

Peer Review File

Manuscript Title: Large-Scale Migration into Britain During the Middle to Late Bronze Age

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

This paper describes a large body of new ancient genomic data, mainly from Britain. This is a significant data legacy contribution - lessened somewhat as it is partial (captured genome) sequence rather than whole genome data but this is balanced by the fact that target coverage is very good in many specimens. The data is clearly carefully obtained, controlled for quality, and described in supplemental tables in good detail. One issue is that only aligned reads are submitted for retrieval - I am not clear why this is, rather than raw fastq data files? At least the alignment parameters and process should be made clear for other researchers who might use these data.

The major finding is evidence for a major migratory influence in Britain which alters the balance of ancestral components and which involved a minimum of 50% contribution from migrants during the Bronze Age. This is tied to M-LBA cultural connections between Britain and mainland Europe. I think that this is an exciting assertion, is additional to previously described migratory episodes, and would not be visible without ancient genomics. My opinion is that this finding merits publication in Nature. However, I have some queries and recommendations regarding the work as presented.

The work is very well presented with clear writing and informative figures. However, the central finding is a dramatic one and I think the reader should be given a closer look at the analyses in a more basic form, in order to judge some inferences.

Firstly, the central assertion that in a temporal transect there is a shift in Early European Farmer ancestry component in England and Wales is illustrated in fig 2a in four period bins, in fig 3 where individual measures are given with the same bins as average lines and in a temporally extended but similar supplement figure. The detail in fig 3 is commendable and a fair amount of noise in this measure is clear. I wonder about timing of this shift in ancestry. The discussion in the paper involves inference around whether this is an episodic (eg M-LBA connections) or gradual, and also asserts little impact of migrants in the Iron Age. I wonder what temporal resolution there actually is within this phenomenon. Rather than use the same temporal bins for averaging as in fig 2, it could be good to use a more agnostic approach (eg sliding windows) in Fig 3 to ask what bins of similar EEF ancestry naturally emerge from the data. How granular is the analysis with regard to estimating the timing of EEF ancestry change and whether it is episodic or gradual? Note some specificity is claimed - line 567 "overwhelming evidence that major movements of people occurred during the M-LBA, particularly during a period of intensification of exchange that led to the 'Match-Mer du Nord complex'"

Second, and related, how sure are the authors that there is no major Iron Age migration that may have, eg, accompanied language influx or La Tene material culture? A shift is easier to detect when there is a higher differential in ancestral component between the donor and recipient population and less distinguishable when they are similar in composition. Given that the occurrence of high EEF outliers through the transect is reasonably constant (to the eye, but this is not examined as a proportion of samples) and that there is a stated evening out of European ancestral components among European regions in later periods, could the plateau in ancestry in the Iron Age be (at least in part) a plateau of detectability rather than stability of population?

Incidentally, some outliers seem odd, eg one just after Thanet ca. 2000 BC seems centrally located

- also the third Carsington sample?

Third, further evidence for the migratory influx is the shift in N_e calculated from ROH plotted in Fig 2b. As an independent aspect of the data this is important support. However as a reader it is impossible to see the trend in the smoothed line reflected in the raw ROH data from which it emerges. Could this be improved, eg by plotting median (or mean) ROH sums for each time bin?

Fourth, an interesting aspect of the EEF ancestral component variation is that it is geographically distributed within Britain, primarily summarised as a Scotland Vs England/Wales contrast. I think it would be good to let the reader see this in some detail, similar to the temporal transect. Eg a plot of EEF ancestry for individuals jittered on a map of Britain could allow one to assess the nature of this - this would also allow the reader to easily judge the distinction of other outliers such as the East Yorkshire Arras samples.

Minor points:

It is nice to see the Amesbury archer included, it might be good to also label his companion on fig 3

Line 430: is the discussion of population size change as being constant followed by 5X expansions too literal? N_e is not census population and an influx of a population with different background could be a disruption of this statistic of a different sort to raw population size increase.

Such a migration does open a window for the influx of Celtic languages. However, how sure is the assertion of evidence of non-IE persistence in Scotland, which is used in support of this process? It is not my expertise but I understand a reasonable settling of opinion has occurred that Pictish is a Brittonic Celtic language? Is there a danger of picking isolated supporting evidence from another field?

I think the introductory review of migrations as an issue in archaeology is welcome. line 239 seems contradictory though, British WHG ancestry is stated as almost entirely due to admixture in Mainland Europe indicating that British local Mesolithic people are absorbed.

Referee #2 (Remarks to the Author):

On the basis of much new aDNA data from Bronze and Iron Age Europe, the authors argue that, subsequent to the so-called 'Beaker' migrations of circa 2450 BC (Olalde et al. 2018), there were, circa 1300-800 BC (in the Middle to Late Bronze Age), massive migrations of people from France into southern Britain (England + Wales). These were followed by a period after circa 800, in the pre-Roman Iron Age, with much less migration into Britain. The authors rightly note the desirability of more aDNA data from northern France and Ireland, but undoubtedly they have made an important advance which should be published as soon as possible.

The authors are careful not to specify the nature and context of their migrations, rightly avoiding a premature interdisciplinary synthesis at this stage. They do, however, cautiously note that migrations across the English Channel would suit the recent hypothesis (Sims-Williams 2020) that the Celtic language spread from France circa 1000 BC rather than spreading later, in the Iron Age, as was long assumed, or developing earlier in situ (from PIE) in about the fourth millennium BC (as in Cunliffe 2018 and earlier publications). Their view seems plausible. If so, the implication of their genetic data would seem to be that Celtic-speakers reached northern Britain relatively late. That, as the authors rightly realise, would be consonant with the current, sober, judgment that many of the Pictish inscriptions of Scotland cannot be explained as Celtic (Rodway 2020).

The question of how Celtic got Ireland is not addressed. That does not seem to be an insuperable problem, however. If it does not turn out that there was a parallel migration from the Continent to Ireland, comparable to that to southern Britain, one could get round that by hypothesising that Ireland, like other peripheral areas (e.g. the Iberian peninsula and perhaps Scotland), had a large number of competing minor/relic languages; if so, a relatively small number of incoming Celtic-speakers (direct from the Continent and/or via Britain), coming in under the archaeological and genetic radar, might be enough to effect the language shift to Celtic that is first attested in the Irish place-names recorded by Ptolemy circa 150 A.D. Sims-Williams (2020) quotes Nettle (1999) as saying that 5000 speakers is the median number of speakers for languages worldwide. Supposing that was the situation in Ireland, even 20,000 Celtic-speaking incomers could have a decisive effect in favour of a language shift to Celtic. The same could apply in Britain, as hypothesised by Sims-Williams, in which case the eventual prevalence of Celtic would not *have* to be due to the 'large scale migration' detected by the authors. Nevertheless, their explanation looks attractively straightforward and deserves publication.

Some minor points:

1. Given that the media tend to misunderstand and sensationalise, the singular 'migration' in the title and abstract may give the unintended impression of an event as opposed to a process. This would be regrettable, as the authors are careful in the text to point out. Would the plural 'Migrations' be better in the title?
2. In connection with the Pictish point above: it would be best not to say 'a non-Celtic language *remained in use into* the first millennium CE in Scotland'. It would be safer to say 'a non-Celtic language *was in use in* the first millennium CE in Scotland'. We can't be 100% sure that Pictish wasn't introduced from elsewhere early in the first millennium.
3. Instead of 'facilitated by the emergence of a Celtic lingua franca' (with a reference to Loog et al. 2017, which seems an inexact reference so far as this phrase is concerned), it might be better either to omit the phrase or to say 'the emergence of a *hypothetical* Celtic lingua franca', just to underline that the existence of any such a lingua franca is more hypothetical than the increase in mobility which it is here supposed to have facilitated. Sims-Williams (2020) cites the following paper as casting doubt on the importance of a lingua franca: Heggarty, P., 2017. Towards a (pre)history of linguistic convergence areas: correlates in genetics, archaeology, history and geography. *Bulletin de la Société de Linguistique de Paris* 24, 135-78.
4. In Table 2 'Hungary IA Celtic 500-1 CE' looks strange beside the other references which include no ethnic or linguistic descriptor. 'Celtic' is being used here as a traditional shorthand for 'La Tène Iron Age' or similar. This tradition in archaeological scholarship can cause endless misunderstandings among readers who do not realise that 'Celtic' generally does not have a provable linguistic or ethnic significance when applied to eastern European sites. Generally there is no linguistic information to back the 'Celtic' up (an exception in this paper is one of the sites in Slovakia from which come some arguably Celtic-language BIATEC coins). It might be better to omit 'Celtic' or to put it in scare quotes. This is a trivial point of course.

References

- Cunliffe, B., 2018. *The Ancient Celts*. Second edition. Oxford: Oxford University Press.
- Loog, L. et al. Estimating mobility using sparse data: Application to human genetic variation. *Proceedings of the National Academy of Sciences* 114, 12213 (2017).
- Nettle, D., 1999. *Linguistic Diversity*. Oxford: Oxford University Press.
- Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555, 190-196 (2018).
- Rodway, S. The Ogham Inscriptions of Scotland and Brittonic Pictish. *Journal of Celtic Linguistics* 21, 173-234 (2020).
- Sims-Williams, P. *An Alternative to 'Celtic from the East' and 'Celtic from the West'*. Cambridge

Referee #3 (Remarks to the Author):

This is clearly an important and significant paper, presenting an essential expanded data set of ancient DNA samples, primarily, but not solely from Britain. The paper uses this data set to argue for a large-scale migration into Britain during the Middle - Late Bronze Age, coupled with significant population increase. The originality of the data is unquestionable, and it has potentially serious significance. There is a divide here, however, between the significance of the individual pieces of data, and their collective interpretation, which I do not think is sufficiently robust, or critically informed.

The paper begins with a significant point, that there is a clear distinction between the manner in which archaeologists understand migration, and how it is used by geneticists. This is clear and important. Yet it swiftly becomes apparent that the paper is not content with describing genetic change, it wants to explain archaeological (and linguistic) shifts. This is a reasonable ambition - however, in so short a paper there is very little attempt to robustly engage with how these different scales of interpretation might be explained. Nor is there any recognition that it is not merely in archaeology that migration is a politically charged concept - but in general society as well. By immediately returning to the language of migration after this opening discussion (and so defaulting to the terms of one discipline over another, and refusing to address the ways in which this data will be deployed) the paper does its conclusion an enormous disservice. It also becomes clear, as the paper broadens its scope, that the large-scale migration of the title does mean the movement of people, and migration in the archaeological as well as genetic scope. E.g. by page 6 migration already refers again to 'the large-scale movement of people'. I would strongly urge that the authors develop a neutral term to describe the genetic shifts their work clearly demonstrates, and thus set out precisely where this can be linked to migration - in the archaeological sense - and where it cannot. This engagement in developing a term not carrying the archaeological or political baggage could also be added to by some acknowledgment of the extensive archaeological critique of these kinds of interpretation which the authors do not cite or engage with. It seems unlikely given the countless authors on the manuscript that none of them have come across this literature. A failure to engage with the critique, coupled with an insistence on using a complex and contested term, demonstrates a clear refusal to recognise how best to position this data to have the most significant and robust impact on archaeological and historical understandings.

