16-Dec-2014

Dear Dr. Fedorov,

In view of the criticisms of the referees found at the bottom of this letter, I fear that I cannot offer to publish your manuscript # GBE-140909.R1 entitled "Inference Of Distant Genetic Relations In Humans Using “1000 Genomes”" which you submitted to Genome Biology and Evolution.

I wish I had better news. I would encourage you to consider the comments below and if you think that a reanalysis cal solve the problems, I would be glad to see a new version of this paper in the future.

In the meantime, I wish to thank you for considering Genome Biology and Evolution for the publication of your research and I hope the outcome of this specific submission will not discourage you from the submission of future manuscripts.

Yours sincerely,

Bill Martin

Editor in Chief, Genome Biology and Evolution

w.martin@uni-duesseldorf.de

Associate Editor Recommendation: decline

Associate Editor: Majumder, Partha

Comments to the Author:

1 Selection of the cut-off of 0.2% for the rare variants is still not adequately explained.

2. There is no clear explanation on whether the authors have analyzed all possible pairs of individuals from the 1000 genomes phase I dataset (as shown in the Figure 1 legend and other places) or whether they restricted their analysis to “96,464 human pairs within and between populations” as stated in the main text (Introduction, Page 4, Line 39). If it was the latter then there is no explanation of the number of individuals that were examined from each population and how they were chosen.

3. The authors do not attempt to validate any of their heterozygous calls, despite admitting that these may be especially prone to sequencing and base call errors, although the genotyping data is publically available from the 1000 Genomes Project ftp website. They could also have validated a proportion of heterozygous sites that were covered by the high coverage exome sequences or those in the high coverage CEU and YRI trios. The authors have stated that the 1000 Genomes Project Consortium was non-cooperative in their attempts to validate their call set of rare heterozygous variants. The genotype and high coverage exome sequences are all publically available from the 1000 Genomes Project website.

4. The authors also explored the proportion of their rare variants that are in linkage disequilibrium and show that when a pair of individuals has a significant excess of shared rare genetic variants, a considerable portion of these are clustered within a single or a few chromosomal regions. This clustering also raises the possibility of structural variation in these regions, but the authors do not address this. The example of regional clustering that they give in Table 2 (chromosome 11 region 90,787,654-90,858,949) has a number of annotated structural variants.

5. It’s not mentioned whether the shared variants between the different continental populations are on the same or different haplotype backgrounds. This would be relevant for the British-Chinese pairs.

6. There are too many typographical, grammatical and labelling errors in the manuscript even after revision. Examples of labelling errors: (a) The three letter population codes are still inconsistent in the main text (Page 20 Lines 45” TSC” and “CEP”; (b) Population labels are missing in the second figure in Supplementary data S2.

Reviewers' Comments:

Referee: 2

Comments to the Author

The revised paper by Al-Khudhair et al. entitled, “Inference of distant genetic relations in humans using “1000 Genomes”, address some the points that were raised in my earlier review but some issues remain.

1. The authors do not attempt to validate any of their heterozygous calls, despite admitting that these may be especially prone to sequencing and base call errors, although the genotyping data is publically available from the 1000 Genomes Project ftp website. They could also have validated a proportion of heterozygous sites that were covered by the high coverage exome sequences or those in the high coverage CEU and YRI trios.

2. The authors also explored the proportion of their rare variants that are in linkage disequilibrium and show that when a pair of individuals has a significant excess of shared rare genetic variants, a considerable portion of these are clustered within a single or a few chromosomal regions. This clustering also raises the possibility of structural variation in these regions, but the authors do not address this. The example of regional clustering that they give in Table 2 (chromosome 11 region 90,787,654-90,858,949) has a number of annotated structural variants.

3. It’s not mentioned whether the shared variants between the different continental populations are on the same or different haplotype backgrounds. This would be relevant for the British-Chinese pairs.

Minor Points:

1. The three letter population codes are still inconsistent in the main text (Page 20 Lines 45” TSC” and “CEP”.

2. Population labels are missing in the second figure in Supplementary data S2.

Referee: 1

Comments to the Author

The authors implemented my suggestions, the paper is acceptable.