

RESPONSE LETTER TO THE COMMENTS OF THE RECOMMENDER AND REVIEWERS

Resubmission MS Title:

Inference of the worldwide invasion routes of the pinewood nematode *Bursaphelenchus xylophilus* using approximate Bayesian computation analysis

Available at: <https://doi.org/10.1101/452326>

Authors:

Sophie Mallez

Chantal Castagnone

Eric Lombaert

Philippe Castagnone-Sereno

Thomas Guillemaud

Submitted to: PCI Zoology

Dear Recommender,

We found the reviewer's comments and criticisms extremely helpful and cogent. We apologise for the long response time, but a number of pertinent remarks have motivated us to redo the entire set of ABC analyses from scratch with a new strategy. It has therefore taken some time, but the new approach (i) is simpler, (ii) is more accurate (all genetic units are now equally used) and (iii) takes into account many of the reviewers' comments. Some new results are very interesting: the origin of the population "Japan2" could be clarified, while the most likely source of the population "Portugal" appear to be "Japan1". We have, of course, revised the abstract and the discussion accordingly.

More generally, we have incorporated most of the minor and major suggestions, and feel our manuscript is much improved. We hope that you will now find the MS appropriate for recommendation in PCI Zoology.

Please find below our detailed responses (in italic characters) to your comments as well as those of the two reviewers (in bold characters).

Best wishes,

Eric Lombaert on behalf of all authors

Recommender's comments:

COMMENTS TO THE AUTHOR:

Minor revisions needed

Dear Colleagues,

This is an important contribution to unravel pine wood nematode invasion routes and provide insights to invasion biology and invasion route inference methods.

The reviewers raised several useful comments that I invite you to consider in a revised version for a second round of revision.

Main comment:

For my concern, this manuscript would need further testing of the most striking result, that is the difference observed between your ABC results suggesting multiple invasion from USA and previous and current classical Bayesian and descriptive population genetics results on this model suggesting a single origin. If confirmed it would open the way for further theoretical and methodological development to explain these differences, and biologically, the result that many of the global invasions may originate directly from particular locality in USA would also be very informative regarding invasion biology and quarantine forecasting. But, the scenario suggested by classical population genetics is actually not fully tested globally in your ABC analysis. This is because you use stepwise method to select scenario from local to global and rule out single origin in Japan from the beginning. Yet it could be that cutting in piece of the ABC analysis and stepwise approach lead to a local optimum. This discrepancy then could result from the fact that the scenario corresponding to the classic population genetic inference has just not been considered as a global hypothesis in ABC? To make sure, since this result has important implication both theoretical and practical, I would suggest to consider this 5th global scenario that correspond to the neighbor joining tree in fig 2.

The ABC analyses have been completely redone. Indeed, the observations raised by the reviewers led us to reconsider our approach. The overall methodology remains similar to the previous one on many aspects (stepwise approach, work on the inferred genetic units, consideration of the historical information). To summarize, the main improvements are the following:

- *We have focused exclusively on the powerful ABC Random Forest method.*

- *The analysis is now more linear, with 6 successive steps that are all treated in the same way. In particular, the tests of different prior sets and different time scales are done equally for each of the 6 steps.*
- *As pointed out by reviewer #1, the genetic unit represented by the Jap212 sample was not present in the previous ABC analyses. This omission has been fixed, and this genetic unit can be found under the name "Japan3".*
- *Also pointed out by reviewer #2 was the confusing treatment of the Japan2 genetic unit. The admixture hypothesis is now more clearly based on the patterns of the Structure results. Moreover, this genetic unit is now present in the analyses until the final step.*
- *The simulation of the native area in the previous version of the article was not satisfying. Regarding our incomplete sampling scheme and the strong genetic structure in the USA, we chose to use four American samples to create a "native group". This allows us to no longer rely on the unrealistic hypothesis that we would have sampled the true native source by the greatest coincidence. In addition, this approach allows a significant amount of genetic information to be added to each analysis. The results of the model checking analyses suggest that the choice we made is wise (e.g. Table S3).*

It is important to note that the five first ABC steps make it possible to deal with all the genetic units, and thus to obtain a comprehensive final scenario. The sixth step, on the other hand, represent an additional analysis which allows us to rule out once more the hypothesis of a single unique introduction from the native area. This allows us to further validate what has been obtained, step by step, previously.