It is on this archaeological side that the handling of the data lacks sufficient precision and care. To give some examples, what does the paper mean when it talks about large-scale migration? What does large-scale mean, over what time frame? The Middle-Late Bronze Age is a matter of centuries, what kind of transformation would this look like if it took place over one century, two or eight? Referring to specific time periods, drawing on the extensive radiocarbon evidence, would be useful. Similarly, the paper argues that an increased population can be revealed by evidence showing the number of reproducing individuals living in 'the last few dozen' generations. We need much greater consideration and precision here as a few dozen generations covers a gap from anything from 250-1500 years. This is not - to be clear - to critique the precision in the handling of the genetic data, but if the authors wish to provide historical-scale explanations it is precisely this they need to engage with, and where if no further precision should be implied they should say so.

The paper reads as though the historical models - the movement or lack of movement of people - have been decided on in advance, and because the genetic data cannot itself demonstrate this the authors attempt to hide behind a different version of the term migration; yet as noted what they really mean consistently emerges. The predisposition to have interpreted the data in advance is most clearly evidenced in the one area where the authors treat their data with a great deal of suspicion. This is where the evidence suggests that women may have been migrating. This is

unlikely, however, we are told, because that does not match the data for Saxon migration. This is a) a strange non-sequitur, and b) suggests a vision where the historical process of movement is already understood and merely needs to be revealed.

The paper would also benefit from a tighter engagement with the archaeological literature. The evidence for increasing pastoralism in the Late Bronze Age is interesting when coupled with the archaeological evidence for field systems in this period. How do these two things interact? How does their evidence for steady population in the Neolithic match with claims for significant population fluctuations in this period? What is the evidence for languages arriving in the Early Iron Age, and other than the presumption that only the arrival of people can spread a language, why is that no longer plausible in the authors' eyes? Other than a catchy claim bound to make headlines, the paper presents no evidence at all for the arrival of a new language at this point, and again their underlying presumption - that the data = new people, and that new people = new language is made clear.

This is clearly an important paper, but it repeats the mistakes that have been made in multiple ancient DNA papers, and are now widely commented on in the literature. If there is insufficient room to explore the nuance of interpretation in a Nature paper then the authors should limit themselves to what they can demonstrate clearly. If they wish to explore the (fascinating) historical implications of their data they need to be much more precise, and potentially choose a journal with a much higher word count.

Referee #4 (Remarks to the Author):

Review of "Large Scale Migration into Southern Britain at the End of the Bronze Age" by Patterson et al.

This paper presents an impressive number of ancient dna samples, mostly from Britain but also some continental samples, treated as reference. It performs a now quite standardized set of analyses in order to perform inference on the pre-history of Britain, presenting several interesting and important pieces of information that extend what we currently know. This focusses on the history of migration into Britain, but also presents an impressive selection result on lactase. Overall the results are compelling, broadly convincing, and of interest to a general audience.

My largest criticism is that the paper as a whole does not feel very cohesive and is rather unclear in places. I give several specifics below, but the nomenclature is not consistently used, the figures don't match, and it is very confusing to read. A related criticism is that the analysis lacks a coherent modelling framework, instead being composed of a large number of related (hence confusing) partial models. For an expert, these are not real issues as everything is broadly as it should be. However, for a general interest paper, more effort needs to go into explaining the motivation and spelling out what the models do and why.

Major points:

1. Modelling.

It would be really helpful to spell out your model, even if just in the supplement, of what is really going on here. You have published models for these, and are constructing qpAdm analyses as a consequence.

All presented results are in terms of EEF and WHG proportions, which are inherently limited in complexity. Some analyses are simply about describing these proportions whereas others are more general. Without reading several papers (where the models are defined, and qpAdm) it is impossible to understand qpAdm results. For example, the difference between Table 1 "distal

model" and "proximal model" is defined only in terms of population tuples, which have an intuitive explanation that is not provided.

In general: there are lots of populations, and complex arguments about what could have happened, that may be simply explained best by placing EEF estimates on them (e.g. Atlantic population, Ireland, etc).

Table 2: "Bonferroni correcting the lowest p-value for the 38 hypotheses tested". This would be correct if the intention were to perform conservative testing for rejecting H_0 . But here, the intention is inverted: we seek to find models in which we fail to reject H_0 . The test is not suitable for this and at best needs caveating. It is possible I misunderstood what was done here, as like many analyses it is not clearly spelt out.

l696 "We used ADMIXTURE to cluster individuals": but admixture is a mixture model, so how does this achieve clustering? How was K chosen? Can we see these results? Please provide details.

2. Writing clarity.

I've enumerated issues below, none of which are fatal alone but together make the paper inaccessible. I suspect these are mostly communication issues rather than actual scientific disagreements.

Figures:

Fig 1: Neolithic/Chalcolithic/EBA/MBA/LBA/IA are not defined in the figure and should be, e.g. using Fig 1c (which is not particularly interpretable with this amount of data and could be better served as a histogram plot.) Generally: we move between 4/5/6 such ages, which should be defined and depicted clearly throughout.

Fig 2: Is valuable but needs work. Use (a,b) notation for consistency. It would help if these plots were consistent within this figure; are differences solely due to QC filtering? The time scales and periods don't seem to match? How was (a) binned, apriori based on periods? Can we see the ranges of times? Colour usage in (right) is not defined; age scale colour not used; y axis association to the two overlaid plots is not clear. (left) has loads of space to define the periods on the graph whereas (right) doesn't define them in the legend either?

Figure 3: Why is there a red outlier individual at 2000BCE whose confidence interval overlaps the mean? I prefer the extended figure version of this. Would be very helpful to have the x-axes comparable (either by lining them up or by showing the ages).

Text:

Introduction: as discussed in regards to figures, when talking of "Late Bronze Age (LBA) and Iron Age (IA)", it helps to clearly define these since "iron age" etc can vary around the world.

The introduction spends a huge amount of time on the history of the definition of migration. This is surely not the way to start the paper? The following para does not even need it as it uses the traditional term. Either move this or make it part of the story.

Line 330 "The only plausible explanation for these changes in EEF ancestry is new migrations from mainland Europe into southern Britain" - seems plausible, but why is this not demonstrable using the model?(is it? is that what Extended Data Fig 3 is showing?) The evidence for this claim needs to be presented clearly. Your data doesn't cover Ireland (much) or Scandinavia at all. Are these ruled out due to the direction of the change? Could the persistence of people with different burial practice (i.e. cremation) with overlapping geography also be ruled out? (missing full stop)

Line 373: "In either case, the Archer's anomalously low Steppe ancestry is important in revealing that Beaker-using people who came to Britain were not socially stratified in such a way that Steppe lineage from eastern Europe necessarily conferred the highest social status." This is a confusing statement, that needs clarification. It's just one person with high status from the Alps (?). Is the companion also isotopically outside of the UK? What data goes into this claim?

MMN is defined but never used and we learn nothing about them except the age. Either explain the context (which I think could be helpful), or just leave off the end of this sentence.

Minor things:

Is the library prep (line 272) all standard? if so better to say less here and more in the methods? ("almost always pretreated" doesn't seem a useful statement...?)

Line 316: pretty boring to list the Scotland percentages present in Table 1 and Figure 2 anyway.

Extended data fig 5: state what the horizontal lines are.

Author Rebuttals to Initial Comments:

Referee #1 (Remarks to the Author):

This paper describes a large body of new ancient genomic data, mainly from Britain. This is a significant data legacy contribution - lessened somewhat as it is partial (captured genome) sequence rather than whole genome data but this is balanced by the fact that target coverage is very good in many specimens. The data is clearly carefully obtained, controlled for quality, and described in supplemental tables in good detail. One issue is that only aligned reads are submitted for retrieval - I am not clear why this is, rather than raw fastq data files? At least the alignment parameters and process should be made clear for other researchers who might use these data.

The deposition of raw data as aligned bam files has emerged as the *de facto* convention for reporting raw data in the ancient DNA literature. There are of course more raw forms of data that could be shared, including fastqs as well as photographic images of flow cells and BCLs. These take up more storage space and are thus harder to download and our experience is that most users have been interested in aligned bams so we were planning to follow the standard approach. If the journal would like us to release fastqs with publication, we could of course do that. We have edited the “Data Availability” section to clarify our approach and to specify that other forms of data are available from the authors upon request:

“The raw data are available as aligned sequences (bam files) through the European Nucleotide Archive under accession number [to be made available on publication]. Less processed versions of the data (fastq files) are available from the corresponding authors on request, whereas more processed versions of the data (the genotype data obtained by random allele sampling used in analysis) are available at <https://reich.hms.harvard.edu/datasets>. Any other relevant data are available from the corresponding authors upon reasonable request.”

We have also edited the Methods section to provide additional details of the alignment parameters and process.

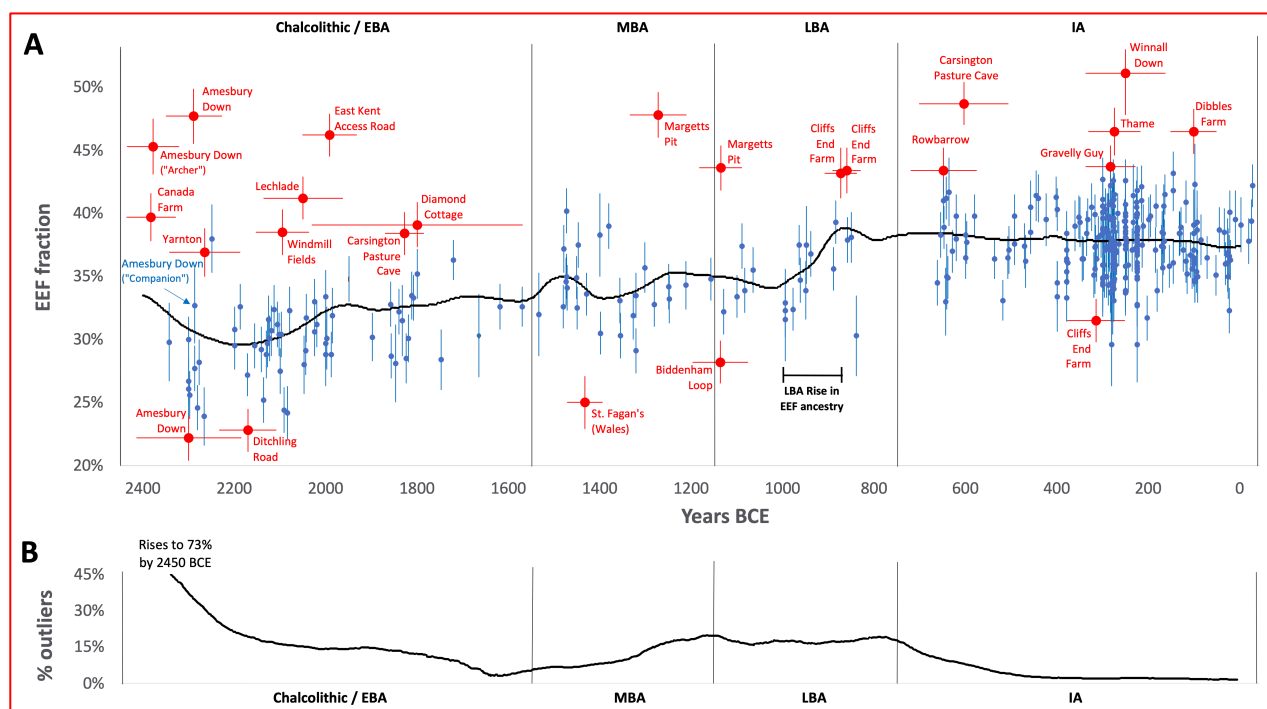
The major finding is evidence for a major migratory influence in Britain which alters the balance of ancestral components and which involved a minimum of 50% contribution from migrants during the Bronze Age. This is tied to M-LBA cultural connections between Britain and mainland Europe. I think that this is an exciting assertion, is additional to previously described migratory episodes, and would not be visible without ancient genomics. My opinion is that this finding merits publication in Nature. However, I have some queries and recommendations regarding the work as presented.

