Author’s response to reviews

Title: Comparison of Sanger sequencing for Hepatitis C Virus Genotyping with a Commercial Line Probe Assay in a tertiary hospital

Authors:

Sylvie Goletti (sylviegoletti@hotmail.com)
Siméon Zuyten (szuyten@gmail.com)
Léonie Goeminne (leonie.goeminne@uclouvain.be)
Hector Rodriguez-Villalobos (hector.rodriguez@uclouvain.be)
Monique Bodéus (monique.bodeus@uclouvain.be)
Peter Stärkel (peter.starkel@uclouvain.be)
Yves Horsmans (yves.horsmans@uclouvain.be)
Benoît Kabamba-Mukadi (benoit.kabamba@uclouvain.be)

Version: 1 Date: 22 Jan 2019

Author’s response to reviews:

Sir,

Would you please consider the submission of our revised manuscript, “Comparison of Sanger sequencing for Hepatitis C Virus Genotyping with a Commercial Line Probe Assay in a tertiary hospital” (INFD-D-18-01576). We responded to the requests of the reviewer as specified in the attached document.

We feel this paper should now be acceptable for publication in BMC Infectious Diseases.

Yours sincerely

Sylvie Goletti

Reviewors comments to the author:
Reviewer’s comment: C
Author’s response: R

Reviewer 1:
Major points

C1: Authors perform HCV genotyping by LiPA Assay and sequencing. Since they have HCV sequences, they should perform HCV genotyping by phylogenetic analysis using HCV reference sequences. Using the HCV genomic blast bank can give an initial genotyping result but it is not useful to identify HCV genotypes, especially when you have unidentified genotypes in your data set. Authors should perform phylogenetic analysis using at least two different phylogenetic methods and present phylogenetic trees using proper dataset of reference sequences.

R1: We agree with the reviewer’s comment, we did phylogenetic analysis on the sequences for the 3 genes (5’UTR, Core and NS5B) using Neighbor-Joining and UPGMA methods. Forty-seven reference sequences were used to make the trees. A paragraph concerning this analysis has been added in the method section (page 6 lines 25 to 27) as well as in the result section (page 8 lines 21 to 27). We also added the phylogenetic analysis in the discussion section (page 9 lines 30 to 33 and page 10 lines 1 to 7). Figures are added to illustrate this point (Figures 1 to 3).

C2: Authors should submit all sequences in GenBank database and provide accession numbers.

R2: All sequences have been submitted to Genebank, the assignment of accession number is ongoing

C3: It would be useful to present their results in a table showing the LiPA assay result in comparison with sequencing of various regions so as the reader to have direct access to their results.

R3: We fully agree with the reviewer’s comment and we have presented the results differently in Table 2 and detailed discrepant results in Table 4. A table has also been added concerning samples for which discrimination between G1 and G6 was not possible by the VERSANT Lipa assay (Table 3).

Minor points
C4: English needs minor editing.

R4: The manuscript has been reviewed by a native English reviewer.

Reviewer 2:

Major points

C1: The authors present their method as being highly innovative, while I do not agree with this statement. Re-using the UTR and core amplicons generated by the LiPA procedure is definitely a large step forward in the feasibility of performing sequencing assays for clinical laboratories to determine HCV genotypes and subtypes. However, to target the NS5B region, a separate assay starting from extraction should be followed in order to obtain these sequences. Sequencing a short and conserved stretch of the NS5B gene is common practice and has been published numerous times in the past, not providing any novelty to the current practice.

R1: We agree with the reviewer’s remark that sequencing NS5B, core and 5’UTR regions has been already reported, we have modified the text (page 4, lines 20 to 24). But the innovative approach reported here being the first report of sequencing of LiPA core and 5’UTR available amplicons in case of undetermined hydridization results. As explained in the text (page 12, lines 17 to 19), we are convinced that using already available LiPA amplicons will result in significant cost and workload savings, does not require additional plasma samples.

C2: Although I'm absolutely in favor of HCV genetic sequencing, I'm often confronted with the unwillingness of clinical laboratories to move towards sequencing assays. I believe the authors should elaborate more on the actual use of these assays in clinical practice, such as when is it advised and to provide an estimation of the % of cases in which they would recommend to perform additional testing. Moreover, as the NS5B amplification and sequencing starts from scratch, do the authors think it is feasible to always have enough plasma volume to perform this additional testing, what about the extra time and cost that this will require?

