Reviewer's report

Title: A low clustering rate with predominance of Haarlem strains in patients without known risk factors for multi-drug resistant tuberculosis in North-Eastern Lima, Peru

Version: 1 Date: 12 February 2013

Reviewer: Midori Kato-Maeda

Reviewer's report:

Major Compulsory Revisions

Abstract:
1. In the conclusions section please clarify what is the meaning of –early introduction-.
2. The current study can't be used to conclude that a low proportion of TB within this study population was due to recent transmission. The 7 month study period is too short to study events of recent transmission with rapid evolution to active TB. Also, the study population was not community-based and the patients with risk of drug resistant TB that were excluded from the study may be in cluster or not with patients that do not have risk for drug resistant TB.

Introduction:
1. Please include that one of the aims of the manuscript is to determine the factors associated with MDR and isoniazid resistant, which is detailed in your tables 2 and 3.
2. Please describe what information is added in this study, that it was not reported in the published paper from Otero L, Krapp F, Tommatis C, Zamudio C, Matthys F, Gotuzzo E, Van der Stuyft P, Seas C. High Prevalence of Primary Multidrug Resistant Tuberculosis in Persons with No Known Risk Factors. PLOS ONE 2011. 6(10):e26276.
3. Please describe why describing the M. tuberculosis diversity (distribution of lineages) in patients without risk factors for drug resistant tuberculosis is important.

Method
1. Study setting and study population. In order to assess the burden of TB due to recent transmission in North Eastern district in the Lima Province in Peru, all (or almost all) patients with culture positive tuberculosis should be included, as transmission can occur between patients with and without risk factors for drug resistance. Therefore the study population chosen can't be used to study the burden of recent transmission. Please reconsider the study population if you would like to assess the burden of TB due to recent transmission.
2. Data collection: Authors collected information about treatment outcome but it is
not discussed in any part of the manuscript. Please describe.

3. Fingerprint analysis: Please describe the goal of generating the dendrogram and describe the dendogram in the result section.

4. Statistical analysis.
   a. Please describe the relationships that were tested with the statistical analysis.
   b. Please clarify in which analysis the lineages the outcome variable.

5. Ethical considerations:
   Please describe if all the patients that were eligible consented. If not all eligible accepted participating in the study, please describe those that did not accept to determine the representation of the study population.

Results.
1. Please review the numbers.
   a. 50 excluded from 314 (50/314) is not 24.8%. Also, 314-50 is not 254.
   b. The result section of the manuscript describes that 147 were grouped in 22 spoligo-based clusters; however in the figure 1, there were just 146 isolates that were in cluster.

2. Please show the results related to the analysis of lineages and its association to age, sex, or residency, so the reader can understand the comments described in the discussion section.

Discussion
1. The authors mention that the study illustrated the possible relatedness of the circulating genotypes with drug resistant patterns and patients characteristics. Please comment on the fact that none of these associations were statistically significant, which may be in part because of the small sample size.

2. The discussion regarding the molecular epidemiology of TB can’t be supported by this study which is not designed to determine the frequency of recent transmission. Although in the limitations the authors mentioned that clustering rates could be underestimated due to the relatively short study period, it is also because the study population is not appropriate: patients with risk factors for drug resistant TB may be in cluster with those that have no risk factors, which were excluded from the study.

3. The comment related to the widespread occurrence of the Haarlem lineage in this setting could be attributed to its higher stability and/or relative superior fitness is very speculative, and out of context and should be deleted.

4. The table 1 is mentioned when discussion the associations of lineage with age. However, table 1 describes the frequency of drug resistance. Please add the table, and discuss it in the results.

5. The authors comment that: “A high MDR rate (7.5%) was found among smear-positive pulmonary TB patients without previous treatment history, or unknown risk factors for MDR”.
   a. Smear status and previous treatment history were not described in the results
or in tables, please add the information in results.
b. Please describe what are the comparison groups to state that a high MDR rate was found among (patients) with unknown risk factors for MDR.

Summary:
1. Authors describe that their findings suggest ongoing transmission: however, the methodology of the study can’t support this observation.
2. The next sentence: “spread of MDR-TB occurs at low rates from multiple patients with Untreated MDR-TB” can’t be supported by this study.
3. Please elaborate and clarify why “Predominance of the Haarlem lineage suggests for the long presence of this lineage in Peru”
4. Please elaborate and define what is the ‘distinct introduction mechanism of the Harlem lineage to Peru.

Figure 2 the dendogram related findings are not commented in the results no in the discussion. Please explain what is in each of the yellow squares, what are the questions marks, where are the MIRU15 results (the text refers to figure 2 to see the MIRU).

Table 2 and 3.
1. Please identify the referent groups.
2. The tables show the factors associated with MDR or INH resistant. This should be described in the aims.

Minor Essential Revisions

Abstract:
1. In the method section, please add how transmission was measured and how MTB diversity was defined.
2. The results section says that spoligotype analysis identified clustering among 166 of 199 isolates. However, the result section of the manuscript says that 147 were grouped in 22 spoligo-based clusters. Please clarify.
3. In the result section, please add the total number of cases included.

Methods:
1. Spoligotyping: Please include the reference formatted as required by the journal.
2. Fingerprint analysis:
a. Please change subtitle to DNA fingerprint analysis.
b. Please clarify how demographic information was compared with the international spolDB4 Data base. If this was performed, please describe in the result section.
c. In the same sentence reads as MIRU VNTR plus software was used to compare with the international Spold DB4.0. Please rephrase.
d. Please review and rephrase the sentence: “Identical spoligotypes and MIRU-VNTR patterns were considered to be genotypically clustered a cluster”.

3. Statistical analysis
   a. Please clarify what was subjected to the quality control described in this section.
   b. Please provide the package used for the statistical analysis.

4. Please add a section with definitions of the following terms:
   a. TB due to recent transmission.
   b. Alcohol and drug abuse.

Results
   a. Please move the following sentence to the end of the paragraph as it refers to the spoligotypes and the sentences before and after are referring to the lineage: “Among the identified strains, 147 (87.0%) were grouped into 22 spoligo clusters, with cluster sizes ranging from 2 to 34 (Figure 1)”.
   b. The percentage of resistant to S is 8.6% not 8.5%.
   c. Figure 1. There are just 24 ST in the LAM lineages and the text describes 25. Please clarify.

Discretionary Revisions

Methods
1. Spoligotyping: this is a standard method so the methodology can just be referred.
2. MIRU VNTR: this is a standard method so the methodology can just be referred.

Level of interest: An article of limited interest

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests