Interactions Hôte-Vecteur-Parasite-Environnement dans les maladies tropicales négligées dues aux Trypanosomatidae, UMR IRD/CIRAD 177

Thierry de Meeûs
DR1 IRD
☎ +33 (0)4 67 59 37 08; +33 (0)6 19 83 52 60; Fax: +33 (0)4 67 59 38 94
thierry.demeeus@ird.fr

Le 11/07/2019

à
Editor in Chief
Peers Community In Evolutionary Biology

Dear Editor,

Please accept the revised version or our manuscript, entitled "Deceptive combined effects of short allele dominance and stuttering: an example with *Ixodes scapularis*, the main vector of Lyme disease in the U.S.A.," for resubmission to Peers Community In Evolutionary Biology.

We have taken into account all referees' comments and highlighted in yellow all material additions in the amended manuscript in the highlighted version of our manuscript, which can be downloaded at http://www.t-de-meeus.fr/Data/DeMeeus-et-al-SAD&StutteringI-scapularisUSA-PCI-EvolBiol-11-07-2019-Highlighted.pdf.

We hope you will now find this article suitable for publication in Peers Community In Evolutionary Biology and we remain at your disposal for any more modifications or questions you may have.

Please note that the raw and cured datasets are available as “supplementary file S1” on my web site at:
http://www.t-de-meeus.fr/Data/ DeMeeus-et-al-SAD&StutteringI-scapularisUSA-PCI-EvolBiol-TableS1

Sincerely,

Thierry De Meeûs

Answers to the Recommander and Referees Round #1

Author's Reply:

Decision

by Aurelien Tellier, 2019-06-11 14:13
Manuscript: 10.1101/622373
Revisions required

Dear authors,
Both reviewers and myself do find the topic of the study and the results to be of interest and relevant. Citing reviewer 1, the quest for interpreting difficult microsatellite data indeed deserves attention. It is thus of special interest to understand the biases which can be introduced during the curation steps of these datasets. If the aims, methods and interpretations are clear, the study would benefit from two major improvements. These would enhance the generality of the paper and its relevance for the wider community.
First, both reviewers point out the lack of theoretical “a priori” expectations in the paper. In comments 1-3, reviewer 1 asks to describe the rationale behind the idea that experimental artifacts should increase LD. Reviewer 2 would like to understand in a more quantitative manner the rationale behind pooling alleles close in size and the effect of the sample size on the results. The latter is important as in the study the authors chose a small sample size, while microsatellites have bene recently applied to much larger datasets (at least on many fungal pathogen species for example). A more thorough comparison with other existing curing methods could be provided. I would suggest as a possible solution to indeed build simulated datasets and apply curing methods revealing the different experimental artifacts. It would thus be possible to reveal general rules and outcomes of applying different curing approaches (including yours), such as changes of basic statistics and LD estimates. The effect of the sample size could also be tested on the same pseudo-observed data by subsampling. This general “theoretical” set-up would allow an in depth discussion of the mechanisms involved and make the article more general in scope. The biological dataset of the tick Ixodes scapularis analyzed here would then be used as an application of these general principles.
If it is not possible to perform such theoretical analysis of the curing of pseudo-observed datasets, several in depth descriptions answering comments of both reviewers should be added to the manuscript.
Second, as reviewer 1 points out (comment 4), most researchers move to other type of markers (GBS, RADseq,…) and it would be helpful to discuss if the effect of curing datasets also apply to those data. As a matter of curiosity, a focus could be on highlighting how population genetics inference combining different types of markers (SSRs, GBS, RADseq) can be affected by curing some markers but not others?
Several minor points are also suggested by the reviewers and need to be addressed for the revision. These include restructuring/reorganizing some parts and providing a flowchart (a schematic description) of the analysis/curing steps (reviewer 2).
I look forward to receive your revised version, and believe that this improved contribution would fit into the scope of PCI Evol Biol and be of general interest to the community.
Best regards.

Reviews

Reviewer 1

Reviewed by anonymous reviewer, 2019-05-29 10:06

Review 622373

Title: Deceptive combined effects of short allele dominance and stuttering: an example with Ixodes scapularis, the main vector of Lyme disease in the U.S.A.

In this paper by de Meeûs et al. the authors use a microsatellite dataset of the tick Ixodes scapularis to test for distortion of population genetic parameters by marker typical problems, i.e. null alleles, short allele dominance and stuttering.

The paper is overall well written, interesting and relevant, but suffers a few structural shortcomings. I found that the methods section in part already includes discussion of the topic,

Response: We apologize, but we did not see what in the Material and Methods section can be considered as "Discussion". The justification for not using immature specimens, might have been considered to be part of a discussion, but in our opinion it is indeed part of Materials and Methods.

and the results and discussion lacks a bit of discussion. I would also prefer to have the results and discussion sections separated.

