The health-disease equilibrium is shifted depending on many factors among which are the virulence of the circulating strain, the resistance/susceptibility of the host and many others still unknown.

‘And causes besets me, daily, invisible. And chance entangles me, powerful, invincible.’

*On Causes and Chance*
Silvio Rodriguez

*Sunset with Birds*
Ileana Mulet
Oil on canvas
CHAPTER 2C.1

INTERACTIONS BETWEEN MYCOBACTERIAL AND HUMAN GENETIC DIVERSITY AND IMPLICATIONS FOR NEW TB VACCINES

Sebastien Gagneux

The co-evolution of man and MTB has endowed MTB with multiple escape mechanisms, and man with complex defence systems.

‘I want to give thanks to the divine labyrinth of causes and effects for the diversity of the creatures that form this singular universe.’

Another Poem of Gifts
Jorge Luis Borges

Rebuilt the Babel Tower
Alfredo Sosabravo
Oil on canvas; 125 × 89.5 cm
TB IS AN ANCIENT HUMAN DISEASE

TB has been affecting humans for millennia (1). While paloepathological evidence suggests that TB might predate the ascend of our own species (2), analyses of ancient DNA indicate that Homo sapiens has been infected with members of the M. tuberculosis complex (MTBC) for at least 9,000 years (3). Large-scale analyses of contemporary DNA extracted from global strain collections suggest that MTBC and its human host have co-existed for a very long time (4–6). Contrary to previous belief, human TB did not emerge as a consequence of animal domestication (7, 8). Instead, human TB appears to have originated as a human disease in Africa, and has since then spread to the rest of the world following human migrations (4, 6, 9). This notion is supported by the fact that Africa harbours the largest diversity of human-associated MTBC (Chapter 2A.4, Figure 2A.4.2) (6). Moreover, whole-genome sequencing of 21 representative MTBC strains (10) has shown that the most ancestral MTBC lineages (Figure 2C.1.1), also known as Mycobacterium africanum (11), are almost exclusively found in West-Africa. Moreover, the ‘Out-of-Africa’ scenario for the evolution of human TB is further supported by multilocus sequence data (4), and the observation that Mycobacterium canettii and other so-called “smooth mycobacteria”, which are believed to be related to the most recent common ancestor of all MTBC, only occur in the Horn of Africa (9).

HOST-PATHOGEN CO-EVOLUTION IN HUMAN TB?

The long-standing association between MTBC and humans raises the question as to whether or not co-evolution has occurred (12). Host pathogen co-evolution can be defined as “reciprocal adaptive changes in interacting host and pathogen species” (13). Co-evolution is difficult to demonstrate formally, particularly in infectious diseases of humans (14), but several examples have been described in invertebrate systems (13). A related concept in the ecology of host-pathogen interactions is known as “local adaptation”, which can be used as indirect evidence for co-evolution (15). Local adaptation refers to the fact that local variants of a given pathogen are often better adapted to cause disease and transmit among
local, i.e. sympatric, host variants than among allopatric hosts (16). Differences in local adaptation are usually observed in host and pathogen variants that are geographically separated (17, 18).

In this respect, the observation that human-adapted MTBC exhibits a phylogeographic population structure is particularly relevant (Chapter 2A.4, Figures 2A.4.1 and 2A.4.2). Indeed, many studies have shown that the various phylogenetic lineages of MTBC (Figure 2C.1.1) are associated with different geographical regions and human populations (6, 19–22). Not only should this geographically structured diversity in MTBC be considered in the design of new TB diagnostics, drugs and vaccines (23), but the phylogeography of MTBC also suggests possible local adaptation of the various MTBC lineages to different human populations (12). This notion is supported by the observation that the phylogeographic host-pathogen association in human TB remains

