Author’s response to reviews

Title: Associations between male infertility and ancestry in South Americans: A case control study

Authors:

Maria Skowronek (fernandaskowronek@gmail.com)
Tatiana Velazquez (tavalentina@hotmail.com)
Patricia Mut (mut.patricia@gmail.com)
Gonzalo Figueiro (vazfigue@gmail.com)
Monica Sans (mbsans@gmail.com)
Bernardo Bertoni (bbertoni@fmed.edu.uy)
Rossana Sapiro (rsapiro@fmed.edu.uy)

Version: 2 Date: 15 Jun 2017

Author’s response to reviews:

Dr. Matteo Pasini
Editor in Chief
BMC
BMC Medical Genetics

Dear Dr,

Please find our revisions and corrections of the manuscript entitled “Associations between male infertility and ancestry in South American: a case control study” by Skowronek et al. to be considered for publication as a research paper in BMC Medical Genetics. We would like to thank again the reviewers. We tried to follow their suggestions and introduce changes that improved the original manuscript.

Main modifications incorporated include the fusions of tables 2 and 3 as well as 4 and 5 saving space and clarifying the data. We also performed the suggested changes in the figure 1.
We decided to include the information of the primers as supplementary Table 1 and include a supplementary Table 2 with the data of the analysis of F(xK). We incorporated as part of the main text the table that shows the ANOVA analysis, now Table 5 (in the previous version it was a Supplementary Table 1).

We think that these changes reflect the reviewers’ comments. We think that the quality and clearness of the manuscript has improved.

Finally, below you will find the specific answers to the reviewers. We thank you for your consideration and look forward to learning of your decision in the near future.

Sincerely,

Dr. Sapiro

Rossana Sapiro M.D. PhD

Histology and Embriology Department

School of Medicine,

Avenida General Flores 2125, CP 11800,

Montevideo-Uruguay,

Tel: (598-2)-9243414/3551,

Fax: (598-2)-9249563,

e-mail: rsapiro@fmed.edu.uy

Specific answer to the Reviewers:

Reviewer 1: The revised version of the manuscript assesses most of the concerns I had with the previous versions and the result is, in my opinion, generally improved.

My main concern was about Y chromosome genotyping and I acknowledge that the authors reported a more detailed Y chromosome haplogroups composition for the individuals in their studies.
In Table 3, the Authors report 16 individuals as belonging to haplogroup H and 5 to haplogroup O, however they do not report any test for markers defining these haplogroups in the text, nor in the figure. How were these 21 individuals assigned to their haplogroups?

Regarding haplogroup O, it was a typewriting mistake because the 5 fertile men are Q and not O, that was corrected in the actual version. We really apologize for the error and subsequent misunderstanding.

In the previous version of the manuscript haplogroup H was assigned to all individuals from Haplogroup F that lacks the SNPs of markers M201, M170 and M172. Following your suggestions, we consider all the individuals without these markers as a heterogenous group not belonging to the G, I, J2, or K haplogroups.

In the same Table, 13 individual are described F(xK) with undefined subclade. I suggest that these are indicated instead as F(xGHIJK) to avoid confusion with the F(xK) individuals used in the tests, which are correctly identified as such in Table 6 and supplementary Table 1. Similarly, I suggest replacing K(xP) with K(xLMNOP) in Table 3.

The 13 individual were already defined as F(xK). Actually we hypothesized that the individuals with undefined subclades probably belong to G,H,I or J2, but for technical reasons (especially bad quality of DNA) we weren’t able to genotype them. In this new version, we tracked 8 out of 13 men and got new samples of DNA, performed the tests and confirmed the hypothesis. Unfortunately, we could not get all the samples so we still have a small group of not determined subclades. We re-analyzed the data and they still are not statistically significant. Moreover we computer simulated the missing data and it looks that distribution tends to be similar between cases and controls.

While the newly reported haplogroups clarify the phylogenetic relations among the individuals in the study, the tests performed do not keep this into account and still consider haplogroup F(xK) as a unique entity. I suggest that the phylogenetic structure is kept into account in the tests as well, i.e. to perform the tests separately for the different haplogroups, or alternatively, that the Authors explain the rationale behind the selection of haplogroup F(xK) as a meaningful entity in their sample.

We decided to incorporate Supplementary Table 2 that shows semen analysis from G, I and J2 men. Interestingly, all of the groups were statistically similar between them and different to the other groups suggesting that it is the whole F(xK) group that is in higher risk of presenting abnormal sperm morphology.

As a minor point, I suggest, for the sake of clarity and accuracy, that Figure 1 should be redrawn to reflect the real phylogenetic relations known (i.e. subsequent ramification in F(xK) branches, rather than parallel clades), even if the authors are not including all the markers. For completeness, markers not included in the study (for instance M429, which marks the root of IJ haplogroup) could be indicated in the figure, making sure they are clearly from the markers used.
in the study. Moreover, the indications (xK) and (xP) should be removed from the figure, as they incorrectly mark the root of haplogroups F and K respectively.

