Reviewer's report

Title: Molecular Characterization of Hepatitis B Virus in Liver Disease Patients and Asymptomatic Carriers of the Virus in Sudan

Version: 1 Date: 13 May 2013

Reviewer: Vincent Thibault

Reviewer's report:

This manuscript provides an original and interesting molecular characterization of HBV isolates from Sudan. There is no concern about the originality of the work as very few data are available on HBV molecular epidemiology in Africa, in general. The paper reads well and is well presented except for some minor points highlighted below. The main weakness lies certainly in the overall description of the methodology and most points are detailed in the following comments. Authors should also be careful when interpreting clinical data from this small cohort as each group includes rather low numbers and statistical interpretation of such data is questionable. It is probably advisable to remain purely descriptive in this context. I am convinced that an effort in the presentation of the data would really improve the quality of the manuscript.

Specific comments

1- Methods (P5-6): in order to strengthen the scientific impact of their work, the authors should provide more details on their methodologies. Please, specify what exact region was amplified in the "short region of the S" (even if it is from a previous paper). The modifications applied to the method described initially by Günther et al. should also be briefly described.

2- In line of previous comment, a general outline of your approach should be proposed either as a figure or in the text (table?). How many genomes could be fully sequenced and if not, how many were only sequenced in the S gene or only analyzed by RFLP? You report on only 4 full genome analyses: is it because you only managed to perform full amplification of these samples or is it because you decided for some reasons (to be specified) to further document these very samples? (Page 8)

3- For each applied strategy, the genotype (and subtype) distribution should be reported. This would insure that your technical approach did not bias the genotype distribution. Please provide as much information on the 18 samples that could finally not be genotyped.

4- All primers used should be gathered in a table indicating exactly the one used as forward or reverse primers. "In addition to the sequencing primers used previously..." is not really informative to the reader and it is therefore difficult to understand the overall strategy. The same applies to the BCP/PC region amplification strategy that is very superficially described.

5- Results: overall, it is very difficult to figure out how many samples could
effectively be sequenced in each region of the genome and from which region the genotype was deduced. An effort should be made to present a clear overview of the sequencing and phylogenetic analyses. Were the S and BCP/PC phylogenetic analyses fully concordant? If not, this information should clearly be stated. If your strategy had been clearly exposed, it would have been easier to understand your analysis and the proposed conclusions. For instance, (discussion p11) you state that some samples were classified as A1 based on BCP/PC analysis; why were those samples not analysed using the S region? Many questions arise from the imprecise description of your strategy.

6- Table 1: it appears that all included samples could not be genotyped. The numbers of not genotyped samples (failure) should be reported and percentages should take into account these failures. So table 1 is misleading as the percentage is calculated according to genotyped samples from each group rather than percentage of the entire included samples of each group.

7- Since previous works have reported genotype D/E recombinants, why did you not include genotype E in your Simplot analysis? Are there more reasons to include other genotype D subtypes than genotype E?

8- The abstract should report how many samples were finally genotyped, as only 81 of the initial 99 (82%) could be genotyped and only 53 of the 81 (65% or 54% of the initial population) could phylogenetically be characterized. The same remark applies to the discussion (page 11). While 99 samples were considered, genotyping could be performed for 81 of them; all numbers should be rigorously reported in order to convey a scrupulous message.

9- Page 9: The rtT237P motif has been first reported in most genotype "D7" (D8?) strains (Meldal, J. Gen. Virol. 2009). This information should be discussed in light of their findings, as well as the preS2 motif I42T.

10- Discussion (page 11). The last part of the first paragraph is not really informative and compares patients from different studies; its pertinence is questionable. The viral load in HCC patients is not significantly higher than in the other groups; it is therefore not an outstanding finding. Could the higher viral load observed in the HCC group, be due to the higher percentage of genotype E in this group? Since high ALT levels were one of the criteria to identify AH, it is expected to be higher in this group than in the other. This finding is not very impressive. Moreover, the AH group includes only 7 patients; it is difficult to conclude anything on so small numbers.

11- Figure 1: the little box representing a phylogenetic tree of complete genomes is not well introduced in the legend. All information regarding this specific tree should be mentioned.

12- Figure 3: the genotype distribution (according to the analyzed region S or BCP/PC) for each mutated region should be shown. This information would help to understand the discussion (page 13) where a discrepancy between S and BCP/PC for 3 isolates is raised.

13- Discussion, page 13: I believe it is rather cloning followed by sequencing rather than the opposite as written.
14- Discussion, page 13: you cannot state "there was a significant correlation between the pre-S2..." as you cannot apply any statistical test to justify it, mainly because you have too small numbers. So, your sentence should remain descriptive; yet, I agree that this deletion was only observed in HBV-E/HCC patients.

Minor point:
1- Table 1: the male:female ratio is improperly expressed on 2 columns (37:05:00...).
2- Discussion, page 12: "In agreement with Sudanese blood donors..."; this sentence requires attention.
3- Is figure 3 clearly needed?

**Level of interest:** An article of importance in its field  

**Quality of written English:** Acceptable  

**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