A major result of the previous version of the manuscript changed: the population of Portugal is probably originating from Japan (Japan1), and not from the native area. This new result seems convincing to us, but it also underlines the limits of the ABC method in the context of this species with a very low genetic diversity, and strong inter-population structure. Therefore, we have taken more precaution in the discussion by suggesting that a certain degree of caution should be maintained in the interpretation of our overall results (lines 493-501). However, this does not change the observation of discrepancies observed between different methods, and thus the interest in discussing them in detail.

From our point of view, the manuscript as a whole appears much more solid and convincing.

Other minor comments:

Line 30. This unknown origin cannot be found in the result section. In the result we see this second introduction in japan 2 is from USA, NE2 or ghost native.

This point is now solved. The “unknown origin” statements have been removed from the manuscript. Overall, the analyses related to the search of the source of Japan2 have been greatly improved and clarified.

Line 63. The evaluation of the confidence in the scenario choice. May have to clarify. Probably talking about type I error rate. Probability to reject the true scenario.

We were referring to our ability to select the true scenario (measured here using the prior error rate). We modified the text accordingly (lines 64-65).

Line 79. Which year in Spain ?

First observation in Spain was in 2008 (e.g. Fonseca et al. 2012). We added this information to the text (line 81).

Line 208. Prior set 1 and 2. Where are they defined (I figured out in Table S1, see below) ? They should be described here since they are mentioned 23 times in the main text.

The table containing the description of the priors has been moved to the main text (Table 2).

Line 217. Not clear what is “95 % confidence intervals of the posterior probabilities of scenarios” and how it is calculated.

*We have chosen to remove the logistic regression ABC analyses from the manuscript, and to focus solely on the ABC-RF analyses. Therefore, the calculation of the confidence interval is no longer relevant. However, more information on the computation of the probability of a scenario and its confidence interval through a weighted polychotomous logistic regression can be found in Cornuet et al (2008, *Bioinformatics*, 24:2713-2719).*

Line 222. Instead of “wrongly identify scenario” change to “do not give the highest probability to the true scenario”. In order to clarify how the scenario is identified.

The text has been modified (line 228).

Line 393. .. to the same scenario with a unique invasive bridgehead : since it is not present in figure 1, explain that this bridgehead is the same scenario as the previously selected but with a unique bridgehead ghost population between native and invasive populations.

The “unique bridgehead invasion scenario” is explained lines 365-369 and in Table S2.

Line 395. The scenario “selected in the global analysis”. Change to “without bridgehead”. To clarify.

Given the important changes made to the ABC analyses, the text has been significantly modified. We believe that the summary of the main results in the text is clearer than it was in the previous version of the manuscript (lines 353-369).

Line 440. Figure 4. Why not representing in this figure the most probable scenario in Japan with multiple introductions in Japan. And in China from USA ? Also the result suggests all the invasion come from a particular USA pop.

Given the new results, Figure 4 has been significantly modified. All geographical routes are now represented (Figure 3B). We have also incorporated a graphical representation of the final scenario, so as to provide more information about the demographic events at play (Figure 3A).

Regarding the "USA" origin of the invasive populations, our data do not allow us to conclude on a particular geographical location within the native area. The NE2 sample can only be considered as the "closest sample we have", but the native area has not been exhaustively covered, and given the high genetic structure observed in this species, it is unlikely that this sample is, by chance, the precise location of the native source populations. As explained above, this is the reason why we have modified the way we simulate the native area in the new ABC analyses.

Line 446. two events of introduction ... in Japan (one from the USA and one with an unknown origin). Please clarify this unknown origin. It is not mentioned in the results anywhere among the different scenarios.

The new ABC analyses have enabled us to shed light on the likely origin of this Japanese population (i.e. Japan2). There is therefore no longer any mention anywhere in the text of an unknown origin of any population.

Table s1. Not clear what is set 1 and set 2. Does it refer to Uniform vs log uniform shape distribution? Why is there set 1 and set 2 as well in the interval if the interval is the same apparently for both ?

We have entirely updated the table presenting the priors, and incorporated it into the main text (Table 2). Moreover, we used a new alternative prior set, because the previous one generated summary statistics much too different from the real dataset. Consequently, the shapes of the distributions no longer change, whereas the ranges do.

