nature portfolio

Peer Review File

Interbreeding between farmers and hunter-gatherers along the inland and Mediterranean routes of Neolithic spread in Europe

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
This submission aims at linking a simulation of dispersal of early farming across Europe to genetic data, in order to evaluate the proportion of interbreeding between incoming and local populations. These are important, if dated, questions and the submission relies extensively upon former work done by the same group, often at the expanse of other datasets and categories of evidence. The modelling framework is interesting but suffers of key flaws, and the comparison with empirical data is hardly convincing. In this sense, I would recommend that this publication is rejected.
The main issue with this approach lies in several of the decisions linked to the model. Model simplicity is always welcome, as long as it retains enough realism and helps furthering knowledge, two qualities not met by the present submission. My main issue here is that this is an analytical model, and therefore directly predicted by several values and parameters, which are hardly explored. For instance, rather than varying the parameter n, I would like to see how much variation in the proportion of mtDNA haplogroup K in the original population (as implied by the large eroor-bars in the crresponding data) influences the outcome of the model. If this is done, I could not find the information either in the text, or in the SI. Anyway, the lack of any envelope on the simulation outputs suggest minimal care in exploring parameter values.
As fundamentally, the model rests upon speed of dispersal, a long-time obsession of this research group, regardless of the fact that several studies have demonstrated, both empirically and computationally, the limited value of this parameter. The range of variation, especially in Europe, is so large that using a mean is pointless. It is not a surprise for instance that, rather than using individual values associated with sites, the linear regression and corresponding plot uses regional averaged values (with very small error bars), which minimise the noise associated with the data. Further, I have extensive doubt regarding the way these averages were used: a suggested onset of farming in Germany c. 5700BC or in Britain by c. 4500BC is massively at odds with data to be found in the literature. In this sense, the precision and accuracy of the dispersal model used here can be extensively questioned.
Based upon fig.2, the authors suggest that there is a clear cline in the proportion of mtDNA haplogroup K; it is, to some extent, ok when considering the medians only, but, once more, I'd like this assumption to be tested formally by considering the variation attached to the data. In the same vein, is there really a demonstrable difference between both regional dispersals?
Another issue lies in the selection of the data: exclusion of Lepenski Vir because of suggested high levels of interaction between hunter-gatherers and farmers? The literature is indeed large, but the associated genetic pattern far less so. Rejecting Britain because megalithic monuments were only used by specific families, and thus not representative of the entire population is wrong for two reasons; firstly, this is a generalisation of a handful of examples; secondly, the fact that a funerary population is only a fraction of the real population holds everywhere, and in this sense, all samples are flawed. This should be tested rather than assumed.
The rejection of the role of drift, as per the IS, rests entirely upon work done by the same group, and is only considered as "plausible". There is however a large body of literature suggesting that this effect is likely to be more important, and this should thus be taken into consideration (once more, I only trust models which display some uncertainty...)
The way the parameter values is set is often arbitrary: it is difficult to evaluate these directly from empirical data, and thus exploring parameter range is fundamental, something that this paper seems to ignore entirely.
Lastly, it is surprising that the results of this model are actually not compared to the entire range of the available literature, especially from a genetic point of view. For instance, some recent work does indicate profound differences in effective population size between regions (both from a HG or a farmer point of view), and this should be acknowledged and considered. I agree that it may not be possible to do this within the modelling framework used here, but the incompatibility of this with data should not be used as an excuse.

Reviewer #3:
Remarks to the Author:
This manuscript by Fort and Pérez-Losada provides a robust, and archaeologically-informed statistical approach to better understand the post-dispersal interaction of farmers with local hunter-gatherer (HG) populations in Europe during the Neolithic transition. They report important results, especially that the percentage of farmers admixing with HG populations was similar along both the inland and Mediterranean routes of the Neolithic dispersal into Europe, despite key differences in the modes and rates of dispersal. This is significant for the field because, while many papers have empirically analyzed the genetic contributions of HG and farmer communities to the genomes of Neolithic Europe, the question being addressed here, namely, what percent of the farming populations moving into Europe during the Neolithic transition contributed genetic material to European Neolithic populations has not been explored. The paper also provides a unique approach to integrating genetic and cultural transmission in studying the Neolithic transition in Europe. Finally, the authors integrated archaeological research into their simulations and models by choosing parameters, such as population density, based on archaeological estimates.

I do think the paper would benefit from more figures in the main text representing these processes and some of the modeling choices made by the authors need more detailed explanations. Additionally, the discussion of the Neolithic transition could draw more from archaeological literature (in addition to ancient DNA research).

Suggestions

Page 8
It is unclear to me what the term "continental averages' means. From the figures, this does not seem to be the whole European continent, and this term is not defined or explained in the supplementary materials. And if this does refer to the entire European continent, is it being compared with continent-wide averages from Asia and Africa

Page 10
The authors state that one of the primary initial conditions of their analyses is choosing Abu Hureyra as the anchor point for the origin and spread of the Neolithic, citing Pinhasi et al. 2005 to support this. Since one of the main features of the paper is the basis of the models used by the authors on archaeological evidence, I think it would be important to provide a paragraph explaining the choice of Abu Hureyra as the origin and giving some additional context about the transition to farming there, as it is extensively studied. For instance, Theya Molleson and co-authors, examine dietary shifts at Abu Hureyra through dental characteristics, and Gordon Hillman and co-authors examine plant and cereal cultivation at the site. I do think it's a good location to choose for the origin, but I think this choice needs more justification and description, as it is an important part of the argument.

Figures
While readers from archaeology may be familiar with the inland/Danubian road of expansion and the Mediterranean route, a visual representation of this in the main text is important, so readers can visualize the differences in the routes being discussed. While Figure S6a. provides a helpful visualization of the coastal route, a visualization of the inland route is needed and both of these should be featured in the main text.

I also think a map showing the geographics units plotted in Figures 1 and 2 using the labels "A", "B", "C", etc., would be important to illustrate visually on a map. I found Figure S1b., which shows the sites and regional designation in Anatolia to be very helpful and I think a similar map covering the entity study area would likewise be helpful for readers, and reinforce the point that there is sufficient data

from these regions to ask these sorts of questions.

Minor Comments
Page 6
Typo in middle of the page: “Due to the fact that [t]he availability of genetic”

Page 14
R0,HG = 1 is referred to both as a “stationary state” and a “steady state.” I would suggest using only “steady state” as “stationary” sounds like it refers to geospatial mobility when, in fact, this metric represents the net fecundity of hunter-gatherers.

DETAILED RESPONSE TO REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

This submission aims at linking a simulation of dispersal of early farming across Europe to genetic data, in order to evaluate the proportion of interbreeding between incoming and local populations. These are important, if dated, questions and the submission relies extensively upon former work done by the same group, often at the expanse of other datasets and categories of evidence. The modelling framework is interesting but suffers of key flaws, and the comparison with empirical data is hardly convincing. In this sense, I would recommend that this publication is rejected.

We disagree absolutely, for the reasons explained below.

The main issue with this approach lies in several of the decisions linked to the model. Model simplicity is always welcome, as long as it retains enough realism and helps furthering knowledge, two qualities not met by the present submission. My main issue here is that this is an analytical model, and therefore directly predicted by several values and parameters, which are hardly explored. For instance, rather than varying the parameter η, I would like to see how much variation in the proportion of mtDNA haplogroup K in the original population (as implied by the large error-bars in the corresponding data) influences the outcome of the model. If this is done, I could not find the information either in the text, or in the SI. Anyway, the lack of any envelope on the simulation outputs suggest minimal care in exploring parameter values.