Response: We apologize again, but we planned this paper as a short note, where the results and discussion sections are usually treated together. In our opinion, discussing each result, one by one, enhances the clarity and logical flow of the text. Nevertheless, if it is mandatory to make modifications into a classical full paper with a separated discussion section we would be happy to comply.

The approach how to address the common microsatellite problems should be provided in more detail in the methods rather than the discussion. Maybe a flowchart could be used to visualize the workflow how to improve microsatellite datasets.

Response: This is a good idea. A flowchart has been added as Figure 6 at the end of the manuscript.

I cannot judge all the statistics in detail, but
they appeared largely sound to me. However, I found the approach to pool alleles close in size a little questionable as one would think that this artificially further reduces heterozygosity.

Response:

Pooling alleles increases homoplasy. Depending on the initial and final number of alleles, everything being equal, it decreases genetic diversity and heterozygosity at the same time, though heterozygosity decreases slightly faster. For F_{IS}, the difference will be invisible and for F_{ST}, it will be weakly visible and only for a very small number of possible alleles and relatively high mutation rates.

Regarding loci affected by stuttering, because heterozygote deficits are due to alleles that are too close in size, pooling such alleles tends to decrease the proportion of misinterpreted homozygous individuals, without affecting significantly the actual (true) *F*-statistics. We have added sentences explaining that pooling alleles could not bias *F*-statistic estimates in the "Results and discussion" section.

I was wondering if large allele dropout and short allele dominance refers to the same thing (I assume so). Maybe this could be clarified.

Response: This was mentioned as "long" instead of "large" line 64 of the manuscript. We have changed this to make it clearer.

Further I would like to see some comparison to other approaches addressing these problems (e.g. Wang et al. 2012 Genetics 192(2): 651–669).

Response: Though we think that dropout has a similar effect as null alleles (e.g. see (Sere et al., 2014), which are very well handled by FreeNA as shown elsewhere (Chapuis and Estoup, 2007; Séré et al., 2017), we have reanalyzed the raw data with MicroDrop as suggested by Referee 1. The results give slightly lower F_{IS} and smaller F_{ST} than the cured dataset and only three significant LD tests. There is, however, no good reason to give up SAD and stuttering correction since we indeed specifically detected these problems. Allelic dropout, as defined in the Wang et al. paper, will have very similar effects to null alleles. As mentioned above, null alleles are very well corrected by FreeNA. The fact that MicroDrop ended with a modified data set that behaved very well, or even better, in terms of F_{IS} and LD, as compared to our cured dataset is very hard to explain. As seen in the new manuscript, it did not fully cure SAD or stuttering and such an issue would require a full simulation project with different scenarios and comparisons of different methods. But this would be out of the scope of the present paper. We have however added a comment on this issue at the end of our manuscript. What counts here is that our specific cures for SAD and stuttering did the job in a satisfactory way.

While the common Microchecker approach is compared, I would like to see some more details on the differences of the approaches as well.

Response: Regarding SAD, the answer is straightforward, MicroChecker almost never detects it. In addition, in my experience (TdM), it never detected it in any of the hundreds of data sets I have analyzed. The regression or correlation methods proposed by De Meeûs et al (De Meeûs et al., 2004) or Manangwa et al (Manangwa et al., 2019) are much more powerful, with a slight advantage for the second. As for stuttering, MicroChecker is the only software that implements stuttering detection. However, it does not propose any cure. Hence, we can think of no comparison that could be made. Finally, regarding null alleles, this matter is already extensively discussed in recent articles (Sere et al., 2014; Séré et al., 2017; De Meeûs, 2018; Manangwa et al., 2019), there is therefore no need to add anything in the present short note.

Maybe it would be possible to test the different approaches using an artificial dataset with known rates of the different problems.

Response: For null alleles, this was already done elsewhere (Sere et al., 2014; Séré et al., 2017). For SAD, modelling is quite complex and many parameters would need to be considered and examined, depending on the shape of allele frequency distribution (De Meeûs et al., 2004). For stuttering, heterozygous samples with 1 repeat difference would have to be recoded as homozygous. In both cases, simulating and recoding data would represent a tedious task that would represent a full paper in itself. This being said, the best simulation example is, naturally, represented by real datasets. Here, we have compelling evidences that amplification problems in *I.*

scapularis are causing SAD, stuttering and null alleles. We also show that the cures proposed result in satisfactory results.

