Dear Recommender of PCI Infections
We have amended our preprint #191 following your comments and the remarks made by the three referees. We have taken into account all these remarks and provided explanation, for a few of those, when we did not exactly follow the recommendations made.
You will find all modifications undertaken in the track-change version of our manuscript, and our rebuttal at the end of the present document.
We hope that you will find the new version of our preprint suitable for recommendation in PCI Infections and remain at your disposal for any modification or question you may find necessary to ask.
Sincerely
Thierry de Meeûs

Recommender comments
Your decision
by Hugues Nana Djeunga, 09 Dec 2023 15:02
Manuscript: https://doi.org/10.1101/2023.07.25.550445 version 1
Major revisions

The manuscript submitted by Kagbadouno and colleagues entitled “Population genetics of Glossina palpalis gambiensis in the sleeping sickness focus of Boffa (Guinea) before and after eight years of vector control: no effect of control despite a significant decrease of human exposure to the disease” aim to investigate the impact of tiny target-based vector control on the population biology of G. p. gambiensis in Boffa, using microsatellite markers genotyping and population genetics tools. This is an important study in the field of public health.

Based on my evaluation and the reports from three invited independent Reviewers, this manuscript presents a number of issues and limitations that need to be appropriately addressed before being considered for publication in PCI Infections. Below are some of the most important issues:

1. The justification of the study and problem statement are unclear to me. There is no information on the sleeping sickness status in Guinea or the Boffa focus that would have justified a vector control. What was the impact of vector control? What is the link with animal reservoirs?

Answer: We apologize for our miscomprehension, but this comment is very hard to understand for us. Indeed, a full paragraph was already devoted to these issues, lines 57-69, with the most relevant literature cited. We have nevertheless added some more precisions.

What is the link with the post-vector control increase of the GPCAG allele in Côte d’Ivoire? Was resistance established in Côte d’Ivoire after vector control or this was just a hypothesis/speculation by the authors?

Answer: We again need to apologize, since we unfortunately failed to understand this comment. Indeed, a full paragraph was already devoted to these issues, lines 70-73 (old manuscript). The hypotheses to interpret the very odd behavior of locus GPCAG before and after control In Bonon (Côte d'Ivoire) were clearly stated in the reference cited, which

was published in a peer reviewed journal with a very good reputation (https://doi.org/10.1016/j.meegid.2019.103963).

The objective of the study is different between the abstract and introduction sections, and seems not achieved in this study (based on the objective presented in the introduction section).

Answer: This remark is hard to follow. In the results section, we exactly provided the results of the effect of 11 years of vector control on the population genetics, and sex ratio of this population of *Glossina palpalis gambiensis*. We have added some more details in the abstract regarding sex ratio, which indeed was not mentioned in version 1 of the manuscript. Nevertheless, except that, we could not find any discrepancy between the Abstract, the Introduction or the Results sections. We looked for any effect of the vector control on the sex-ratio and the population genetics of this population and found none.

2. The study design is not clearly presented. It is unclear how vector control was conducted. For example, (i) how many traps deployed on which area, (ii) for how long tiny targets were deployed before being changed (knowing that their efficacy relies on the baited insecticide), (iii) were the traps set at the same position, (iv) what do the authors consider as cohort? .... The study area is not presented in a comprehensive way, and the figure provided is not that informative. These details are useful for the interpretation of the results.

Answer

Figure 1, Table 1 and Table S1 provide all information on the geographic position of each trap, where genotyped tsetse flies were captured, and when. Regarding vector control, which is not the focus of the present paper, trap deployment strategies and success in Boffa were partly detailed in Courtin et al 2018 that we already cited, and in Camara et al 2021 (new reference). We have added a sentence to specify that in the Introduction section and at the beginning of the Material and methods section.

3. There is not enough information in the discussion section on the effect of vector control on population genetics metrics. For example, it is unclear what can be the influence of the areas not covered by the vector control, knowing that flies' dispersion is about 40 km. Also, the conclusion of the manuscript is a hypothesis rather than a real conclusion, and it is unclear why the authors are raising the hidden human and/or animal reservoir in this context.