In Supp. Methods, Sec. S5 (Sensitivity analysis) we included 5 pages and 16 figures exploring in full detail the effect of parameter values. We also mentioned this 4 times in the main paper (p. 7, 11, 15 and 16 in the first version).

Furthermore, in Supp. Methods, Sec. S4 (Effect of the initial genetic con~~ditions) we~~ included 4 figures analyzing the effect of the uncertainty in the proportion o~~f haplog~~roup K in the original population (error bar 1 in Fig. 3). We also mentioned this twice in the main paper (p. 7 and 16 in the first version).

~~In the~~ new version we also include (Supp. Methods, Sec. S6) 28 new figures with envelopes on the simulation outputs (4 of them also in the main pap~~er, ne~~w Fig. 4). These envelopes have been computed not only for the observed proportion of haplogroup K in the original population (square 1 in Fig. 3), but also for its lower and upper bounds (upper and lower values of error bar 1 in Fig. 3), i.e., taking into account the uncertainty in the proportion of haplogroup K in the original population. They agree with our conclusions, and refine the percentage of farmers that interbred with HGs to about 1.2-8.3% (end of Sec. S6). We explain this in the new version (new Fig. 4 and text in the Results section, last two paragraphs).

As fundamentally, the model rests upon speed of dispersal, a long-time obsession of this research group, regardless of the fact that several studies have demonstrated, both empirically and computationally, the limited value of this parameter. The range of variation, especially in Europe, is so large that using a mean is pointless.

We are not aware of a single published paper that has demonstrated, neither empirically nor computationally, the limited value of estimating the speed of dispersal. Moreover, this reviewer does not provide any references to proof his/her claim. A previous study in which we estimated the mean speed of Neolithic spread in Europe (Pinhasi, Fort & Ammerman, *PLoS Biology,* 2005) has 484 citations (google scholar), and this also contradicts the claim that estimating the mean speed is of "limited value". Of course there is a "range of variation" in the speed, as noted

already by Ammerman and Cavalli-Sforza (1971) and studied in detail by Bocquet-Appel et al. (*JAS* 2012), Fort (*JRSI* 2015), etc. However, the mean speed along each route is very useful as it provides a method to estimate (i) the average importance of demic and cultural diffusion (Fort, *PNAS* 2012) and (ii) the average percentage of early farmers that interbred with hunter-gatherers (HGs) and/or acculturated them along each route (this manuscript). Moreover, the largest variation for the speed of the Neolithic in Europe is that between the inland and Mediterranean routes. This difference is properly taken into account in this paper because we analyze each route separately.

It is not a surprise for instance that, rather than using individual values associated with sites, the linear regression and corresponding plot uses regional averaged values (with very small error bars), which minimise the noise associated with the data.

Contrary to what this reviewer believes, we do not use "regional averaged values" at all. Instead, we do use "individual values associated with sites", namely the oldest Neolithic site for each region. In the first version we stressed this point at several places, e.g., in Results, line 4 we wrote "in Fig. 1 we compare the date of the oldest Neolithic site in several regions", in Methods, end of paragraph 2 we wrote "the oldest archaeological site for each of the regions", in the caption to Fig. 1, line 5 we wrote "each square and error bar gives the radiocarbon date and error of the oldest archaeological site for a region", in the legend to Data S4 we wrote "oldest Neolithic site in each region", etc.

Furthermore, the first version did not contain any linear regression.

Further, I have extensive doubt regarding the way these averages were used: a suggested onset of farming in Germany c. 5700BC or in Britain by c. 4500BC is massively at odds with data to be found in the literature. In this sense, the precision and accuracy of the dispersal model used here can be extensively questioned.

Again, we use the oldest Neolithic date in each region. We do not use any "averages". We have found the oldest Neolithic site per region from existing archaeological databases with hundreds of sites. In order to be totally sure on the reliability of our data, we have now performed a very careful analysis of the selected dates. This has led to minor improvements but the results in the manuscript do not change, as we next explain.

First, the reviewer criticizes our date for Germany. In the first version (Data S4) it was the oldest Neolithic date in Germany from a database (Fort, Pujol & vander Linden, *Amer. Antiq.* 2012) where its material is described as "miscellaneous". This could be criticized because it might include, e.g., charcoal from oak, which would make this date uncertain due to the old-wood effect. Under this criterion, the oldest reliable date from the same database is 6580±20 uncal BP=5611-5478 cal BC (human bone from the Schwanfeld site). We have used this date in the new version, Data S4 and Fig. 1b. It agrees very well with an onset of farming in Germany at about 5,550 cal BC according to archaeologists who are expert in the earliest LBK sites in Germany (see, e.g., the map in Gronenborn *et al.*, *J. Arch. Sci.* 71, 73-83, 2014, Fig. 1, which even highlights the Schwanfeld site used by us). For the benefit of the reader, we now include this explanation in Data S4.

Second, the reviewer criticizes our date for Britain. In the first version (Data S4) it was the oldest Neolithic date in Britain from a database (Betti et al., *Nature Human Behav.* 2020) where its material is described as "charcoal". Under the same criterion as above, the oldest reliable date from that database is 5190±100 uncal BP=4319-3771 cal BC (human bone from the Whitwell Long Cairn site). It is encouraging that this is also the oldest reliable date in the other database mentioned above (Fort, Pujol & vander Linden, *Amer. Antiq.* 2012). We have used this date in the new version. It agrees very well with an onset of farming in Britain at about 4,000 cal BC according to the literature (Cubas et al., *Nature Commun.* 2020; Shennan, The first farmers

of Europe, Cambridge Univ. Press 2018, p. 183; Larson, *Nature* 2015). We include this explanation in Data S4.

Third, we have revised the rest of regions to assure that *all* data are highly reliable (Data S4) by following the same strict criteria in our recent paper on the Western Mediterranean (Fort, *Archaeol. Anthropol. Sci.* 2022), i.e., all samples are either cereal seeds or bones from humans or domesticated animals, thus highly reliable following the guidelines by Zilhao (2011) and Bernabeu (2006), full citations are included in Data S4. For the benefit of the reader, we now include the material and ID of each sample (Data S4).

In this way we have obtained Fig. 1b, which is similar to Fig. 1 in the first version. For this reason, all results (namely, the lines in Figs. 1b and 3, obtained from our simulations) are the same as in the first version.

In summary, in the second version we have improved the dates after a careful analysis, but the results do not change and the precision and accuracy of our dispersal model are absolutely clear.

Finally we stress that, as explained already in the first version (Results, end of first paragraph), naturally in some regions the Neolithic arrived somewhat earlier and in others somewhat later than according to our simple model. However, more complicated models (with non-uniform speeds) are unlikely to change our conclusions because: 1) the latter follow from a genetic cline at the continental level, not in a specific region; and 2) the percentage of haplogroup K changes very slowly after the arrival of the first farmers (Sec. S3-C "Delayed and advanced regions").

Based upon fig.2, the authors suggest that there is a clear cline in the proportion of mtDNA haplogroup K; it is, to some extent, ok when considering the medians only, but, once more, I'd like this assumption to be tested formally by considering the variation attached to the data.
Unfortunately this reviewer does not suggest any specific methods to apply, neither provides any references. Although from Fig. 3 (Fig. 2 in the first version) the existence of a cline along each route seems intuitively clear to us, in the new Sec. S8 we now include 8 figures with 4 correlograms that confirm formally the existence of the cline along both routes. This analysis combines spatial correlograms with the Bonferroni technique, and is performed not only for the observed data but also taking into account their uncertainty by means of bootstrap resampling (Sec. S8). We mention this in the new version (Results section, end of paragraph 4).