Reviewer #1: Aude Gilabert's comments

The invasion history of the pinewood nematode in Europe and Asia remaining unresolved through the use of descriptive population genetics analyses, Mallez et al. undertook approximate Bayesian computation analyses to clarify it. To this aim, the authors analysed samples from the USA representing the native area, and for the invaded area from Japan, China and Portugal. Most of the samples analysed in this paper had already been analysed in a previous paper from the same authors, which had the same objective (Mallez et al. 2015): of the 48 sites (consisting of 1080 individuals), 14 locations were new representing 310 individuals (< 1/3). It should be stated that the study was to complete the sampling of a previous study, and to use a methodology more adapted here. Since you add new locations, it is a pity not to have samples from Canada and western USA, especially as it seems the origin of the invasive populations has not been captured here.

The link between Mallez et al (2015) and the present study has been better highlighted in the introduction section (lines 91-93).

About the Canada/Western USA samples, we agree with Reviewer #1 that it is a pity that we don't have some. It should be noted, however, that the new version of the discussion highlights more clearly the weaknesses of the sampling scheme, and their consequences (lines 498-501).

Mallez et al. identified multiple introductions into Japan from USA and an unknown origin, and into China from USA and Japan. Authors claim that the Portuguese populations originated from America. Maybe this is due to misinterpretations/mistakes from my part but I do have issues/interrogations on the ABC analyses, the steps of the procedure and the choice of some models, which I will detail below. The authors also observed discrepancies between the different methods they used, and a lack of power. They hypothesized it could be due to the extremely low level of genetic diversity within the pinewood nematode populations. This can be tested using simulations, as they mention it lines 516-522. I agree and think it should be included in the study to strengthen the results, and justify the paper. Implementing such simulations to test the effect of a low level of genetic diversity on ABC model choice analyses would be a very good addition to the paper to increase value to it.

The ABC analyses have been thoroughly redone, in order to take into account some important comments of this review (see above and below).

The idea of doing a comprehensive simulation work to better assess the power of ABC in the PWN context is appealing. However, we chose not to do it for several reasons. First of all, the work of Lombaert et al (2018, Heredity) already provides a lot of information about the role of low genetic diversity in the potential biases of the method implemented in Structure. We have added details on this matter in the manuscript (lines 486-489). Second, carrying out a simulation work to evaluate and

quantify the effect of low level of genetic diversity on ABC model choice is far from a trivial work, and we believe that it warrants a dedicated study, and a methodologically oriented publication. We therefore chose to keep this idea as a perspective in the discussion (lines 502-509). In the context of this manuscript on the PWN, the PERs (which are based on simulations) provide some, albeit incomplete, elements of answer. We have also modified the discussion to suggest a more cautious interpretation of the results (lines 493-501).

Material & methods:

Sampling & genotyping: I found the sampling and genotyping section too succinct, although I could find some answers in the previous paper by Mallez et al. 2015.

First, the previously analysed and new sites should be indicated more precisely. It would be nice to have a map of the sampling locations, to get an idea of the sampling scales within each area.

The samples that were already used in Mallez et al. (2015) are now highlighted in Table 1 with the symbol “§”. Maps showing the location of each sample are now available as supplementary information (Figure S1).

I could not find the years of the samplings. Have the samples been collected at the same or different time? Could you give the information, for example in the table 1 if there are different dates of collection or in the text?

The sampling years have been added to the text (lines 111-112)

The sampling scheme should be a little bit more detailed, or at least should refer to Mallez et al. (2015; however the access to the paper does not seem to be free, so I would suggest to give details here). If I understood correctly from Mallez et al. 2015, one site sample corresponds to individuals collected from one tree. Has anyone already studied the pinewood nematode population genetics at a local scale using a hierarchical sampling to see how much of the diversity is captured in a single tree relatively to a local scale?

A precision has been added to the text (lines 108-109).

Mallez et al (2013) provide information about the distribution of the genetic variability at various geographical scales. They found that an important part of the genetic variance detected corresponded to inter-individual variation within a tree, where sexual reproduction appears to take place, and that spatial differentiation can exist at very short scale, with neighboring trees significantly differentiated.

