Author’s response to reviews

Title: Multidrug-resistant Mycobacterium tuberculosis genetic diversity characterization and molecular epidemiology in Minas Gerais State, Brazil

Authors:

Nayanne G T Dantas (ngtdantas@gmail.com)
Phillip N Suffys (psuffys@gmail.com)
Wânia S Carvalho (waniasilvacakarvalho@gmail.com)
Harrison M Gomes (magdinier@gmail.com)
Isabela N Almeida (isabelanalmeida@gmail.com)
Lida J Assis (lidajouca@hotmail.com)
Pedro D Navarro (pedrodaibert@yahoo.com.br)
Michel K Gomgnimbou (gomikir@yahoo.fr)
Guislaine Refregier (guislaine.refregier@u-psud.fr)
Christophe Sola (christophe.sola@igmors.u-psud.fr)
Silvana S Miranda (silvanaspindola@gmail.com)

Version: 3  Date: 6 June 2015

Author’s response to reviews: see over
Dear Dr. Eyal Oren,

Please find attached herewith a revised version of our manuscript entitled :

“Multidrug-resistant *Mycobacterium tuberculosis* genetic diversity characterization and molecular epidemiology in Minas Gerais state, Brazil”, by N. Dantas *et al.* which was improved thanks to the two reviewer’s comments.

We are pleased to see that both reviewers considered this paper to be a timely contribution to molecular typing of *Mycobacterium tuberculosis* and should be published into the BMC Infectious Diseases Journal.

We carefully considered all suggestions raised by the reviewers and provide herewith an accompanying letter with point by point answers to their comments.

We hope that our paper will now be acceptable for a rapid publication into the BMC Infectious Diseases Journal

Sincerely yours,

Nayanne Dantas
**Answers for reviewers report**

The answers are in highlighted in blue color.

**Reviewer: Omar Caceres**

**MAJOR COMPULSORY REVISIONS**

**ABSTRACT**

- Line 51. The authors not explain the 22.2% of the remaining samples

  answer: We now explained in line 54. *The 22.2% remaining samples were classified as unknown profiles.*

- Line 52-53 and 288-290: The paragraph is not clear, the 32 spoligotypes were found in 88 isolates (16 unique and 16 common)? Therefore 16 isolates are missing? Please verify

  answer: We clarified this paragraph in the lines 48-50 and 283-285. “Thirty-two different spoligotyping profiles were found: 16 unique patterns (n=16) and 16 clustered profiles (n=88).”

- Lines 54-56. The sum of the drug resistant profiles of the 104 samples is 100, four samples profiles are missing. Please verify.

  answer: We corrected this sum in the lines 55-57. The four samples corrected were two strains susceptible for both drugs and two strains that were not successfully amplified (sum=104).

- Lines 56-57: The authors report rpoB mutations found in their samples but these results are not described in the corresponding section (TB SPRINT, line 283). Please describe it in more detail. This paragraph raise the following question, only one RIF mutation was found in each sample? Usually more than one RIF mutation are found in this kind of samples. Similarly, this occurs with INH mutations too. Please describe in the corresponding section.

  answer: We now describe the rpoB mutations found in detail the corresponding section (in the lines 304-307). No, it happened in one sample that we found three RIF mutations (in the lines 307-308). In INH mutations we found more than one mutation in 15 isolates (14 katG 315ACC + inhA-15 and 1 katG 315AAC + inhA-15) as described in the lines 311-315. We did not characterize double mutations on *rpoB* by TB-SPRINT, given the design of the method, such isolates with double mutations should also have been characterized if present, however the reviewer raised an interesting issue relative to such isolates, that, if characterized previously by sequencing, could later be studied deeply by TB-SPRINT.
METHOD
Strain isolation and drug susceptibility testing

To have a better idea of this study, it is necessary that 104 analyzed had epidemiological data as age of the patients from where the samples were obtained, gender, clinical data, DST, etc. The authors refers that they collected demographic data but they don't show these data. Please put theses data in result sections.

