DOTTORATO DI RICERCA

SCIENZE PNEUMO-CARDIO-TORACICHE
DI INTERESSE MEDICO E CHIRURGICO

Ciclo XXI

Settore scientifico disciplinari di afferenza: MED 10

TITOLO TESI:

IMMUNOGENETICS OF GRANULOMATOUS LUNG DISEASE:
A PHENOTYPIC APPROACH TO SUSCEPTIBILITY

Presentata da: Silvia Contini

Coordinatore Dottorato
Prof. S. Mattioli

Relatore
Prof. C. Saltini

Esame finale anno 2009
• **Background**

Immunogenetics of Tuberculosis

Evaluation of a multi-antigen test based on B-cell epitopes peptides for the serodiagnosis of pulmonary tuberculosis.

A model of phenotypic susceptibility to Tuberculosis. Deficient in silico selection of mycobacterium tuberculosis epitopes by HLA alleles.

• **Immunogenetics of sarcoidosis**

M. avium binding to HLA-DR expressed alleles in silico: a model of phenotypic susceptibility to Sarcoidosis.
**BACKGROUND**

**Immunogenetics**

Immunogenetics is a scientific discipline that uses immunological methods to study the inheritance of traits. Traditionally, immunogenetics has been concerned with moieties that elicit immune response, that is, with antigens (antigenic determinants). It has now broadened its scope to study also the genetic control of the individual's ability to respond to an antigen. The immunological methods used in immunogenetics are of two principal kinds, serological and histogenetical. In serological methods, antibodies are used to detect antigens, either in solution or on a cell surface. In histogenetical methods, immune cells (lymphocytes) are used to detect antigens on the surface of other cells. In modern immunogenetics research, the serological and histogenetical methods are combined with molecular methods in which the researcher isolates and works with the genes that code for the traits. This approach of going back and forth from classical to molecular methods has proved to be very successful and has led to the elucidation of several complex genetic systems.

**Granulomatous Lung Disease**

Granulomatous lung diseases, such as sarcoidosis, hypersensitivity pneumonitis, Wegener's granulomatosis, and chronic beryllium disease, along with granulomatous diseases of known infectious etiologies, such as tuberculosis, are major causes of morbidity and mortality throughout the world. Clinical manifestations of these diseases are highly heterogeneous, and the determinants of disease susceptibility and clinical course (e.g., resolution vs. chronic, progressive fibrosis) are largely unknown. The underlying pathogenic mechanisms of these diseases also remain poorly understood. Within this context, these diseases have been approached using genomic and proteomic technologies to allow us to identify patterns of gene/protein expression that track with clinical disease or to identify new pathways involved in disease pathogenesis. The results from these initial studies highlight the potential for these "-omics" approaches to reveal novel insights into the pathogenesis of granulomatous lung disease and provide new tools to improve diagnosis, clinical classification, course prediction, and response to therapy. Realizing this potential will require collaboration among multidisciplinary groups with expertise in the respective technologies, bioinformatics, and clinical medicine for these complex diseases.
IMMUNOGENETICS OF TUBERCULOSIS

EVALUATION OF A MULTI-ANTIGEN TEST BASED ON B-CELL EPITOPE PEPTIDES FOR THE SERODIAGNOSIS OF PULMONARY TUBERCULOSIS

Abstract

SETTING: Two sample panels: (i) twenty pulmonary tuberculosis (TBp) patients and ten healthy subjects (HS) from a country with low incidence of TB (Italy) and (ii) forty-seven TBp patients and 26 HS from a country with high incidence of TB (Morocco).

OBJECTIVE: To identify a combination of Mycobacterium tuberculosis (Mtbd) peptides useful for the serodiagnosis of active TBp.

METHODS: Fifty-seven B-cell epitopes peptides of Mtbd were evaluated by immunoenzymatic assay and the data were analyzed using logistic regression analysis and Random Forest method.

RESULTS: The best discriminating peptide between TBp patients and HS from the sample of the country with low incidence of TB has been the 23 amino acid peptide of the Rv3878 protein. Thus, the sensitivity and specificity was 65 % and 100 %, respectively. In contrast, the same peptide showed 47% and 100% as sensitivity and specificity respectively in the country with high incidence of TB. In addition, the best peptides combination was a pool of nine peptides which showed a sensitivity of 70.2% and a specificity of 100% in the country with high incidence of TB.

CONCLUSIONS: The 9-peptides pool can be useful in identifying patients with active TBp

Keywords: Tuberculosis, peptides, serodiagnosis, B-cell epitope, ELISA
Introduction

Tuberculosis (TB) remains a major public health problem in most developing countries. The global annual incidence of tuberculosis has been estimated at 8 to 10 million cases, with approximately 1.7 million deaths (1). The control of the disease depends largely on early detection and the treatment of active cases.

At present, efforts at developing immunological tests are directed towards the identification of novel antigens that are associated with active disease. Thus the major objective of current immunological studies in TB is the identification of species-specific antigens and determination of the significance of corresponding immune responses. In this context, it is reasonable to hypothesize that proteins preferentially expressed by \textit{Mtb} in models of intracellular growth and infection might be ideal targets for the design of new, highly sensitive diagnostic tests of \textit{Mtb} infection and, eventually, for vaccine design (2, 3).

Serodiagnosis tests like enzyme linked immuno-sorbenent assay (ELISA) are promising in view of their ease of performance and cost-effectiveness, many studies have been carried out by using quite complex antigens, such as whole bacteria, culture filtrates, bacterial extracts, and tuberculin or their purified protein derivatives (PPD) (4-6). However, a large variability in diagnostic accuracy has been reported, depending on the antigen employed and on the heterogeneity of the antibody response in TB patients(7).

The use of biochemical purification methods and recombinant DNA techniques of natural and recombinant proteins to obtain mycobacterial antigens is a challenging task with high cost and low specificities of the assays based on these antigens (8, 9). Therefore, we used synthetic peptides to assess their antibody responses in tuberculosis patients. Furthermore, the \textit{in silico} approach based on screening of \textit{Mtb} coding sequences from the proteins databases by using the prediction programs algorithms to identify B-cell epitopes is promising, since it reduce the cost of synthesizing peptides by reducing the number of peptides required for experimental evaluation for antibody reactivity (10, 11).

In this study, we assessed the potential use of the in-silico selection approach of \textit{M. tuberculosis}-specific synthetic peptides in the serodiagnosis of active tuberculosis.
Materials and methods

Study population
To identify the immunological relevant peptides between the in-silico selected peptides, a first panel of sera from a country with low incidence of tuberculosis (Italy) were analysed. Specifically, 10 sera from skin test negative (TST-negative) healthy subjects without any history of TB exposure and 20 sera from active TB patients enrolled at Division of Respiratory Medicine of the University of Rome “Tor Vergata”, at the “L. Spallanzani” Institute (Roma, Italy). The diagnosis of active TB was microbiologically confirmed by culture isolation of \textit{Mtb} in all TB cases. Patient’s details are shown in Table 1.

A second panel of sera from a country with high level of TB incidence has been used to evaluate the relevant peptide and the pool of peptides selected with the first study population. To this end, 47 patients with pulmonary active tuberculosis from Morocco were evaluated (Table 1). All patients underwent to smear and culture examination for \textit{Mtb} according to standard procedures. Specifically, 31 Smear and culture positive (S+C+) patients and 16 smear-negative, culture-positive (S-C+) TB patients were evaluated in the study (table 1). Further, sera from twenty-six skin test positive healthy voluntary donors were collected from the Blood transfusion Regional Center (CRTS) of Rabat, Morocco.

All study subjects were found to be HIV negative. Sera from all patients were collected before anti-tuberculous treatment and were stored at -20°C until use.

The B-cell epitopes selection
Fifty-seven synthetic peptides containing potential B-cell epitopes were selected by in-silico strategy from 25 proteins of \textit{Mtb}. Briefly, the sequence of \textit{Mtb} proteins belonging to RD1 genomic region (12) and of proteins expressed during \textit{Mtb} growth in human macrophage (2, 3), was obtained by the Swiss-Prot database (Table 2). Each single protein was screened for potential B-cell epitopes by using three computer algorithms of antigenicity: the Jameson-Wolf index, the Hopp-Woods method and the Parker index. These algorithms, available with Genrunner software [Hastings Software, Inc., Hastings, N.Y.] and Antheprot software (13-15), predict antigenic regions based on probability of surface exposure, local hydrophobicity, beta turns amino acid sequence, atomic flexibility and experimental HPLC retention times of synthetic peptides. Only protein regions identified as containing B-cell epitopes by at least two algorithms were selected.

Finally, the sequence of each epitope was screened for homology search of all known protein sequences by BLAST (\url{http://www.ebi.ac.uk/blast/}) through the selection of peptides that are uniquely restricted to the \textit{Mtb}-complex (homology lower than 80% compared to other bacteria).
The sequence of all the selected peptides is reported on Table 2.

**Peptides**

The peptides were synthesized by Fmoc chemistry as free amino acid termini by ABI (http://www.abi.it/). All peptides were purified by reverse-phase chromatography (RPC). The purity of the purified peptides was higher than 90%. Sequence and purity were confirmed by mass spectrometry and analytical RPC. Lyophilized peptides were diluted in DMSO at stock concentrations of 10 mg/ml for each peptide and stored at –80°C until use.

**Peptide-based ELISA**

Ninety-six wells high-binding capacity flat-bottom microplate (Greiner) was coated with 1 µg per well of each peptide in a Phosphate Buffered saline (PBS) buffer (pH 7.4) and kept at 4°C overnight. Then the plate was washed with 0.05% (v/v) Tween-20 in PBS and blocked with 3% of bovine serum albumin (BSA, Sigma) for 1 hour. Each sample was assayed at a 1:50 dilution (2 hours, at room temperature). A horseradish peroxidase (HRP) conjugated anti-human IgG was used as the second antibody. The reaction was revealed using Ortho-Phenyl-Diamine (OPD)/H₂O₂ (Sigma) and quantified using an ELISA reader at 450 nm after stopping the reaction with 0.5M H₂SO₄. Blank wells were included as negative control in this assay.

**Statistical analysis**

Test results are presented as means ± standard deviation (SD) of the mean. ELISA tests were scored positive when individual readings were greater than two standards deviations above the optical density average of healthy subjects. Comparisons between different groups were made by Mann-Whitney test.

The best peptides discriminating between TB patients and controls sera were identified by using univariate logistic regression model using Statistical Package for Social Science™ (SPSS version 15), the Random Forests Method [http://genesrf.bioinfo.cnio.es/] (16) and stepwise multivariate logistic regression model using SPSS program.

The Receiver Operating Characteristic (ROC) curve was used to evaluate the performance of the ELISA tests with the two categories (TB patients and controls); the Area Under the Receiver Operating Characteristic curve (AUC) was used as a measure of diagnostic quality. The cut-off value used here is the one that leads to a ≥50% probability of being a case.
Results
In order to identify peptides capable of being recognized by TB patients, a first panel of TB and control sera from a low incidence country for TB was tested. Figure 1 shows the IgG response against each one of the fifty-seven selected peptides containing potential B–cell epitope (Figure 1). The antibody responses against the peptides were variable (figure.1). Univariate logistic regression method showed that the best discrimination score value between TB and healthy controls was 12.128 obtained with the peptide Pep51 of the Rv3878 protein (see for sequence table 2). With a 2SD cut-off value, the sensitivity and specificity was 65 % and 100% respectively and the area-under-the-ROC curve was 0.912 (Figure 2).

When multivariate logistic regression method and random Forest method were used, the best peptides combination identified was: Pep5 (Rv0747), Pep11 (Rv1114), Pep14 (Rv1434), Pep26 (Rv1979c), Pep42 (Rv3736), Pep48 (Rv3874 also known as CFP10), Pep51 (Rv3878), Pep55 (Rv3883) and Pep57 (Rv3883).

The peptides identified were then used in a second population from a country with high incidence of TB in order to evaluate their performances in the diagnosis of active TB.

The results of the ELISA based on the peptide Pep51 alone among the forty-seven TB patients and the twenty-six healthy controls showed a sensitivity and specificity of 47% and 100% respectively with a cut-off value at 2SD of the controls and the area-under-the-ROC curve was 0.842 (Figure 3). However, the results among the same population but with the ELISA based on the selected 9-peptides showed a higher value (0.938) of the area-under-the-ROC curve (P value=0.04) (Figure 4). Further, figure 5 shows the box plots of the antibody titer obtained with the second panel sera of TB and healthy controls. With a cut-off value at 2SD of the controls, the sensitivity and specificity was 70.2 % and 100% respectively (Figure4).
Discussion

The availability of mycobacterial genome sequences, the development of chemical synthesis techniques of peptides and the development of algorithms for the identification of immunodominant B-cell epitopes, might determine the rapid identification and testing of epitopes with diagnostic potential for active TB with the advantage of reducing the number of experimental evaluations for antibody reactivity. In fact, the use of ELISA test based on synthetic peptides would in principle circumvent the technical complexity and potential antigen variability associated with antigen purification. Further, synthetic peptides used as antigens have the added advantage of avoiding contamination with low-level impurities derived from cloning vectors required to produce recombinant proteins. Such impurities could decrease the specificity of the assay if they were cross-reactive with M. tuberculosis antigens. Finally, using synthetic peptides as the diagnostic antigen permit the uniformity and standardization of antigen preparations (17).

However, the main problem in setting immunodiagnostic systems is the identification of immunodominant antigen(s) specific of the active disease phase. In this context, previous studies (2, 3), have shown that Mtb is capable of expressing differential transcription programs in response to different environments. In particular, Mtb can express different gene sets in synthetic media compared to macrophage cultures. For instance, RD1 proteins are expressed both in early and late phase cultures, are variably expressed in in vivo in murine models, but are expressed at very low level in macrophage cultures (2, 3, 18-20).

With this background, we hypothesized that the Mtb proteins overexpressed in a model of Mtb infection in macrophage cultures could be the specific target of a strong immune response in active TB subjects. Thus, we combined the use of relevant Mtb antigen in a model of Mtb infection with in silico approach to identify B-cell epitopes to evaluate their potential use in TB serodiagnosis.

We were able to identify a peptide belonging to the Mtb protein Rv3878, as the best discriminating peptide between TB patients and healthy controls among a population of low exposure level of TB. Rv3878 is one the proteins encoded by RD1, a region present in Mtb and virulent M. bovis genomes but missing from the DNA of all substrains of M. bovis Bacillus Calmette-Guerin (BCG) (12).

Although our findings are not in agreement with previous observation by Brusasca et al (21) who reported that only 3–7% of human TB patients responded to RD1 region proteins, including Rv3878, it is well known that all RD1 antigens elicited a high antibody response in guinea-pigs infected with Mtb (21). Further, in a recent study (22), it has been shown that the Rv3878 protein elicited strong humoral immune response in TB patients in Indian population, supporting the notion
that the antibody response to a specific antigen often depends on the geographical location and ethnic background of the population being studied (23).

Several studies indicate that combinations of mycobacterial antigens are likely to have better diagnostic value (24, 25). Using both multivariate logistic regression analysis and random Forest method, we were able to identify a pool of 9-peptides belonging to the proteins Rv0747, Rv1114, Rv1434, Rv1979c, Rv3874, Rv3736, Rv3878 and Rv3883 as the best discriminating peptides combination between TB patients and healthy controls.

Recent reports indicate that the antigen Rv3874 is frequently recognized by antibodies in patients with TB, from regions where TB is not endemic (21) and from regions where TB is endemic (26). Further, Rv3883c or Mycosin-1 is an extra-cellular protein that is membrane- and cell wall-associated, and is shed into the culture supernatant. The protein is expressed after infection of Macrophages (27), but Rv3883C is not expressed in the attenuated M. bovis strain BCG, although the gene for mycosin-1 is present in the genome of this organism. This gene was found to be situated 3700 bp from the RD1 deletion region (28).

Moreover, recombinant Rv1979c has been identified as an antigen capable to induce IFNγ in correlation with extent of lesions at necropsy (29). Finally, no information is available in the literature about the immunoreactivity of the other proteins identified as relevant in discriminating TB patient sera.

Both genetic and environmental factors also influence the antibody response. For example, the varied antibody response to Mtb is managed by HLA types. That fact, therefore, that our pool of peptides identified in an population has strong diagnostic value also for another population with a different genetic background, suggest that these peptides may form the basis of a test that can be used independent of location.

Our results, in accordance with those obtained by other investigators, confirmed higher sensitivity for the smear-positive patients than for the smear-negative patients (P<0.001), due probably to a greater exposure to antigens in patients with high bacillary loads (30, 31). However, in developing countries such as Morocco where TB disease is prevalent, serological tests might play a major role in the setting of the diagnosis of patients with suspect TB where acid-fast bacilli examination is negative. Our results showed that with the ELISA based on the peptide pool, TB can be diagnosed with 44% sensitivity in acid fast bacilli smear-negative TB patients. This results is interesting given the facts that there are no simple fast alternatives (e.g. culture takes 6 weeks) in these clinical settings, and even thought that x-ray is a standard tool of TB diagnosis it could be quit questionable in some cases (32).
Finally, it is worth to consider that in the study population with high TB prevalence we evaluated, all controls were TST-positive. Thus, the peptide based ELISA was able to distinguish between active and inactive tuberculosis infection, while other immunological based tests such as TST and gamma–interferon based assays are not capable (33, 34). Nevertheless it would be important to explore how the ELISA based on the peptide pool performs in recently exposed individuals as well as in subjects with cured tuberculosis or under chemotherapy treatment that are known to present high anti-\textit{Mtb} IgG levels (35).

