

Current Topics in Microbiology and Immunology

Jean Pieters

John D. McKinney *Editors*

Pathogenesis of  
*Mycobacterium  
tuberculosis* and  
its Interaction  
with the Host  
Organism

 Springer

# Current Topics in Microbiology and Immunology

Volume 374

## *Series editors*

Klaus Aktories

Medizinische Fakultät, Institut für Experimentelle und Klinische Pharmakologie und Toxikologie, Abt. I Albert-Ludwigs-Universität Freiburg, Albertstr. 25, 79104 Freiburg, Germany

Richard W. Compans

Department of Microbiology and Immunology, Emory University, 1518 Clifton Road, CNR 5005, Atlanta, GA 30322, USA

Max D. Cooper

Department of Pathology and Laboratory Medicine, Georgia Research Alliance, Emory University, 1462 Clifton Road, Atlanta, GA 30322, USA

Jorge E. Galan

Boyer Ctr. for Molecular Medicine, School of Medicine, Yale University, 295 Congress Avenue, room 343, New Haven, CT 06536-0812, USA

Yuri Y. Gleba

ICON Genetics AG, Biozentrum Halle, Weinbergweg 22, 06120 Halle, Germany

Tasuku Honjo

Faculty of Medicine, Department of Medical Chemistry, Kyoto University, Sakyo-ku, Yoshida, Kyoto 606-8501, Japan

Yoshihiro Kawaoka

School of Veterinary Medicine, University of Wisconsin-Madison, 2015 Linden Drive, Madison, WI 53706, USA

Bernard Malissen

Centre d'Immunologie de Marseille-Luminy, Parc Scientifique de Luminy, Case 906, 13288 Marseille Cedex 9, France

Fritz Melchers

Max Planck Institute for Infection Biology, Charitéplatz 1, 10117 Berlin, Germany

Michael B. A. Oldstone

Department of Immunology and Microbial Science, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Rino Rappuoli

Novartis Vaccines, Via Fiorentina 1, Siena 53100, Italy

Peter K. Vogt

Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, BCC-239, La Jolla, CA 92037, USA

Honorary Editor: Hilary Koprowski (deceased)

Biotechnology Foundation, Inc., Ardmore PA, USA

For further volumes:

<http://www.springer.com/series/82>

Jean Pieters · John D. McKinney  
Editors

Pathogenesis of  
*Mycobacterium tuberculosis*  
and its Interaction with the  
Host Organism

Responsible Series Editor: Fritz Melchers

 Springer

*Editors*

Jean Pieters  
Biozentrum  
University of Basel  
Basel  
Switzerland

John D. McKinney  
Laboratory of Microbiology and  
Microsystems  
Swiss Federal Institute of Technology  
Lausanne (EPFL)  
Lausanne  
Switzerland

ISSN 0070-217X

ISSN 2196-9965 (electronic)

ISBN 978-3-642-40231-9

ISBN 978-3-642-40232-6 (eBook)

DOI 10.1007/978-3-642-40232-6

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013949198

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

Tuberculosis has been with mankind ever since the dawn of civilization, with evidence of human infection as early as 9,000 years ago. Although the causative agent, *Mycobacterium tuberculosis*, has been identified more than a century ago, tuberculosis remains one of the most prevalent diseases on the globe, with currently one third of the population being latently infected with *M. tuberculosis*. Despite concerted efforts from scientists with diverse backgrounds, including medical doctors, microbiologists, cell biologists and immunologists, global control of tuberculosis continues to be highly challenging. However, research over the past few years has highlighted several issues that may contribute to overcoming some of the difficulties of eradicating *M. tuberculosis*. In this volume, a number of experts discuss various topics related to the virulence of *M. tuberculosis*.

*Mycobacterium tuberculosis* is an obligate intracellular pathogen that has gained the capacity to survive within those cells of the immune system that are in fact designed to eradicate microbial pathogens, the macrophages. As a result, pathogenic mycobacteria are able to cope with a plethora of antimicrobial activities, against which these bacilli have evolved numerous mechanisms of resistance. In addition, one emerging problem is the appearance of strains of *M. tuberculosis* that show various degrees of drug resistance. This problem has become of increasing concern also because of the recent discovery of extremely drug-resistant (XDR) as well as totally drug resistant (TDR) strains of *M. tuberculosis*. The emergence of these highly problematic strains may be a result of the slow generation time of *Mycobacterium spp.*, their exposure to a battery of drugs for prolonged time periods, the sometimes poor compliance of patients with anti-tuberculosis therapy or most likely a combination of these factors. One realization however has been that there is a need for a better understanding of the genetic diversity in different strains of *M. tuberculosis*, as outlined by Gagneux (Chap. 1). Also, recent work suggests that *M. tuberculosis*, unlike many other bacterial pathogens, relies heavily on chromosomal mutagenesis to drive its evolution within the human host. The intricate balance between maintenance of chromosome integrity and evolutionary progress is discussed in detail by Warner and colleagues in Chap. 2.

The emerging problem of multi-drug resistance requires a far better understanding of the mechanisms involved in drug resistance, and Smith and colleagues (Chap. 3) review the current understanding of the molecular basis of the

development of resistance mechanisms. Furthermore, despite the fact that *M. tuberculosis* is extensively exposed to diverse antimycobacterial drugs as well as host antimicrobials, the bacilli can survive exposure to many of these compounds through the upregulation of antimicrobial efflux pumps. In [Chap. 4](#), Szumowski and colleagues propose that drug tolerance has mainly evolved as a result of the need of intracellularly residing mycobacteria to employ efflux pumps to protect against environmental toxins.

To successfully survive under challenging conditions, *M. tuberculosis* not only needs to be able to fight off attacks from toxic components through efflux pumps and other drug resistance mechanisms, but it also requires a diverse array of secretion systems in order to communicate with and modulate its environment, the macrophage. Van der Woude and colleagues ([Chap. 5](#)) review the different secretion systems that are employed by *M. tuberculosis*, which have recently attracted much interest as determining factors for intracellular survival.

*Mycobacterium tuberculosis* is notoriously able to survive in a so-called “dormant” or non-growing state within host macrophages. The capacity to survive without growth also creates challenges in terms of finding anti-mycobacterial drugs that are able to interfere with survival of non-proliferating bacteria rather than controlling growth per se. Manina and McKinney ([Chap. 6](#)) highlight the importance of analyzing bacteria that are non-growing but metabolically active, knowledge that might be important for designing better strategies to eliminate non-replicating mycobacteria. Metabolism becomes also important when considering the natural niche of *M. tuberculosis*, the macrophage phagosome. In [Chap. 7](#), Ehrh and Rhee discuss the importance of the metabolic network of intracellularly residing *M. tuberculosis* for survival within its host.