One other larger problem may be the sample sizes which are quite low for population genetic analyses. Maybe this could at least be mentioned. How many populations are included (in terms of population structure)? This again may make a difference for the analyses.

Response: We are sorry but we disagree with this view. As we now mention specifically in the amended manuscript, there are 12 subsamples that contained at least 10 individuals and 5 subsamples that contained at least 20 individuals. With 9 highly polymorphic loci, this is more than enough to undertake good population genetics analyses. The best argument is the numerous and highly significant tests obtained, in particular for LD, known to be weakly powerful (De Meeûs et al., 2009).

Besides I have a few smaller comments and suggestions which I provide in chronological order below.
Line 30: maybe refer to ascertainment bias here.

Response: We have added "ascertainment bias" in the text.

Line 77: I guess this would be more commonly referred to as Hardy-Weinberg proportions?

Response: We have added a few words to refer to Hardy-Weinberg. We just wanted here to restore the true creators of this model from a historical perspective. Castle derived the model 5 years before Weinberg and Hardy. He just generalized it only verbally, while Weinberg derived the generalized model, 6 months before Hardy. It seemed fairer to speak of Castle-Weinberg, since Hardy brought nothing new to the story.

Lie 103f: I do not understand this sentence. Maybe it could be clarified.

Response: We have changed this sentence and added more explanations. We hope the sentence is clearer now.

Line 155: Why and how was this subset chosen? Why were not all used? Here some more details would be useful.

Response: We have added some explanations and comments that, we hope, will meet with Referee 1 approval.

Line 178: Here it says nine markers, later (line 284) you refer to 22 sets. The table also includes more.
What was now actually used? Some more details are needed here and some more structure, which would make it more easy to follow.

Response: We have added some details that should make the situation clearer.

Line 199f: This sentence is odd. Could this be reworded?

Response: We have changed this sentence.

Line 258ff: This all reads like discussion. This needs to be described more in a “methods way” in order to be reproducible.

Response: We are not sure we understand what Referee 1 means here. We have added "to our knowledge". This is not a discussion but the introduction to the methodology we used.

Line 284: See above. How many loci were actually used? What was the proportion of missing data?

Response: We now provide these details.

Line 289: Which data? The whole discussion is a bit confusing and would benefit from some clearer structure. Separating the results and the discussion may help here.

Response: We are sorry, but we did not understand what Referee 1 means here. The raw data are the cured data and are available in supplementary file S1.

Line 308: What is meant by blanks?

Response: This is meant to describe the missing genotypes (now written within brackets in the amended manuscript).

Figure 4: It would be nice to label the scored peaks more clearly.

Response: Done

The discussion actually lacks discussion and barely includes any reference to other studies which had similar aims. The results need to be put in a broader context.

Response: We have now further discussed our results and added some references to provide more perspectives for the reader.

Line 380: I find it difficult to say that one estimate is more accurate than the other. In order to do that the real value would need to be known.

Response: This was demonstrated for null alleles at least, so we mention that in the amended manuscript with the corresponding references.

I hope my comments are of help.

Response: Yes they were, thank you.

Reviewer 2

Reviewed by Eric Petit, 2019-06-08 11:29
Dear Thierry and colleagues,
Your study is an attempt to get the best interpretation out of difficult microsatellite data, a quest that deserves attention. In general, the paper's aims, methods and interpretations are fair and clear. You use a whole set of diagnostic statistics to disentangle experimental (null alleles, SAD, stuttering) from sampling (Whalund) effects that may explain deviations from expected intra- and inter-locus allelic associations, and you actually end up with an FIS value that is close to what you expect given the biology of your biological model (you however do not discuss that such a value would correspond to Ne~8.7). The cured data set still shows odd behaviors that can be linked to both null alleles and Wahlund effects. The main weakness of the paper is that you do not provide any expectation for the effect of experimental errors on LD. I detail this and other comments below.

1. You do not explain to the reader what is the rationale behind the idea that experimental artefacts should increase LD, though this is the main question of your study (lines 93-94). It is difficult to understand if stuttering, on the one hand, and SAD and null alleles, on the other hand, could have similar effects on LD. This may be an interesting question per se (you explain lines 87-88 that it is still an open question), but it is not tackled in the present paper, neither through theoretical arguments, nor by using simulations. The only argument that I could find in the paper is that you observe a decrease in LD after curing the data set. But whether this is a general behavior or specific to your data set is questionable.