In conclusion, with the limitation of the small study population evaluated, this study indicates the applicability of the combination of nine well-defined synthetic peptides to the serodiagnosis of active TB. Their use in an ELISA would be important because of their stability, reproducibility and capability to identify strong humoral immune responses in patients with active pulmonary TB. The \textit{Mtb}-complex specificity of these peptides strongly encourages further evaluation of this pool of peptides as a reagent for TB-specific immunodiagnostic assays on larger study populations including other disease groups such as patients with pulmonary disease other than TB, or seropositive patients for HIV.
REFERENCES


Table 1: Demographic characteristics of the study subjects

<table>
<thead>
<tr>
<th>Study group</th>
<th>Number</th>
<th>Mean Age (years)</th>
<th>Sex-ratio (M/F)</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>First panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary TB patients</td>
<td>20</td>
<td>39 (33-53)</td>
<td>1.22</td>
<td>Caucasians</td>
</tr>
<tr>
<td>Healthy subjects (TST-)</td>
<td>10</td>
<td>30 (26-42)</td>
<td>1.00</td>
<td>Caucasians</td>
</tr>
<tr>
<td>Second panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary TB patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPM+</td>
<td>31</td>
<td>35 (25-45)</td>
<td>2.1</td>
<td>North Africans</td>
</tr>
<tr>
<td>TPM0</td>
<td>16</td>
<td>39 (28-51)</td>
<td>2.2</td>
<td>North Africans</td>
</tr>
<tr>
<td>Healthy subjects (TST+)</td>
<td>26</td>
<td>39 (34-48)</td>
<td>2.5</td>
<td>North Africans</td>
</tr>
</tbody>
</table>

TP = Pulmonary tuberculosis, TPM+ = Smear positive pulmonary tuberculosis, TPM0 = Smear positive pulmonary tuberculosis, M/F = male /female, TST+ = Tuberculin Skin Test positive, TST- = Tuberculin Skin Test negative.
### Table 2: List of the peptides used in this study

*Peptide epitopes starting from amino acid position 1 in the protein sequence where two Alanine were added to the N-terminus of the peptide in order to increase binding affinity to the epitope.*

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide Number</th>
<th>Starting amino acid</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0418</td>
<td>Pep1 at 1</td>
<td>AAMVKNRSMPAV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep2 at 216</td>
<td>IPVSVTKSVPQFGQSGPTTYK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep3 at 377</td>
<td>AGIERTPEVALKSMAGKTAQDT</td>
<td></td>
</tr>
<tr>
<td>Rv0747</td>
<td>Pep4 at 1</td>
<td>MSWVMVSPELVVAAAADLAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep5 at 385</td>
<td>HTLQDFVIMNYDPQETLGRPLG</td>
<td></td>
</tr>
<tr>
<td>Rv0755c</td>
<td>Pep6 at 107</td>
<td>VYHPAVYQANVRTRLLAVSYPQVQ</td>
<td></td>
</tr>
<tr>
<td>Rv0934 (FA)</td>
<td>Pep7 at 140</td>
<td>EHEKLINFLKANMYIQ</td>
<td></td>
</tr>
<tr>
<td>Rv0956</td>
<td>Pep8 at 12</td>
<td>ARLVLASGTVGGSLR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep9 at 176</td>
<td>ETLHERIKYTERLVAVAAALATH</td>
<td></td>
</tr>
<tr>
<td>Rv1114</td>
<td>Pep10 at 46</td>
<td>RDVVLDDLNNYLQHPVY</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep11 at 60</td>
<td>FPVTVDDLVRNLKVRGQW</td>
<td></td>
</tr>
<tr>
<td>Rv1382</td>
<td>Pep12 at 8</td>
<td>GSLIAAFLVMLHVLARLMMRGRW</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep13 at 120</td>
<td>RGVAGKVAGGILARWRRLPSGT</td>
<td></td>
</tr>
<tr>
<td>Rv1434</td>
<td>Pep14 at 1</td>
<td>AAMRASPAERVDGAFAAGPHTQSV</td>
<td></td>
</tr>
<tr>
<td>Rv1582: (RPO-10)</td>
<td>Pep15 at 143</td>
<td>SYVPPQRELARQTFG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep16 at 254</td>
<td>AFWRRRVYFVPFADEQD</td>
<td></td>
</tr>
<tr>
<td>Rv1666</td>
<td>Pep17 at 91</td>
<td>DRYRHLVALISITDFGAAGPRESWRA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep18 at 149</td>
<td>AAVQAAAWAVLYLAVYFMRRLCGTDYG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep19 at 222</td>
<td>DGYRRHVCMARQPKWQGLRWRLLGEP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep20 at 372</td>
<td>YFPEGLRLIDRGLIGVAGGELS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep21 at 502</td>
<td>TRMGYLGLVRKAAVTRVWTDEAQ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep22 at 538</td>
<td>TTFPGHVGVAGGIALAAHLHBDR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep23 at 568</td>
<td>AHDYHSQAEVYVYNQLYDMMVEA</td>
<td></td>
</tr>
<tr>
<td>Rv1979c</td>
<td>Pep24 at 138</td>
<td>LTFLQFSVVLAINLFLONRAIKWAN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep25 at 413</td>
<td>DKVPLAVAVSVVGLAVSYD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep26 at 466</td>
<td>HYYRIRRVGGDRSPTR</td>
<td></td>
</tr>
<tr>
<td>Rv1983</td>
<td>Pep27 at 348</td>
<td>NGHVTPAYVVEELISPTPPAI</td>
<td></td>
</tr>
<tr>
<td>Rv1986</td>
<td>Pep28 at 36</td>
<td>LPVPVAllctvSHDLVIAAGIAFGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep29 at 102</td>
<td>TPVRLAEVLVTCATGILRGLETPPGWS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep30 at 165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2994</td>
<td>Pep31 at 13</td>
<td>IMIVSLQVSASSFLNPQTVLFP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep32 at 56</td>
<td>PSWLGVYTMTFAGAYLLD184VMR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep33 at 98</td>
<td>HSLWQGYFVLFLGMAAAGGCGNS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep34 at 204</td>
<td>SPYRGISSLIRHIAASALLMMP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep35 at 228</td>
<td>VTVTFLVLWLLHHEVWSVAQ</td>
<td></td>
</tr>
<tr>
<td>Rv3042c</td>
<td>Pep36 at 137</td>
<td>LIIRVSVSSPCVGQGRQLIKTVVAEE</td>
<td></td>
</tr>
<tr>
<td>Rv3219</td>
<td>Pep37 at 68</td>
<td>RALKRARNATK</td>
<td></td>
</tr>
<tr>
<td>Rv3619c</td>
<td>Pep38 at 28</td>
<td>QAISDVLTASDF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep39 at 72</td>
<td>QKQVVQAGNNMaT</td>
<td></td>
</tr>
<tr>
<td>Rv3736: (38 KD)</td>
<td>Pep40 at 24</td>
<td>TLLRAAGVRQDOVQV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep41 at 34</td>
<td>QDVGNYDAFIRSARAJAES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep42 at 145</td>
<td>VALQVIRLLGGADYAPLAVH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep43 at 226</td>
<td>GIVSERTVRQVQQLPTGAANLV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep44 at 269</td>
<td>TPVILDEVRKDVADYLR</td>
<td></td>
</tr>
<tr>
<td>Rv3872</td>
<td>Pep45 at 56</td>
<td>EQQDLNASAAQ</td>
<td></td>
</tr>
<tr>
<td>Rv3873</td>
<td>Pep46 at 68</td>
<td>TFMVVWLQTASTQAKTR</td>
<td></td>
</tr>
<tr>
<td>Rv3874 (CFP-10)</td>
<td>Pep47 at 54</td>
<td>AVYRVFSQEAANKQK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep48 at 74</td>
<td>TNIQGAGQYSHRA</td>
<td></td>
</tr>
<tr>
<td>Rv3875 (ESAT-6)</td>
<td>Pep49 at 2</td>
<td>QQQWFAGIEEAASAAHQYNTSHSIL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep50 at 65</td>
<td>NALQNLARTISEA</td>
<td></td>
</tr>
<tr>
<td>Rv3878</td>
<td>Pep51 at 139</td>
<td>PRVVAATPVQHLQAPHAVQMSQ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep52 at 216</td>
<td>GDVQPAELYVAAAR</td>
<td></td>
</tr>
<tr>
<td>Rv3879 (UF)</td>
<td>Pep53 at 135</td>
<td>TAINLVLTHGVAHNPLAAELQPG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep54 at 325</td>
<td>PHVRPAALAEPQG</td>
<td></td>
</tr>
<tr>
<td>Rv3883c (MycP1)</td>
<td>Pep55 at 219</td>
<td>IDYAVNVKVGTVVVVAAAGNTGGDC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep56 at 337</td>
<td>DDLVGAGVIDAYA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pep57 at 372</td>
<td>APVYNRRLPPPVVEP</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Graphical value representation of the optical density (OD) obtained for each peptide and sera from the first panel by ELISA showing the heterogeneity of the antibody responses in TB patients and healthy subjects against *Mycobacterium tuberculosis* peptides. Antibody level (OD):  
- OD<0.100
- 0.100≤OD<0.200
- 0.200≤OD

The right column shows the score value for each peptide obtained by univariate regression analysis.

<table>
<thead>
<tr>
<th>Peptides</th>
<th>TB patients</th>
<th>Healthy subjects</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pep01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pep57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: The ROC curve of the peptide Pep51 evaluated with the panel of TB and control sera from a low incidence country for TB. (AUC = Area under the curve)

\[ \text{AUC(Pep51)} = 0.912 \]
Figure 3: The ROC curves of the peptide Pep51 and the 9-peptides pool evaluated with the panel of TB and control sera from a high incidence country for TB. (AUC = Area under the curve)
Figure 4: Box plot representation of the distribution of levels of antibodies (IgG) to the 9-peptides pool of *Mycobacterium tuberculosis* in TB patients and healthy subjects originated from a high incidence country for TB. (Error bars: 95% confidence interval for mean.)
A MODEL OF PHENOTYPIC SUSCEPTIBILITY TO TUBERCULOSIS:
DEFICIENT IN SILICO SELECTION OF MYCOBACTERIUM TUBERCULOSIS EPITOPES BY HLA ALLELES

Abstract.

HLA-DR allelic variants have been associated with tuberculosis (TB) susceptibility in different populations with risk ratios of 3.7 to 7.2. We hypothesized that the genetic susceptibility to TB depends upon the reduced capability of HLA-class II alleles of TB patients to bind and select peptide antigen from the Mycobacterium tuberculosis (MTB) expressed genome. To test this hypothesis, we developed a software that can predict HLA-DR restricted epitopes within the whole MTB genome based on quantitative peptide binding matrices. We analyzed the number of MTB epitopes recognized in two previously described populations of TB patients and matched controls and in a control population comprised of individuals affected by a sarcoid-like granuloma induced by beryllium and by healthy exposed controls. The number of putative epitopes within the whole MTB genome which could be bound by any HLA-DR allele (HLA-DR immunome of MTB) was 405,422 out of 1,304,277 possible 9-mers i.e., 31.08% of the global capability, instead of the expected 35%. When tested at an affinity level equivalent of the 1% of the best binder peptides, the HLA-DR alleles (HLA-DRB1*0801, *0802, *1401, *1501 and *1502) associated with TB susceptibility recognized a significantly lower mean number of MTB-epitopes (7,862±4,258) than the MTB-epitopes recognized by HLA-DR alleles (HLA-DRB1*0301, *0701, *1101, *1102, *1301 and *1302) negatively associated with TB (11,376±1,984, p<0.032). The number of epitopes bound at high affinity out of the whole MTB genome by the combination of the two HLA-DR alleles carried by each individual was lower in TB patients [TB-population 1: 11,341±908 (mean+SEM); TB-population 2: 15,303±657] than in matched healthy controls (CTR-population 1: 13,587±605, p<0.03 vs TB-population 1; CTR-population 2: 1,6841±555, p<0.04 vs TB-population 2). No difference was seen in individuals with the sarcoid-like granuloma induced by beryllium compared to the exposed healthy (beryllium-hypersensitivity: 17,593±447; controls 18,014±421; p=0.57). The data suggest that HLA-DR alleles associated with susceptibility to tuberculosis may be endowed with a reduced capability to bind at high affinity T-cell epitopes and select them for antigen presentation. The same alleles may contribute to determine the reaction to mycobacteria in non tuberculous granulomatous disorders. (Sarcoidosis Vasc Diffuse Lung Dis 2008; 25: 21-28).
Key words: tuberculosis, susceptibility, HLA, epitope prediction, T-cell response, sarcoid-like granulomas
Introduction

Classic tuberculosis (TB) epidemiologic studies indicate that the genetic background may play an important role in susceptibility to TB (1). With the expansion of molecular genetics studies, a large number of genes have been associated with TB, leading to think that susceptibility might be the result of an imbalance between the effects of susceptibility genes and of protective genes such as the natural resistance-associated macrophage protein Nramp1, the IFN-α receptor 1, the IL-12 receptor β1 genes and the HLA genes. In this regard, a number of allelic variants of the HLA locus which have been associated with TB risk in population studies indicating DQB1*05, *06 and DRB1*08, *14, *15 and *16 as susceptibility genes with risk ratios ranging from 3.7 to 7.2, and HLA-DRB1*03, *07, *11 and *13 as “protective” genes (2-10). HLA genes code for surface receptors that are known to play a pivotal role in the generation of antimicrobial immunity (11). HLA class II proteins (HLA-DP, -DQ and -DR) bind peptides derived from the digestion of microbes in the phago-lysosome of antigen presenting cells, carry them to the cell surface and present them to cytokine-producing CD4 T-cells. Structure-function studies of HLA class II molecules have indicated that the selection of antigenic peptides by the HLA receptors is dictated by the chemico-physical interaction between the amino acid side chains lining receptor-like pockets on the floor of the HLA antigen binding groove and the the agretopes i.e., the aminoacid side chains of the antigenic peptides (12, 13). As most of the polymorphisms generating allelic variability of the HLA molecules code for aminoacid changes in the peptide binding groove’s pockets, each HLA allele will bind a unique set of aminoacid side chains hence selecting a discrete set of antigenic peptides for antigen presentation (14, 15). As a consequence, the ability of HLA alleles to select a peptide antigen repertoire from a given microbe for antigen presentation, hence to induce a protective immune response, may vary widely, leading to greater susceptibility to infection of the subjects carrying the HLA alleles less efficiently binding and presenting antigens (16). In this context though, since each individual coexpresses at least 2 HLA-DR molecules on the cell surface of the APC, it is reasonable to think that individual susceptibility to infection shall be determined by the capability of each subject’s two HLA-DR molecules combined to recognize microbial antigen epitopes, rather than the carriage of a single susceptibility allele. Moreover, among the HLA-DR alleles, there is a predominant expression of HLA-DRB1 alleles, being expressed at a level five times higher than its paralogues DRB3, DRB4 and DRB5 (17-19). The assessment of HLA-associated susceptibility to TB, as model for susceptibility to granulomatous disorder mediated by mycobacteria, might thus require an analysis of disease associated phenotypes instead of disease associated alleles. To assess this hypothesis, we took advantage of bioinformatics tools allowing the
identification of antigenic peptides in whole microbial genomes by quantitative peptide binding motifs analysis for the HLA alleles (20). We developed software that can predict HLA-DR restricted epitopes in the whole MTB genome based on quantitative implemented peptide binding matrices and used this tool to determine the number of epitopes potentially recognized in the MTB genome in two already described populations of TB patients and matched healthy controls (21, 22). In addition, a population of patients affected by the sarcoid-like granulomatous reaction induced by beryllium and matched beryllium-exposed subjects (23).
Methods

Patients’ characteristics: The study populations was composed by the TB patients and matched controls of two already described separate reports on the genetic susceptibility to TB in which HLA-DR high resolution typing was available for all study subjects (21,22). A population of patients with beryllium hypersensitivity and matched beryllium-exposed unaffected subjects were used as disease control population (23). They were 160 patients with tuberculosis (TB population 1), and 200 controls (CTR-population 1) (see table 1) included in the Kim HS et al. study (22) and the 127 patients with tuberculosis (TB-population 2) and 120 matched controls (CTR-population 2) included in the Vejbaesya S et al. study (21) (see 22 A model of phenotypic susceptibility to Tuberculosis (table 1). The disease control population included 74 subjects with beryllium hypersensitivity (BeH) and 86 beryllium exposed matched controls (Be-CTR) from the Amicosante et al. study (23). The quantitative implemented peptide binding motifs are available only for 52 over more than 300 HLA-DRB1 alleles (24, 25). All together they may cover, with at least one allele, about 90% of the HLA-DR variability of different human populations (26). For the purpose of this study, only the subjects carrying both HLA-DR alleles with an available HLA-DR binding motif were used (see table 1). Specifically, 106 out of 320 (33%) control subjects and 85 out of 287 (30%) TB patients could be subjected to the analysis, as 168 (52%) controls and 151 (53%) TB patients had only one HLA-DR allele with an available binding motif, while 48 (15%) controls and 51 (18%) TB patients had both HLADR alleles without an available binding motif. Consequently, the alleles which were analyzed in the study populations were: HLA-DRB1* 0101, 0102, 0301, 0401, 0402, 0404, 0405, 0408, 0410, 0701, 0802, 0806, 1101, 1102, 1104, 1106, 1301, 1302, 1307, 1401, 1501, 1502. The selected subgroups did not differ for demographical characteristics from the subgroups of subjects excluded from the study for having one or both HLA-DR alleles without a known binding motif (data not shown).