R2: We agree with the reviewer and have therefore added a paragraph concerning the place that sequencing could have in the clinical routine of laboratories, specifying in which cases sequencing could be considered. “HCV genotyping results that would be confirmed represent approximately 5% in our cohort. Taking in account the cost of these confirmation tests, we report here an assay that uses LiPA amplicons already available allowing a significant saving of costs, workload, set up of a RT-PCR steps and would not require extra plasma samples. (…) Homemade or commercial Sanger sequencing or Next Generation sequencing assays should be performed at least in specialized laboratories such as National Reference Centers.” (page 12, lines 16 to 25)
C3: Do they think the NS5B success rate of less than 90% is sufficient to implement in routine practice?

R3: We are aware that this percentage seems low for routine laboratory analysis, but as noted in the revised manuscript (page 11, lines 24 to 27), we believe that the percentage is underestimated in our study given that most of the samples included in this study were complex samples. In addition, the matrices used have undergone some freeze-thaw cycles than routine samples.

C4: While the authors describe in the methods that several samples have been selected for sequencing using different genetic regions, it is not clear why this was done for certain samples and not for all samples. I guess this might be an issue of sufficient sample volume, however this is not properly explained.

R4: We agree with the reviewer’s comment, this point is not clear in the initial manuscript. We have therefore clarified it by reformulating the paragraph under the section Methods – Samples (page 5 lines 3 to 8) “In this retrospective study, we evaluated the determination of hepatitis C genotype of samples from HCV infected patients. In order to make the comparison of the sequencing of the 5’UTR, Core and NS5B regions we select 100 genotype pre-characterized samples from HCV infected patients. Samples were selected to obtain at least 5 samples per genotype or G1 subtype except for G5 and G7, responsible for less than 2% of HCV infection in Belgium [Bouacida et al, PlosONE 2018]. We also added undetermined LiPA samples to challenge the sequencing assays“.

We also added the paragraph under the section Results - Characteristics of analysed samples (page 8, lines 7 to 9): “A total of 100 patients were included in the study among which 41 were sequenced for the 3 regions, 5’UTR, core and NS5B, 23 patients were sequenced for 5’UTR and Core regions and 36 were sequenced for NS5B region only. This repartition was made in function of the available matrix.”

C5: Table 2 provides insight in which samples have been sequenced using more than one genetic region, and how the LiPA and sequencing assays compare. However, in the results the different genetic regions are discussed separately, only comparing LiPA results with the results of each sequencing assay separately (so gene per gene), not providing any clue for the reader whether results obtained between sequencing of 5’UTR, Core and/or NS5B were consistent or not. The current representation of the results is therefore quite confusing for readers, only Table 2 is currently of value. Please consider to re-organize Table 2. I would suggest to add a column for the final assignment of each sample and to organize the samples according to LiPA assignment (starting from 1a) instead of by sample number, as the latter has no meaning to readers. Highlights or colors to visualize whether results were consistent among assays would be very useful. Also explain the abbreviations more in-depth, NR (not realized, why?), NI (not interpretable, this does mean that an actual sequencing attempt was performed and why not interpretable? Low quality or ?) and PA = ?
R5: We fully agree with the reviewer's comment and we have presented the results differently in Table 2 and detailed discrepant results in Table 4. A table has also been added concerning samples for which discrimination between G1 and G6 was not possible by the VERSANT Lipa assay (Table 3).

We also added a paragraph on HCV genotype results of NS5B region as reference in comparison with results obtained from the other assays (page 9, lines 20 to 29).

C6: Why do the authors determine the HCV genotype and subtype after sequencing only with the so-called HCV genomic blast bank? Of note, this is not how you should refer to the use of ncbi blast. Why don't they use a proper way of subtyping, such as constructing a phylogenetic tree? That way they might be able to distinguish between certain HCV4 subtypes for NS5B sequences (and especially when more than one genetic region was sequenced), as in Table 2 often more than one subtype was reported. Moreover, which threshold for similarity was used when performing a blast search? As you are using highly conserved regions across the HCV genome and moreover short sequences, this is of importance.

R6: We agree, the reviewer is right, phylogenetic analysis reinforces the overall message of the article and provides additional genotype verification. We have responded to similar suggestions from reviewer 1 (see response R1).