The work is very well presented with clear writing and informative figures. However, the central finding is a dramatic one and I think the reader should be given a closer look at the analyses in a more basic form, in order to judge some inferences.

Firstly, the central assertion that in a temporal transect there is a shift in Early European Farmer ancestry component in England and Wales is illustrated in fig 2a in four period bins, in fig 3 where individual measures are given with the same bins as average lines and in a temporally extended but similar supplement figure. The detail in fig 3 is commendable and a fair amount of

noise in this measure is clear. I wonder about timing of this shift in ancestry. The discussion in the paper involves inference around whether this is an episodic (eg M-LBA connections) or gradual, and also asserts little impact of migrants in the Iron Age. I wonder what temporal resolution there actually is within this phenomenon. Rather than use the same temporal bins for averaging as in fig 2, it could be good to use a more agnostic approach (eg sliding windows) in Fig 3 to ask what bins of similar EEF ancestry naturally emerge from the data. How granular is the analysis with regard to estimating the timing of EEF ancestry change and whether it is episodic or gradual? Note some specificity is claimed - line 567 “overwhelming evidence that major movements of people occurred during the M-LBA, particularly during a period of intensification of exchange that led to the ‘Match-Mer du Nord complex”

This is a good suggestion. We have revised Figure 3 to replace the period averages with a Gaussian-smoothed average of ancestry at each time point, weighting contributions of individuals both based on the precision of their ancestry estimate (weighted as one divided by the squared standard deviation), and the precision of their date determination (weighted by the amplitude of the posterior distribution of the date estimate at each evaluated time point). The resulting revised figure, excerpted below, suggests that the major rise in ancestry in our sample series occurred ~1000-875 BCE and leveled off after that point.



“Figure 3: By-individual analysis of the England and Wales time transect. (A) Estimates of EEF ancestry and one standard error for all individuals that fit a three-way admixture model (EEF + WHG + Yamnaya) at $p > 0.01$ using qpAdm and date to 2450 BCE–43 CE (we plot the mean of the posterior interval of the 14C date or the mean of the archaeological context range). Most individuals are shown in blue while significant outliers at the ancestry tails are in red with a horizontal error bar to show one standard error for the date (outliers are identified either as $p < 0.005$ based on a qpWave test from the main cluster from their period and $|Z| > 3$ for a difference in EEF proportion, or $p < 0.1$ and $|Z| > 3.5$) (Online Table 5). Outliers like the four MBA and LBA individuals at Margetts Pit and Cliffs End Farm fit as the source for the increase in EEF ancestry in the IA. The black line shows an estimate of population-wide EEF ancestry proportion at each time obtained by weighting each individual’s EEF estimate by the inverse square of their standard error as well as the probability that their date falls at each point in time (based on the date

mean and error given in Online Table 5; we filter out individuals with date errors >120 years). The incorporation of increased EEF ancestry into the majority of samples occurred ~1000-875 BCE. (B) Proportion of outliers measured over 300-year sliding windows centered on each point; to estimate this we randomly sample dates of all individuals 100 times assuming normality and their mean and standard deviation in Online Table 5 (we remove individuals with EEF errors >0.022 and date standard errors >120 years). Major periods of migration into Britain are evident as periods when elevated proportions of individuals are outliers: between 2450-1800 BCE (17% outliers) and 1300-750 BCE (17% outliers again). Our ability to temporally resolve the period of outliers is poorer than our ability to resolve the period of incorporation of increased EEF ancestry into the broader population; thus, for example, the earliest outlier with elevated EEF ancestry in the M-LBA at Margetts Pit 113716 could have died as late as 1129 BCE based on the uncertainty of her 14C date.

Second, and related, how sure are the authors that there is no major Iron Age migration that may have, eg, accompanied language influx or La Tene material culture? A shift is easier to detect when there is a higher differential in ancestral component between the donor and recipient population and less distinguishable when they are similar in composition. Given that the occurrence of high EEF outliers through the transect is reasonably constant (to the eye, but this is not examined as a proportion of samples) and that there is a stated evening out of European ancestral components among European regions in later periods, could the plateau in ancestry in the Iron Age be (at least in part) a plateau of detectability rather than stability of population?

This was a valuable suggestion. In the revised Figure 3 (excerpted above) we have addressed this issue by adding a second panel (B) which shows the proportion of significant ancestry outliers among sampled individuals in each period. This panel shows that outliers are observed at a high rate of 17% in the Chalcolithic and early part of the EBA (2450-1800 BCE), that the sampled population then homogenized with a low rate of outliers of 4% in the later part of the EBA and the early MBA (1800-1300 BCE); that the outlier rate rose again to 17% at the end of the MBA and LBA (1300-750 BCE); and that it fell again in the IA to a low level of 3%.

Indeed, homogenization of ancestry in northern and southern Europe by the LBA and IA in principle could make outlier detection in the IA in Britain more difficult. However, even in the IA, the ancestry of the great majority of Britain's neighbours was sufficiently different in many regions from that in Britain that we would have expected to have detected migrants relatively easily given the precision of our by-individual EEF ancestry estimates (<2% for most individuals). Thus, in Iberia EEF proportion was ~15% higher in this period, in France it was ~10% higher, and in the Netherlands it was ~5% lower. We describe this evidence and associated caveats in our revised manuscript as follows:

“[T]he fraction of individuals whose ancestry is significantly different from the main group is 17% over the first part of the C/EBA (2450-1800 BCE), much lower at 4% from the end of the EBA through the beginning of the MBA (1800-1300 BCE), high again at 17% between the end of the MBA through the LBA (1300-750 BCE), and low again at 3% through the IA (Figure 3). This is consistent with two major periods of migration into southern Britain in the Chalcolithic and then again in the M-LBA. We considered the possibility that our failure to observe a high rate of outliers in the IA compared with the immediately preceding period was because ancestry had homogenized to an extent between northern and southern Europe by this time, which could make outliers more difficult to detect. However, the average EEF ancestry in Britain in the IA was $37.9 \pm 0.4\%$, which is substantially different from much of contemporary Western and Central Europe— $52.6 \pm 0.6\%$ in Iberia, $49.8 \pm 0.4\%$ in Austria, Hungary, and Slovenia, $45.4 \pm 0.5\%$

in the Czech Republic, Slovakia and Germany, $45.6 \pm 0.5\%$ in France and Switzerland, and $34.4 \pm 1.2\%$ in the Netherlands (Figure 4A)—which would have made the majority migrants from these regions detectable given the small standard errors in most individual ancestry estimates of $<2\%$ (Online Table 5). Our sampling from western France and Belgium is poor, and it is possible that EEF ancestry proportions there were similar to Britain, so we cannot rule out migration from this region in the IA. Nevertheless, the lack of a change in EEF ancestry proportion in the IA is consistent with reduced migration from continental Europe and suggests relatively closed and self-sustaining social communities in Britain during the IA³⁰.”

Incidentally, some outliers seem odd, eg one just after Thanet ca. 2000 BC seems centrally located - also the third Carsington sample?”

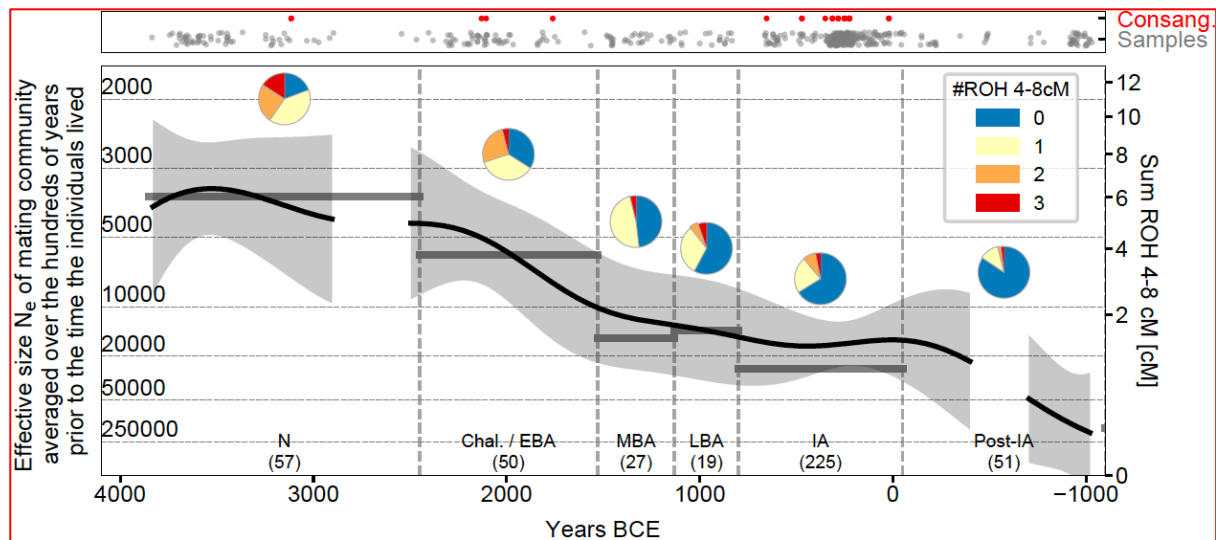
The original presentation of the outliers in Figure 3 had some visually odd (but not actually problematic) features. Individuals were labeled as outliers entirely based on a *qpWave* analysis that tested if they were significantly different in ancestry from the main cluster of individuals during their period with respect to differential relatedness to three outgroups: a pool of Yamnaya-like Steppe pastoralists, a pool of Early European Farmers (EEF), and a pool of Western Hunter Gatherers (WHG). The plot only shows proportion of EEF ancestry, but individuals can also be outliers if their EEF ancestry is typical for the period but the ratio of Yamnaya-like and WHG-like ancestry is not, as was the case of the two individuals the referee mentions. In the revised manuscript we have a new outlier labeling system for Figure 3, which is specified in the Figure 3 legend and focuses on EEF ancestry. It is explained as follows:

*“outliers are identified either as $p < 0.005$ based on a *qpWave* test from the main cluster from their period and $|Z| > 3$ for a difference in EEF proportion, or $p < 0.1$ and $|Z| > 3$.”*

Third, further evidence for the migratory influx is the shift in N_e calculated from ROH plotted in Fig 2b. As an independent aspect of the data this is important support. However as a reader it is impossible to see the trend in the smoothed line reflected in the raw ROH data from which it emerges. Could this be improved, eg by plotting median (or mean) ROH sums for each time bin?

We agree the figure in its originally presented form was hard to understand than it should have been and we have revisualized it. Our changes include:

- We now have relabeled the Y-axis to a much clearer (albeit verbose) “Effective size N_e of mating community averaged over the hundreds of years prior to the time the individuals lived” rather than “Effective population size N_e .” This new label makes it clear that the quantity being estimated should not be conflated with island-wide population size.
- We now plot the mean ROH in time bins, along with pie charts showing the fractions of individuals with 0/1/2/3 short ROH segments. This makes it easy to appreciate the striking genetic change over time.
- We now plot the ages of the individuals included in the ROH analysis



Legend excerpt: “(B) Detecting runs of homozygosity (ROH), we find that close kin unions were rare at all periods as reflected in the paucity of individuals harbouring >50 centimorgans (cM) of their genome in ROH segments of >12 cM (red dots in top panel). The number of ROH of size 4-8 cM per individual (bottom panel) reflects the rate at which distant relatives have children, providing information about the sizes of mate pools (N_e) averaged over the hundreds of years prior to when individuals lived; thus, the broad trend of an approximately four-fold drop in N_e from the Neolithic to the IA is robust, but we may miss fluctuations on a time scale of centuries. The thick black line is N_e from a Gaussian process with a 600-year smoothing kernel (gray area 95% confidence interval); horizontal lines show period averages from maximum likelihood; and we interrupt the fitted line for periods with too little data for accurate inference (<10 individuals in a 400-year interval centered on the point).”