Genomes The genome of Mycobacterium tuberculosis H37Rv strain (NC 000962.2) composed of 4048 genes transcribed into 3,989 proteins and the genome of Escherichia coli K12 (NC 000913.2) were used in this study for the immuno-informatic analysis. Software for the identification and enumeration of epitopes in whole genomes To enumerate the T-cell epitopes present in data sets of proteins as large as a microbial genome, we developed a software for the identification and enumeration of peptide binding epitopes to HLA-DR molecules. This software was developed on LabView platform (National Instruments, US) using a graphic language. Basically, it is an open system in which two different databases are uploaded and crossed. Briefly, the first input database is represented by the protein sequences that can be uploaded from a file in FASTA format. The second database is represented by matrice(s) describing...
the peptide binding capabilities of the HLA-DR alleles under analysis. The present version of the software is equipped with a set of 52 additional matrices for HLA-DR peptide binding profiles (15, 27). The matrices database includes also the threshold values for different affinity levels as reported in the original packages. The distribution of the matrix results of all the 209 possible peptides that can theoretically bound HLA-DR molecule is automatically generated as well as theoretical affinity thresholds. The analysis can be customized by selecting single/multiple proteins and HLA alleles among the set of data loaded, threshold and other parameters. The software generates all the possible nonamer peptides in a protein sequence and analyse them on the HLA-peptide binding matrices in analysis. For each protein and each HLA allele, all the peptides presenting a permissive 23 aminoacid (AA) in relative position P1 are stored in memory together with its matrix score and relative position in the protein. For the purpose of this work, two analysis were implemented in the software. A first analysis is represented by the identification of the epitopes recognised, in each single protein of the data set at the affinity threshold applied for each HLA-DR allele in analysis. This allows the enumeration of the epitopes in the whole data set and the identification and enumeration of proteins that present a defined number of epitopes, such as the proteins that are putatively not recognised by the HLA-DR allele at the affinity threshold used, as they present zero epitopes. A second specific analysis for the identification of the epitopes recognised in a set of proteins by 2 HLA-allelestogether has been developed to mime the situation of the HLA-DR recognition in a single subject. In this analysis, the software identify and enumerate the common and the different peptide epitopes recognised by the two HLA-DR alleles of the subject under evaluation by the position in the test set of proteins at the threshold of affinity applied. The whole MTB and E. coli genomes were analysed for the HLA-DR alleles negatively and positively associated to TB and for the enumeration of epitopes recognized by single subjects in analysis at the different thresholds of affinities equivalent to the 1%, 2%, 3%, 4% and 5% of the best binding natural peptides for HLA-DR alleles (27). The Threshold of affinity is a preselected numerical value used to differentiate between binders and non binders, any peptide frame scoring higher than this value is predicted as binder or vice versa; it correlates with the peptide score (15) and therefore with HLA-ligand interaction, therefore it is an indicator for the likelihood that predicted peptide is capable of binding to a given HLA-molecule. To express the results, we have chosen, the percentage of 1% in order to lower the false positive rate. The number of epitopes recognized by each study subject has been evaluated both as absolute number and as a relative number respect the amount of MTB epitopes in the whole genome. Statistical analysis All the data are expressed as
mean + standard deviation of the mean (SD). Comparisons between groups are made by Student’s t
test. Results 1. The immunome of MTB H37Rv HLA-DR molecules bind a core of nine
aminoacids long protein fragments, when they carry in the relative position 1 (hereafter named P1)
non-polar residues (I, L, M, F, W, Y, V) i.e., 7 out of 20 aminoacids or 35% of all the aminoacids.
Analysis of the genome of M. tuberculosis H37Rv strain (NC_000962), which comprises 4,048
genes that can be transcribed into 3,989 proteins, allowed to estimate that, independently of the
relative affinity of the epitopes for the HLA-DR molecules, the number of putative epitopes capable
of binding any HLA-DR allele in the whole MTB genome was 405,422 out of 1,304,277 possible 9-
mers, that we define as the HLA-DR MTB-immunome. This indicates that the HLA-DR MTB-
immunome encompasses 31.08% of the whole MTB nonamer population. This number is
equivalent to the 88% of the expected theoretical recognition that is equivalent to 35% of the all the
peptides, thus suggesting that MTB encompasses a lower number of epitopes than expected if its
genoce presented a normal aminoacid distribution. Differently from MTB, the number of putative
epitopes capable of binding any HLA-DR allele in the E.coli genome was 537,294 out of
1,566,080 nonamers (34.30%). This is equivalent to 98% of the expected theoretical recognition, a
fraction that is significantly higher than that of the MTB HLA-DR immunome (p<0.0001). 2.
Impaired MTB proteins recognition by HLA-DR alleles associated with TB susceptibility. HLA-
DR alleles HLA-DRB1*0801, *0802, *1401, *1501, *1502, which have been associated with TB
susceptibility in previous studies, recognized significantly lower number of MTB-epitopes
(7,862+4,258) than the HLA-DR alleles HLADRB1* 0301, *0701, *1101, *1102, *1301 and
*1302, associated with TB resistance [11,376+1,984 (P<0.032)], at the affinity level of 1%. Consequently, there was a significantly higher number of MTB proteins (1,268+686) which could
not be recognized by the HLA-DR alleles associated with TB susceptibility compared to HLA-DR
alleles negatively associated with TB (776+232; p<0.001). 24 A model of phenotypic susceptibility
to Tuberculosis 25 3. Phenotypic analysis of the ability of HLA-DR alleles of TB patients to bind
whole MTB genome peptides. When this analysis was applied to those individual subjects carrying
a pair of HLA-DR alleles with known peptide-binding motifs, the number of MTB epitopes
recognized by TB patients in both populations was significantly lower than that recognized by
controls [TB-population 1: 15,303±657, CTR-population 1: 16,841±555; p=0.038 compared to TB
population 1; TB-population 2: 11,341±908, CTR-population 2: 13,587±605, p=0.035 compared to
TB-population 2 (Figure 1 panel A)]. Interestingly, when a population of individuals affected by the
sarcoid-like granulomatous reaction to beryllium was analyzed as a control, no differences where
observed between granuloma-affected and unaffected subjects in their recognition ability of MTB
genoce peptides (beryllium-hypersensitivity: 17,593±447; beryllium-exposed controls 18,014±421;
p=0.608). Finally, when the three population groups were tested for their ability to recognize E. coli genome peptides, no differences were seen between TB and their matched controls nor between Be-hypersensitive subjects and their matched controls (figure 1 panel B), suggesting that the epitope binding defect seen in TB patients was restricted to, or more pronounced for, the MTB genome.
Discussion

The binding of antigenic peptides by the host HLA proteins expressed by antigen presenting cells is thought to represent a limiting step in the development of an antimicrobial immune response. In the context of the importance of HLA-DR genes in the immune response to MTB (28, 29) and of the observations positively or negatively linking different alleles of the HLA-DR, HLA-DQ, HLA-DP, and HLA class I genes to TB susceptibility in HLA association studies (10), it is reasonable to think that altered peptide binding by susceptible HLA molecules may be the cause of susceptibility to disease (11).

With this as a background and in the context that TB susceptibility is increased in homozygous, compared to heterozygous twins (30), it is reasonable to hypothesize that a deficient antigen recognition capability of the immune system might be at the basis of the ineffecacious response to MTB in susceptible individuals, and that the determinant of susceptibility to MTB infection and disease may be the combination of the recognition abilities of both HLA-DR alleles expressed by each subject, i.e., the HLA-DR phenotype. The finding of this in silico model that the number of MTB epitopes recognized by the combination of the two HLA-DR alleles by TB subject was significantly lower than the matched Fig. 1. Total number of epitopes recognized by M. tuberculosis and E. coli. The number of M. tuberculosis (panel A) and E. coli (panel B) epitopes recognised in silico by each study subject with the combination of the two carried HLA-DR alleles at the threshold affinity of 1% in the two genomes. The three populations evaluated (TB-population 1, TB-population 2, and Be-exposed control population) are presented separately. Open circles, control subjects; closed circles, TB patients; grey circles Be-exposed control population. p value has been determined by Student’s t-test. 26 S. Contini, M. Pallante, S. Vejbaesy, M. Hee Park, N. Chierakul, et al controls in both study populations suggest that TB patients present may be affected by a deficient capability of recognition of the MTB proteome compared to the non affected subjects in the population. Interestingly, Delgado et al. have recently reported a highly significant association between progressive pulmonary TB and homozygosity for HLADQ beta57-Asp alleles where a single polymorphism in the HLA-DQ beta chain played a critical role in the binding of ESAT-6, a highly immunogenic MTB protein and in the ensuing CD4+ T-cells immune response. Although they do not explain the mechanism of susceptibility to the development active TB, these data provided a functional link between an HLA polymorphism and susceptibility to progressive tuberculosis infection (31). In contrast to infection, current concepts are that in hypersensitivity and
autoimmune diseases susceptibility is associated with excessive HLA binding or the binding of specific (neo)-antigens by the HLA allele associated to the diseases (32).

In this context, it has been shown that the immune reaction to non-tuberculous mycobacteria can be characterized by an exaggerated reaction leading to the hypersensitivity pneumonias of the hot tub lung (33), to the metalworking fluid-associated hypersensitivity pneumonitis (34), or to the formation of granulomas within the bronchial walls leading to the formation of bronchiecstasis (35), a condition that’s been associated with HLA-DR 6, i.e., the alleles 13 and 14 (36). Mycobacteria have also been implicated in sarcoidosis, where acid fast rods have been seen in affected tissues (37), wall-deficient form (l-form) of mycobacteria have been isolated (38), and mycobacterial DNA has been detected (39). Interestingly, a reaction to mycobacterial antigens has been described in sarcoidosis as T-cell and antibody responses to MTB ESAT-6 and KatG protein (40), HSP70 (41), and superoxide dismutase (SodA) (42) have been observed in patients with sarcoidosis and some antigens such as the Heat Shock Protein (HSP) and Catalase-Peroxidase (KatG) have been detected by immunohistochemistry in sarcoid tissues (43, 44). It is worth noticing in this regard, that the same HLA allelic variants which have been associated with susceptibility or resistance to tuberculosis have also been implicated in susceptibility to sarcoidosis. The HLA-DRB1*03 alleles, which have been negatively associated with TB, being therefore dubbed as “resistance” genes (45), have been associated with acute, self limiting, sarcoidosis (46). In contrast, the HLA-DRB1*15 alleles, which have been associated with the susceptibility to develop active TB (45), have been associated with stage III, or chronic sarcoidosis (47). Thus, one might hypothesize that the lack of HLA binding and presentation of mycobacterial antigens could determine disease progression in tuberculosis as well as in sarcoidosis, the type of reaction –extensively necrotizing versus non-necrotizing, being possibly driven by the expression of allelic variants of the host’s genes of the innate immune response (48) or by variants expressed in the infected organism of the parasite virulence genes (49). The virtual approximation to truth made by in silico models, together with the limitation imposed by the reduced number of HLA-DR phenotypes that can be analyzed in the different populations, require that in silico results be confirmed by studies using conventional biological techniques. With this caveats, is conceivable that the use of immuno-informatic tools for the prediction of T-cells epitopes also on other HLA class I and II alleles on data sets as large as an the entire MTB genome, will help with generating mechanistic hypothesis on the determinants of HLA-associated genetic susceptibility to TB, as well as to other granulomatous disorders caused, or induced, by mycobacteria.
References


DESIGN OF IMMUNOGENIC PEPTIDES FROM M. TUBERCULOSIS GENES EXPRESSED DURING MACROPHAGE INFECTION

Abstract
In vitro diagnosis of MTB-infection uses MTB proteins coded for by genes of the region of differentiation 1 (RD1) of the MTB genome. This study wants to test if proteins preferentially expressed by during MTB intracellular growth might provide new targets for the diagnosis of MTB infection.

To this end Seventy-five multiepitopic HLA-promiscuous MTB-peptides were designed by quantitative implemented peptide binding motif analysis from 3 MTB-protein genes expressed in activated human macrophages (MA), 4 genes expressed during growth in non-activated human macrophages (MN-A), 12 housekeeping genes (HKG) and 6 genes of the RD1 region (RD1) as control. ELISpot for IFN- was performed to measure the responses of PBMCs deriving from 45 patients affected by active Tuberculosis and 34 controls. In active TB patients, the mean response to RD1 derived peptides was higher than that to either MA (p<0.01), MN-A (p<0.008) or HKG (p<0.01) derived peptides. In TST-positive subjects all selected peptides elicited significant IFN- T-cell responses (p<0.02 compared to TST-negatives), but without differences between the subgroups. Further, T-cell responses to RD1 peptides were lower in the 23 active-TB treated patients than in the untreated ones (p<0.01). The response to MA peptides in treated active-TB was higher than when untreated (p<0.01). These results demonstrate that the use of in vitro models of MTB-intracellular infection to select MTB gene products for further in silico and in vitro assessment of their immunogenicity has the potential to identify novel antigens amenable to the design of new tools for diagnosis and monitoring of tuberculosis.

KEYWORDS:
Tuberculosis, peptide binding motifs, ELISpot, macrophage-induced mycobacterial genes.

34
INTRODUCTION

*Mycobacterium tuberculosis* (MTB) is a facultative intracellular pathogen endowed with the ability to adapt to different growth milieus, either in cell culture or synthetic media, by altering its transcriptional profile. The ability of MTB to evoke a strong type-1 immune response, hence allowing the rapid detection of MTB-specific interferon-gamma (IFN-γ producing T-cells, has been exploited to design *in vitro* diagnostic blood tests for MTB latent infection (LTBI). Following the demonstration that proteins coded for by genes of the region of differentiation 1 (RD1) of the MTB genome i.e., the early secreted antigenic target 6-kDa protein (ESAT-6) and the culture filtrate protein (CFP-10), can stimulate IFN- γ release by blood T-cells in 81 to 97% of tuberculosis (TB) affected individuals, IFN-γ -based blood tests have been designed for the diagnosis of LTBI. In this context, it is reasonable to hypothesize that proteins preferentially expressed by MTB in models of intracellular growth and infection might be ideal targets for the design of new, highly sensitive diagnostic tests of MTB infection and, possibly, for vaccine design. However, a strong limitation in evaluating the efficacy of T-cell response-based assays is represented by the restriction of the host HLA haplotypes. The “reverse immune-genetics” approach by using the quantitative implemented HLA peptide-binding motifs has been successfully applied for the identification and selection of HLA-restricted peptide epitopes. It is worth noticing that the peptide binding motifs are now available for a large number of HLA alleles covering more than 90% of the HLA haplotypes present in different human populations. The present study was designed to determine whether the selection of MTB proteins upon their expression in different conditions of intracellular growth in *in vitro* models might help with the identification of immunogenic epitopes amenable to *in vitro* diagnosis of LTBI. To this end, a panel of HLA-restricted, promiscuous multiepitopic MTB-peptides, identified by quantitative implemented HLA peptide-binding motifs, has been evaluated in a study population including subjects of MTB exposed and not-exposed controls, BGC and not-BCG vaccinated, and patients with active TB.
MATERIALS AND METHODS

Study design
To measure the *ex vivo* immune response to the products of MTB genes expressed within activated and non-activated human macrophages, or belonging to RD1 proteins, as control, a set of antigenic peptides were designed by bioinformatic analysis and tested in patients with active tuberculosis as well as in subjects infected and not with MTB. These responses were also measured in active tuberculosis during standard chemotherapy.

Study population
The study population included 22 patients with newly diagnosed, untreated active pulmonary TB and 23 TB patients undergoing standard treatment for 2 to 4 months, 10 of whom were evaluated both at presentation and after 2 months of treatment. Thirty-four healthy individuals, 9 tuberculin skin test (TST)-negative subjects without any history of TB exposure (hereafter referred to as TB unexposed controls) and 25 TST-positive health-care workers, with a history of previous or current professional exposure (hereafter referred to as TBexposed controls) were also evaluated. Subjects were recruited at the Division of Respiratory Diseases of the University of Rome “Tor Vergata” at the Spallanzani Hospital of Roma (Italy), the Forlanini Hospital of Rome (Italy) and at the National Center of Infectious and Parasitic Diseases (University Hospital for Lung Diseases "St. Sofia"; Sofia, Bulgaria) after informed consent. The diagnosis of active TB was confirmed by *M. tuberculosis* culture isolation in all cases. A detailed analysis of the main demographic features, including the BCG vaccination status and the TST results, of each study group is shown in Table 1. An EDTA-peripheral venous blood sample was collected from all the participants of the study. Epitope prediction and design of antigenic peptides Epitope prediction and design of antigenic peptides was performed by quantitative implemented HLA peptide-binding motif analysis as previously described with minor modifications. Briefly, the sequence of MTB proteins belonging to the RD1 genomic region and of proteins expressed during MTB growth in human macrophage, was obtained using the Swiss-Prot database. Each protein sequence was then analyzed by quantitative implemented HLA peptide-binding motif analysis using the ProPred database (http://www.imtech.res.in/raghava/propred/page2.html), which allowed the identification and scoring of all putative epitopes for any of the 51 considered HLA-class II alleles. Finally, to verify the specificity to MTB, the sequence of each pre-selected peptide was screened for homology of all known protein sequences using BLAST (http://www.ebi.ac.uk/blast/) through the selection of peptides that are restricted to the MTB complex with an identity higher than 85% respect to other mycobacteria and higher than 75% respect to
other microbes. Table 2 shows the list of the peptides evaluated in the study. Each selected peptide included: (i) one or more epitopes able to bind at least four different HLA-DR alleles, with a putative binding capability equivalent to that of the 3% of the best binding peptides for any allele belonging to a given HLA-class II specificity and (ii) a flanking region of 2 aminoacids at the N- and C-terminus to increase the peptide binding affinity14.

**Peptides**

Designed peptides were synthesized by Fmoc chemistry as free amino acid termini with ABI technology (http://www.abi.it/). All peptides were purified by reverse-phase chromatography * At the time of the selection analysis the Mycobacterium marinum, Mycobacterium ulcerans and Mycobacterium smegmatis genomes were not available yet. However, the homology data presented in table 2 report also the identity to these mycobacteria although for some peptides the homology is higher than the 85% used as threshold during the selection steps. (RPC) to >90% purity. The sequence and purity were confirmed by mass spectrometry and analytical RPC. Lyophylized peptides were diluted in DMSO at stock concentrations of 10 mg/mL for each peptide and stored at −80°C till use.

**Cell isolation and culture conditions**

Peripheral blood mononuclear cells (PBMCs) were obtained by Ficoll separation and resuspended in complete medium [RPMI, 10% heat inactivated fetal calf serum (FCS), 10 mM HEPES, 2 mM L-glutamine and 10U/ml penicillin-streptomycin, all from Euroclone Ltd, United Kingdom]. Then, 2x10^5 PBMCs per well were incubated in duplicate in a 96-well plate and stimulated with: (i) 10µg/ml PPD (batch RT47; Statens Serum Institut, Copenhagen, Denmark); (ii) 5µg/ml Phytohemagglutinin (PHA) (Sigma, St Louis, MO, USA); and (iii) 18 pools of the 75 selected peptides. Briefly, to assess the T-cell response to a given peptide, 18 pools was prepared in a square matrix fashion set up with 9 rows and 9 columns pools. Each pool was prepared to contain 8 or 9 peptides (a part pool 9 that contained only 2 peptides) so that each single peptide (final concentration of 5µg/ml) was tested in two different pools, once in a row-grouped pool and once in a column grouped pool in order to determine the specific contribution to the T cell response of each single peptide as previously described15. Both unstimulated and DMSO-treated (at the same concentration present in the peptide solution) PBMCs were included as negative controls. Data were rejected if the sum of the peptide pool result in in columns and in rows of the square matrix presented a difference above 10%. Evaluation of peptide specific single T-cell IFN-γ release. The enumeration of peptide specific T-cell producing IFN-γ was performed at the single cell level using a ELISpot assay, as previously describeds. The number of spot-forming cells (SFCs) was measured in each well using an EliSpot reader (AID EliSpot reader system ELR03; Autoimmune Diagnostika
GmbH, Strassberg, Germany). Evaluated spots had a size >15U (1U=500μm²). The average number of SFCs of duplicate wells was then calculated and the value of appropriate controls subtracted from the SFCs of the stimuli. The absolute value of the SFCs for a single peptide was obtained by resolving the square matrix with the tested peptide pools as previously described. Further, in some subjects the response to single peptides were tested individually to confirm the results of the matrix analysis method. An overall concordance above 90% between the methods were observed.