Lines 208-214: I understand from the results section that the influence of the number of generations per year has been tested for the analysis on the origin of the European invasion (with the Japan unit 1) only. If true, it should be indicated here.

We have made the necessary clarifications in the text (lines 220-225). It was true, but it isn't anymore. We are now testing different generation times in all steps of the ABC analyses.

Evolutionary relationships between the different site samples:

Both FST and DEST are computed to estimate population differentiation. If you use the two measures, the explanation on why and the interpretation of both measures should be given. Otherwise one cannot understand why both measures are considered.

We chose to present both Fst and Dest for the following reasons: From one side, Fsts are the classical statistics used in empirical population structure studies. We think that readers should know that PWN samples display extraordinarily high values of Fsts. From the other side, Dest is shown here because they may be considered more informative as stand-alone statistics in the PWN context. This is because they are less sensitive to a low level of intra-sample variability. The two measures of differentiation (Fst and Dest) are presented because they do not always lead to the same conclusions and do not provide information on the same processes (Alcala et al. 2014; Gerlach et al. 2010; Jost 2009). We added this information with new references in the text (lines 153-157).

Structure analyses: I would first talk about the global analysis, with all the samples considered, and then present the analyses at the country scale. I think it is easier to visualize like this. What about an analysis that would include all the invasive populations to explore relationships between those populations? Is there any specific reason for which it has not been performed? Or eventually, the analysis of the American samples could be used to select a few sites representatives of the America to include in an analysis with the invasive samples?

The order in which the STRUCTURE results are presented (i.e. either beginning or ending with the overall analysis) is mainly a matter of personal judgement. From our point of view, ending with the global analysis seems more relevant, as it is the only one that can give indications on the invasion routes, which corresponds to the general objective of our study. We have therefore chosen not to change the order in which our results are reported.

An analysis that would include all the invasive populations, without the native ones, appears to us to be irrelevant. Indeed, for a given invasive population, the other geographical areas do not present any theoretical distinction regarding whether they correspond to a native or invasive area: all are potential sources. Therefore, the intra-continental approach (within-Japan, within-China, within-

Europe, etc.) on the one hand, and the global approach (with all available samples, whatever their invasive or native nature) on the other hand, seem to us sufficient to answer our questions.

Regarding the analysis of U.S. samples, there are indeed many ways to address how they can be included and studied in the analyses. In our new ABC analyses, we have chosen to use several native samples (NE2, NE5, NE9 and NE22) simultaneously to anchor the native area. This allows us to no longer rely on the hypothesis (unrealistic given the very strong genetic structure) that we would have sampled the true native source by the greatest coincidence. This choice allows us to simulate very realistic data with respect to what we observe (see model checking results, Table S3).

Lombaert et al. (2014) recommend using several clustering methods to define the genetic clusters in a sampling. Especially given the HW disequilibrium observed in many populations, don't you think that it would be a good idea to use at least one complementary clustering method?

This is indeed an interesting idea. However, the extremely strong genetic structuring of this species makes some clustering methods uninformative. For example, we tested the BAPS software (Corander et al. 2003, Genetics, 163:367-374) on our data (results not shown): unlike STRUCTURE, BAPS uses the number of cluster K as a parameter, and the result appears meaningless since each sample (with a few rare exceptions) is found to belong to its own private cluster. Furthermore, we believe that the combined information from the STRUCTURE analyses on the one hand, and the other analyses (e.g. the neighbor joining tree) on the other hand, constitutes a satisfactory generic approach.

ABC analyses: I am not sure to understand the step-by-step ABC analysis, and the choice of some of the models. The choice of the scenarios to test should be more detailed and explained. Based on the historical records, the first analysis should focus on Japan, and then the Chinese genetic units should be included in the model, and then the Portuguese one. Why China is first treated independently (a detail here: Japan is the first introduction so Fig. S1 should be the figure of the Japan scenarios and Fig. S2 that of the China ones...)? Then, I don't understand the different scenarios considered in Fig.S3 and S4 (why those scenarios, and why in that stepwise manner)... For an easier reading, it could be helpful to have a more precise description of the scenarios tested. There is a hypothesis for admixture for the Japanese genetic unit 2 in Fig.S4, which comes out of the blue: unless I missed it, it has never been mentioned anywhere. I guess it comes from the results of the Structure analyses. However, if we rely on historical records, which is done here and allow the stepwise procedure, Japan was first invaded, then China. I may have missed something and/or be wrong but could it be the Japanese unit that originated from an admixture event (e.g. between a native population and a Japanese unit close to Kosa –from Fig.3D in Mallez et al. 2015), and then the Chinese genetic unit 2 originated from that Japanese unit...? Why only two Japanese genetic units are considered, and what