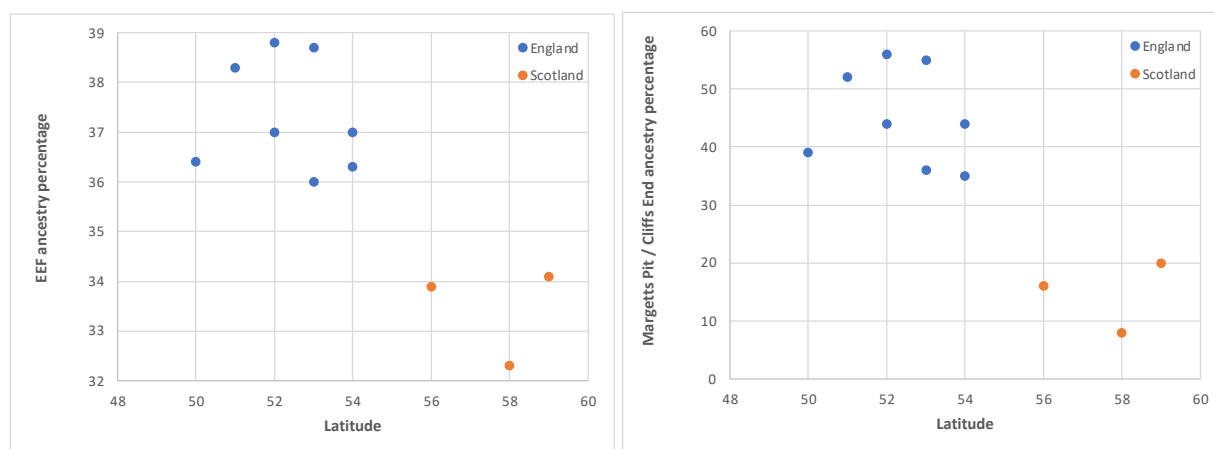
Fourth, an interesting aspect of the EEF ancestral component variation is that it is geographically distributed within Britain, primarily summarised as a Scotland Vs England/Wales contrast. I think it would be good to let the reader see this in some detail, similar to the temporal transect. Eg a plot of EEF ancestry for individuals jittered on a map of Britain could allow one to assess the nature of this - this would also allow the reader to easily judge the distinction of other outliers such as the East Yorkshire Arras samples.

In fact, the East Yorkshire Arras individuals do not stand out from their neighbors in terms of overall proportion of EEF ancestry or evidence of new migration from the continent (Table 1). They do stand out from the perspective of genetic drift (Extended Data Table 2). We clarify this in the following revised passage:

“We highlight the case of East Yorkshire, where most individuals we analysed are associated with ‘Arras Culture’ contexts comprising square-ditched barrows and occasional chariot burials^{40,41}; similarities between Arras funerary traditions and those of IA societies in the Paris Basin and the Ardennes/Champagne regions of France and Belgium have led to suggestions that East Yorkshire was influenced by direct migration from continental Europe in the IA⁴². Our estimate of the Margetts Pit/Ciffs End Farm ancestry source for East Yorkshire burials is $44 \pm 4\%$ (Table 1) which is typical for middle latitudes of Britain at this time (East Anglia is similar). However, the East Yorkshire burials are distinctive in another way: regional differentiation in IA Britain, as measured by F_{ST} , is higher between East Yorkshire and other groups than between any other pairs of IA

populations in England and Wales (Extended Data Table 2). Without ancient DNA from the putative IA sources for the Arras culture in continental Europe, we cannot determine if the genetic differentiation we observe of people in Arras burial contexts is attributable to the subsequent isolation in the IA of the population from the rest of southern Britain, or later streams of migration specifically affecting East Yorkshire that we have limited power to detect.”

The referee suggests plotting the Iron Age regional ancestry variation on a map. While there is indeed statistically significant differentiation in ancestry proportion within England and Wales in the IA as shown in Extended Data Table 2, the estimates are too noisy on an individual-specific level for a map to be helpful. For example, here are plots of latitude against ancestry proportion (we remove Wales because the sample sizes are so small that the inferences are extremely noisy). Within England, this does not reveal a visually apparent correlation.



We appreciate that the extent to which there is a real correlation within England will be a question of interest to the reader, however, and we have therefore added a “latitude” column to Table 1 to allow readers to appreciate how noisy the correlation to geography is. We also reference this in a new line in the main text:

“Within England, EEf ancestry proportions inferred over the mostly later IA individuals do not correlate to latitude (Table 1).”

Table 1: Regional variation in ancestry in Iron Age Britain

Region	N	Lat- it- ude	Modeling Ancestry With Pre-Bronze Age Sources			Modeling Ancestry with Middle to Late Bronze Age Sources	
			P- value	WHG	EEF	Steppe	P- value Margetts Pit and Cliffs End Farm-like ancestry
Scotland Orkney	2	59	0.22	14.2 ± 1.1%	34.1 ± 1.2%	51.6 ± 1.6%	0.10 20 ± 9%
Scotland West	4	58	0.12	13.0 ± .8%	32.3 ± 1.0%	54.7 ± 1.2%	0.19 8 ± 7%
Scotland Southeast	12	56	0.67	12.1 ± .6%	33.9 ± .7%	54.0 ± .9%	0.39 16 ± 5%
England North	10	54	0.35	13.4 ± .6%	36.3 ± .8%	50.3 ± 1.0%	0.76 35 ± 5%
England East Yorkshire	47	54	0.61	13.2 ± .4%	37.0 ± .5%	49.8 ± .6%	0.86 44 ± 4%

England Midlands	18	53	0.66	12.6 ± .5%	36.0 ± .6%	51.4 ± .8%	0.77	36 ± 4%
England Southwest	84	53	0.30	13.7 ± .4%	38.7 ± .4%	47.6 ± .6%	0.56	55 ± 5%
England East Anglia	21	52	0.44	13.5 ± .5%	37.0 ± .5%	49.5 ± .7%	0.52	44 ± 4%
England Southcentral	38	52	0.32	13.9 ± .4%	38.8 ± .5%	47.2 ± .6%	0.35	56 ± 5%
England Southeast	3	51	0.13	13.9 ± .5%	38.3 ± .5%	47.8 ± .6%	0.40	52 ± 5%
England Cornwall	16	50	0.40	13.5 ± .5%	36.4 ± .7%	50.1 ± .8%	0.64	39 ± 5%
Wales North	1	53	0.20	12.1 ± 1.6%	34.7 ± 2.0%	53.2 ± 2.5%	0.53	22 ± 14%
Wales South	2	51	0.66	14.2 ± 1.2%	38.6 ± 1.5%	47.2 ± 1.8%	0.57	53 ± 11%

Notes: Regions are ordered first by large grouping (Scotland-England-Wales), then reverse order of latitude. For modeling ancestry with pre-Bronze Age sources the right set is (OldAfrica, WHGA, Balkan_N, OldSteppe); for modeling of ancestry with M-LBA sources it is OldAfrica, OldSteppe, Turkey_N, Netherlands_C.EBA, Poland_Globular_Amphora, Spain.Portugal_4425.to.3800BP, CzechRepublic.Slovakia.Germany_3800.to.2700BP, Sardinia_8100.to.4100BP, CzechRepublic.Slovakia.Germany_4465.to.3800.BP, Sardinia_4100.to.2700BP, Spain.Portugal_6500.to.4425BP. We separate “England East Yorkshire” from “England North” because of East Yorkshire’s distinctive cultural context (Arras).

Minor points:

It is nice to see the Amesbury archer included, it might be good to also label his companion on fig 3

We have done this; see Figure 3 excerpted above.

Line 430: is the discussion of population size change as being constant followed by 5X expansions too literal? Ne is not census population and an influx of a population with different background could be a disruption of this statistic of a different sort to raw population size increase.

We agree the statement in the originally submitted manuscript could lead to misinterpretation, and have revised the manuscript to clarify. The quantity N_e we are estimating represents the effective sizes of mating communities, not census sizes. Our analyses reveal something not previously documented—that the effective size of mating communities increase by around four-fold from the Neolithic to the Iron Age among our sampled individuals from England and Wales—a type of finding that can only be revealed by a dataset of the magnitude reported in this study and so we think it is important to report. However, better contextualization was needed about what this does and does not mean. For example, are we estimating the population size of small communities, or the entire island of Britain, or an even larger region with which people in Britain were intensively exchanging mates? Another challenge for interpretation is that a high degree of local endogamy in a community (limited interaction with neighbors) can depress estimates of N_e even if the population density is high; this is plausible for the British Iron Age where there is also archaeological evidence of more localized cultural phenomena and limited long-distance interaction. We have clarified this in the revision by heavily modifying the section describing the results in a way that highlights the caveats:

“Demographic change in Britain over our period of study is also evident from another aspect of the data: the rate of observations of runs of homozygosity (ROH), which can occur when a person’s parents are relatively closely related. The larger the pool of people from which individuals draw their mates, the less likely it is for parents to be closely related, and thus we can average the number of 4-8 centimorgan (cM) ROH segments to estimate the effective size of the pool of people within which people were mating in the ~600 year period prior to the time when

the analysed individuals lived³¹. We find that the size of the mating pool increased by roughly four-fold from the Neolithic to the IA, but we caution that this inference should not be interpreted as an estimate of population size changes over this period as mating pool sizes are also affected by changing social customs. First, if the distance over which people ranged to find their mates was higher in some cultural contexts than in others it would cause the mating pool sizes to be different even if there was no difference in actual population size; for example, mating pool size may be less than the island-wide population size if members of communities mixed little with their neighbours as is plausible in the British IA where there was high regional variation in material culture³⁰, or larger if individuals mated not only with people outside their local communities but also outside Britain. Second, we have a major gap in sampling at the end of the Neolithic (roughly 3000-2450 BCE) and thin sampling at the beginning of the IA, which means that demographic processes in these periods may be missed. Third, due to the method effectively averaging over centuries, this analysis may also fail to detect major population declines over the space of a few decades, if these were followed by rapid population recovery. Future work that complements measurement of changing degree of parental relatedness over time with estimates of changing degree of relatedness of individuals across communities³² should make it possible to examine how much of this signal of increasing size over time was driven by true growth in population size, and how much by increasing rates of gene exchange across geographically dispersed communities.”

Such a migration does open a window for the influx of Celtic languages. However, how sure is the assertion of evidence of non-IE persistence in Scotland, which is used in support of this process? It is not my expertise but I understand a reasonable settling of opinion has occurred that Pictish is a Brittonic Celtic language? Is there a danger of picking isolated supporting evidence from another field?

Referee #2 who is indicated by the editor as an expert on Celtic languages suggests that our highlighting of the existence of non-Celtic elements in Scotland into the first millennium CE. They write: “as the authors rightly realise, would be consonant with the current, sober, judgment that many of the Pictish inscriptions of Scotland cannot be explained as Celtic (Rodway 2020).”