Statistical analysis Data are expressed as the mean ± standard deviation (SD). Comparison between groups was made using Mann-Whitney tests. A p value below 0.05 was considered significant. All tests were performed using the GraphPad Prism 4.0 (Graphpad software, San Diego, CA, USA) software package.
RESULTS

Peptide selection

Out of 25 MTB proteins, three expressed during growth in activated human macrophages (hereafter grouped as MA), four expressed during growth in non-activated human macrophages (hereafter grouped as MN-A), twelve coded for by constitutively expressed housekeeping genes (hereafter grouped as HKG) and six RD1 region proteins (hereafter grouped as RD1, which were used as comparators), 75 peptides were selected and used in this study (table 2). Specifically, 8 peptides were selected from the MA, 5 peptides from the MN-A, 47 from the HKG and 15 from the RD1 proteins respectively (Table 2). Each peptide contain one or more HLA-DR restricted epitopes (Table 2). Each peptide, in its overall sequence, is predicted to bind HLA-DR alleles belonging at least to three different serological groups of HLA-DR1, 3, 4, 7, 8, 11, 13 and 15, with an affinity equivalent or higher than the 3% of to the best binder epitopes known for the HLA-DR alleles themselves (Table 2). Further, BLAST analysis for these multiepitopic peptides reveals a sequence identity with other mycobacteria varing from 100% to lower than 50% with an identity respect to other bacteria and viruses lower than 75%. Analysis of human T-cell responses to in silico selected peptides To evaluate the cell-mediated immune response to in silico selected peptides, the number of IFN-γ producing peptide-specific T-cell precursors in PBMCs of subjects with active pulmonary TB, in comparison with MTB-exposed and non-exposed healthy controls were determined by ELISpot. Further, the effect of standard chemotherapy on the response against these peptides were assessed. In its overall, 16 out of 75 peptides (n: 6, 7, 21, 25, 26, 28, 29, 33, 34, 38, 39, 40, 41, 49, 59 and 66) did not show any significant difference of response between the three study groups. For the remaining 59 peptides, patients with active TB and TB-exposed controls presented a mean number of SFU/million PBMC significantly higher than the TST-negative controls (Figure 1). Analysis of T-cell responses to selected peptides in active TB patients as positive control. The T-cell responses in active TB subjects presented a greater intensity of response compared to MTB-unexposed controls [TST-negative 1.4+2.1 mean spots per peptide/million PBMCs; active-TB 12.6+9.8; p<0.006 (Figure 1, panel A)]. Further, in active TB patients, the average intensity of the responses to either MA (7.1+3.5), MN-A (6.9+1.1) or HKG (8.2+3.2) derived peptides was lower that that of RD1 derived peptides (21.3+10.8; p<0.01 all comparisons) (Figure 1, panel A). Analysis of T-cell responses to selected peptides in TB-exposed subjects. TST-positive controls, elicited a significantly higher IFN-γT-cell responses against the in silico selected peptides (MA 7.6+3.5; MN-A 8.9+3.5; HKG 7.9+4.5; RD1 12.9+6.8) than TST-negative (p<0.02; Figure 1, panel B). As already shown with ESAT-6 and CFP-10 multiepitopic promiscuous peptides[15,17] the average
intensity of the T-cell responses to RD1 derived peptides was lower in TB-exposed TST-positive subjects (12.9+6.8) i.e., in the LTBI-affected, compared to active-TB patients (21.3+10.8; p<0.003). To the contrary, there was no difference in the response to MA, MN-A and HKG derived peptides in the TST-positive compared to the active TB patients affected (p>0.05, all comparisons). Analysis of T-cell responses to selected peptides during the course of active TB. In the context of the observation that the response to ESAT-6 and CFP-10 multiepitopic promiscuous peptides in active TB patients declines during chemotherapy, T-cell responses to the MA, MN-A and HKG derived peptides were also analysed in subgroups of untreated and of treated active-TB patients, as well as in a selected group of active TB patients which were tested before and after two months of standard chemotherapy. As expected, the T-cell responses to RD1-derived peptides were higher in the 22 untreated active-TB treated patients (mean total RD1 peptides SFC 251+116) compared to the 23 chemotherapy treated ones (mean total RD1 peptides SFC 155+129, p<0.01 compared to untreated patients). Strikingly, we found an increased response to MA peptides in chemotherapy treated patients (mean total MA peptides SFC 121.8+44.6) compared to the untreated (mean total MA peptides SFC 52+29; p<0.01), while no differences were found for the MN-A (untreated active-TB mean total MNA peptides SFC 38+12; active-TB treated 33+9; p>0.05) and HKG-derived peptides (untreated active-TB mean total HKG peptides SFC 389+45; active-TB treated 306+78; p>0.05). This observation was confirmed in the ten active TB patients which were followed up during the two months of chemotherapy treatment. In eight out of ten patients IFN-γ release in response to RD1-derived peptides markedly declined over time (figure 2 panel A, p<0.02). Instead, it did strikingly augment in 9 out of ten (figure 2 panel B, p<0.01) in response to MA-derived peptides. As above, no change was seen with MN-A or HKG-derived peptides (p>0.05, all comparisons).
DISCUSSION

The finding that the measure of IFN-γ-release as marker of type-1 immune response, in response to MTB-proteins is capable of identifying subjects with MTB-infection, either latent or active, has led to the production of a number of tests such as the RD1-derived antigen based ELISA “Quantiferon-TB” and the ELISpot “T-spot TB” tests. These tests, that are endowed with sensitivities ranging from 81 to 97% in the identification of patients with active MTB proliferation in tissue lesions, have spurred microbial immunology and molecular biology research into MTB immunogenic proteins. However, critical aspects in this context are represented by the choice of immunodominant and specific antigens and, for T-cell response evaluation, in the restriction imposed by the patients' HLA background. In this context, the use of synthetic peptides in TB has been successfully demonstrated for diagnostic applications in humans. Overlapping synthetic peptides covering the entire sequence of a protein have been used to identify the peptides recognized by T-cells. However, since Th1 cells recognize mycobacterial antigens and epitopes in association with MHC class II molecules, an alternative approach has been to screen proteins for identification of regions that can associate with MHC molecules and then test the peptides predicted to bind the MHC molecules for T-cell reactivity. Such approach has been shown to reduce drastically the number of peptides to be tested and thus the cost to screen for T-cell reactivity. Although several prediction programs have been proposed to identify peptides capable of binding to MHC class II molecules, recently a virtual matrix-based prediction server (ProPred) has been described, which can predict the binding capability of peptides to 51 HLA–DR alleles and has been shown to identify epitopes with a confidence of 95%. Although these prediction systems are far from being fully precise, the result of this study confirm that the \textit{in silico} selected HLA-promiscous multiepitopic area of proteins can be the region presenting the highest capability of immune response, as shown comparing the result of this study with classic epitope mapping for some of the RD1 proteins such as ESAT-6, CFP-10, Rv3873 and Rv3878. Furthermore, previous studies from our laboratories have shown that MTB is capable of expressing differential transcription programs in response to different environments. In particular, MTB can express starkly different gene sets in synthetic media compared to macrophage cultures. RD1 proteins are expressed both in early and late phase cultures, are variably expressed \textit{in vivo} in murine models, but are expressed at very low level in macrophage cultures.

With this background, we hypothesized that the MTB proteins expressed at the highest level in macrophage cultures could induce a stronger T-cell response in MTB infected subjects. The result of this study show that a group of peptides derived from a panel of MTB proteins selected for their expression in macrophage cultures are highly immunogenic, as they seem capable of eliciting Th1
T-cell responses in LTBI affected subjects that are as high as those elicited by RD1 derived peptides i.e., the present day diagnostic standard. A number of studies have analyzed the in vivo or ex vivo mycobacterial gene expression profile17. However, the inter-relationship between RNA and protein expression profiles, immunogenicity, either in silico or in vitro of the expressed proteins and the induction of T-cell responses has not been yet investigated in sufficient depth. Further, this study confirm the immunogenicity of the RD1 proteins in their overall and not only confined to ESAT-6 and CFP-10. In fact, only 4 out of the 15 RD1 derived peptides were designed on ESAT-6 and CFP10 and these peptides account only for 19-43% of the response to the whole RD1 peptides (data not shown). Furthermore, this study identified 19 proteins which elicit a strong T-cell response in LTBI and TB affected subjects as high as in RD1 peptides. However, few of these proteins have been characterised so far. Specifically, none of the MA proteins has been immunologically characterized. The MA proteins: Rv3042 and Rv1114; the M-NA proteins: Rv0198c, Rv1382 and Rv0755; the HKG proteins Rv1582, Rv0418, Rv1434, Rv1866, Rv0956 and Rv3219 have not yet been cloned and expressed for biochemical and functional studies34. The MA protein Rv3883c (MycP1) is a membrane-anchored mycosin serine protease35 and Rv3042 (serB2) is a phosphoserine phosphatase34. The only characterized M-NA protein is Rv0125 (PepA), another protein belonging to the MTB serine protease gene family, is also known as MTB32A and has been found in culture filtrates. Consistent with our data, PepA is immunogenic as it has been shown to elicit memory T-cell responses in latent and active TB36,37. Among the 12 HKG proteins evaluated in the study, three are known to be immunogenic in an animal model and two in an in silico model. Rv1979c and Rv1986, that’s belong to the RD2 gene region and Rv1769 that’s belong to the RD14 gene region, have been shown to elicit effector T-cell responses in 59-86% of M. bovis-infected and/or BCG-vaccinated cattle38. Interestingly, two other HKG gene products, Rv1983 and Rv0747, belong to the PE-PGRS gene family34 i.e., a multigene family that, together with the PPE gene family, makes up about 10% of the MTB genome and whose gene products comprise an extraordinarily high number of peptides displaying high-affinity for HLA class I alleles in silico39 supporting the hypothesis of their immunogenicity and potential role in immunodiagnosis as suggested by our data by the reports on other proteins belonging to these multigene families10,40,41. Further, the comparison between the time course of the responses elicited by the MA, MN-A and the HKG derived peptides and that against RD1-derived peptides in active TB patients, whereby the response to RD1 peptides declines with the reduction of active lesion subsequent to treatment, may suggest that the IFN-γ response to certain RD1 peptides may be mycobacterial load, hence antigen concentration, dependent and may wane with the its rapid mycobacterial load reduction achieved by standard chemotherapy15,18. Interestingly, the response to MA, MN-A and the HKG derived
peptides seems to be more stable or may increase with time, as shown for MA-derived peptides. In conclusion, the results of our study suggest that promiscuous peptides capable of Th1 cell reactivity in HLA heterogeneous population can be predicted from the in silico HLA-DR binding analysis of complete proteins overexpressed in model M. tuberculosis infections in macrophage. It has been predicted that the number of putative T-cell epitopes in M tuberculosis are about 1,500,000, which can be reduced efficiently reduced to few hundreds peptides by identifying proteins overexpressed during MTB infection in human and prescreening the coding sequences for HLA-DR binding using the prediction programs. This has the potential to identify novel antigens eventually be amenable to the design of new diagnostic tools.
REFERENCES


Figure 1.

IFN-γ response to the 75 multiepitopic HLA promiscuous MTB peptides obtained from 24 MTB proteins. Each column represents the enumeration of IFN-γ producing cells induced by one selected peptide. The measurement was performed by a ELISpot assay in active TST-negative (panel A), TST-positive (panel B) and TB patients (panel C). Data are reported as the mean of the SFCs per million PBMCs for each peptide. Peptides are grouped as belonging to genes expressed in activated human macrophages (MA), expressed during growth in non-activated human macrophages (MN-A), housekeeping genes (HKG) and RD1 region (RD1). Black bars represent peptides showing responses with significant differences between the three study groups (p<0.05) while open bars represent those did not. Peptides derived from CFP-10 (*) and ESAT-6 ($) are represented in the last four positions.
Figure 2.

Evaluation of the *in vitro* response to RD1 derived peptides (panel A) and to peptides designed from genes expressed in activated human macrophages (MA, panel B) at the time of diagnosis of active-TB and after 2 months of standard anti-tuberculous therapy in a subgroup of active-TB patients. Each study subject is represented by one line and the ordinate represents the sum of the number of IFN-γ producing spot forming cells (SFC) per million PBMCs in response to the peptides of each group.
# Table 1. Study group

<table>
<thead>
<tr>
<th>STUDY GROUP</th>
<th>n</th>
<th>Mean Age (years)</th>
<th>M/F</th>
<th>Nationality</th>
<th>Vaccination</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTB-Not Exposed TST-negative controls</td>
<td>9</td>
<td>30</td>
<td>5/4</td>
<td>Italian</td>
<td>0/9</td>
</tr>
<tr>
<td>MTB-Exposed TST-positive controls</td>
<td>25</td>
<td>46</td>
<td>13/12</td>
<td>Italian</td>
<td>18/25</td>
</tr>
</tbody>
</table>
| Untreated active-TB patients          | 22 | 39               | 13/9 | 3 Bulgarian  
  9 Italian  
  7 Romanian  
  1 Brazilian  
  1 Moldavian  
  1 Polish | 12/21  
  (1 not known) |
| Treated active-TB patients          | 23 | 38               | 20/3 | 10 Bulgarian  
  2 Italian  
  2 Romanian  
  1 Albanian  
  1 Macedonian  
  1 Egyptian  
  1 Indian  
  1 Uruguayan  
  1 Somali  
  1 Colombian  
  1 Chinese  
  1 Filipino | 15/18  
  (5 not known) |
<table>
<thead>
<tr>
<th>#</th>
<th>MIB protein</th>
<th>Ga accession number</th>
<th>AA starting position in protein sequence</th>
<th>PEPTIDE SEQUENCE</th>
<th>HLA-DR GROUP VARIANTS RECOGNISED BY THE EPITOPES PRESENTED IN THE PEPTIDE</th>
<th>HOMOLOGY BY BLAST*</th>
<th>GROUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rv3883c</td>
<td>156111019</td>
<td>1*</td>
<td>ALA ALA ALA ALA</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae 79%</td>
<td>MA</td>
</tr>
<tr>
<td>2</td>
<td>Rv3883c</td>
<td>156111019</td>
<td>176</td>
<td>GULRPLAVAAALNQGOV</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae 79%</td>
<td>MA</td>
</tr>
<tr>
<td>3</td>
<td>Rv3883c</td>
<td>156111019</td>
<td>219</td>
<td>IDIAAVKVGKVRVAVG TGOAGDC</td>
<td>1, 3, 4, 7, 8, 11, 13</td>
<td>M. leprae and M. tuberculosis 80%</td>
<td>MA</td>
</tr>
<tr>
<td>4</td>
<td>Rv3883c</td>
<td>156111019</td>
<td>257</td>
<td>DOLAYTVGVDAVA</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae and M. tuberculosis 78%</td>
<td>MA</td>
</tr>
<tr>
<td>5</td>
<td>Rv3883c</td>
<td>156111019</td>
<td>472</td>
<td>APDANVRGLVAGPVP</td>
<td>1, 3, 4, 8, 11, 13</td>
<td>Oryza sativa 66%</td>
<td>MA</td>
</tr>
<tr>
<td>6</td>
<td>Rv3043c</td>
<td>156101017</td>
<td>157</td>
<td>LCVQNPSPCPPCQLTV KAAE</td>
<td>1, 3, 4, 8, 11, 13</td>
<td>M. leprae 82%</td>
<td>MA</td>
</tr>
<tr>
<td>7</td>
<td>Rv1114</td>
<td>15608254</td>
<td>46</td>
<td>RPQVTRLLANVYEQPVV</td>
<td>3, 4, 8, 11, 13</td>
<td>Rodent ehrlichia sp. 64%</td>
<td>MA</td>
</tr>
<tr>
<td>8</td>
<td>Rv1114</td>
<td>15608254</td>
<td>60</td>
<td>FTVTVTHDEFLVVEQ URLWVG</td>
<td>1, 3, 7, 8, 11, 13, 15</td>
<td>M. paratuberculosis and J. abdomini 28%</td>
<td>MA</td>
</tr>
<tr>
<td>9</td>
<td>Rv0135</td>
<td>15607267</td>
<td>218</td>
<td>GQGQMVMTAADDNFP</td>
<td>1, 3, 4, 8, 11, 13</td>
<td>M. leprae 83%</td>
<td>MA-N</td>
</tr>
<tr>
<td>10</td>
<td>Rv0198c</td>
<td>15607259</td>
<td>56</td>
<td>TGVQALDOAQAGQAA</td>
<td>1, 3, 4, 11, 13</td>
<td>M. leprae 10%</td>
<td>MA-N</td>
</tr>
<tr>
<td>11</td>
<td>Rv0142c</td>
<td>15608512</td>
<td>9</td>
<td>GQGLADVYKAMQKLA RNVNGW</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae 78%</td>
<td>MA-N</td>
</tr>
<tr>
<td>12</td>
<td>Rv1182</td>
<td>15608522</td>
<td>110</td>
<td>RQAVGKVVRGGLAHKLP SG</td>
<td>1, 7, 8, 11, 13</td>
<td>M. leprae and M. tuberculosis 75%</td>
<td>MA-N</td>
</tr>
<tr>
<td>13</td>
<td>Rv0755c</td>
<td>57116777</td>
<td>107</td>
<td>VYQHPAVQPCR TLLASV NFGQ</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae and M. tuberculosis 59%</td>
<td>MA-N</td>
</tr>
<tr>
<td>14</td>
<td>Rv1582c</td>
<td>15608720</td>
<td>143</td>
<td>LDDLNLKLELPHADPEIT</td>
<td>3, 4, 8, 12</td>
<td>Brucella melitensis 50%</td>
<td>KHO</td>
</tr>
<tr>
<td>15</td>
<td>Rv1582c</td>
<td>15608720</td>
<td>254</td>
<td>GQVGDGLKQRVAYVSKER</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae 92%</td>
<td>MA-N</td>
</tr>
<tr>
<td>16</td>
<td>Rv1769</td>
<td>15608070</td>
<td>131</td>
<td>AFVPMDQSVLSLE</td>
<td>1, 3, 4, 7, 8, 11, 15</td>
<td>M. leprae 100%</td>
<td>MA-N</td>
</tr>
<tr>
<td>17</td>
<td>Rv1769</td>
<td>15608070</td>
<td>215</td>
<td>DNQAERGERMAHMRQ MFE</td>
<td>1, 3, 7, 8, 11, 15</td>
<td>M. leprae 85%</td>
<td>KHO</td>
</tr>
<tr>
<td>18</td>
<td>Rv1769</td>
<td>15608070</td>
<td>265</td>
<td>PRLDEGQKVRTHTKAGLCE</td>
<td>1, 3, 7, 8, 13</td>
<td>M. leprae 85%</td>
<td>KHO</td>
</tr>
<tr>
<td>19</td>
<td>Rv1769</td>
<td>15608070</td>
<td>501</td>
<td>HNYSFPGYFDVSN</td>
<td>1, 3, 4, 11, 15</td>
<td>M. leprae 83%</td>
<td>KHO</td>
</tr>
<tr>
<td>20</td>
<td>Rv0179c</td>
<td>15609116</td>
<td>158</td>
<td>LQVFGQLLLNLGNNRAIKWN</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae 60%</td>
<td>KHO</td>
</tr>
<tr>
<td>21</td>
<td>Rv0179c</td>
<td>15609116</td>
<td>187</td>
<td>GQLQDNLATAWSY</td>
<td>4, 8, 11, 13</td>
<td>M. leprae 71%</td>
<td>KHO</td>
</tr>
<tr>
<td>22</td>
<td>Rv0179c</td>
<td>15609116</td>
<td>207</td>
<td>ATPETTIVGALISMEGQVNAASEGA</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. leprae 92%</td>
<td>KHO</td>
</tr>
<tr>
<td>Peptide</td>
<td>Sequence</td>
<td>HLA-DR group variants</td>
<td>M. avium %</td>
<td>M. bovis %</td>
<td>M. marinum %</td>
<td>M. microti %</td>
<td>M. porcinum %</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>-----------------------</td>
<td>------------</td>
<td>------------</td>
<td>--------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>31 Re1866 15609003 149</td>
<td>AAVQAASAVLAVXAYNLROCGTDQY</td>
<td>1, 7, 8, 11, 13, 15</td>
<td>M. avium 82%</td>
<td>M. bovis 88%</td>
<td>M. marinum 91%</td>
<td>M. microti 83%</td>
<td>M. porcinum 85%</td>
</tr>
<tr>
<td>32 Re1866 15609003 232</td>
<td>DGTVECVMAIPQVRGWRWLGEP</td>
<td>1, 4, 7, 8, 11, 13, 15</td>
<td>M. avium 55%</td>
<td>M. bovis 83%</td>
<td>M. marinum 81%</td>
<td>M. microti 77%</td>
<td>M. porcinum 85%</td>
</tr>
<tr>
<td>33 Re1866 15609003 357</td>
<td>YPFEGLEISLQIVAGGELS</td>
<td>1, 3, 5, 7, 8, 11, 13, 15</td>
<td>M. avium 95%</td>
<td>M. bovis 88%</td>
<td>M. marinum 91%</td>
<td>M. microti 85%</td>
<td>M. porcinum 89%</td>
</tr>
<tr>
<td>34 Re1866 15609003 457</td>
<td>BAYLAVILHAGLASSGAK</td>
<td>1, 3, 5, 7, 8, 11, 13, 15</td>
<td>M. avium 95%</td>
<td>M. bovis 88%</td>
<td>M. marinum 91%</td>
<td>M. microti 85%</td>
<td>M. porcinum 89%</td>
</tr>
<tr>
<td>35 Re1866 15609003 507</td>
<td>IPXXIYQXRERATQKPRVYTSFAQ</td>
<td>1, 3, 5, 7, 8, 11, 13, 15</td>
<td>M. avium 95%</td>
<td>M. bovis 88%</td>
<td>M. marinum 91%</td>
<td>M. microti 85%</td>
<td>M. porcinum 89%</td>
</tr>
<tr>
<td>36 Re1866 15609003 538</td>
<td>TIFDPYVNGQYALLAALHHD</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. avium and M. bovis 84%</td>
<td>M. marinum and M. bovis 84%</td>
<td>M. microti and M. bovis 84%</td>
<td>M. porcinum and M. bovis 84%</td>
<td>M. smegmatis 82%</td>
</tr>
<tr>
<td>37 Re1866 15609003 568</td>
<td>AYHIHSGAEVQGVQDITMVVE</td>
<td>1, 3, 4, 7, 8, 11, 13, 15</td>
<td>M. avium and M. bovis 84%</td>
<td>M. marinum and M. bovis 84%</td>
<td>M. microti and M. bovis 84%</td>
<td>M. porcinum and M. bovis 84%</td>
<td>M. smegmatis 82%</td>
</tr>
</tbody>
</table>

