

Review of the reviewed Master thesis of Klára Kopicová "Phylogeny of human populations in Papua New Guinea, a genetic and linguistic diversity hotspot"

One of the common jobs of scientists is to review the work of others before it is accepted for publishing in scientific journals. My favorite part of it is to read the second/higher-order revisions (provided they are necessary) to see how the authors implemented my suggestions and improved their manuscript. Well, I never thought I would have a similar opportunity with diploma theses, but I guess never-say-never has proven to be true again.

After the unfortunate events of January, where the thesis of Miss Kopicová had to be rejected, she was allowed to submit another version and had several months to make it better than its predecessor. So, did she succeed? Short answer: In my eyes Yes, at least to some extent.

The Introductory review has improved. Some less relevant chapters are gone. Despite the length, it is easy to read and nicely introduces the fascinating cultural diversity of the Papua New Guinea (PNG) from several angles. In the following chapters, some methodological principles used in the study of human evolution are presented. I remember commenting on the good language skills in the first version of the thesis, and I think the English even improved. My needs for typo/grammar errors hunting are fully saturated by my kids' homework, but during the careful reading of the thesis, I discovered only several issues, definitely not enough to point it out as the shortcoming. I just hope it is all original.

Next come the Aims, and these are improved now too. Both aims and hypotheses are laid-out in concise points, without vague sentences and word ballast. I have little to criticize also in Methods. The methodology used here is obviously complex, but in combination with Chapter 5, I guess the educated reader can understand most of it on the first try. Maybe only in the first part of Table I (page 32), the last column is headed as an "ISO639-3 ", while in second and third parts as "Language names ". Moreover, I think it contains language name abbreviations rather than full names. Also, I think the column is not that relevant for the thesis and could have been omitted anyway.

Then come the Results combined with Discussion. I remember me commenting on this crucial part to be too short without an obvious discussion. Well, that improved a bit too. The results are discussed in more detail with recent works (mostly Bergström et al., 2017 and Lipson et al., 2017). I am not an expert on human archeogenomics, but I suspect there are maybe other relevant sources to discuss? Also, the linguistics is mentioned in the title, and occupies two chapters of the Introduction, I am disappointed to see it is still barely mentioned in Results/Discussion. In the previous version, I was complaining about the way the results are presented. Although some of the issues are now resolved, I think the author could pay more attention to explaining the results. For example, on page 39. the legend of Fig. 3 (the PCA) mentions "Remote and Near Oceanians including Polynesians (POL), Papuans and Australians ". Neither Papuans nor Australians are on the Actual PCA plot as a discrete

category. Are they included within Near Oceanians (OCE)? If so, then why and it has to be mentioned somewhere, not just expect the reader will figure it out somehow!

Similarly, on pages 43 and 45, what exactly do Tables II and IV show/say? The legend is unacceptably shy on details, which are not revealed in the text either. Lastly, I remember me complaining about the missing explanation of numbers in qpGraphs. Well, it is still missing, so once again, what do these numbers mean?

Conclusion nicely sums everything up in a concise but fitting manner. To illustrate my previous points about more attention that should be paid to the way the results are presented-explained, this part also contains the best topology obtained by the author. However, instead of a nice little schematic cladogram, which would actually perfectly conclude the results for those not accustomed to reading the qpGraphs, she used text/Newick format, which needs a lot of readers' imagination to get the same information.

In my previous evaluation, I asked three general questions, which were (to some extent) answered during the presentation. Today, I will add one more, answering which will require a bit of speculation from the author.

Q How would the inclusion of samples from the Indonesian part of NG island affect the results of your analyses? The text does not mention the western part much, is there anything known about the genetic composition of their inhabitants? If not, try to speculate about it.

At the beginning of my review, I have used the analogy of the second round of review of the scientific paper. If I should keep to it also at the end of my review, my verdict would likely be another round of revision. While I believe this version of the thesis is better than the previous, there are still some aspects lacking. I will suggest the final grade after the defense, but I think the thesis fulfills the requirements, and Klára deserves the degree.

In Boršov nad Vltavou, July 11th 2020

Aleš Horák