Once within macrophage phagosomes, *M. tuberculosis* can rely on several distinct mechanisms to subvert the antimicrobial activities of the macrophages. Many of these mechanisms allow *M. tuberculosis* to prevent intraphagosomal destruction, and in [Chap. 8](#) Jayachandran and colleagues discuss work that has deciphered several of these mechanisms. Besides intraphagosomal survival, *M. tuberculosis* can also extensively interact with other host components, and Stanley and Cox ([Chap. 9](#)) review recent work that has added several layers of complexity to the interaction of *M. tuberculosis* with its host.

Together, these Chapters provide a topical overview on the current understanding of diverse mechanisms that are involved in the virulence of *M. tuberculosis*, ranging from their genetic, metabolic and molecular makeup as well as the exquisite strategies these bacteria utilize to circumvent host innate immune responses. The work described in this volume may therefore provide a stimulus for further exploration of the intricate biology of *M. tuberculosis* and its interaction with their hosts.

Jean Pieters  
John McKinney

# Contents

<b>Genetic Diversity in <i>Mycobacterium tuberculosis</i></b> . . . . .	1
Sebastien Gagneux	
<b>DNA Metabolism in Mycobacterial Pathogenesis</b> . . . . .	27
Digby F. Warner, Tone Tønjum and Valerie Mizrahi	
<b>Molecular Biology of Drug Resistance in <i>Mycobacterium tuberculosis</i></b> . . . . .	53
Tasha Smith, Kerstin A. Wolff and Liem Nguyen	
<b>Antimicrobial Efflux Pumps and <i>Mycobacterium tuberculosis</i> Drug Tolerance: Evolutionary Considerations</b> . . . . .	81
John D. Szumowski, Kristin N. Adams, Paul H. Edelstein and Lalita Ramakrishnan	
<b>Getting Across the Cell Envelope: Mycobacterial Protein Secretion.</b> . . . . .	109
Aniek D. van der Woude, Joen Luirink and Wilbert Bitter	
<b>A Single-Cell Perspective on Non-Growing but Metabolically Active (NGMA) Bacteria</b> . . . . .	135
Giulia Manina and John D. McKinney	
<b><i>Mycobacterium tuberculosis</i> Metabolism and Host Interaction: Mysteries and Paradoxes</b> . . . . .	163
Sabine Ehrt and Kyu Rhee	

**Surviving the Macrophage: Tools and Tricks Employed  
by *Mycobacterium tuberculosis* . . . . . 189**  
Rajesh Jayachandran, Somdeb BoseDasgupta and Jean Pieters

**Host–Pathogen Interactions During *Mycobacterium tuberculosis*  
Infections . . . . . 211**  
Sarah A. Stanley and Jeffery S. Cox

**Index . . . . . 243**

# Genetic Diversity in *Mycobacterium tuberculosis*

Sebastien Gagneux

**Abstract** Recent years have witnessed an increased appreciation of the extent and relevance of strain-to-strain variation in *Mycobacterium tuberculosis*. This paradigm shift can largely be attributed to an improved understanding of the global population structure of this organism, and to the realisation that the various members of the *M. tuberculosis* complex (MTBC) harbour more genetic diversity than previously realised. Moreover, many studies using experimental models of infection have demonstrated that MTBC diversity translates into significant differences in immunogenicity and virulence. However, linking these experimental phenotypes to relevant clinical phenotypes has been difficult, and to date, largely unsuccessful. Nevertheless, emerging high-throughput technologies, in particular next-generation sequencing, offer new opportunities, and have already lead to important new insights. Given the complexity of the host-pathogen interaction in tuberculosis, systems approaches will be key to define the role of MTBC diversity in the fight against one of humankind's most important pathogens.

## Contents

1	Introduction.....	2
2	Genomic Diversity Among Human-Associated MTBC .....	3
2.1	The Current Definition of MTBC.....	3
2.2	Current Genotyping Methodologies for MTBC .....	4
2.3	From Genotypes to Whole Genomes.....	6
2.4	What is the Role of HGT and Recombination in MTBC?.....	9
2.5	Using DNA Sequence Data to Predict the Impact of Genome Diversity .....	10
2.6	From MTBC Genotype to Phenotype.....	12

---

S. Gagneux (✉)  
Swiss Tropical and Public Health Institute (Swiss TPH),  
University of Basel, Basel, Switzerland  
e-mail: Sebastien.Gagneux@unibas.ch

3	Micro-Evolution of Drug-Resistant MTBC.....	13
3.1	Compensatory Evolution.....	14
3.2	The Role of the Strain Genetic Background.....	15
4	Conclusions.....	16
	References.....	17

## 1 Introduction

The outcome of tuberculosis (TB) infection and disease is extraordinarily diverse, ranging from lifelong asymptomatic infection to pulmonary TB and disseminated disease. In the past, this variable outcome has largely been attributed to host and environmental factors, but there is mounting evidence suggesting variation in the causative agent might also play a role. Indeed, the fact that this book, which is primarily focusing on the pathogenesis and host-pathogen interaction in TB, includes a chapter on “genetic diversity in *M. tuberculosis*” is a testimony to the recent paradigm shift with respect to the importance of studying strain variation (Comas and Gagneux 2009). Early sequencing studies revealed little DNA diversity across clinical isolates (Musser et al. 2000; Sreevatsan et al. 1997) leading to the dogma that strain variation in *M. tuberculosis* was “negligible” and therefore clinically irrelevant. Even though genetic diversity in *M. tuberculosis* is limited compared to other bacteria (Achtman 2008), this does not necessarily mean that the existing variability, as minor it may be, will not translate into any relevant phenotypes. Already in the 1960s, Mitchison and colleagues observed differences in virulence in clinical strains of *M. tuberculosis* when infecting guinea-pigs (Mitchison et al. 1960). At the time, no molecular tools were available to classify these strains into phylogenetically meaningful groupings. When the first genotyping tools became available in the early 1990s (van Embden et al. 1993), epidemiologist noticed very soon that *M. tuberculosis* strains also differed in their propensity to spread between individuals, and that some of these strains were associated with prolonged outbreaks (Rajakumar et al. 2004; Valway et al. 1998; Zhang et al. 1999). In some cases, experimental studies identified molecular features that might contribute to the success of outbreak strains (Newton et al. 2006; Reed et al. 2004). However, the relevance of these characteristics relative to non-bacterial factors remains unclear.

Over the last 10 years, several reviews have been published on the extent and relevance of strain diversity in *M. tuberculosis* (Gagneux and Small 2007; Kato-Maeda et al. 2001a; Malik and Godfrey-Faussett 2005; Nicol and Wilkinson 2008; Parwati et al. 2010). One of the most comprehensive articles published recently reviewed 100 papers on the topic, and concluded that even though the phenotype of clinical strains of *M. tuberculosis* undoubtedly differs significantly in vitro and in animal models of infection, if and how these differences are reflected in clinical settings is less clear (Coscolla and Gagneux 2010). Two years

later, this view largely still holds. Nevertheless, recent technological advances, in particular large-scale DNA sequencing and other-omics platforms, are now able to approach the subject at an unprecedented scale, with the potential to generate important novel insights. Given the biological complexities in TB, it is too simplistic to think that variable strain behaviour will be the product of just one or few genomic differences. It is much more likely that the various genomic features of a given strain will have combined effects. Such epistatic interactions will be difficult to detect and link to relevant phenotypes, but systems biology approaches might offer a way forward (Comas and Gagneux 2011).

