003; Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>β–methyl–D–galactoside – C07</td>
<td>Galactose Metabolism</td>
<td>(Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>Sebacic acid – F08</td>
<td>Carboxylic acid metabolism</td>
<td>(Nai et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>L–cysteine – A11</td>
<td>Amino acid metabolism</td>
<td>(Baloni et al. 2014)</td>
</tr>
<tr>
<td></td>
<td>D–galactosamine – E09</td>
<td>Amino-sugar pathway</td>
<td>(Baloni et al. 2014)</td>
</tr>
<tr>
<td>Phosphorous and sulphur</td>
<td>Carbamyl phosphate – B05</td>
<td>Urea cycle and Pyrimidine synthesis</td>
<td>(Nelson 2004)</td>
</tr>
<tr>
<td></td>
<td>Sodium pyrophosphate – A03</td>
<td>Phosphoric acid synthesis</td>
<td>(Nelson 2004)</td>
</tr>
</tbody>
</table>
Figure 1

Heatmap showing the 20 substrates that were differently metabolized by the ten MAH isolates analyzed in this study.

The color key scale for each substrate is based on dye reduction quantified by Omnilog units. A yellow color indicates strong positive substrate metabolism, a green color moderate metabolism and a blue color indicates no substrate metabolism. Regarding the MAH isolates, environmental isolates are marked in orange, while clinical isolates are marked in blue.
Figure 2

Clustering of the 10 MAH isolates.

The tree was generated using RAxML 8.1. The alignment comprised 1,658 genes constituting the maximum common genome of our ten MAH isolates. Two reference strains were also included (MAH 104 and MAH TH135). The genome sequence of \textit{M. avium} subsp. \textit{paratuberculosis} K10 (Accession Number: AE016958) was used as outgroup. Isolate origin is also represented by blue for clinical origin and orange for environmental origin. The percentage of trees in which the associated taxa clustered together is shown adjacent to the branches.
4. Discussion

This thesis contributes to our knowledge about genetic and metabolic diversity of MAH, which are emerging opportunistic human pathogens also widespread in the environment. We identified and characterized the hvGI in a population of MAH isolates from Germany. This hvGI contributes to the genetic diversity of MAH isolates. Our study expanded the knowledge on genome plasticity of MAH. In addition, we also studied the metabolic properties of MAH isolates; specifically we gave an overview on the use of substrates by MAH isolates from different sources.

4.1 Implications of the hvGI for MAH

We identified and characterized the hvGI in more than forty MAH isolated from clinical and environmental sources. The hvGI is a dynamic region; we identified eight types of hvGI, also organized in combinations of multiple hvGIs inserted in the same position of the MAH chromosome. For these reasons the hvGI was termed as hypervariable. GIs have been previously identified in mycobacteria. For example, three GIs were recently identified in MTB, named MP-1, MP-2 and MP-3, carrying genes important for bacterial pathogenesis. The MP-1 GI carries the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system, a system responsible for the persistence of MTB infections (Hille & Charpentier 2016). The MPI-2 and MPI-3 GIs carry other important genes such as genes encoding for PE/PPE family proteins, genes encoding for a type VII secretion system and genes responsible for the mycolic acids biosynthesis (Xie et al. 2014). The importance of mycolic acids for mycobacterial virulence and host-pathogen interaction has been discussed above. In NTM more evidences regarding GIs are coming: in *M. massiliense*, one of the subspecies of *M. abscessus*, the GI MmGI-1 has been described, with putative function in lipid metabolism. The genes of the MmGI-1 were similar to genes of MAC, suggesting an HGT between *M. massiliense* and MAC isolates (Xie et al. 2014). An alternative possibility is that both species acquired these genes from other bacteria. Another GI has been analyzed in MAP, with a possible role in zinc metabolism (Eckelt et al. 2014). However, all these studies have been conducted only on single isolate; therefore it is not known whether these GIs are
shared by all isolates of the same species or only by specific subpopulations or isolates. The hvGI that we discovered has been found in both clinical and environmental isolates. We did not find any correlation between presence of a specific hvGI and isolation from either patients or environment. The same evidences came out also when we further stratified the analysis for specific isolation source (pulmonary infections, lymphadenitis, water, soil or dust). When we compared the maximum common genome of the MAH isolates containing the hvGI, we observed no clustering between clinical and environmental MAH isolates. Clinical and environmental MAH seem to represent a heterogeneous mixture of isolates. In conclusion the hvGI seems not to be a marker or to determine whether an MAH isolate is clinical or environmental. Another study reported a different conclusion: Barthpo and co-authors demonstrated that GIs differentiated among clinical and environmental *Burkholderia pseudomallei* isolates (Bartpho et al. 2012). One can speculate that in case of MAH host factors such as immunological reactions determine if colonization with MAH will cause an infection.