In the same vein, is there really a demonstrable difference between both regional dispersals?
Again, this reviewer gives no specific methods neither references. Although from Fig. 1b it is intuitively clear that the slopes of both dispersals are different, in the new Sec. S7 we report linear regressions of the archaeological data (error bars in Fig. 1b). For the inland route the spread rate is 0.89 km/yr and its range is 0.69-1.01 km/yr with 95% confidence level (CL). For the sea route the spread rate is 1.64 km/yr and its range is 1.27-2.01 km/yr with 95% CL. These ranges do not overlap, which confirms that there is a statistically significant difference between both dispersals. We mention this in the new version (Results section, end of paragraph 3).

Another issue lies in the selection of the data: exclusion of Lepenski Vir because of suggested high levels of interaction between hunter-gatherers and farmers? The literature is indeed large, but the associated genetic pattern far less so.
It is very important to note that in this manuscript we consider the simplest possible model of interbreeding, namely a uniform value of the interbreeding intensity (parameter η) along each route. The results of such a simple model agree reasonably well with the genetic data for each route (Figs. 3a-b) by fitting a single parameter (η, which interestingly turns out to have the same value for both routes). The fact that a single-parameter model works so well implies that: (i) such a single model is reasonable at the continental level; and (ii) it would make no sense to complicate the model, because this is unnecessary in order to explain the clines at this scale and

estimate the average percentage of farmers that interbred with HGs and/or acculturated them along each route.

Nevertheless, there can of course be some regional or local exceptions. A case in point is the site of Lepenski Vir (Sec. S1-D). This is an exceptional site because it was not founded exclusively by an incoming population of Neolithic ancestry and culture. Instead, there is a clear co-existence of Mesolithic and Neolithic culture and genes during the Mesolithic-Neolithic transitional phase. During that period there are no domestic animals neither crops but there is, e.g., pottery characteristic of the Balkan early farmers (Shennan, The first farmers of Europe, 2018, p. 73). New haplogroups appear due to incoming individuals with Neolithic ancestry, but a substantial portion of individuals have Mesolithic haplogroups during this phase and individuals with Neolithic ancestry are primarily buried in the habitation area (either in houses or between them), whereas those with HG ancestry are preferentially buried in the upslope area at the rear of the village (Brami et al., *J. World Prehist.* 2022). This is totally different than what we see all over early Neolithic Europe, where there is an abrupt replacement in the economy, culture and genes due to the foundation of new sites by incoming Neolithic populations. It is not reasonable to expect that interbreeding will take place at a similar rate in so different situations, namely a single site with high fractions of individuals with Neolithic and Mesolithic genetic ancestry and culture (Lepenski Vir) versus settlements founded by Neolithic individuals with only occasional incorporation of HGs (in most of Europe). Therefore, it is clearly very reasonable not to include in our analysis the site of Lepenski Vir and similar sites such as Malak Preslavets (Mathieson et al., *Nature* 2018). In future papers, we plan to deal with such exceptional cases by using a local rather than continental approach, i.e., by considering that η may depend on position. For the reasons explained above, we expect that in the area where those sites are located the interbreeding intensity η will turn out to be much higher that the range estimated at the continental level in our manuscript. Such additional studies will require a lot of work and including them in this paper is not realistic neither convenient because it is much clearer to focus our attention in our main point, namely to compare the continental trends of interbreeding along the inland and Mediterranean routes.

Anyway, as mentioned above, the Mesolithic/Neolithic transformational phase in Lepenski Vir cannot be considered as Neolithic because there were no domestic animals neither crops (see Boric & Dimitrijevic 2007, Marchi *et al.* 2022). For this reason, we know the mitochondrial haplogroup of only 2 Neolithic individuals in Lepenski Vir, namely those that lived during the early-middle Neolithic phase (Data S1). Including these 2 individuals would not change the conclusions, because the %K in Data S3 (region 5 Romania and Serbia) would be 15.6% instead of 16.7%, so Fig. 3 would be essentially the same. But for clarity and consistency, we prefer not to include these 2 individuals from Lepenski Vir (for the reasons in the previous paragraph). We now explain this in Sec. S1-D.

Rejecting Britain because megalithic monuments were only used by specific families, and thus not representative of the entire population is wrong for two reasons; firstly, this is a generalisation of a handful of examples; secondly, the fact that a funerary population is only a fraction of the real population holds everywhere, and in this sense, all samples are flawed. This should be tested rather than assumed.

We disagree absolutely. We have now tested explicitly that including the UK and Ireland would not be reasonable, in the following way. Our new table S1 reports all Neolithic individuals in the UK and Ireland for which relatives have been found. They are a total of 63 individuals, i. e. 47% of individuals for which kinship analyses have been performed (Sec. S1-D, text below table S1). This implies that a substantial part (about half) of the sampled individuals are not representative of the whole population but of specific families. Therefore, it would not be reasonable to include data (error bars) from the UK neither Ireland in Fig. 3 in our main paper, where the simulations (lines) correspond to the whole population.

Although this is probably unnecessary, for completeness we mention that it is not worth to complicate the simulations with regional assumptions, which would be unclear and are totally unnecessary for our purposes (namely, to explain the cline of haplogroup K and estimate the average percentage of farmers that interbred with HGs or acculturated them along each route).

Concerning the second reason, the pattern of kinship in the UK and Ireland, as described above, is totally different from that in most of continental Europe, where the first farmers did not build megaliths and such a high percentage of individuals who were relatives of other sampled individuals has never been detected. Thus, in continental Europe it is very reasonable to assume that a random sample of early farmers is representative of the whole early Neolithic population, whereas in the UK and Ireland it is not. In summary, it is absolutely justified to exclude the UK and Ireland from our analysis.

The rejection of the role of drift, as per the IS, rests entirely upon work done by the same group, and is only considered as "plausible". There is however a large body of literature suggesting that this effect is likely to be more important, and this should thus be taken into consideration (once more, I only trust models which display some uncertainty...)

The effect of drift can be important in some specific genetic clines but not necessarily in all of them. And once more, unfortunately this reviewer does not provide any references. In the first version of our manuscript, Supp. Methods, p. 3, we already explained that: (i) previous simulations have shown that a drift effect (surfing) can explain some clines but they cover distances of only about 500 km, so they cannot explain the continent-wide cline of haplogroup K analyzed by us; (ii) a completely independent and fully conclusive reason why drift cannot explain the genetic data is that we observe an increase of hunter-gatherer haplogroups in farming populations as a function of distance from the near East along both routes. This is predicted by our model but not by alternative models of the cline of haplogroup K that do not include interbreeding between farmers and hunter-gatherers (e.g., drift models such as surfing and isolation by distance). For clarity in the new version, Sec. S1-B, point (4), we reiterate this and also explain that our simulations do include drift effects, because we use integer population numbers (not population densities) and random dispersal (i.e., the number of individuals dispersing in not necessarily the same in all directions), so each simulation run yields slightly different results (Sec S2-D). Thus drift is indeed included in our simulations, but its effect turns out to be negligible, and this reviewer is wrong to claim that our simulations do not display any uncertainty.

The way the parameter values is set is often arbitrary: it is difficult to evaluate these directly from empirical data, and thus exploring parameter range is fundamental, something that this paper seems to ignore entirely.