Response: We have added more explanations as to why amplification problems may artificially increase LD between loci. In reference to the last sentence of Referee 2, the fact that our results are specific to our data is obvious. Nonetheless, this would formulate the question in the wrong way. When analyzing these data, we were shocked by the prohibitive proportion of locus pairs in significant LD in a dioecious sexually reproducing species. To my (TdM) experience, I have only encountered this once. This was in a population of a tsetse fly where we found that two sibling species were sampled at the same sites, resulting in a massive Wahlund effect. Here, using the recently published criteria, there is no evidence of any Wahlund effect. On the contrary, numerous obvious amplification problems were present. This is where the idea of the influence of these amplification problems on LD came from. To this extent, this is specific to our data. Curing the data dropped the number of locus pairs in significant LD. For us, this is proof that amplification problems contributed to a significant degree of this puzzling result.. We also added the results of the correlation between the number of missing genotypes and the number of heterozygous sites in each individual ticks. As expected, we found a negative and significant correlation.

2. By curing your data set from stuttering and SAD, LD decreases, but there is also a change in the sign of the correlation between NLD and HT, which becomes positive, as expected for the Wahlund effect. Does this mean that the Wahlund effect and experimental errors both increase LD but have different effects on the correlation between NLD and HT? Here again, because expectations are not provided, it is difficult to understand whether there is information in these patterns or not. This adds to the difficulty that both of the mentioned correlations are not significant (as many others in the paper), which may be linked to the relative limited size of your data set.

Response: The correlation between NLD and HT becomes positive, but, without any other clues, this alone cannot represent any evidence for a Wahlund effect. Here the F_{IS}<0, even when locus IS11 (many null alleles) is included and significantly negative without it, which is incompatible with a Wahlund effect. Moreover, no test stays significant after BY correction, which is also not in favor of a Wahlund effect. The lack of significance of the correlations is less due to the limited size of the data set than it is to the weak power of such tests (see (Manangwa et al., 2019)). This is why a single non-significant correlation, cannot be used alone but as an element in a panel of different clues. As discussed above, our data set is not so small.

3. Waples proposed in 2015 (a paper you cite) that the proportionality of LD and the product of locus-specific FST could be used as diagnostic tool for the Wahlund effect. How does this apply to your data? Does it help understand that LD is not homogeneously distributed across loci in your case (lines 317-319)? Is it a relationship that is also influenced by experimental artefacts?

Response: I (TdM) have already discussed the problem of Waple's criterion in a recent paper (De Meeûs, 2018). In Waples' model, the F_{ST} before Wahlund effect must be known from an independent sample and F_{IS} and LD measured in another sample with Wahlund effect. "The knowledge of true F_{ST} is a rare situation that I only know of from a few studies on fairly long-lived mammal species: the Leadbeater's possum in Australia (Waples 2015), and the North Pacific minke whale (Waples 2011). Long-lived organisms indeed allow sampling the same cohort after several years or months so that each locus keeps the same characteristic over time between 2 sampling campaign, one of which displays a Wahlund effect, and the other displays the true population subdivision.". A more detailed discussion can be found in (De Meeûs, 2018). Here, as shown by the very small F_{IS} and the fact that no locus pair stays in significant LD after BY correction, there is no evidence for any Wahlund effect...

4. Curing data sets is common place when working with microsatellites, especially so when starting, as here, with a new set of markers. Very often though, rather than keeping markers that have odd behaviors, only markers that can reliably genotyped are kept (see for instance recommendations by Manangwa et al. 2019). Here, among the nine microsatellites that were genotyped, two (IS15 and IS17) could be under selection (but see comment #3), and another one is showing large frequencies of null alleles (IS11). Nowadays, I expect that most research teams would choose alternative genotyping strategies (microsatellite genotyping by sequencing, SNP genotyping) rather than stay with such problematic data sets, which reduces the scope of the present paper, unless the authors could explain how their protocols could help improve data sets obtained with different kinds of genotyping strategies.

Response:
As Referee 2 will see, following Referee 1's recommendation, we have inserted an additional figure with a workflow of actions and decisions, where removing incurable loci with odd behavior is one of possible outcomes. In Manangwa et al, there was in fact two sibling species in the dataset, one of these species did not or hardly amplified 5 loci that had to be removed from the analyzed data (for this species named Clade A). Therefore, this was not exactly the same problem. Here, we show that the different loci that initially represented a problem could be efficiently cured,

and thus kept for F_{ST} estimations. The apparent non-neutrality of some loci also vanished after being cured. Keeping as many loci as possible is important. Sequencing and SNP still represents expensive alternatives in time, money and expertise, which lies beyond the reach of many laboratories and most of the time at the expense of the sample sizes, especially so for non-model organisms. Three decades ago, microsatellite markers were presented as the most powerful genetic markers (Jarne and Lagoda, 1996), and then researchers began to detect the different problems that can arise and developed different kind of cures. The last kind of detection tools and cures only arose very recently. It is probable that SNP will experience the same fate. Null alleles are known to exist in SNPs (Vignal et al., 2002) and these markers may display more frequent allelic dropouts (Bayerl et al., 2018). Ascertainment bias represent a very serious issue for non-model organisms (Garvin et al., 2010). The number of SNPs needed is at least 200 to compete with microsatellite information (Séré et al., 2017). Given the weak power of LD tests on locus pairs for such kind of markers, detecting a significant amount of significant LD would have required a prohibitive number of genotyped ticks to observe results comparable with those obtained with 9 microsatellites.

