KELLERSBERGER MEMORIAL LECTURE

EVOLUTION OF MYCOBACTERIUM TUBERCULOSIS AND IMPLICATIONS FOR VACCINE DEVELOPMENT

Sebastien Gagneux, PhD

ABSTRACT

Tuberculosis (TB) is a growing public health threat, particularly in the face of the global epidemics of multidrug resistance. Given the limited efficacy of the current TB vaccine and the recent clinical failure of the most advanced new TB vaccine candidate, novel concepts for vaccine design should be explored. Most T cell antigens in the human-adapted Mycobacterium tuberculosis complex (MTBC) are evolutionarily conserved and under strong purifying selection, indicating that host immune responses targeting these antigens might not be protective. By contrast, a few highly variable T cell epitopes have recently been discovered, which could serve as alternative vaccine antigens. Moreover, there is increasing evidence that the human-adapted MTBC has been co-evolving with the human host for a long time. Hence, studying the interaction between bacterial and human genetic diversity might help identify additional targets that could be exploited for TB vaccine development.

INTRODUCTION

Tuberculosis (TB) remains a global public health emergency. According to the World Health Organization, there were an estimated 1.5 million deaths and 9.6 million new cases of TB in 2014 (1). The global TB epidemics are worsening due to increases in antibiotic resistance in many parts of the world (2). TB is caused by a group of genetically closely related bacteria known as the Mycobacterium tuberculosis complex (MTBC). The MTBC can be divided into seven human-adapted phylogenetic lineages (known as Lineage 1 to 7) and two lineages adapted to various wild and domestic animal species (3). In humans, the outcome of TB infection and disease is highly variable, ranging from latent infection with no symptoms to classical pulmonary TB, extra pulmonary disease and TB meningitis (4).

In the past, this variation in outcome has primarily been attributed to host and environmental factors (5). However, there is growing evidence that bacterial variation also plays a role. Indeed, many experimental studies have shown that MTBC clinical strains differ in virulence and immunogenicity (reviewed in Ref. 6), and there is increasing evidence that some of these laboratory phenotypes are also reflected in epidemiological settings (reviewed in Ref. 3). For example, the Beijing family of strains, which belongs to the evolutionarily “modern” MTBC Lineage 2 (also known as the East Asian Lineage), has been associated with hyper-virulence and a delayed pro-inflammatory immune response in infection models (7, 8). In clinical settings, Lineage 2 has been associated with faster progression to active disease (9), increased transmissibility (10) and recent emergence (11, 12). Lineage 2 has also repeatedly been associated with antibiotic resistance (reviewed in Ref. 13), which might be a consequence of its higher transmission potential (14) and/or its elevated mutation rate compared to other MTBC lineages (15). By contrast, Lineage 6 (also known as Mycobacterium africanum West Africa II) and other members of the evolutionarily “ancient” MTBC lineages have been associated with lower virulence (16,17) and early pro-inflammatory responses (8). Lineage 6 also exhibits a slower progression to active disease (9), and is less likely to be drug-resistant (18). Because a large proportion of the MTBC genome is dedicated to lipid metabolism (19), and lipids are important cell wall components involved in virulence and immunity to TB (20), it is likely that some of these lineage-specific phenotypic differences are mediated by variation in lipid content. Indeed, some of the strain-specific differences in virulence and pro-inflammatory immune responses in TB have been linked to the presence or absence of particular bacterial lipids (7,21). For example, quantitative and qualitative lineage-specific differences have been reported in mycolic acids (22), which are the most abundant cell wall lipids in the MTBC.

One of the biggest problems for global TB control is the lack of a universally effective vaccine (23). The only currently available vaccine against TB, Bacille-Calmette-Guerin (BCG) is an attenuated derivative of Mycobacterium bovis (24). BCG protects infants well against TB meningitis and military TB, but it is essentially ineffective against regular pulmonary disease in adults, which is the transmissible form of the disease (25). Efforts to develop a better TB vaccine have been ongoing for many decades, but our fundamental