The hvGIs carry interesting genes annotated as known/putative virulence factors or genes known to be involved in host-pathogen interaction or antibiotic resistance. For example, the *mce* genes are organized in an operon encoding for the mammalian cell entry (MCE) proteins (Forrellad et al. 2013). MCE proteins are known virulence factors in MTB, being responsible for the invasion and the survival of MTB within human macrophages (Arruda et al. 1993; Stewart et al. 2005). In addition, *mce1A* and *mce3A* are important for bacterial penetration also in other types of cells such as epitheliocytes (El-Shazly et al. 2007). Another attractive set of genes is represented by the genes encoding for PE/PPE proteins, present for example in the hvGI of the reference strain MAH 104. Genes encoding for PE/PPE proteins are largely represented in the mycobacterial genomes, since they represent almost 10% of the whole mycobacterial genome (Mukhopadhyay & Balaji 2011). These proteins are located in the mycobacterial cell wall (Sampson 2011). The different PE/PPE proteins might have different roles in mycobacteria. In one example, two mutagenesis studies indicate that these proteins inhibit the maturation and the acidification of the phagosome, allowing for MTB survival within human cells (Akhter et al. 2012; Li et al. 2005b). Similar to our study, Saini and co-
authors also found PE/PPE genes located within mobile genetic elements such as GI in *M. indicus pranii* isolates (Saini et al. 2012). In our hvGI there are also genes encoding for the nitric oxide reductase activation protein which might be involved in the resistance to the nitric oxide produced by the macrophages after mycobacterial ingestion (Jung et al. 2013).

The hvGI contains also a gene annotated as *mmpL10* that stimulated our curiosity. The *mmpL* genes encode for the Mycobacterial membrane protein Large (MmpL) proteins that are transmembrane proteins and members of the Resistance, Nodulation and Cell division protein family. MmpL proteins are responsible for different functions, such as the transport of glycolipids through the mycobacterial cell wall or the heme uptake (Bailo et al. 2015). MmpL proteins have been studied almost exclusively in MTB where they demonstrated virulence potential. Indeed, experiments using mouse models revealed that MmpL4 and MmpL7 are required for early stages of infection by MTB, while MmpL8 and MmpL11 are crucial at a chronic stage of infection (Bailo et al. 2015; Domenech et al. 2005). In addition, the impact of MmpL proteins on antibiotic-resistance is controversially discussed. Domenech and co-authors reported that MmpL proteins do not affect antibiotic-resistance in MTB (Domenech et al. 2005). On the contrary, recent studies suggested an association of MmpL5 with cross-resistance to clofazimine (used as third-line drug for TB) and to bedaquiline (a diarylquinolines recently approved for TB treatment) in MTB (Andries et al. 2014; Bailo et al. 2015; Jang et al. 2008).

For all these reasons, and because there are scarce knowledge on the role of MmpL proteins in MAH, we decided to investigate some of the functions of the *mmpL10* gene present in the hvGI (data not shown). We reduced the *mmpL10* gene expression in one of our MAH isolates (MAH 27-1) by about 50% by introducing a *mmpL10* antisense plasmid into this strain. Then we performed functional studies on the wild-type MAH 27-1 and its antisense derivative with reduced *mmpL10* gene expression. We measured the minimal inhibitory concentrations (MICs) towards 13 clinically relevant antibiotics in order to investigate whether the *mmpL10* gene influences the antibiotic-resistance. We found that the antisense strain was less resistant (two-fold difference in MICs) to Rifampin and Linezolid when compared to the wild-type strain, suggesting that the *mmpL10* gene might play a role in antibiotic-resistance. We
hypothesize that the MmpL10 protein in our MAH isolates functions as transmembrane channel and can be involved in the extrusion of antibiotics form the cell (data not shown). In the literature it has been shown that in *M. avium* the induction of transmembrane efflux pumps is an important step in the development of antibiotic-resistance (Schmalstieg et al. 2012). The performance of laboratory tests on additional isolates, along with the generation of an *mmpL10* knockout mutant may further confirm this hypothesis.