This chapter summarises some of the recent advances in our understanding of the nature and consequence of strain variation in *M. tuberculosis*, with a special focus on comparative whole-genome sequencing. I will start by reviewing the current definition of the *M. tuberculosis* complex (MTBC), and briefly discuss the main genotyping methods for MTBC as well as their limitations. I will then review our current understanding of the global phylogeography of human-associated MTBC based on the available whole-genome data, discuss some of the new forms of genomic variation that have become evident through comparative genome sequencing, and review recent examples where large-scale DNA sequence data was used to predict putative phenotypic effects. Moving from the genome diversity to phenotypic variation, I will review some recent insights into strain-specific transcriptomes and strain-specific differences in innate immune recognition. Finally, I will discuss recent findings on the microevolution of drug-resistant MTBC, and end with a few thoughts on possible future research directions.

## 2 Genomic Diversity Among Human-Associated MTBC

### 2.1 The Current Definition of MTBC

MTBC consists of several closely related species and sub-species of acid-fast bacteria also referred to as ‘ecotypes’ (Smith et al. 2005). Even though these ecotypes share identical 16S ribosomal RNA sequences and 99.9 % nucleotide identity at the whole-genome level (with the exception of *Mycobacterium canettii* and the other so-called “smooth tubercle bacilli” further discussed below), they appear to be adapted to different host species. *M. tuberculosis sensu stricto* and *Mycobacterium africanum* are the main agents of TB in humans. *M. africanum* is limited to West Africa for reasons that are unknown, but causes up to 50 % of human TB in parts of that region (de Jong et al. 2010). In addition to these human pathogens, several animal-associated members of MTBC are found in various domestic and wild animals. These include *Mycobacterium bovis*, the agent of bovine TB (Garnier et al. 2003), *Mycobacterium caprae* (sheep and goats) (Niemann et al. 2002), *Mycobacterium microti* (voles) (Frota et al. 2004), *Mycobacterium pinnipedii* (seals and sea lions) (Cousins et al. 2003), *Mycobacterium mungi* (mangoose) (Alexander et al. 2010),

*Mycobacterium orygis* (antelope) (van Ingen et al. 2012) and the “dassie bacillus” (rock hyrax) (Mostowy et al. 2004a).

*M. canettii* and the other smooth TB bacilli (also referred to as “*Mycobacterium prototuberculosis*”) are regarded by some as human-adapted pathogens (Gutierrez et al. 2005), but several factors point to a likely environmental reservoir for these microbes (Koeck et al. 2010). First, only about 60 isolates have been described in the literature since the original discovery of *M. canettii* in 1969 (van Soolingen et al. 1997). Second, the majority of these organisms have been isolated from patients in Djibouti (Fabre et al. 2010). Third, no human-to-human transmission has been documented thus far (Koeck et al. 2010). And forth, these bacteria show clear evidence of extensive ongoing horizontal gene transfer (HGT) (Gutierrez et al. 2005), which stands in contrast to the other members of MTBC (the role of HGT and recombination in MTBC is further discussed below).

For the remainder of this chapter, I will focus on *M. tuberculosis* sensu stricto and *M. africanum*, which are the most important agents of human TB. Based on the various strain genotyping techniques, *M. tuberculosis* and *M. africanum* have been further subdivided into strain lineages and families. For example, the “Beijing” family of strains was originally defined based on characteristic IS6110 RFLP patterns (van Soolingen et al. 1995). This strain family has gained considerable attention because of its association with drug resistance (reviewed in Borrell and Gagneux 2009) and hypervirulence in some experimental models (reviewed in Parwati et al. 2010). Other human-associated MTBC lineages have been studied less, partially because of the lack of a standardised and phylogenetically robust classification system and associated nomenclature (Gagneux and Small 2007). Ultimately, such a system should be based on whole-genome sequencing of a large and representative collection of MTBC clinical strains. However, until genome-sequencing becomes more readily available, the different genotyping methodologies and the corresponding nomenclatures are likely to continue to be used in parallel (for a comparison of the various nomenclatures for human-associated MTBC refer to Coscolla and Gagneux 2010). Let me thus briefly review these different genotyping techniques, and discuss some of their advantages and disadvantages.

## 2.2 Current Genotyping Methodologies for MTBC

Strain typing in MTBC is usually performed for one of three main purposes: (i) “classical” molecular epidemiology, (ii) phylogenetic and evolutionary studies and (iii) strain classification. However, the existing genotyping techniques are not equally suited for all of these applications. Classical molecular epidemiological studies in TB usually involve measuring ongoing transmission, differentiating between relapse and re-infection, or detecting laboratory cross-contamination (Kato-Maeda et al. 2011). Such studies require highly discriminatory genotyping

tools such as IS6110 RFLP analysis (van Embden et al. 1993) or MIRU-VNTR typing (Supply et al. 2006). These techniques rely on mobile and repetitive DNA elements, respectively, which exhibit a rapid rate of change (i.e. a fast “molecular clock”), providing a high discriminatory power for differentiating between patient isolates. However, whilst useful for genotyping closely related strains, these high rates of change lead to convergent evolution and homoplasmy (i.e., emergence of identical patterns in phylogenetically unrelated strains), which complicates phylogenetic inference and strain classification (Comas et al. 2009). Spoligotyping is another genotyping technique which has been widely used for both molecular epidemiology and evolutionary studies. It usually shows a lower discriminatory power than IS6110 RFLP or MIRU-VNTR typing, and is therefore often used as a complementary genotyping method in epidemiological studies (Kamerbeek et al. 1997). Spoligotyping is based on the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) region, also known as the Direct Repeat (DR) region of MTBC. Unique so-called spacer sequences are interspersed between repetitive sequences and are variably present or absent in a given MTBC isolate. Interrogation of 43 of these unique spacer sequences results in a strain-specific fingerprint which is easily digitalised and compared across laboratories. Several large international databases have been compiled containing spoligotypes of thousands of clinical isolates from many countries (Allix-Beguec et al. 2008; Demay et al. 2012). Based on some characteristic spoligopatterns, several MTBC lineages and families have been defined. For example, the Beijing family shows a typical spoligopattern in which the first 34 unique spacers are deleted (Bifani et al. 2002). Unfortunately, spoligotyping patterns, too, are prone to homoplasmy as individual spacers can be deleted independently in phylogenetically unrelated strains (Comas et al. 2009). For example, a recent study reported clinical strains exhibiting a “pseudo-Beijing” spoligotype, referring to the fact that these strains showed a spoligotype pattern characteristic of Beijing strains but belonged to a different phylogenetic lineage (Fenner et al. 2011). Moreover, some spoligotyping patterns do not exhibit any of the recognised signatures and can therefore not be classified into any of the known strain families (Flores et al. 2007).