Finally, the legend to Figure 1 should be amended to reflect the increase in the number of markers assayed.

As the reviewer suggested, the figure 1 was totally redrawn following the ISSOG phylogeneic tree. We only included the haplogroups defined by the 14 markers that we used, to give a clear picture of the relationship between them. The legend was amended.

Some additional minor suggestions:

page 5, line 2 from top: I wouldn't include the fact that "one-half consists of tandemly repeated DNA" among the reasons why the Y chromosome is used in population genetics.

The line was changed, the new version is “It is because of its particular properties, (paternally inherited, lack of recombination and abundance of genetic polymorphisms) that the Y chromosome has been used as the tool for investigating recent human evolution from a male perspective”

Page 8: I suggest including an explanation of WHO guidelines about motility class: what do a, b, c, d mean?

We added the explanation. The meaning of each kind of motility following WHO definitions was included in this new version

page 16: I suggest editing the sentence "The SNPs that determine haplogroup F(xK) are very close to candidate genes that may be involved in the determination of sperm morphology and infertility, thus the possibility that mutations of these genes may be inherited together may be further investigated". As the Y does not recombine, all mutations on the chromosome are inherited together regardless of position, function, or vicinity of genes, unless back-mutation, gene conversion or deletion events remove them.

We understand the issue and completely agree with the reviewer so we modified the sentence. We want to point out that the analysis of the architecture of Y chromosome could lead us to understand several issues related to fertility, populations, etc.

Answer to Reviewer 2

The Authors have competently addressed most of my concerns and I have only minor suggestions remaining. Most importantly, I suggest saving space by

(1) moving Table 1 to the Supplement since this is a methodological detail that presumably holds no new information (i.e. the primer sequences have probably been published before)  
(2) merging Tables 2 and 3 by adding two more columns to Table 3. This table could then have the
columns "Main haplogroup", "Subclade", and "Ancestry", followed by the counts of fertile and infertile men

(3) likewise, merging Tables 4 and 5 by introducing the "Ancestry" column into Table 5. The two different Fisher's exact tests could both be reported under this Table, where one refers to using "Ancestry" as a factor with 3 levels, and the other using "Haplogroup" with 13 levels. Together, this should create enough space to allow incorporating the Supplementary Table into the body of the paper, since, in my opinion, this Table is the most useful representation of the findings of the study (I maintain that Tables 6 and 7 lack statistical power compared to this Supplementary Table and they show no metric values that could be compared to those of other studies).

We really appreciate the suggestions. The modifications are the following: the table that shows the primers is the supplementary table 1. Table 2 and 3 were fusion and now is Table 1. Table 4 and 5 is now Table 2 and the former Supplementary 1 is now Table 5.

Specific minor comments:

(1) Page 7, last line and first line of page 8: Here it is said that 6 parameters were measured, but only 3 of them get reported later in the Tables. Were the remaining three parameters also subjected to testing for association with genotypic data? If the Authors regard this data to be potentially valuable for future meta-analyses they may want to consider incorporating it into the current Supplementary Table. Note that discarding potentially valuable information because of non-significance of associations introduces a systematic bias into the published literature which makes meta-analyses questionable.

We agree with the reviewer that some data of the semen analysis may be useful for this or other studies so we included main data in the Results section. However part of the missing data of the former version are routine in semen analysis and they are only valuable in particular cases. E.g., viability is important if after motility analysis the physician finds a lot of non motile sperm (class d) so he/she suspects necro-zoospermia, meaning that it is usually not so important regarding fertility in the whole infertile population (as morphology, concentration or sperm motility are).

(2) Page 8, line 20: I suggest reporting the repeatability of motility measurements (e.g. Pearson correlation between first and second measure)

We added the correlation between measurements

(3) Page 8, line 27: Is "vitality" the same as "viability" in line 1?

The assessment (Eosin-nigrosin) is for sperm vitality, which is the correct name. Viability is incorrect so we changed. We really appreciate the reviewer’s observation since it was skipped even by the language correction

(4) Page 10, line 30: In my experience, statistical testing becomes most powerful when the dependent variable is transformed to approach normality (e.g. log(proportion normal sperm) *100
and then parametric tests (e.g. ANOVA) are conducted, rather than having to rely on non-parametric tests, but this decision should be up to the Authors.

After receiving this revision, we performed the analysis in the way suggested but we did not obtained any differences so we preferred to maintain the former version.

(5) Page 10, Line 54: delete "the" before "60%"

This modification was done

(6) Page 11, Line 29: "paternal Native American ancestry was not present at all among the cases" This modification was done

(7) Page 13, line 31: "inherited"

This modification was done

(8) Page 14: line 7: "depending on the region"

This modification was done

(9) Page 14, line 12: insert dot after "[61]"

This modification was done

(10) Page 15, line 17: "In this study we focussed on male infertility, trying to discard… This modification was done

(11) Page 15, line 37: insert comma after "[69, 70]"

This modification was done

(12) Tables 6 & 7: consider replacing "mill" by "x10^6"

We did the replacement

(13) Supplement Table 1: Legend: "standard"

It was corrected