are these two units? The figures 2 and 3 in this paper and figures 2B and S1B seem to indicate at least 3 genetic units (Jap212, 120 and 308 / Kosa, Kasuming 2 and 5 / Kasuming3), don't they? From the analysis of the whole dataset (Fig.3 and lines 318-327), there is one cluster grouping together Portuguese samples, the Japanese sample Jap212 and Laoshan. Although Jap212 and Laoshan seem to belong to the same genetic cluster, and therefore the scenario 4 from Fig.1 could be a test of a Japan origin from the Iwate area, don't you think it could be worth testing a Japanese origin considering Jap212 for the Japanese sample? Also, there is no hypothesis of the Portuguese samples originating from an admixture event. Don't you think it could be possible?

As explain above, the ABC analyses have been completely changed and redone. As a result, the text has been substantially modified, and we believe that our new approach, which is more linear than before, is a satisfactory response to most, if not all, of your comments. For example, we now consider three Japanese genetic units, with Jap212 becoming the representative sample of Japan3. In addition, the treatment of Japan2 is less confusing: we specify that the admixture hypothesis is indeed based on the Structure results (line 360), and this genetic unit is now present in the analyses until the final step. You will find a more complete description of these new analyses above. In addition, you will find new illustrations in the manuscript (Figure 3A, Figure S2), and all the new ABC configuration files are available at <https://doi.org/10.5281/zenodo.4604958>

About an admixed origin of the Portuguese sample, we do not see any reason to test this hypothesis: the Structure results show no signals of admixture, and the very low diversity of this genetic unit further suggests that it is highly unlikely.

To come back to the methods, lines 229-231: the F_{ST} and $d\mu^2$ distances between pairs of populations have not been considered in the ABC analyses but in the quality checking is that right? If it is, why so?

We have chosen to remove the logistic regression ABC analyses from the manuscript, and to focus solely on the ABC-RF analyses, using the software DIYABC Random Forest v1 (Collin et al., Manuscript under revision for Molecular Ecology resources). Because ABC-RF is more robust than polychotomous logistic regression to the choice and number of summary statistics, we now use all those available, as described lines 204-212. The same statistics are used in the model checking analyses.

Results:

Table S2: This is a very large table, difficult to read. Could you use a gradient colour to fill the cells according to their value?

The Table (Table S1) has been modified as proposed.

Even if it has been done in the previous analyses, I think it would be nice to still add a figure or at least mention the results of the structure analysis of the Japanese samples. I would not insist too much on the results of Structure. I am not convinced by the consistency between runs.

The results of the STRUCTURE analyses of the Japanese and European samples (initially shown in Mallez et al., 2015) have been added to the supplementary file (Figure S5).

Figure 2: would it be possible to add to the colour code the results of the genetic clustering using Structure, i.e. to use two colours for the American samples, 2 for the Chinese samples, and 2 if 2 it is for the Japanese ones?

We believe that colors by geographical area is simpler to apprehend than by genetic units on this figure (Figure 1). We think that many readers will prefer at this point to be able to visualize at a glance the geographical origin of the samples on this tree. However, and as suggested below, we have added an asterisk for each of the population sample that is used in the ABC analyses. This enables a link to be made between this figure, and the genetic units described later in the manuscript.

Lines 343-349: maybe I missed it but I did not see a clear description of the genetic units considered in the ABC analyses. It could help to add the list of samples each genetic unit is the representative of. You could also highlight the samples used in the ABC analyses as representatives of their genetic cluster in the figures 2 and/or 3, for example using an asterisk or an arrow. That would help the understanding.

The genetic units are listed lines 342-349. As suggested, we added to both Figure 1 and Figure 2 an asterisk for each of the population sample that is used in the ABC analyses.