We did phylogenetic analyses on the sequences for the 3 genes (5’UTR, Core and NS5B) using Neighbor-Joining and UPGMA methods. Forty-seven reference sequences were used to make the trees. A paragraph concerning this analysis has been added in the method section (page 6 lines 25 to 27) as well as in the result section (page 8 lines 21 to 27). We also discussed this phylogenetic analysis in the discussion section (page 9 lines 30 to 33 and page 10 lines 1 to 7). Figures were added to illustrate this point (Figures 1 to 3). However, we have decided not to go further in the analysis of subtypes than the subtypes that must be identified according to EASL recommendations. (page 3, lines 28 and 29) “Therefore, we focused this work only on subtyping G1, no analysis was done for the subtype of the other genotypes.”

C6a: Of note, this is not how you should refer to the use of ncbi blast.

R6a: We modified the referencing page 6 line 23

C6b: Moreover, which threshold for similarity was used when performing a blast search? As you are using highly conserved regions across the HCV genome and moreover short sequences, this is of importance.

R6b: We add this information in the section Methods (page 6 lines 23 to 24) “A threshold for similarity of minimum 85% was used to consider the genotype or subtype”.

Minor points
C7: How was the sample selection done? The authors mention that they wanted to cover the different HCV genotypes and subtypes that are prevalent in more than 2% of cases in Belgium, however they also state that the selection is not representable for the Belgian epidemic. (...)and subsequently explain more in detail how these 100 samples were selected? A substantial high proportion of these samples, respectively 8 and 27, was not properly classified with LiPA, either not discriminating between G1 and 6, or being undetermined. Compared to the percentages that are reported throughout the paper, only for between 2-7% of samples LiPA is not able to determine the HCV1 subtype, so these 100 samples are not just a random set of samples, being not representative for what is generally seen in the clinic.

R7: We agree with the reviewer’s comment, this point is not clear in the initial manuscript. We have therefore clarified this point by reformulating the paragraph under the section Methods – Samples (page 5 lines 3 to 8) “In this retrospective study, we evaluated the determination of hepatitis C genotype of samples from HCV infected patients. In order to make the comparison of the sequencing of the 5’UTR, Core and NS5B regions we select 100 genotype pre-characterized samples from HCV infected patients. Samples were selected to obtain at least 5 samples per genotype or G1 subtype except for G5 and G7, responsible for less than 2% of HCV infection in Belgium [Bouacida et al, PlosONE 2018]. We also added undetermined LiPA samples to challenge the sequencing assays“ and the paragraph under the section Results - Characteristics of analysed samples (page 8 lines 7 to 9): “A total of 100 patients were included in the study among which 41 were sequenced for the 3 regions, 5’UTR, core and NS5B, 23 patients were sequenced for 5’UTR and Core regions and 36 were sequenced for NS5B region only. This repartition was made in function of the available matrix.”

C8: First of all, can the authors provide more information on the Belgian epidemic in the introduction

R8: We agree with the reviewer that epidemiological data are important. We have added this section in the introduction (page 3 lines 6 to 9): “In Belgium, a recent publication showed the same distribution of HCV genotypes with 53.6% of genotype 1 (31.6% subtype 1b and 19.7% subtype 1a) followed by genotype 3 with 22.0% [Bouacida L. et al PlosOne 2018].”

C9: Please choose between a numerical value (23) or writing in full (twenty-three) instead of using both in the Abstract.

R9: this miswriting has been corrected (page 2, lines 16 and 17)

C10: Introduction: please note that already more than 80 subtypes were identified.

R10: We have corrected this information based on the article of Smith et al, Hepatology, 2014 (page 3 line 4)
C11: Introduction: the authors give the impression that HCV genotype determination has become key since the approval of DAAs, however this has been always highly important to initiate treatment, already in the IFN era. I would suggest the authors to add a section in the introduction about the impact of HCV genotype misclassifications on the treatment of patients. There have been published several reports stating virological failure in patients due to misclassification of the HCV genotype and/or subtype, and therefore subsequently the wrong choice of treatment regimen.

R11 : We agree with the reviewer and we have rephrased the sentence (page 3 line 9 and 10) and added a short section about the impact of HCV genotype misclassifications on the treatment of patients (page 3 lines 13 to 18) “ Indeed, several studies have shown that a misclassification of the HCV genotype can lead to therapeutic failure. Starace et al showed that 14.9% of DAA failure are related to a genotyping error [Starace et al, Journal of medical virology 2017] and Di Maio et al. report that 6/197 (3%) of DAA-failing patients and in particular 4/7 non responders to a DAA INF-free regimen were impacted on a wrong genotype assignment [Di Maio et al, Liver international, 2017].”

C12: The authors do not mention anything about mixed infections in the paper. Please elaborate on this aspect of HCV genotyping, as this is another issue associated to the use of LiPA.

R12 : We added this notion in the introduction section (page 4 lines 8 to 10) “Other limitations are that the commercial assays are not always able to differentiate between G6 and G1b [], coinfection [del Campo et al, 2018] or recombination of HCV genotypes []”

C13: The authors use Sanger population sequencing in their assay, however how long do they think it is going to be still used in clinical practice? What about NGS and its compatibility with the proposed method using LiPA amplicons? In general, the discussion is very limited.

R13: We agree with the reviewer's opinion and have included this point in the discussion (pages 9 and 10). We approached the notion of the NGS at the end of the discussion section (page 10 lines 23 to 25)“Homemade or commercial Sanger sequencing or Next Generation sequencing assays should be performed at least in specialized laboratories such as National Reference Centers.”

C14: Ethics: can the authors provide a study or approval number? Have the generated sequences been submitted to Genbank?

R14: We add a Ethic section in the Method section on page 6 line 30 to 32 “The study was approved by the local ethical committee “Comité d’Ethique Hospitalo-Facultaire Saint-Luc – UCL” under number 2018/30AVR/197.” And all sequences have been submitted to Genebank, the assignment of accession number is ongoing
C15: The amplification success rate for NS5B is rather low, the authors claim that it is comparable to what has been reported in one study. However, many other studies do much better, hence why a more in-depth discussion on this aspect is required. Can the authors identify other reasons why the amplification was not successful in more than 10% of samples, next to the somewhat higher genetic diversity compared to 5'UTR and Core? What about viral load? And have the samples been tested in duplicate or triplicate? The authors should be aware that far more diverse regions of the HCV genome have been successfully sequenced.

R15: We have already responded to this comment (see R3)

C16 : Conclusion: referring to a study using the Abbott assay while the entire manuscript compares the determination success of LiPA and sequencing based assays is not appropriate, this should be moved to the discussion section.

R16 : We have moved this reference to page 11 line 11 to 14.

Reviewer 3:

Major points

C1: Please define how you are determining "true" genotype of the LiPA indeterminate samples. If discrepant analysis is not required, please justify.

R1: The phylogenetic analysis we have conducted allowed us to determine the true genotype. we did phylogenetic analysis on the sequences for the 3 genes (5'UTR, Core and NS5B) using Neighbor-Joining and UPGMA methods. Forty-seven reference sequences were used to make the trees. A paragraph concerning this analysis has been added in the method section (page 6 lines 25 to 27) as well as in the result section (page 9 lines 31 to 33 and page 10lines 1 to 7). We also added the phylogenetic analysis in the discussion section (page 2 lines 21 to 23). Figures were added to illustrate this point (Figures 1 to 3).

C2: Methods –There is no reference to optimisation of the sequencing protocols. Please state if there is supporting data for these protocols elsewhere. Analysis of third party reference material, including some reproducibility testing would strengthen the dataset.

R2: During the development of the protocol, the reproducibility was evaluated but the data are not shown. The similar results obtained from different regions for the same samples show the robustness of the protocol (see core and NS5B phylogenetic analysis).

C3: Results : The data is presented as a single table containing all results. This is interesting but could be supplementary data if there is lack of space. The manuscript needs a summary of this so the reader can quickly compare how the sequencing assays performed. Possibly using a table by genotype and LiPA result with of the number of samples successfully genotyped by each of the three sequencing regions. Either within this table or as a separate table there could be a sub-
analysis of the LiPA indeterminate results which were successfully resolved by each region, especially since this is the primary aim.

R3: We fully agree with the reviewer's comment and we have presented the results differently in Table 2 and detailed discrepant results in Table 4. A table has also been added concerning samples for which discrimination between G1 and G6 was not possible by the VERSANT Lipa assay (Table 3).

C4: Explain discrepancies between the methods and highlight which individual or combination of protocols is better at resolving the genotype.

R4: We agree with the reviewer and add a paragraph on our recommendation in the Discussion section (page 12 lines 12 to 25) “Based on the literature and our own data, we would be very cautious in interpreting HCV genotyping result if the Abbott test is used in the first line and we would not hesitate to do the NS5B sequencing assay as a confirmation test. When using the VERSANT LiPA assay, G1 subtyping without core region information should be confirmed by a sequencing method. All undetermined results must also be sequenced. HCV genotyping results that would be confirmed represent approximately 5% in our cohort. Taking in account the cost of these confirmation tests, we report here an assay that uses LiPA amplicons already available allowing a significant saving costs, workload and would not require extra plasma samples. We would therefore recommend the sequencing of the core region when the LiPA amplicons are available and the sequencing of the NS5B region starting from plasma sample. As shown by the analysis of phylogenetic trees, the sequencing of these 2 regions allows a much more discriminating genotyping than the sequencing of the 5' UTR region. Homemade or commercial Sanger sequencing or Next Generation sequencing assays should be performed at least in specialized laboratories such as National Reference Centers”.

C5: Are the viral loads known for the samples? Are you able to state anything about the sensitivity of each assay?

R5: According to the LiPA protocol, the efficacy of the RT-PCR was low for viral load below 2000 IU/mL, consequently, all analysed samples had a viral load above this limit. (Page 5 lines 11 to 13)

C6: Within the text it is not clear if you think these methods are comparable or one is better than another. Or do you recommend both are run simultaneously on indeterminate samples? You could state your proposed algorithm for dealing with indeterminate samples, especially if this has been implemented in the Belgian reference laboratory, and how many results you expect this to resolve. It may be clearer to structure the results to fit with the algorithm you are proposing.

R6: We agree with the reviewer and have added a paragraph on our recommendations in the Discussion section (page 12 lines 20 to 25) “We would therefore recommend the sequencing of the core region when the LiPA amplicons are available and the sequencing of the NS5B region
starting from plasma sample. As shown by the analysis of phylogenetic trees, the sequencing of these 2 regions allows a much more discriminating genotyping than the sequencing of the 5' UTR region. Homemade or commercial Sanger sequencing or Next Generation sequencing assays should be performed at least in specialized laboratories such as National Reference Centers”.

C7: Abstract line 24 - Is misleading stating all the samples were characterised by the LiPA assay but it seems only 65% gave a genotype? Please clarify.

R7: We agree with this comment, we changed “pre-characterised” by analysed. (Abstract page 2 line 10 and Results section page 8 Line 3)

C8: Background line 29 - the statement conflicts with line 4.

R8: We agree with this comment, we changed NS5A into NS5B (page 3 line 3)

C9: Results line 5 - see comment for abstract line 24.

R9: We agree with the reviewer, this point is not clear in the initial manuscript. We have therefore clarified this point by reformulating the paragraph under the section Methods – Samples (page 5 lines 3 to 8) “In this retrospective study, we evaluated the determination of hepatitis C genotype of samples from HCV infected patients. In order to make the comparison of the sequencing of the 5’UTR, Core and NS5B regions we select 100 genotype pre-characterized samples from HCV infected patients. Samples were selected to obtain at least 5 samples per genotype or G1 subtype except for G5 and G7, responsible for less than 2% of HCV infection in Belgium [Bouacida et al, PlosONE 2018]. We also added undetermined LiPA samples to challenge the sequencing assays“ and the in the section Results - Characteristics of analysed samples (page 8, lines 7 to 9): “A total of 100 patients were included in the study among which 41 were sequenced for the 3 regions, 5’UTR, core and NS5B, 23 patients were sequenced for 5’UTR and Core regions and 36 were sequenced for NS5B region only. This repartition was made in function of the available matrix.”

C10: Results line 15 - Only 64 samples were sequenced by UTR and core. Why was this so low? Was viral load a factor or is it likely interference of the two amplicons in the sequencing reaction? Potential reasons should be raised in the discussion. Results line 16 - why were 77 samples analysed for NS5B? Did 23 not amplify or were these not attempted.

R10: We agree with the reviewer’s observations, this part is not clear in the first version of the manuscript. The explanation has been added in the paragraph under the section Results - Characteristics of analysed samples (page 8, lines 7 to 9): “A total of 100 patients were included in the study among which 41 were sequenced for the 3 regions, 5’UTR, core and NS5B, 23 patients were sequenced for 5’UTR and Core regions and 36 were sequenced for NS5B region only. This repartition was made in function of the available matrix.”
C11: Table 1 - Please cite the source of these oligonucleotides.

R11: We have added the reference of oligonucleotides in the Table 1 and in the text (page 5 line 32 and page 6 line 12)

C12: Table 2 - states that sequencing enabled discrimination of all 27/100 samples not identified by LiPA yet table 2 contains 2 samples which have a LiPa result only. Therefore the comparison is of 98 samples? Also the footnotes "not realized" and "undetermined" require clarification.

R12: Samples included in the study are represented in tables 2 to 5 and the fact that these samples could not be sequenced seemed important to us as information. Table’s legends have been improved.