On reflection, though, we think that discussing Pictish by name is more confusing than helpful to readers, since as the referee states, many linguists now think Pictish was Celtic. The important point is that there is evidence for first millennium CE persistence of non-Celtic languages in Scotland (evident in non-Celtic elements in Pictish inscriptions) which could be consistent with our genetic finding of limited impact of the M-LBA migrations on northern Britain. Our genetic findings are thus fully consonant with an existing set of observations from historical linguistics. To avoid misinterpretation, we have revised our discussion of this to the following:

“While the lack of evidence for M-LBA EEF ancestry change in Scotland (Figure 2A) could be interpreted as a concern for the case that Celtic language spread into Britain at this time, in fact a later arrival of Celtic languages in Scotland is entirely consistent with evidence that non-Celtic and Celtic languages coexisted there into the first millennium CE⁵².”

I think the introductory review of migrations as an issue in archaeology is welcome. line 239 seems contradictory though, British WHG ancestry is stated as almost entirely due to admixture

in Mainland Europe indicating that British local Mesolithic people are absorbed.

We have now rephrased this sentence to clarify our meaning:

“The WHG ancestry in the first British farmers was almost entirely due to admixture in continental Europe, indicating that migrants in the Neolithic brought this WHG ancestry with them and local WHG contributed negligibly to the ancestry of later people in Britain.”

Referee #2 (Remarks to the Author):

On the basis of much new aDNA data from Bronze and Iron Age Europe, the authors argue that, subsequent to the so-called 'Beaker' migrations of circa 2450 BC (Olalde et al. 2018), there were, circa 1300-800 BC (in the Middle to Late Bronze Age), massive migrations of people from France into southern Britain (England + Wales). These were followed by a period after circa 800, in the pre-Roman Iron Age, with much less migration into Britain. The authors rightly note the desirability of more aDNA data from northern France and Ireland, but undoubtedly they have made an important advance which should be published as soon as possible.

The authors are careful not to specify the nature and context of their migrations, rightly avoiding a premature interdisciplinary synthesis at this stage. They do, however, cautiously note that migrations across the English Channel would suit the recent hypothesis (Sims-Williams 2020) that the Celtic language spread from France circa 1000 BC rather than spreading later, in the Iron Age, as was long assumed, or developing earlier in situ (from PIE) in about the fourth millennium BC (as in Cunliffe 2018 and earlier publications). Their view seems plausible. If so, the implication of their genetic data would seem to be that Celtic-speakers reached northern Britain relatively late. That, as the authors rightly realise, would be consonant with the current, sober, judgment that many of the Pictish inscriptions of Scotland cannot be explained as Celtic (Rodway 2020).

The question of how Celtic got Ireland is not addressed. That does not seem to be an insuperable problem, however. If it does not turn out that there was a parallel migration from the Continent to Ireland, comparable to that to southern Britain, one could get round that by hypothesising that Ireland, like other peripheral areas (e.g. the Iberian peninsula and perhaps Scotland), had a large number of competing minor/relic languages; if so, a relatively small number of incoming Celtic-speakers (direct from the Continent and/or via Britain), coming in under the archaeological and genetic radar, might be enough to effect the language shift to Celtic that is first attested in the Irish place-names recorded by Ptolemy circa 150 A.D. Sims-Williams (2020) quotes Nettle (1999) as saying that 5000 speakers is the median number of speakers for languages worldwide. Supposing that was the situation in Ireland, even 20,000 Celtic-speaking incomers could have a decisive effect in favour of a language shift to Celtic. The same could apply in Britain, as hypothesised by Sims-Williams, in which case the eventual prevalence of Celtic would not *have* to be due to the 'large scale migration' detected by the authors. Nevertheless, their explanation looks attractively straightforward and deserves publication.

Some minor points:

1. Given that the media tend to misunderstand and sensationalise, the singular 'migration' in the title and abstract may give the unintended impression of an event as opposed to a process. This would be regrettable, as the authors are careful in the text to point out. Would the plural 'Migrations' be better in the title?

We have changed the title to “Large-Scale Migration into Southern Britain During the Middle to Late Bronze Age,” using the word “during” rather than “in” to convey our finding that what we have detected was spread out over hundreds of years, and thus could not have been a singular

pulse. We have revised the abstract to emphasize that this migration was a process that while transformative demographically unfolded over an extended period of time:

“Between 1000-875 BCE, EEF ancestry increased in southern Britain (England and Wales) but not in northern Britain (Scotland), due to incorporation into the population of a stream of migrants who arrived at this time and in previous centuries and who were genetically most similar to ancient individuals from France. These migrants cumulatively contributed about half the ancestry of Iron Age people of England and Wales, thereby documenting a previously unknown demographic process that is a plausible vector for the spread of early Celtic languages into Britain.”

2. In connection with the Pictish point above: it would be best not to say 'a non-Celtic language *remained in use into* the first millennium CE in Scotland'. It would be safer to say 'a non-Celtic language *was in use in* the first millennium CE in Scotland'. We can't be 100% sure that Pictish wasn't introduced from elsewhere early in the first millennium.

We have made this change, now writing:

“While the lack of evidence for M-LBA EEF ancestry change in Scotland (Figure 2A) could be interpreted as a concern for the case that Celtic language spread into Britain at this time, in fact a later arrival of Celtic languages in Scotland is entirely consistent with evidence that non-Celtic and Celtic languages coexisted there into the first millennium CE⁵².”

3. Instead of 'facilitated by the emergence of a Celtic lingua franca' (with a reference to Loog et al. 2017, which seems an inexact reference so far as this phrase is concerned), it might be better either to omit the phrase or to say 'the emergence of a *hypothetical* Celtic lingua franca', just to underline that the existence of any such a lingua franca is more hypothetical than the increase in mobility which it is here supposed to have facilitated. Sims-Williams (2020) cites the following paper as casting doubt on the importance of a lingua franca: Heggarty, P., 2017. Towards a (pre)history of linguistic convergence areas: correlates in genetics, archaeology, history and geography. Bulletin de la Société de Linguistique de Paris 24, 135-78.

We have eliminated the sentence.

4. In Table 2 'Hungary IA Celtic 500-1 CE' looks strange beside the other references which include no ethnic or linguistic descriptor. 'Celtic' is being used here as a traditional shorthand for 'La Tène Iron Age' or similar. This tradition in archaeological scholarship can cause endless misunderstandings among readers who do not realise that 'Celtic' generally does not have a provable linguistic or ethnic significance when applied to eastern European sites. Generally there is no linguistic information to back the 'Celtic' up (an exception in this paper is one of the sites in Slovakia from which come some arguably Celtic-language BIATEC coins). It might be better to omit 'Celtic' or to put it in scare quotes. This is a trivial point of course.

We no longer use the term “Celtic” in any group names.

References

- Cunliffe, B., 2018. *The Ancient Celts*. Second edition. Oxford: Oxford University Press.
- Loog, L. et al. Estimating mobility using sparse data: Application to human genetic variation. *Proceedings of the National Academy of Sciences* 114, 12213 (2017).
- Nettle, D., 1999. *Linguistic Diversity*. Oxford: Oxford University Press.
- Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555, 190-196 (2018).
- Rodway, S. The Ogham Inscriptions of Scotland and Brittonic Pictish. *Journal of Celtic Linguistics* 21, 173-234 (2020).
- Sims-Williams, P. An Alternative to 'Celtic from the East' and 'Celtic from the West'. *Cambridge Archaeological Journal* 30, 511-529 (2020).

Referee #3 (Remarks to the Author):

This is clearly an important and significant paper, presenting an essential expanded data set of ancient DNA samples, primarily, but not solely from Britain. The paper uses this data set to argue for a large-scale migration into Britain during the Middle - Late Bronze Age, coupled with significant population increase. The originality of the data is unquestionable, and it has potentially serious significance. There is a divide here, however, between the significance of the individual pieces of data, and their collective interpretation, which I do not think is sufficiently robust, or critically informed.

The paper begins with a significant point, that there is a clear distinction between the manner in which archaeologists understand migration, and how it is used by geneticists. This is clear and important. Yet it swiftly becomes apparent that the paper is not content with describing genetic change, it wants to explain archaeological (and linguistic) shifts. This is a reasonable ambition - however, in so short a paper there is very little attempt to robustly engage with how these different scales of interpretation might be explained. Nor is there any recognition that it is not merely in archaeology that migration is a politically charged concept - but in general society as well. By immediately returning to the language of migration after this opening discussion (and so defaulting to the terms of one discipline over another, and refusing to address the ways in which this data will be deployed) the paper does its conclusion an enormous disservice. It also becomes clear, as the paper broadens its scope, that the large-scale migration of the title does mean the movement of people, and migration in the archaeological as well as genetic scope. E.g. by page 6 migration already refers again to 'the large-scale movement of people'. I would strongly urge that the authors develop a neutral term to describe the genetic shifts their work clearly demonstrates, and thus set out precisely where this can be linked to migration - in the archaeological sense - and where it cannot. This engagement in developing a term not carrying the archaeological or political baggage could also be added to by some acknowledgment of the extensive archaeological critique of these kinds of interpretation which the authors do not cite or engage with. It seems unlikely given the countless authors on the manuscript that none of them have come across this literature. A failure to engage with the critique, coupled with an insistence on using a complex and contested term, demonstrates a clear refusal to recognise how best to position this data to have the most significant and robust impact on archaeological and historical understandings.

We appreciate the referee bringing up this point which touches on a topic that is one of the areas where archaeologists and geneticists often have misunderstandings: the use of the term “migration.” This is why we confront the issue head-on in our first paragraph, which now reads:

“Archaeologists generally use “migration” to refer to processes that result in significant demographic change due to directed and permanent translocation of people from one region to another^{2,3}. In European archaeology, discussions of prehistoric migrations have become fraught due to the ways in which theories of migration were exploited for political purposes in the early-mid twentieth century—when movement of large numbers of people over a short period of time was sometimes argued to be a primary mechanism for the spread of ethnic groups⁴ and archaeological reconstructions of such events were used to justify claims on territory⁵.”

However, a large number of archaeologists, including in our view the great majority of British archaeologists (and the more than hundred archaeologist co-authors of this manuscript), believe that “migration” is an important term whose use needs to be maintained in the archaeological literature in order to convey a broad set of processes including large-scale movements of people that take place possibly over longer periods of time and may have had crucially important roles in prehistory. The fact that this is a widely held view explains why “migration” continues to be the standard term used in the archaeological literature to describe processes like the large-scale streams of gene flow into Britain that have been documented in association with the start of the Neolithic and Chalcolithic periods, as well as early medieval movements of people. Stating that the term “migration” should be avoided in order to describe such processes would be, using David Anthony’s analogy in his 1990 “Migration in Archaeology: The Baby and the Bathwater,” throwing away an important and logically coherent explanatory framework because it has been politically misused in a particular context (by Nazi propagandists using archaeological theories of migration to justify claims on territory). Because of this, the referee’s assertion that “there is a clear distinction between the manner in which archaeologists understand migration, and how it is used by geneticists,” conveys an inaccurate sense of homogeneity in how archaeologists understand “migration.” This is why in our first paragraph we felt it was important to explicitly discuss what we believe is the primary understanding that most archaeologists have about what “migration” means in the discipline of archaeology, and to be explicit that our findings match the broader definition but should not be misunderstood as evidence for a movement that occurred in a short period of time since our data indicate the events unfolded over hundreds of years:

“Setting a high bar for theorizing migration, however—for example by only using it when there is clear evidence for organized movements of people over a short period of time—minimizes the important effects that cumulatively large-scale movements are likely to have had in shaping prehistory², such as the westward migration of people from the Steppe beginning in the third millennium BCE that genetic data have unambiguously shown contributed much of the ancestry of later Europeans including eventually people from Britain^{6,7,8,9}. We emphasize that the large-scale Middle to Late Bronze Age migrations we document here unfolded over hundreds of years as attested by our radiocarbon dating data and so cannot all be explained by movement over a short time.”