Notes:
- Peptides are grouped according to the capability of the belonging protein to be selectively over expressed during growth in activated human macrophages (MA), expressed during growth in non-activated human macrophages (MN-A), constitutively expressed housekeeping genes (HKG) and belonging to the RD1 region proteins (RD1).
- Peptide epitopes starting from aminoacid position 1 in the protein sequence where two alanines were added to the N-terminus of the peptide in order to increase binding affinity (see also Method section).
- Peptide sequence reported as single letter code. The underlined letter(s) in the peptide sequence indicate the aminoacid acting as P1 of the epitope(s) presented along the peptide sequence as for quantitative implemented peptide binding motif analysis.
- HLA-DR group variants capable to bind in silico the peptide sequence with an affinity equivalent or higher than the 3% of the best binder peptides for the alleles present in the variant.
- Homology analysis of the peptide sequence performed by BLAST reporting reported the first microorganisms not belonging to the M. tuberculosiscomplex with the higher percentage of identity (expressed as percentage) to the selected protein sequence.
IMMUNOGENETICS OF SARCOIDOSIS

*M. avium* binding to HLA-DR expressed alleles *in silico*: A model of phenotypic susceptibility to sarcoidosis

Abstract

Sarcoidosis is a systemic granulomatosis disease of unknown origin where a number of microbes, in particular *M. tuberculosis* and non-tuberculous mycobacteria, have been hypothesized to play a role in disease pathogenesis, possibly through bacterial antigen-driven hypersensitivity. To test this concept, we used bioinformatic tools allowing the identification of antigenic peptides in whole microbial genomes to analyze the interaction between the expressed HLA-DR gene allelic variants and the HLA-DR immunome of all pathogenic bacteria in a population of 149 sarcoidosis affected subjects and 447 controls, all HLA-typed at high resolution.

We show here that patients with the Loefgren’s syndrome, express HLA-DR alleles that recognize *in silico* a significantly higher number of bacterial antigen epitopes compared to the control population (18,496+9,114 vs 17,954+8,742; p<0.00001), and the chronic sarcoidosis affected population (17,954+8,742; p<0.00001 vs Loefgren’s and controls). Further, the analysis of the ability of the HLA-DR allele combinations expressed by the Loefgren’s and the chronic sarcoidosis affected subjects to recognize *M. avium* epitopes demonstrates that a significantly larger number of Loefgren’s are capable of top affinity recognition, compared to chronic sarcoidosis (45% vs 17%, p<0.0037). Finally, both Loefgren’s and chronic sarcoidosis subjects expressed HLA-DR allele combinations capable of *M. tuberculosis* and *M. avium* epitope recognition at higher affinity than tuberculosis affected subjects (p<0.01 all comparisons).

In conclusion, we propose that -- at least in a subgroup of affected subjects-- sarcoidosis might be part of a spectrum of granulomatous responses to several agents where the Loefgren’s syndrome represents the hyper-reactive end of the spectrum while pulmonary tuberculosis and atypical mycobacterial infections might represent the opposite end.
Introduction

Sarcoidosis is a chronic disorder characterized by the accumulation of increased numbers of T-lymphocytes and activated mononuclear phagocytes in the lower respiratory tract and other sites of disease, leading to the formation of noncaseating granulomas in involved tissues such as the lung, skin and eyes. Several lines of evidence suggest that the accumulation of activated T-cells and macrophages in the lower respiratory tract is the result of chronic stimulation by some (yet unknown) antigen(s) (1). Firstly, in sarcoidosis activated alveolar macrophages possess increased ability of presenting antigens to T-cells (2) and express HLA-DR molecules at higher level at the surface of the cell (3). Secondly, T-cells accumulate in the lower respiratory tract where they release interleukin 2 (4, 5) and interferon-γ at exaggerated rates (2). Thirdly, sarcoid lung T-lymphocytes display activated memory T-cell phenotypes (6, 7) and express oligoclonal Vβ T-cell receptor genes as they were locally expanded in response to a classical antigenic stimulus rather than a more broad-based stimulus such as from a superantigen (8, 9).

A number of potential causation agents have been implicated in the pathogenesis of sarcoidosis (10) and some sarcoidosis-like disease and reactions could be identified as hypersensitivities to compounds such as beryllium (11), titanium (12), zirconium (13) and aluminum (14). Recently, it has been shown that, in addition to chemicals, microbial agents otherwise capable of causing infection and disease may be implicated in the pathogenesis of granulomatous hypersensitivity pneumonias such as the hot tub lung, where M. avium has been identified as a causative agent and (15) and metal working fluid pneumonitis, where a variety of bacteria have been implicated (16). Furthermore, in a chronic granulomatous disease such as Chron’s disease M. avium paratuberculosis has been implicated (17).

In sarcoidosis, since the application of the the polymerase chain reaction to the detection of mycobacterial genomes in clinical samples (18), the presence of nucleic acids of M. tuberculosis, non tuberculcous mycobacteria and of Propionibacterium acnes has been observed both in bronchoalveolar lavage and tissue samples in different sarcoidosis patient populations (19-23). Collectively, these studies indicate that mycobacteria can be found in tissue samples in 23 to 30% of sarcoild patients, with a ten-fold higher probability compared to controls, and suggest an association between mycobacteria and sarcoidosis (24-26).

In this context and with the knowledge that certain microbial agents could cause hypersensitivity reactions leading to granuloma formation in the absence of clinical infection, the observations that antibody and T-cell mediated immune responses against mycobacterial antigens, such as catalase-peroxidase (mKatG) (27), heat shock protein (Mtb-hsp) (28), and Antigen 85A (29)
could be demonstrated in sarcoidosis and in the Kweim reaction, led to hypothesize that exposure to certain microbes might be causative of sarcoidosis in hyperreactive individuals (30-32).

Antigen presentation to CD4 T-cells is mediated by a class of peptide cell surface receptors, the HLA class II proteins (HLA-DP, -DQ and -DR), which are capable of binding peptides derived from the digestion of microbes and other agents in the phago-lysosome of antigen presenting cells and to carry them to the cell surface and to present them to the CD4 T-cell expressing the appropriate T-cell antigen receptor. Importantly, the selection of antigenic peptides presented to, hence recognized by, T-cells is driven by the chemico-physical interaction between (i) the amino acid side chains which are lining the receptor-like pockets on the floor of the HLA antigen binding groove and (ii) the aminoacid side chains of the antigenic peptides (33, 34).

The HLA class II genes have been implicated in susceptibility to sarcoidosis (35, 36) and it has been shown that the expression of different HLA allelic variants is associated with different clinical presentations and disease outcomes. The Loefgren’s syndrome, an acute form of sarcoidosis characterized by fever, erythema nodosum, bilateral hilar lymphadenopathy and/or ankle arthritis is associated with the expression of the HLA-DRB1*0301 allele and has a favourable clinical outcome (37). On the other hand, chronic forms of sarcoidosis with less favourable outcomes have been associated with the expression of the DRB1*1501 allele (37).

Key to the understanding of susceptibility to immune disorders is the knowledge that HLA alleles can select peptide antigens for antigen presentation with widely varying affinities, thus leading to different immune responses against the same agent. In this regard, it is well known that since the HLA polymorphisms characterizing allelic variants code for aminoacid changes in the peptide binding groove’s, each HLA allele binds a unique set of aminoacid side chains. Each HLA class II allelic variant is therefore capable of selecting a discrete set of antigenic peptides for antigen presentation (38, 39) and to allow for the formation of unique HLA molecule/T-cell receptor/antigenic peptide complexes activating specific cytokine producing CD4 T-cells (40). Thus, since each individual coexpresses at least two HLA-DR molecules on the cell surface of the antigen presenting cell, it is reasonable to think that individual ability to generate immune responses to a given antigen shall be determined by the capability of each subject’s two HLA-DR molecules combined to recognize antigen epitopes (41-43).

By using bioinformatics tools allowing the identification of antigenic peptides in whole microbial genomes (44), we have recently demonstrated that the expression of HLA-DR allele combinations endowed with a lower affinity for the *M. tuberculosis* genome and for its derived peptide antigens is associated with susceptibility to active tuberculosis (45). Interestingly, the HLA-DR alleles implicated in this reduced ability to bind and select for mycobacterial antigenic peptides
are the same which have also been associated with chronic sarcoidosis. Even more interestingly, the alleles endowed with higher affinity for *M. tuberculosis* antigenic peptides, and with “protection” against tuberculosis, are the same which have been previously associated with the acute forms of sarcoidosis (45).

With the background that the affinity for peptide antigens of the HLA-DR expressed allelic variants plays a key role in determining the ability of the individual to recognize and react to bacterial antigens, hence in generating the ineffective immune response leading to bacterial invasion and active disease, we hypothesized that it may also play a role in determining susceptibility to hypersensitivity reactions to the same bacteria. To test this hypothesis, we set out to analyze the interaction between the expressed HLA-DR gene allelic variants and the genomes of all pathogenic bacterial in the Loefgren’s syndrome and in chronic sarcoidosis, two immunologically contrasting forms of sarcoidosis, in order to determine whether hypersensitivity to mycobacterial, propionibacterial or other bacterial antigens may play a potential role in the pathogenesis of sarcoidosis.
Methods

Patients’ characteristics:

The study population was composed by the patients with sarcoidosis and matched controls of an already described population evaluated for the genetic susceptibility factors in sarcoidosis. HLA-DR high resolution typing was available for all study subjects (35).

They were 149 patients with sarcoidosis, of which 39 with Loefgren’s syndrome and 110 with chronic sarcoidosis, and 447 matched controls (see table 1). The quantitative implemented peptide binding motifs i.e., the algorithms with which to assess the ability of a given HLA-DR alleleic variant, are available only for 49 out of the more than 300 HLA-DRB1 alleles (46, 47) and cover, with at least one allele, about 90% of the HLA-DR variability of different human populations (48). For the purpose of this study, only the subjects carrying both HLA-DR alleles with an available HLA-DR binding motif were included (see table 1). Specifically, 349 out of 447 (78.1%) control subjects, 31 out of 39 (79.5%) patients with Loefgren syndrome and 93 out of 110 (84.5%) patients with chronic sarcoidosis, could be subjected to the analysis.

The selected subgroups did not differ for demographical characteristics from the subgroups of subjects excluded from the study for having one or both HLA-DR alleles without a known binding motif (data not shown).

Finally, a group of patients with active tuberculosis and matched controls already evaluated previously (45) were used as comparison. They include: (i) 303 patients with active tuberculosis of which 92 carrying both HLA-DR alleles with an available HLA-DR binding motif and (ii) 345 controls of which 111 carrying both HLA-DR alleles with an available HLA-DR binding motif (Table 1).

Genomes.

One hundred and twenty four, non redundant, bacterial complete genomes, classified as “relevant human pathogens” in the Genome OnLine Database v2.0 (GOLD database update October 30th 2008, http://www.genomesonline.org/), were used to analyse the number of potential epitopes in each genome which were recognised by the study population. Bacterial genome characteristics (accession code, species, family, human disease and number of open reading frames) are listed in table 2.

Software for the identification and enumeration of epitopes in whole genomes
To enumerate the T-cell epitopes present in the protein data sets of each microbial genome examined, we used a software specifically designed for the identification and enumeration of peptide binding epitopes to HLA-DR molecules in large set of protein databases, as already described (45).

For the purpose of this study, the bacterial genomes were analysed for enumerating the epitopes in silico recognised by the combination of the HLA-DR alleles carried by single subjects at an affinity level equivalent or higher than the 1% of the best binder peptides for the allele in analysis (49). The Threshold of affinity is a preselected numerical value used to differentiate between binders and non binders, any peptide frame scoring higher than this value is predicted as binder or vice versa; it correlates with the peptide score (39, 49) and therefore with HLA-ligand interaction, therefore it is an indicator for the likelihood that predicted peptide is capable of binding to a given HLA-molecule.

Statistical analysis.

All the data are expressed as mean ± standard deviation of the mean (SD), and frequency expressed as percentage. Interquartile distribution of epitope number recognised was used to classify allele combination. Comparisons between means are made by Student’s t test. Comparison between frequencies is made by Chi-square, with Yates and Bonferroni’s correction, Chi-square for trend and Fisher’s exact test when appropriate. GraphPad Prism version 4.0 was used for all statistical analyses and graphs.
Results

Differential capability of chronic sarcoidosis and Loefgren’s syndrome affected subjects to recognize genomes of human pathogen.

The capability of the HLA-DR expressed allelic variants of the study subjects to bind bacterial peptide epitopes was determined as the number of epitopes bound in silico at the top 1% affinity. At an affinity level equivalent to that of the top 1% of the binder peptides for the HLA-DR molecules of the study subjects, Loefgren syndrome affected subjects recognized a higher number of epitopes (mean epitope number for all the genomes: $18,496+9,114$) than chronic sarcoidosis affected subjects ($17,515+8,448$; $p<0.00001$) and controls ($17,954+8,742$; $p<0.00001$). In addition, the comparison between chronic sarcoidosis and healthy controls showed that the capability of affected subjects to recognize bacterial pathogen genomes was significantly reduced compared to controls ($p<0.00001$).

Interestingly though, the capability of the HLA-DR allelic variants expressed by Loefgren’s syndrome affected subjects to recognize a higher number of epitopes than chronic sarcoidosis affected subjects was not generalized to all the genomes examined, but it was more prominent for certain bacterial pathogens. In particular, statistically significant differences in the mean number of epitopes recognized by Loefgren syndrome affected subjects with respect to chronic sarcoidosis affected subjects was observed for the genomes of bacteria belonging to the genders Bordetella, Burkholderia, Granulibacter, Mycobacterium, Nocardia and Propionibacterium (table 2 and figure 1). The largest difference was observed for *Mycobacterium avium*, where there was 12% difference between the average HLA-DR immunomes, i.e. the number of putative epitopes deduced from the whole bacterial genome that were capable of being bound by the HLA-DR expressed allele of each individual subject, of the two disease groups (figure 1).

Contribution of the different HLA-DRB1 alleles combination in the hyper-recognition of genomes by patients with Loefgren’s syndrome.

In order to analyze the nature of the exaggerated epitope recognition of bacterial genomes seen in the subjects with Loefgren’s syndrome compared to those with chronic sarcoidosis, the data obtained in the study population for the HLA-DR immunome of *M. avium* i.e., the best binder of all tested genomes, were used.

The distribution of HLA-DR allelic variants in the study populations is shown in table 3. As previously reported (37), alleles HLA-DRB*0301 and *0407 were significantly more frequent in
the Loefgren’s syndrome affected while the allele *1501 was significantly over-represented in the chronic sarcoidosis affected group (table 3).

