Reviewer's report

Title: Diversity of HIV-1 genotypes and high prevalence of pretreatment drug resistance in newly diagnosed HIV-infected patients in Shanghai, China

Version: 0 Date: 14 Dec 2018

Reviewer: Joakim Esbjörnsson

Reviewer's report:

Summary

In this study, Wang and colleagues aimed to determine the subtype/CRF distribution and TDR frequency based on HIV-1 pol sequences collected from 317 ART-naïve and newly HIV-1 diagnosed participants at the Shanghai Public Health Clinical Center during 2017. The main conclusions were that the subtype/CRF distribution was extremely diverse (dominated by CRF01_AE [50.8%]) and that the TDR frequency was very high (in total 17.4%).

General comments

Relatively few detailed reports has been presented on the HIV epidemic in China and the reported TDR prevalence of 17.4% is both very high and worrying. Even if the report are based on a very small fraction of the newly HIV-1 diagnosed individuals in China (less than 0.5%; no data on the number of newly HIV-1 diagnosed individuals in Shanghai is given), this result is relevant and calls for further investigation. However, and unfortunately, the study is generally poorly executed and some fundamental information is missing in the manuscript. The subtype/CRF analysis that constitutes about half of the results section is not executed with appropriate methodology. This is a major flaw of the report.

Major points

1. The conclusion that the HIV-1 distribution in Shanghai is extremely diverse (consisting of five main subtypes/CRFs and seven minor variants) is not true. This level of diversity is very common. In fact, many countries and regions around the world display a much more diverse HIV-1 distribution than the one reported here.

2. Both the introduction and discussion sections are missing references in several places to support the written statements. For example, the statement that HIV-1 genotypes vary remarkable in transmission route is questionable and need to be supported by proper referencing.

3. It is stated that the "with the development of social economy and the rapid growth of migrants, the epidemiological pattern of HIV/AIDS may have changed". This is a very vague statement.
Changed from what? Also, it would be helpful to exemplify what the authors mean by "epidemiological pattern".

4. P. 6, lines 120-121: What does newly diagnosed within 3 months mean? Please clarify.

5. The state-of-the-art is to determine the HIV subtype/CRF using phylogenetics with reference to the Los Alamos Sequence Database reference dataset (preferably maximum-likelihood or Bayesian with a robust branch support method). A BLAST approach (which is poorly explained in the report) is not sufficient for proper subtype/CRF assignment. This part of the study needs to be entirely redone with proper methodology and reporting.

6. Please add definitions for the degree of resistance on p. 7, lines 140-142.

7. P. 7, lines 145-146: How was normality assessed?

8. One major limitation with the generalizability of the results is that 95% of the participants were male and that 69% were MSM. This needs to be discussed much more in detail. Also, the authors may want to consider to highlight this already in the abstract and title.

9. Another point is the CD4 status of the newly diagnosed participants. A median CD4 count of 275 indicates that the majority of participants are very immunosuppressed and they have presumably been infected for a relatively long time. How does this affect the interpretation of the data and reflection of the current HIV epidemic in Shanghai?

10. P. 8, lines 175-176: The reporting of the numbers are wrong. It should be (8.5%, 4/47) and (0.5%, 1/218).

11. A number of different comparisons have been made. Was correction for multiple testing done? If not, motivate why this was not necessary.

12. P. 9, lines 193-196 (also in abstract): This paragraph describes the results from a Chi2 test comparing the distribution between three groups. The author interpret this as subtype B having higher proportions of TDR mutations than CRF01_AE and CRF07_BC. This is wrong! The test only asses the overall difference in distributions between the three groups and does not say anything about which specific groups that differ. This must be tested by pairwise testing (and correction for multiple testing, e.g. Bonferroni). Inspection of the effect estimates rather suggest that it is the CRF07_BC that is lower than the other two groups. However, this must be verified by pairwise comparisons.

13. P. 10, lines 202-216: This paragraph is very descriptive. It would be informative to see some statistics and comparisons in this section.

14. P. 11, lines 221-222: Mutation rates have not been assessed in this study. Please correct.

15. P. 11, lines 222-223: "...half of the mutated virus strains...". Mutated from what? Please use appropriate terminology.
16. P. 12, lines 250-252: The authors state that their findings indicate an increase among MSM in Shanghai. How can they know that? The study that they refer to was performed in Fujian which is situated 800 km from Shanghai. The authors need to rephrase and motivate how this comparison is relevant.

17. The sequences does not appear to have been submitted to Genbank which is highly recommended and common standard in the field.

18. Table 3: Here the authors report that they have used Fisher’s exact test. This was not mentioned in the Methods section. Please clarify.

Minor points

1. The manuscript contains several typos and would benefit from some language editing in some sections. It also needs to be reviewed for consistency in terminology and use of symbols and the same font throughout.

2. The section in the introduction that describes the HIV types and groups can be written more concise. For example, the description of HIV-2 and non-M groups on p. 5 are not very relevant for the current study.

3. The term "gender" refers to e.g. each individual's personal identification of one's own gender which is not necessarily the same as the biological sex of that individual. This may seem like semantics, but in scientific writing most journals recommend to use sex instead of gender (see e.g. Lancet's instructions for authors).

4. P. 10, lines 199-200: Revise p-values to p<0.0001.

5. P. 11, 234-235: This part is redundant from the beginning of the discussions section.

6. Table 1: Why is some letters small and some capitalized?

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

No

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

No

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