Author's response to reviews

Title: The significance of Lactobacillus crispatus and L. vaginalis for vaginal health and the negative effect of recent sex: a cross-sectional descriptive study across groups of sub-Saharan women.

Authors:

Vicky Jespers (vjespers@itg.be)
Janneke van de Wijgert (j.vandewijgert@liverpool.ac.uk)
Piet Cools (piet.cools@ugent.be)
Rita Verhelst (riet7@icloud.com)
Hans Verstraelen (hans.verstraelen@ugent.be)
Sinead Delany-Moretlwe (sdelany@wrhi.ac.za)
Mary Mwaura (mwaura.maryw@gmail.com)
Gilles F Ndayisaba (ndayisaba@gmail.com)
Kishor Mandaliya (kishor@icrhk.org)
Joris Menten (jmenten@itg.be)
Liselotte Hardy (lhardy@itg.be)
Tania Crucitti (tcrucitti@itg.be)

Version: 3
Date: 16 December 2014

Author's response to reviews: see over
Dear Editor,

Thank you for the review of the manuscript “The significance of Lactobacillus crispatus and L. vaginalis for vaginal health and the negative effect of recent sex: a cross-sectional descriptive study across groups of African women”.

Please find enclosed a thoroughly revised version of our manuscript. We recognize that the previous version was suffering from a number of weaknesses, which we have corrected in the present version. We have addressed the comments point-by-point and have listed them in the cover letter. The most important changes are the following:
- The result and conclusion section of the abstract have been changed to more accurately summarize all data as suggested by reviewer 1.
- The diverse groups of African women with the large two reference groups in several countries are now the focus of the abstract and throughout the manuscript.
- The result section has been substantially expanded.
- The data analysis section has been amended to state the tests and variables used.
- The discussion section has been expanded and includes the items suggested by reviewer 2.
- Table 2 includes the respective values and significance tests.

We thank the reviewers for their in depth review and valuable comments and the Editor for the opportunity to revise the manuscript. We hope that the present version is acceptable for publication in your journal.
**Reviewer comments:**

**Reviewer’s report 1**

We thank the reviewer for the in depth review and suggestions to improve the manuscript.

This paper is a highly significant contribution to the literature related to vaginal microbiota in diverse groups of African women in several countries, allowing for comparisons and insights that have not been described before. The main strength of the paper is the quantitative estimation of abundance for well-known dominant organisms in the human microbiota, allowing for extensive profiling of a relatively large number of samples. However, I have serious reservations about the way the authors have organized and presented their data in the manuscript. The results section is thinly described, leaving the reader to struggle through interpretation of large tables with little help from the text. In my opinion, the manuscript requires extensive re-organization and editing to remove awkwardly worded and repetitive sections that mitigate the impact that this paper could have given its exciting scope and obvious value for researchers in the field.

Answer: We thank the reviewer for the thorough review of the manuscript. We agree with the shortcomings as mentioned and have revised the paper extensively.

The major limitation of the current study is the lack of resolution available with single primer sets to detect functionally different groups of micro-organisms from the same species. For example, the authors discuss G. vaginalis subgroups (lines 461-4) but offer no insight into whether the primers used in this study target all or only one subgroup. An enhanced discussion of this limitation, in reference to deep sequencing approaches for dissecting the vaginal microbiome, would be much appreciated.

Answer: Our methods did not study the functionally different subgroups of any of the species quantified as this was not an objective of the study. Nevertheless, this could be regarded as a shortcoming and therefore we are currently investigating the role of the different genotypes for G. vaginalis (to be reported at a later stage). Further, the primer set we used for the amplification of G. vaginalis amplifies the three distinct genotypes as described by dos Santos Santiago et al. We verified this in another study for which the results are not yet published. We confirm that the primer set as used in this manuscript amplifies G. vaginalis genotypes containing and lacking the sialidase gene. We validated the different primer sets in silico first and then determined their analytical sensitivity, limit of detection and specificity using reference isolates of various species and genera and clinical specimens. At present, our results do not indicate that we have missed specific subgroups.

We applied single species quantitative PCR, as this method is more sensitive compared to next generation sequencing and provides us concentrations. We considered adding a section to the discussion of the manuscript on this topic but finally we concluded that it seemed beyond the scope of the stated objectives. Nevertheless, we plan to include this topic in a next paper dealing with the methodology of the tests.
Major compulsory revisions:

1. line 55-6: The fact that two large groups of “reference women” in 2 different countries (as well as two groups of pregnant women and two groups of adolescents) were assessed is totally under-emphasized in this paper (no mention until Additional File 3). I understand that few differences may have been observed between Kenyan and South African women, but this finding is striking in itself. These comparison groups are a major strength of this paper and deserve to be highlighted in more detail.