This is totally wrong. As explained at the beginning of this response, already in the first version, Supp. Methods, Sec. S5 (Sensitivity analysis) we included 5 pages and 16 figures exploring in full detail the effect of parameter values. We also mentioned this 4 times in the main paper (p. 7, 11, 15 and 16 in the first version).

In any case, as also mentioned at the beginning, we now include envelopes due to parameter uncertainly (28 new figures in Sec. S6, 4 of them in then new Fig. 4) computed not only for the observed proportion of haplogroup K in the original population (square 1 in Fig. 3), but also for its lower and upper bounds (upper and lower values of error bar 1 in Fig. 3).

Lastly, it is surprising that the results of this model are actually not compared to the entire range of the available literature, especially from a genetic point of view. For instance, some recent work does indicate profound differences in effective population size between regions (both from a HG or a farmer point of view), and this should be acknowledged and considered. I agree that it may not be possible to do this within the modelling framework used here, but the incompatibility of this with data should not be used as an excuse.

The manuscript has about 80 references. The instructions for authors suggest as a guide that they should be less than 70. Including the entire range of the available literature would imply a hundred or several hundred references and a much longer paper. In our opinion this would be fine for a review but not for a research paper. On the other hand, it is true that some genetic work has estimated different population densities in different regions (both for farmers and hunter-gatherers), specially in African populations. Again, this reviewer does not provide any references that could be useful according to him/her. We are not aware of any conclusive estimations for the early Neolithic in the regions analyzed by us in Europe and the Near East. Anyway, in the sensitivity analysis of the carrying capacity of farmers (Sec. S5-C) and hunter-gatherers (Sec. S5-D) we use their upper and lower bounds according to direct ethnographic observations, and we repeat the simulations in the main paper. Thus any non-homogeneous approach would necessarily lead to values of the interbreeding intensity η intermediate between those already found in Secs. S5-C,D and S6. For this reason, obviously the conclusions would not change.

Finally, in our simulations it would be easy to allow for different population densities in different regions (both for HGs and farmers). We do not do this for the two reasons explained above, namely: (i) conclusive data for our regions are not available; and (ii) even if they were, such a change would not add any important insight and would obscure the important points in our manuscript.

Reviewer #2 (Remarks to the Author):

This manuscript by Fort and Pérez-Losada provides a robust, and archaeologically-informed statistical approach to better understand the post-dispersal interaction of farmers with local hunter-gatherer (HG) populations in Europe during the Neolithic transition. They report important results, especially that the percentage of farmers admixing with HG populations was similar along both the inland and Mediterranean routes of the Neolithic dispersal into Europe, despite key differences in the modes and rates of dispersal. This is significant for the field because, while many papers have empirically analyzed the genetic contributions of HG and farmer communities to the genomes of Neolithic Europe, the question being addressed here, namely, what percent of the farming populations moving into Europe during the Neolithic transition contributed genetic material to European Neolithic populations, has not been explored. The paper also provides a unique approach to integrating genetic and cultural transmission in studying the Neolithic transition in Europe. Finally, the authors integrated archaeological research into their simulations and models by choosing parameters, such as population density, based on archaeological estimates.

We appreciate that this reviewer has read carefully the manuscript and Supplementary Methods, and has understood its main points and novelty.

I do think the paper would benefit from more figures in the main text representing these processes and some of the modeling choices made by the authors need more detailed explanations. Additionally, the discussion of the Neolithic transition could draw more from archaeological literature (in addition to ancient DNA research).

We have taken all of these suggestions into account, as explained in detail below.

Suggestions

Page 8

It is unclear to me what the term “continental averages'' means. From the figures, this does not seem to be the whole European continent, and this term is not defined or explained in the

supplementary materials. And if this does refer to the entire European continent, is it being compared with continent-wide averages from Asia and Africa

As noted by this reviewer, we are not dealing with the whole European continent (for reasons explained in Sec. S1-D), so we agree that the term "continental average" is not appropriate and we no longer use it.

Page 10

The authors state that one of the primary initial conditions of their analyses is choosing Abu Hureyra as the anchor point for the origin and spread of the Neolithic, citing Pinhasi et al. 2005 to support this. Since one of the main features of the paper is the basis of the models used by the authors on archaeological evidence, I think it would be important to provide a paragraph explaining the choice of Abu Hureyra as the origin and giving some additional context about the transition to farming there, as it is extensively studied. For instance, Theya Molleson and co-authors, examine dietary shifts at Abu Hureyra through dental characteristics, and Gordon Hillman and co-authors examine plant and cereal cultivation at the site. I do think it's a good location to choose for the origin, but I think this choice needs more justification and description, as it is an important part of the argument.

We now justify in detail the use Abu Hureyra as distance origin and we also describe previous work on this site concerning cultivation (Hillman et al.) and dietary shifts (Molleson et al.). We are very thankful for these citations, because they make this point more complete and understandable.

Figures

While readers from archaeology may be familiar with the inland/Danubian road of expansion and the Mediterranean route, a visual representation of this in the main text is important, so readers can visualize the differences in the routes being discussed. While Figure S6a provides a helpful visualization of the coastal route, a visualization of the inland route is needed and both of these should be featured in the main text.

The new Fig. 1a illustrates both routes and their differences, following the example in Fig. S6a as suggested.

I also think a map showing the geographic units plotted in Figures 1 and 2 using the labels "A", "B", "C", etc., would be important to illustrate visually on a map. I found Figure S1b, which shows the sites and regional designation in Anatolia to be very helpful and I think a similar map covering the entity study area would likewise be helpful for readers, and reinforce the point that there is sufficient data from these regions to ask these sorts of questions.

Fig. 1a gives the locations of the oldest regional sites using the labels "A", "B", "C", etc. The new Fig. 2 shows all early Neolithic sites with genetic data and identifies their regions using the labels "1", "2", etc., similarly to Fig. S1b but covering the entity study area.

Minor Comments

Page 6

Typo in middle of the page: "Due to the fact that [t]he availability of genetic"

This typo has been corrected.

Page 14

R0,HG = 1 is referred to both as a "stationary state" and a "steady state." I would suggest using only "steady state" as "stationary" sounds like it refers to geospatial mobility when, in fact, this metric represents the net fecundity of hunter-gatherers.

We now use "steady" instead of "stationary".

In summary, we have made an earnest effort to prepare this version by including envelopes to the simulation outputs (28 new Figs. in the new Sec. S6 and Fig. 4), a careful revision of the oldest Neolithic site for each region based on strict criteria for sample selection (Data S4), 8 figures with 4 correlograms that confirm formally the existence of the cline along both routes (new Sec. S8), linear regressions showing that both routes have different spread rates (new Sec. S7), three new figures in the main paper (Figs. 1a, 2 and 4), and detailed explanations on (i) the exclusion of some data (including the new table S1), (ii) the role of drift, and (iii) the transition to farming in Abu Hureyra.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Although the authors have made some noticeable effort at conveying their methods better, and also made some key empirical changes to their datasets, the rationale and issues in only using mtDNA (which by definition only provides a very biased, and thus not representative genomic signal) remain poorly explained. In terms of literature, the text still lacks discussion of some of the more recent genomic papers (e.g. https://doi.org/10.1016/j.isci.2022.105387)

Reviewer #2:
Remarks to the Author:
The authors have addressed my concerns about the paper and I think the revisions have have improved the paper. I congratulate the authors on an important and interesting manuscript.