The demographical and epidemiological data (age, gender, clinical data and DST) were added in the result sections in the lines 241-247. Within the Methods section (in the lines 232-234) we explain that city and address were used and the only personal data that were analyzed was the mother's name. Due to ethical restrictions we could not put these two types of data in results section. However, the geographical results (city) are now shown in Figure 2.

- Line 156: How many clinical isolates were analyzed? This clinical isolate belong to the 104 samples analyzed or are others samples. Please clarify

We analyzed 104 clinical isolates. Yes, this clinical isolates belong to the 104 samples analyzed. We clarified this phrase in the lines 156-160.

- Line 169: Please put the concentration of the CTAB used

The concentration of CTAB now is indicated: 10% (in line 173).

- Lines 188-189: The authors refers that they evaluated the samples using MIRU-VNTR 24 following a previous report but they put four bibliographic references. Which of the references they followed? Or the authors improved the method, taking data of these references. If the latter is true, the authors need to describe this method.

Answer: We now clarified this point in the lines 190-192. The technique was performed following the reference [43] and using the twenty-four loci MIRU-VNTR specified in the reference [18].

- Lines 204-205: The paragraph is confuse, the authors need to improve the idea.

Answer: We now improved the idea of this paragraph (lines 201-204).

- Line 205: Please verify the reference

Answer: We corrected the reference (line 204).

- Line 223-224: The paragraph is confuse, the authors need to improve the idea.
RESULTS AND DISCUSSION

- Lines 252-254: What happened with the 35% of the samples remaining? The authors need to explain the question

  Answer: The 35% of samples remaining didn’t show any IS6110-RFLP profile. To explain this question we added one sentence (lines 256-258).

- Line 273: Please verify the reference 60 is related to the paragraph

  Answer: We corrected this reference and replaced it by reference 53 (line 272).

- Line 279: Please verify the reference 50 is related to the paragraph

  Answer: We corrected this reference and replaced it by reference 64 (line 277).

- Lines 294-295: the redaction of the paragraph is not clear. Please clarify.

  Answer: We now improved this paragraph (lines 287-288).

- Line 295: The clinical isolates are not visualized in the MST graphic. Please improve the graphic.

  Answer: We improved this figure by introducing specific SIT label for each group that were associated with the clinical isolates shown in the Figure 2.

- Line 304: The authors must explain what happened with the 2 isolates in the RIF-INH typing

  Answer: These 2 isolates didn’t amplifed at any position tested. We now explain this fact within the text (lines 300).

- Lines 307-308: From 98 samples 71 have RIF resistance mutations, the 27 samples remaining are susceptible or has other RIF mutations? Please clarify. Similarly, from 102 isolates evaluated for RIF resistance 98 had non-wild-type genotype, the 4 samples remaining had wild-type genotype? Please clarify too.

  Answer: We corrected these results. From 98 samples 78 have RIF resistance mutations, 20 carry other RIF mutations and 4 had no mutations (lines 300-304).
Line 313-314: According to the sentence, the reference 64 is not related. Please verify

Answer: We corrected these references.

- Lines 316-317: The paragraph is confuse, please describe in more detail.

Answer: We now improved this paragraph (lines 317-319).

- Line 323: What happened with the 2 samples remaining? Please clarify. Similarly, in the line 325 Why only 94 of 104 isolates were evaluated by 3R-SNP typing?

Answer: The 2 samples remaining were classified as unknown lineage (line 321-322). In fact all of 104 isolates were evaluated by 3R-SNP typing but 10 isolates could not be typed successfully (lines 325-326).

- Lines 327-328: These 12 samples with unknown lineage were analyzed by other methods described in this study?

Answer: These 12 samples were analysed by the other methods described in this study. Nine of them showed discordant classification by MIRU-VNTR and TB-SPRINT whereas three were concordant (2 S and 1 LAM). Discordance between MIRU-VNTR and spoligotyping classification is the focus of an accepted study by J. Aze et al. to be published in PloS ONE soon (PONE-D-15-07333R1, Genomics and machine learning for taxonomy consensus: the Mycobacterium tuberculosis complex paradigm). Studying discordance between MIRU-VNTR and spoligotyping was not the focus of this study, hence, these discrepancies were not investigated further within this study.