Table 4 shows the distribution of M. avium genome recognition levels in the study groups, according to the percentile distribution from bottom to top recognition levels. Considering the affinity level equivalent to 1% of the higher affinity peptides, the 1,225 different combinations generated by the 49 HLA-DRB1 alleles for which the quantitative implemented peptide binding motifs is available, can recognize 6,149 to 43,781 epitopes in the genome of M. avium. Subjects with Loefgren’s syndrome showed significantly higher levels of M. avium epitope recognition than subjects with chronic sarcoidosis as 45% of the Loefgren’s affected were in the top 76-100° percentile compared to 17% of the chronic sarcoidosis affected (p=0.0037) and 26% of the controls (p=0.0386). Conversely, subjects with chronic sarcoidosis showed a significantly higher level of M. avium epitope recognition in the lower 26°-50° percentiles distribution with 39% compared to the 16% of the Loefgren’s affected (p=0.0363) and the 24% of the controls (p=0.0059; table 4).

Furthermore, in order to determine whether the exaggerated recognition of microbial genomes such as those of mycobacteria, shown by Loefgren’s syndrome affected subjects, was associated with the expression of specific HLA-DR allelic variants rather than with the expression of combinations of HLA-DR alleles endowed with higher binding capabilities, we analyzed the distribution of HLA-DR alleles in the M. avium recognition percentile subgroups.

This analysis showed that there were no biases in the expression of HLA-DR variants in the Loefgren’s syndrome affected population, neither in the 26-50° nor in the 76-100° percentile groups (table 5). Only the association of the HLA-DRB1*1501 allele with chronic sarcoidosis remained statistically significant both in the higher and the lower affinity percentile groups (table 5).

Assessment of HLA-DR M. avium and M. tuberculosis epitope affinity in sarcoidosis and tuberculosis.

Altogether the data suggests that Loefgren’s syndrome immune responsiveness might be characterized by the expression of HLA-DR genomes endowed with an exaggerated ability to recognize bacterial genomes while, in contrast, chronic sarcoidosis appears to be characterized by the expression of HLA-DR alleles with significantly lower ability to bind and present bacterial antigens. As already reported in tuberculosis affected patients a biased expression of the same HLA-DR alleles expressed in sarcoidosis could be observed (45).

Strikingly, the comparison of epitope binding affinity of the HLA-DR allelic combinations expressed by sarcoidosis affected subjects with those of tuberculosis patients showed a statistically significant lower binding affinity, both toward the M. avium and the M. tuberculosis immunomes
than that of sarcoidosis affected subjects. In addition, when the patient groups were ranked by their binding affinities, the Loefgren’s affected ranked the highest, with the tuberculosis affected the lowest. Chronic sarcoidosis, although ranking lower than controls ranked higher than tuberculosis (table 6), thereby indicating that in tuberculosis the susceptibility to infection and active disease is likely associated with the expression of a defective epitope recognition repertoire.
**Discussion**

We show here that patients affected by the Loefgren’s syndrome express HLA-DR allelic combinations that recognize a significantly higher number of bacterial antigen epitopes compared to healthy individuals, as they do in comparison to chronic sarcoidosis.

With regard to sarcoidosis, the data are consistent with the observation of Grosser et al. (50) who recently reported the association of chronic sarcoidosis, the presence of mycobacterial DNA in affected tissue and the expression of DRB1*15, whereas acute sarcoidosis was associated with the absence of mycobacterial DNA and with the expression of HLA-DRB1*03. These data add further support to the concept that the manifestations and outcomes of sarcoidosis are driven by a complex interaction between antigen presenting molecules, antigens(s) and T.-cell receptors leading to antigen clearance or persistence (51). In this hypothesis, HLA Class II molecules with the higher affinity for mycobacterial antigenic epitopes might present to T cells wider antigen repertoires with more elevated efficiency. This would initiate a hypersensitivity response leading to acute onset of disease, antigen clearance and spontaneous disease remission. To the contrary, HLA Class II molecules, such as DRB1*1501, due to their lower affinity for bacterial epitopes, would express a less efficient antigen presentation process, thereby dictating a sluggish immune response resulting in insidious clinical presentation, antigen persistence, continuing granuloma formation: in summary, to chronic sarcoidosis.

Following the same line of interpretation, also for the *in silico* data presented here on tuberculosis patients, our observation would suggests that patients bearing even less recognizing HLA-DR allele combinations could be unable to mount an effective immune response, would fail to achieve bacterial containment and would, consequentially, develop active infectious disease.

Interestingly, it has been proposed that in tuberculosis a spectrum of immune reactions may be recognized, where poor cellular mediated immunity, associated with exuberant antibody production, represents one end of the spectrum while good cellular mediated immunity, with little or no antibody formation, would be at the other end (52). Similarly, in leprosy the immune spectrum is represented by non-reactive disseminated lepromatous leprosy, associated with HLA-DR 15 alleles (53, 54), at one end, while tuberculoid leprosy --the disease presentation characterized by stronger T-cell mediated immunity and a lesser invasiveness—is associated with HLA-DR3 alleles (55) and represents the opposite end. In this context, it is enticing to hypothesize that sarcoidosis may be part of a spectrum of granulomatous response to one or more closely related bacteria where the Loefgren’s syndrome represents hypersensitivity, similarly to the hot tub lung
(15) and the metalworking fluid associated pneumonitis (16), while pulmonary atypical mycobacterial infection and tuberculosis might represent the other end of the spectrum.

In conclusion, even though these in silico immunogenetic data still need to be confirmed in other sarcoid populations with different ethnical backgrounds and to be validated by in vitro and ex vivo confirmatory analysis, our findings support a pathogenetic role of specific groups of bacteria in the hypersensitivity granulomatous reaction(s) of sarcoidosis.
References


Figure 1. Relative recognition of all the Loefgren’s syndrome affected subjects with respect to chronic sarcoidosis patients of the HLA-DR immunomes deduced from the genomes of pathogenic bacteria. The ratio of the total number of the recognized epitopes (on the ordinate) in the two study groups is reported for 124 non redundant genomes of bacterial pathogens analysed (shown in alphabetic order on the abscissa). The bacterial groups showing statistically significant differences in HLA-DR immunome recognition are indicated on the graph.
Table 1. Study population.

<table>
<thead>
<tr>
<th>Study population</th>
<th>Subject number</th>
<th>Subjects with both HLA-DR alleles with known peptide binding motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>447</td>
<td>349 (78.1%)</td>
</tr>
<tr>
<td>Loefgren</td>
<td>39</td>
<td>31 (79.5%)</td>
</tr>
<tr>
<td>Chronic Sarcoidosis</td>
<td>110</td>
<td>93 (84.5%)</td>
</tr>
</tbody>
</table>

1. Numbers of subjects for each study population used for the epitope prediction analysis.
Table 2. List of non redundant, complete bacterial genomes classified as “relevant human pathogens” in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations.

