

Reviewer Recommendation and Comments for Manuscript Number PONE-D-14-14972

Detecting Genome-wide Signals of Human Adaptation to Tropical Forests in a Convergent Evolution Framework

Original Submission
Simon Aeschbacher (Reviewer 3)

[Back](#)
[Edit Review](#)
[Print](#)
[Submit Review to Editorial Office](#)

Recommendation Minor Revision

Custom Review Question(s)

Comments to the Author

Response

Partly

1. Is the manuscript technically sound, and do the data support the conclusions?

The manuscript must describe a technically sound piece of scientific research with data that supports the conclusions. Experiments must have been conducted rigorously, with appropriate controls, replication, and sample sizes. The conclusions must be drawn appropriately based on the data presented.

2. Has the statistical analysis been performed appropriately and rigorously?

Yes

3. Does the manuscript adhere to the PLOS Data Policy?

Yes

Authors must follow the [PLOS Data policy](#), which requires authors to make all data underlying the findings described in their manuscript fully available without restriction. Please refer to the author's Data Availability Statement in the manuscript. All data and related metadata must be deposited in an appropriate public repository, unless already provided as part of the submitted article or supporting information. If there are restrictions on the ability of authors to publicly share data—e.g. privacy or use of data from a third party—these reasons must be specified.

4. Is the manuscript presented in an intelligible fashion and written in standard English?

Yes

PLOS ONE does not copyedit accepted manuscripts, so the language in submitted articles must be clear, correct, and unambiguous. Any typographical or grammatical errors should be corrected at revision, so please note any specific errors here.

5. Review Comments to the Author

Please use the space provided to explain your answers to the questions above. You may also include additional comments for the author, including concerns about dual publication, research ethics, or publication ethics. (Please upload your review as an attachment if it exceeds 20,000 characters)

5.1 Major Concerns, and Comments Relating to Answers 1 to 4

@1: In general, it seems that experiments have been conducted rigorously and the necessary controls have been conducted. Methods are very well written and choices and assumptions made are stated. However, I missed a discussion of the sensitivity to several choices that had to be made. It would be desirable to know how sensitive the final results are to these choices. Examples are the false discovery rates, the nature of permutation tests to judge significance in enrichment analyses, or the width and overlap of windows in the sliding-window approach for defining significant clusters of candidate SNPs. See specific comments below.

@4: The paper is generally very well written, and procedures, results and potential issues are clearly described. However, there was a tendency for long sentences. See specific minor comments below.

a) It seems that the four scenarios of selection described on p.6, §2, do not include cases where $\alpha < 0$ in at least one comparison of forest vs. control population. On p.7, §2, it is stated that "...convergent selection occurred when selection presents a higher probability than neutrality and model 4 has a higher posterior probability than models 2 and 3". This should be clarified. I was worried that a case of $\alpha_{\text{America}} < 0$ and $\alpha_{\text{Africa}} > 0$ could be classified as "convergent", although it is not. In the Discussion (§1), you mention that you looked for SNPs with overly large or low differentiation between populations. In the rest of the paper (in particular in Results), I got the impression that overly low differentiation is not referred to at all.

b) I would have liked to see a distinction between

synonymous and non-synonymous SNPs. The authors suggest that several of these candidate SNPs could be directly or indirectly affected by selection. Their analysis also suggests that multiple sets of gene ontologies are enriched for candidates. It would be of high interest to know whether these candidates are potential causal variants, linked to causal variants, or structural/regulatory variants. Even though a conclusive answer to these questions seems beyond the scope of this paper, a distinction between above-mentioned classes of SNPs could give initial hints. See respective comment below.

c) I was uneasy about whether the permutation tests performed to assess significant enrichment with candidate SNPs or genes did control well enough for potential biases. First, in the test for enrichment of genic versus non-genic SNPs, clusters of closely linked candidate SNPs could introduce a bias and it might be better to pick one representative SNP per cluster. Second, in the test for enrichment of certain gene ontology classes, I missed control for, or at least discussion of, the impact of gene length. Even though only one candidate SNP was chosen per gene, isn't it that long genes per se have a higher chance of containing an outlier SNP even if being an outlier SNP was uncorrelated to gene ontology class? See specific comments below.