To further head off misunderstanding from readers who might be coming from perspectives similar to the referee, we decided to modify our title from the original to “**Large-Scale Migration into Southern Britain During the Middle to Late Bronze Age**”. We changed “In” to During” to indicate that we are describing a process that took place over an extended period of centuries and should not be interpreted as a singular pulse event.

After extensive discussion among our co-authors in response to the Referee’s comments, we have chosen to retain the use of the term “migration”, but always carefully explain the use of the term to make it clear that we are using the term in the broad sense in which it is widely used in the archaeological literature—a period of intensified movement of people from one region to another—and not to describe a more abrupt migratory movement. We appreciate the critical feedback that has helped us to create a revised manuscript that most effectively conveys our findings to as wide a group of readers as possible.

It is on this archaeological side that the handling of the data lacks sufficient precision and care. To give some examples, what does the paper mean when it talks about large-scale migration? What does large-scale mean, over what time frame? The Middle-Late Bronze Age is a matter of centuries, what kind of transformation would this look like if it took place over one century, two or eight? Referring to specific time periods, drawing on the extensive radiocarbon evidence, would be useful.

In our revised manuscript we are everywhere careful to indicate our degree of uncertainty about the time-scale of the processes we describe occurred.

We have also revised the title to make it clear that we are describing is a process that took place over an extended period of centuries and should not be interpreted as a singular pulse event: “Large-Scale Migration into Southern Britain During the Middle to Late Bronze Age” (we changed “In” to During”).

Similarly, the paper argues that an increased population can be revealed by evidence showing the number of reproducing individuals living in 'the last few dozen' generations. We need much greater consideration and precision here as a few dozen generations covers a gap from anything from 250-1500 years. This is not - to be clear - to critique the precision in the handling of the genetic data, but if the authors wish to provide historical-scale explanations it is precisely this they need to engage with, and where if no further precision should be implied they should say so.

The previous phrasing could be interpreted as implying that we might be talking about a time scale of up to ~50 generations which is more than is in fact consistent with our inferences. There is indeed uncertainty in the time frame over which we are estimating effective size of the mating community using our current genetic methodology and it is important to be extremely clear about this, and in our revised manuscript we now spell this this out, writing:

“The larger the pool of people from which individuals draw their mates, the less likely it is for parents to be closely related, and thus we can average the number of 4-8 centimorgan (cM) ROH segments to estimate the effective size of the pool of people within which people were mating in the ~600 year period prior to the time when the analysed individuals lived³¹.... [Another caveat is that] we have a major gap in sampling at the end of the Neolithic (roughly 3000-2450 BCE) and thin sampling at the beginning of the IA, which means that demographic processes in these periods may be missed. [Another caveat is that] due to the method effectively averaging over centuries, this analysis may also fail to detect major population declines over the space of a few decades, if these were followed by rapid population recovery.”

The paper reads as though the historical models - the movement or lack of movement of people - have been decided on in advance, and because the genetic data cannot itself demonstrate this the authors attempt to hide behind a different version of the term migration; yet as noted what they really mean consistently emerges. The predisposition to have interpreted the data in advance is most clearly evidenced in the one area where the authors treat their data with a great deal of suspicion. This is where the evidence suggests that women may have been migrating. This is unlikely, however, we are told, because that does not match the data for Saxon migration. This is

a) a strange non-sequitur, and b) suggests a vision where the historical process of movement is already understood and merely needs to be revealed.

We do not see where in our manuscript it sounds as if we have decided on demographic models in advance. In the narrative of our manuscript, we have attempted to be faithful to the way our research unfolded. The history of our research which indeed we discuss at the beginning of the manuscript was that—motivated by a previous finding that today southern Britons have more EEF ancestry than northern Britons or than Early Bronze Age Britons and that this cannot be explained by Saxon immigration (Galinsky et al. AJHG 2016)—we hypothesized that there must have been a period from the Middle Bronze Age onward when this increase in EEF ancestry occurred. Because of this hypothesis, we set out to collect a time-transect of samples to detect this change. Our paper is about how we can use ancient DNA data to narrow down this period of change to the end of the Bronze Age.

Regarding sex-biased migration and comparison to the Saxon evidence, we have removed this discussion as the evidence for sex-bias was on reflection weaker than the evidence we used to support the other findings in the study. The revised manuscript now only emphasizes results for which there is compelling genetic data.

The paper would also benefit from a tighter engagement with the archaeological literature. The evidence for increasing pastoralism in the Late Bronze Age is interesting when coupled with the archaeological evidence for field systems in this period. How do these two things interact?

Our finding that the lactase persistence allele rose to appreciable frequency about a millennium earlier in Britain than in central Europe has many possible interpretations, only one of which is a greater reliance on pastoralism in Britain than in Central Europe in the LBA and IA. We have removed from the manuscript all statements that could be read as suggesting that our results imply a greater reliance on pastoralism in Britain than in the continent in the Iron Age; for example, the relevant sentence in the abstract has now changed to:

“The distinct genetic trajectories of Britain and continental Europe in the Iron Age are also exemplified by the fact that the allele conferring lactase persistence rose to ~50% frequency in Britain by this time whereas it was only ~7% in Central Europe and underwent a comparable rise in frequency only a millennium later, a pattern that could only occur if there were qualitative differences in how dairy products were used in Britain and in Central Europe.”

It is not clear to us that any part of the archaeological literature has advanced hypotheses for or against more reliance in pastoralism in Britain than in Central Europe at this time, so we feel it would require us to carry out new critical interpretation of the archaeological data in order to make such a connection to the archaeological literature, something that would be interesting but is beyond the scope of this paper. We think that our framing our result in terms of a new question it raises for archaeologists to explore further is the best approach here, thus writing:

“This raises questions about whether dairy products were consumed in a qualitatively different way or were economically more important in LBA-IA Britain than in Central Europe.”

How does their evidence for steady population in the Neolithic match with claims for significant population fluctuations in this period?

This is an important point. To address this in our revision, we have now added caveats:

“Second, we have a major gap in sampling at the end of the Neolithic (roughly 3000-2450 BCE) and thin sampling at the beginning of the IA, which means that demographic processes in these periods may be missed. Third, due to the method effectively averaging over centuries, this analysis may also fail to detect major population declines over the space of a few decades, if these were followed by rapid population recovery.”

What is the evidence for languages arriving in the Early Iron Age, and other than the presumption that only the arrival of people can spread a language, why is that no longer plausible in the authors' eyes? Other than a catchy claim bound to make headlines, the paper presents no evidence at all for the arrival of a new language at this point, and again their underlying presumption - that the data = new people, and that new people = new language is made clear.

Referee #2 who is described by the editor as an expert on Celtic languages is clear that our genetic findings are relevant to debates about the origin of Celtic languages in Britain, and affirms that our discussion of these topics is well-justified: “The authors are careful not to specify the nature and context of their migrations, rightly avoiding a premature interdisciplinary synthesis at this stage. They do, however, cautiously note that migrations across the English Channel would suit the recent hypothesis (Sims-Williams 2020) that the Celtic language spread from France circa 1000 BC rather than spreading later, in the Iron Age, as was long assumed, or developing earlier in situ (from PIE) in about the fourth millennium BC”.

We have also been consulting with independent linguist experts outside the review process, and they too feel that our discussion of the linguistic implications is appropriate (these linguistic experts, Guus Kroonen and John Koch, are now explicitly mentioned in our Acknowledgments section). Nevertheless, to head-off misunderstanding from readers coming from perspectives similar to the referee, we are now even more explicit that the documentation of movement of people should never be seen as proof of spread of languages

“Population movements are often a significant driver of cultural change, including in the languages people speak. While periods of intense migration such as the one we infer here do not always result in language shifts³⁵, genetic evidence of significant migration is important because it documents demographic processes that are plausible conduits for language spread^{48,49}.”

Finally, we emphasize that we are not presenting a new hypothesis, and instead are simply stating that our data would support one of the existing hypotheses proposed (from historical linguistics) for when early Celtic languages arrived in Britain. Our genetic data add weight to this scenario but do not prove it, as we make clear in our revised manuscript.

This is clearly an important paper, but it repeats the mistakes that have been made in multiple ancient DNA papers, and are now widely commented on in the literature. If there is insufficient room to explore the nuance of interpretation in a Nature paper then the authors should limit

themselves to what they can demonstrate clearly. If they wish to explore the (fascinating) historical implications of their data they need to be much more precise, and potentially choose a journal with a much higher word count.

We appreciate the invitation to provide a clearer integration with the most relevant processes in this period known from the archaeological literature. We have revised the paragraph where we contextualize the genetic findings in light of the archaeological literature as follows:

“The period from 1500-1150 BCE has long been recognized as a time when cultural connections between Britain and regions of continental Europe intensified, and when societies on both sides of the Channel shared cultural features including domestic pottery, metalwork and ritual depositional practices⁴³⁻⁴⁷. From around 750 BCE there is more limited archaeological evidence of contact between Britain and the continent, with little that would suggest the significant movement of people⁴³. Both the genetic and archaeological data concur in showing that, by the beginning of the IA, there is little evidence of demographically significant migration into Britain. Our findings do not establish whether the population movements we infer were a cause or consequence of M-LBA exchange networks, but they do suggest that interactions between local populations of Britain and new migrants bringing ideas from continental Europe could have been a vector for some of the cultural change we see in M-LBA England and Wales. Much of northern and central France is currently almost entirely unsampled, and thus we cannot at present test if the gene flow between the two regions in this period was largely unidirectional.

Referee #4 (Remarks to the Author):

Review of "Large Scale Migration into Southern Britain at the End of the Bronze Age" by Patterson et al.

This paper presents an impressive number of ancient dna samples, mostly from Britain but also some continental samples, treated as reference. It performs a now quite standardized set of analyses in order to perform inference on the pre-history of Britain, presenting several interesting and important pieces of information that extend what we currently know. This focusses on the history of migration into Britain, but also presents an impressive selection result on lactase. Overall the results are compelling, broadly convincing, and of interest to a general audience.

My largest criticism is that the paper as a whole does not feel very cohesive and is rather unclear in places. I give several specifics below, but the nomenclature is not consistently used, the figures don't match, and it is very confusing to read. A related criticism is that the analysis lacks a coherent modelling framework, instead being composed of a large number of related (hence confusing) partial models. For an expert, these are not real issues as everything is broadly as it should be. However, for a general interest paper, more effort needs to go into explaining the motivation and spelling out what the models do and why.

Major points:

1. Modelling.

It would be really helpful to spell out your model, even if just in the supplement, of what is really going on here. You have published models for these, and are constructing qpAdm analyses as a consequence.

All presented results are in terms of EEF and WHG proportions, which are inherently limited in complexity. Some analyses are simply about describing these proportions whereas others are more general. Without reading several papers (where the models are defined, and qpAdm) it is impossible to understand qpAdm results.

