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: 2 Date: 17 April 2013

Reviewer: Midori Kato-Maeda

Reviewer’s report:

- Major Compulsory Revisions

Overall concept of the manuscript:

1) The authors incorporated our suggestions to eliminate the aim related to the molecular epidemiology of MDR among the individuals with no risk factors for MDR. However, there are several sections in the paper in which this concept was left - as a result, the manuscript confusing. For example, the title includes the phrase “low clustering rate” which, as commented by the 2 reviewers, can’t be measured in their study. The conclusion in the abstract also mentions the phrase “low clustering” and similar phrases are in the discussion, conclusion and summary. Based on the aim of the current manuscript there is no reason to mention this topic.

2) The current aim is to identify predominant circulating TB lineages in a population without factors for MDR TB. However, the manuscript does not contain information in the introduction or in the discussion to justify why their study is important. The authors should modify their introduction and discussion to describe why it is important to determine the lineages of MTB in ONLY individuals without risk factors for MDR.

3) In the result and discussion sections, the authors describe the association of MDR (or resistance to INH which is what the authors analyzed in this paper) and a specific lineage of MTB. However, this association was measured in just a fraction of the patients with MDR. Unless the authors provide a clear justification of why MDR MTB diversity in individuals without MDR risk factor is important, the current description is incomplete. Therefore, the authors need to strengthen their introduction to clarify this point.

4) Introduction: Authors should re-write the introduction justifying why it is important to determine the lineages of MTB in individuals without risk factors for MDR. It should also include why age (which is one of the variables studied in relation with lineage) is important.

5) Methods: Specify that the quality control was done for the double entry of the data to the database.

6) Methods: Authors should include in their analysis variables that have been associated with bacterial diversity such as the race/ethnicity and geographic origin of the patient.
7) Results: The description of study subjects is not very clear. The text says that 376 were eligible, but based on the point to point response document, these patients were eligible for the cohort study but not for the study described in the manuscript. The authors should clarify and distinguish between the study population of the original cohort, and the study population described in the manuscript.

8) Results: Please explain why the association of lineage was just restricted to isoniazid resistance and not rifampin. Based on the rest of the text, MDR is one of the topics of the manuscript.

9) Results: Please present the data analysis using the other data obtained in the study (sex, risk factors for TB, household contacts are listed in the methodology).

10) Discussion: The authors should re-write the discussion and focus it on the bacterial population diversity. It should describe why it is important to know the diversity ONLY in individuals without risk factors and why their results matters.

11) Discussion: the sentence regarding that - these data suggest a trend for an increasing prevalence of the Haarlem genotype in the North-Eastern area of Lima- should be made with caution, as this study just included patients without risk factors for MDR. In order to determine the validity of this comment, the authors should provide information to determine about the representativeness of their sample regarding all cases of TB in the North-Eastern area of Lima in the results and probably comment on them in the discussion (i.e what is the % of TB cases diagnosed in the public sector vs not public sector 2) what is the % of TB cases with and without risk factors for MDR, 3) describe the characteristics of those included in the study vs not included, etc.

12) Discussion. All comments regarding molecular epidemiology should be eliminated, including the first sentence of the discussion.

13) Discussion. In page 8, the first paragraph, clarify what is the risk factor it is referring to.

14) Discussion. In page 8, last paragraph should state clearly that there was no evidence of a predominant clone because the study was not designed to study the transmission and pathogenesis dynamics.

15) Discussion. In the limitations section, the comment about clustering rate should not be included as this is not a study designed for this purpose.

16) Discussion. The last paragraph in page 9 is related to a molecular epi study and should be eliminated.

17) The summary and conclusion should be re-written focused on the bacterial diversity in this specific population and the impact of their findings.

Minor essential revisions

18) English: There are several grammatical errors through the document. For example, Isoniazide (should be isoniazid), sensible (should be susceptible), among others.
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