5. You do not provide any biological interpretation of your data, though this may help understand whether the cured data set is biologically "plausible". In particular, the results obtained from the cured data set are compatible with a Wahlund effect (correlation between NLD and HT, LD). Is this plausible given the subsamples were sorted according to clade, site and cohort? Is this consistent with a slightly negative FIS?

Response: The biological interpretation was explicitly stated in the conclusion: this tick population is strongly subdivided into small subpopulations with meager immigration. The precise population structure will be the matter of a future article (in preparation). As discussed above, Wahlund effect can be safely dismissed as a possible explanation here. The persistence of 19% of locus pairs in significant LD is probably due to population subdivision (but note that none of those remain significant after BY correction). We have added another source of amplification of extant LD perception: the presence of null alleles at several loci. As explained in the text, amplification problems tend to preferentially affect specific individuals at several loci. This was tested with a correlation approach between the number of missing data and the number of heterozygous sites. We have added these new analyses (of the raw and cured datasets) in the amended manuscript.

Additional comments

6. AL2 and VA2 are missing from Fig. 1

Response: Indeed, some problems apparently arose during data manipulation, leading to the incorrect labelling of some individuals and subsamples in Texas and Wisconsin. We thus entirely reanalyzed the data with the correct subsample labelling. Some values changed a little (or not), but the main results remained unchanged or even reinforced.

7. Lines 228-229: There are 25 subsamples with 5 ticks or more, and 23 subsamples with 6 ticks or more

Response: Due to mislabeling of a few subsamples and individuals from Texas and Wisconsin, these values have changed (see also our answer to the previous remark of Referee 2).

8. Line 361: odd format for the confidence interval

Response: Sorry but we do not see what is odd in this format.

References

Bayerl, H., Kraus, R.H.S., Nowak, C., Foerster, D.W., Fickel, J., Kuehn, R., 2018. Fast and cost-effective single nucleotide polymorphism (SNP) detection in the absence of a reference genome using semideep next-generation Random Amplicon Sequencing (RAMseq). Mol. Ecol. Res. 18, 107-117.

Chapuis, M.P., Estoup, A., 2007. Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. 24, 621-631.

De Meeûs, T., 2018. Revisiting F_{IS}, F_{ST}, Wahlund effects, and Null alleles. J. Hered. 109, 446-456.

De Meeûs, T., Guégan, J.F., Teriokhin, A.T., 2009. MultiTest V.1.2, a program to binomially combine independent tests and performance comparison with other related methods on proportional data. BMC Bioinformatics 10, 443.

De Meeûs, T., Humair, P.F., Grunau, C., Delaye, C., Renaud, F., 2004. Non-Mendelian transmission of alleles at microsatellite loci: an example in *Ixodes ricinus*, the vector of Lyme disease. Int. J. Parasitol. 34, 943-950.

Garvin, M.R., Saitoh, K., Gharrett, A.J., 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. Mol. Ecol. Res. 10, 915-934.

Jarne, P., Lagoda, P.J.L., 1996. Microsatellites, from molecules to populations and back. Trends Ecol Evol 11, 424-429.

Manangwa, O., De Meeûs, T., Grébaut, P., Segard, A., Byamungu, M., Ravel, S., 2019. Detecting Wahlund effects together with amplification problems : cryptic species, null alleles and short allele dominance in Glossina pallidipes populations from Tanzania. Mol. Ecol. Res. 19, 757-772.

Sere, M., Kabore, J., Jamonneau, V., Belem, A.M.G., Ayala, F.J., De Meeûs, T., 2014. Null allele, allelic dropouts or rare sex detection in clonal organisms : simulations and application to real data sets of pathogenic microbes. Parasites and Vectors 7, art. 331.

Séré, M., Thévenon, S., Belem, A.M.G., De Meeûs, T., 2017. Comparison of different genetic distances to test isolation by distance between populations. Heredity 119, 55-63.

Vignal, A., Milan, D., SanCristobal, M., Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34, 275-305.