

Klára Kopicová 2020

**Phylogeny of human populations in Papua New Guinea,
a genetic and linguistic diversity hotspot**

Master thesis

Opponent's review

Klára Kopicová's master thesis includes 46 pages of text (excluding references), five original figures, and four tables. The reference list contains 136 sources. The aim of the thesis is to investigate history of human populations in Papua-New Guinea (PNG hereinafter) using previously published genomic data and new methods (mainly the qpGraph).

It is not possible to evaluate this thesis out of context. It is a rewrite of a thesis that has been defended unsuccessfully half a year ago. The original thesis was oddly structured, riddled with mistakes, and lacking in presentation and discussion of the results. On top of that, it included huge chunks of text that were copy-pasted from other sources. Rewritten and resubmitted papers are judged differently than first submissions. Since this thesis is just that, I will partly base my review upon a comparison with the original version.

The author made an effort to either exclude or reformulate sections of the text that were copy-pasted from other sources. I am happy to report that this problem has been largely eradicated. A few problematic sections remained: The section on PNG languages begins with the sentence that is copied from Palmer (2018), the introductory section on ABO blood groups contain chunks of text copied from Bodmer (2015) and descriptions of f3- and f4-statistics are copied from online repositories (GAWorkshop and GitHub) with only minor changes. I have compared the new version of the thesis with the original and google-searched many random sections of the text. These were all the only problematic sections I have found.

There are several formal improvements in this new version of the thesis. It doesn't feel like such a rush job. There are not as many formatting and spelling mistakes, the most glaring issues (unfinished contents section, edits in track changes mode...) have been fixed. The introduction section has been appropriately restructured. Geography, history, and anthropology of PNG is now introduced prior to the introduction of population-genetic methods, which feels more natural. I like that after each sections on different population-genetic markers, relevant studies of the Papuans are mentioned (sadly, this applies only to blood groups and uniparental markers). I also appreciated the inclusion of additional tables, particularly Table 1, which lists the studied populations and their linguistic affiliation.

(There are some claims in the thesis that are factually wrong. Papuans do not share hunting-gathering lifestyle with Negritos of Philippines (p. 5, paragraph 3), their lifestyles are very

different. The grouping of Highlanders into 3 clusters described on p. 10, paragraph 3, is not linguistic, those are population-genetic clusters. One of them corresponds to linguistic grouping – the Angan family, which is considered by some to be a part of Trans–New Guinean.)

Many of the issues of the original version are still very much present. The introduction is very long, technical, sort of textbookish (a textbook of human evolutionary genetics by Jobling et al. is cited 17 times). Some sections of it bear little relevance to results and discussion, which are very brief.

Since the thesis relies entirely on previously analyzed data – the dataset of Bergström et al., merged with 5 other previously published SNP genotyping datasets – it doesn't offer new insights. The author attempts to formulate explicit hypotheses (p. 30), which only highlights the weaknesses of her work. The stated hypotheses (widespread Southeast Asian ancestry in Papuans, sharp highland-lowland division, and west–east division of highlanders) are not interesting hypotheses, because there are no reasonable alternatives. No one doubts the SEA ancestry of Papuans. The intra-PNG population structure is – again – what Bergström et al. discovered recently and since the author is using the same data, she can only come to the same conclusion, provided the results reported in Bergström et al. are correct. To “re-check” the results described in the paper that was recently published in Science is an odd goal for a master thesis. Sure you must do that, but it shouldn't be a goal in and of itself.

I can merely repeat what I stated in my original review: The results and discussion are disappointing. The discussion is virtually absent, it is a highly technical description of the results, which only confirm (repeat would be the more appropriate word) those of previously published studies. Some findings seem like new but they really are not (e.g., Kankanaey as the best genetic proxy for early Austronesians; Mörseburg et al. 2016). The qpGraphs are described (several times in parenthetical format within the text, which is just annoying) but the implications of their topologies are not discussed. The conclusions section is poorly written.

My conclusion is that while several formal improvements have been made, the new version of the thesis is as unengaging as was the original. It is passable, in my opinion, nothing more or less.

České Budějovice, 8 July 2020

RNDr. Pavel Duda, Ph.D.