Extracting the maximum historical information on pine wood nematode worldwide invasion from genetic data

Stéphane Dupas (iD) *based on peer reviews by* **Aude Gilabert** *and 1 anonymous reviewer*

Submitted: 15 September 2020, Recommended: 28 April 2021

Cite this recommendation as:

Dupas, S. (2021) Extracting the maximum historical information on pine wood nematode worldwide invasion from genetic data. *Peer Community in Zoology*, 100006. 10.24072/pci.zool.100006

Published: 28 April 2021

Redistribution of domesticated and non domesticated species by humans profoundly affected earth biogeography and in return human activities. This process accelerated exponentially since human expansion out of Africa, leading to the modern global, highly connected and homogenized, agriculture and trade system (Mack et al. 2000, Jaksic and Castro 2021), that threatens biological diversity and genetic resources. To accompany quarantine and control effort, the reconstruction of invasion routes provides valuable information that help identifying critical nodes and edges in the global networks (Estoup and Guillemaud 2010, Cristescu 2015). Historical records and genetic markers are the two major sources of information of this corpus of knowledge on Anthropocene historical phylogeography. With the advances of molecular genetics tools, the genealogy of these introductions events could be revisited and empowered. Due to their idiosyncrasy and intimate association with the contingency of human trades and activities, understanding the invasion and domestication routes require particular statistical tools (Fraimout et al. 2017).

Because it encompasses all these theoretical, ecological and economical implications, I am pleased to recommend the readers of PCI Zoology this article by Mallez et al. (2021) on pine wood nematode invasion route inference from genetic markers using Approximate Bayesian Computation (ABC) methods.

Economically and ecologically, this pest, is responsible for killing millions of pines worldwide each year. The results show these damages and the global genetic patterns are due to few events of successful introductions. The authors consider that this low probability of introductions success reinforces the idea that quarantine

measures are efficient. This is illustrated in Europe where the pine-worm has been quarantined successfully in the Iberian Peninsula since 1999. Another relevant conclusion is that hybridization between invasive populations have not been observed and implied in the invasion process. Finally the present study reinforced the role of Asiatic bridgehead populations in invasion process including in Europe.

Methodologically, for the first time, ABC was applied to this species. A total of 310 individual sequences were added to the Mallez et al. (2015) microsatellite dataset. Fraimoult et al. (2017) showed the interest to apply random forest to improve scenario selection in ABC framework. This method, implemented in the DiYABC software (Collin et al. 2020) for invasion route scenario selection allows to handle more complex scenario alternatives and was used in this study. In this article by Mallez et al. (2021), you will also find a clear illustration of the step-by-step approach to select scenario using ABC techniques (Lombaert et al. 2014). The rationale is to reduce number of scenario to be tested by assuming that most recent invasions cannot be the source of the most ancient invasions and to use posterior results on most ancient routes as prior hypothesis to distinguish following invasions. The other simplification is to perform classical population genetic analysis to characterize genetic units and representative populations prior to invasion routes scenarios selection by ABC.

Yet, even when using the most advanced Bayesian inference methods, it is recognized by the authors that the method can be pushed to its statistical power limits. The method is appropriate when population show strong inter-population genetic structure. But the high number of differentiated populations in native area can be problematic since it is generally associated to incomplete sampling scheme. The hypothesis of ghost populations source allowed to bypass this difficulty, but the authors consider simulation studies are needed to assess the joint effect of genetic diversity and number of genetic markers on the inference results in such situation. Also the need to use a stepwise approach to reduce the number of scenario to test has to be considered with caution. Scenarios that are not selected but have non negligible posterior, cannot be ruled out in the constitution of next step scenarios hypotheses.

Due to its interest to understand this major facet of Anthropocene, reconstruction of invasion routes should be more considered as a guide to damper biological homogenization process.

References:

Collin, F.-D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., Marin, J.-M. and Estoup, A. (2020) Extending Approximate Bayesian Computation with Supervised Machine Learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. Authorea. doi: https://doi.org/10.22541/au.159480722.26357192

Cristescu, M.E. (2015) Genetic reconstructions of invasion history. Molecular Ecology, 24, 2212–2225. doi: https://doi.org/10.1111/mec.13117

Estoup, A. and Guillemaud, T., (2010) Reconstructing routes of invasion using genetic data: Why, how and so what? Molecular Ecology, 9, 4113-4130. doi: https://doi.org/10.1111/j.1365-294X.2010.04773.x

Fraimout, A., Debat, V., Fellous, S., Hufbauer, R.A., Foucaud, J., Pudlo, P., Marin, J.M., Price, D.K., Cattel, J., Chen, X., Deprá, M., Duyck, P.F., Guedot, C., Kenis, M., Kimura, M.T., Loeb, G., Loiseau, A., Martinez-Sañudo, I., Pascual, M., Richmond, M.P., Shearer, P., Singh, N., Tamura, K., Xuéreb, A., Zhang, J., Estoup, A. and Nielsen, R. (2017) Deciphering the routes of invasion of Drosophila suzukii by Means of ABC Random Forest. Molecular Biology and Evolution, 34, 980-996. doi: https://doi.org/10.1093/molbev/msx050