**Figure 2C.1.1** Global phylogeny of human-associated MTBC based on the whole genome sequences of 21 representative clinical strains (adapted from Ref. (10)).
stable in the cosmopolitan environments of London, San Francisco, and Montreal, despite a presumed degree of intermingling among patients and their pathogens (6, 19, 24, 25). Moreover, a study in San Francisco has shown that active TB transmission tends to occur more frequently in sympatric compared to allopatric host populations (6). Interestingly, the transmission of allopatric strains occurs disproportionately in patients with strong risk factors for TB, including HIV co-infection and homelessness (6). Unfortunately, this study could not adjust for all social and environmental variables that might influence particular host-pathogen combinations (26). For example, it is known that immigrants tend to interact more extensively among themselves, which will increase the chances of them getting infected with a sympatric, as opposed to an allopatric strain. At the same time, this reduced social mixing across ethnically distinct patient populations will limit the exposure of the local population to imported allopatric strains. Nevertheless, while social factors clearly play a role, the phylogeography of MTBC and the current epidemiological data are consistent with local adaptation of MTBC variants to different human populations (12). Future work will determine whether the host-pathogen association characteristic of human TB can be linked to some underlying biology or whether it represents a mere sociological phenomenon (27). If biological factors are involved, this association then, would have implications for the control of new tools to control TB (23).

**EVOLUTIONARY ARMS RACE OR IMMUNE SUBVERSION?**

Assuming the host-pathogen association in human TB reflects co-evolution and local adaptation, we would expect to observe the relevant changes in parts of the host and mycobacterial genomes which are known to interact with each other. Specifically, antigenic variation and host immune escape, followed by reciprocal adaptive changes in the host immune system, are an important characteristic of host-pathogen co-evolution (13). This phenomenon has been referred to as an “evolutionary arms race” (28), and has been described for a variety of infectious diseases (29, 30). However, whether or not MTBC utilises a similar strategy of antigenic variation to overcome host defences is unknown (31).

In a recent study, Comas *et al.* used next-generation genome sequencing to study the sequence variation of 78 experimentally confirmed human T cell antigens in 21 clinical strains representative of MTBC’s global diversity (10). The authors found that on average, these antigens were not more diverse and equally conserved than essential genes (Figure 2C.1.2). Strikingly, when comparing the specific T cell
epitopes to the corresponding non-epitope regions in these antigens, the authors found that the epitopes were not only more conserved than the non-epitope regions, but also the most evolutionarily conserved regions of the MTBC genome (Figure 2C.1.2). Indeed, more than 95% of the 491 epitopes analyzed harboured no amino acid change at all. The authors interpreted their results by proposing a model in which the adaptive immune responses elicited by these hyper-conserved T cell epitopes contribute to the lung pathology characteristic of pulmonary TB, thereby increasing the transmission efficiency of MTBC. This model is supported by the observation that the formation of lung cavitations usually renders TB patients more infectious (32). Moreover, CD4 T cells seem to be directly or indirectly involved in the formation of cavitary TB disease because HIV co-infected TB patients with low CD4 T cell counts are significantly less likely to present lung cavitations (33).

Figure 2C.1.2 Human T cell epitopes in MTBC are evolutionarily hyperconserved

Note: The ratio of non-synonymous to synonymous nucleotide changes in known T cell epitopes is the lowest compared to other regions of the MTBC genome, indicating that purifying selection is acting to preserve these epitopes, perhaps because the host immune responses they elicit contribute to the spread of MTBC (adapted from Ref. (10))
As a result, the transmission potential of HIV co-infected TB patients is often (but not always) reduced compared to HIV uninfected individuals (34). If confirmed, this model suggests that vaccine-induced immunity to any of these hyper-conserved epitopes might be detrimental to the host, perhaps even leading to an increase in MTBC transmission. While this disturbing thought remains speculative at this stage, the hyperconservation of T cell epitopes included in the analysis by Comas et al. need to be considered in the development of new TB vaccines.