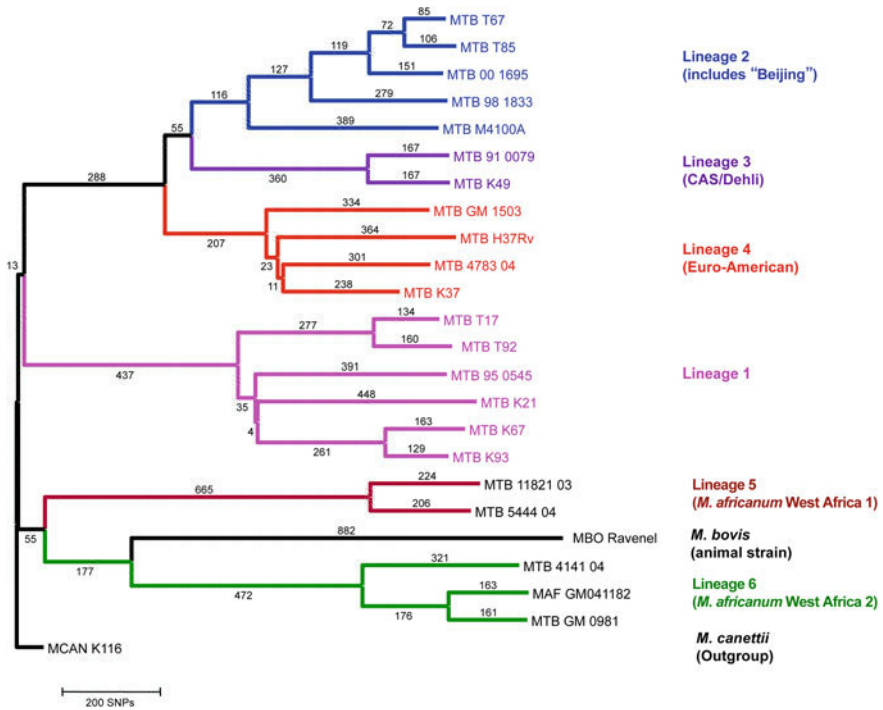
Two additional genotyping approaches have been developed primarily for phylogenetic studies and strain classification in MTBC; one is based on genomic deletions (also referred to as large sequence polymorphisms (LSPs) or regions of difference (RDs)), and the other on single nucleotide polymorphisms (SNPs). Starting about 12 years ago, scientists have used bacterial artificial chromosome libraries and DNA microarrays to interrogate the genome of MTBC strains (Behr et al. 1999; Gordon et al. 1999; Kato-Maeda et al. 2001b; Tsolaki et al. 2004). These studies identified many genomic regions that were absent compared to the H37Rv reference genome, which at the time was the only whole-genome sequence of MTBC available (Cole et al. 1998). Many of these genomic deletions can be used as robust phylogenetic markers for assigning MTBC strain into meaningful groupings. This is because due to the virtual absence of large-scale HGT (further discussed below), once a particular genomic region is lost in MTBC, this region

cannot be reacquired, and all progeny of the strain experiencing the loss initially will inherit this deletion. The observation that phylogenetic trees constructed using such genomic deletion markers showed no homoplasy was per se strong support for the already established notion that MTBC exhibits no ongoing HGT and a highly clonal population structure. Several groups used such an approach to revisit the evolutionary history of MTBC and define sets of discrete strain lineages (Brosch et al. 2002; Gagneux et al. 2006b; Hirsh et al. 2004; Mostowy et al. 2002, 2004b; Tsolaki et al. 2005). From a strict phylogenetic point of view; however, these evolutionary scenarios do not represent actual phylogenetic trees but mere cladograms, as their branches do not reflect actual evolutionary distances (i.e., genomic deletions are stochastic events for which no evolutionary models are available). Moreover, phylogenetic inference using genomic deletions is limited by the fact that these markers are usually defined through a one-way comparison with the H37Rv reference genome.

SNP typing is rapidly gaining importance for MTBC genotyping, primarily because of available technologies originally developed for other organisms. It is important to note however that, similarly to genomic deletion analysis, SNP typing does not provide the necessary resolution to be applicable in classical molecular epidemiological studies of TB transmission (Kato-Maeda et al. 2011). Nevertheless, SNP typing is ideal for classifying MTBC strains into lineages as it exhibits negligible levels of homoplasy (Comas et al. 2009). Moreover, various typing methodologies have already been proposed (Dos Vultos et al. 2008; Filliol et al. 2006; Gutacker et al. 2006; Mestre et al. 2011). Importantly, in contrast to de novo SNP discovery through DNA sequencing (Baker et al. 2004; Hershberg et al. 2008), mere SNP typing is of limited use in phylogenetic studies because of the problem known as ‘phylogenetic discovery bias’ (Pearson et al. 2004). This bias refers to the fact that if only a few reference genomes are used to identify the SNPs used for subsequent genotyping, the resulting phylogeny will be biased, because the genetic diversity among strains not included in the initial SNP discovery will not be detected. As a consequence, such strains will automatically fall in intermediate positions on the phylogenetic tree, leading to the problems known as ‘branch collapse’ and ‘linear phylogeny’ (Achtman 2008; Alland et al. 2003; Smith et al. 2009). Recently, two new complementary SNP-typing assays have been developed to screen for the main lineages of MTBC (Stucki et al. 2012). In contrast to most other methods available, the SNPs used in these new assays have been identified by comparing many genomes representative of the known global MTBC diversity (Fig. 1).

### ***2.3 From Genotypes to Whole Genomes***

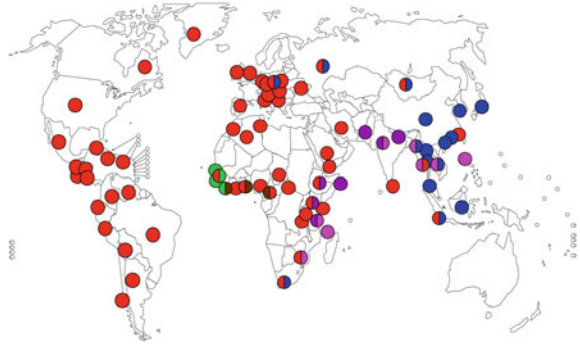
Whole-genome sequencing offers many advantages over the current methods used for genotyping MTBC. The main disadvantage remains the relatively high cost and the bioinformatics capacity required to analyse the data. However, both of these