<table>
<thead>
<tr>
<th>ORGANISM1</th>
<th>DISEASE</th>
<th>GOLD DATABASE Entry</th>
<th>Genome (Kb)²</th>
<th>Number of ORFs³</th>
<th>Mean Number of potential epitopes (1% affinity) in the study population⁴</th>
<th>t-test (p value)⁵</th>
<th>Controls</th>
<th>Chronic sarcoidosis</th>
<th>Loefgren</th>
<th>Chronic Sarc vs CTR</th>
<th>Loefgren vs CTR</th>
<th>Loefgren vs Chronic Sarc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acinetobacter baumannii</td>
<td>Meningitis, Pneumonia, Septicemia</td>
<td>Gc00522</td>
<td>3976</td>
<td>3352</td>
<td>17871±3963</td>
<td>17460±3881</td>
<td>18156±4008</td>
<td>0.1868</td>
<td>0.3504</td>
<td>0.1964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acinetobacter calcoaceticus</td>
<td>Nosocomial infection</td>
<td>Gc00201</td>
<td>3598</td>
<td>3325</td>
<td>21074±4645</td>
<td>20626±4573</td>
<td>21373±4622</td>
<td>0.2036</td>
<td>0.3657</td>
<td>0.2168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actinobacillus pleuropneumoniae</td>
<td>Necrotizing pleuropneumonia</td>
<td>Gc00708</td>
<td>2242</td>
<td>2036</td>
<td>13015±2854</td>
<td>12798±2808</td>
<td>13221±2833</td>
<td>0.2565</td>
<td>0.3499</td>
<td>0.2346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeromonas hydrophila</td>
<td>Gastroenteritis, Septicemia, Food poisoning</td>
<td>Gc00456</td>
<td>4744</td>
<td>4122</td>
<td>24703±5644</td>
<td>24118±5315</td>
<td>25373±5553</td>
<td>0.1843</td>
<td>0.2633</td>
<td>0.1355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeromonas salmonicida</td>
<td>Salmonicida</td>
<td>Gc00536</td>
<td>4702</td>
<td>4086</td>
<td>23482±5284</td>
<td>22944±4991</td>
<td>24103±5246</td>
<td>0.1890</td>
<td>0.2653</td>
<td>0.1355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaplasma marginale</td>
<td>Anaplasmosis</td>
<td>Gc00239</td>
<td>1197</td>
<td>949</td>
<td>6919±1536</td>
<td>6723±1466</td>
<td>7152±1639</td>
<td>0.1350</td>
<td>0.2108</td>
<td>0.0867</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaplasma phagocytophilum</td>
<td>Anaplasmosis</td>
<td>Gc00351</td>
<td>1471</td>
<td>1264</td>
<td>7080±1571</td>
<td>6916±1544</td>
<td>7282±1617</td>
<td>0.1842</td>
<td>0.2470</td>
<td>0.1301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arcobacter butzleri</td>
<td>Bacteremia, Gastroenteritis</td>
<td>Gc00657</td>
<td>2341</td>
<td>2259</td>
<td>16597±3902</td>
<td>16502±4006</td>
<td>16797±3699</td>
<td>0.1842</td>
<td>0.2470</td>
<td>0.1301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus anthracis</td>
<td>Anthrax</td>
<td>Gc00189</td>
<td>5227</td>
<td>5309</td>
<td>31589±7108</td>
<td>31033±7157</td>
<td>31971±6753</td>
<td>0.3655</td>
<td>0.3868</td>
<td>0.3245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus cereus</td>
<td>Food poisoning</td>
<td>Gc00173</td>
<td>5224</td>
<td>5603</td>
<td>33351±7428</td>
<td>33086±7480</td>
<td>33715±7080</td>
<td>0.3804</td>
<td>0.3964</td>
<td>0.3410</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus licheniformis</td>
<td>Food poisoning</td>
<td>Gc00213</td>
<td>4222</td>
<td>4152</td>
<td>23972±5311</td>
<td>23715±5265</td>
<td>24447±5125</td>
<td>0.3388</td>
<td>0.3165</td>
<td>0.2507</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteroides fragilis</td>
<td>Diarrhea, Abscesses</td>
<td>Gc00260</td>
<td>5205</td>
<td>4184</td>
<td>30522±6601</td>
<td>30276±6557</td>
<td>31142±6555</td>
<td>0.3746</td>
<td>0.3083</td>
<td>0.2628</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteroides thetaiotaomicron</td>
<td>Peritonitis</td>
<td>Gc00126</td>
<td>6260</td>
<td>4778</td>
<td>35884±7660</td>
<td>35563±7597</td>
<td>36516±7751</td>
<td>0.3674</td>
<td>0.3154</td>
<td>0.2658</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteroides vulgatus</td>
<td>Opportunistic peritoneal disease</td>
<td>Gc00584</td>
<td>5163</td>
<td>4065</td>
<td>28893±6119</td>
<td>28651±6052</td>
<td>29445±6250</td>
<td>0.3597</td>
<td>0.3302</td>
<td>0.2742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bartonella henselae</td>
<td>Bacillary angiomatosis</td>
<td>Gc00192</td>
<td>1931</td>
<td>1488</td>
<td>9206±1997</td>
<td>9048±1958</td>
<td>9427±2013</td>
<td>0.2483</td>
<td>0.2778</td>
<td>0.1780</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bartonella quintana</td>
<td>Bacillary angiomatosis, Endocarditis, Trench fever</td>
<td>Gc00191</td>
<td>1581</td>
<td>1142</td>
<td>7710±1699</td>
<td>7544±1666</td>
<td>7896±1705</td>
<td>0.1998</td>
<td>0.2801</td>
<td>0.1564</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bdellovibrio bacteriovorus</td>
<td>Cell lysis</td>
<td>Gc00168</td>
<td>3782</td>
<td>3587</td>
<td>21211±4611</td>
<td>20600±4493</td>
<td>21603±4929</td>
<td>0.1654</td>
<td>0.2899</td>
<td>0.1478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bordetella parapertussis</td>
<td>Respiratory infection</td>
<td>Gc00147</td>
<td>4773</td>
<td>4185</td>
<td>22589±5550</td>
<td>21833±5067</td>
<td>23661±5670</td>
<td>0.1177</td>
<td>0.1520</td>
<td>0.0469</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bordetella pertussis</td>
<td>Respiratory infection</td>
<td>Gc00146</td>
<td>4086</td>
<td>3436</td>
<td>18390±4527</td>
<td>17731±4121</td>
<td>19260±4681</td>
<td>0.1024</td>
<td>0.1534</td>
<td>0.0432</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrelia afzelii</td>
<td>Lyme disease, Acrodermatitis chronica atrophicans</td>
<td>Gc00408</td>
<td>905</td>
<td>856</td>
<td>7545±1708</td>
<td>7616±1778</td>
<td>7590±1624</td>
<td>0.3624</td>
<td>0.4436</td>
<td>0.4718</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
<td>Lyme disease</td>
<td>Gc00012</td>
<td>910</td>
<td>851</td>
<td>7635±1733</td>
<td>7699±1803</td>
<td>7690±1652</td>
<td>0.3767</td>
<td>0.4331</td>
<td>0.4896</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrelia garinii</td>
<td>Lyme disease</td>
<td>Gc00211</td>
<td>904</td>
<td>832</td>
<td>7395±1673</td>
<td>7446±1734</td>
<td>7447±1603</td>
<td>0.3963</td>
<td>0.4338</td>
<td>0.4996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrelia hermsii</td>
<td>Tick-borne relapsing fever</td>
<td>Gc00792</td>
<td>922</td>
<td>819</td>
<td>7605±1724</td>
<td>7668±1791</td>
<td>7695±1634</td>
<td>0.3786</td>
<td>0.3905</td>
<td>0.4707</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Disease(s)</td>
<td>Gene ID</td>
<td>Accession</td>
<td>Length (nt)</td>
<td>Similarity (nt)</td>
<td>InterproScan</td>
<td>BlastP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
<td>--------------</td>
<td>--------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Borrelia turicatae</em></td>
<td>Tick-borne relapsing fever</td>
<td>Gc00791</td>
<td>9173</td>
<td>818</td>
<td>7763±1784</td>
<td>7828±1859</td>
<td>0.3778</td>
<td>0.3851</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Brucella abortus</em></td>
<td>Brucellosis</td>
<td>Gc00263</td>
<td>3287</td>
<td>3085</td>
<td>15773±3805</td>
<td>16508±3809</td>
<td>0.1484</td>
<td>0.1516</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Brucella melitensis</em></td>
<td>Brucellosis</td>
<td>Gc00074</td>
<td>3294</td>
<td>3198</td>
<td>16050±3949</td>
<td>17265±3953</td>
<td>0.1605</td>
<td>0.1524</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Brucella suis</em></td>
<td>Brucellosis, Fever, Infectious abortions</td>
<td>Gc00699</td>
<td>3323</td>
<td>3241</td>
<td>16270±3886</td>
<td>17580±3886</td>
<td>0.1574</td>
<td>0.1585</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Burkholderia cenocepacia</em></td>
<td>Chronic infection, Necrotizing Pneumonia</td>
<td>Gc00744</td>
<td>7900</td>
<td>7008</td>
<td>34542±8237</td>
<td>36080±8521</td>
<td>0.1131</td>
<td>0.1604</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Burkholderia mallei</em></td>
<td>Glanders, Pneumonia</td>
<td>Gc00518</td>
<td>5900</td>
<td>5852</td>
<td>24519±5794</td>
<td>25689±6100</td>
<td>0.1316</td>
<td>0.1420</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Burkholderia multivorans</em></td>
<td>Cepacia syndrome</td>
<td>Gc00689</td>
<td>6820</td>
<td>6121</td>
<td>32560±7692</td>
<td>33960±8004</td>
<td>0.1202</td>
<td>0.1668</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Burkholderia pseudomallei</em></td>
<td>Pneumonia, Bacteremia, Melioidosis</td>
<td>Gc00520</td>
<td>7100</td>
<td>7183</td>
<td>32157±7571</td>
<td>33613±7981</td>
<td>0.1216</td>
<td>0.1537</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Burkholderia vietnamiensis</em></td>
<td>Pneumonia</td>
<td>Gc00531</td>
<td>7430</td>
<td>7617</td>
<td>33383±7916</td>
<td>34852±8335</td>
<td>0.1036</td>
<td>0.1624</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Burkholderia xenovorans</em> (fungorum)</td>
<td>Opportunistic infection</td>
<td>Gc00365</td>
<td>9279</td>
<td>8702</td>
<td>46606±10914</td>
<td>48476±11425</td>
<td>0.1010</td>
<td>0.1815</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter concisus</em></td>
<td>Gingivitis, Periodontitis, Periodontosis, Gastroenteritis</td>
<td>Gc00644</td>
<td>2052</td>
<td>1929</td>
<td>11849±2753</td>
<td>12204±2635</td>
<td>0.3881</td>
<td>0.2453</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter curvus</em></td>
<td>Gastroenteritis, Periodontal infection</td>
<td>Gc00628</td>
<td>1971</td>
<td>1931</td>
<td>12086±2771</td>
<td>12425±2642</td>
<td>0.3799</td>
<td>0.2564</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter fetus fetus</em></td>
<td>Bacteremia, Infertility, Septicemia, Meningitis</td>
<td>Gc00466</td>
<td>1733</td>
<td>1719</td>
<td>11959±2849</td>
<td>12283±2689</td>
<td>0.3809</td>
<td>0.2716</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter jejuni jejuni</em></td>
<td>Diarrhea, Gastroenteritis, Food poisoning</td>
<td>Gc00491</td>
<td>1616</td>
<td>1653</td>
<td>11426±2562</td>
<td>11577±2383</td>
<td>0.4583</td>
<td>0.3761</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter lari</em></td>
<td>Bacteremia, Diarrhea, Food poisoning, Gastroenteritis</td>
<td>Gc00851</td>
<td>1562</td>
<td>1599</td>
<td>10676±2320</td>
<td>10803±2214</td>
<td>0.4603</td>
<td>0.3854</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Candidatus Protochlamydia amoebophila</em></td>
<td>Pneumonia</td>
<td>Gc00185</td>
<td>2414</td>
<td>2031</td>
<td>13565±2812</td>
<td>13470±2836</td>
<td>0.3928</td>
<td>0.4286</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chlamydia muridarum</em></td>
<td>Pharyngitis, Bronchitis, Pneumonia, Respiratory infection</td>
<td>Gc00028</td>
<td>1072</td>
<td>904</td>
<td>6170±1313</td>
<td>6241±1360</td>
<td>0.2159</td>
<td>0.3859</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chlamydia trachomatis</em></td>
<td>Bronchitis, Heart disease, Pharyngitis, Pneumonia, Respiratory infection, Trachoma, Veneereal disease</td>
<td>Gc00017</td>
<td>1042</td>
<td>895</td>
<td>6007±1291</td>
<td>6084±1331</td>
<td>0.2044</td>
<td>0.3743</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chlamydia abortus</em></td>
<td>Pharyngitis, Bronchitis, Pneumonia, Respiratory infection</td>
<td>Gc00262</td>
<td>1144</td>
<td>932</td>
<td>6630±1421</td>
<td>6683±1430</td>
<td>0.2550</td>
<td>0.4157</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chlamydia pneumoniae</em></td>
<td>Pharyngitis, Bronchitis, Pneumonia, Heart disease, Respiratory infection</td>
<td>Gc00143</td>
<td>1225</td>
<td>1113</td>
<td>7387±1560</td>
<td>7438±1582</td>
<td>0.2572</td>
<td>0.4305</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chromobacterium violaceum</em></td>
<td>Diarrhea, Septicemia</td>
<td>Gc00157</td>
<td>4751</td>
<td>4407</td>
<td>22494±5283</td>
<td>23367±5441</td>
<td>0.1431</td>
<td>0.1898</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Citrobacter koseri</em></td>
<td>Bacteremia, Brain abscesses, Meningoencephalitis, Neonatal meningitis</td>
<td>Gc00042</td>
<td>4720</td>
<td>5003</td>
<td>28306±6188</td>
<td>28928±6208</td>
<td>0.1664</td>
<td>0.2961</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium botulinum</td>
<td>Botulism</td>
<td>Gc00741</td>
<td>3992</td>
<td>3655</td>
<td>2397±2559</td>
<td>23730±2581</td>
<td>23730±5592</td>
<td>0.3563</td>
<td>0.3178</td>
<td>0.2639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium difficile</td>
<td>Diarrhea, Colitis, Peritonitis</td>
<td>Gc00420</td>
<td>4290</td>
<td>3742</td>
<td>2656±6412</td>
<td>2612±6506</td>
<td>27255±6245</td>
<td>0.2791</td>
<td>0.2826</td>
<td>0.1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium perfringens</td>
<td>Dysenteria, Enterocolitis, Enterotoxemia, Food poisoning, Gas gangrene</td>
<td>Gc00406</td>
<td>3256</td>
<td>2876</td>
<td>1887±4506</td>
<td>28730±4584</td>
<td>19400±4242</td>
<td>0.3989</td>
<td>0.2672</td>
<td>0.2370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium tetani</td>
<td>Tetanus</td>
<td>Gc00120</td>
<td>2799</td>
<td>2373</td>
<td>18240±4199</td>
<td>18631±4105</td>
<td>0.3730</td>
<td>0.3096</td>
<td>0.2656</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corynebacterium diphtheriae</td>
<td>Diphtheria, Respiratory infection</td>
<td>Gc00163</td>
<td>2488</td>
<td>2272</td>
<td>12322±3052</td>
<td>12830±3196</td>
<td>0.0584</td>
<td>0.1882</td>
<td>0.0447</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corynebacterium jeikeium</td>
<td>Endocarditis, Septicemia, Nosocomial infection</td>
<td>Gc00423</td>
<td>2708</td>
<td>2437</td>
<td>1369±3305</td>
<td>1433±3446</td>
<td>0.1017</td>
<td>0.1540</td>
<td>0.0457</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corynebacterium urealyticum</td>
<td>Cystitis, Pyelitis, Urinary tract infection</td>
<td>Gc00755</td>
<td>2369</td>
<td>2024</td>
<td>10616±2640</td>
<td>11118±2738</td>
<td>0.0790</td>
<td>0.1560</td>
<td>0.0401</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coxiella burnetii</td>
<td>Food poisoning</td>
<td>Gc00692</td>
<td>2016</td>
<td>1930</td>
<td>12824±2643</td>
<td>12901±2601</td>
<td>0.4375</td>
<td>0.4386</td>
<td>0.4086</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ehrlichia chaffeensis</td>
<td>Ehrlichiosis, Human monocytic ehrlichiosis</td>
<td>Gc00353</td>
<td>1176</td>
<td>1105</td>
<td>7907±1729</td>
<td>7993±1720</td>
<td>0.3058</td>
<td>0.3958</td>
<td>0.3023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterobacter sakazakii</td>
<td>Meningitis, Septicemia, Necrotizing enterocolitis</td>
<td>Gc00638</td>
<td>4368</td>
<td>4277</td>
<td>24007±5218</td>
<td>24594±5261</td>
<td>0.1912</td>
<td>0.2746</td>
<td>0.1438</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterococcus faecalis</td>
<td>Bacteremia, Endocarditis, Urinary infection</td>
<td>Gc00831</td>
<td>2739</td>
<td>2701</td>
<td>18533±4090</td>
<td>18760±3981</td>
<td>0.3306</td>
<td>0.3833</td>
<td>0.3026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>Diarrhea</td>
<td>Gc00716</td>
<td>4686</td>
<td>4126</td>
<td>29392±6442</td>
<td>28654±6199</td>
<td>0.1618</td>
<td>0.2931</td>
<td>0.1433</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franciscella philomiragia</td>
<td>Bacteremia, Pneumonia, Septicemia</td>
<td>Gc00722</td>
<td>2045</td>
<td>1911</td>
<td>14775±3443</td>
<td>14931±3225</td>
<td>0.3523</td>
<td>0.4040</td>
<td>0.3326</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franciscella tularensis</td>
<td>Tularemia</td>
<td>Gc00241</td>
<td>1892</td>
<td>1603</td>
<td>11937±2715</td>
<td>12039±2605</td>
<td>0.3443</td>
<td>0.4206</td>
<td>0.3424</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusobacterium nucleatum nucleatum</td>
<td>Periodontal infection</td>
<td>Gc00085</td>
<td>2170</td>
<td>2067</td>
<td>14465±3377</td>
<td>14784±3182</td>
<td>0.4891</td>
<td>0.3064</td>
<td>0.3309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granulibacter bethesdensis</td>
<td>Chronic granulomatous</td>
<td>Gc00423</td>
<td>2708</td>
<td>2437</td>
<td>1369±3305</td>
<td>1433±3446</td>
<td>0.1017</td>
<td>0.1540</td>
<td>0.0457</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemophilus ducreyi</td>
<td>Genital ulcer, Chancroid</td>
<td>Gc00142</td>
<td>1698</td>
<td>1717</td>
<td>9470±2049</td>
<td>9291±2014</td>
<td>0.2260</td>
<td>0.3986</td>
<td>0.2543</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemophilus influenzae</td>
<td>Otitis, Meningitis, Septicemia, Sinusitis, Bronchitis</td>
<td>Gc00001</td>
<td>1830</td>
<td>1657</td>
<td>10640±2330</td>
<td>10852±2324</td>
<td>0.2560</td>
<td>0.3138</td>
<td>0.2077</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemophilus somnus</td>
<td>Arthritis, Myocarditis, Pneumonia</td>
<td>Gc00743</td>
<td>2263</td>
<td>1980</td>
<td>11524±2533</td>
<td>11710±2538</td>
<td>0.2313</td>
<td>0.3476</td>
<td>0.2201</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>Gastric inflammation, Ulcer</td>
<td>Gc00879</td>
<td>1673</td>
<td>1567</td>
<td>10030±2103</td>
<td>10098±2064</td>
<td>0.4475</td>
<td>0.4316</td>
<td>0.4091</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>Bacteremia, Pneumonia, Urinary tract infection</td>
<td>Gc00841</td>
<td>5641</td>
<td>5425</td>
<td>2917±6561</td>
<td>28491±6262</td>
<td>0.1855</td>
<td>0.2740</td>
<td>0.1406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legionella pneumophila</td>
<td>Legionellosis</td>
<td>Gc00226</td>
<td>3345</td>
<td>2878</td>
<td>2107±4478</td>
<td>21141±4445</td>
<td>0.3014</td>
<td>0.4681</td>
<td>0.3571</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leptospira borgpeterseni hardjobovis</td>
<td>Leptospirosis</td>
<td>Gc00433</td>
<td>3920</td>
<td>2945</td>
<td>19500±4144</td>
<td>19667±4101</td>
<td>0.3306</td>
<td>0.4150</td>
<td>0.3290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listeria monocytogenes</td>
<td>Food poisoning, Listeriosis</td>
<td>Gc00186</td>
<td>2905</td>
<td>2821</td>
<td>17937±4140</td>
<td>18305±3994</td>
<td>0.2431</td>
<td>0.3173</td>
<td>0.2044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium abscessus</td>
<td>Broncho-pulmonary infection, Respiratory infection, Wound infection</td>
<td>Gc00729</td>
<td>5067</td>
<td>4920</td>
<td>24543±6186</td>
<td>23410±5792</td>
<td>0.0562</td>
<td>0.1502</td>
<td>0.0309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Description</td>
<td>GenBank ID</td>
<td>Blot Score 1</td>
<td>Blot Score 2</td>
<td>Blot Score 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium avium</td>
<td>Respiratory infection, Tuberculosis type pulmonary infection</td>
<td>Gc00462</td>
<td>5475</td>
<td>5120</td>
<td>2335±6182</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium avium paratuberculosis</td>
<td>Johne's disease, Paratuberculosis, Enteritis</td>
<td>Gc00169</td>
<td>4829</td>
<td>4350</td>
<td>2167±5713</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium bovis</td>
<td>Tuberculosis</td>
<td>Gc00138</td>
<td>4345</td>
<td>3920</td>
<td>1975±5050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium bovis BCG</td>
<td>Bovine tuberculosis</td>
<td>Gc00489</td>
<td>4374</td>
<td>3952</td>
<td>1993±4510</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium leprae</td>
<td>Leprosy, Hanson's disease</td>
<td>Gc00045</td>
<td>3268</td>
<td>1605</td>
<td>9120±2335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium marinum</td>
<td>Tuberculosis</td>
<td>Gc000784</td>
<td>6636</td>
<td>5423</td>
<td>2894±7417</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium smegmatis</td>
<td>Soft tissue lesions</td>
<td>Gc00461</td>
<td>6988</td>
<td>6716</td>
<td>32660±8582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium tuberculosis</td>
<td>Tuberculosis</td>
<td>Gc00578</td>
<td>4424</td>
<td>3950</td>
<td>2004±5081</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium ulcerans</td>
<td>Buruli ulcer</td>
<td>Gc00469</td>
<td>5631</td>
<td>4160</td>
<td>2021±5218</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycoplasma arthritidis</td>
<td>Arthritis</td>
<td>Gc00825</td>
<td>820</td>
<td>631</td>
<td>4728±998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycoplasma genitalium</td>
<td>Urogenital infection, Respiratory infection, Non-gonococcal urethritis</td>
<td>Gc00002</td>
<td>580</td>
<td>477</td>
<td>4034±836</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycoplasma penetrans</td>
<td>Urogenital infection, Respiratory infection</td>
<td>Gc00110</td>
<td>1358</td>
<td>1037</td>
<td>8755±1969</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycoplasma pneumoniae</td>
<td>Pneumonia, Tracheobronchitis, Respiratory infection</td>
<td>Gc00005</td>
<td>816</td>
<td>689</td>
<td>6227±1330</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria gonorrhoeae</td>
<td>Gonorhea</td>
<td>Gc00258</td>
<td>2153</td>
<td>2002</td>
<td>1071±2295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria meningitidis</td>
<td>Meningitis, Septicemia</td>
<td>Gc00026</td>
<td>2272</td>
<td>2063</td>
<td>1010±2187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neorickettsia sennetsu</td>
<td>Sennetsu fever</td>
<td>Gc00352</td>
<td>859</td>
<td>932</td>
<td>5766±1265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nocardia farcinica</td>
<td>Nocardiosis, Mastitis</td>
<td>Gc00224</td>
<td>6021</td>
<td>5683</td>
<td>26116±6921</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parabacteroides distasonis</td>
<td>Opportunistic peritonneal disease</td>
<td>Gc00583</td>
<td>4811</td>
<td>3850</td>
<td>2885±2624</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pasteurella multocida</td>
<td>Pasteurellosis, Cholera, Septicemia</td>
<td>Gc00048</td>
<td>2257</td>
<td>2015</td>
<td>13526±2959</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porphyromonas gingivalis</td>
<td>Dental plaque, Periodontal infection</td>
<td>Gc00809</td>
<td>2354</td>
<td>2090</td>
<td>1333±2871</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propionibacterium acnes</td>
<td>Acne</td>
<td>Gc00204</td>
<td>2560</td>
<td>2297</td>
<td>1299±3347</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteus mirabilis</td>
<td>Ecephalitis, Pneumonia, Pyelonephritis, Septicemia, Surgical wound infection, Ulcer, Urinary tract infection, Urolithiasis</td>
<td>Gc00758</td>
<td>4063</td>
<td>3693</td>
<td>2280±5068</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudomonas aeruginosa</td>
<td>Opportunistic infection, Nocosomial infection</td>
<td>Gc00432</td>
<td>6537</td>
<td>5892</td>
<td>32928±7633</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudomonas mendocina</td>
<td>Endocarditis, Spondylodiscitis</td>
<td>Gc00544</td>
<td>5072</td>
<td>4594</td>
<td>25846±6022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rickettsia bellii</td>
<td>Epidemic typhus</td>
<td>Gc00661</td>
<td>1528</td>
<td>1476</td>
<td>9556±2129</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rickettsia canadensis</td>
<td>Epidemic typhus</td>
<td>Gc00660</td>
<td>1159</td>
<td>1093</td>
<td>7112±1594</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rickettsia conorii</td>
<td>Rocky Mountain Spotted Fever</td>
<td>Gc00062</td>
<td>1268</td>
<td>1374</td>
<td>8045±1775</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Disease(s)</td>
<td>GenBank Accession</td>
<td>GC1</td>
<td>GC2</td>
<td>GC3</td>
<td>GC4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rickettsia prowazekii</strong></td>
<td>Typhus, Rocky Mountain Spotted Fever</td>
<td>Gc00018</td>
<td>1111</td>
<td>835</td>
<td>7260±1621</td>
<td>7249±1677</td>
<td>7277±1536</td>
<td>0.4777</td>
<td>0.4784</td>
<td>0.4684</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rickettsia rickettsi</strong></td>
<td>Rocky Mountain Spotted Fever</td>
<td>Gc00736</td>
<td>1257</td>
<td>1345</td>
<td>7436±1648</td>
<td>7374±1687</td>
<td>7479±1583</td>
<td>0.3747</td>
<td>0.4447</td>
<td>0.3810</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rickettsia typhi</strong></td>
<td>Typhus</td>
<td>Gc00208</td>
<td>1111</td>
<td>838</td>
<td>7162±1602</td>
<td>7164±1653</td>
<td>7212±1528</td>
<td>0.4544</td>
<td>0.4645</td>
<td>0.4432</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Salmonella enterica</strong></td>
<td>Food poisoning, Gastroenteritis, Salmonellosis</td>
<td>Gc00833</td>
<td>4685</td>
<td>4318</td>
<td>25874±5635</td>
<td>25253±5414</td>
<td>26418±5671</td>
<td>0.1723</td>
<td>0.2980</td>
<td>0.1502</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Salmonella enterica sv Paratyphi A</strong></td>
<td>Food poisoning, Gastroenteritis, Salmonellosis, Typhoid fever</td>
<td>Gc00853</td>
<td>4585</td>
<td>4093</td>
<td>28451±5697</td>
<td>27834±5959</td>
<td>29034±6155</td>
<td>0.1746</td>
<td>0.3163</td>
<td>0.1606</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Salmonella enterica sv Typhi</strong></td>
<td>Typhoid fever, Salmonellosis, Food poisoning</td>
<td>Gc00066</td>
<td>4809</td>
<td>4395</td>
<td>25456±5596</td>
<td>24843±5356</td>
<td>26012±5603</td>
<td>0.1951</td>
<td>0.3081</td>
<td>0.1688</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Serratia proteamaculans</strong></td>
<td>Pneumonia</td>
<td>Gc00647</td>
<td>5448</td>
<td>4891</td>
<td>30454±6854</td>
<td>29710±6536</td>
<td>31514±6733</td>
<td>0.1740</td>
<td>0.2929</td>
<td>0.1462</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shigella boydii</strong></td>
<td>Dyserenteria, Food poisoning</td>
<td>Gc00779</td>
<td>4615</td>
<td>4246</td>
<td>22229±4780</td>
<td>21699±4539</td>
<td>22648±4871</td>
<td>0.1686</td>
<td>0.3205</td>
<td>0.1621</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shigella dysenteriae</strong></td>
<td>Dyserenteria, Food poisoning</td>
<td>Gc00324</td>
<td>4369</td>
<td>4274</td>
<td>21088±4422</td>
<td>20522±4135</td>
<td>21444±4595</td>
<td>0.1331</td>
<td>0.3344</td>
<td>0.1488</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shigella flexneri</strong></td>
<td>Dyserenteria, Gastroenteritis, Food poisoning</td>
<td>Gc00130</td>
<td>4599</td>
<td>4068</td>
<td>23543±5078</td>
<td>22894±4839</td>
<td>24010±5130</td>
<td>0.1729</td>
<td>0.3121</td>
<td>0.1592</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shigella sonnei</strong></td>
<td>Dyserenteria, Food poisoning</td>
<td>Gc00323</td>
<td>5039</td>
<td>4461</td>
<td>24441±5249</td>
<td>23862±4983</td>
<td>24912±5377</td>
<td>0.1703</td>
<td>0.3163</td>
<td>0.1606</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Staphylococcus aureus</strong></td>
<td>Mastitis, Nosocomial infection</td>
<td>Gc000327</td>
<td>2515</td>
<td>2665</td>
<td>18106±4153</td>
<td>17948±4241</td>
<td>18367±3890</td>
<td>0.3727</td>
<td>0.3684</td>
<td>0.3140</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Staphylococcus haemolyticus</strong></td>
<td>Opportunistic infection</td>
<td>Gc000324</td>
<td>2685</td>
<td>2676</td>
<td>17340±3945</td>
<td>17207±4023</td>
<td>17582±3713</td>
<td>0.3869</td>
<td>0.3711</td>
<td>0.3238</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptococcus agalactiae</strong></td>
<td>Meningitis</td>
<td>Gc000643</td>
<td>2196</td>
<td>2051</td>
<td>13014±2945</td>
<td>12859±2958</td>
<td>13286±2829</td>
<td>0.3647</td>
<td>0.3106</td>
<td>0.2602</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptococcus gordonii</strong></td>
<td>Dental plaque, Endocarditis, Periodontal infection</td>
<td>Gc00285</td>
<td>1838</td>
<td>1865</td>
<td>10563±2399</td>
<td>10376±2396</td>
<td>10801±2403</td>
<td>0.2530</td>
<td>0.2985</td>
<td>0.1976</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptococcus mutans</strong></td>
<td>Dental caries</td>
<td>Gc00032</td>
<td>2030</td>
<td>1960</td>
<td>12205±2735</td>
<td>12118±2770</td>
<td>12415±2651</td>
<td>0.3939</td>
<td>0.3406</td>
<td>0.3015</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptococcus pneumoniae</strong></td>
<td>Meningitis, Pneumonia, Otitis media</td>
<td>Gc000437</td>
<td>2000</td>
<td>1914</td>
<td>12092±2702</td>
<td>11967±2722</td>
<td>12331±2619</td>
<td>0.3458</td>
<td>0.3182</td>
<td>0.2580</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptococcus pyogenes</strong></td>
<td>Bone infection, Endocarditis, Mastoiditis, Meningitis, Myositis, Necrotizing fasciitis, Otitis, Pharyngitis, Pneumonia, Sinusitis, Tonsillitis</td>
<td>Gc00285</td>
<td>1838</td>
<td>1865</td>
<td>10563±2399</td>
<td>10376±2396</td>
<td>10801±2403</td>
<td>0.2530</td>
<td>0.2985</td>
<td>0.1976</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptococcus suis</strong></td>
<td>Meningitis, Endocarditis, Septicemia, Arthritis</td>
<td>Gc000546</td>
<td>2096</td>
<td>2186</td>
<td>12022±2696</td>
<td>11867±2694</td>
<td>12293±2660</td>
<td>0.3108</td>
<td>0.2955</td>
<td>0.2225</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Treponema pallidum pallidum</strong></td>
<td>Syphilis</td>
<td>Gc000789</td>
<td>1139</td>
<td>1028</td>
<td>6849±1410</td>
<td>6676±1310</td>
<td>7002±1533</td>
<td>0.1435</td>
<td>0.2829</td>
<td>0.1264</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tropheryma whipplei</strong></td>
<td>Whipple's disease</td>
<td>Gc000123</td>
<td>925</td>
<td>783</td>
<td>5486±1270</td>
<td>5331±1245</td>
<td>5654±1276</td>
<td>0.1467</td>
<td>0.2404</td>
<td>0.1079</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ureaplasma parvum</strong></td>
<td>Respiratory infection, Urinary tract infection</td>
<td>Gc000742</td>
<td>751</td>
<td>609</td>
<td>5510±1196</td>
<td>5469±1221</td>
<td>5498±1212</td>
<td>0.3847</td>
<td>0.4797</td>
<td>0.4533</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vibrio cholerae</strong></td>
<td>Cholera, Food poisoning</td>
<td>Gc00057</td>
<td>4148</td>
<td>3875</td>
<td>22495±4906</td>
<td>21872±4736</td>
<td>22891±5091</td>
<td>0.1368</td>
<td>0.3337</td>
<td>0.1552</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vibrio parahaemolyticus</strong></td>
<td>Gastroenteritis</td>
<td>Gc000124</td>
<td>5165</td>
<td>4832</td>
<td>27525±6192</td>
<td>26701±6047</td>
<td>28124±6419</td>
<td>0.1263</td>
<td>0.3035</td>
<td>0.1330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yersinia enterocolitica</td>
<td>Gastroenteritis, Food poisoning</td>
<td>Gc00481</td>
<td>4615</td>
<td>3979</td>
<td>25630±5744</td>
<td>24958±5547</td>
<td>26182±5722</td>
<td>0.1696</td>
<td>0.2958</td>
<td>0.1485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yersinia pestis</td>
<td>Bubonic and Pneumonic plague</td>
<td>Gc00538</td>
<td>4517</td>
<td>3850</td>
<td>23982±5310</td>
<td>23393±5121</td>
<td>24516±5316</td>
<td>0.1566</td>
<td>0.3041</td>
<td>0.1466</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yersinia pseudotuberculosis</td>
<td>Gastroenteritis</td>
<td>Gc00776</td>
<td>4695</td>
<td>4150</td>
<td>25461±5702</td>
<td>24847±5530</td>
<td>26020±5664</td>
<td>0.1766</td>
<td>0.3005</td>
<td>0.1556</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Non redundant, complete bacterial genomes classified as “relevant human pathogens” in the Genome OnLine Database v2.0 (GOLD database update October 30\textsuperscript{th} 2008, [http://www.genomesonline.org/](http://www.genomesonline.org/)) under analysis
2. Genome size in Kilobases
3. Number of Open Reading Frame in the genome
5. Mean Number of potential epitopes of the genome under analysis, equivalent to the affinity of 1\% of the best binder peptides, recognized by the different study populations
5. p value for the comparison between groups indicated. Students’ t-test. Statistically significant comparisons (p<0.05) are indicated in bold.
Table 3. HLA-DRB1 allelic frequency analysis of the study population.