We subjected the wild-type MAH 27-1 and its antisense derivative strain to phenotype microarray screening using the Phenotype MicroArray™ (PM) (BIOLOG, Hayward CA) assay (Bochner et al. 2001; Khatri et al. 2013), in order to determine if the *mmpL10* gene determines metabolic activities (data not shown). The wild type strain metabolized stronger the sugar D-Psicose when compared to the antisense strain. This evidence suggests that the MmpL10 protein might be involved in the transport of certain carbohydrates across the membrane. This result of our study is similar to results shown by Belardinelli and co-authors, who showed that the MmpL10 protein in MTB is involved in the transport of acylated trehaloses (Belardinelli et al. 2014).

These functional studies suggest that the *mmpL10* gene contained in the hvGI of MAH might have a role in antibiotic-resistance and in carbohydrates transport across the mycobacterial cell wall supporting the hypothesis that the hvGI plays a role in metabolism and adaptation of MAH. GIs are genetic elements integrated in the chromosome able to spread to other species via HGT. The mobility of the hvGI might transfer genes with known or potential virulence or antibiotic-resistance capacity (such as *mmpL10*, *mce* or PE/PPE) to other species. This is of special awareness for *M. avium* because it is a pathogen well-known for acquiring and transmitting genetic material from and to other species, including phylogenetically distant organisms. Indeed Reva and co-authors, through WGS, analyzed the GI distribution and phylogeny in several mycobacterial species, including MTB and other NTM. They observed a frequent exchange of genes especially in *M. avium*, which was the most promiscuous among the analyzed mycobacteria (Reva et al. 2015). Our study showed also evidences of HGT, by comparing the DNA sequences of the hypervariable GI with sequences of other mycobacteria. DNA similarities were found between the hvGI from MAH and sequences of
*M. intracellulare*, *M. marinum* and *M. canettii*, a member of the MTBC (Fabre et al. 2010; Supply et al. 2013). This evidence suggests once again that MAH can acquire or donate genetic material, also from/to highly pathogenic mycobacteria.

The acquisition of mobile genetic element via HGT is particularly important for the evolution of mycobacteria: it has been hypothesized that harmless mycobacteria present in the environment acquired genetic elements from other bacteria present in their surroundings and thus become adapted to and pathogenic for humans (Arnold 2007; Banuls et al. 2015; Behr 2013; Bottai et al. 2014; Coros et al. 2008; Gutierrez et al. 2009). This concept is explained by the model of MTB evolution. MTB is supposed to derive from an ancient species, close to the *M. canettii*, that later acquired GIs from alfa- and gamma- proteobacteria (such as *Pseudomonas, Burkholderia*) and thus became adapted to humans (Becq et al. 2007; Kinsella et al. 2003; Reva et al. 2015; Rosas-Magallanes et al. 2006). The transfer of GIs and other genetic material might have increased the adaptation of the recipient to cope with several possible stress conditions. One of the possible stress conditions was the resistance to ingestion by protozoa, which are spread for example in soil. Acquisition of genetic material by environmental bacteria might have provided to them genes conferring resistance to ingestion and/or degradation by protozoa, and therefore it is hypothesized that in this way the mycobacteria learned to survive within human macrophages (Jang et al. 2008). The ability to survive within protozoa such as *Acanthamoeba* was also demonstrated for *M. avium* (Cirillo et al. 1997; Drancourt et al. 2007; Mura et al. 2006; Whan et al. 2006). Adaptation of MTB to humans (and animals) is supposed to have taken part along with loss of genetic material, while *M. avium* did not undergo relevant loss of genetic material, why it is adapted to a wider range of environmental and clinical conditions (Ignatov et al. 2012).