Answer: The main information is that there is not any kind of genetic signature of VCC, in any of the analysis undertaken: no subdivision between dates for neutral loci or GPCAG, no difference of effective population sizes, even after more than 100 generations, and absence of any bottleneck signature. We have tried to expound it a little more in the discussion, and added a reference (Camara et al 2021). Regarding "the influence of the areas not covered by the vector control, knowing that flies' dispersion is about 40 km", we are not sure of what the recommender was thinking of there, since we indeed evidenced an absence of any spatio-temporal genetic structure. It means that sampled spots, which are also, among others, submitted to VCC, are always and continuously re-populated by flies that are representative of the whole population. We have added two supplementary sentences to insist on that point (see track changes in the Discussion section).

Also, the conclusion of the manuscript is a hypothesis rather than a real conclusion, and it is unclear why the authors are raising the hidden human and/or animal reservoir in this context.

Answer: We respectfully happen to disagree with the Recommender's opinion. We have added references that clearly explain the role of reservoirs, and our conclusion is much more a useful recommendation than a hypothesis. We indeed already discussed that VCC offers a clear protection against the disease, despite its weak or even absent effect on the population biology of this vector population, as proven by our population genetics analysis. To avoid further misunderstanding, we have insisted again on such points in our conclusion (last paragraph).

4. A few minor editing needed to be addressed: (i) please avoid some unusual abbreviations (for example “aka” at line 50, “in minimax” at line 455) ...; (ii) the citation of references in the text should be harmonized a follow the guidelines of the journal ...

Answer: we have replaced all "aka"'s with "also known as". For minimax, since we clearly defined this term in the Material and Method, *Effective population sizes and effective population densities* section, end of first paragraph, we chose to keep it as such in the subsequent sections of the manuscript. The reference cited and the reference list were undertaken with Zotero, using the PC Journal style provided in the corresponding website.

In addition to my comments and suggestions, the reports of three independent Reviewers provide detailed appreciation of the manuscript. If you are able to fully address these points, we would encourage you to submit a revised manuscript to PCI Infections. Once you have made the necessary corrections, please include a cover letter with a point-by-point response to the comments, including a detailed rebuttal of any criticisms or requested revisions that you disagreed with. Please also ensure that all changes to the manuscript are indicated in the text by highlighting or using track changes. A decision will be made once we have received your revised manuscript.

We look forward to receiving your revised manuscript and please do not hesitate to contact us if you have any questions.

Best wishes,

Hugues C. Nana Djeunga, PhD
Recommender, PCI Infections.
I

Reviews

Reviewed by Fabien HALKETT, 22 Nov 2023 10:00

Kagbadouno and colleagues present a fine population genetics study aimed at testing the effect of a policy control on the population size of the fly responsible for sleeping sickness in a Guinean focus in Boffa.

They exploit the full power of microsatellite loci to test and produce a high quality dataset. On these data they use various methods to calculate effective size, which can be

extrapolated to insect density in the study area. Their precise study does not appear to show any difference in population size between samplings performed before and after the onset of the control policy, even though the latter effectively reduced the outbreak of HAT disease in Boffa. The authors conclude their study with a message of prevention, arguing that the control campaign should not be interrupted prematurely, otherwise the number of cases of the disease will start to rise again as quickly.

The study is sound and the methods employed are robust. The data pre-processing step can be cited as an example. I have no doubts about the conclusions of this study. I have only two major comments to improve the presentation of the context and provide more detail on the population size estimates.

1. The introduction is rather short, a little too short, and could go into some detail about the policy of controlling the insect vector.

Answer: We have added an additional reference on the ongoing VCC occurring in Boffa (and elsewhere in the Guinean mangrove) and a few modifications. The strategy used is explained in details in the references we have cited. Our manuscript does not deal with this aspect but more on the consequences of VCC on the population genetics and population biology of this vector. We thus prefer avoiding a detailed description of VCC here, because it would require a very long paragraph that would simply repeat what is already explained in details in other and cited papers.

2. The different methods of estimating population sizes show wide variations, with temporal methods in particular giving much higher values. I think it's a matter of regret that the authors don't go into more details about these results. The analyses of the various cohorts are aggregated in a single table that presents the mean values over all the samples. This does not allow us to visualize the differences before and after the onset of the control. This result is only stated in one sentence (without reference to the underlying method).
Even if the analysis techniques differ, it seems possible to me to contrast the before/after estimates for each method, including the temporal method, e.g. by distinguishing between pairs of cohort sampled before/after the onset of the control.