I will not discuss the results of ABC analyses too much. Do you have an explanation for the different results you obtained according to the prior set for the analyses between China 2 and Japan 2? Are the dataset correctly simulated using both prior sets?

The results obtained with the new ABC analyses are much more consistent, including when we look at the link between Japan2 and China2 (see step 4, Table S2). For information, the “prior set 2” used in the previous version of the manuscript was indeed unable to simulate datasets close to the real one (i.e. very poor model checking results).

Tables 2/3 and others: for the “Analysis and scenario” column, highlight in bold the selected scenario rather than the prior set/nb of generations per year considered. I think you could reorganize the tables for clarity/easiest visualization. For example for the table 2, you could re-organize the results

according to the prior set by columns and not lines, and include the mention of the number of generations in the legend rather than in the table for each analysis since there is only one option considered here. There is no need to indicate in the legend the method for the posterior probabilities, it's written in the material & methods section.

Since the ABC analyses have been completely revised and rerun, the tables have significantly changed. We took this opportunity to reduce their number and to simplify them, taking into account the suggestions (Table 3 and Table S2).

Lines 382: all invaded areas, except Japan2.

Japan2 (as well as Japan3) is now included in the ABC analyses up to the end.

Discussion:

Even the closest native samples appeared to be quite differentiated from the invaded ones. Don't you think that it could also be a reason of the disagreement between the analyses and of the "mixed" results?

We agree with this hypothesis. Because of the very strong genetic structure in the USA, we have changed the way we simulate the native area. Rather than using a single sample, we now use four American samples to create a "native group". This allows us to no longer rely on the unrealistic hypothesis that we would have sampled the true native source by the greatest coincidence. As explained above, this approach allows a significant amount of genetic information to be added to each analysis, and the results of the model checking analyses suggest that the choice we made is wise (e.g. Table S3).

In this manuscript, the authors want to investigate the invasion routes of the pinewood nematode from North America to Japan, China and Europe (Portugal). They obtained microsatellite markers over large sample sizes representing these different populations. Based on classic descriptive summary statistics and ABC analysis, they could infer that there have been three different invasion from North America: 1) to Japan, 2) to China and 3) to Portugal. A second invasion occurred to Japan from an unknown population, and a secondary invasion occurred from Japan to China. The manuscript is well written, the descriptive analyses are well performed and correctly interpreted. The ABC analyses are sound and adequate and well described and reproducible. The authors find similar outcome using the classic ABC and the Random Forest ABC. The results are clear and well explained. The interpretations are accurate and follow well the results, and the discussion is well organized and clear. The discussion has some interesting points about the lack of power of the ABC under some invasion scenarios. Overall it is a very nice contribution and an interesting read which is of general interest for readers interested in population genetics analyses of invasive species. I have only a few minor comments for improving clarity.

1) Introduction: line 69-70, it would be useful to know if this nematode also reproduce asexually or what is the frequency of sexual/asexual reproduction or selfing. Are two sexes needed? This would influence the diversity during invasion bottlenecks and the establishment of the new populations? Does this nematode produce large variance in offspring production (if this is known, see theory by Wapples). For some nematodes parasite of plants this has been shown (Montarry et al. Proc Roy Soc B 2019) and this has a strong influence on the rate of genetic drift.

Unlike some other nematodes, Bursaphelenchus xylophilus reproduces exclusively sexually. This information, with a reference (Futai 2013), was added to the text (line 74).

Females of PWN only produce a low number of offspring (80 to 150) during the primary transmission phase, so the hypothesis of a large variance in offspring production is not justified for this species.

2) Methods: lines 141-146, would the Fst per population as computed in Weir and Goudet (Genetics 2017) help to gain information on population differentiation?

A simple and maybe naïve answer is that we do not know. Many estimates of fixation index exist and we were not aware of this specific one. It is noteworthy that we did not use estimates Fst as stand-alone statistics to infer population histories (NJ tree based on Cavalli-Sforza and Edwards distances + Structure analyses + ABC-RF analyses). They were just displayed as descriptive statistics to give a rough

information about genetic differentiation among samples to readers. They were also used in the ABC-RF analyses but it is important to note that the current ABC-RF package does not allow using the Weir and Goudet's estimate of F_{st} .