Reviewer #3:
Remarks to the Author:
Comments on ms NCOMMS-22-50491A entitled : « Interbreeding between farmers and hunter-gatherers along the inland and Mediterranean routes of Neolithic spread in Europe" by Joaquim Fort and Joaquim Pérez-Losada

I have read with interest the ms of Fort and Pérez-Losada. I think that it is a very interesting contribution and I also think that the results and conclusions should be of interest to a wide community from archaeologists to population geneticists and to modelers interested in spatio-temporal processes. I was not 100% convinced by the strength of the conclusion as I try to explain below, not because of the mathematical treatment but because of the use of the genetic data, or rather because of the possibly biased choice of the genetic data. I thus felt divided regarding whether I should support publication in Nature Communications or not. Despite these doubts, I feel that the study contributes to our understanding of the Neolithic transition in Europ. It shows how simple models can explain average differences between coastal and inland routes. It suggests that the relationships and interactions between inland and coastal Neolithic communities on one side and local HG communities, on the other side, were not very different, at least in terms of "demographic contributions". It suggests that the transition to farming communities involved similar processes and a similarly low contribution of HG people. Altogether this is one more brick in the building of the Demic diffusion model of Cavalli-Sforza and Ammerman. At the same time, I am convinced that reconstructing the past remains one of the most difficult challenges that we are facing and our community should thus be open to arguments, even if we identify weaknesses in these arguments, or in the data produced or models used. This is why I support its publication despite its weaknesses (see below). It is a study that will generate interest and discussions, and criticisms (including from me). I however would like to see a revised version where the authors explain how they answer my questions, and clarify a few important statistical points.

Below I present what I think are potentially serious weaknesses in the article's premises, and then I identify more specific issues in the manuscript (questionable statements, missing references, etc.). The authors should thus try and correct at least some of the issues I identified and mention some of the potential biases.
I should stress that I am not competent in a number of technical issues identified by the previous referees and to which the authors seem to provide reasonable answers. I will focus on the population genetics aspects of the study. Prof. Fort has significantly contributed to our understanding of spatial processes, human movement and how it can be modelled. He has continued a line of research started

by Cavalli-Sforza and Ammerman, and others, and has gone further. This study fits nicely in this dynamic research and is a natural continuation of his 2017 article (ref. 18: Isern et al., 2017). My worry is that the authors seems to take this article as a solid basis from which the new article would be an extension, without discussing the possible weaknesses of the previous study. This is dangerous as this is how house-of-cards are built. I think that both articles suffer from the same general weakness related to the use of a haplogroup and taking this as a way to access demographic processes quantitatively. There is a difference between the fact the haplogoup K is a good indicator of a process and the fact that we can use it to obtain quantitative estimates of demographic parameters.

This article uses the haplogroup K data as if there were no potential statistical bias(es). As population geneticists we try to understand how genetic data can be used to infer specific parameters of a demographic model which we think could be appropriate to understand the recent evolutionary history of populations of interest. For instance, we could have an admixture or interbreeding model for the Neolithic transition and wish to estimate the admixture parameter across Europe to determine whether the process is most likely demic (limited HG contribution during the admixture process between farmers and HG) or cultural (major HG genetic contribution) and whether it varies spatially.
To do that we try to use as many independent genetic markers as possible (assuming that they are neutral), where each one of them may have a few or many different alleles or haplotypes. It is thus a limitation when one selects (i) only one locus (mtDNA behaves as a single locus) and then (ii) a subset of alleles from that one locus (i.e. that reduces even further the information). An additional statistical issue arises because the genetic markers (haplogroup K) are selected because they happen to have a spatial distribution that fits the process one is interested in.
From a population genetics point of view, this selection process not only throws away potentially interesting information but it also can create several statistical issues. One is a bias of the average effect (if I only select haplogoups that exhibit a cline) and the other is an underestimation of the variance of the process one wishes to study. The risk is that one will thus be overconfident of a biased signal.
To clarify my point, there are many types of genetic markers and they will all be influenced in similar ways by the demographic processes such as expansions, contractions, admixture, etc. If we choose markers that look like what we expect we will tend to estimate parameters that favour our hypothesis, if we have a good intuition (if we select the right type of markers).
To conclude on this general issue, the work is of potential interest to a wide community. It is of quality in the sense that the work seems mathematically solid but I am afraid that it is misleading because it focuses on a biased data set. The authors would be more convincing if they could clarify and sort these statistical issues. One way would be to simulate genetic data under the scenario they identify, and then select from the many independent markers those that follow the pattern they detected in haplogroup K, then infer the process again and see if the estimated parameters are biased or not. It is possible that the bias would be limited, but the variance would likely be underestimated. I hope that this makes sense. There may be other ways, this is just a suggestion.

Below are more specific comments with citations from the manuscript:

The authors write that ancient DNA established that the Demic Diffusion played a primary role in the spread of the Neolithic: “About 15 years ago, the technical possibility to analyze ancient DNA unveiled large genetic differences between local HGs and early Neolithic farmers in central Europe, thereby establishing that demic diffusion played a primary role in the spread of the Neolithic”.
I have to disagree with the second part of that statement. The Demic Diffusion was established way before by many studies based on modern genetic data. From Barbujani et al (1995, AJPA) who used spatial simulations to reproduce spatial patterns to Chikhi et al (1998, PNAS) who used autocorrelation patterns or Chikhi et al (2002, PNAS) who used admixture models and spatial analyses or Currat et al., (2005, P) who used spatial simulations. It is important to remind the authors that there is a rich literature using population genetics theory that showed that no cultural diffusion process could explain the spatial patterns of genetic data. The ancient DNA data confirmed these studies, they did not establish the role of Neolithic farmer.

I am thus particularly surprised by the citation of the Bramanti et al. 2009 paper (ref. 9) in that context. I suggest that the authors ignore that study (or read it carefully, and agree that it can be ignored!) and that they read the study by Rasteiro et al (2013) (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0060944) who among other things deconstructed the Bramanti et al study. For instance, the Bramanti paper uses simulation approaches that completely ignore spatial processes. They assume that the whole of Europe was a panmictic population since its colonization 45,000 years ago.

Later the authors define genetic markers "A genetic marker can be defined as a set of genetic features that may de different for different individuals.". First of all, to explain "genetic markers" by using "genetic features" as if this was clearer than "genetic markers" is disputable to say the least.
My understanding of the phrase "genetic marker" is that it is a vague term whose meaning changes depending on the type of genetic studies. It is thus a good idea to try and clarify what definition the authors use, but here my feeling is that the definition used by the authors is too vague to be useful. I may be wrong but it seems to me that what the authors call a genetic marker is a set of mtDNA alleles (haplotypes here because they are part of a non-recombining DNA molecule) that are thought to have a shared ancestor not shared with other mtDNA alleles/haplotypes. If I am wrong then the authors should provide another, possibly better definition. They can also consider that genetic markers would be the mutations associated to that haplogroup. Their decision.