5.2 Minor, and More Specific Comments

p.: page, followed by number

§: paragraph, followed by number

l.: line, followed by (range of) number(s); negative values are counted from the end of the paragraph

If there is no page and/or paragraph number, these are identical as in the most recent comment where these numbers were given.

p.1 | Title: I did not particularly like the phrase "convergent evolution framework". I agree that you are answering intriguing questions about convergent evolution in human adaptation to tropical forests. However, I do not think that the approach used here needs to be termed "convergent evolution framework". The idea of comparing levels of population differentiation (as measured in terms of F_{ST} or related quantities) at various levels of hierarchy is not restricted to applications in the context of convergence, but could equally well be used in the context of divergence. How about "Detecting Signals of (Convergent) Human Adaptation to Tropical Forests Using a Stratified [or: Hierarchical] Outlier Approach"?

p.2 | Abstract: It would be good to state earlier that you are studying both African and American populations. This is mentioned in the last two sentences of the Abstract, but could be made more clear.

p.2 | Abstract | l.6: "with overly" -> "with an overly"

p.2 | Abstract | l.13-14: I was hoping that "pygmy phenotype" would be defined in more detail (e.g. w.r.t. body size, habitat, continents) in the Introduction, but then did not find this. I think there could be potential confusion as to whether the term "pygmy" refers to African tropical forest populations exclusively, or includes those of the Americas. Even if the literature is ambiguous, it should be clear what you are referring to in this study. I realise that Perry and Dominy (2009) explicitly include South American populations. For a general readership, clarification / an explicit definition might be useful.

p.3 | §2 | l.3: "higher pathogen diversity" -> "higher human pathogen diversity"

p.5 | §3 | l.2-6: It was not clear to me why the Y chromosome was excluded, but the X chromosome included. It may be good to add a justification.

p.5 | §3 | l.6: "After selecting the different PSs as described above,..." -> "Within each PS as defined above,..."

l.-3: "at all" -> "in all"

l.-2: Replace short dash by en-dash

p.5 | §4 | l.2: "implemented with this software" -> "in BayeScan" [avoid repeating "implemented"]

p.6 | §1 | l.1: "algorithm [20] and uses" -> "algorithm [20]. The approach assumes"

l.3: "population specific" -> "population-specific"

§2 | l.2: "relative closer" -> "relatively closer"

l.3: "the case of populations" -> "the case for populations"

l.-4: "In this regard,..." -> "Thus,..."

l.-3: "two different continents..." -> "two different continents,..."

p.7 | §2 | l.1-2: How sensitive are the results to the choice of a FDR = 0.1?

§3 | l.1-3: I was missing a distinction between synonymous and non-synonymous sites, especially in the analysis and interpretation of candidate SNPs (see major comment above).

l.5: "<50kb": insert spaces before and after "50"

l.7: "ahead" -> "above"

l.7-10: I presume it was assumed that SNPs are unlinked in the permutation test. However, this might introduce a bias when candidate SNPs occurred in clusters, which might be a consequence of linkage disequilibrium. Then, would it not be better to summarise a cluster of candidate SNPs just by one representative SNP? The SNPs of that cluster might not be independent.

p. 8 | §1 | l.-2: How far apart were the various positions? Are these most likely gene duplications?

§3 | l.3: "500 kb-wide" -> "500 kb wide"

l.4: The 95% quantile of what? Of the distribution of individual q-values?

l.5: "the average SNP density per chromosome" -> "the average SNP density per 500 kb for the given chromosome"

l.1-6: How sensitive is the analysis to these choices?

p. 9 | §1 | l.1: "distribution of the SNPs" -> "physical position of the SNPs" (It seemed to me that "distribution" was often used to mean "physical position", which is potentially confounding).

§2 | l.1-4: It was not clear to me why, a priori, a significant SNP in a region of low SNP density should be of less interest. Please clarify.

§3 | l.5-6: The term and definition of "any selection" is potentially confounding, as it is not clear whether cases of $\alpha < 0$ are ignored or included. See previous related comments.

l.-4: Insert space between "50" and "kb".

p.10 | §2 | l.1-2: I missed a justification / rationale for the SUMSTAT enrichment score. It seems to me that SUMSTAT overrates gene sets for which there is evidence of enrichment in at least two selection models, as then the value of p_{co} and/or p_{sl} will naturally also be high.

l.1: "For all selection models, we..." -> "We..."

p.11 | §2 | l.2: It would be nice to have the percentage of identified SNPs relative to the original number of SNPs given in brackets, for each PS.