**Fig. 1** Global phylogeny of MTBC based on 24 whole-genome sequences. *M. canettii* is used as outgroup. Coloured branches indicate the six main human-associated lineages. Numbers on branches indicate the number of SNPs (adapted from Bentley et al. 2012; Comas et al. 2010)

obstacles will likely be surmounted in the near future thanks to further advances in DNA sequencing technologies and the development of more user-friendly analytical tools. Several recent publications support the notion that genome sequencing is already on its way to become the new gold standard for molecular epidemiological studies of MTBC. Specifically, these studies have shown that genome sequencing has a much higher discriminatory power than the standard genotyping tools, and that clinical strains exhibiting identical genotyping profiles can sometimes harbour extensive genetic diversity, with obvious implications for the interpretation of transmission patterns (Casali et al. 2012; Gardy et al. 2011; Niemann et al. 2009). Comparative genome sequencing has also been used to study closely related strains isolated sequentially from a single patient (Comas et al. 2012; Saunders et al. 2011), from chains of ongoing transmission (Sandegren et al. 2011; Schurch et al. 2009, 2010), or from experimentally infected macaques (Ford et al. 2011), in the hope of learning more about the mechanisms driving the micro-evolution of MTBC. Intriguingly, the study in macaques found that the mutation rate of MTBC in latently infected animals was not significantly different from that of bacteria isolated from macaques with active TB. This suggests that

**Fig. 2** Global phylogeography of human-associated MTBC. The geographic distribution of the six main human-associated MTBC lineages is indicated. Each dot corresponds to a country and the dominant MTBC lineage(s) are indicated by colours corresponding to Fig. 1 (adapted from Gagneux et al. 2006b)



dormant and/or persister cells are also prone to mutational damage, and therefore drug resistance could also be acquired during latent infections (Ford et al. 2011).

With respect to the global diversity of MTBC, comparative genome sequencing is revealing a degree of heterogeneity which was previously unrecognised. Specifically, because DNA sequence data (as opposed to e.g. genome deletions) can be used to infer genetic distances, this heterogeneity can be measured quantitatively and appreciated in the context of a robust phylogenetic framework. Based on recently published genome sequences (Bentley et al. 2012; Comas et al. 2010), six main phylogenetic lineages can be distinguished among the human-associated MTBC (in addition to all the animal-associated variants not further discussed here; Fig. 1). This includes two lineages traditionally referred to as *M. africanum* (de Jong et al. 2010). These six lineages have already been described using various genotyping methods and shown to be distributed non-randomly around the world (Fig. 2) (Brudey et al. 2006; Gagneux et al. 2006b). However, as mentioned above, DNA sequence data can be used to compute evolutionary distances and construct robust phylogenetic trees. As shown in Fig. 1, the genetic distance between two human-associated MTBC strains can reach close to 2,000 SNPs, which is equivalent to the evolutionary distance between H37Rv and *M. bovis* (Garnier et al. 2003). Moreover, about two-thirds of amino acid coding SNPs in MTBC are non-synonymous, and a large proportion of these nSNPs have been predicted to affect gene function (Hershberg et al. 2008). Many more genomes will be necessary to better define the global diversity of MTBC (Fig. 2). In particular, based on the currently available genomes, geographic coverage is limited, and potential phylogenetic substructures within the main lineages are not well understood (Fig. 1). Moreover, we know from genotyping studies that much more MTBC diversity exists that has not yet been characterised by whole-genome sequencing (Demay et al. 2012; Gagneux et al. 2006b).

In addition to genomic deletions and functional SNPs, recent data suggest that gene duplications might also be an important source of genome plasticity in MTBC. In 2000, Brosch et al. reported, two duplicated regions in *M. bovis* BCG Pasteur (Brosch et al. 2000). Until recently, genomic duplications were considered rare events limited to a few BCG sub-strains (Leung et al. 2008), possibly

reflecting the result of long-term in vitro evolution (Brosch et al. 2007). However, two recent studies reported gene duplications in clinical isolates of *M. tuberculosis*, including one large-scale duplication of more than 500 Kb (Domenech et al. 2010; Weiner et al. 2012). Intriguingly, these duplications occurred multiple times independently and across different MTBC lineages, suggesting that they were positively selected. Moreover, some of these duplications are unstable when strains are grown in vitro, arguing against a phenomenon linked to prolonged in vitro growth, but rather in support of in vivo selection and adaptation. Future studies will help define the extent and significance of genomic duplications in MTBC.

## ***2.4 What is the Role of HGT and Recombination in MTBC?***

There is strong evidence that past instances of HGT have contributed to the emergence of MTBC as a successful pathogen (Behr and Gagneux 2011). Specifically, genomic comparisons of MTBC and other mycobacteria have identified many genes that the common ancestor of all MTBC appears to have acquired horizontally (Becq et al. 2007; Rosas-Magallanes et al. 2006; Veyrier et al. 2009). Some of these genes have been implicated in the pathogenesis of TB. Similarly, analysis of housekeeping genes in *M. canettii* and other smooth TB bacilli demonstrated mosaic structures, suggesting ongoing HGT among these rare and extant members of MTBC (Gutierrez et al. 2005). However, whether ongoing (as opposed to past) HGT plays any role in the evolution of MTBC is controversial (except in the case of *M. canettii* and the other smooth TB bacilli discussed above). For a long time, the common view was that MTBC exhibited a strictly clonal population structure with little evidence for ongoing HGT. With a few exceptions largely considered anecdotal (Hughes et al. 2002; Liu et al. 2006), most of the available evidence supported this view until recently. This included the observation of a rather stable G + C content across most of the MTBC genome (Cole et al. 1998), strong linkage disequilibrium between minisatellite loci (Supply et al. 2003), congruence of phylogenies derived from many different molecular markers including genomic deletions (Baker et al. 2004; Filliol et al. 2006; Gagneux et al. 2006b; Gagneux and Small 2007; Gutacker et al. 2006; Hirsh et al. 2004; Wirth et al. 2008), and negligible homoplasy in DNA sequence data (Comas et al. 2009; Hershberg et al. 2008). In particular, the fact that genomic deletions in MTBC can be used as robust phylogenetic markers has been demonstrated multiple times (Brosch et al. 2002; Hirsh et al. 2004; Mostowy et al. 2002), and suggests that once a particular genomic region has been lost in MTBC, the corresponding genes cannot be reacquired via HGT. Moreover, all known drug resistance determinants in MTBC represent *de novo* acquired chromosomal mutations (Sandgren et al. 2009), indicating that HGT plays no significant role in the emergence of drug resistance in this organism.