We now present an intuitive description of *qpAdm* in the main text and also how we have optimized *qpAdm* modeling in this paper to provide improved precision in estimates of the three primary ancestry sources in Britain, writing:

“To estimate the proportions of EEF, Steppe, and WHG ancestry in any “Target” population²⁶, we leveraged the fact that the magnitudes of f_4 -statistics can be interpreted in terms of mixture fractions. Specifically, *qpAdm* models the f_4 -statistics in each “Target” population as a linear combination of the f_4 -statistics in populations we use as proxies for the mixing “Sources”. *qpAdm* computes statistics of the form $f_4(\text{Target}, \text{Source}_i; \text{Outgroup}_j, \text{Outgroup}_k)$ that compare Target and Sources to a set of four “Outgroups” that we optimized to provide more precise estimates of ancestry proportions than previous *qpAdm* setups (Supplementary Information section 2). Here, our proxies for the “Sources” are 22 early Balkan Neolithic farmers with

minimal hunter-gatherer admixture (EEF); 20 Yamnaya and Poltavka pastoralists (Steppe); and 18 Mesolithic hunter-gatherers from across Western Europe (WHG).”

For example, the difference between Table 1 "distal model" and "proximal model" is defined only in terms of population tuples, which have an intuitive explanation that is not provided.

We now no longer use the terms “distal model” and “proximal model” which are unnecessary jargon in the context of this paper. Table 1 now describes these models as “Modeling Ancestry With Pre-Bronze Age Sources” and “Modeling Ancestry with Middle to Late Bronze Age Sources”.

In general: there are lots of populations, and complex arguments about what could have happened, that may be simply explained best by placing EEF estimates on them (e.g. Atlantic population, Ireland, etc).

We have tried to keep the explicit mentioning of populations to a minimum except where mentioning those populations adds meaningfully to our argument.

Table 2: "Bonferroni correcting the lowest p-value for the 38 hypotheses tested". This would be correct if the intention were to perform conservative testing for rejecting H_0 . But here, the intention is inverted: we seek to find models in which we fail to reject H_0 . The test is not suitable for this and at best needs caveating. It is possible I misunderstood what was done here, as like many analyses it is not clearly spelt out.

We no longer reference a Bonferroni correction in Table 2.

We do describe a Bonferroni-style correction as part of our narrowing down of potential source populations in the revised Supplementary Information section 6. The revised description in which we discuss this approach in an appropriately careful way as follows:

“We wished to reduce the length of this list and therefore carried out a follow-on analysis where for each of the candidate source populations for the ancestry shift with $P > 0.05$, we tested 69 populations from Europe, North Africa or the Levant dating to older than 1000 BCE, in each case carrying out a *qpAdm* analysis in which the basic set of 11 *Right* populations was supplemented with the additional population. We performed a total of $1311 = 19 \times 69$ *qpAdm* runs after merging the Margetts Pit and Cliffs End individuals into a group we called “MargettsPit_CliffsEnd_MLBA” ($n=4$) that increased power to study the possible proximate source for the ancestry shift in LBA and IA Britain. For each of the 19, we recorded the lowest P-value (P_{lowest}) and applied a conservative Bonferroni-style correction for 69 hypotheses tested: $P_{\text{corrected}} = 1 - (1 - P_{\text{lowest}})^{69}$. Ten populations have $P_{\text{corrected}} > 0.05$ (Table S6.2).”

l696 "We used ADMIXTURE to cluster individuals": but admixture is a mixture model, so how does this achieve clustering? How was K chosen? Can we see these results? Please provide details.

We do not use the ADMIXTURE software in any of the analyses in the current version of the manuscript (this passage referred to a figure included in early drafts). Thus, we have removed mention of ADMIXTURE from the manuscript.

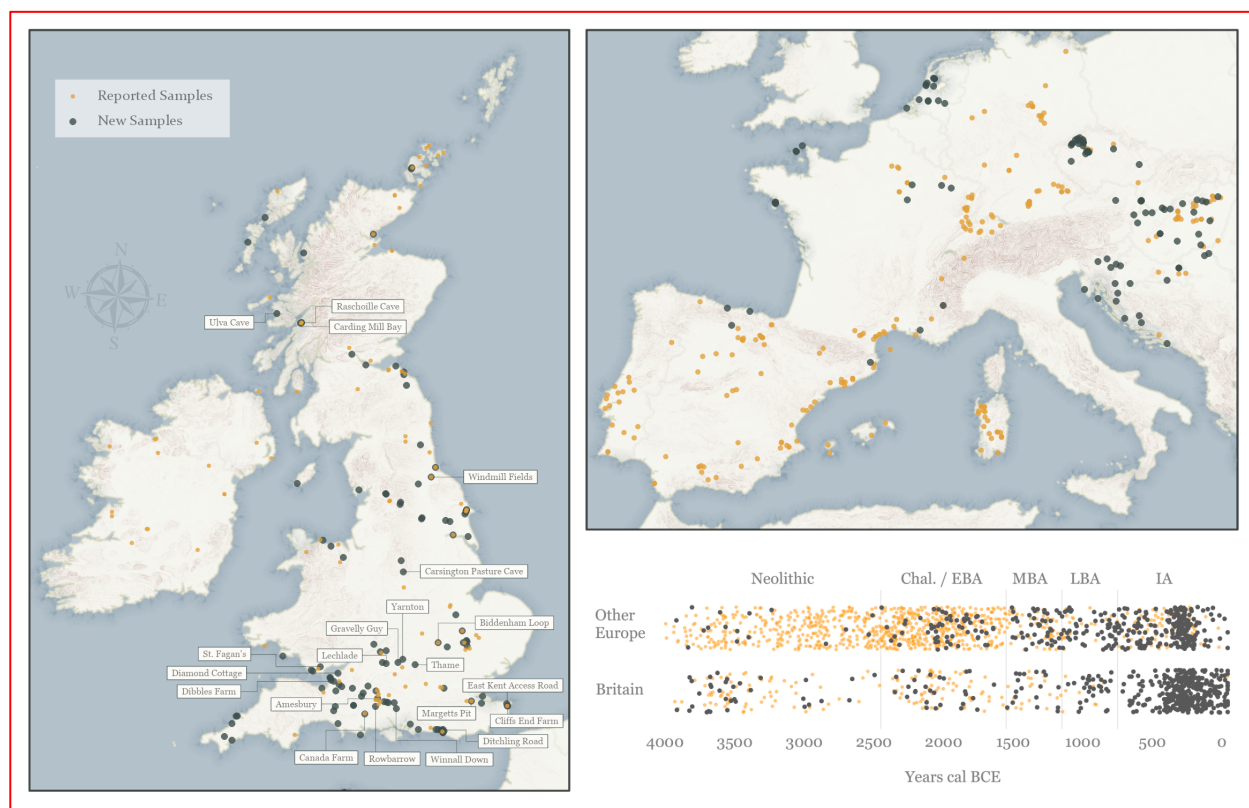
2. Writing clarity.

I've enumerated issues below, none of which are fatal alone but together make the paper inaccessible. I suspect these are mostly communication issues rather than actual scientific disagreements.

Figures:

Fig 1: Neolithic/Chalcolithic/EBA/MBA/LBA/IA are not defined in the figure and should be, e.g. using Fig 1c (which is not particularly interpretable with this amount of data and could be better served as a histogram plot.) Generally: we move between 4/5/6 such ages, which should be defined and depicted clearly throughout.

We modified Figure 1 to explicitly show the periods being discussed, as excerpted below. We think the showing of the distribution of samples over time with dots is clear with respect to the density of new and previously reported individuals in different time periods. We also like how the revised figure ties the timeline directly to the map (via shared color coding and use of dots).



“Figure 1: Ancient DNA Dataset. Geographic distribution of sites and temporal distribution of individuals in the period 4000 BCE-43 CE. Newly reported data are in black; published data are in orange. In the map of Britain we label sites that harbour individuals who are ancestry outliers relative to others of the same period. In the timeline we denote the archaeological periods according to the British chronology: Neolithic (3950-2450 BCE), Chalcolithic and Early Bronze Age (C/EBA, 2450-1550 BCE),

Middle Bronze Age (MBA, 1550-1150 BCE), Late Bronze Age (LBA, 1150-750 BCE), and pre-Roman Iron Age (IA, 750 BCE-43 CE). To aid visualisation, we add jitter on the Y axis and randomly sample dates from their probability distributions given in Online Table 1."

Fig 2: Is valuable but needs work. Use (a,b) notation for consistency.

This is now done.

It would help if these plots were consistent within this figure; are differences solely due to QC filtering?

The plots are now consistent.

The time scales and periods don't seem to match?

The time scales and periods now match.

How was (a) binned, apriori based on periods? Can we see the ranges of times?

In the main text and Figure 1 legend, we now specify what we mean by the time periods. For example, in the Figure 1 legend we write:

"In the timeline we denote the archaeological periods according to the British chronology: Neolithic (3950-2450 BCE), Chalcolithic and Early Bronze Age (C/EBA, 2450-1550 BCE), Middle Bronze Age (MBA, 1550-1150 BCE), Late Bronze Age (LBA, 1150-750 BCE), and pre-Roman Iron Age (IA, 750 BCE-43 CE)."

We also explicitly show these periods in the timeline of Figure 1, in Figure 2A, in Figure 2B and in Figure 3.

Colour usage in (right) is not defined; age scale colour not used; y axis association to the two overlayed plots is not clear. (left) has loads of space to define the periods on the graph whereas (right) doesn't define them in the legend either?

We have redone panel B (formerly the right panel), and it is now much clearer.

Figure 3: Why is there a red outlier individual at 2000BCE whose confidence interval overlaps the mean? I prefer the extended figure version of this. Would be very helpful to have the x-axes comparable (either by lining them up or by showing the ages).

We have redone this figure in response to feedback both from this referee and Referee #1. Repeating our response to Referee #1's comment on this point, the original presentation of the outliers in Figure 3 had some visually odd (but not actually problematic) features. Individuals were labeled as outliers entirely based on a *qpWave* analysis that tested if they were significantly different in ancestry from the main cluster of individuals during their period with respect to differential relatedness to three outgroups: a pool of Yamnaya-like Steppe pastoralists, a pool of Early European Farmers (EEF), and a pool of Western Hunter Gatherers (WHG). The plot only

shows proportion of EEF ancestry, but individuals can also be outliers if their EEF ancestry is typical for the period but the ratio of Yamnaya-like and WHG-like ancestry is not, as was the case of the two individuals the referee mentions. In the revised manuscript we have a new outlier labeling system for Figure 3, which is specified in the Figure 3 legend and focuses on EEF ancestry. It is explained as follows:

“outliers are identified either as $p < 0.005$ based on a qpWave test from the main cluster from their period and $|Z| > 3$ for a difference in EEF proportion, or $p < 0.1$ and $|Z| > 3$.”

Text:

Introduction: as discussed in regards to figures, when talking of "Late Bronze Age (LBA) and Iron Age (IA)", it helps to clearly define these since "iron age" etc can vary around the world.

We now explicitly discuss the time periods we are referring to in the text. In the Figure 4 legend, we note the challenge of the chronological period boundaries differing across regions, writing:

“we use the N/C/EBA/MBA/LBA/IA division into archaeological periods for Britain, but avoid using it for most of the rest of Europe because chronological boundaries differ”

The introduction spends a huge amount of time on the history of the definition of migration. This is surely not the way to start the paper? The following para does not even need it as it uses the traditional term. Either move this or make it part of the story.

Referees #1 and #3 agreed with us that addressing archaeological and genetic definitions of the meaning of migration head-on is a welcome aspect of this study given the history of misunderstandings around this topic that could directly lead to misinterpretation of the findings in this paper by one of our crucial target audiences: the archaeological community. We have therefore retained the approach of starting with this paragraph in our revised manuscript.