Later again the authors actually make a statement that goes rather against what I wrote above, namely: "Therefore, the spatial patterns of the frequencies of genetic markers can be very different, due to the different processes that cause each of them. This is why studies that aggregate markers with many different spatial patterns (due to different processes, as explained above), such as genome-wide studies, are not appropriate to explain the spatial pattern of a single marker."
In my opinion the authors are making a possibly serious mistake. The processes are necessarily the same across the genome (assuming neutrality). The only reason why the processes would be different is if the process was driven in relation to the alleles of specific markers. For instance, if admixture was higher depending on say, their blood groups, that could make sense to ignore such markers. But I cannot see how humans could guess each other's haplogroups, and thus why different haplogroups would be more or less influenced by admixture. What happens is that haplogroups may have different spatial patterns at the moment of the Neolithic diffusion and thus they will be more or less able to exhibit this signature. But they will all be influenced in the same way. Using aDNA to confirm that a haplogroup had the right pattern at the right time makes senses but does not take away the risk to create statistical biases or reduce variance of the process.
Thus, the authors should prove (or convince us) that the selection of haplogroup K provides an unbiased signal and that the variance is properly estimated.

Another point that needs to be addressed is the fact that the authors seem to refer to a linear decrease of frequencies. This is surprising as I would expect a geometric decrease. Chikhi et al (2002) used a simple rationale to suggest that if admixture took place according to a "stepping stone" model similar to the implicit model of the authors, admixture proportion should display a geometric (not a linear) decline of admixture proportions as one moves away from the expansion source. Can the authors clarify why they seem to implicitly assume a linear decline even though the figures do actually represent a geometric decline. It may be just me who misunderstood something. I am in any case surprised that the Chikhi et al 2002 study was not at all cited as it was, to my knowledge, the first to actually make the point that the local admixture process should be separated from the spatial effect of admixture. The same process (say 90% farmers and 10% HG) can lead to a high or very low genetic contribution. For instance, if there are 10 or 20 stops during the expansion process, with the same admixture proportions at each encounter point, the final "Neolithic contribution" will be $0.9^{10} = 0.34$ or $0.9^{20} = 0.12$, and the average genetic contribution 0.586 and 0.395 respectively. This will be true for any genetic marker, any allele, etc. If we look at genetic markers that had the same frequencies in HG and farmers, there will be no information of the process. If they look at markers which are only present in one group the information will be maximized but will underestimate the uncertainty on the

process. Finally, Chikhi et al (2002) looked at how admixture proportions varied across Europe and used this to identify what kid of spatial process was reasonable (number of stops and local admixture rate).
One interesting question could thus be whether the haplogroup K could be explained by different HG contributions depending on the number of stops, assuming that the actual process was identical at each location. How would that differ from the Chikhi et al results.

Later in the manuscript the authors use PCA plots as an argument for clines observed in Neolithic individuals. I would be very cautious here. PCA plots are just a representation and there is an increasing number of studies identifying serious issues in the interpretation of such plots in terms of demographic processes. Genetic drift alone can "move" points across the board and the percentage of variance explained may be much lower that one wishes to see.

In the supplementary material the authors fail to mention the study of Rasteiro et al 2012 which used spatial simulations using a simulation framework similar to that Currat and Excoffier 2005 but in a forward (not coalescent) framework. This allowed them to study male and female behaviour, in terms of philopatry and admixture. That study also looked a patterns in ancient samples (in the supplementary material: they showed that some patterns were visible little after admixture, i.e. in aDNA but would disappearing present-day data). These authors did not use aDNA but stressed that aDNA would allow them to separate scenarios because they tested it with simulations.

To conclude, I think that this manuscript provides interesting new approaches, that it needs to justify better why using the haplogroup K is acceptable, and rewrite some sections.

If the authors feel that I misunderstood something or was simply wrong, they should not hesitate to tell me. I will be happy to read the revised version of the manuscript.

POINT-BY-POINT RESPONSE TO THE REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author)

Although the authors have made some noticeable effort at conveying their methods better, and also made some key empirical changes to their datasets, the rationale and issues in only using mtDNA (which by definition provides a very biased, and thus not representative genomic signal) remain poorly explained.

We believe that we now justify this point appropriately. Please see our detailed answer below to reviewer 3, point 1), Sec. b), where we explain that we now include also Y-chromosome results along the inland route (new Fig. 5) and that they support the conclusions obtained from mtDNA, not only qualitatively but also quantitatively. In the new version we also mention that when more data for Y-chromosome data along the sea route are known, they will provide another test of our results.

In the last part of the Discussion we explain that our method cannot be applied to the complete genome, because we cannot define haplogroups in the sense used in the present paper due to recombination. However, when more data are available additional tests of our results may be possible for regions of autosomes that do not recombine or have essentially no recombination.

In terms of literature, the text still lacks discussion of some of the more recent genomic papers (e.g., Arzelier et al. 2022).

We have extended the discussion on the more recent genomic papers, including Arzelier et al. 2022, Yu et al. 2022, Ariano et al. 2022, Hofmanová et al. 2022, Cox et al. 2024, Allentoft et al. 2024, etc. We have had to cut other parts of the manuscript to keep its length and number of references within the guidelines of this journal.

Reviewer #2 (Remarks to the Author):

The authors have addressed my concerns about the paper and I think the revisions have improved the paper. I congratulate the authors on an important and interesting manuscript.

Thank-you.

Reviewer #3 (Remarks to the Author):

I have read with interest the manuscript of Fort and Pérez-Losada. I think that it is a very interesting contribution and I also think that the results and conclusions should be of interest to a wide community from archaeologists to population geneticists and to modelers interested in spatio-temporal processes.

Thank-you.

I was not 100% convinced by the strength of the conclusion as I try to explain below, not because of the mathematical treatment but because of the use of the genetic data, or rather because of the possibly biased choice of genetic data ... Despite these doubts, I feel that the study contributes to our understanding of the Neolithic transition in Europe. It suggests that the relationships and interactions between inland and coastal Neolithic communities on one side and local HG communities, on the other side, were not very different, at least in terms of "demographic contributions". It suggests that the transition to farming involved similar processes and a similarly low contribution of HG people. Altogether this is one more brick in the building of the demic diffusion model of Cavalli-Sforza and Ammerman. At the same time, I am convinced that reconstructing the past remains one of the most difficult challenges that we are facing ... This is why I support its publication despite its weaknesses (see below) ... I however would like to see a revised version.

We have taken into account the suggestions by this reviewer, as explained in detail below.

1)

Below I present what I think are potentially serious weaknesses in the article's premises, and then I identify more specific issues ...

The authors should thus try and correct at least some of the issues I identified, and mention some potential biases.

... Prof. Fort has significantly contributed to our understanding of spatial processes, human movement and how it can be modelled. He has continued a line of research started by Cavalli-Sforza, Ammerman, and others, and has gone farther. This study fits nicely in this dynamic research and is a natural continuation of his 2017 article (ref. 18: Isern et al., *Sci. Rep.* 2017). My worry is that the authors seem to take this article as a solid basis from which the new article would be an extension, without discussing the possible weaknesses of the previous study ... I think that both studies suffer from the same general weakness related to the use of a haplogroup and taking this as a way to access demographic processes quantitatively. There is a difference between the fact that haplogroup K is a good indicator of a process and the fact that we can use it to obtain quantitative estimates of demographic parameters.

This article uses the haplogroup K data as if there were no potential statistical bias(es). As population geneticists ... we try to use as many independent markers as possible ... It is thus a limitation when one selects: (i) only one locus (mtDNA behaves as a single locus) and then (ii) a subset of alleles from that one locus (i.e., that reduces even further the information). An additional statistical issue arises because the genetic markers (haplogroup K) are selected because they happen to have a spatial distribution that fits the process one is interested in. From a population genetics point of view, this selection process not only throws away potentially interesting information but it also can create several statistical issues ... The risk is that one will thus be overconfident of a biased signal.

