JPHE JOURNAL OF PUBLIC HEALTH AND EMERGENCY
AN OPEN ACCESS JOURNAL FOR HIGH-QUALITY RESEARCH IN PUBLIC HEALTH AND EMERGENCY MEDICINE

Peer Review File

Article information: http://dx.doi.org/10.21037/jphe-20-10

Reviewer A
In the underlying manuscript, the authors sought to define the rate of TB recurrence in Stockholm region, and differentiate between relapse and re-infection by using whole genome sequencing of first and second episode TB isolates.
Overall the manuscript is very well written and clearly structured. I have few comments that needs to be considered.

Comment 1
A crucial question in apparent re-infections is always a mixed infection at baseline with one strain only present at very low frequencies, especially when the MDR strain is found in the second episode. The authors should investigate the presence of resistance-related and few synonymous benign SNPs that are specific for the second strain in the isolate from the first episode. Indeed, they discuss this point later, but an analysis is missing.

Reply
Our "resistance pipeline" reports all variants (in the pre-determined set of genes) which fulfill the quality criteria. Minorities with a frequency ≥10% in the isolate from the first episode would therefore have been reported in our output data (for example, a SNP with a frequency between 10-20% in the first isolate and then ~100% in the second isolate), but we did not see any such cases. Also, going below 10% might not be very informative if the average sequencing depth lies between 20 and 30x. Moreover, we did not have any case where a patient was re-infected with an MDR strain.

Changes in the text
No change was made in the text following Comment 1.

Comment 2
The fact that the pair with 7 SNPs distance has an identical very rare katG deletion indeed suggests that this is a relapse case as the authors point out. This touches the problem with strict cut-offs in molecular epidemiology. Also, here it would be interesting if the seven new SNPs are not already present at low frequency in the first episode and just by chance another sub-population was detected at baseline. Further,

Walker et al 2013 showed that even in well documented transmission chains in the UK, differences up to 12 SNPs are possible.

Reply
To double-check this specific case, we manually investigated the mapping of the first isolate. Six of the SNPs were not detectable even at very low frequencies, but for the remaining SNP we noted some ambiguous calls and when we repeated the SNP analysis using a different reference genome (CP003248.2), the SNP difference between the first and second isolate was reduced to six (the "ambiguous" seventh SNP was not reported).

Changes in the text
The manuscript has been updated: See Page 10, line 242-246 (changes in the text are highlighted in yellow).

Comment 3
The authors mention a minimum spanning tree in the method, used to analyze the samples. That should be presented in the manuscript.

Reply
The analysis was not dependent upon the minimum spanning trees.

Changes in the text
The method section has now been rephrased: See Page 8, line 179.

Comment 4
In the logistic regression analysis, the response variables are a bit confusing e.g. TB recurrence should be the outcome of interest. And then old age would be indicative as a protective factor for TB recurrence. Further, the results are not mentioned in the text.

Reply

We are grateful for the comment. However, that TB recurrence is the dependent variable in the logistic regression analysis is mentioned both in the statistical analysis paragraph in the methods section and in the table text to Table 2. We are embarrassed to admit that we by mistake presented the two age groups in the wrong order. This fault has been corrected in the table. It is correct that persons diagnosed with TB <35 years of age had a somewhat higher risk of TB recurrence (odds ratio 2.20 (95%

confidence interval, 0.91–5.34), nevertheless we have chosen not to comment this finding as the difference in risk for the two age groups did not reach the level of statistical significance (p=0.08).

Changes in the text
Table 2 has now been revised.

Comment 5
The WGS results paragraph could be improved by better describing relapse vs re-infection cases. How many SNP differences in both groups, how often did the genotype change or remained the same in re-infection cases. It is also interesting that some former MDR patients got re-infected with a susceptible strain. What was the average time between first and second TB episode in both categories?

Reply
We believe that the information regarding SNP differences and genotypes is clearly stated in Table 3 and that the time between first and second TB episode in both categories is presented in earlier in the results section (see Page 10, line 229-231).

Changes in the text
No change was made in the text following Comment 5.

Comment 6
The SNP threshold for distinguishing between re-infection and relapse is only first mentioned in the results part. Please also include it in the method section.

Reply
The SNP threshold is for sure described already in the method section under the paragraph, Study definition (Page 7, line 153)

Changes in the text
No change was made in the text following Comment 6.

Comment 7
The contamination in one sample seems to be an issue for the SNP calling pipeline. Normally non-TB reads are already filtered out in the reference mapping, and particularly with the high frequency cut-off of 90% should not occur anymore in the

extracted SNP alignment. The authors should re-evaluate their mapping parameters for the IonTorrent reads or implement a test for mixed infections beforehand.

Reply
The 124 non-TB specific SNPs are located within the *rrs* gene (information added on Page 10, line 245) and we assume that is the reason why the non-TB reads were included in this specific sample. Although we agree that filtering out non-TB reads is desirable, we also think that a "noisy" *rrs* can be used as an indicator of contamination and therefore we do not want to be excessively strict in this sense.

Changes in the text
No change was made in the text following Comment 7.

Comment 8
There are still some placeholder words "XXX"

Reply
We are somewhat unfamiliar with the meaning of this wording and hope that the reviewer (or Editor) might explain to us what changes are suggested.

Changes in the text
No change has yet been made following Comment 8.