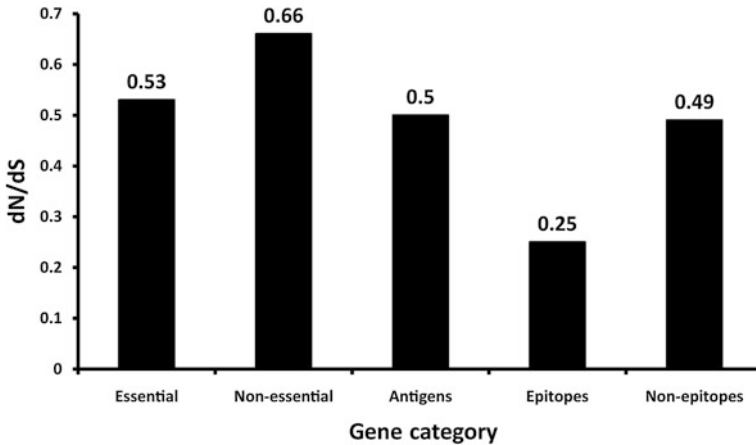
A recent study has now challenged the dogma of strict clonality in MTBC (Namouchi et al. 2012). The authors used 24 MTBC genomes and performed a

series of analyses designed to detect ongoing HGT. Based on their results, the authors concluded that MTBC does in fact show evidence of recombination, mostly involving short DNA fragments of around 50 bp. This observation is striking, given that in other bacteria, recombination usually involves much longer stretches of DNA (Didelot et al. 2010). This short size of recombination tracts might explain why the genomic deletions used to genotype MTBC, which comprise much larger DNA segments, show no evidence of homoplasmy linked to HGT or recombination. The authors found that 150 kb (i.e. about 3 %) of the MTBC genomes showed significant recombination tracts. However, the nature of the putative donor organisms could not be determined, except for some instances where the recombining tracts matched regions in the genome of *M. canettii*. More work is needed, including larger collections of high-quality genomes, to define the role and relevance of HGT and recombination in MTBC, and to identify the sources and mechanisms of acquisition of foreign DNA in this organism.

## ***2.5 Using DNA Sequence Data to Predict the Impact of Genome Diversity***

DNA sequences can also be used to analyse the relative strength and direction of selection acting on these sequences. For example, a recent study used whole-genome sequences from 21 clinical strains covering the global diversity of MTBC to study the genetic diversity of 491 experimentally confirmed human T cell epitopes (Comas et al. 2010). In contrast to other pathogens where immune pressure drives increased antigenic diversity (Deitsch et al. 2009), the authors found that in MTBC, T cell epitopes were evolutionarily hyperconserved, showing the lowest dN/dS values (i.e. ratio of non-synonymous to synonymous mutations) in the genome (Fig. 3), with more than 95 % of these epitopes harbouring no amino acid substitution at all. The authors concluded from their findings that the host immune responses directed towards these hyperconserved T cell epitopes might offer a net benefit to the bacteria rather than to the host, by promoting tissue damage, ultimately leading to enhanced transmission of MTBC. This notion is supported by the fact that patients with cavitary disease are particularly likely to generate secondary cases (Rodrigo et al. 1997). Also, in TB patients co-infected with HIV, CD4 T cell counts are inversely correlated with the likelihood of developing cavitations (Kwan and Ernst 2011). The latter suggests that CD4 T cells directly or indirectly contribute to the formation of cavitations, and that TB patients with low CD4 T cell counts might be less likely to transmit MTBC (Brites and Gagneux 2011; Cruciani et al. 2001).

The observation that known human T cell epitopes in MTBC are evolutionarily hyperconserved does of course not exclude the possibility that other, as yet unknown epitopes might be involved in antigenic variation. Indeed, one of the limitations in the study by Comas et al. was that due to technical reasons, the



**Fig. 3** Human T cell epitopes of MTBC are evolutionarily hyperconserved. The ratios of non-synonymous to synonymous nucleotide substitutions are indicated for each gene category (adapted from Comas et al. 2010)

members of the PE/PPE gene family had to be excluded from the analysis (Brennan and Delogu 2002). Some of these genes have previously been shown to be highly diverse (Talarico et al. 2005, 2008), prompting the view they may be implicated in antigenic variation (Banu et al. 2002). A recent study addressed this possibility by analysing the genetic diversity of various PE/PPE genes across a panel of 40 phylogenetically diverse clinical strains (McEvoy et al. 2012). The authors confirmed that PE/PPE genes were more genetically diverse than other genes, but the observed dN/dS values were in average around 1.0, suggesting these genes evolve neutrally. The latter finding argues against the notion that PE/PPE diversity is a result of immune escape. One of the inherent problems in studying PE/PPE genes is that hardly any information exists as to their specific function(s). Future studies will show whether or not the findings by McEvoy et al. can be extrapolated to other PE/PPE genes.

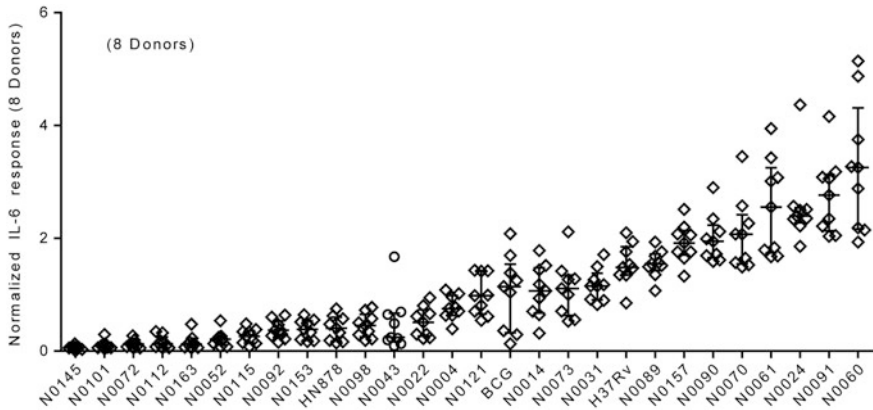
One of the striking findings of comparative DNA sequencing in MTBC has been the observation that in average about two-thirds of coding SNPs are non-synonymous, thus leading to a change in the encoded amino acid (Comas et al. 2010; Fleischmann et al. 2002). Moreover, in silico analyses have shown that 58 % of nonsynonymous SNPs in MTBC are predicted to affect gene function (Hershberg et al. 2008). Recent data from our laboratory using whole-genome sequences have confirmed that even in the deep phylogenetic branches that define the main lineages of MTBC (Fig. 1), close to 50 % of non-synonymous SNPs are predicted to be functional (unpublished data). Taken together, these data highlight the extent of genetic diversity among MTBC strains and lineages, and suggest that this diversity is likely to have phenotypic consequences.

## 2.6 From MTBC Genotype to Phenotype

Some evidence that genetic diversity in MTBC does indeed translate into phenotypic diversity comes from studies of gene expression. One early study used DNA microarrays to study the transcriptome of 10 clinical strains and found about 500 genes differentially expressed (Gao et al. 2005). More recently, Homolka et al. studied a panel of 17 MTBC strains covering several of the main phylogenetic lineages of MTBC (Fig. 1) using DNA microarrays, and determined their transcriptional profile when grown *in vitro* and in resting or activated mouse macrophages (Homolka et al. 2010b). The authors detected both strain-specific and lineage-specific gene expression patterns. These patterns were consistent with the phylogenetic position of the corresponding strains, supporting the association between phylogenetic diversity discussed above and phenotypic variation observed experimentally. Moreover, the authors were able to define a core transcriptome that was common to all strains when grown intracellularly. Considering recent technological advances, including the development of RNA sequencing (RNA-seq), future studies will likely uncover additional MTBC diversity at the level of the genome-wide transcriptome. RNA-seq offers many advantages over DNA microarrays, as it allows for detection of unknown transcripts, including novel regulatory RNAs in a strand-specific manner (Sorek and Cossart 2010). A recent RNA-seq analysis of H37Rv grown *in vitro* revealed many 5' and 3' untranslated regions, antisense transcripts and intergenic small RNAs (Arnvig et al. 2011), supporting a role for non-coding RNA in the biology of MTBC (Arnvig and Young 2012).