Answer: We have added a Figure (Figure 6) that presents all effective population size estimates with each method and for each cohort and cohort pair.

Detailed comments:

L38 indicate the number (instead of several that is quite vague).

Answer: Done

L38 end of the line, remove the s at genetics, before tools

Answer: Done

L66 same grammatical mistake, remove the s at the end of individuals before clones (and I prefer the term clonal lineage ou clonemates)

Answer: Done

L62 Please provide more details on this vector control campaign. What does it involve?

Answer: Done, please see our answer to Recommender's point #2.

L86 Remove the name of the author in the parenthesis (only the date).

Answer: Done

L113 two month generation time (without s at month)

Answer: Done

117 considered as distinct time sample (rather than separate entities).

Answer: we have replaced this by " as belonging to distinct subsamples".

L139 described in Berté et al (2019) – brackets around the date only

Answer: Done

L145 to keep

Answer: changed into "and to keep"

L146 population genetic analyses (remove ‘s’ and data).

Answer: Done.

L153 note that the appropriate level of population delineation (or the test for a lack of population structure) can also be performed using assignment methods (e.g. DAPC analyses). It could be interesting to further cross the Fst based and DAPC method (also considering the slightly positive Fis value, which may indicate a Wahlund effect behind the contribution of null alleles).

Answer: We respectfully disagree with this opinion. When geographic data are available, Bayesian, or pseudo-Bayesian clustering algorithms will never find the exact appropriate partition. Moreover, we know that these procedures can be very sensitive to deviations from the null model of the populations under investigation (local deviation from panmixia, isolation by distance, linkage disequilibrium, and amplification problems). In particular,

DAPC can provide spurious results (e.g. Figure 7 in De Meeûs T, Chan CT, Ludwig JM, Tsao JI, Patel J, Bhagatwala J, Beati L (2021) Peer Community Journal, 1, e40. https://doi.org/10.24072/pcjournal.34). We thus tend to use such procedures, and in particular DAPC, with extreme caution, and only when we really have to, which is not the case here. We have added a sentence at the beginning of the Discussion section, which argue on the absence of any Wahlund effect or of any deviation from the random mating assumption.

L 157 (here and line 187), remove B. S. in the citation of Weir & Cockerham.

Answer: The switch to Zotero fixed that issue.

L159 add “slightly” before “negative”. (not the same order compared to the Fis values obtained in the case of clonal populations).

Answer: We respectfully happen to disagree with Dr Halkett on that matter. The situation that Dr Halkett describes holds for purely clonal small populations, with rather small mutation rates, and rather big dioecious populations. We may use Eq 25 in De Meeûs (2015) and Eq 22 in De Meeûs and Noûs (2023) (De Meeûs T (2015), Infection Genetics and Evolution, 33, 227–232. https://doi.org/10.1016/j.meegid.2015.05.008; De Meeûs T, Noûs C (2023), Peer Community Journal, 3, e51. https://doi.org/10.24072/pcjournal.280). In a purely clonal population of size 2000 with a mutation rate of 0.001, F_{IS}=-0.1111, and a dioecious pangamic population of size 6 with an even sex-ratio will display exactly the same value.

L172 from different origins (and place in brackets traps… cohorts) to simplify the sentence

Answer: Done

L170-176 Concerning this test, I wonder about the effect of the differences in sample size, with “traps” suffering from very low sample size, which can distort estimates (e.g. Barrès et al. 2013). In this case, it is best to apply rarefaction. From Figure 2, it seems it is not the case. What are the mean and range of sample sizes according to the different origins?

Answer: I (TdM) am not familiar with the rarefaction procedure. Nevertheless, I do not think it applies here at all. Indeed, F_{IS} was estimated with Weir and Cockerham's *f*, which is unbiased, i.e. its expectancy is independent of sample size. Only the variance will decrease with sample size. Here, the global sample size is the same from one subsampling strategy to the other, and what Fstat computes is the weighted average across subsamples, which is the best way to account for differences in subsample sizes. Consequently, we do not expect much variations of the variance across subsampling strategies, e.g. the width of 95%CI should not vary very much. This is exactly what we observe in our results (Figure 2).