The hyperconservation of these T cell epitopes does, of course, not preclude the possibility of antigenic variation occurring in other parts of the MTBC genome. One of the limitations of the study by Comas et al. lies in the fact that the authors have had to exclude the PE and PPE gene families due to technical limitations (31). As discussed in Chapter 2A.4 these genes are known to be highly diverse and some have been shown to be immunogenic, which together has led to the hypothesis that PE/PPE genes might be involved in antigenic variation (35). A recent study analyzed the sequence variation of a number of these genes (36). Intriguingly, the authors found that PE/PPE genes were indeed hyper-variable compared to other genes in MTBC, but the ratio of nonsynonymous- to synonymous nucleotide changes in these genes is consistent with neutral evolution. In other words, the diversity in PE/PPE genes may be due to factors intrinsic to the gene structure rather than to immune selection (36). One of the main handicaps in studying PE/PPE genes is that the function of most of these genes remains unknown. More work is needed to understand the role of the different members in these gene families, as well as the relevance of the genetic diversity between and within them.

INTERACTION WITH HUMAN GENETIC DIVERSITY

TB used to be a major killer in Europe until only a few decades ago (37). Yet, according to one theoretical study (38), the selective pressure imposed by MTBC seems to have been insufficient to explain the striking decrease in TB mortality in Europe, which started at the end of the 19th century, long before the introduction of the first vaccines and antibiotics against TB (27). The authors have used a time window of 300 years for their analysis, and concluded that this was too short of a time period for the selection of increased host resistance. However, we now know that MTBC has been affecting humankind for much longer, and as discussed above, there is increasing evidence that co-evolution might have occurred (12). We also know that humans differ in their susceptibility to TB, as demonstrated by early twin studies (39). In addition, many susceptibility loci for TB have been discovered in different human populations (40), suggesting that TB has indeed imprinted on human evolution (41). Alternatively, these human variants might be the result of
SECTION 2 MAIN PLAYERS

selection pressures imposed by other infectious diseases or mere genetic drift, and the variable susceptibility to TB might have emerged incidentally.

Further support for the role of MTBC in shaping human diversity comes from several recent studies, which report associations between human immune variants and particular MTBC lineages (42-46). For example, one study in Vietnam showed that TB patients with a particular variant of Toll-like receptor-2 were more likely to carry an MTBC strain belonging to Lineage 2 (also known as the East-Asian/Beijing lineage, Figure 2C.1.2) (42). Similarly, in Indonesian patients, this MTBC lineage was associated with human polymorphisms in SLC11A1 (formerly known as "NRAMP1") (46). Studies in Ghana reported that a human gene variant in the autophagy gene IRGM was associated with protection against TB caused by MTB sensu stricto but not against M. africanum (i.e. the West-African MTBC lineages in Figure 2C.1.1) (43). Conversely, the same authors found that a gene variant of the Mannose Binding Lection protected against TB was caused by M. africanum but not by MTB sensu stricto (44). Taken together, these studies support the view that genetic diversity in MTBC is interacting with human variation. As discussed in Chapter 2A.4 there is increasing evidence that the variation in MTBC impacts on the outcome of TB infection and disease (47). Given the intimate dialog between MTBC and the human immune system (48), variation in both MTBC and the human host is likely to have an effect on the host-pathogen interaction. These interactions will also be of relevance for the development of new TB vaccines.

CONCLUSION

MTBC is more genetically diverse than previously thought. Moreover, TB is an ancient disease, and the phylogeographic population structure and epidemiology of MTBC suggest local adaptation of particular MTBC lineages to different human populations. There is mounting evidence that MTBC genetic diversity and human variation interact, possibly as a consequence of the long-standing host-pathogen association in TB. While more work is required to define and understand the complexities of these interactions, the diversity in both MTBC and its human host will have to be considered in the design and deployment of new TB vaccines.

Acknowledgments

I thank 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 (PP0033-119205), the Leverhulme-Royal Society Africa Award (AA080019), and the National Institutes of Health (HHSN266200700022C and AI090928).
REFERENCES