3) Methods line 229: It is unclear why the authors did not include F_{st} , Jost D or DeltamuSquare as summary statistics in the first ABC but some were then used in the random forest ABC (if I understand correctly)?

As detailed above, we have thoroughly modified the ABC analyses. We have taken this opportunity to focus exclusively on the RF methodology, so we now use all the statistics available in the software DIYABC Random Forest v1.0. See lines 204-212 for details.

4) Methods: in a paper recommended in PCI, de Meus et al. (2020) have studied how to tackle the issue of missing data and null alleles, maybe it is useful to cite this and check if their approach is similar to that used here?

In our study, we use the established method proposed by Chapuis & Estoup (2007; cited almost 2500 times on February 1st 2021) to take into account the problems related to null alleles in the computation of F_{ST} and Cavalli-Sforza and Edwards distances. The method described in De Meeûs et al. (2020) is not the same, so we have no reason to cite this paper here.

5) Format references lines 402-403: remove "J.M." in front of Cornuet.

Table 3 has been removed, so this typo no longer appears.

6) In Figure 4, one could even be more precise and represent the origin of the invasions from Nebraska (versus other US populations).

The former Figure 4, now called Figure 3B, has been reworked in light of the results of the new ABC analyses. Regarding the native origin of the outbreaks, it is not realistic to consider that we are in possession of samples from the exact true sources, when our sampling scheme is incomplete and the genetic structure is very strong. Our new ABC analyses take explicitly into account this uncertainty (see above), and the new simulated data are close to the real data (Table S3). See also lines 498-501.

7) Discussion: lines 454-464. I find the argument somehow convoluted and not fully convincing. It is indeed possible that the observed pattern results from the effect described here (from Guillemaud et al. 2010), while this has not been explicitly tested. Could one imagine more complex scenarios also accounting for this pattern: e.g. several unsampled source populations in America which present gene flow/admixture with one another, or there has been several introductions within a short

amount of time from several geographically close populations in Nebraska (unsampled populations). Would the conclusion change if Canadian samples would be available? A solution to this lack of resolution would probably be to resolve the spatial structure of populations (and past demographic history) in the ancestral range at a greater resolution.

This comment is relevant, but unfortunately, it is difficult to address it in a very satisfying way. Simulating a native area is always a challenge, because the observed genetic diversity is the result of a long and often complex evolutionary history. Nevertheless, it is clear to us that the choices we made in the previous version of the manuscript was not satisfying. In the new ABC analyses, we chose to get closer to your suggestion by using four American samples to create a "native group", and ghost populations are used in a more consistent way. The results of the model checking analyses are satisfactory, suggesting that we are not getting too far away from the reality (e.g. Table S3).

8) Discussion, lines 486-512: Discrepancy of ABC with the Structure results. Could this also be due to sexual/asexual reproduction and/or the effect of genetic drift (as the HWE does not account for it)? Using ABC and structure in domesticated crops (rye in Parat et al. Mol Ecol 2016) the effect of bottlenecks in crop domestication did not create a discrepancy and structure could be used in a sequential manner consistently with the ABC results. Therefore it seems unlikely that bottlenecks alone generate such discrepancy. Could we interpret this discrepancy by the existence of high rates of genetic drift in nematodes due to large variance in offspring production and/or difference in sexual/asexual reproduction across the ancestral and derived populations? Or is there cryptic structuring in the ancestral populations which influences the composition of the invasive populations? In the case of large variance in offspring production, it would be interesting to compare known census sizes compared to estimates of N_e (see work by Wapples, Montarry) to test this hypothesis.

The discrepancies between ABC and Structure results is thoroughly discussed in the manuscript (lines 464-492). We do not agree with reviewer #2 when he states "it seems unlikely that bottlenecks alone generate such discrepancy". Lombaert et al. (2018, Heredity) have shown that strong bottleneck, but also (and most of all) low genetic diversity in source population could lead to misleading results with Structure. The very low diversity in the native area of the PWN, together with founder effects, is very likely to result in those discrepancies.

The alternative hypotheses proposed in this comment are interesting and it should be interesting to test them in the future. However we are not comfortable to discuss them in the present article: although they are plausible, we currently have no elements that tend to agree with them in the case of PWN (for instance the PWN is a strictly sexual species - we added this information to the text, line 74).