Answer: The result and conclusion section of the abstract have been changed to more accurately summarize all data as suggested. The diverse groups of African women with the large two reference groups in several countries are now the focus of the abstract (line 63-78) and throughout the manuscript (line 290-353).

2. line 62-72: I am unclear as to why the most general information about the entire study group is presented as the most important results in the abstract. No mention is made of the different study groups and comparisons between them although this is obviously a critical strength of this study. If there are no real differences between the subgroups, I think this is really important to mention here. Simply confirming that the major organisms are associated with Nugent score is fine, but seems more important as an overall validation of your method rather than a fundamental insight of your study.

Answer: As above; the result and conclusion section of the abstract have been changed to more accurately summarize all data as suggested. The diverse groups of African women with the large two reference groups in several countries are now the focus of the abstract (line 63-78).

3. line 63 (and throughout): I am uncomfortable with these measurements being called “counts” or “cell counts” since no cells were actually counted in these experiments. I would prefer the term “genome copies” which better reflects the indirect method used, and that the term “concentrations” (for copies/ml) be used instead of counts throughout. Also, some discussion of the quantitative limitations of using 16S, present in multiple copies in most organisms, to represent single genomes. I think it would be better to use the term “presence” or “detection” instead of “prevalence” to describe observation of these organisms in samples.

Answer: We agree with the reviewer and have now used the terms “genome copies”, “concentrations” and “presence” throughout the manuscript. We considered adding a section to the discussion of the manuscript on this topic but finally we concluded that it seemed beyond the scope of the stated objectives.

4. line 73-77: Again, these conclusions do not address the different subgroups studied. The second sentence is also very confusing – “prevalence of low... was unexpectedly high”. I would like to see a more convincing concluding statement.

Answer: As above; the conclusion section of the abstract has been changed to more accurately summarize all data as suggested (line 74-78).

5. line 140-146: Description of recruitment should be provided in more detail, or ref,
since these diverse study groups are such a major strength of this study.

Answer: A detailed description of the recruitment has been added (line 140-149).

6. lines 158-211: I find these 2 sections to be awkward and confusingly organized. I would reorganize as follows: Put shipping info, Nugent scoring, swab processing, PSA measurements and RTI diagnostics first, in a section called “Swab processing and laboratory procedures” or at the end of the previous section (renamed “Clinic visit and laboratory procedures”). Selection of organisms doesn’t need its own section, just start a new section called “Quantitative PCR of selected organisms” with “We designed or selected primers targetting the following genus and species which have previously been shown to be important members of the vaginal microbiota (9,14)”.

Answer: We thank the reviewer for the suggestions. The sections have now been re-organized (line 150-183).

7. line 212-258: The data analysis section is too wordy, and the vague term “predictor analysis” is over-used. Just state which tests were used on which variables. Justification of variables and tests used should go into the results or discussion. The entire last paragraph (lines 242-58) should be cut down to a sentence or two and moved to results.

Answer: We have stated the tests and the variables used as suggested by the reviewer and removed the “predictor” wording throughout the manuscript (line 214-247). The reasoning behind the modelling has now been explained including the choice of the variables in the model (line 240-247).

8. lines 273-286: The first paragraph of results section simply states the contents of all the figures and tables, which is unnecessary. Just refer to the tables and figures as you describe the highlights and most important findings that they contain.

Answer: We removed this section as suggested.

9. lines 318-338 and throughout results: Much of these 2 sections simply repeats values that are already provided in Table 1. I would remove this redundant information and provide an enhanced description of what these numbers are telling us about the different subgroups of women. Also, all the odds ratios in text make the text difficult to read and would be better off in a table in the main text, at least for the most important findings.

Answer: We have joined the two sections into one and improved the reading by removing redundant information as suggested by the reviewer (line 290-312).

10. line 381: It is not clear to me why this is an “alternative” modelling approach or how it is different from the previous analysis shown, since both seem to involve univariate followed by multivariate comparisons. Is one better than the other? Are they so different from each other that seeing both at the same time is helpful? A better justification of this approach is needed.

Answer: The modelling has the advantage that it includes all variables in the same
analysis and that variable can be selected based on assumptions. The reasoning behind the modelling has now been explained including the choice of the variables in the model (line 240-247).

11. Table 2: I don’t see the utility of simply showing significance levels for all the possible associations with no data that allows the reader to judge the associations by. I would like to see the actual differences in values, with significance tests included, please. Rather than showing all the tests in a table and all the values in supplementary, just show the most important values or “model” along with significance tests in the main article.