- Lines 336-337: The authors analysed 4 molecular methods to determine genetic diversity of M. tuberculosis (see line 194) but in these lines the authors point out 5 methods, please clarify. Moreover, there is a mistake in the number of table, must be table 3 instead of Table 2.

Answer: the difference between 4 and 5 methods stems from the fact that we computed HGDI separately for spoligotyping and RIF-INH typing whereas these two methods can be run simultaneously. It was clarified in the lines 334-335.

- Lines 338-339: The redaction of the paragraph is not clear, please clarify.

Answer : We now improved this paragraph (lines 335-337).

- Lines 340-341: The authors must explain the discriminatory power of the other methods described in this study.
Answer: We now provide more interpretation on the discriminatory power of each method used (lines 338-340).

- Lines 346-347: How the authors demonstrate that the transmission is not recent in the study area? If the authors not show epidemiological data of samples analyzed.

Based on a tight or smooth cluster definition and cluster analysis, we now provide a much deeper epidemiological interpretation of our results, based on the computation of the recent transmission index (RTI) with one or the other cluster definition. Facts are that, if we adopt a strict cluster definition, we have no case of recent MDR-TB transmission as shown by the final dendrogram (no strain perfectly identical by all methods simultaneously). In that case, we could exclude case of MDR-TB recent transmission on Minas Gerais. However, the smooth cluster definition allows to suggest up to 20% of recent transmission of MDR-TB. Consequently, it is worth to complement this study by epidemiological inquiry on the clusters defined at >85% similarity. It was clarified in the lines 342-369.

- Lines 350-352: Similarly to the paragraph above, how the authors infer these results if they not show epidemiological data.

Answer: We improved the result section adding some epidemiological and demographic data (lines 240-247 and 364-369).

- Lines 368: Apparently there is a contradiction between the descriptions pointed in lines 346-347 and this sentence. Please clarify.

Answer: We now corrected this contradiction (lines 384).

- Lines 368-369: Apparently there is a contradiction between this paragraph and the samples with unknown lineage not resolved by 3R-SNP typing. Please clarify

Answer: We deleted this paragraph.

REFERENCES

- Please verify the correct separation of lines between the following references: 1 and 2; 36, 37 and 38; 40 and 41; 58, 59 and 60.

    Corrections were done.

MINOR ESSENTIAL REVISION

- Line 77: Please delete: “M. tuberculosis”
Done, as suggested by the reviewer (line 84).

- Line 82: I suggest change “M.tuberculosis” word by MTb
  Done, as suggested by the reviewer (in all paper).

- Line 112: Please correct the letter “e” by “be”
  Done, as suggested by the reviewer (line 113).

- Line 301: the reference 54 are repeated
  Deleted the reference repeated.

- The reference 67 is missing
  We introduced this reference number.

- Line 771: In Table 1 the letter “a” is missing
  Done, as suggested by the reviewer.

- Line 796-797: Please correct the title of the table according to the correction pointed out in lines 336-337
  Done, as suggested by the reviewer.

- Line 838: Please correct the word “perio”
  Done, as suggested by the reviewer.

- Line 846: Please complete the word “3R”
  Done, as suggested by the reviewer.

- The authors need to improve the quality and resolution of the MST graphic and dendrogram (figure 1 and 2)
  Done, as suggested by the reviewer.

- Figure 1: The legend of Haarlem lineage (H) is missing in the MST graphic
  The legend of Haarlem lineage (H) was added in the MST graphic.
Answers for reviewer report

The answers are in highlighted in blue color.

Reviewer: Sabrina Rodriguez-Campos

Major Compulsory revisions

1. In general, the conclusions are poorly worked out. The authors merely focus on the usefulness of genotyping, which has been shown in numerous reports. The authors should highlight the importance of their findings instead: in the context of previous genotyping studies from Brazil and in the context of the global M. tuberculosis situation.

   We have considerably improve this part (lines 379-393) by a deeper epidemiological interpretation of our results, at the light of RTI (Recent Transmission Index) computation based on two clusters definition, one smooth, one tight, and a deeper discussion on how to interpreter these results (lines 342-374).