Jaksic, F.M. and Castro, S.A. (2021). Biological Invasions in the Anthropocene, in: Jaksic, F.M., Castro, S.A. (Eds.), Biological Invasions in the South American Anthropocene: Global Causes and Local Impacts. Springer International Publishing, Cham, pp. 19-47. doi: https://doi.org/10.1007/978-3-030-56379-0_2

Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., Loomans, A., Malausa, T., Nedved, O., Rhule, E., Staverlokk, A., Steenberg, T. and Estoup, A. (2014) Complementarity of statistical treatments to reconstruct worldwide routes of invasion: The case of the Asian ladybird Harmonia axyridis. Molecular Ecology, 23, 5979-5997. doi: https://doi.org/10.1111/mec.12989

Mack, R.N., Simberloff, D., Lonsdale, M.W., Evans, H., Clout, M., Bazzaz, F.A. (2000) Biotic Invasions : Causes , Epidemiology , Global Consequences , and Control. Ecological Applications, 10, 689-710. doi: https://doi.org/10.1890/1051-0761(2000)010[0689:BICEGC]2.0.CO;2

Mallez, S., Castagnone, C., Lombaert, E., Castagnone-Sereno, P. and Guillemaud, T. (2021) Inference of the worldwide invasion routes of the pinewood nematode Bursaphelenchus xylophilus using approximate Bayesian computation analysis. bioRxiv, 452326, ver. 6 peer-reviewed and recommended by Peer community in Zoology. doi: https://doi.org/10.1101/452326

Reviews

Evaluation round #2

Authors' reply, 22 April 2021

Download author's reply

Decision by **Stéphane Dupas** ⓘ, posted 01 April 2021

Evaluate one versus multiple bottleneck sources in the native area

I would like to thank the authors for the quality of the new analyses performed. A lot of work responding to almost all reviewer issues, and many improvements in the analysis have been included in this second version.

I would have only one issue to raise before recommendation. In the first version, I had pointed out the interest of the result that, based on the analyses not only there would have been several independent introductions from the native area in USA to the different regions in Asia, but that these introductions would come from the same one local population in the USA. The authors changed the analysis by including ghost populations between the USA source and the invasive populations. This is justified since the probability to have sampled the very population source is low. Yet, this removes from the present version the single population source results that had come out, since the authors group all the USA population and perform several independent bottlenecks after, to generate different introductions. Since there is a strong population structure in the USA, this result of a single pop source certainly had strong signals. In addition it has implications for quarantine efforts on this population, and for biological invasion theory in general. It may thus be of interest to address again the issue of a single versus multiple source populations in the USA. This can be easily combined with the ghost source hypothesis by performing a single bottleneck on the joint USA pop before the introductions (single pop hypothesis) versus multiple bottlenecks before the introductions (multiple pop hypothesis).

This request is very minor but might improve the impact of the paper that makes use of the most recent inference technologies to unravel invasion routes from population genetics data, and might be seminal for many students and researchers.

In addition some minor comments :

- remove "of" line 38

- Line 206. Number of independent introduction events instead of "independently" later.

- Line 251 not clear what is main and alternative in the text. Although it's clear on the table.

- Line 427 May have derived

- Line 528 Change "and being" by "which was"

- Line 628 :   "For instance, if the native area is weakly diversified so that it exhibits a few very frequent alleles, it is probable that two independent introductions from this native area (native ⧠ invasive 1 and native ⧠ invasive 2) lead to samples closer to each other than to their native area."

In the present case the native area is diversified. Only the source population if unique would be weakly diversified

Evaluation round #1

DOI or URL of the preprint: `https://doi.org/10.1101/452326`

Authors' reply, 16 March 2021

Download author's reply
Download tracked changes file

Decision by Stéphane Dupas ⓘ, posted 18 November 2020

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

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.

I added 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.

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.

Line 79. Which year in spain ? 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 it they are mentioned 23 times in the main text.

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

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.

Line 395. The scenario "selected in the global analysis". Change to "whithout bridgehead". To clarify.

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 suggest all the invasion come from a particular USA pop.

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 scenario.

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 ?

With regards,
Stéphane Dupas

Reviewed by **Aude Gilabert**, 12 November 2020

Download the review

Reviewed by anonymous reviewer 1, 10 November 2020

Review Mallez et al. Title: Inference of the worldwide invasion routes of the pinewood nematode Bursaphelenchus xylophilus using approximate Bayesian computation analysis.

In this manuscript, the authors want to investigate the invasion routes of the pinewood nematode from North America to Japan, China and Europe (Protugal). 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 Chine 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.

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?

3) Methods line 229: It is unclear why the authors did not include Fst, 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)?

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?

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

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

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.

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 Ne (see work by Wapples, Montarry) to test this hypothesis.