Line 330 "The only plausible explanation for these changes in EEF ancestry is new migrations from mainland Europe into southern Britain" - seems plausible, but why is this not demonstrable using the model?(is it? is that what Extended Data Fig 3 is showing?) The evidence for this claim needs to be presented clearly. Your data doesn't cover Ireland (much) or Scandinavia at all. Are these ruled out due to the direction of the change? Could the persistence of people with different burial practice (i.e. cremation) with overlapping geography also be ruled out? (missing full stop)

We agree this statement was too strong as we have not adequately surveyed places like Ireland in this period. We have changed the statement to:

“The most plausible explanation is migration from outside Britain into southern Britain in the M-LBA.”

Line 373: "In either case, the Archer's anomalously low Steppe ancestry is important in revealing that Beaker-using people who came to Britain were not socially stratified in such a way that Steppe lineage from eastern Europe necessarily conferred the highest social status." This is a

confusing statement, that needs clarification. Its just one person with high status from the Alps (?). Is the companion also isotopically outside of the UK? What data goes into this claim?

We have clarified this passage as follows:

“The ‘Companion’ (I2565), a burial found next to the Archer whose isotopic profile like most others at the site is consistent with a local upbringing, was not an ancestry outlier for C/EBA Britain ($32.7 \pm 3.0\%$ EEF; $P=0.47$ for consistency with the period average; Figure 3). The Archer and the Companion shared a rare tarsal morphology and similar grave goods that have been hypothesized to reflect a close genetic relationship (Supplementary Information section 4)²⁸, but our results rule out first- or second-degree genetic relatedness.”

MMN is defined but never used and we learn nothing about them except the age. Either explain the context (which I think could be helpful), or just leave off the end of this sentence.

In the revised manuscript, te no longer refer explicitly to the ‘Manche-Mer du Nord complex’ (MMN).

Minor things:

Is the library prep (line 272) all standard? if so better to say less here and more in the methods? ("almost always pretreated" doesn't seem a useful statement...?)

The uracil-DNA glycosylase (UDG) pretreatment we carried out greatly drives down the error rate in the data and many papers publishing ancient DNA data do not use this treatment, so we feel it is important to mention in the text of this paper that this pretreatment was performed.

We reviewed the list of libraries we newly report in the paper, and in fact 100% of them are UDG-pretreated. The sentence in question is therefore simplified to:

“To generate these data, we prepared powder from bones and teeth, extracted DNA¹⁴⁻¹⁶, and generated 1033 sequencing libraries all pretreated with the enzyme uracil-DNA glycosylase to reduce characteristic cytosine-to-thymine errors of ancient DNA (Online Table 2)^{14,15,17}.”

Line 316: pretty boring to list the Scotland percentages present in Table 1 and Figure 2 anyway.

We no longer list these percentages and just refer readers to Figure 2 and Extended Data Table 1.

Extended data fig 5: state what the horizontal lines are.

The red horizontal lines give the present-day frequency; we now specify this in the figure legend.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

The authors have amended this manuscript and responded to my queries. My previous positive comments stand and I think the work is improved and acceptable for publication.

In particular I mentioned three issues that also resonated with comments by reviewers 2 and 3.

First, regarding the nature of language in Scotland in the first Millennium - comments by reviewer 2 and the rebuttal allay my concern.

Second, I queried if there was some circularity in assigning samples to archaeological bins and then making archaeological inference from these. Fig 3 now has a smoothed temporal ancestry plot and readers may judge for themselves the nature of variation in ancestry with time. This central figure also has a % outliers plot added which clarifies the relative proportions in the Iron Age. My opinion is that the agnostic display of ancestry through time suffices to guard against preselected models and that the progression of logic through the paper is clear.

Third, I mentioned and welcomed the discussion of migration as a contested issue in archaeogenomics. This discussion has been amended - I think improved. It has a clear distinction of the use of the word in the two disciplines and outlines how it used in this manuscript. I agree with the authors that this is a useful term and should be used in this context.

Extended figures:

I particularly like ext. Data fig 2 - even though it is somewhat repetitive of fig 3. It has a very clear illustration of the cliff edge of ancestry change at the end of the Neolithic.

In the earlier version there was an interesting ext. figure (1) of an extended family uncovered in Hazelton North. I think this is useful for wider discussions of patrilocality and I wonder why it is now removed.

Signed: Dan Bradley

Referee #2 (Remarks to the Author):

Summary of the key results

This is accurate.

Originality and significance: Yes

Data & methodology: validity of approach, quality of data, quality of presentation. All OK in the area of my expertise.

Appropriate use of statistics and treatment of uncertainties. OK so far as I can judge.

Conclusions: robustness, validity, reliability. OK in my opinion.

Suggested improvements: experiments, data for possible revision.

None needed at this stage. Needless to say, more data from more geographical areas would (as always) be desirable, as the authors say, but there is more than plenty of new data here already.

References: appropriate credit to previous work?

Yes, appropriate.

Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions

All fine.

Referee #3 made no comments to the authors.

Referee #4 (Remarks to the Author):

The authors have addressed issues raised in the first round appropriately and the paper is now much smoother and easier to follow. It is generally carefully written, with minor exceptions below, providing careful context for all other claims. I am generally satisfied with the ms and it will clearly be an important contribution to the literature.

Minor but vital issue at abstract L21: "a pattern that ****could only**** occur if there were qualitative differences in how dairy products were used in Britain and in Central Europe". This is a strong causal selection claim, but SNP frequencies could have changed for other reasons including migration from a population that engaged in destructive burial (cremation), (very!) low effective population size, or another associated function of the gene. Of course the claim is the most likely but "could only" seems too strong given the possible evidence. (Elsewhere this is appropriately careful.)

Line 162: the concept of "statistically significant" differences in ancestry is based on independent contributions to ancestry, but the data are obtained in specific locations and times which will have correlated ancestry. Testing is therefore not strictly valid, though clearly helpful and indicative. A caveat would resolve this.

A caveat request: aDNA comes from human remains and cannot say anything about remains that were not in the sample, which in theory could lead to arbitrary selection bias if populations have different practice. Cremation is known to vary by culture and time in this period, which as I understand it (as a non-archaeologist) is still documented via e.g. urns, etc so a quantitative approach is in principle possible. Given the very long digressions elsewhere on what is knowable, a short sentence on the data bias would be appropriate.

I do see the value in defining terms, but I stand by my concern that the introduction is too distracted with abstract notions of migration which gets in the way of the point. There are lots of other worthy digressions that would belong equally well here.

Finally, congratulations on another lovely paper!

Author Rebuttals to First Revision:

Referee #1 (Remarks to the Author):

The authors have amended this manuscript and responded to my queries. My previous positive comments stand and I think the work is improved and acceptable for publication.

In particular I mentioned three issues that also resonated with comments by reviewers 2 and 3.

First, regarding the nature of language in Scotland in the first Millennium - comments by reviewer 2

and the rebuttal allay my concern.

Second, I queried if there was some circularity in assigning samples to archaeological bins and then making archaeological inference from these. Fig 3 now has a smoothed temporal ancestry plot and readers may judge for themselves the nature of variation in ancestry with time. This central figure also has a % outliers plot added which clarifies the relative proportions in the Iron Age. My opinion is that the agnostic display of ancestry through time suffices to guard against preselected models and that the progression of logic through the paper is clear.

Third, I mentioned and welcomed the discussion of migration as a contested issue in archaeogenomics. This discussion has been amended - I think improved. It has a clear distinction of the use of the word in the two disciplines and outlines how it used in this manuscript. I agree with the authors that this is a useful term and should be used in this context.

Extended figures:

I particularly like ext. Data fig 2 - even though it is somewhat repetitive of fig 3. It has a very clear illustration of the cliff edge of ancestry change at the end of the Neolithic.

In the earlier version there was an interesting ext. figure (1) of an extended family uncovered in Hazelton North. I think this is useful for wider discussions of patrilocality and I wonder why it is now removed.

We have removed this figure as we decided to pull the Hazelton North individuals out of the present manuscript and to write a separate manuscript entirely about the family structure at Hazelton North to do proper justice to the extraordinary genetic patterns at that site. That manuscript is now also in press at *Nature* and is likely to come out before this manuscript (or possibly at the same time).

Signed: Dan Bradley

Referee #2 (Remarks to the Author):

Summary of the key results

This is accurate.

Originality and significance: Yes

Data & methodology: validity of approach, quality of data, quality of presentation. All OK in the area of my expertise.

Appropriate use of statistics and treatment of uncertainties. OK so far as I can judge.

Conclusions: robustness, validity, reliability. OK in my opinion.

Suggested improvements: experiments, data for possible revision.

None needed at this stage. Needless to say, more data from more geographical areas would (as always) be desirable, as the authors say, but there is more than plenty of new data here already.

References: appropriate credit to previous work?

Yes, appropriate.

Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions

All fine.

Referee #3 made no comments to the authors.

Referee #4 (Remarks to the Author):

The authors have addressed issues raised in the first round appropriately and the paper is now much smoother and easier to follow. It is generally carefully written, with minor exceptions below, providing careful context for all other claims. I am generally satisfied with the ms and it will clearly be an important contribution to the literature.

Minor but vital issue at abstract L21: "a pattern that ****could only**** occur if there were qualitative differences in how dairy products were used in Britain and in Central Europe". This is a strong causal selection claim, but SNP frequencies could have changed for other reasons including migration from a population that engaged in destructive burial (cremation), (very!) low effective population size, or another associated function of the gene. Of course the claim is the most likely but "could only" seems too strong given the possible evidence. (Elsewhere this is appropriately careful.)

[We have rewritten the section in the abstract on lactase persistence](#)

[“Britain’s independent genetic trajectory was also reflected in the rise of the allele conferring lactase persistence to ~50% frequency by this time compared to ~7% in Central Europe where it rose rapidly a millennium later. This suggests that dairy products were used in qualitatively different ways in Britain and in Central Europe over this period.”](#)

Line 162: the concept of "statistically significant" differences in ancestry is based on independent contributions to ancestry, but the data are obtained in specific locations and times which will have correlated ancestry. Testing is therefore not strictly valid, though clearly helpful and indicative. A caveat would resolve this.

[We think that our current assessments of statistical significance are valid. The Block Jackknife procedure appropriately assesses whether there are statistically significant differences in the ancestry of two samples being compared, taking into account any correlations in their history.](#)

A caveat request: aDNA comes from human remains and cannot say anything about remains that were not in the sample, which in theory could lead to arbitrary selection bias if populations have different practice. Cremation is known to vary by culture and time in this period, which as I understand it (as a non-archaeologist) is still documented via e.g. urns, etc so a quantitative approach is in principle possible. Given the very long digressions elsewhere on what is knowable, a short sentence on the data bias would be appropriate.

We amended the text to add this caveat as follows:

“We considered the possibility that the rise in EEF ancestry in southern Britain was due to a resurgence of archaeologically less visible populations with more ancestry from people living in Britain in the Neolithic, which we missed either due to geographic biases in sampling, or variation across cultural contexts in the way groups treated their dead for example through cremation. However, models of IA people of England and Wales as a mixture of groups in Neolithic and C/EBA Britain failed at high significance (Extended Data Fig. 1). This is due to IA populations in Britain sharing alleles with some Neolithic populations in continental Europe that was not present in early Neolithic or C/EBA groups in Britain (Supplementary Information section 3). The most plausible explanation for these patterns is migration of people carrying this distinctive ancestry into southern Britain in the M-LBA.”

I do see the value in defining terms, but I stand by my concern that the introduction is too distracted with abstract notions of migration which gets in the way of the point. There are lots of other worthy digressions that would belong equally well here.

We have revised our manuscript to remove the first paragraph on migration into a standalone “Box 1” entitled “Reconciling archaeological and genetic understandings of ‘migration’”

Finally, congratulations on another lovely paper!