Reviewer's report

Title: Molecular epidemiology of hepatitis B virus infection in Norway

Version: 0 Date: 14 Dec 2018

Reviewer: Asgeir Johannessen

Reviewer's report:

This study presents the genotype distribution and related epidemiological data from more than 1100 patients with HBV infection in Norway. The methods seems adequate and the results are presented with informative figures including a phylogenetic tree, which is the main result of the paper. The paper is worth publishing; however, I have some minor essential points for clarification, and importantly the conclusion chapter needs to be re-written:

1. Abstract:
   
a. Results: it is very general. I would prefer some more details. B, C and E were from Asia and Africa, can you specify? For example that xx% of B and C were from East Asia (or South-East Asia). Or that E was from West Africa.

b. Conclusion seems completely unlinked to the Results? You have not mentioned anything in Results about where patients said they were infected. Don't introduce new information in the Conclusion. Please re-phrase based on your results. The sentence currently in Conclusion must either be moved to results or deleted.

2. Introduction:
   
a. 2nd paragraph, line 63-64: According to reference 5 there are not that many countries with more than 8% prevalence. Please be more precise.

b. 3rd paragraph, line 94-96: You might want to give a few examples of clinical implications of genotype, for example that geno C is associated with an increased risk of HCC, or that geno A has a favourable response to treatment with pegylated interferon.

c. Last paragraph, line 99-101: "In Norway" twice in one sentence, remove one.

d. Last paragraph, last sentence, line 104-106: I don't really see how your epidemiological study will help in patient management?
3. Methods:

a. Study group:

i. Better "Study subjects" or "Study population"

ii. Line 112-115: Do you need to specify this? They were very few samples (14 + 8) and from the same time period. And the Norwegian name of the clinic is not relevant. I suggest to delete the whole sentence.

b. Sequence alignment, line 151-153: The number of patients who were infected in Norway vs. abroad belongs in the Results section.

c. You should include a subsection called "Statistical analysis"

4. Results:

a. 1st paragraph, line 178: it is ambiguous, are B and C together 18% or are they 18% each. Please correct.

b. 1st paragraph, line 185-186: The sentence "Among these cases 50%..." is not very informative, I suggest to delete it. Figure 1 gives this information.

c. 2nd paragraph, line 198: "...different countries in Africa". All over Africa? Or only sub-Saharan or West or Central or Southern? Please specify (if possible). I thought E was rare in East Africa for example.

d. 3rd paragraph, line 200: "Information on transmission route..." should be changed to "...likely transmission route...".

e. 3rd paragraph, line 204-207: The sentence "In addition, the 14 strains..." could be omitted, since I assume these are included in the patients who claimed to be infected in Norway? If not, then I suggest to lump these together. It is only confusing to add another separate group. The next sentence (gender distribution) should appear at the very beginning of the Results section (among the initial 2-3 sentences), not here. Finally, the genotype distribution in men and female is not very interesting, so I suggest to delete this sentence together with figure 3.

f. Instead of the gender analysis perhaps you could say something about the time trend over these 30 years? Was the genotype distribution different in the first decade compared to the last?

5. Discussion:

a. 1st paragraph, line 232-233: No need to repeat the percentage again and again. You can simply write that geno D was the most prevalent, followed by geno A, B and C.
b. 2nd paragraph, line 235: Write "...in Europe (including Norway), Afghanistan..." instead of "...in Europe, Norway, Afghanistan..."

c. 2nd paragraph, line 242-245: This is one of your main findings and I think it deserves more explanation. How do you see that there have been single origin outbreaks for example? Try to explain so that the paper can be read by people with less molecular biology background. This also applies to the same topic in the final paragraph in the Results chapter.

d. 2nd paragraph, line 246: Delete "extensive" as it is more political than scientific.

e. 3rd paragraph, line 256-259: This sentence can be incorporated in the next paragraph.

f. 4th paragraph, line 266-268: Is it not possible that the clustering with non-Norwegian strains means that they have been infected in Norway by someone from Asia? Vietnamese people in Norway are probably more likely to have close contact with other Vietnamese people?

g. 4th paragraph, line 275-280: These data are so incomplete (76% missing data) that you better delete it.

6. Conclusion: This is weak. I think you should work on this to make it a summary of your study, rather than general remarks. For example, the need to genotype patients with chronic HBV infection is controversial, I personally don't think it is needed in inactive carriers (which is in line with the Norwegian HBV guidelines). The last sentence belongs in background, not conclusion.

*Are the methods appropriate and well described?*
If not, please specify what is required in your comments to the authors.

Yes

*Does the work include the necessary controls?*
If not, please specify which controls are required in your comments to the authors.

Yes

*Are the conclusions drawn adequately supported by the data shown?*
If not, please explain in your comments to the authors.

No
Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal