Author's response to reviews

**Title:** Molecular epidemiology and phylogenetic analysis of Hepatitis B virus in a group of migrants in Italy.

**Authors:**

Umbertina Villano (umbertina.villano@iss.it)
Alessandra Lo Presti (alessandra.lopresti@guest.iss.it)
Michele Equestre (michele.equestre@iss.it)
Eleonora Cella (ele.28@hotmail.it)
Giulio Pisani (giulio.pisani@iss.it)
Marta Giovanetti (smarties.88@hotmail.it)
Roberto Bruni (roberto.bruni@iss.it)
Elena Tritarelli (elena.tritarelli@iss.it)
Massimo Amicosante (amicosan@uniroma2.it)
Alba Grifoni (alba.grifoni@uniroma2.it)
Carmelo Scarcella (carmelo.scarcella@aslbrescia.it)
Issa El-Hamad (issa1957@libero.it)
Maria Chiara Pezzoli (chiarapezzoli@yahoo.com)
Silvia Angeletti (s.angeletti@unicampus.it)
Anna Rita Ciccaglione (annarita.ciccaglione@iss.it)
Massimo Ciccozzi (massimo.ciccozzi@iss.it)

**Version:** 3  
**Date:** 25 May 2015

**Author's response to reviews:** see over
Reviewer’s report
Title: Molecular epidemiology and phylogenetic analysis of Hepatitis B virus in a group of migrants in Italy.

Version: 2 Date: 5 May 2015

Reviewer: HSIN-FU LIU

Reviewer’s report:

In this manuscript, the authors reported molecular epidemiology and phylogenetic analysis of Hepatitis B virus in a group of migrants in Italy. This is an interesting issue for epidemiology of infectious diseases and also is important for monitoring viral spread and evolution. The manuscript is well written and suitable for publication. There are only few miner comments.

Minor Essential Revisions

1. The authors indicated that the sequences will be submitted to GenBank after the acceptance of the article. However, in most of cases, sequences shall submit to GenBank database to obtain accession numbers before publication, and they shall provide in the article. Authors are strongly suggested to do this before publication rather than after. Accession numbers can keep confidential by GenBank if the authors prefer not to release them before publication.

Reply: The sequences have been submitted to Gen Bank but the Accession Numbers will be written in the manuscript, when the proof will be available.

2. Sequence quality control is essential for phylodynamic analysis. The authors has performed likelihood mapping for phylogenetic signal, but how about recombination? Authors shall explain how they exclude the possibility of sequence recombination by means of any specific method (e.g. RDP).

Reply: We have done this in a preliminary analysis, anyway there is not uncertainty about the genotype of the sequences. There were only three sequences, as described in the text, that appear outside the clade of a specific genotype; it is a common procedure to report them as “unclassified” (however the aim of this study is to focus and to study in detail the HBV D1 genotype Moldovan sequences). Specifically, these three unclassified sequences, were analyzed for recombination with specific softwares (like SplitsTree and Simplot) but the statistical test, the phi test was not statistically significant and were indicated as “unclassified”.

3. In figure 2, authors shall describe the meaning of the vertical line appeared between 1997 and 2000 in the BSP.

Reply: We analyzed the evolutionary dynamics with the Bayesian Skyline Plot (BSP) model and the vertical line appeared between 1997 and 2000 in the BSP corresponds with the middle of the exponential phase. Generally it is not necessary to describe this.
Level of interest: An article whose findings are important to those with closely related research interests

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.
Reviewer’s report

Referee’s 2

Title: Molecular epidemiology and phylogenetic analysis of Hepatitis B virus in a group of migrants in Italy.

Version: 2 Date: 5 May 2015

Reviewer: Yi-Ming Arthur Chen

Reviewer’s report:

The paper by Villano et al. entitled “Molecular epidemiology and phylogenetic analysis of Hepatitis B virus in a group of migrants in Italy” addresses the HBV genotype and phylodynamics among migrants who were HBV chronic infection in Italy. There are major issues in the results and methodology which I list as followings. Therefore, I recommended this article to be accepted pending major compulsory revision.

Major points: 1. The authors used only maximum likelihood method to construct phylogenetic tree (Supplementary Figure 2 and 3). The authors should use other methods (e.g. Bayesian inference) to confirm the tree topology.

Reply: Regarding the phylogenetic analysis performed for Supplementary Figure 2 and 3, in phylogenesis the maximum likelihood method is normally used and is an adequate and sufficient method to assign the genotype and to establish the phylogenetic relationships. Bayesian analysis is usually performed when has been applied coalescent theory and molecular clock to identify the way of transmission or perform phylogeographic analysis but this do not was our aim. Anyway we usually confirm with Bayesian analysis by Mr. Bayes software and we had the same topology.

2. In line 269, the authors described that there are 3 sequences were “unclassified”, how are unclassified sequences defined? Please describe in detail.

Reply: In phylogenetic analysis it is a normal procedure to report that some sequences are unclassified, especially if BioAfrica (Oxford HBV Subtyping Tool Web Interface) report, classify them as NA and also NCBI Genotyping tool, classify them as not pure genotypes. Anyway, a supplementary analysis can be do using a split decomposition tool and NeighborNet, to check the recombination. Looking for the tree is also very easy to the expert eyes to recognize eventual recombinant if the position of the sequences are out of the cluster specific for a genotype.

The three unclassified sequences, “putative recombinant”, were analyzed for recombination with specific softwares (like SplitsTree and Simplot) but the statistical test, the phi test was not statistically significant and were therefore indicated as “unclassified”. In this study we focalized the analysis only on HBV subgenotype D1 (the more prevalent ones) to better investigate the temporal dynamics and phylogenetic relationships.
3. The authors described that the Moldavian sequences (17 sequences) were intermixed and most correlated with the sequences from Russian, East European and Asian countries. Why the authors did not add the reference sequences from Moldova to confirm the transmission relationship? If there the sequences from Moldova are available, I suggest that the author should add these sequences to the phylogenetic tree analysis.

Reply: We disagree with the Reviewer because the aim of the study was not the relationship between Moldavian sequences but to understand the phylogenetic relationships between our Moldovan isolates (migrant in Italy) and other reference sequences sampled from different countries (foreign sequences). This analysis was focused in this case on the presence of the infection before or after the venue in Italy from Moldavian migrants: in Effect only coalescent analysis and molecular clock principles if applied can give information about the “probable” relationship of transmission that do not was our aim.

Minor points:

1. The author should provide the GenBank accession numbers.

The sequences have been submitted to Gen Bank but the Accession Numbers will be written in the manuscript, when the proof will be available.

2. The author should check the spelling carefully such as in line 271 (Among) and line 345 (East- Europe).

Reply: The sentences have been improved, and the spelling have been corrected.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being Published

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.

Reply to Reviewer’s Comments

Referee’s 3
Reviewer's report

Reviewer's report:

Major Compulsory Revisions

There are two major problems about the phylogenetic analysis which might effect the whole conclusion in this research:

1. In Figure 1, does the scale at the bottom of the tree represent the time in years with 0 indicating present year? If yes, these sequences seems sampling from the same year for their branch end at the same time point (year 0). But, the name of viral sequences represent the viruses sampling from different year. This problem might cause by the year setting process in the BEAST analysis. If the year setting of each sequences had some problem, this would effect the interpretation.

Reply: The scale at the bottom of the tree represent time in years with 0 representing the present year, considering the most recent sampling date. When estimating the evolutionary rate, the year of sampling of the sequences is considered in Beast (with the Guess Dates option); for example in our analysis we previously estimated the evolutionary rate on the second dataset and then, this evolutionary rate was imposed and fixed in the run of the 17 HBV Moldovan sequences. The Bayesian phylogenetic tree obtained with this method appears like the phylogenetic tree reported in Figure 1. This method is used when estimating the evolutionary rate with dataset that are not big enough, and when it is necessary to reinforce the dataset with an external dataset to estimate the evolutionary rate.

2. About the maximum likelihood phylogenetic analysis in the supplementary figure 2, why the genotype D1 didn’t shown as the monophyletic clade? How do the author recognized these were all belong to D1 if they were not clustering together? This tree might had problem for making any conclusion.

Reply: Regarding the maximum likelihood analysis this sometimes happens when sequences have different lengths. In the phylogenetic analysis, this happens also for the reference sequences that didn’t appear as monophyletic clade probably for the different similarity between sequences. However we have improved the Supplementary Figure 2 eliminating some symbols.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.