Answer: Table 2 and also table 3 have been adapted to accommodate the values, odds ratios and confidence intervals. The manuscript reflects the given data in table 2. As a result, the supplementary tables (additional file 4-8) have been removed.

Minor essential revisions:

All the revisions have been adapted and clarified as suggested by the reviewer. We thank the reviewer for his thorough review. We have answered the remaining queries point-by-point below.

12. line 50: “potentially protective or harmful” instead of “protective and disruptive”
13. line 57: “species to subgroups” is confusing – How about “compared species variation in subgroups”?
14. line 60-1: “proximate local” and “more distal underlying” are awkward and redundant – how about “in relation to several factors of known or theoretical importance”
15. lines 85-90: Very long sentence – please split into 2 sentences. line 97: Remove the word “said”.
16. line 98-99: Clarify what you mean by “different HIV epidemiums”. line 100: Remove the words “of normality”.
17. lines 101-103: I would reword this sentence to something like: “Further research is needed to address how composition of the vaginal microbiome determines optimal vaginal and/or reproductive health”.
18. line 104: “STI infections” is redundant.
19. line 105: Should read “address biological vulnerability”
20. line 106: I think this sentence is unnecessary. Maybe ref could go after “neonatal infections” in line 88?
22. line 111: How do you know these are the 5 most prevalent Lactos? Reference dataset or are you relying on common knowledge?

Answer: The references have been added.
23. line 115-116: “species variation in subgroups”.
24. line 121: No need to call these factors “more distal underlying”.
25. line 126: Remove “per protocol” and line 127: “and either”.
26. line 129-131: Since you did an 8-month study why aren’t you showing the data for the whole thing?? I wouldn’t mention it otherwise - delete last 2 sentences.
27. line 138-139: Is this the CSW group? Each group should be called one thing consistently throughout the study.
28. line 139-140: Can you really say that HIV+ women are in good health?
“Showing no symptoms”, “high CD4 counts”, “no OIs”? 
30. line 155: “by a clinician, including colposcopy. Vaginal swabs…” 
31. line 157: “previously” instead of before 
32. line 171: “indicates” instead of “reflects” 
33. lines 213-4: Why is it important to state that the data analysis plan was prepared prior to data analysis?? 
Answer: this sentence has been removed. 
34. lines 214-5: Obviously cross-sectional analyses involve single timepoints – no need to state explicitly. 
35. line 216: No need to state that RTIs were not re-tested at enrollment. “Data from the enrolment visit was used except for RTI diagnoses and information about sex partners, which was collected at screening”. 
36. line 221-2: “variables” instead of “characteristics” 
37. line 222-4: This could be stated more economically as “qPCR data was expressed categorically as presence/absence, or alternatively presence was divided into 3 separate categories: Not quantifiable, <10-6 copies/ml, >10-6 copies/ml. For quantifiable levels, mean and SD were calculated”. What is the difference between not quantifiable and <10-6 copies/ml ?? 
Answer: We are able to detect the bacteria above 1600 genome equivalents/ml. The quantification process starts at 16000 geq/ml as low concentration measurements are less accurate. This category was therefore defined as present but not quantifiable. The cut off of <10-6 geq/ml was set to distinguish between high and low concentrations and was not related to the limit of detection. This has now been added to the manuscript (line 222). 
38. line 226-7: So why mention it? 
39. line 228: I don’t like the term “predictor analysis”. “We performed logistic regression on the presence/absence of each bacterial species (Lactobacillus genus and Candida data not included).” 
40. line 230-1: Why are there “consequently” small differences in p-values from chi-squared tests? Different from what? 
Answer: This sentence has been removed. The statistical tests have been described in the methods section (line 215-240). 
41. line 231-233: You already described the Nugent score. 
42. line 233-5: Why exactly was antibiotic use “collinear” with HIV? This should be stated somewhere in results? 
Answer: Most HIV-positive women were using long term cotrimoxazole prophylaxis for the prevention of HIV-associated opportunistic infections. This has been added to the statistical methods section (line 230-231). 
43. line 235-6: You already said you excluded genus-level. 
44. line 236-9: Long confusing sentence – “excluding other indicators...were considered for inclusion”?? “the vaginal microbiome health” is unclear. Please reword to clarify. 
45. line 240-2: Weren’t “sub-analyses” carried out for all the subgroups? 
Answer: The analysis included a comparison between groups by including the variable group or group-site but a full subgroup analysis was only performed on the reference group. The sample size of the other groups was too small to allow for relevant conclusions. As before, we improved the reading of the stats methods section. 
46. line 245-6: So did you correct for multiple comparisons? If not, why mention it? 
Answer: This had been removed. 
47. line 250: Why is logistic in brackets?
Answer: The brackets referred to the analysis of the *Lactobacillus* concentrations. The stats methods section has now been changed.