L182 brackets around the date only for the reference De Meeûs et al.

Answer: Done.

L192 tests were one sided (past tense)

Answer: Done.

L233 here and line : add the references (Fox). Not obvious that you refer to the package R commander when reading quickly.

Answer: Done.

L243 choose between “measured” or “tested”

Answer: In fact there was a typo and a "and" was missing.

L260 what do you mean by infinite. If the distribution is skew with very large value, it would be more accurate to report the median value rather than the mean.

Answer: "Infinite" is an output of the different softwares to specify that the algorithm could not converge to a finite value. It most of the time means "very large". We have replaced "results" with "outputs" to make it clearer. Regarding the average, as we explained this at the end of this paragraph, we computed the averages weighted with the number of usable values, and there is not a simple way to compute weighted medians for the averages and minimum and maximum values. The rationale behind this strategy was given in the reference cited (De Meeûs & Noûs, 2023).

L303 replace TdM by De Meeûs.

Answer: Done.

L357-358 not clear to me. Loci B3 and pGp24 are two outlier loci (Fis value not explained by null alleles, Figure 4). Please provide more details.

Answer: We have provided some explanations in this section. We already explained that an excess of missing data does not make these loci outliers for the null allele explanation, as they display more blank genotypes than expected (i.e. more than needed). These loci are outliers for the regression F_{IS}~N_b. This implied that we did not recoded missing data as null homozygotes for these three loci. After doing so, the regression with null allele frequencies as estimated by FreeNA F_{IS}~p_n was good and loci pGp24, B3 and C102 were not outliers anymore.

L399: I wonder whether the larger population size of cohort 10 does not reflect an “outbreak” of tsetse flies (doubling compared with other estimates) which would have motivated the control.

Answer: This was the first time we computed these quantities in this focus. Consequently, there was no way this would have triggered the VCC. After the preliminary and exploratory survey undertaken in the East side (left bank) of the Pongo River, in 2009, and on which nothing was ever published, a more global survey was undertaken on both sides of the river in 2011 (published in Kagbadouno et al., 2012). Then, a VCC was triggered on the East side and not on the West side, for comparison, with surveys undertaken in 2012, and 2013 (Courtin et al., 2015). Then VCC was continued and extended to the West side, with surveillance studies in 2016, 2017, 2018 (Camara et al 2021), 2019 and 2020 (present study). Comparison between Cohort 0 and 10 can only be done with Hetrozygote excess, Coancestries and Sibship methods. And even if the three values are higher in C10 than in C0, the resulting *p*-value cannot be significant (0.25). It is nevertheless more reasonable to interpret this as the effect of sampling variance. Indeed, it is quite unlikely that, as brutal it could have been, an increase in the size of the population would present a significant signature in its effective population size after only 10 generations. If so, it is then hard to understand why no signature of any bottleneck was observed at generation 66. We have added such considerations in the "Results" and "Discussion" sections. Following Dr Halkett's remark, we have added a table (new Table 2), at the beginning of the Results section, which provides the evolution of the number of flies captured and the apparent density of flies per trap and day. One cannot see any real difference between 2009 and 2011.

L400: I don’t understand why it would not be possible to perform comparisons with the temporal method. You have estimates for each pair of cohorts, so you can compare pairs of cohorts before and after the onset of the control, no?
More generally, I see a discrepancy between the estimates of population size and the variation in pairwise Fst that you present in figure 5. You obtained slightly negative value for all pairs of population that includes cohort 67. Why this signal in population structure does not translate into different population size estimates? For me, it's a puzzling result that deserves to be discussed.

Answer: We have added graphics that display effective population sizes for each method and each cohort or between each cohort (temporal methods). Single sample methods do not really allow any significant testing of difference between cohorts, especially between 0 and 10. Values with temporal methods, even if variable, were always big, and the differences hard to interpret. It is in line with a very slow genetic drift, leading to an absence of genetic differentiation, whatever the cohort pair considered. This implies that VCC only affects flies locally. Because of large dispersal distances at each generation, tsetse flies from neighboring sites, which are simply subsamples of the total population, with no noticeable genetic differentiation, reinvade the sites that were emptied by the VCC. We have added some comments on these aspects in the discussion.