In addition to differences in gene expression, MTBC phenotypic diversity has also become evident at the level of the host-pathogen interaction, at least in experimental settings. Many studies have reported MTBC strain-specific differences in immunogenicity and virulence using various models of infection; these have been reviewed in detail in (Coscolla and Gagneux 2010). However, only a few studies have been able to link the observed immune- or virulence-phenotypes to particular molecular characteristics of the infecting MTBC strain (Manca et al. 1999; Newton et al. 2006; Reed et al. 2004). In a recent study, Portevin et al. managed to link strain-specific immune-phenotypes with the phylogenetic classification of these strains using a human monocyte-derived macrophage infection model (Portevin et al. 2011). Specifically, the authors infected macrophages from multiple donors with one of 28 strains of MTBC and measured the cytokine profiles at several time points. They found that these strains varied widely in their stimulation of inflammatory cytokines, suggesting that human macrophages perceive and react to these strains very differently (Fig. 4). Despite this wide range of immune responses, the authors detected statistically significant differences between the different MTBC lineages, suggesting that the phylogenetic diversity of MTBC is also reflected at the host-pathogen interface.

For a long time, one of the main challenges in the field has been linking strain phenotypes defined in the laboratory to relevant clinical phenotypes; so far no



**Fig. 4** IL-6 response in human monocyte-derived macrophages infected with 28 different MTBC strains. Each strain was tested in eight different human donors (adapted from Portevin et al. 2011)

consistent picture has emerged (Coscolla and Gagneux 2010). Perhaps one of the most promising avenues has been the study of *M. africanum* West Africa 2 (i.e. MTBC Lineage 6 in Fig. 1), which causes up to 50 % of TB in some countries of West African (de Jong et al. 2010). Recently, it was shown that Lineage 6 was attenuated in mice compared to H37Rv (Bold et al. 2012). Consistent with these laboratory findings, individuals infected with Lineage 6 strains were less likely to progress towards active TB compared to individuals infected with other strains, even though no difference was observed at the level of transmission (de Jong et al. 2008). Moreover, Lineage 6 has been associated with HIV co-infection in the Gambia, suggesting Lineage 6 might behave as an opportunistic pathogen (de Jong et al. 2005). No such association was observed in Ghana (Meyer et al. 2008), suggesting that even within a particular MTBC lineage, additional phylogeographic variation exists that might be clinically relevant.

### 3 Micro-Evolution of Drug-Resistant MTBC

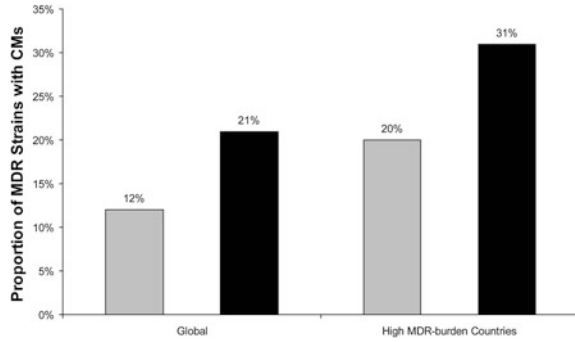
A particular aspect of genetic diversity in MTBC that is of obvious clinical relevance relates to drug resistance. MTBC strains resistant to an ever increasing number of antibiotics are threatening global TB control (Gandhi et al. 2010). Over the last years, many of the molecular mechanisms leading to resistance have been elucidated for many of the first- and second-line drugs (Zhang and Yew 2009). Similarly, many of the mutations causing resistance have been identified (Sandgren et al. 2009). What is much less understood is how drug-resistant MTBC strains evolve over the longer run (Borrell and Gagneux 2009). In particular, little is known on how different drug resistance-conferring mutations might interact

with each other, with pre-existing strain-specific mutations, or other mutations acquired subsequent to the resistance mutations (Borrell and Gagneux 2011). A particular class of mutations represented by the latter are referred to as compensatory mutations.

### 3.1 Compensatory Evolution

Acquisition of drug resistance determinants in bacteria is often associated with reduced Darwinian fitness in absence of the drug (Andersson and Levin 1999). However, this fitness cost can be mitigated by subsequent mutations at secondary sites, a phenomenon known as compensatory evolution. Compensation has been described in many bacterial species and antibiotics (reviewed in Andersson and Hughes 2010). In MTBC, a widely cited example has been mutations in the promoter of *ahpC* leading to over-expression of this gene, believed to compensate for the loss of the katalase-peroxidase (KatG) activity in isoniazid-resistant strains (Sherman et al. 1996). KatG activates isoniazid but also protects the bacterial cell against host-mediated oxidative stress. Over-expression of *ahpC*, which encodes an alkyl hydroperoxide reductase is thought to partially protect against excess oxidative damage in strains lacking KatG activity. However, the actual role of *ahpC* promoter mutations in isoniazid resistance remains unclear (Heym et al. 1997). Moreover, molecular epidemiological studies have found these mutations to be rare in clinical settings (Gagneux et al. 2006a), suggesting they play a minor role in the epidemiology of drug-resistant MTBC. Similarly, a compensatory mechanism in the 16S ribosomal RNA of MTBC strains resistant to aminoglycosides was reported (Shcherbakov et al. 2010). However, the corresponding mutation is only rarely observed in clinical settings (Georghiou et al. 2012). By contrast, Comas et al. described a set of novel compensatory mutations in *rpoA* and *rpoC*, which encode the alpha- and beta-prime subunits of the RNA polymerase (Comas et al. 2012). These mutations occurred exclusively in rifampicin-resistant strains with resistance-conferring mutations in *rpoB*, and were associated with an increased competitive fitness in vitro. Moreover, they occurred in up to 30 % of clinical MDR strains, and were significantly over-represented in MDR strains from countries known to suffer from a high burden of MDR-TB (Fig. 5). This suggests that these compensatory mutations contribute to the spread of drug-resistant MTBC strains. Subsequently, the same mutations were reported from XDR strains from Russia (Casali et al. 2012), and in experimentally evolved *Salmonella* (Brandis et al. 2012). While the mechanism(s) leading to compensation are still unknown, genetic reconstruction of some these *rpoA* and *rpoC* mutations in *Salmonella* showed that they were directly responsible for the improved growth rate of strains carrying rifampicin resistance-conferring mutations in *rpoB* (Brandis et al. 2012).