1Department of Medical Parasitology and Infectious Disease, Swiss Tropical and Public Health Institute / University of Basel, Switzerland, e-mail: sebastien.gagneux@unibas.ch
Evidence for host-pathogen co-evolution in human TB: There is mounting evidence that the human-adapted MTBC lineages have co-evolved with humans for thousands of years (32). Similar to humans, the genetic diversity of MTBC is largest in Africa and decreases with increasing distance from Africa (33). Moreover, there is strong phylogeographical evidence that the MTBC originated in sub-Saharan Africa (34). This is also supported by the observation that Mycobacterium canettii and the other so-called “smooth mycobacteria” that share a common ancestor with the MTBC, have only been isolated in the Horn of Africa (35). Similarly, several human-adapted lineages of the MTBC are geographically restricted to parts of Africa. Specifically, Lineages 5 and 6 (also known as M. africanum West Africa I and II) are mainly found in West Africa (36) or in patient of West African descent (37), whereas Lineage 7 almost exclusively occurs in Ethiopia and recent Ethiopian immigrants (38). Hence, these MTBC lineages may be particularly well adapted to infect and transmit among their respective host populations (39). In support of this notion, the sympatric host-pathogen association between the different MTBC lineages and their respective human host populations are maintained in low-incidence cosmopolitan settings (40-43), even during active TB transmission abroad (43, 44).

Furthermore, allopatric host-pathogen combinations in these settings are strongly associated with HIV coinfection, an association that is further enhanced with increasing immune-suppression of the host (43). Even though social factors are likely contributing to these phenomena, the stable associations between different host and pathogen populations, and the fact that these associations are disrupted by HIV-induced immune-deficiency are consistent with host-pathogen co-evolution in human TB (30). Little work has been done so far comparing the genetic population structures of the MTBC and humans. Comas et al. found qualitative and quantitative similarities between the main MTBC lineages and the three main human mitochonial macro-haplogroups (34). Pepperell et al. found no association between the MTBC lineages and human Y chromosomes (45), but when focusing on the main Y chromosome haplogroups, possible associations became evident (29). Finally, additional parallels between the MTBC and human population structures emerge when considering the frequency spectra of single nucleotide polymorphisms. Both humans and the human-adapted MTBC feature a high proportion of so-called singleton mutations, i.e. single nucleotide mutations that occur only in one individual (45-47). In humans, this phenomenon has been linked to strong population increases in the past and a reduction in selective constraint, leading to an accumulation of slightly deleterious mutations, which has important implications for human medical genetics (48).

In the human-adapted MTBC, signals of population expansion have also been detected and connected to the Agricultural and Industrial Revolutions (34, 49-51). Given the strong link between the MTBC and its human host (28), it is perhaps not surprising that bacterial population expansions would follow naturally from similar expansions in humans (29). A reduction in selective constraint linked to serial transmission bottlenecks has also been invoked for the MTBC to explain the high proportion of non synonymous mutations (46), about 50% of which have been predicted to impact protein function (52). Deciphering if and how these predicted changes in protein function are linked to some of the lineage-specific differences in experimental and clinical phenotypes outlined in the first section of this review will require considerably more work. Ideally, this bacterial variation should be studied in the context of the corresponding host variation (31).

It is well known that humans differ in susceptibility to TB (48). However, even though several human genetic loci have been implicated in TB susceptibility (53), replicating these findings across different human populations has proven difficult (54). Part of the reason for this poor reproducibility is that most studies in human genetics of TB have not taken into account pathogen diversity (54), which is particularly problematic in infectious diseases in which the host and the pathogen have a long co-evolutionary history (55). Yet, several studies have reported interactions between particular human polymorphisms and MTBC genotypes (56-61). This supports the notion that to be able to understand clinical phenotypes and predict the outcome of TB infection and disease, the genetic background of both the host and the pathogen need to be considered in conjunction (31). Similarly, host and pathogen diversity need to be considered for vaccine development.

Consequences of host-pathogen co-evolution for TB vaccine development: Assuming the notion of host-pathogen co-evolution in human TB is valid, this has important implications for our understanding of the biology and epidemiology of the disease. Moreover, this will influence the way we think about developing new tools to combat TB, in particular vaccines. Tradi-
tionally, TB vaccine development has focused on dominant antigens that elicit strong T cell responses, in particular interferon-γ (INF-γ, 62). From other infectious diseases we know that pathogens vary their antigens to evade host immunity (63). This kind of evolutionary arms race is the hallmark of many chronic disease systems in which host and pathogen species interact over extended periods of time (64). Does this type of arms race also occur in human TB? Unexpectedly, we found that in the MTBC the known T cell epitopes were evolutionary highly conserved. First we showed this in 2010 using 491 experimentally confirmed T cell epitopes known at the time (65), which included many candidate vaccine antigens, and 21 MTBC genomes covering six of the seven main human-adapted MTBC lineages (66).