Table 2: consider the median rather than the mean estimates.

Answer: As discussed above for line 260, we prefer computing averages weighted with the number of usable values. As discussed in another paper (De Meeûs & Noûs, 2023: https://doi.org/10.24072/pcjournal.280, the most accurate average N_e (weighted, unweighted, algebraic, harmonic, median…) will require many simulations with different scenarios. Here, this would not change much the results: temporal methods outputted very big values and single sample methods rather small ones.

L485 affect the biodiversity is over conclusive. You only test the effect on tsetse fly populations. Please rephrase.

Answer: We have rephrased this sentence.

Reviewed by anonymous reviewer 1, 28 Sep 2023 01:11
Dear Editor,
The manuscrit entitled ''Population genetics of Glossina palpalis gambiensis in the sleeping sickness focus of Boffa (Guinea) before and after eight years of vector control: no effect of control despite a significant decrease of human exposure to the disease'' presents interesting base-line data on the population genetics of Glossina palpalis Gambiensis in Boffa (Guinea).
The title of this manuscrit clearly reflects the content of the articleand the abstract present the main findings of the study. The methodology is adequate to the main objectives of the study and the details provided in this section are sufficient to allow easy reproduction of the work by others scientists. The differents figures and tables provided by the authors in results section confirm that the statistical analysis are appropriate.
Major Comments.
≠1. Regardless, I remain thirsty regarding the protocols that led to obtaining tsetse DNA. The authors just gave us the microsatellite loci they worked on. Please, for the reproductibility of your work other scientists, it is necessary to provide information even by referencing your previous work.

Answer: We have added a reference describing how tsetse DNA was obtained.

≠2. From the line 85 to 87, you said that ‘All the sites used for the entomological survey, the number of trapped flies and their gender…can be seen in Kagbadouno et al., 2012
Is the data obtained in 2019 and 202 recorded in this article? If no, rephrase the sentence from 85 to 87.

Answer: We have rephrased this paragraph to make things clearer. We hope this will meet Referee 1's satisfaction.

Minor comments.
Line 85 number of captured flies instead of number of trapped flies

Answer: Done.

Line 90 river Rio Pongo, Line 92 River Rio Pongo, Line 104 Rio Pongo River. Please harmonise.

Answer: Done.

Reviewed by anonymous reviewer 2, 27 Sep 2023 17:30
Abstract:

-Adding some quantitative data, such as changes in tsetse fly populations or disease prevalence rates before and after VCC, would strengthen the abstract and provide a clearer picture of the study's outcomes.

Answer: Done.

- The abstract mentions the need to continue vector control measures until HAT is entirely eradicated from the focus area. While this is a logical conclusion, it might be useful to highlight any specific recommendations or implications arising from the study's results.

Answer: We have amended the least sentences of the abstract. We hope we have been able to reach what Referee 2 was expecting.

Introduction:
- The introduction effectively introduces the topic of HAT, its causative agent, and the vector responsible for its transmission. However, it would be beneficial to include some contextual information about the global impact of HAT, especially in the endemic countries mentioned. This could help readers understand the broader significance of the study.
-While the introduction mentions Guinea and the mangrove ecosystem as significant for HAT, providing a brief geographical context of the study area and its relevance in the fight against HAT could enhance the reader's understanding.

Answer: Following Recommender's and other referees' remarks, we have added some more precisions on that matter in the Introduction. We hope that this will meet Referee 2 's satisfaction.

Results:
-The Results section provides a comprehensive analysis
Discussion:
- The discussion effectively summarizes the genetic data and analyses, showing that despite a significant decrease in fly densities after VCC, there is no genetic signature of control. However, the discussion could be enhanced by providing a more detailed interpretation of what this implies for the effectiveness of the VCC and the overall dynamics of the tsetse population.
-While the discussion acknowledges the efficiency of VCC in protecting human populations, it could expand on the implications of these findings for future vector control strategies. For example, should VCC be continued, modified, or supplemented with other approaches?

Answer: We have added several sentences and precisions in the Discussion section. We hope that these amendments will meet Referee 2's satisfaction.

- Please remove the website link from the discussion section.

Answer: We have replaced this web link with two references.