48. line 252: Don’t refer to yourselves as “the authors” – just say “we”
49. line 256-8: These statements about HSV-2 seem to come out of nowhere – what do you mean by “temporal causal relation”?
Answer: The statement has been removed. The limitation to define a cause effect relationship has now been mentioned in the discussion section under shortcoming of the study (line 531).

50. line 261: What is a legally authorized representative?
Answer: This has been removed.

51. lines 264-5: “by signing or marking the consent form”
52. lines 279-280: You already stated this in methods.
53. lines 289-90: remove “by design”.
54. line 299-305: This section is over-explained. You already mentioned that you excluded antibiotics for HIV+ women in methods, and you repeat it twice here. Could you just describe antibiotic usage in one way? No specific antibiotics are mentioned – do you have that information?
Answer: This has been summarized (line 275-277).

55. line 308: Spelling out “fifty seven and a half” looks awkward. How about: “Overall, 57.5% of women...”
56. line 319: “proportion of 8.3%” is redundant, “detected” instead of “present”
57. line 320: “measured by genus-level qPCR.”
58. lines 340-51: Collapse to a single section.
59. line 379: Maybe a sentence to indicate the biological relevance of these findings?
Answer: This has been added to the discussion (line 489-491).

60. line 479: “sex workers” (plural)
61. line 480: remove “solely”
62. line 485-6: “confounding...confounders”
63. line 496-8: Seems unnecessary to say.
64. line 512-3: This is almost identical to a sentence in the introduction.
65. line 545: Thiong’o should be capitalized.
66. line 546: “capturer”
67. Table 1: Row, G. vaginalis, Col, Pregnant Women, Kenya - % value is miscalculated

Quality of written English: Needs some language corrections before being published
Answer: the manuscript has now been reviewed by an English native speaker.
Reviewer's report 2

We thank the reviewer for the review and suggestions to improve the manuscript.

Overall this is a well-written, large cross-sectional study that aimed to characterize key members of the vaginal microbiome across groups of African women. The question posed by the authors is well defined. The methods are appropriate and well described. The data are sound and the figures appear to be genuine. There is a section that outlines the hurdles faced during the data analysis. The title and abstract accurately convey the findings of the study.

The manuscript could benefit from additional discussion and comparisons to recently published work in this arena and those suggestions are indicated below. In addition, the authors could point to future directions or areas in which they will expand based off of the data generated in this report.

Answer: We have revised the discussion section as suggested by the reviewer and included future research directions.

- Discretionary Revisions
  1. For the additional files, the references are not integrated into the overall reference section of the manuscript. Suggest incorporating.
  Answer: The references have been included in the reference section of the manuscript.

- Minor Essential Revisions
  1. It would enhance the conclusions to include a future studies or directions at the end of the manuscript.
  Answer: The following has been added: “Longitudinal studies are needed to study health outcomes, including acquisition of STI and HIV. Additionally, research is needed to understand the role and function of different microbial species and sub-species (e.g. G. vaginalis) in relation to the protection and susceptibility to infection.” (line 545-548).

- Major Compulsory Revisions
  1. In results section the percentages are clearly outlined, but it does not include references to the particular tables or figures where the data can be found. Suggest including those references throughout.
  Answer: The references to tables and the figure have been added.

  2. In results section there is no mention of the correlation or lack of correlation between educational or marital status, please incorporate these results in this section.
  Answer: The variables education and marital status were not withheld for this cross-sectional analysis. In a previous report, published after the initial submission of this manuscript, no association between the presence of bacterial vaginosis and education (p=0.619) or marital status (p=0.196) was seen. The publication reported the education levels and the marital status for the different groups in the study:
  Jespers V et al. Prevalence and correlates of bacterial vaginosis in different sub-populations of women in sub-Saharan Africa: a cross-sectional study. *PLoS One* 2014, 9:e109670. We agree that the additional table 3 is confusing and we
therefore have removed the variables that are not presented in this manuscript (including education and marital status) from this table. Further, we have now listed and clarified all the variables that were included in the analysis in the stats methods section (line 213-247).