2. lines 367-368: the authors conclude that one third of the MDR strains belonging to the LAM lineage showed recent transmission. In lines 346-347 the fact that two thirds did NOT show recent transmission in contrast to other studies from Brazil seems more important (?). Moreover, it is unclear how the authors define recent/not recent transmission (see also). Please clarify.

   This question raises the fact that LAM is extremely difficult to characterize genotypically, and similarly as it is the case for Beijing in Asia, molecular epidemiology in Brazil requests either combinations or methods or would request whole genome sequencing for a better understanding of tuberculosis epidemiology. We emphasize the recent transmission definition in our methods and discussion (lines 219-220 and 342-353).

Minor Essential Revisions

1. The two and a half pages of background could easily be reduced by summarizing the techniques, not reviewing them.

   We could not reduce the background too much but try to reduce the one on the techniques by summarizing ancient information.

2. Please revise the use of italics, e.g. lines 39, 182, 364 -> IS6110, line 144, 306 -> kat G, inhA, line 65, 368 -> typing, line 370 -> spoligotyping.

   We revised the use of italics as suggested by the reviewer.

3. Use a period not a comma for decimal points, e.g. line 258, 325 ->
5.97%, 90.38%.

Done, as suggested by the reviewer (lines 255 and 323).


We added this reference (line 121).

5. line 111: unless you want to specify the detection limit "(nanograms)" is redundant.

Done, as suggested by the reviewer (line 112).

6. line 152: what is meant by "convenience samples"? Were these chosen at random? Arbitrarily? Are the samples representative?

We define our sample as a "convenience samples" since the design of our study was not done based on sampling calculation based strict epidemiological statistical design. Essentially, our study is descriptive. However, having said that, our sampling is fully representative of clinical isolates identified as MDR-TB collected in Minas Gerais State between 2008 to 2013. We used all MDR-TB samples from Ezequiel Dias Foundation that were collected between 2008 to 2013; this Institution is responsible to identify the MDR-TB strains in Minas Gerais State and all these samples are conserved in the Mtb clinical isolates collections. It was clarified in the lines 162-163.

7. lines 228-229: this sentence is incomprehensible.

We rewrote this sentence (221-224).

8. lines 257-259: check sentence structure "number (5.97%) number (very high...)"?

We improved this sentence in the lines 254-256.

9. line 278: "As well as found in this study..." Do you want to say "As supported by this study..."?

Yes. We changed according as suggested by the reviewer (line 274-275)

10. lines 308-309: I would suggest to write "A SNP referred to as 531TTG mutation" for better comprehension of the SNP paragraph. Delete "like" at the end of sentence.

We improved this sentence in the lines 304-308.

11. line 325: I would suggest "Out of 104 isolates 94 (90.38%)..."
Done, as suggested by the reviewer (line 327).

12. lines 361-362: "the effect of this genotype" Which genotype?

We deleted this sentence.

13. lines 370-371: TB-SPRINT was used as a typing method and was not carried out in parallel to test its performance, the authors could by the most conclude that this technique was useful to type the isolates.

The usefulness of the TB-SPRINT technique was already shown in a recent study done in Pakistan (cf. M. Yasmin et al. Infect. Genet. Evol. 2014). Comparing the TB-SPRINT performances to the ones of existing methods (HAIN-MTBDR, GenExpert, etc…) will be the focus on another study.

14. line 768 (Table 1): Do you mean "diversity"?

Yes. We replaced “diversification” by “diversity” in the Table 1.

15. lines 792-793 (Table 2): "Phylogenetic classification by spoligotyping patterns of multidrug resistant Mycobacterium tuberculosis strains (n=104)" Does this table only refer to the TBSPRINT results as suggested in line 286? Then please specify this in the table heading.

Yes, this table only refer to the TB-SPRINT results. We specified the table 2 heading: “Table 2. Phylogenetical classification by spoligotyping patterns of Mycobacterium tuberculosis multidrug resistant (n=104) by TB-SPRINT method".