Our results imply that more awareness on genetic diversity and more studies on genomic surveillance of GIs or other mobile genetic elements in MAH and other NTM are needed, because the shuffling of these elements may generate new variants potentially able to cause infections or outbreaks of public health relevance.
4.2 Implications of the metabolic studies on MAH

Our study contributes to understand MAH at phenotypic-metabolic level. For the first time we described and compared the range of metabolic substrates that can be utilized by MAH isolates from clinical and from environmental sources. The PM method is a high-throughput method for the rapid screening of metabolic substrates used by bacteria. We found that our MAH isolates mostly used fatty acids as energy sources. By clustering the MAH isolates based on their metabolic substrates utilization, we found that clinical and environmental isolates represent two distinct clusters. The major difference between clinical and environmental MAH isolates consisted in the usage of the two substrates propionic acid and butyric acid, which are stronger metabolized by environmental than clinical isolates. Interestingly, we did not observe this clustering when we analyzed the maximum common genome of all our MAH isolates. This means that although MAH isolates from clinical and environmental origin do not cluster at this genetic level, they differed at metabolic level. This might suggest that bacteria sharing the same genetic information are able to regulate differently their metabolism depending on the conditions that they have to face in their habitat (Archuleta et al. 2005; Honer zu Bentrup & Russell 2001; Marri et al. 2006). Moreover, it has to be considered that the metabolism is influenced not only by the gene composition, but also by epigenetic effects (such as DNA methylation) influencing the gene expression (Kathirvel & Mahadevan 2016). This reinforces the statement that the accessory genome should be analyzed to understand its role in bacterial evolution and survival. Other studies also showed that in bacteria the use of fatty acids plays a role in adaptation to different environmental conditions. For example, de Sarrau and co-authors investigated the growth, the metabolism and the bacterial cell wall composition under anaerobiosis and low temperature of *Bacillus cereus*. They observed a reduced growth in these conditions and a modified pattern of fatty acids production, which was responsible for an increase of the cell wall permeability, allowing bacterial survival (de Sarrau et al. 2012; de Sarrau et al. 2013). Similar researches have been done by Diomande and co-authors, which showed that several *Bacillus* species modify their fatty acids pattern in response to different media compositions, temperatures or pH conditions (Diomande et al. 2015).
We also compared the substrate utilization between MAH and MTB (Khatri et al. 2013; Lofthouse et al. 2013). Interestingly our MAH isolates utilized substantially less substrates when compared to MTB. However, there is a need of harmonization of the PM data analysis algorithms, in order to adjust methodology for determining positivity of a substrate or setting of positivity thresholds for achieving comparable and reproducible results. A total of nine substrates differentiate MAH from MTB (specifically propionic acid, D-psicose, L-alaninamide, L-histidine, γ-hydroxy-Butyric Acid, β-Methyl-D-Galactoside, Sebacic Acid, D,L-α-Amino-Caprylic Acid and Carbamyl Phosphate). More detailed knowledge on substrate utilization of mycobacteria might improve culture and identification techniques.

More indications are coming about the importance of metabolism for mycobacterial pathogenesis. For example, researchers determined that MTB and *M. avium* use fatty acids and cholesterol of the host-cells as energy source during infections to persist within macrophages (Caire-Brandli et al. 2014; Daniel et al. 2011). During infection mycobacteria switch from the usage of own metabolic substrates to the usage of host-derived fatty acids (Schnappinger et al. 2003). Host lipids represent the primary carbon source during the first phase of infections (Caire-Brandli et al. 2014; Daniel et al. 2011). Another example is the usage of asparagine by MTB during host infection both as nitrogen source and for resistance towards pH stress mediated by production of ammonia through asparagine hydrolyzation (Gouzy et al. 2014a; Gouzy et al. 2014b).

Understanding how bacteria utilize own or host-derived substrates during host-pathogen interaction might open new therapeutic options for treatment of bacterial infections. We therefore encourage phenotypic testing of microbial isolates from different ecological niches in order to identify pathogen-specific key metabolites or pathways that can be used as targets for the development of new drugs.
5. Conclusions and outlook

The exploration of the biology of MAH is important because (i) it is a pathogen adapted to a wide range of environments and conditions, (ii) it is an emerging opportunistic human pathogen and (iii) it is prone to transfer genetic material such as GIs to other species. Our study highlighted the high degree of the genetic diversity of the newly identified hvGI of MAH. Preliminary functional studies suggest that the hvGI might have an impact on the adaptation to certain types of selective pressure and consequently on evolution of MAH. Further studies on the functions of genes present in hvGI may contribute to understanding of its role for evolution and pathogenesis of MAH. Mobile elements shaping the accessory genome of mycobacteria should be given more attention since the transfer of these elements drives bacterial evolution and adaptation.