To clarify my point, there are many types of genetic markers and they will all be influenced in similar ways by demographic processes such as expansions, contractions, admixture, etc. ... If we choose markers that look like what we expect, we will tend to estimate parameters that favor our hypothesis ...

To conclude on this general issue, the work is of potential interest to a wide community. It is of quality in the sense that the work seems mathematically solid but I am afraid that it is misleading because it focuses on a biased data set. The authors would be more convincing if they could clarify and sort these statistical issues. One way would be to simulate genetic data under the scenario they identify, and then select from the many independent markers those that follow the pattern they detected in haplogroup K, then infer the process again and see if the estimated parameters are biased or not ... I hope this makes sense. There may be other ways, this is just a suggestion.

It is true that our manuscript analyzes the cline of a specific marker (mitochondrial haplogroup K). In the new version we acknowledge the limitations of these data (last 3 paragraphs of the Discussion section). As suggested by the reviewer, it is certainly possible to simulate many clines under the scenario that we identify ($\eta = 0.07$, from Fig. 3) by varying the percentage p_O of the haplogroup considered at the origin region of the Neolithic expansion and its percentage p_{HG} in hunter-gatherers. For clarity let us consider the case in which p_{HG} is approximately uniform (although this is not necessary, it is simpler). Firstly, if $p_{HG} = p_O$ then HGs and famers have the same percentage of the haplogroup considered, so interbreeding will not cause any increase neither decrease of its frequency in farmers. Thus the cline will be horizontal. Secondly, if $p_{HG} < p_O$ then relatively few HGs transfer (via interbreeding) the considered haplogroup to the populations of farmers, so HGs will mostly transfer other haplogroups, and the percentage of the considered haplogroup in early farmers will decrease with increasing distances (as in Fig. 3). And thirdly, if $p_{HG} > p_O$ then relatively many HGs transfer the haplogroup considered, so the percentage of the marker in early farmers will increase with increasing distances. We have checked these intuitive expectations using simulations. Now, if we follow the suggestion by the reviewer and select markers with the decreasing pattern observed for haplogroup K, we will always obtain the same cline that we have simulated ($p_{HG} = 0$) because all other clines ($p_{HG} \neq 0$) will obviously have less negative slopes (or even positive ones) due to the transfer of the haplogroup considered from HGs to the populations of farmers. Obviously changing other parameter values does not modify this result, because for consistency we have to use the same parameter values to generate a cline and to infer the value of p_{HG} from it. Thus, in our opinion this approach does not make it possible to verify that using the cline of haplogroup K yields reasonable results. For this reason, in the new version we apply the following two alternative ways to verify this point.

a) In our first approach, we have tested if random effects could be responsible for the observed clines of haplogroup K along the inland and sea routes. By generating clines at random, we find that a spatial decrease similar to that of haplogroup K is not obtained (with 90% confidence level) along any of both routes (Sec. S9). This suggests that the formation of the clines of haplogroup K is driven by a non-random process. Our model includes such a process (interbreeding and/or acculturation), so it is reasonable to apply our

model to analyze the clines of haplogroup K. We summarize these results in the main paper, Discussion section.

b) In our second approach, we consider the Y chromosome. In the new Fig. 5 we find that its highest-frequency marker (haplogroup G2a) is consistent with our model along the inland route, not only qualitatively (because it exhibits a decreasing cline) but also quantitatively (because the range obtained from mitochondrial DNA in Fig. 3a, namely $\eta = 0.07 - 0.08$, is consistent with the Y-chromosome cline in Fig. 5). In our opinion, this is strong support for our model in the sense that is independent from mitochondrial DNA data. We explain this point in the main paper, Discussion section. There we also explain that when more data become available, the cline of haplogroup G2a along the sea route will provide a further test of our model, and that future additional tests of our proposal could be possible for regions of autosomes that do not recombine or have essentially no recombination.

c) An important related point is the following. The new Sec. S11 contains an example (haplogroup T2) showing very clearly that for low-frequency haplogroups we cannot detect the decreasing clines expected from our model, because of the uncertainty due to the small regional sample sizes available at present (Fig. S24b). Therefore, with the data available at present we can only detect clear clines for markers that display high frequencies (about 40% or more). In practice this means that we have to use haplogroup K because all other haplogroups display low frequencies in all regions (below 20%). We believe that this clarifies substantially why we have used the highest-frequency marker in our database (haplogroup K, which attains frequencies above 40%). We summarize this point (Sec. S11) in the main paper, Introduction and Discussion.

2)
Below are more specific comments ...

The authors write that ancient DNA established that Demic Diffusion played a primary role in the spread of the Neolithic: "About 15 years ago, the technical possibility to analyze ancient DNA unveiled large differences between local HGs and early Neolithic farmers in central Europe, thereby establishing that demic diffusion played a primary role in the spread of the Neolithic". I have to disagree with the second part of that statement.

We have deleted the second part of that statement.

The Demic Diffusion was established way before by many studies based on modern genetic data. From Barbujani et al (1995, *AJPA*) who used spatial simulations to reproduce spatial patterns to Chikhi et al (1998, *PNAS*) who used autocorrelation patterns or Chikhi et al (2002, PNAS) who used admixture models and spatial analyses or Currat et al (2005, *PRSB*) who used spatial simulations. It is important to remind the authors that there is a rich literature using population genetics theory that showed that no cultural diffusion process could explain the spatial patterns of genetic data. The ancient DNA data confirmed these studies, they did not establish the role of Neolithic farmers.

In the Introduction we now explain that those works analyzed modern DNA data and found them consistent with a mainly demic diffusion spread of the Neolithic.

I am thus particularly surprised by the citation of the Bramanti et al. 2009 paper (ref. 9) in that context. I suggest that the authors ignore that study (or read it carefully, and agree that it can be ignored!) and that they read the study by Rasteiro et al (2013) who among other things deconstructed the Bramanti et al. study. For instance, the Bramanti paper uses simulation approaches that completely ignore spatial processes ...

We have deleted the citation of the Bramanti et al. 2009 paper. We agree on the importance of the study by Rasteiro et al (2013) in this context and summarize its main conclusion in the new version.

Later the authors define genetic markers ... my feeling is that the definition used by the authors is too vague to be useful. I may be wrong but it seems to me that what the authors call a genetic marker is a

set of mtDNA alleles (haplotypes here because they are part of a non-recombining DNA molecule) that are thought to have a shared ancestor not shared with other mtDNA alleles/haplotypes ...

We now use the definition of genetic marker suggested by this reviewer.

Later again the authors actually make an statement that goes rather against what I wrote above, namely: "Therefore, the spatial patterns of the frequencies of genetic markers can be very different, due to the different processes that cause them. This is why studies that aggregate markers with many different spatial patterns (due to different processes, as explained above), such as genome-wider studies, are not appropriate to explain the spatial pattern of a single marker".
In my opinion the authors are making a possibly serious mistake. The processes are the necessarily the same across the genome (assuming neutrality) ... What happens is that haplogroups may have different spatial patterns at the moment of the Neolithic diffusion and thus they will be more or less able to exhibit this signature. But they will be influenced in the same way.

We agree that this text was misleading and have deleted it.

Using aDNA to confirm that a haplogroup had the right pattern at the right time makes sense, but does not take away the risk to create statistical biases or reduce the variance of the process. Thus, the authors should prove (or convince us) that the selection of haplogroup K provides an unbiased signal and that the variance is properly estimated.

Please see our response to point 1) above, where we have explained the two ways a) and b) that we use in the new version to verify that using the cline of haplogroup K yields reasonable results.