<table>
<thead>
<tr>
<th>HLA-DRB1 Allele</th>
<th>Controls</th>
<th>Chronic Sarcoidosis</th>
<th>Loefgren</th>
<th>CTR vs Chr Sarc</th>
<th>CTR vs Loef</th>
<th>Chr Sarc vs Loef</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Freq</td>
<td>N</td>
<td>Freq</td>
<td>N</td>
<td>Freq</td>
</tr>
<tr>
<td>0101</td>
<td>57</td>
<td>0.064</td>
<td>12</td>
<td>0.055</td>
<td>6</td>
<td>0.077</td>
</tr>
<tr>
<td>0102</td>
<td>4</td>
<td>0.004</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0103</td>
<td>3</td>
<td>0.003</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0301</td>
<td>103</td>
<td>0.115</td>
<td>21</td>
<td>0.095</td>
<td>16</td>
<td>0.205</td>
</tr>
<tr>
<td>0302</td>
<td>2</td>
<td>0.002</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0401</td>
<td>57</td>
<td>0.064</td>
<td>13</td>
<td>0.059</td>
<td>3</td>
<td>0.038</td>
</tr>
<tr>
<td>0402</td>
<td>6</td>
<td>0.007</td>
<td>1</td>
<td>0.005</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0403</td>
<td>10</td>
<td>0.011</td>
<td>0</td>
<td>0.000</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td>0404</td>
<td>41</td>
<td>0.046</td>
<td>4</td>
<td>0.018</td>
<td>2</td>
<td>0.026</td>
</tr>
<tr>
<td>0405</td>
<td>9</td>
<td>0.010</td>
<td>1</td>
<td>0.005</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0406</td>
<td>1</td>
<td>0.001</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0407</td>
<td>9</td>
<td>0.010</td>
<td>1</td>
<td>0.005</td>
<td>4</td>
<td>0.051</td>
</tr>
<tr>
<td>0408</td>
<td>8</td>
<td>0.009</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0701</td>
<td>74</td>
<td>0.083</td>
<td>15</td>
<td>0.068</td>
<td>6</td>
<td>0.077</td>
</tr>
<tr>
<td>0801</td>
<td>18</td>
<td>0.020</td>
<td>8</td>
<td>0.036</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td>0803</td>
<td>6</td>
<td>0.007</td>
<td>1</td>
<td>0.005</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0804</td>
<td>2</td>
<td>0.002</td>
<td>2</td>
<td>0.009</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0812</td>
<td>1</td>
<td>0.001</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0901</td>
<td>10</td>
<td>0.011</td>
<td>4</td>
<td>0.018</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td>1001</td>
<td>6</td>
<td>0.007</td>
<td>3</td>
<td>0.014</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td>1101</td>
<td>78</td>
<td>0.087</td>
<td>29</td>
<td>0.132</td>
<td>4</td>
<td>0.051</td>
</tr>
<tr>
<td>1102</td>
<td>5</td>
<td>0.006</td>
<td>0</td>
<td>0.000</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td>1103</td>
<td>7</td>
<td>0.008</td>
<td>2</td>
<td>0.009</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1104</td>
<td>51</td>
<td>0.057</td>
<td>5</td>
<td>0.023</td>
<td>2</td>
<td>0.026</td>
</tr>
<tr>
<td>1201</td>
<td>10</td>
<td>0.011</td>
<td>2</td>
<td>0.009</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1202</td>
<td>4</td>
<td>0.004</td>
<td>1</td>
<td>0.005</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1301</td>
<td>94</td>
<td>0.105</td>
<td>16</td>
<td>0.073</td>
<td>9</td>
<td>0.115</td>
</tr>
<tr>
<td>1302</td>
<td>52</td>
<td>0.058</td>
<td>18</td>
<td>0.082</td>
<td>3</td>
<td>0.038</td>
</tr>
<tr>
<td>1303</td>
<td>13</td>
<td>0.015</td>
<td>0</td>
<td>0.000</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td>1305</td>
<td>1</td>
<td>0.001</td>
<td>1</td>
<td>0.005</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1326</td>
<td>1</td>
<td>0.001</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1401</td>
<td>33</td>
<td>0.037</td>
<td>13</td>
<td>0.059</td>
<td>3</td>
<td>0.038</td>
</tr>
<tr>
<td>1404</td>
<td>1</td>
<td>0.001</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1407</td>
<td>2</td>
<td>0.002</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1501</td>
<td>86</td>
<td>0.096</td>
<td>44</td>
<td>0.200</td>
<td>13</td>
<td>0.167</td>
</tr>
<tr>
<td>1502</td>
<td>9</td>
<td>0.010</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1601</td>
<td>19</td>
<td>0.021</td>
<td>2</td>
<td>0.009</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td>1602</td>
<td>1</td>
<td>0.001</td>
<td>1</td>
<td>0.005</td>
<td>0</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*Bonferroni corrected p>0.05

^Bonferroni corrected p=0.0013
Table 5. Analysis of the HLA-DRB1 allele frequency in the study population subgroups ranked by the percentile recognition of *M. avium* immunome.

A. 76°-100° percentile of *M. avium* immunome

<table>
<thead>
<tr>
<th>HLA-DRB1 Allele</th>
<th>Loefgren N</th>
<th>Freq</th>
<th>Chronic Sarcoidosis N</th>
<th>Freq</th>
<th>Controls N</th>
<th>Freq</th>
<th>χ-square p</th>
<th>χ-square p</th>
<th>χ-square p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0101</td>
<td>4</td>
<td>0.142857</td>
<td>2</td>
<td>0.0625</td>
<td>12</td>
<td>0.065934</td>
<td>0.752637</td>
<td>0.29569</td>
<td>0.545972</td>
</tr>
<tr>
<td>0102</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0.016484</td>
<td>0.933209</td>
<td>0.86417</td>
<td>NA</td>
</tr>
<tr>
<td>0301</td>
<td>11</td>
<td>0.392857</td>
<td>12</td>
<td>0.375</td>
<td>71</td>
<td>0.39011</td>
<td>0.972137</td>
<td>0.856908</td>
<td>0.901165</td>
</tr>
<tr>
<td>0401</td>
<td>1</td>
<td>0.035714</td>
<td>1</td>
<td>0.03125</td>
<td>5</td>
<td>0.027473</td>
<td>0.644641</td>
<td>0.714702</td>
<td>0.532173</td>
</tr>
<tr>
<td>0404</td>
<td>2</td>
<td>0.071429</td>
<td>1</td>
<td>0.03125</td>
<td>6</td>
<td>0.032967</td>
<td>0.62522</td>
<td>0.645852</td>
<td>0.905486</td>
</tr>
<tr>
<td>0405</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.005495</td>
<td>0.324589</td>
<td>0.279601</td>
<td>NA</td>
</tr>
<tr>
<td>0408</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.010989</td>
<td>0.688938</td>
<td>0.625796</td>
<td>NA</td>
</tr>
<tr>
<td>0701</td>
<td>1</td>
<td>0.035714</td>
<td>3</td>
<td>0.09375</td>
<td>14</td>
<td>0.076923</td>
<td>0.976218</td>
<td>0.693497</td>
<td>0.703662</td>
</tr>
<tr>
<td>0801</td>
<td>1</td>
<td>0.035714</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0.021978</td>
<td>0.889536</td>
<td>0.824374</td>
<td>0.94628</td>
</tr>
<tr>
<td>1101</td>
<td>2</td>
<td>0.071429</td>
<td>2</td>
<td>0.0625</td>
<td>11</td>
<td>0.06044</td>
<td>0.721656</td>
<td>0.844175</td>
<td>0.703662</td>
</tr>
<tr>
<td>1102</td>
<td>1</td>
<td>0.035714</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0.279601</td>
<td>0.94628</td>
</tr>
<tr>
<td>1104</td>
<td>1</td>
<td>0.035714</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0.054945</td>
<td>0.366004</td>
<td>0.975771</td>
<td>0.94628</td>
</tr>
<tr>
<td>1301</td>
<td>3</td>
<td>0.107143</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0.093407</td>
<td>0.147756</td>
<td>0.908241</td>
<td>0.191528</td>
</tr>
<tr>
<td>1305</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.005495</td>
<td>0.324589</td>
<td>0.279601</td>
<td>NA</td>
</tr>
<tr>
<td>1401</td>
<td>1</td>
<td>0.035714</td>
<td>4</td>
<td>0.125</td>
<td>12</td>
<td>0.065934</td>
<td>0.419583</td>
<td>0.844175</td>
<td>0.435253</td>
</tr>
<tr>
<td>1501</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0.21875</td>
<td>13</td>
<td>0.071429</td>
<td>0.020827</td>
<td>0.298833</td>
<td>0.025735</td>
</tr>
</tbody>
</table>

B. 26°-50° percentile of *M. avium* immunome

<table>
<thead>
<tr>
<th>HLA-DRB1 Allele</th>
<th>Loefgren N</th>
<th>Freq</th>
<th>Chronic Sarcoidosis N</th>
<th>Freq</th>
<th>Controls N</th>
<th>Freq</th>
<th>χ-square p</th>
<th>χ-square p</th>
<th>χ-square p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0401</td>
<td>1</td>
<td>0.1</td>
<td>6</td>
<td>0.083333</td>
<td>31</td>
<td>0.186747</td>
<td>0.067584</td>
<td>0.788224</td>
<td>0.669285</td>
</tr>
<tr>
<td>0402</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.013889</td>
<td>2</td>
<td>0.012048</td>
<td>0.60619</td>
<td>0.235259</td>
<td>0.245069</td>
</tr>
<tr>
<td>0404</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.013889</td>
<td>11</td>
<td>0.066265</td>
<td>0.169496</td>
<td>0.866468</td>
<td>0.245069</td>
</tr>
<tr>
<td>0405</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.013889</td>
<td>3</td>
<td>0.018072</td>
<td>0.750289</td>
<td>0.407113</td>
<td>0.245069</td>
</tr>
<tr>
<td>0408</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0.018072</td>
<td>0.60619</td>
<td>0.407113</td>
<td>NA</td>
</tr>
<tr>
<td>0701</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0.041667</td>
<td>18</td>
<td>0.108434</td>
<td>0.155789</td>
<td>0.574302</td>
<td>0.809451</td>
</tr>
<tr>
<td>0801</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0.083333</td>
<td>6</td>
<td>0.036145</td>
<td>0.227866</td>
<td>0.775284</td>
<td>0.763971</td>
</tr>
<tr>
<td>0804</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.006024</td>
<td>0.666602</td>
<td>0.054871</td>
<td>NA</td>
</tr>
<tr>
<td>1101</td>
<td>2</td>
<td>0.2</td>
<td>13</td>
<td>0.180556</td>
<td>22</td>
<td>0.13253</td>
<td>0.446225</td>
<td>0.897053</td>
<td>0.773789</td>
</tr>
<tr>
<td>1102</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0.018072</td>
<td>0.60619</td>
<td>0.407113</td>
<td>NA</td>
</tr>
<tr>
<td>1104</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.013889</td>
<td>3</td>
<td>0.018072</td>
<td>0.750289</td>
<td>0.407113</td>
<td>0.245069</td>
</tr>
<tr>
<td>1301</td>
<td>3</td>
<td>0.3</td>
<td>11</td>
<td>0.152778</td>
<td>18</td>
<td>0.108434</td>
<td>0.456291</td>
<td>0.189295</td>
<td>0.477119</td>
</tr>
<tr>
<td>1302</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0.083333</td>
<td>15</td>
<td>0.090361</td>
<td>0.941676</td>
<td>0.681216</td>
<td>0.763971</td>
</tr>
<tr>
<td>1501</td>
<td>4</td>
<td>0.4</td>
<td>23</td>
<td>0.319444</td>
<td>28</td>
<td>0.168675</td>
<td>0.01502</td>
<td>0.155656</td>
<td>0.881651</td>
</tr>
<tr>
<td>1502</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.012048</td>
<td>0.871005</td>
<td>0.235259</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 6. Comparison of the recognition of the HLA-DR immunomes of *M. avium* and *M. tuberculosis* in the study populations in comparison to tuberculosis patients.

<table>
<thead>
<tr>
<th>Genome</th>
<th>Mean Number of potential epitopes (1% affinity)¹</th>
<th>Student’s t-test (p-value)²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Chronic Sarcoidosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loefgren</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TB vs Control</td>
</tr>
<tr>
<td><em>M. avium</em></td>
<td>23353±6182</td>
<td>22194±5736</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>20040±5081</td>
<td>19073±4713</td>
</tr>
</tbody>
</table>

1. Mean Number of potential epitopes in the different study population of the *M. avium* and *M. tuberculosis*, equivalent or higher than the affinity of 1% of the best binder peptides for the subject's HLA-DR molecules
2. p value for the comparison between groups indicated. Students' t-test. Statistically significant comparisons (p<0.05) are indicated in bold
3. Matched controls of the tuberculosis patient populations did not differ from the control population of this study in terms of the number of potential epitopes recognized (*M. avium*: 23353±6182 vs 22591±7612; *M. tuberculosis*: 20040±5081 vs 19472±4972; p>0.05 all comparisons).
THANKS TO

My family, my dear husband Simone, and my beloved son Davide

My friends and coworkers Alessia Comandini, and Floriana Berretta, for it has been a pleasure to work with them.

The head of laboratory Massimo Amicosante for the enthusiasm he transmitted to me.

And a special thanks to my professor and my tutor prof. Cesare Saltini for his constant support and education for my career.