In 2015, we repeated this analysis using 1,226 experimentally confirmed T cell epitopes and 216 MTBC genomes from global sources covering all seven MTBC lineages (67). In both studies we found that the large majority (>70%) of T cell epitopes showed no amino acid substitution among the MTBC strains analyzed (66, 67). Moreover, the T cell epitopes were significantly more evolutionary conserved than essential genes. These results were consistent with a previous study, in which we analyzed 27 pe_pgrs genes in 94 phylogenetically diverse MTBC clinical isolates (68). We found that although some pe_pgrs genes were highly polymorphic, their sequence variation was largely confined to the C-terminal PGRS domain, whereas their T cell epitopes were concentrated in the conserved N-terminal PE domain (68).

Hence, despite being under constant host immune pressure, T cell epitopes in the MTBC are under strong purifying selection to remain unchanged, as opposed to experiencing diversifying selection for antigen variation. Based on these observations, we hypothesized that because the MTBC is an obligate human pathogen in which virulence contributes to pathogen transmission (28), the T cell responses elicited by most dominant T cell antigens will be beneficial to the pathogen rather than protecting the host (67). In support of this notion, lung cavitation is a hallmark of TB and a consequence of immune-pathological processes that strongly increase the transmission potential of TB patients (69).

Until now, efforts in TB vaccine development have been focusing on conserved antigens. The results summarized above suggest an alternative strategy, namely to focus on variable antigens, as these might reflect genomic loci that are evolving to evade the host immune responses deleterious to the pathogen. Stimulating host responses to such bacterial targets might therefore promote protective immunity. In an attempt to identify variable T cell epitopes, we recently screened 216 whole genome sequences from MTBC clinical strains covering all seven MTBC lineages (67).

We then selected seven genes that showed a high genetic diversity and a high proportion of non-synonymous variation, and computationally predicted CD4 and CD8 T cell epitopes in these seven candidate antigens using HLA class I and HLA class II alleles that are prevalent in diverse human populations (67). Finally, we synthesized 14 corresponding candidate peptides and used these to stimulate whole blood from 82 HIV-negative TB patients. We found that each of our candidate epitopes was immunogenic in at least one subject as assessed by INF-γ release assay. Moreover, we evaluated the impact of naturally occurring amino acid substitutions in these candidate peptides with respect to their capacity to elicit an immune response in the same individual. We found that 72% of the subjects showed a differential response to the ancestral versus the alternative epitope variant. Taken together, these results show that even though the large majority of the known T cell epitopes in the MTBC are conserved, some variable epitopes exist and this variation can influence host immune recognition.

Whether this natural variation in MTBC results from immune escape and whether the immune responses to such variable epitopes could be exploited for TB vaccine development should be explored further. In addition, a "genome-to-genome" approach could be a new way forward for identifying novel vaccine antigens with a high probability of eliciting protective immune responses in TB (31). In the context of the co-evolution between humans and the MTBC, interrogating both the host and the pathogen genome in conjunction might reveal particular loci that are in conflict with each other (55). Focusing TB vaccine development efforts on such conflicting loci could be particularly rewarding.

In conclusion, a novel, more effective vaccine is urgently needed to control TB globally, particularly in view of the growing epidemics of multidrug resistance. Future efforts to develop new vaccine candidates should consider the natural variation of MTBC populations. Given the mounting evidence that this variation partially reflects host-pathogen co-evolution, studying the interaction between human and MTBC diversity might point to novel targets and strategies for the development and deployment of new TB vaccines.

ACKNOWLEDGEMENTS

I thank Joel Ernst, Richard Copin and all the members of my group for the stimulating discussions. Work in my group is supported by the Swiss National Science Foundation (grant PP00P3_150750), the European Research Council (309540-EVODRTB), and SystemsX.ch.
REFERENCES