Another point that needs to be addressed is the fact that the authors seem to refer to a linear decrease of frequencies. This is surprising as I would expect a geometric decrease. Chikhi et al (2002) used a simple rationale to suggest that if admixture took place according to a "stepping stone" model similar to the implicit model of the authors, admixture proportion should display a geometric (not a linear) decline of admixture proportions as one moves away from the expansion source. Can the authors clarify why they seem to implicitly assume a linear decline even though the figures do actually represent a geometric decline. It may be just me who misunderstood something.

We do not assume any linear decline of genetic frequencies. In fact, as noted by the reviewer, the decrease of genetic frequencies in Figs. 3-5 (curves) is not linear. Perhaps the confusion is due to the fact that in Fig. 1 the variations are linear. The reason is that Fig. 1 does not display genetic frequencies but arrival times of the Neolithic front, for which well-known reaction-diffusion theory and numerical simulations imply linear dynamics (see, e.g., Eq. (S1) and Fig. S2b). In any case, to avoid confusion we now mention (in the Introduction) that the frequency decrease is non-linear and provide a cite with a clear justification [Chikhi et al, PNAS 2002]. Also to avoid confusion, we have replaced 'slope' by 'average slope' in 3 places.

I am in any case surprised that the Chikhi et al 2002 study was not at all cited as it was, to my knowledge, the first to actually make the point that the local admixture process should be separated from the spatial effect of admixture. The same process (say 90% farmers and 10% HGs) can lead to a high or very low genetic contribution. For instance, if there are 10 or 20 stops during the expansion process, with the same admixture proportions at each encounter point, the final “Neolithic contribution” will be 0.910= 0.34 or 0.920= 0.12, and the average genetic contribution 0.586 and 0.395 respectively. This will be true for any genetic marker, any allele, etc. If we look at genetic markers that had the same frequencies in HG and farmers, there will be no information of the process. If they look at markers which are only present in one group the information will be maximized but will underestimate the uncertainty on the process. Finally, Chikhi et al (2002) looked at how admixture proportions varied across Europe and used this to identify what kind of spatial process was reasonable (number of stops and local admixture rate). One interesting question could thus be whether the haplogroup K could be explained by different HG contributions depending on the number of stops, assuming that the actual process was identical at each location. How would that differ from the Chikhi et al results.

We agree that Chikhi et al (PNAS 2022) is relevant to several issues discussed in our manuscript. Among other things, they considered a model in which there are n admixture events along the Neolithic spread in Europe and considered explicitly the values n=3, 10 and 20. In that model, the Neolithic contribution

decreases by the same factor (e.g., 0.9) in each admixture event. In the spatial simulations reported in our manuscript, the decrease in each admixture event is computed using Eqs. (1)-(12) (Methods, (iii) Cultural transmission) and farmers move 50 km per generation along the inland route and 70 km per generation along the sea route. These values (50 km and 70 km) lead to agreement between our simulations and the archaeological data (Fig. 1b and Results section). Since the total distance between sites with genetic information along the inland route is about 3,000 km (Fig. 3a), there are about 3,000/50=60 admixture events (one per generation) on a straight line along the inland route. Similarly along the sea route (Fig. 3b) there are about 6,000/70=86 admixture events. In our spatial model we cannot change these values (60 and 86 admixture events) because even small changes would imply changing the distances moved per generation (50 and 70 km) in such a way that our simulations (lines in Fig. 1b) would disagree with the archaeological data (error bars in Fig. 1b). Additionally, the value 50 km has been independently estimated from ethnographic data (first paragraph in the Results section). Thus we cannot change arbitrarily the number of admixture events in our homogeneous spatial model.

Alternatively, we could assume a non-homogeneous model, in which there is interbreeding only in specific cells of our simulation grid (although we have no reliable data to justify in which cells there is interbreeding and in which ones there is not, e.g., because they were empty of HGs when the first farmers arrived). Such a model would lead to clines that, in contrast with Figs. 3-5, would not be smoothly decreasing lines but a series of steps (i.e., a series of horizontal lines with $\eta = 0$ alternated with sloping lines with values of η larger than those found by us in Fig. 3, i.e., larger than $\eta = 0.06 - 0.08$). This could be the topic of future work, but in this paper we are interested in using the simplest possible model (homogeneous dispersal, reproduction and interbreeding along each route) because our aim is to compare the interbreeding behaviour along both routes. In fact we honestly believe that analyzing such the effect of such and/or other local or regional processes would make the main point of the paper less clear. In the Discussion we explain that in future work, non-homogeneous models would be of interest to discuss scenarios in which admixture events do not take place in all locations but only in specific ones [Chikhi et al., PNAS 2002]. This will require substantially lengthier simulations and analyses, which implies a lot of work and text to explain the new simulations and results carefully. This is an additional reason why such a study seems more appropriate for a separate paper (the manuscript of the present submission cannot be longer according to the journal guidelines, and the Supp. Methods are already 57 pages long).

Later in the manuscript the authors use PCA plots as an argument for clines observed in Neolithic individuals. I would be very cautious here. PCA plots are just a representation and there is an increasing number of studies identifying serious issues in the interpretation of such plots in terms of demographic processes. Genetic drift alone can “move” points across the board and the percentage of variance explained may be much lower than one wishes to see.

We changed the sentence "It is well-known that in principal-component plots, as well as in estimations of HG ancestry, early Neolithic individuals display clines ..." into "It is well-known from estimations of HG ancestry that early Neolithic individuals display clines ...".

In the supplementary material the authors fail to mention the study of Rasteiro et al 2012 which used spatial simulations using a simulation framework similar to that Currat and Excoffier 2005 but in a forward (not coalescent) framework. This allowed them to study male and female behaviour, in terms of philopatry and admixture. That study also looked a patterns in ancient samples (in the supplementary material: they showed that some patterns were visible little after admixture, i.e. in aDNA but would disappearing present-day data). These authors did not use aDNA but stressed that aDNA would allow them to separate scenarios because they tested it with simulations.

In Sec. S1a, point (5), we now summarize the work by Rasteiro et al (PRSB 2012) and compare it to the other simulation studies reviewed.

To conclude, I think that this manuscript provides interesting new approaches, that it needs to justify better why using the haplogroup K is acceptable, and rewrite some sections.

We believe that we have justified why using haplogroup K is acceptable in point 1) above, Secs. a)-c). We have also rewritten the sections suggested.

Reviewers' Comments:

Reviewer #3:
Remarks to the Author:
see attached pdf

I have now read the revised ms of Fort and Pérez-Losada (and I am sorry for my slowness in making these comments). The authors have done their best to answer my questions and requests and those of the referees. They have clarified a number of issues that bothered me, cited the relevant literature and the new ms reads really well.

As I noted in my previous comments, this is a relevant study that maintains a tradition started by Cavalli-Sforza and Ammerman several decades ago and pursued by the authors and a few others. **We need more studies like this where modelling tries to clarify demographic processes**. I ***was*** less convinced by the genetic analyses and I think that the interpretation of genetic data remains a controversial issue which will likely last a few additional decades. **But the authors have done their best to satisfy my queries and I feel that it is difficult to do more. They have much improved the reference to previous studies, clarified the controversy and contributions of previous studies. They have also clarified their arguments for the use of the K haplogroups.**

I am thus happy to suggest acceptance and publication. I congratulate the authors for their work

Response to the reviewers' comments

We are thankful for the reviewer comments on version 3.

She/he does not suggest any changes.