3. Many behavioral factors outlined in the additional files are not presented in the results section nor are they discussed later in the manuscript. It is not clear why this data is included if it is not going to be included in the results or discussion. Suggest revising.
Answer: We agree with the reviewer and as explained and mentioned above, we have now removed the variables that are not included in the methods and results section.

4. The pH findings should be discussed in the context of the literature in other reported/published data sets and among women from different geographical areas (Europe, US, etc.).
Answer: We added the following to the manuscript (line 476-487): “We detected a higher presence of Lactobacillus species, with the exception of L. iners, for women with a low pH. The association of pH with individual vaginal species has not been described previously but a higher abundance of lactobacilli has been associated with a lower vaginal pH among 100 cycling Chinese women [43], 494 asymptomatic Estonian women [44], and 396 asymptomatic North American women [20]. The median pH observed in the reference groups was well above 4.2, the cut-off for which values below are reported as normal [51]. The pH, measured in clinical studies as one of the Amsel criteria, is not often described separately in the literature and data is lacking for women in Africa and in general. A median pH of 3.6 was described in a Belgian healthy population of 141 women [52]. Moreover, the vaginal pH in different ethnic groups in North America was as follows: Hispanic 5.0; black 4.7; Asian 4.4 and white 4.2 [20] and for Estonian women the mean value was 4.7 [44]."

5. The increase in Lactobacilli in pregnant populations should be discussed in the context of recent reports (Romero et al, Microbiome, 2014 and others) on the vaginal microbiome in pregnancy.
Answer: The following has been added to the discussion (line 438-446): “Pregnancy has been shown to be associated with low bacterial diversity and high levels of lactobacilli, particularly L. crispatus [35, 36]. The higher Lactobacillus concentrations that we observed in the pregnant women agrees with a longitudinal study in 22 pregnant, mostly African American women, showing a higher abundance in 16S rRNA V1-V3 of Lactobacillus vaginalis, L. crispatus, L. gasseri and L. jensenii as compared to 20 non-pregnant women, of whom 10 were African American [35]. Similarly, a US longitudinal study of 12 Caucasian healthy women showed a stable Lactobacillus dominant vaginal microbiome (16S rRNA V3-V5) throughout pregnancy [36]. In future, larger and longitudinal studies are needed to adequately characterise the vaginal microbiota among women in sub-Saharan Africa.”

6. A discussion on parity and changes in vaginal microbiota should be added along with a discussion as to why the authors think that L.vaginalis and L.iners prevalence decreases with increased parity along with the strong association with Candida. Have other groups shown these associations?
Answer: We have added the following to the manuscript (line 488-497): “To our knowledge, the association of parity with a decrease in *L. vaginalis* and *L. iners* presence has not been described previously. This suggests that pregnancy, a period of high oestrogen status with high lactobacilli presence, is followed by a reduction of certain strains of the lactobacilli species. It is possible that the delivery period could disrupt the stable vaginal lactobacillus population attained during pregnancy. This hypothesis needs further study. *C. albicans* was positively associated with parity in our study. It is known to be associated with pregnancy which may indirectly explain the association with parity. However, a study in 500 Australian pregnant women showed no difference for vaginal candida colonization and parity history [53]. This may indicate that candida colonization does not normalize after delivery.”

7. In the section (lines 422-435) there is a recent report by Benning et al, PLoS ONE, 2014 that investigates the vaginal microbiota profiles in HIV women from Rwanda. Those findings should be discussed in the context of this manuscript and in this section of the discussion.
Answer: We thank the reviewer for pointing this out. This publication has been added to the discussion.

8. Other vaginal microbiome studies have shown that douching and other feminine hygiene practices impact the vaginal microbiota profile, as such the authors should discuss and hypothesize why they did not observe this in their study. Is this due to different practices, timing, etc.?
Answer: We have added the following to the discussion (line 498-510): “Though research has shown that intravaginal cleansing is a risk factor for BV [54, 55], we did not observe a negative effect of the use of products to externally wash, internally cleanse, or use of products for drying or tightening on the presence or concentrations of species or on BV status by Nugent scoring. On the contrary, in the reference group, a negative association with *A. vaginae* and a positive association with *Lactobacillus* genus were present among those who reported cleansing during bathing. A systematic review of longitudinal studies concluded that intravaginal cleansing with soap was associated with the development of intermediate vaginal flora and bacterial vaginosis in women with normal vaginal flora at baseline (pooled adjusted odds ratio 1.24, 95% CI 1.04–1.47) [55]. However, there is also evidence that vaginal practices are highly heterogeneous and, therefore different study populations use different practices and products [56]. Indeed, women in this study mostly performed washing with water during bathing as opposed to using soap or detergents, which may explain our findings.”