**Fig. 5** Compensatory evolution in rifampicin-resistant MTBC. The proportion of MDR strains carrying compensatory mutations (CMs) in *rpoA* or *rpoC* is indicated for strains isolated globally or from countries with a high MDR-TB burden. Light bars indicate mutations very likely to have a compensatory role; black bars indicate all putative compensatory mutations identified (adapted from Comas et al. 2012)



### 3.2 The Role of the Strain Genetic Background

The interaction between a drug resistance mutation and a compensatory mutation is an example of epistasis (Borrell and Gagneux 2011). Epistatic interaction can generally be defined as a case in which the phenotypic effect of one mutation is modified by the presence or absence of another mutation (Phillips 2008). In addition to epistatic interactions represented by compensatory evolution, interactions between different drug-resistance mutations have also been described. Importantly, work in *Escherichia coli* and *Pseudomonas aeruginosa* has shown that the fitness cost associated with a mutation conferring resistance to one drug can be ameliorated by a mutation conferring resistance to a second drug (Trindade et al. 2009; Ward et al. 2009). This phenomenon is referred to as sign epistasis, as the “sign” of the fitness effect associated with the second mutation switches from “negative” to “positive” (Weinreich et al. 2005). Currently, we do not know whether sign epistasis plays any role in the evolution of MDR-TB, but work is needed to explore this disturbing possibility.

There is increasing evidence that the strain genetic background can influence the evolutionary trajectory of drug-resistant MTBC (Borrell and Gagneux 2011). Presumably, this is due to epistatic interactions between pre-existing strain-specific genomic features, and subsequently acquired drug-resistance conferring mutations and compensatory mutations (Mueller et al. 2012). Experimental work in *E. coli* has shown that such interactions can limit the available mutational pathway(s) towards drug resistance (Toprak et al. 2012). In MTBC, one particular phylogenetic lineage known as “Beijing” has repeatedly been associated with drug resistance in clinical settings (reviewed in Borrell and Gagneux 2009). The underlying basis of this phenomenon remains unknown, but several hypotheses have been proposed (Parwati et al. 2010). For example, based on various

non-ysynonymous mutations found in DNA repair genes of Beijing strains, it was hypothesised that these strains might exhibit an intrinsically elevated mutation rate, leading to an enhanced likelihood of acquiring resistance mutations (Dos Vultos et al. 2008). However, fluctuation assays have yielded conflicting results with respect to the in vitro mutation rate of Beijing compared to other MTBC strains (de Steenwinkel et al. 2012; Werngren and Hoffner 2003).

Several studies have reported associations between particular drug resistance-conferring mutations and MTBC lineages, suggesting that the strain genetic background, as defined by MTBC lineage, can influence the mutational pathway to resistance (Gagneux et al. 2006a; Homolka et al. 2010a). For example in isoniazid-resistant MTBC, *inhA* promoter regions, which are one of the main causes of isoniazid resistance in clinical settings (Zhang and Yew 2009), have been associated with Lineage 1 of MTBC (Fig. 1). The fact that this association was observed in three independent studies supports a biological basis for this phenomenon (Baker et al. 2004; Fenner et al. 2012; Gagneux et al. 2006a). Moreover, Fenner et al. showed recently that putative epistatic interactions between specific isoniazid resistance-conferring mutations and different strain genetic backgrounds can influence the level of resistance to isoniazid in vitro (Fenner et al. 2012).

## 4 Conclusions

The past few years have witnessed much progress in our understanding and appreciation of strain-to-strain diversity in MTBC. This has partially been a consequence of new opportunities arising from access to novel technologies, including DNA microarrays, and more recently next-generation DNA sequencing (Comas and Gagneux 2009). Fresh insights have also been gained by learning from work in other bacteria. Fifteen years ago, human-associated MTBC was essentially regarded as a “clone”, and strain variation was generally considered irrelevant by the majority of the basic science community (Kato-Maeda et al. 2001a). Today, we know that MTBC harbours more genetic diversity than previously realised, and experimental studies show very clearly that part of this diversity translates into important phenotypic variation (Coscolla and Gagneux 2010). Yet, the role of this diversity for clinical TB remains largely elusive, partially because of the difficulty of linking MTBC strain genotypes to relevant clinical phenotypes. In addition to differences at the DNA level, MTBC strains have been shown to differ at the transcriptome level (Homolka et al. 2010b). Yet, nothing is known with respect to strain variation at the level of regulatory RNAs, which are highly abundant in this organism (Arnvig et al. 2011). Similarly, epigenetics is a rapidly emerging field that has recently been proposed as an avenue to discover host-derived biomarkers for TB (Esterhuysen et al. 2012). However, little work has been done on epigenetics in MTBC, and it is unknown whether strain variability will be reflected at the epigenetic level as well. Much effort today is dedicated to the development of new diagnostics, drugs and vaccines against TB (Young et al. 2008). Based on the

existing evidence from other bacterial pathogens, strain genetic diversity in MTBC should be considered during the development of new tools and strategies to better control TB (Gagneux and Small 2007). Increasingly, systems approaches should be used (Comas and Gagneux 2011), not only to determine if and how diversity in MTBC matters for disease control, but to better understand the biology and epidemiology of one of the most important human diseases.

**Acknowledgments** I thank Douglas Young, Iñaki Comas and all the members of my group for the stimulating discussions and comments on the manuscript. Work in my laboratory is supported by the Swiss National Science Foundation (grant number PP0033-119205), the National Institutes of Health (AI090928 and HHSN266200700022C) the Leverhulme-Royal Society Africa Award (AA080019) and the European Research Council (309540).

## References

- Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 62:53–70
- Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, Gey van Pittius NC (2010) Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis* 16:1296–1299
- Alland D, Whittam TS, Murray MB, Cave MD, Hazbon MH, Dix K, Kokoris M, Duesterhoeft A, Eisen JA, Fraser CM, Fleischmann RD (2003) Modelling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *J Bacteriol* 185:3392–3399
- Allix-Beguec C, Harmsen D, Weniger T, Supply P, Niemann S (2008) Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *J Clin Microbiol* 46:2692–2699
- Andersson DI, Hughes D (2010) Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol* 8:260–271
- Andersson DI, Levin BR (1999) The biological cost of antibiotic resistance. *Curr Opin Microbiol* 2:489–493
- Arnvig KB, Young D (2012) Non-coding RNA and its potential role in *Mycobacterium tuberculosis* pathogenesis. *RNA Biol* 9:427–436
- Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, Rose G, Perkins TT, Parkhill J, Dougan G, Young DB (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog* 7:e1002342
- Baker L, Maiden MC, Drobniowski F (2004) Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg Infect Dis* 10:1568–1577
- Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol* 44:9–19
- Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P (2007) Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* 24:1861–1871
- Behr M, Gagneux S (2011) The rise and fall of the *Mycobacterium tuberculosis* complex. In: Tibayrenc M (ed) *Genetics and evolution of infectious diseases*. Elsevier, London, pp 651–667