§3 | I.1-3: It was unclear whether $\alpha < 0$ never occurred only in the sliding-window procedure, or also for the SNP-based analysis, i.e. before identification of 'significant clusters'. If there were no negative alphas at all, this should be stated before the definition of 'convergent evolution' on p.7. This would anticipate a result there, but it is necessary to understand your choices.

p.12 | §2 | I.1-2: Again, it would be great to know how many and which of these candidate SNPs were non-synonymous. I am aware of the fact that synonymous SNPs could be tightly linked to non-synonymous ones, and so a distinction between synonymous and non-synonymous would only be partial confirmation.

§3 | I.1: "distribution" -> "physical position"
 I.1-2: "...in the genome for the different PSs and their estimated q-value, as well as..." -> "...in the genome and their estimated q-values for the different PS, as well as..."

I.3: "into the sliding..." -> "in the sliding..."

p.14 | §1 | I.-4: "nervous system-related" -> "nervous-system related"

p.15 | §2 | I.3: "Native American" -> "Native Americans"

p.16 | §1 | I.6-7: "Two of these pathways remain significant after pruning for..." -> "After pruning, two of these pathways remain significant for..."

I.8: "...events, indicating..." -> "...events. This indicates..."

p.18 | §1 | I.1: "...pathways, what..." -> "...pathways. This..."

I.1-4: Could this be tested by screening for elevated divergence between African and American non-forest populations? These could be representative of different genetic backgrounds and/or population history.

I.-4: "among others SNPs" -> "among other SNPs"

§2 | I.2: "...with positive selection signals suggesting..." -> "...with signals of positive selection, suggesting..."

I.5. "signal" vs. "they": Should both be either singular or plural?

p.19-25 | References: Page ranges should be given using en-dashes instead of shorter dashes.

p.20 | Reference 8: Omit spaces in "U S A" (cf. reference 42)

p.22 | Reference 25: Insert space before "URL".

p.26 | Figure Legends | Fig. 1: "distribution for SNPs" -> "physical position of SNPs"; if appropriate, say that q-values are on the log10 scale

p.28-29 | Figure Legends | Suppl. Figs. 1--6: "distribution of SNPs" -> "physical position of SNPs"

Suppl. Fig. 7: "selection signals" -> "signals of selection"

Figure 1: I suggest increasing the font size for axis labels, chromosome labels and cluster labels. It would also be great if you could report estimates of α and β in a supplementary table, at least for 'significant' SNPs. Does this figure only show SNPs kept after the sliding-window smoothing / filtering? Comparison to Figures S1--S6 suggests so. Please clarify in the caption of Figure 1.

Figure 2: Why is there not always a shaded area for "any selection" if there is a coloured area in any of the other sectors? Again, does "any selection" include or ignore cases where at least one out of α_{Africa} and α_{America} was smaller than 0?

6. If you would like your identity to be revealed to the authors, please include your name here (optional)

your name here (optional).

Your name and review will not be published with the manuscript.

Confidential to Editor

1. Do you have any potential or perceived competing interests that may influence your review? Please review our [Competing Interests policy](#) and declare any potential interests that you feel the Editor should be aware of when considering your review. If you have no competing interests, please write "I have no competing interests."

2. Did you receive any assistance in preparing this review (e.g. from a post-doc or graduate student)? If yes, please provide his/her name and title here. If you have chosen to reveal your identity to the authors in question 6 above, please include the name of the person who assisted you in the response to question 6 as well. (optional)

3. If accepted, do you think this submission should be highlighted on the PLOS ONE website? PLOS ONE does not evaluate manuscripts based on perceived significance or readership. We aim to provide tools for readers to filter and evaluate our publications. (optional)

Thank you very much for the opportunity to review this article. This is a very well-written and thorough manuscript. I

think the paper deserves publication and I do not want to delay this unnecessarily. However, I have taken the freedom to state several concerns w.r.t. to the statistical analysis that I would like the authors to clarify. In the end, it is an editorial decision whether and to what extent these concerns need to be addressed.

The results are carefully discussed and interpreted. Previous studies relating to current results are mentioned. I was left with the impression that this study confirmed previously published candidates, but did not find a substantial number of novel candidates or regions.

I discussed particular points with Alisa Sedghifar (PhD student) and Graham Coop (PhD).

No

[Back](#)[Edit Review](#)[Print](#)[Submit Review to Editorial Office](#)