Our study suggests that MAH isolates might adapt their metabolism to the different conditions they encounter during their life. The study of the role of metabolism in bacterial virulence is a relatively new but expanding field (Lundgren et al. 2015; Njoroge et al. 2012; Santic & Abu Kwaik 2013; Schoen et al. 2014).

Important questions to answer in the future:

- How does MAH regulate the mobility and the transfer the hvGI?
- What is the function of the genes, especially genes annotated as hypothetical proteins, present in the hvGI?
- Which metabolic substrates are used by MAH during human infection?

The identification of new GIs and new genes, and their function in colonization or host-pathogen interaction might help defining novel targets for the development of new antimicrobials or develop specific vaccines (Ahmed et al. 2008; Jang et al. 2008; Turenne et al. 2007). Functional studies, for example generating gene-specific knockout mutants, might clarify the role of the hvGIs and its genes. The screening of the utilization of metabolic substrates suggests specific substrates to be further investigated for their importance in bacteria.
6. Summary

*Mycobacterium avium* subsp.*hominissuis* (MAH) is an opportunistic human pathogen that can cause lymphadenitis, pulmonary infections and disseminated infections. In addition, MAH is widespread in the environment, since it has been isolated from dust, soil and water. MAH isolates are characterized by high genetic diversity. Recently a new genomic island (GI), later re-named as hypervariable GI (hvGI), was identified in few MAH isolates, contributing to the genetic diversity of the MAH isolates.

It is not known whether clinical or environmental MAH isolates differ from each other. In this thesis we analyzed 41 MAH isolates from Germany isolated from clinical (n=20) and from environmental (n=21) source. First we identified and characterized the hvGI in all isolates, in order to see if the hvGI differentiates clinical and environmental MAH isolates. Then we investigated the function of the *mmpL10* gene of the hvGI in order to get insights on the function of one of the genes present in the hvGI. We identified the hvGI in 39/41 isolates. We found high genetic diversity in the hvGI: eight types of hvGI have been identified (size 6.2-73.3 kb). Two types shared more than 80% sequence identity with *Mycobacterium canettii* responsible for Tuberculosis. We identified 253 different genes in all hvGIs, among which the previously documented virulence genes *mmpL10* and *mce*. Functional studies on the *mmpL10* gene suggest its involvement in antibiotic-resistance and in sugar transport. Our study expands the knowledge on MAH genome plasticity. The diversity of the hvGIs and the similarities with other mycobacteria suggests cross-species transfer. The shuffling of virulence/drug-resistance genes via the hvGIs may generate new variants able to can cause new outbreaks.

In addition, we analyzed ten MAH isolates at metabolic level using the BIOLOG Phenotype Microarray method, in order to see whether clinical or environmental MAH isolates show any metabolic differences. We found that MAH metabolized mostly fatty acids such as Tween, caproic, butyric and propionic acid. Clinical MAH metabolized stronger butyric (*p = 0.0209*) and propionic acid (*p = 0.00307*) compared environmental MAH. Our study provides new insight into the metabolism of MAH. Understanding how bacteria utilize substrates during infection might support the development of strategies to fight such infections.
7. Zusammenfassung

*Mycobacterium avium* subsp. *hominissuis*: die Bedeutung der genetischen und metabolischen Diversität


Antibiotikaresistenz-Genen durch die hvGIs kann neue Varianten von *Mycobacterium*-Arten erzeugen, die neue Ausbrüche verursachen könnten.

8. References


Coros A, DeConno E, and Derbyshire KM. 2008. IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis, suggestive of lateral gene transfer among mycobacterial species. *J Bacteriol* 190:3408-3410. 10.1128/JB.00009-08


Coros A, DeConno E, and Derbyshire KM. 2008. IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis, suggestive of lateral gene transfer among mycobacterial species. *J Bacteriol* 190:3408-3410. 10.1128/JB.00009-08

Coros A, DeConno E, and Derbyshire KM. 2008. IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis, suggestive of lateral gene transfer among mycobacterial species. *J Bacteriol* 190:3408-3410. 10.1128/JB.00009-08


Kaneda K, Naito S, Imaiizumi S, Yano I, Mizuno S, Tomiyasu I, Baba T, Kusunose E, and Kusunose M. 1986. Determination of molecular species composition of C80 or longer-