JPHE JOURNAL OF PUBLIC HEALTH AND EMERGENCY
AN OPEN ACCESS JOURNAL FOR HIGH-QUALITY RESEARCH IN PUBLIC HEALTH AND EMERGENCY MEDICINE

Reviewer B

In this work, the authors explored 2552 TB patients from Stockholm Country diagnosed between 1996 and 2016 to look for relapses or reinfection. WGS of Mtb isolates was performed in order to discriminate between these two situations. TB recurrence was defined as positive Mtb culture recovered more than 180 days after successful treatment completion. Following this definition, Tb recurrence was seen in 24 (0.7%) patients. WGS on both isolates (first and second episode) was possible for 17 patients out of 24. From these 17 patients, 12 (71%) cases were classified as relapse and five (29%) were regarded as reinfection, resulting in a relapse frequency of 0.5%, corresponding to an annual risk of relapse of 0.06%. Of note, half of the recurrent cases were linked to drug resistance and no additional mutations for drug resistance were detected at the second TB episode in patients with relapse.
The authors conclude that the recurrence frequency is low and mainly observed in drug resistant cases, thus stressing the need of improved treatment control in this group.
Overall this work conveys a useful massage, however the authors should revise some points prior to publication.

Comment 1

Authors should better define "successful treatment completion for MDR-TB patients".

Reply

Following the comment, we have clarified the definition of successful treatment completion in the Method section (see below). The definition includes MDR TB cases. We used the limit >180 days after treatment completion for recurrence also for MDR-TB patients. We have chosen not to specifically comment on MDR in this matter.

Changes in the text

“Treatment success was defined as *cure* if smear- or culture-negative in follow up samples or *treatment completion* if without evidence of failure and in abundance of follow up samples, according to the WHO definitions.” See Page 6-7. Line 149-151.

Comment 2

Reference 15 should be updated as it has been revised in 2020.

Reply

Notion taken.

Changes in the text
Reference 15 was updated as suggested.

Comment 3
A relapse case was defined as having a maximum difference of five SNPs. However, in table 3, Relapse case 1 displays 7 SNP distance while the re-infection cases had more than 150 SNP distance. In my opinion setting a cut-off at 5 SNP is probably too restrictive, as within-host variability may account for more than 4 SNP (Casali et al 2016). I would suggest that the authors adopt the 12 SNP cut-off proposed by T. Walker for cross-transmission investigation.

Reply
The 12 SNP cut-off could be adopted and we agree that it might be suitable in a low-incidence setting (lower risk of re-infection with a closely related strain), but we would also argue that the risk of true relapses ending up at the "wrong" side of the cut-off remains even if a 12 SNP cut-off is adopted (particularly when considering the intra-host variability and the large time-span between the episodes for some of our patients, eleven years in one case).

Changes in the text
No change was made in the text following Comment 3.

Comment 4
The authors state that "Variants were filtered for sequencing depth: ≥10x and frequency: ≥0.9", whilst lines 140 - 142 there is stated that for the detection of resistance mutations "variants were filtered for a minimum frequency of reads calling SNPs: 25%". Why would the authors operate a lesser filter for resistant determinants? By setting high filters on the regions of genome not known as canonical resistant genes, don't the authors risk missing SNP of difference and thus biasing the conclusions of the study?

Reply
We agree that this might look confusing (the explanation is that we used a previous version of the "resistance pipeline" in the analysis) and we have now re-analyzed the resistance genes with a minimum frequency set at 10% obtaining the same results.

Changes in the text
The manuscript has been updated: See Page 8, Line 188.

Comment 5
"In the remaining 36 patients." unfinished sentence

Reply
Notion taken. The unfinished sentence was left by mistake

Changes in the text
The line was deleted.

Comment 6
Line 170 - 171 "Out of the 24 cases with a culture confirmed second episode the isolate could be obtained for WGS", whilst line 183 "In patients analyzed with WGS (n=17)..." this is confusing for the reader, were 24 or 17 recurrence cases analyzed by WGS?

Reply
Out of the 24 cases with a culture confirmed second episode 17 were analyzed by WGS. To minimize the risk for confusion we have added a short explanation to the first sentence.

Changes in the text
The manuscript was changed: See Page 9, Line 218-220.

Comment 7
In the text, there is a description of a contamination by NTM resulting in a first misidentification of 125 SNP, which in fact were only one SNP difference after re-analysis. In my opinion, these details are not of interest for the reader, for relapse case 3 only 1 SNP should be mentioned in the table, as it is expected that only "pure" Mtb isolates should be used for WGS.

Reply
The notion is taken and the statement was removed from the text in Method section and moved to a footnote to Table 3.

Changes in the text
See above.

Comment 8

The authors should also take into account intra-host variability when discussion genotypic discrimination between relapse and reinfection based on SNP distance.

Reply
Notion taken. Text addressing the intra-host variability and the importance of sequence quality has now been added

Changes in the text
Page 13-14, line 322-325.

Comment 9
The authors should stress the importance of the quality of sequencing when searching for minority populations

Reply
See our response to Comment 7 by the first reviewer.

Changes in the text
No change was made in the text following Comment 9.

Comment 10
"In 24 of the registered 60 cases with a second TB episode, the second case was not culture verified. This could be explained by early diagnosis made on a positive PCR result or clinical suspicion, without culture confirmation." Diagnosis of TB recurrence based on PCR should be cautious as PCR could remain positive in some patients several month or years after a first TB episode, even in cases of cure.

Reply
Also this notion is well taken and the statement that “PCR can also remain positive several month after treatment completion without viable *M.tb*.” has been added to the text

Changes in the text
Page 15, line 364-365.

Comment 11
Importantly, Mtb genome sequences of the 34 isolates should be deposited and referenced prior to publication.

JPHE JOURNAL OF PUBLIC HEALTH AND EMERGENCY
AN OPEN ACCESS JOURNAL FOR HIGH-QUALITY RESEARCH IN PUBLIC HEALTH AND EMERGENCY MEDICINE

Reply

The sequences have now been deposited at ENA under study accession number PRJEB38721.

Changes in the text

Page 7, line 171-172.