Kathirvel M, and Mahadevan S. 2016. The role of epigenetics in tuberculosis infection. *Epigenomics* 8:537-549. 10.2217/epi.16.1


Thegerstrom J, Marklund BI, Hoffner S, Axelsson-Olsson D, Kauppinen J, and Olsen B. 2005. Mycobacterium avium with the bird type IS1245 RFLP profile is commonly found in wild and domestic animals, but rarely in humans. Scand J Infect Dis 37:15-20. 10.1080/00365540510026850


Treangen TJ, and Rocha EP. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet 7:e1001284. 10.1371/journal.pgen.1001284


91


Whan L, Grant IR, and Rowe MT. 2006. Interaction between Mycobacterium avium subsp. paratuberculosis and environmental protozoa. BMC Microbiol 6:63. 10.1186/1471-2180-6-63


9. List of publications

9.1 Scientific articles


9.2 Poster presentations at conferences

Andrea Sanchini, Torsten Semmler, Lei Mao, Narender Kumar, Flavia Dematheis, Kshitij Tandon, Vidyullatha Peddireddy, Niyaz Ahmed, Astrid Lewin. A Genomic Island identified in clinical and environmental *Mycobacterium avium* subsp. *hominissuis* isolates from Germany. 11th International Meeting on Microbial Epidemiological Markers (IMMEM XI), Lisbon, Portugal, 09-12/03/2016.


### 9.3 Scientific presentations at conferences


10. Acknowledgments

The first and the biggest acknowledgment is for my supervisor Dr. Astrid Lewin for her great availability and support during my PhD, for all the advice I received from her, for being the guide of my project also in difficult and less motivated moments and for the great patience she demonstrated in the three years. Many thanks for this wonderful supervision.

I would also like to thank my supervisor Prof. Dr. Lothar Wieler, first for having initiated the IRTG1673 PhD graduate school, which was a great learning opportunity and an enjoyable experience for me, and second for supervising my project during the three years.

I would also like to thank my supervisor Prof. Dr. Rupert Mutzel of the Biology Department and PD Dr. Volker Rickerts for supporting my project and for helping me during my professional development.

My special thanks go to Elisabeth Kamal, for teaching and helping me in all the laboratory procedures and protocols and for the great patience she demonstrated with me in the laboratory. Also, I would like to thank all the other colleagues of Dr. Astrid Lewin’s group, especially Dr. Annesha Lahiri, PD Dr. Lei Mao, Greana Kirubakar and Sarah Berger for their advice in the lab and for the enjoyable atmosphere we had while working. I also thank all the other colleagues of the FG16 department of the Robert Koch Institute for their assistance during my PhD.

My gratitude also goes to two special collaborators, Torsten Semmler for the help and assistance with genome sequencing data and Dr. Flavia Dematheis for her assistance with the BIOLOG experiments and for the statistical support.

I would also like to thank the coordinators of the IRTG1673 graduate school Dr. Elisabeth Otto and Dr. Esther-Maria Antáo for their invaluable help with the PhD logistics and for their assistance during the PhD.

Special thanks to the Indian collaboration team, especially to Dr. Niyaz Ahmed for having initiated the IRTG1673 PhD graduate school. Big thanks also to Shankar Doddam and
Gangadhar Dharmapuri who made me feel at home while I was in India for my research stay. I would also like to acknowledge Dr. Narender Kumar, Kshitij Tandon, Aditya Kumar and all the other members of Dr. Niyaz Ahmed’s group for their support during my research stay in India.

Also I would like to acknowledge the colleagues from the Free University Berlin, especially PD Dr. Soroush Sharbati for reviewing my thesis and for the support during the PhD, Univ.-Prof. Dr. Thomas Alter for reviewing my thesis, Dr. Katharina Schaufler for the support with the BIOLOG experiments and Inga Eichhorn and Petra Krienke for their support with the whole genome sequencing data.

I also express my gratitude to all the other members of the IRTG1673 PhD graduate school family; it was wonderful to share all these experiences with you.

I would like to thank the Deutsche Forschungsgemeinschaft and the Robert Koch Institute for funding my PhD.

And last but not least, I would like to thank my girlfriend Hannah Ruhm, my family and my friends for their love and encouragement during the PhD and because they always believe in me.
11. Selbstständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe.

Andrea Sanchini

Berlin, 10/07/2017